

Concept Paper

Physicochemical Foundations of Life that Direct Evolution: Chance and Natural Selection are not Evolutionary Driving Forces

Didier Auboeuf

Laboratory of Biology and Modelling of the Cell, Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, 46 Allée d'Italie, Site Jacques Monod, F-69007, Lyon, France

Correspondence: didier.auboeuf@inserm.fr

Received: 22 November 2019; Accepted: 16 January 2020; Published: 21 January 2020

Abstract: The current framework of evolutionary theory postulates that evolution relies on random mutations generating a diversity of phenotypes on which natural selection acts. This framework was established using a top-down approach as it originated from Darwinism, which is based on observations made of complex multicellular organisms and, then, modified to fit a DNA-centric view. In this article, it is argued that based on a bottom-up approach starting from the physicochemical properties of nucleic and amino acid polymers, we should reject the facts that (i) natural selection plays a dominant role in evolution and (ii) the probability of mutations is independent of the generated phenotype. It is shown that the adaptation of a phenotype to an environment does not correspond to organism fitness, but rather corresponds to maintaining the genome stability and integrity. In a stable environment, the phenotype maintains the stability of its originating genome and both (genome and phenotype) are reproduced identically. In an unstable environment (i.e., corresponding to variations in physicochemical parameters above a physiological range), the phenotype no longer maintains the stability of its originating genome, but instead influences its variations. Indeed, environment- and cellular-dependent physicochemical parameters define the probability of mutations in terms of frequency, nature, and location in a genome. Evolution is non-deterministic because it relies on probabilistic physicochemical rules, and evolution is driven by a bidirectional interplay between genome and phenotype in which the phenotype ensures the stability of its originating genome in a cellular and environmental physicochemical parameter-depending manner.

Keywords: Evolution; Darwinism; biophysics; RNA; Origin of life

1. Introduction

The current framework of evolutionary theory that comes from the modern synthetic theory of evolution postulates that evolution relies on random mutations, generating a diversity of phenotypes on which natural selection acts. This concept is widely accepted in the scientific community despite the fact that some important issues concerning it have been raised. [1–3]. The notion of random mutations can lead to multiple interpretations. It can mean that the nature or the location of mutations is random, and indeed, mutations are often described as “errors” during replication [4,5]. However, many factors influence the rate, nature, and location of mutations [5–10]. Thus, the appearance of a mutation is probabilistic and depends on multiple cellular- and environment-dependent physicochemical parameters. The notion of random mutations can also mean that the probability of a mutation is independent of the phenotype it generates. One of the objectives of this article is to show that if cellular- and environment-dependent physicochemical parameters influence the frequency, nature, and location of mutations, the probability of a mutation should depend on the

phenotype it generates. Indeed, a continuum between physiological and genetic adaptation is shown, as already has been proposed in [11–13], which indicates that physiological adaptation facilitates and guides genetic variations in an environment-depending manner (see below).

The notion of natural selection is also subject to multiple interpretations because it can be negative, positive, or neutral [14–17]. How can evolution be explained, if chance generates a large number of possibilities, and natural selection can be positive, negative, or neutral? In addition, the notion of natural selection is of limited interest, because a living organism observable over more or less long time periods is necessarily adapted to its environment; otherwise, it disappears without leaving any descendants. The second objective of this article is to replace the notion of natural selection with the notion that the role of the phenotype in evolution is to maintain the physicochemical integrity and stability of its originating genome. This establishes feedforward and feedback loops between the genome and the phenotype, i.e., the genome generates a phenotype that exerts a feedback loop by either maintaining genome stability or guiding genome variations depending on its efficiency to relax environment-dependent physicochemical constraints.

Another issue raised by the current model of evolution relies on the fact that the combination of chance and natural selection does not provide a single conceptual framework that simultaneously explains both evolution and organism activities, since evolution, but not organism activities, would rely on chance and natural selection. However, life and evolution are inseparable and must depend on the same fundamental principle. Indeed, living organisms are hierarchically structured since multicellular organisms are composed of cells that are composed of molecules. The lower levels of organization have necessarily appeared before more complex living forms and fundamental principles of evolution must be applicable from molecules to complex organisms. In this setting, the current model of evolution was established using a "top-down" approach originating from Darwinism that is based on observations made of complex multicellular organisms [18] which was modified after the discovery of DNA. In other words, there was no other choice than to define "chance" and "natural selection" as evolutionary driving forces in order to explain the emergence of complex phenotypes of multicellular organisms prior to understanding their molecular origin or the underlying physicochemical mechanisms. However, to uncover evolutionary driving forces, one needs to define them from the physicochemical processes at the molecular origin of life. Therefore, in this article, a "bottom-up" approach is used according to the facts that (i) life started with the emergence of nucleic and amino acid polymers before the emergence of more complex forms of life and (ii) physicochemical laws are the foundations of cellular processes and complex organism activities. A "bottom-up" approach redefines (i) the notion of chance in a precise context of physicochemical laws and (ii) the notion of adapted phenotype not in terms of organism fitness but rather in terms of the impact on genome stability.

2. Overview

Life is based on two types of polymers, nucleic acid polymers (DNA and RNAs) and amino acid polymers (proteins). Over the last decades, the emphasis has been on a functional dichotomy in which nucleic acids are described as supporting genetic information, while proteins perform the cellular activities. As a consequence, nucleic acids are often "simply" considered as the carrier (DNA) or vectors (RNAs) of genetic information and are represented in the form of a suite of letters A, C, G, and T, corresponding to the four nucleotides that composed them. Thus, the physicochemical properties of nucleic acids are obscured in the context of the current synthetic theory of evolution, in which evolution corresponds basically to the random substitutions of one letter by another one. In Section 3, it is underlined that the physicochemical properties of nucleic and amino acid polymers depend on their composition, which is constrained by cellular- and environment-dependent physicochemical parameters. Next, it is shown that this principle has consequences for the way evolution proceeds.

Indeed, in Section 4, it is highlighted that the emergence of life probably corresponds to the establishment of the interdependency between RNAs (or similar molecules) and proteins (or similar molecules). This means that, as observed in modern organisms, protein synthesis depends on RNAs

as templates and RNA synthesis depends on proteins. This interdependency generates a positive feedback loop, as an RNA (the proto-genome) generates a protein (the proto-phenotype) that contributes to the synthesis of the RNA on which it depends (Figure 1A). Moreover, RNAs are unstable and are rapidly degraded by hydrolysis. Therefore, the RNA/protein system can only self-reproduce and self-amplify if proteins (the proto-phenotypes) maintain the physical integrity of the template (the proto-genome) from which they are derived. This fundamental principle can be expressed as follows: The genome (in this case, an RNA) generates a phenotype (here, a protein) that contributes to the reproduction and stability of its originating genome by relaxing environment-dependent physicochemical constraints (Figure 1A). In this model, the stability of RNAs (the genome) and proteins (the phenotype) is interdependent, since by maintaining the stability of RNAs, proteins maintain their own sequence over time. In other words, a phenotype can only be reproduced across generations in a given environment if it maintains the stability of its originating genome.

In Section 5, it is shown that this principle captures the mechanisms by which cellular- and environment-dependent physicochemical parameters direct evolution. Indeed, it is shown that specific cellular- and environment-dependent physicochemical parameters exert constraints on some nucleic and amino acid polymers, which trigger a specific cellular response through the regulation of the expression of a selected set of genes. If the resulting cellular activities (the phenotype) relax the initiating constraints (i.e., return to equilibrium), this corresponds to physiological adaptation (Figure 1B, Path 1). Otherwise, it is shown that physicochemical constraints persist and challenge the physical integrity of nucleic and amino acid polymers inducing more-or-less direct mutations in the corresponding genome locations. This process, corresponding to genetic adaptation, only stops when new sequences directly or indirectly relax the initiating constraints (Figure 1B, Path 2). Therefore, cell activity (or physiological adaptation) and evolution (genetic adaptation) are based on the same principle, that is, the phenotype is derived from a genome in response to physicochemical constraints and relaxes the initial constraints. If the phenotype is adapted to a given environment (i.e., a set of physicochemical constraints), the genome is stable and will be reproduced “identically” (see Discussion). If the phenotype is unsuitable, the genome is unstable at specific locations (when directly or indirectly challenged by environmental fluctuations) and will be modified until a phenotype is generated that relaxes the initial constraints (i.e., that ensures the stability of the challenged genomic locations).

If the main role of a phenotype is to maintain the stability (see Glossary) of its originating genome, the phenotype (as the sum of all cellular activities) in turn creates physicochemical constraints on the genome. For example, the metabolic activities of cells produce a diversity of molecules (e.g., reactive oxygen species) that can interact with DNA and induce mutations. This means that a genome can generate a phenotype that can, in turn, generate physicochemical constraints on its originating genome (Figure 1C). In Section 6, this feedback loop is illustrated by showing that UV-radiation triggered the emergence of photosynthesis, which then triggered the emergence of cell respiration followed by eukaryogenesis. In addition, it is shown that this principle helps to explain the emergence of epigenetic modifications, multicellular organisms, and germline cells. Finally, it is highlighted that the interplay between the genome and the phenotype described above at the RNA/protein level is still operating in multicellular organisms when considering that the phenotype generated from the germline cell DNA corresponds to the production of somatic cells whose function is to protect the originating genome (i.e., the germline cell DNA) from environment-dependent physicochemical constraints.

It is concluded that the activities and evolution of living organisms are governed by the same physicochemical rule which is in disagreement with the current framework of evolutionary theory. Indeed, according to the current framework there is no direct relationship between physiological and genetic adaptation since physiological adaptation is based on physicochemical principles of homeostasis, whereas genetic adaptation would be fueled by random mutations (Figure 1D, left panel). In contrast, in the model proposed in this article, genetic adaptation is the consequence of physiological adaptation that takes place as long as fluctuations in environment-dependent physicochemical parameters do not exceed a physiological range. Above this physiological threshold,

the integrity of nucleic and amino acid polymers, in particular DNA, is challenged leading to targeted mutations until the emergence of a phenotype that maintains the integrity of its originating DNA with regards to environmental constraints (Figure 1D, right panel).

It must be emphasized that rejecting the concepts of “random mutations” and “natural selection” as evolutionary “driving forces” does not at all exclude the possibility that these processes contribute to evolution, for the simple reasons that (i) all physicochemical (and therefore, biological) processes are stochastic; thus, mutational processes are probabilistic rather than deterministic and (ii) a living organism is necessarily adapted to its environment; therefore, natural selection constantly operates as a filter. In this light, the aim of this manuscript is not to define evolutionary driving forces that would explain all the diversity of living organisms. Instead, the evolutionary physicochemical driving forces, depicted here, should generate the fundamental shape of biological objects whose diversity could well depend on other phenomena as well, such as plasticity and natural selection. The depicted evolutionary driving forces allow us to describe evolution of living beings similarly to geological driving forces, in the way that plate tectonics describe earth evolution. Indeed, plate tectonics corresponding to physicochemical processes in the deep and superficial layers of the earth is a driving force explaining the formation of mountains and plateaus, but it does not explain the diversity of landscapes resulting from multiple contingent phenomena (e.g., wind, rain). This notion is further discussed in Section 6.

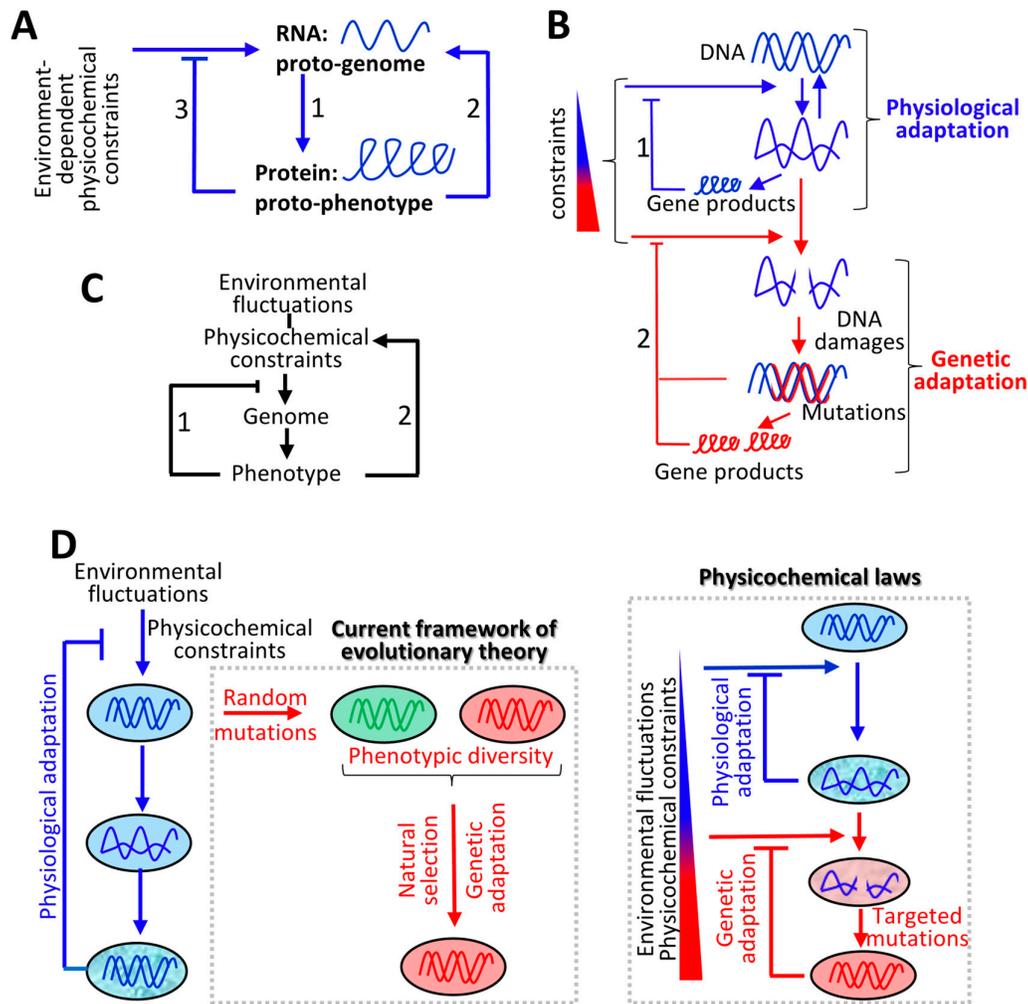


Figure 1. (A) The origin of life probably relies on simpler forms of organization than those observed in modern living organisms. The commonly accepted hypothesis postulates that the origin of life corresponds to the emergence of polymers, such as RNAs and proteins. These two polymers are interdependent as RNAs serve as templates for protein synthesis (1) and proteins are necessary for RNA synthesis (2). This interdependence, which can be represented in the form of feedforward and feedback loops between the proto-genome (RNAs) and the proto-phenotype (proteins) can only be maintained if proteins relax environment-dependent physicochemical constraints triggering for example RNA degradation (3). This interdependence is the foundation of life and evolution. (B) Variations in the extracellular environment induce constraints on cellular polymers sensitive to these variations, triggering the cellular response that results in the biogenesis of gene products, whose activity relaxes the initial constraints (1). However, if these variations exceed a certain amplitude or persist over time, they challenge the integrity of the targeted polymers, which ultimately lead to mutations. This process stops only when new sequences (directly or indirectly) relax the initial constraints (2). (C) A DNA molecule is subjected to environmental fluctuations of physicochemical parameters, which triggers the biogenesis of polymers (gene products) whose activities correspond to the phenotype. Cellular activities (the phenotype) allow a return to equilibrium by relaxing the initial constraints (1). Nevertheless, these activities also generate constraints directly or indirectly on their originating genome, meaning that a genome is adapted to the constraints generated by its own activities (2). (D) According to the current framework of evolutionary theory (left panel), there is no direct relationship between physiological and genetic adaptation because physiological adaptation is based on physicochemical principles of homeostasis as a function of environmental fluctuations, while genetic adaptation would be fueled by random mutations generating a diversity of phenotypes on which natural selection acts. In contrast, in the model proposed in this article (right panel), genetic adaptation is the consequence of physiological adaptation. Indeed, physiological adaptation can take place as long as fluctuations in environment-dependent physicochemical parameters do not exceed a certain threshold. Above this physiological threshold, the integrity of nucleic and amino acid polymers, in particular DNA, is challenged which leads to targeted mutations. This mutational process stops when the mutations generate a phenotype that maintains the integrity of the DNA with regard to environmental constraints (genetic adaptation).

3. Environment-Dependent Physicochemical Constraints on Nucleic and Amino Acid Polymer Composition

Nucleic and amino acid polymers have emergent physicochemical properties (e.g., solubility, folding, and stability) that depend on intrinsic parameters relying on their composition and extrinsic parameters (e.g., temperature). The aim of this part is to show that cellular- and environment-dependent physicochemical parameters constrain the composition of cellular nucleic and amino acid polymers. In the next part, it is shown that this principle has consequences on the way evolution proceeds.

3.1. Physicochemical Constraints on Protein Composition

The amino acid composition of proteins is constrained by intrinsic parameters impacting on protein solubility, folding, and stability, for example, proteins with too many hydrophilic amino acids tend to unfold, whereas proteins with too many hydrophobic amino acids tend to aggregate [19,20]. Extrinsic physicochemical parameters, such as temperature or cellular and environment chemical composition, also constrain the amino acid composition of proteins, notably owing to the chemical modifications of amino acid side chains. Indeed, protein amino acids can react with and be modified by various chemical compounds, for example, lipids, sugars, and reactive oxygen species (ROS) [21]. Protein chemical modifications (see Glossary) (i) are spontaneously or enzymatically generated, (ii) change the physicochemical properties of the modified amino acids, and (iii) contribute to cellular regulatory processes or induce protein mis-folding and aggregation [21]. For example, amino acid oxidation in proteins plays a role in many cellular processes, yet high rates of it induces protein damage and aggregation [21]. In agreement with the fact that biochemical compounds constrain the amino acid composition of proteins, organisms growing under different levels of salt or oxygen, or

with different metabolic activities, produce proteins with different amino acid composition biases [20,22] (see Glossary). Finally, the amino acid composition of proteins depends on the environmental supply of key elements that are required for the biogenesis of amino acids. For example, organisms growing in nitrogen- or sulphur-poor environments produce proteins that contain low amounts of nitrogen-rich amino acids (e.g., arginine) or sulphur-containing amino acids (e.g., cysteine), respectively [23,24]. In summary, the amino acid composition of proteins obeys physicochemical laws and depends on cellular- and environment-physicochemical parameters.

3.2. Physicochemical Constraints on Nucleic Acid Polymer Composition

RNA and DNA molecules are polymers composed of four monomers or nucleotides that interact with each other when they follow each other in a sequence (i.e., stacking interactions) or when they face each other in two different strands (i.e., base-pairing interactions). These interactions have consequences on the structural and physicochemical properties of nucleic acid polymers. G:C pairs form stronger base-pair interactions than A:T pairs, purine-purine dinucleotides (e.g., GpA) form strong stacking interactions, and GC dinucleotides form polymorphic structures [25–27]. As a consequence, GC- and purine-rich polymers are thermodynamically stable, and the genome of thermophilic organisms is enriched in GC or purine nucleotides [25,28]. Nucleotide composition also determines DNA mechanical properties (e.g., flexibility and bendability), contributing to its cellular functions and its resistance to torsional stresses (Figure 2A, B, see Box 1). For example, increasing the GC content increases the B-DNA-to-Z-DNA conformational transition, DNA bendability, and DNA resistance to torsional stresses; accordingly, highly transcribed genes are GC-rich [26,29–31].

Similar to proteins, DNA and RNA molecules undergo dozens of spontaneous or enzyme-dependent chemical modifications (including oxidation, methylation, deamination, alkylation, and glycation) that (i) affect the physicochemical properties of nucleic acid polymers and (ii) contribute to regulatory processes but can also induce damages [32–34] (Figure 2C). For example, DNA oxidation not only plays a role in gene expression regulation but also increases DNA damages and mutations [32,33]. The reason why DNA chemical modifications can generate mutations during replication is because a chemically modified nucleotide can “mimic” another nucleotide. For instance, an oxidized guanine can base pair with an adenine rather than a cytosine, which can result in mutations during replication [32,33]. As for proteins, environmental physicochemical parameters (temperature and chemical composition) constrain the nucleotide composition of genomes and this can be observed in thermophilic, halophilic, acidophilic, aerobic, and radiation-exposed organisms [28,35]. Finally, the amount of DNA per cell (genome size and ploidy) is constrained by the environmental availability of phosphorus, and the genomes of organisms growing in a nitrogen-poor environment are enriched in A:T pairs, which require seven nitrogens instead of the eight used in G:C pairs [36–38].

In summary, the properties of nucleic and amino acid polymers depend on intrinsic parameters, and these polymers can undergo reversible structural and chemical modifications triggered by extrinsic physicochemical constraints. Above a certain threshold, these modifications can challenge their integrity (Figure 2A–D) and raises a major issue for coding sequences, which are under both nucleic acid-related and protein-related constraints (Figure 2E).

3.3. Interdependency between the Physicochemical Properties of Nucleic Acid Polymers and their Cognate Amino Acid Polymers

Coding sequences accommodate different constraints, by not only the encoded protein sequence but also cellular processes such as chromatin organization (e.g., nucleosome positioning), transcription (e.g., DNA flexibility), RNA folding, splicing, and RNA–RNA or RNA–protein interactions [39–41]. It has been assumed until now that variations of the third nucleotide of codons (the “wobble” position, which is usually the only variable between all codons encoding an amino acid) solves nucleic acid-related constraints without affecting the encoded amino acids [39]. However, this assumption is challenged by the diversity of the constraints described above, and by the fact that the third nucleotide of codons changes the thermodynamic property of codon–anticodon

interactions, with consequences on translation fidelity, speed, and cotranslational protein folding [39,42]. Accommodation of different constraints in coding sequences also relies on the fact that the genetic code is not randomly organized, as amino acids that share physicochemical properties (e.g., hydrophathy) correspond to codons with a similar nucleotide composition bias. For example, hydrophilic or hydrophobic amino acids correspond to A- or T-rich codons, respectively, and small or large amino acids are encoded by GC-rich or GC-poor codons, respectively [43,44]. Consequently, two codons with only one nucleotide difference (either at the first or third position) can encode either the same amino acid or different amino acids with similar physicochemical properties.

However, the organization of the genetic code implies that the nucleotide composition bias of a nucleic acid polymer affects the physicochemical properties of the encoded protein and, conversely, that the amino acid composition bias in a protein affects the physicochemical properties of the cognate nucleic acid polymer. For example, compact proteins comprising small amino acids correspond to GC-rich coding sequences (as small amino acids correspond to GC-rich codons), while proteins comprising hydrophobic regions (i.e., containing stretches of hydrophobic amino acids) correspond to T-rich coding sequences (as hydrophobic amino acids correspond to T-rich codons) [30,45,46]. Along the same lines, the global nucleotide composition bias of a genome (e.g., GC content) is associated with a global amino acid composition bias of the encoded proteome [47,48].

Supporting the notion that constraints on the physicochemical properties of one kind of polymer affects composition biases of the cognate polymers are the following observations: (i) Nucleosome positioning leaves a footprint in protein sequences, (ii) splicing sites and splicing factor binding motifs constrain the amino acid composition of peptides encoded by splicing-regulated exons, and (iii) mRNA secondary structures depending on base complementarity have consequences on the secondary structures of the encoded protein [40,49–51]. Conversely, protein secondary structures leave a footprint in the nucleotide composition bias of coding sequences, for example, amino acids that favor alpha-helices and beta-sheets correspond to codons ending with purines and pyrimidines, respectively [52]. Along the same line, alternations of hydrophobic and hydrophilic amino acids in amphipathic alpha-helices that rely on a periodicity of ~3.5 amino acids correspond to a specific detectable ~10 bp periodicity in DNA, with consequences for the helical pitch of nucleosome-wrapped DNA [53]. Finally, purine enrichment in coding sequences is determined by protein-related physicochemical constraints, such as solubility and folding [54].

In summary, the common use of letters to symbolize biopolymers (such as DNA) obscures their physical nature which implies that their composition is constrained by environmental and cellular physicochemical parameters. This raises concerns regarding coding sequences that are constrained directly and indirectly by nucleic acid-related and protein-related parameters. While the organization of the genetic code “buffers” these constraints, nucleotide and amino acid composition biases affect each other above a certain threshold. As nucleotide or amino acid composition biases determine the physicochemical properties of polymers, the composition-dependent physicochemical properties of these interdependent polymers must be adapted to the same fundamental physicochemical constraints (Figure 2E), as is shown in the next Section.

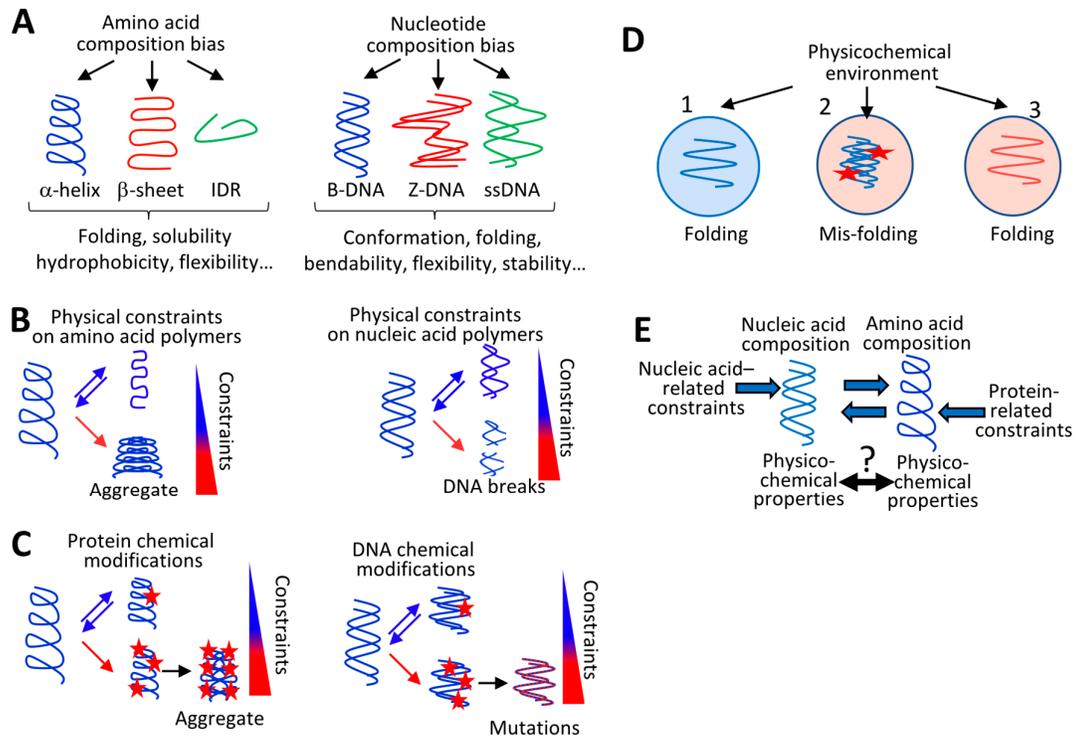


Figure 2. (A) The amino acid or nucleotide composition of proteins (left) or DNA (right), respectively, determines the physicochemical properties of these polymers with consequences on their physical and chemical properties. IDR, intrinsically disordered region and ssDNA, single-stranded DNA. (B) The composition of a polymer determines its physicochemical properties, and therefore its folding and physical resistance to specific constraints. Depending on the composition of a given polymer, some physicochemical constraints induce reversible structural changes (blue arrows), and others induce irreversible damages (e.g., aggregation and breaks) (red arrows). (C) Chemical modifications of amino acids (left) or nucleotides (right) change the physicochemical properties of polymers. These chemical modifications are reversible (blue arrows) or induce irreversible damages (red arrows). (D) Any given polymer (blue) is stable in a physicochemical environment and unstable in another one (1 vs. 2). A different polymer (red) reacts differently under the same constraints (3 vs. 2). (E) Composition determines the physicochemical properties of nucleic or amino acid polymers. Nucleotide or amino acid composition is constrained by physicochemical parameters, which suggests that their composition must correspond to the same fundamental physicochemical parameters as the sequence of nucleic acid polymers determines the composition of proteins.

4. Molecular Origin of Life and Evolution of the Genetic Code: Defining Evolutionary Driving Forces.

Life relies on the interdependency between nucleic and amino acid polymers, since the biogenesis of proteins requires a nucleic acid polymer as a template (RNA), whereas biogenesis of nucleic acid polymers requires proteins. In addition, nucleic and amino acid polymers are in “competition” with each other, as their biogenesis requires the same elements (e.g., nitrogen) and requires the same template, i.e., single-stranded RNA (ssRNA) before the emergence of DNA. The aim of this section is to propose that these fundamental principles (interdependency and competition) constrain the genetic code evolution to match the fundamental physicochemical properties of both polymers. In other words, specific codons correspond to specific amino acids because their presence in nucleic acids and cognate proteins allows both polymers to deal with the same fundamental physicochemical parameters. This Part (i) defines the physicochemical forces driving evolution from

the molecular origin of life and (ii) describes the bidirectional interplay in terms of stability between genomes and phenotypes.

4.1. Molecular Origin of Life: Interdependency between RNAs and Proteins

While there is an ongoing debate about whether the origin of life started with only RNAs (“RNA world”) or with RNAs and peptides (“RNP world”), there is a consensus that the activity of ribozymes (e.g., RNAs that catalyze nucleic acid polymerization) has been enhanced at some points of evolution by their interactions with amino acids or with randomly generated small peptides that could have also stabilized RNAs, which would otherwise be rapidly degraded by hydrolysis (Figure 3A) [55–57]. For example, randomly generated peptides composed of abiogenetically produced amino acids, such as Gly and Asp, can increase the efficiency of replicating ribozymes, as these amino acids play a very important role in the AsnAlaAspPheAspGlyAsp (NADPDGD) peptides found in all polymerases. In particular, binding of these amino acids to the catalytic metal ion Mg^{2+} could have enhanced polymerization and protected RNA from Mg^{2+} -dependent hydrolysis [58]. Coincidentally, a primeval genetic code corresponding to GC-rich and RNY codons (e.g., GGC, GCC, GAC, and GTC) has been proposed, because these codons are the most frequent in coding sequences and correspond to the most metabolically simple amino acids (Gly, Asp, Ala, and Val) that are sufficient to produce stable, folded, and functional proteins [56,59]. A positive feedback loop between nucleic and amino acid polymers could have been initiated by primeval GC-rich RNAs, by first interacting with randomly generated peptides (e.g., made of Gly and Asp), which would then favor the production of more complex peptides through a primeval GC-rich genetic code.

However, such a cooperation would be inefficient if RNA and protein polymerization were physically uncoupled, for two main reasons. First, after replication, the ssRNA templates give rise to stable double-stranded RNAs (dsRNAs) which can decrease the rate of other rounds of replication, as well as that of protein synthesis that requires ssRNA templates [60,61]. Secondly, freely diffusible RNAs and proteins limit their cooperation, as freely diffusible peptides generated from an RNA template can stabilize and enhance the replication of potential “parasitic” or mutated RNA replicators [62,63]. Simulation and in vitro selection experiments demonstrate that ribozyme-dependent replication cycles rapidly end when molecules freely diffuse, due to the appearance of mutated RNAs, which become smaller, and therefore are more rapidly amplified while simultaneously losing their enzymatic activity [62,63]. Proto-cell compartmentalization and physical coupling between replication (i.e., RNA production in an RNP world) and amino acid polymerization (i.e., protein production) could solve these two main issues [64–68]. First, interactions between the nascent RNA and the nascent peptide could decrease the formation of stable dsRNAs while protecting ssRNAs from degradation. This is observed in modern prokaryotes or eukaryotes. Protein binding to nascent RNAs during cotranscriptional and translation or co-RNA processing prevents RNAs from interacting with the DNA template and increases transcription [69–72] (Figure 3B). Secondly, the physical proximity between replication and translation could increase the probability that the neo-synthesized proteins “protect” and enhance replication of its originating RNAs, which would have facilitated the increase in replication and translation fidelity [64–68].

While the physical coupling between RNA and protein polymerization seems to be a “sophisticated” molecular process, several authors have proposed straightforward models [65–68,73]. For example, it has been proposed that amino acylated trinucleotides corresponding to proto-tRNAs (tRNA ancestors) composed of three nucleotides (the proto-anticodons) and bound by one amino acid could have been used simultaneously for replication and translation. The three anti-codon nucleotides would have been used as building blocks during replication, while the attached amino acids would have chemically enhanced the triplet polymerization and been used as building blocks for protein production [74,75] (Figure 3C). Of note, (i) the use of tri-nucleotides rather than mono-nucleotides increases the efficiency and fidelity of replication by ribozymes [76]; (ii) phylogeny analyses suggest that tRNAs originated in replication [77]; and (iii) the coupling between transcription (i.e., RNA biogenesis) and translation is still operating in prokaryotes, and many features are shared by transcription and translation in modern cells [69–72] (Box 2).

To summarize, life likely emerged from the cooperation between nucleic and amino acid polymers. The interdependency between RNAs and proteins defines the main evolutionary driving force. A genome (e.g., RNAs) generates a phenotype (i.e., proteins) that protects from environment physicochemical constraints its originating genome that can be replicated (Figure 1A). Therefore, the stability of the genome and the phenotype across generations is interdependent. This principle can help to understand the evolution of the genetic code.

4.2. Evolution of the Genetic Code: Co-Adaptation of Nucleic Acid Polymers and their Encoded Proteins to the Same Fundamental Physicochemical Parameters

In a proto-cell without complex compartments and protein-dependent compensatory mechanisms, nucleic acids and proteins are exposed to the same physicochemical parameters (e.g., chemical compounds). As both polymers depend on each other, their composition and the processes that lead to their biogenesis need to satisfy the same physicochemical constraints. Supporting this model, GC- or purine-composition biases increase the thermostability of nucleic acid polymers and, in turn, correspond to codons of amino acids that increase protein thermostability [28,46,78].

Along the same line, the organization of the genetic code allows nucleic and amino acid polymers to co-adapt to the bioavailability of nitrogen. Indeed, A:T pairs require fewer nitrogen atoms than G:C pairs (7 vs. 8, respectively) and correspond to amino acids that also require fewer nitrogen atoms [23,36,37]. Accordingly, plant genomes and proteomes are AT-rich and contain amino acids that require fewer nitrogen atoms as compared with animal genomes and proteomes. This fits well with the fact that nitrogen sources are limited for plants, while animals can access organic sources of nitrogen [23,36,37]. Of note, in eukaryotic genes containing introns, exons and introns have the same composition bias [79]. Thus, AT-rich genes (N-poor) produce AT-rich mRNAs that code for N-poor proteins as compared with GC-rich genes.

Oxygen is highly toxic, and oxygen derivatives (e.g., ROS) can damage nucleic and amino acid polymers and induce their cleavage or aggregation [21,32,33]. Stepwise increases of oxygen in the biosphere over time (see Section 6 for more details) could have given the impulse for the late incorporation into the genetic code of amino acids that can act as “ROS scavengers” (including Trp, Tyr, Met, Cys, and His), and thereby protect biopolymers from oxidative damages [80]. Interestingly, the most frequent mutations in ROS-producing cancer cells affect Arg codons (and in particular, the CGN codons), with mutations producing codons for Cys (TGY), Trp (TGG), stop codons (TGA), and His (CAR) [81–85]. It has been proposed that these mutations (i) are induced by ROS-mediated deamination of (methyl) cytosine, leading to C > T mutations, or incorporation of oxidized guanine during replication, leading to G > A mutations and (ii) protect cancer cells from the high levels of ROS by increasing the global antioxidant capacity of the cancer proteome [81–85]. Interestingly, CGN codons (that encode Arg) seem to be particularly sensitive to oxidation because of their physicochemical properties [81,86]. Furthermore, (methyl)cytosine deamination produces CG > TA mutations, and thereby can increase the nitrogen availability required by proliferative cancer cells, as it reduces the nitrogen-richer C:G pairs in favor of the nitrogen-poorer T:A pairs, as well as the GC-rich sequences that contain codons (e.g., CGN) corresponding to nitrogen-rich amino acids (e.g., Arg) [81]. Therefore, ROS-induced cytosine deamination would simultaneously save nitrogen atoms (at both genome and proteome levels) and protect cells from ROS. Although speculative, one possibility is that the assignment of “ROS scavenger” amino acids to codons that originate from the oxidation of Arg codons generates antioxidant proteins from oxidized nucleic acids. Supporting such a possibility, assignment of Met to the ATA codon (in addition to the ATG codon) in most animal mitochondria lineages explains the high frequency of Met in mitochondria-encoded respiratory chain complexes, and it represents an adaption to high ROS level in mitochondria [87].

If physicochemical constraints shaped the genetic code, it is very likely that the universal genetic code is the result of horizontal transfers of “code fragments”, as proposed by several authors [57,88]. For example, protocells that are present in extreme environments (e.g., hot or cold, nitrogen rich or poor, oxygen rich or poor) could have developed “code fragments” adapted to their specific environment. At the frontiers of these habitats where physicochemical parameters fluctuate, genetic

horizontal transfers between cells using different “code fragments” could have led to the emergence of the modern genetic code (Figure 3D).

To summarize, the organization of genetic code has been constrained to match general physicochemical properties of the interdependent nucleic and amino acid polymers. This means that RNAs and proteins could only coevolve if RNAs and their cognate proteins are adapted to the same physicochemical constraints (temperature, N availability, etc.). Moreover, cooperation between RNAs and proteins must have required that proteins preferentially interact with their encoding RNAs in order to (i) protect their cognate RNA from damage and (ii) limit protein diffusion, and thus limit their use by parasitic RNAs (see above for description of the coupling between nucleic and amino acid polymerization). Supporting this possibility, RNA binding proteins bind preferentially to mRNA coding sequences that have similar nucleotide composition biases as their own encoding mRNAs [89–91]. In other words, and as an extension of the stereochemical hypothesis, the genetic code could have been shaped over evolutionary time to allow the interactions between proteins and their cognate mRNAs, making their cooperation possible by protecting and ensuring the stability of each other (Box 2). Somehow, self-assembly of proteins with their encoding RNA in viral capsids could reflect this primitive feature and function of proteins (i.e., proto-phenotype) to protect their encoding RNAs (i.e., proto-genome) [92].

The intimate cooperation and interdependency between nucleic and amino acid polymers imply that the genetic code has been shaped over evolutionary time to ensure that nucleic acid polymers and their cognate proteins share more physicochemical and structural properties than previously anticipated (Figures 2E and 3E). In Section 5, the consequences of these sharing properties are shown in terms of evolution, after having described the interplay between the cell metabolism and nucleic and amino acid polymers.

4.3. Feedforward and Feedback Loops between Gene Products and their Products (i.e., Metabolites)

Nucleic and amino acid polymers depend on the availability of molecules containing elements, such as nitrogen (see Section 3). This dependency would have favored the emergence of polymers with metabolic activities that modify environmental resources and allow elements to be incorporated into amino acids and nucleotides. In an RNP world, the proto-cell phenotype would be a positive feedback loop between biosynthetic pathways (i.e., metabolism in modern cells) and polymerization of RNAs and proteins (i.e., the gene expression process in modern cells). Polymer production depends on biosynthetic pathways that in turn depend on polymerization products (Figure 3F). Supporting this interplay between gene expression and biosynthetic pathways, the genetic code likely coevolved with amino acid biosynthetic pathways, as codons starting with A, C, U, or G correspond to amino acids synthesized from oxaloacetate, alpha-ketoglutarate, pyruvate, or from the reductive amination alpha-keto acid, respectively [56,93–95]. A possibility is that the simple amino acids or metabolites that covalently attached to polynucleotides (e.g., proto-tRNAs made of three nucleotides) could have been chemically transformed to give rise to more complex amino acids [95]. Another possibility is that chemical modifications of simple amino acids occurred after their incorporation into proteins, and that chemically modified amino acids (i.e., complex amino acids) were, thus, available after protein hydrolysis. Then, the complex amino acids would have been incorporated into the genetic code.

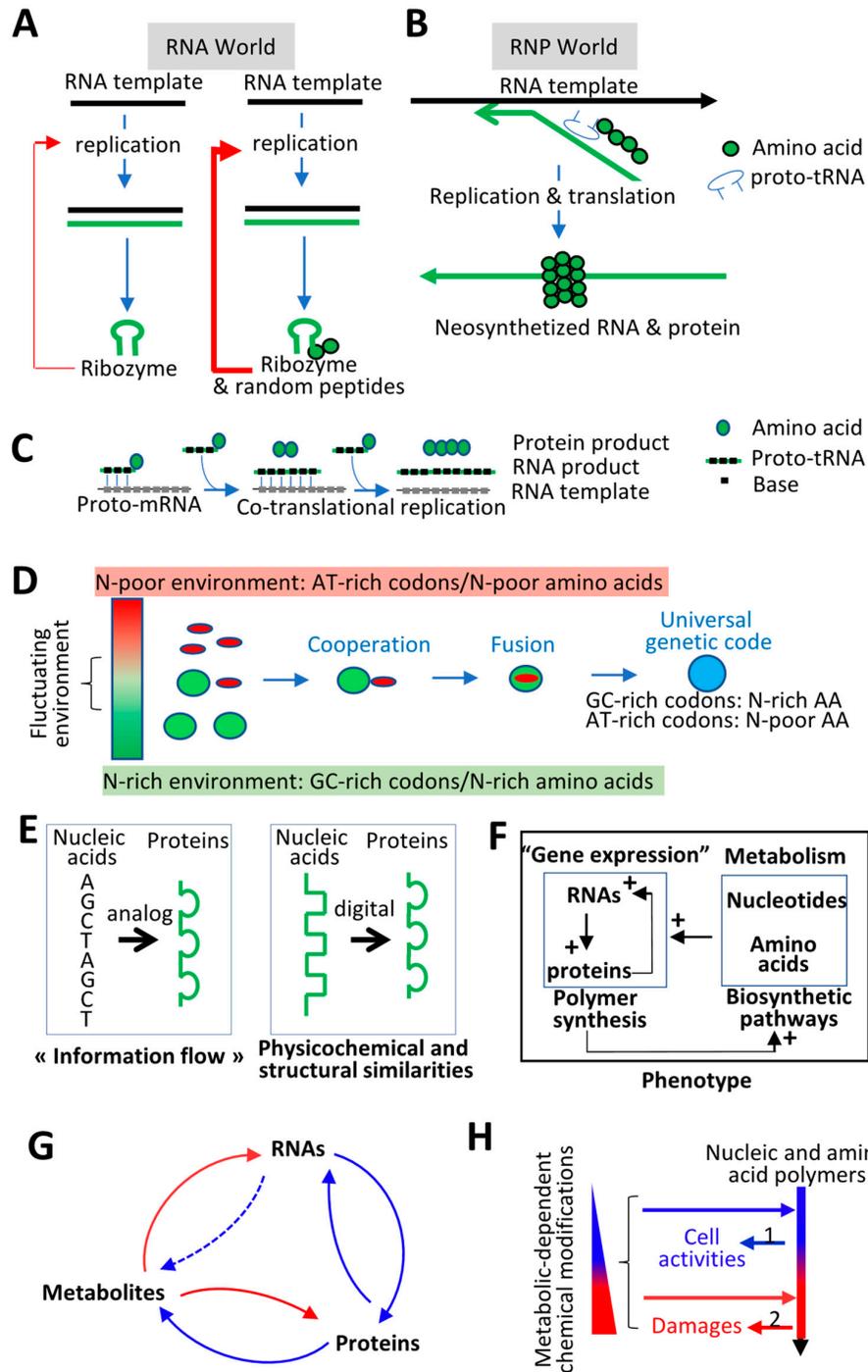


Figure 3. (A) In an RNA world, an RNA molecule is replicated thanks to the product of replication (e.g., ribozyme, left panel). This process is enhanced by amino acids or small peptides (right panel). (B) In an RNP world, replication (RNA production) and translation (protein production) could have been performed simultaneously as in modern prokaryotes, thereby avoiding the nascent RNA to interact back to the template and increasing the probability that the neo-synthesized protein interacts with the RNA replication product. (C) Cotranslational replication in an RNP world could have been performed owing to amino acids attached to proto-tRNAs that enhanced the polymerization of proto-

tRNAs and that were simultaneously incorporated into the nascent protein. (D) Two different proto-cells (red and green circles) growing in different chemical environments (e.g., N-poor vs. N-rich environment) could have developed different proto-genetic codes. Their cooperation in fluctuating environments could have led to horizontal transfer, leading to the emergence of the universal genetic code. (E) Nucleic acid polymers are often represented as linear strings of letters that are translated into proteins with physicochemical properties unrelated to those of nucleic acid polymers. If the genetic code has been constrained over evolutionary time to match nucleic acid polymers and their cognate amino acid polymers to the same fundamental physicochemical constraints (e.g., temperature, element availability), nucleic and amino acid polymers share more physicochemical properties than previously anticipated. (F) In an RNP world, cooperation between interdependent polymers (i.e., RNAs and proteins) relies on their activities toward the biogenesis of nucleotides and amino acids necessary for their synthesis. The phenotype of a proto cell in an RNP world corresponds to the polymerization of nucleotides and amino acids. The polymerization products produce nucleotides and amino acids. (G) RNAs give rise to proteins, which allow the biogenesis of RNAs and metabolites through transformation of molecules captured from the environment (blue lines). RNAs probably played a role, at least early in evolution, in metabolite biogenesis (blue broken lines). Metabolites are required, in turn, to give rise to RNAs and proteins (red lines). (H) Metabolic-dependent chemical modifications of nucleic and amino acid polymers contribute to the cell activities and to the cellular physiological adaptation in response to environment-dependent constraints (1). However, metabolic-dependent chemical modifications can also induce irreversible damages of nucleic and amino acid polymers (2).

In this context, it must be underscored that the chemical modifications of proteins and nucleic acids establish a direct bridge between cell metabolic activities and gene expression and gene product functions (i.e., the cellular physiological adaptation). Indeed, chemical modifications of nucleic and amino acid polymers (e.g., post-translational modifications, DNA, or RNA methylation) that change their physicochemical properties also affect their cellular activities, and these chemical modifications depend on the cell metabolic activities. For example, DNA, RNA, and protein oxidation depend on cellular oxidative metabolism, while DNA, RNA, and protein methylation depend on the production of *S*-adenosyl methionine (SAM), the universal methyl-group donor produced in the one-carbon cycle [96–98].

In summary, the genetic code could only coevolve with metabolic pathways that provide amino acids required for protein synthesis. The necessity of this coevolutionary process comes from the fact that a cell's autonomy is based on the biogenesis of biopolymers (RNAs and proteins), which allows the transformation of molecules from the environment into metabolites (nucleotides and amino acids) that are required for biopolymer synthesis (Figure 3F, G). This concept corresponds to the notion of autopoiesis proposed by H. Maturana and F. Varela [99,100]. Indeed, autopoiesis was defined as the property of a system to produce itself, and therefore to be autonomous by maintaining its organization despite any change in its components. In addition, the interplay between metabolites and biopolymers (RNAs and proteins) (Figure 3G) shares features with hypercycles described by Eigen and Schuster [63,101,102]. Hypercycles that correspond to an organization model of molecules connected in a cyclic and autocatalytic manner have been proposed to play a major role in self-organization and self-reproduction, increasing the fidelity of interdependent processes while limiting the reproduction of parasitic elements.

Very importantly, metabolites are not only necessary for polymer biogenesis but can also react with them, leading to polymer biochemical modifications. These biochemical modifications (e.g., epigenetic modifications of DNA and RNAs, or post-translational modifications of proteins) change the physicochemical properties of the polymers and play a major role in the physiological adaptation of cells in response to environmental variations; however, at the same time, biochemical modifications can trigger polymer damages (e.g., protein aggregates and DNA mutations). One hypothesis is that the composition of biopolymers in cells is adapted to the nature of the metabolites that the cells produce (see Section 3). Thus, within a physiological range of metabolite concentration, metabolite-dependent biochemical modifications of polymers contribute to cellular physiological

adaptation; however, beyond a certain physiological threshold, biochemical modifications directly or indirectly induce mutations (Figures 2C and 3H). This suggests that biochemical modifications of nucleic and amino acid polymers establish a continuum between physiological and genetic adaptation.

5. Continuum between Physiological and Genetic Adaptation

The aim of this section is to show that environmental fluctuations induce physical and chemical constraints on nucleic and amino acid polymers, which trigger the cellular physiological adaptation that maintains cellular homeostasis (Figure 1B, Path 1). However, if the cellular response to environmental fluctuations does not return to equilibrium, constraints persist and can challenge the physical and chemical integrity of DNA, leading to DNA damage, mutations, and genetic variations. This process only ends when mutated sequences allow the direct or indirect relaxation of the initial constraints (Figure 1B, Path 2). First, how physical and chemical constraints originating from the environment or from cellular activities increase the probability of inducing directed and adaptive mutations, thus, “directing” genetic adaptation is shown and, then, the role of RNAs in genetic adaptation, including in multicellular organisms is discussed.

5.1. Genetic Adaptation Directed by Transcription: Transcription-Replication Conflicts

Several authors have already proposed that cellular stresses (i.e., environmental fluctuations above a physiological range) induce “adaptive mutations” (i.e., mutations that occur at a high rate) or “directed mutations” (i.e., mutations occurring at specific genomic locations) [103–105]. A possible underlying mechanism is that stress-directed transcriptional activation of specific loci (as part of the cellular physiological adaptation) increases the probability of mutations occurring within these loci because transcription induces mechanical stresses (e.g., formation of supercoiling) that challenge the physical integrity of transcribed DNA [12,34,104–107]. Next, how this straightforward principle establishes a continuum between physiological and genetic adaptation is described below.

Highly transcribed genes are enriched in GC nucleotides that can “absorb” mechanical stresses by favoring the B- to Z-DNA transition, owing to the physical properties of both base-pairing and base-stacking interactions between G and C nucleotides (see Section 3). Critically, this GC enrichment is explained by several transcriptional-dependent mutational biases (see Glossary), which result, in part, from conflicts between transcription and replication [107,108]. Indeed, while the act of transcription increases DNA accessibility to DNA polymerases, the simultaneous transcription and replication of a locus creates conflictual physical stresses leading to DNA breaks [107–109]. DNA breaks can be repaired by heteroduplex DNA recombination, which is a process known to favor G:C more than A:T base pairs [110–113]. This phenomenon, in part, relies on the fact that a T in a mismatched base-pair (T:G or T:C) within heteroduplex DNA can spontaneously flip out of the dsDNA, increasing the probability of its removal [114]. In addition, transcription-replication conflicts have been shown to induce adenine deamination, giving rise to hypoxanthine, a nucleotide that mimics guanine and leads to A:T > G:C mutations during replication [115]. Different mutational biases also occur in early and late replicating regions [7]. For example, the accumulation of free oxidized-dGTP before replication can result in its incorporation in place of Ts (as oxidized-dGTP mimics adenine), leading to A:T > G:C mutations during the early phase of replication [116]. Additionally, as DNA cytosine methylation is associated with transcription repression, heavily methylated regions correspond to late replicating regions; these regions could more frequently undergo C:G > T:A mutations because of the high rate of spontaneous deamination of methylcytosine [117]. Replication-dependent mutational bias can also be due to decreases during the cell cycle of concentrations of free dGTPs and dCTPs (as producing these nucleotides requires more energy than producing dATP and dTTP), which leads to a higher incorporation rate of A or T nucleotides in late-replicating regions [118]. Therefore, mutational biases associated with replication timing and replication-transcription conflicts could increase the GC- and AT-content in early and late replicating regions, respectively.

While the GC-content of loci can increase as a consequence of DNA breaks resulting from transcription-replication conflicts, then, this increase exerts positive feedback loops by (i) synchronizing transcription and replication, (ii) increasing local DNA stability during transcription or replication, and (iii) increasing the transcription activity of modified loci as well as many downstream steps of the gene expression process [27,31,41,119]. Indeed, the mechanical properties of GC-rich DNA regions favors transcription efficiency (but not elongation speed) and avoids nascent RNAs from interacting with the DNA template due to the formation of stable secondary structures in the nascent GC-rich RNAs [69,70]. High GC-rich content can also increase the local rate of recombination and deletion, leading eukaryotic GC-rich genes to bear smaller introns (as compared with AT-rich genes) [29,30,112,120]. Of note, small GC-rich introns are more efficiently spliced, likely because of intronic RNA secondary structures [79,121]. In addition, high GC content in RNAs increases the efficiency and fidelity of translation by smoothing translation elongation and favoring cotranslational protein folding [51,122–124]. Finally, as the genetic code is not random, increasing gene GC content leads to the biogenesis of proteins with small amino acids, which in turn leads to decreases in: (i) protein volume, (ii) concentration-dependent aggregate formation, and (iii) the energetic cost of protein production [30,125–127]. Therefore, the “over-stimulation” of the transcriptional activity of some genes under sustained stresses could result in (i) replication-dependent mutational bias; (ii) increases in the GC-content of stress-induced genes; and (iii) decreases in the local transcription-dependent DNA instability, while improving gene product biogenesis at multiple levels (Figure 4A). Another genetic process that increases the biogenesis of specific gene products under stress situations is gene duplication, which is also linked to transcription-replication conflicts [12,128]. For example, stress-induced promoter activity can stimulate gene duplication by destabilizing stalled replication forks [128].

These observations, therefore, support a model in which environmental fluctuations induce physical constraints on DNA during transcription, leading to the biogenesis of RNAs and proteins, and, then, leading to re-establishing equilibrium through the cellular physiological adaptation (Figure 4A, 1). However, if the constraints persist, sustained transcriptional activation can result in transcription-replication conflicts resulting in GC-mutational biased or gene duplication, which could relax the initiating constraints by increasing gene product levels (Figure 4A, 2).

5.2. Genetic Adaptation Directed by Transcription: Role of ssDNA Formation and DNA Folding

Transcription-dependent chromatin relaxation and ssDNA formation increase the accessibility of transcribed DNA regions to mutational agents or so-called “transcription-associated mutations” [5,12,34,105–107,129]. For example, ROS-mediated oxidation of cytosines and methylcytosines is enhanced in ssDNA, which can lead to C > T mutations during replication [34]. As a consequence, CGN codons (corresponding to Arg) of transcriptionally activated genes under ROS exposure have a higher probability to be mutated into TGG, TGA, and TGY codons (corresponding to Trp, stop, and Cys codons, respectively), allowing the synthesis of proteins containing amino acids that protect cells from oxidation (see Section 4 and Figure 4B).

Remarkably, a genome-wide analysis of mutations occurring in organisms growing under different environmental constraints (e.g., different metabolic resources or temperatures) shows that each challenging condition is associated with a specific mutational bias [130–134] (see Glossary). This is in agreement with the fact that each mutagenic agent (i) affects sequences with specific physicochemical properties and (ii) induces nucleotide modifications toward a particular pattern, as has now been well established in cancer genetics [135]. Of note, in addition to ROS-associated mutational signatures in cancers (see Section 4), mutational bias induced by high intracellular pH has recently been shown to favor Arg (CGY) > His (CAY) mutations that confer pH-regulated protein functions [136]. It would be very interesting to systematically characterize the relationship between (i) the physicochemical properties of mutagenic agents as well as of their most affected sequences and (ii) the nature of the induced mutations and the resulting physicochemical consequences at the nucleotide and amino acid levels [8]. Each mutagenic agent could trigger a specific mutation bias that more or less directly relaxes the initial physicochemical constraints (Figure 1B, 2).

It is of particular interest that transcription-dependent ssDNA formation also increases the probability of insertion of repeated elements, such as transposons and retrotransposons, a process known to play a major role in the genomic plasticity under sustained stresses [137–141]. It has been proposed that cellular stresses leading to a global chromatin relaxation could, on the one hand, de-repress (retro) transposon activity and, on the other hand, increase the likelihood of their insertion in specific stress-transcriptionally activated genes [137–141]. Insertion of (retro) transposons in stress-activated genes can influence gene expression at multiple levels, particularly by playing a role in the spatial genome organization. Indeed, it is now well recognized that regulation of gene transcription is based on the three-dimensional (3D) genome organization, which roughly corresponds to DNA folding (Box 1). DNA folding plays a critical role in co-regulating genes by bringing them closer together in space [142–146]. Factors binding to repeated sequences dispersed in various genomic locations could facilitate 3D genome organization and promote co-regulation of repeated sequence-hosting genes [147,148] (Figure 4C). Spatial clustering of co-regulated genes could also increase the probability of their recombination, a process facilitated by the presence of repeated elements [149–152]. Recombination between transcriptionally co-regulated genomic regions can lead to the formation of new gene products but also can facilitate the expression coordination of costimulated genes (Figure 4C). The importance of gene position in genomes with respect to their regulation and cellular functions is clearly established in operons from bacteria and in the so-called topologically-associated domains (TADs) in eukaryotes [153–156]. An emerging theme is that the one-dimensional (1D) and 3D locations of genes play a major role in coordinating the expression of genes whose products are involved in the same cellular pathways (Box 1).

In summary, environment-dependent physicochemical constraints on DNA trigger cellular physiological adaptation and a continuum between physiological and genetic adaptation is established when environment-dependent physicochemical parameters above a physiological range challenge DNA physicochemical integrity.

5.3. Physiological Adaptation Facilitates Genetic Adaptation: Role of RNAs

DNA chemical modifications can contribute to physiological adaptation (e.g., the role of DNA methylation and oxidation in transcription regulation) and induce mutations during replication when a chemically modified nucleotide “mimics” another nucleotide (see Section 3). This mimicry process can also result in transcription “infidelity” by affecting the base-pair rules, thereby leading to biogenesis of RNAs with different sequences than those encoded by the DNA template [157–159]. This so-called “transcriptional mutagenesis” is more frequent than previously anticipated and could play a major role in both physiological and genetic adaptations. For example, chemical modifications in transcribed DNA can lead to the biogenesis of new RNAs and proteins, which could contribute to cell survival. As a consequence, only cells bearing the DNA chemical modifications can divide, and therefore having the DNA chemical modifications would increase the probability of giving rise to genetically adapted daughter cells [157–162] (Figure 4D). The same principle can be used to establish a direct link between genetic adaptation and epigenetic modifications (i.e., DNA or histone chemical modifications that impact gene expression), as environmental fluctuations induce epigenetic modifications of specific loci as part of the cells’ physiological adaptation. If transcription-dependent epigenetic modifications at specific loci increase cell survival, the surviving cells have a higher probability of undergoing mutations within these genes, as chromatin organization and epigenetic marks more or less directly impact both DNA damage and DNA repair [34,163,164].

Along the same line, RNAs produced from loci as part of the cellular physiological adaptation could, under certain circumstances, induce mutations in the loci they originated from. For example, the spatial proximity of co-regulated genes (see above) could promote the biogenesis of chimeric RNAs via a mechanism called trans-splicing, thereby “fusing” RNAs produced from two genes [165]. These chimeric RNAs can give rise to new proteins that could contribute to the survival of cells in a stressful situation. As RNAs can be used as a matrix during DNA break repair, surviving cells expressing chimeric RNAs could use these RNAs during the repair of loci broken by the stress-induced transcription, increasing the probability of recombination between specific loci [166,167].

RNAs can increase the likelihood of genetic variations in numerous ways [168,169]. Recently, spectra of molecular mechanisms have been described by which physicochemical constraints on RNAs and proteins, at the time of their synthesis, could trigger mutations in their originating genes through the biogenesis of RNA fragments [72,170,171]. Briefly, environmental fluctuations, on the one hand, induce the transcriptional activity of target genes, and thereby generate a greater amount of mRNAs and proteins, and, on the other hand, generate constraints on nascent RNAs and nascent proteins during transcription and translation (Figure 4E, 1 and 2). Perturbation of mRNA or protein synthesis leads to the biogenesis of RNA fragments, for example, mRNA cleavage occurs when the dynamics of ribosomes (e.g., as a consequence of nascent protein misfolding) along the mRNA template is disturbed [172,173]. RNA fragments generated during transcription or translation could then interact with their originating genomic regions and induce genomic instability and mutations in the targeted regions (Figure 4E, 3). Therefore, RNA-directed mutations could increase the likelihood of mutations occurring in specific loci when cells experience constraints during the biogenesis of specific RNAs or proteins.

In summary, environment-dependent physicochemical parameters trigger cellular physiological adaptation through changes of the cellular activities that leave traces or footprints on nucleic acid polymers through physical damages (e.g., DNA breaks), chemical modifications (e.g., DNA oxidation), and biogenesis of RNAs that can, next, target specific genomic locations. If the physiological adaptation allows a return to equilibrium, polymer modifications are temporary and reversible (Figure 1B, 1). If not, the footprints left on DNA (i.e., the “cell experience”) can have a more-or-less direct effect on replication, potentially leading to mutations at specific loci. This mutational process triggered by physicochemical constraints only stops when new sequences relax the initiating constraints (Figure 1B, 2). This principle is well suited to unicellular organisms in which the same DNA molecule is used as a template during (i) the physiological response to environmental fluctuations and (ii) replication and transmission across generations. Could such a principle apply to multicellular organisms?

5.4. Somatic Physiological Adaptation and Germline Genetic Adaptation: Role of RNAs

It has been proposed above that environment-dependent physicochemical constraints on DNA trigger a cellular physiological adaptation that can leave “marks” (e.g., DNA chemical modifications or RNAs), which, then, potentially result in mutations during replication. As the replicated-DNA that is transmitted across generations in multicellular organisms is no longer directly involved in physiological adaptation, do physicochemical constraints exerted on somatic cell phenotype induce modifications on germline cell DNA?

To answer this question, first, it must be stressed that germline cells depend on and are exposed to the activities of somatic cells. For example, metabolic disorders are associated with the overproduction by somatic cells of small sugars or lipids that can react with and induce damages in the germline cell DNA. Furthermore, metabolic activity of somatic cells, for example under nutrient constraints of the parents, can induce epigenetic changes on specific loci in germline cell DNA with consequences on the development and activity of the offspring [174–177]. There is also a consensus regarding the fact that molecular exchanges between somatic and germline cells are more complex than previously anticipated. For example, somatic cells produce extracellular vesicles that (i) contain proteins, metabolites, and a diversity of small RNAs, for example, microRNAs, tRNA-derived small RNAs (tsRNAs) and (ii) are internalized by germline cells with consequences on the development and activity of offspring after fertilization [176,178–182] (Figure 4F, 1). These processes, collectively called “transgenerational epigenetic inheritance”, establish that a somatic cell’s “experiences” in the parent organism can be transmitted at the molecular level to germline cells, with consequences on the offspring’s phenotype [176,178–183]. Another process, called “genomic imprinting” that also depends on germline epigenetic marks and small RNAs leads to the selective expression of alleles transmitted from one of the two parents sometimes in an environment-dependent manner [184]. Epigenetic marks of one allele can also be transferred to the other allele, a process called “paramutation”, which is based on the biogenesis of biochemically modified small RNAs and which

underlines the plasticity associated with epigenetic mechanisms [185,186]. All the processes described above maintain a diversity of alleles by transcriptionally activating or repressing some of them in the offspring, depending on the parental somatic cells' experiences (Figure 4F, 1 and 2) and combined with meiosis, these processes could "purge" some deleterious alleles [187–193].

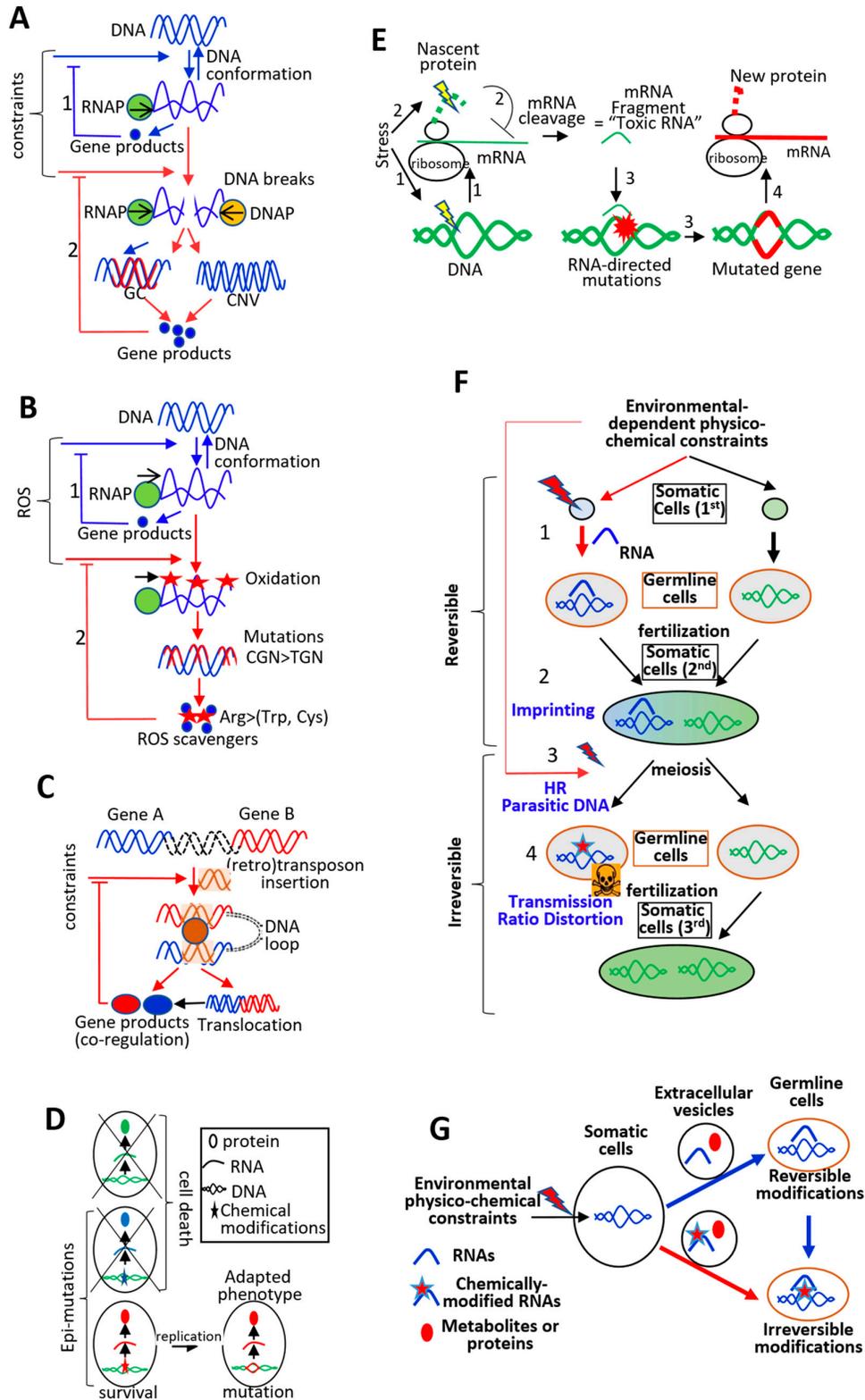


Figure 4. (A) Variations in the extracellular environment increase the expression of target genes which is a process associated with physical constraints on DNA generated by RNA polymerases (RNAP). These physical constraints are transient if the gene activity relaxes the initial constraints (e.g., through the biogenesis of gene products) (1). If not, physical constraints persist, and conflicts between RNAP and DNA polymerases (DNAP) induce DNA damages. DNA damage is repaired by homologous recombination, which favors GC over AT nucleotides, but which can also induce gene copy number variation (CNV). Both of these processes, in turn, generate more gene products, and therefore relax the initial constraint (2). (B) Cellular stresses activate the expression of specific target genes while increasing the production of intracellular ROS. One strand of transcriptionally induced genes is exposed to ROS, promoting deamination of methyl cytosine, which gives rise to thymine. C > T mutations change Arg codons into Trp and Cys codons. Trp- and Cys-containing proteins can play a role in protecting cells from ROS, therefore, relaxing the initial constraints (2). (C) Transcriptional activation of genes induces neo-insertion of repeated sequences within transcription-dependent ssDNA. Neo-insertion of repeated sequences can facilitate co-regulation of two genes by bringing them closer to each other in space, and it also promotes recombination. Both processes coordinate the production of gene products and contribute to relaxing the initial constraints. (D) Cellular stress induces chemical modifications of target genes, which affects chromatin organization and transcription fidelity (“epi-mutations”). Chemical modifications that induce biogenesis of new RNAs and proteins could allow survival of the cells in which these modifications took place. Chemical modifications in surviving cells lead to mutations that increase the survival probability of daughter cells. (E) Cellular stress induces physical constraints simultaneously on target genes and on proteins produced from these genes (1 and 2). By disrupting translation, for example, by inducing nascent protein unfolding, a stress can induce translation stopping and cotranslational cleavage of mRNAs. RNA fragments generated during translation or transcription hybridize on the complementary DNA strand and locally generate DNA or chromatin modifications, thus, increasing the probability of mutations occurring in the targeted regions (3). This process stops only when the gene and its products obtain physicochemical properties that relax the initial constraints (4). (F) Somatic cells are constrained by environmental fluctuations, and their activities can have consequences for germ cells, for example, through the transfer of metabolites or small RNAs from somatic to germline cells (1). These compounds can change the activity of germ cells, affect the development of the body after fertilization, and cause mutations in germ cells (2 and 3). These mutations could lead to the emergence of somatic cells (3 and 4) whose activities maintain the integrity of the germ cell genome of the following generations. (G) Somatic cell genes that are constrained by environmental parameters produce extracellular vesicles containing RNAs that can either induce reversible epigenetic mutations or irreversible genetic mutations.

Indeed, the frequency of transmission of some alleles in offspring can vary depending on their parental origin. This mechanism, known as the “transmission ratio distortion”, is an exception to Mendel’s laws of equal segregation and seems to rely on a diversity of mechanisms, such as “selection” among chromosomes during meiosis (e.g., non-random crossovers during meiotic recombination), allele-dependent elimination of gametes, or selective elimination during early zygote development [191–193]. Although the underlying molecular mechanisms of transmission ratio distortion are not yet fully understood, such a process establishes that sexual reproduction allows the selective elimination of some alleles. One possibility is that genes targeted by somatic RNAs in germline cells have a lower probability to be transmitted in next generations depending on the parents’ experiences (Figure 4F, 3 and 4). Therefore, sexual reproduction allows genes to be turned on and off, and for some alleles to be eliminated across generations without the need of de novo mutations in germline cells.

In this context, several studies have shown that, although the occurrence of germline de novo mutations per generation is very low in some species, their distribution across genomes is biased [9,10,194,195]. For example, an association between de novo mutation occurrence, replication timing, transcription, and chromatin organization has been observed in germline DNA [9,10,194,195]. Furthermore, de novo “mutational clusters” corresponding to multiple de novo mutations in very close vicinity in a single individual, as well as “mutational hotspots” corresponding to de novo

mutations occurring at the same location in several individuals, have been reported [194–196]. As the distribution of de novo mutations across genomes is biased, an important issue is to decipher whether their occurrence could depend on the somatic cell experiences. One possibility could rely on the fact that RNAs produced by somatic cells induce local and targeted epigenetic modifications in the germline DNA, which next induces more or less directly targeted de novo mutations because of the interplay between the chromatin environment and the local mutational rate (Figure 4G, blue arrows) [34,163,164,197,198]. Although very speculative, there is also the possibility that the DNA of somatic cells challenged by environment-dependent physicochemical parameters produces “parasite-mimicking RNAs” that could form DNA:RNA hybrids in their complementary loci in the germline DNA and locally trigger mutations (Figure 4G, red arrow). The following supports this possibility: (i) A convergence between RNA-containing extracellular vesicles and viral particles has been described; (ii) RNAs are widely used in all living organisms to cleave or mutate parasitic nucleic acids; and (iii) RNA:DNA hybrids can be genotoxic, for example, RNA:DNA hybrids can induce DNA adenine deamination [72,168–171,185,186,199–204]. It would be very interesting in the future to investigate whether some of the 150 chemical modifications of RNAs identified so far could trigger selective mutations in RNA:DNA hybrids [72,168–171,185,186,201,202,204].

To summarize, in unicellular organisms, the cell’s experience leaves footprints on specific DNA locations (e.g., breaks and chemical modifications) that can lead to local mutations during replication, thus, establishing a continuum between physiological and genetic adaptation. In multicellular organisms, somatic cells challenged by environment-dependent physicochemical parameters may not properly protect germline cell DNA from physicochemical constraints and could produce compounds (e.g., RNAs) that target specific locations of germline cell DNA. Section 6 provides a description explaining how these mutational processes, and the interplay between the genome and phenotype stability contribute to the emergence of more-or-less complex phenotypes, including those in multicellular organisms.

6. Interplay between Environment-Dependent and Cell-Dependent Physicochemical Constraints and the Emergence of Complex Phenotypes

On the basis of the relationship between RNAs and proteins at the origin of life, in Section 4, it is proposed that evolution relies on the fact that the phenotype maintains the physiochemical stability of its originating genome in a manner that depends on environmental, physicochemical constraints (Figure 1A). One of the objectives of Section 6 is to show how this principle leads to molecular innovations that allow the emergence of new physicochemical properties at higher scales of life organization, i.e., complex phenotypes that can next act on the initial constraints (Figure 5A, blue lines). Furthermore, the aim is to show that the phenotype (irrespective of the scale of organization) generates constraints on its own genome. In other words, environmental constraints induce molecular innovations that can also directly or indirectly generate constraints on their originating genome (Figure 5A, red lines), and therefore shows how this principle creates an evolutionary dynamic. Throughout Section 6, it is stressed that the mutational processes described in Section 5 are not deterministic but rather probabilistic, and therefore are the source of variability, and therefore diversity. Similarly, the fact that the emergence of certain properties under physicochemical constraints can be considered as “side effects” that contribute to life form diversity is highlighted.

6.1. From Molecular Innovations to Emergence of New Properties at Multiple Scales of Life Organization: Metabolic Activities and Cell Organization

The phenotype (i.e., the sum of all cellular activities) necessarily generate physicochemical constraints on its originating genome since numerous compounds are produced by the cell in response to environment-dependent physicochemical variations. The compounds can interact and react with nucleic and amino acid polymers with potentially deleterious effects on their stability, folding, or solubility [21,205] (Figure 5B and see Section 3). Therefore, cellular activities represent a source of physicochemical constraints on the originating genome. The emergence of photosynthesis,

oxidative metabolism, and eukaryogenesis illustrate the evolutionary dynamic generated by the interplay between phenotypic and genomic stability.

Before the emergence of the ozone layer, cells were exposed to strong solar UV radiations [206–208]. By inducing genomic instability, UV favored the emergence of genomes that produce pigments absorbing UV-damaging radiations, which would give these genomes a greater probability to be “accurately” reproduced [206–208] (Figure 5C, 1). Interestingly, it has been proposed that the molecular ancestors of photosynthetic light acceptors (e.g., chlorophyll) were pigments that protected nucleic and amino acid polymers from UV irradiation [206–208]. One possibility is that UV-absorbing pigments generated genomic instability because of the free dispersion of energy (heat) released from UV-absorbing pigments, favoring the emergence of pigment-interacting proteins containing photosynthetic reaction centers that could concentrate energy into complex molecules (e.g., sugars or “energy tanks”) by fixing CO₂ [206–208] (Figure 5C, 2). This gave rise to photosynthesis that in turn generated new constraints, as photosynthesis produces oxygen, a highly toxic compound that damages nucleic and amino acid polymers. In this setting, the ancestors of the gene products and metabolites involved in cell respiration have been proposed to have originated from oxygen scavenger compounds or oxidases that did not conserve energy and that protected their originating genome from the rise of oxygen [209–211]. Therefore, the emergence of cellular respiration could have resulted from a “detoxification” process concentrating oxygen-derived energy in biosynthetic pathways, in the same way that photosynthesis emerged from energy concentration from UV radiations [209–211] (Figure 5C, 3).

The rise of oxygen produced by photosynthetic cyanobacteria in the earth atmosphere could also have impelled anaerobic cells (e.g., archaeas) and aerobic cells (e.g., alphaproteobacteria) to cooperate, as aerobic cells could protect anaerobic cells by scavenging environmental oxygen and because both cell types exchanged intermediate metabolites [212–214]. This cooperation might have resulted in the internalization of aerobic bacteria (the ancestors of mitochondria) by anaerobic archaea, resulting in the emergence of eukaryotes [212–214]. However, as mitochondria, subsequently, generated intracellular toxicity by producing intracellular ROS, this detoxification would have favored the biogenesis of new cellular compounds [215,216]. In this setting, biogenesis of sterols that requires oxygen-dependent enzymes could have first played a role in oxygen detoxification [216–218]. In addition, these molecules have specific properties when incorporated into membranes that contribute to the development of eukaryotic intracellular membranes, such as the nuclear membrane, which could have initially protected intracellular polymers (e.g., DNA) from ROS [216–219]. Indeed, intracellular biomembranes can fold up into three-dimensional periodic arrangements (“cubic membranes”), representing antioxidant defense [220].

In summary, the step-by-step emergence of photosynthesis, oxidative metabolism, and eukaryogenesis over evolutionary time could have been triggered by extracellular (e.g., UV radiation) and intracellular (e.g., ROS) physicochemical constraints that destabilize genomes (i.e., induce genetic variations). This resulted in the emergence of new genomes that produced molecules that relaxed the initiating constraints (molecular innovations), therefore, stabilizing their originating genomes but, simultaneously, generating new constraints (Figure 5C). The emergence of intracellular membranes triggered by the increase of intracellular ROS also illustrates that molecular innovations in response to physicochemical constraints (e.g., sterol metabolism as a ROS detoxification process) supports the emergence of new properties at the upper level of life organization (e.g., cell organization based on intracellular membranes). At the cellular level, these properties (i.e., cell compartmentalization), then, contribute to reducing ROS-dependent genomic instability. Furthermore, the interplay between molecular innovations and the emergence of new properties at the (multi)cellular level is illustrated below.

6.2. From Adaptation to a Diversity of Environment-Dependent Constraints, to Side Effects: Genome Organization, Epigenetics, and Multicellularity

The increase in intracellular ROS levels produced by mitochondria in eukaryotes could have been relevant to the origin of eukaryotic spliceosomal introns from group II introns found in archaea

and bacteria, which are in fact mobile retroelements that use the combined activities of an autocatalytic RNA and an intron-encoded reverse transcriptase to propagate within genomes [221,222]. It has been proposed that retromobility of group II introns can be stimulated by oxidative stress and that the presence of introns could have contributed to “trapping” ROS in introns, thereby, decreasing the probability of nucleotide oxidation in coding exons [221,223–225]. Supporting this model, nucleotide oxidation increases the probability of GC > AT mutations, and the frequency of GC nucleotides is higher in exons than in introns [225,226]. This implies that exons have been protected from ROS, which could have been achieved by histones. Indeed, histones are preferentially found in GC-rich sequences because of the flexibility of the G–C stacking interactions (see Section 3), and they protect DNA from a variety of mutagenic stresses by (i) binding to and stabilizing dsDNA, (ii) compacting DNA, and (iii) providing a “shield” through their C-terminal tails that are rich in Arg and Lys, i.e., amino acids that are basic and positively-charged and can act as ROS scavenger [227–231]. Therefore, binding of histones to GC-rich exons can protect them from oxidation, while intronic GC > AT mutations induced by ROS would ultimately lead to exclude histones from introns. If introns contributed to maintaining the stability of exons, a side effect of intron invasion is that it increased the diversity of proteins produced by eukaryotic genomes through alternative splicing. Similarly, the diversity of genome-driving phenotypes could be a side effect of the emergence of histones.

Indeed, the cell physiological adaptation to environmental fluctuations relies on the biogenesis of gene products and metabolites. Nevertheless, all possible biochemical reactions cannot take place simultaneously in a cell as (i) each reaction depends on specific physicochemical conditions and (ii) the diversity of the generated biochemical products would be highly toxic, due to their ability to interact and react spontaneously with each other [21,205] (see Section 3). Therefore, a cell can physiologically adapt to a limited number of fluctuating physicochemical parameters, which pushes towards the cooperation between unicellular organisms performing complementary biochemical reactions [232] (Figure 5D).

In this context, histones could have allowed an increase in the diversity of metabolic activities encoded by a single genome [233]. Eukaryotic histones evolved by compacting DNA and by acting as a “chemical” shield, therefore, maintaining the stability of the genome that produces them (see above). Histone chemical modifications (i.e., epigenetic marks) could first have been triggered as chemical “shields” and played a role in maintaining DNA stability against chemical DNA “attacks” [234]. However, by affecting DNA accessibility, histone chemical modifications would not only have protected specific genes but also coordinated their activity depending on the intracellular chemical composition. Indeed, different epigenetic marks can protect different parts of the genome from different chemical compounds, while simultaneously adapting gene transcriptional activities with respect to these chemical compounds (Figure 5E). Two pieces of evidence support such a possibility. First, epigenetic marks are directly dependent on the cell metabolism, for example, methylation depends on SAM produced by the one carbon cycle, and demethylation relies on oxidation of methylated residues, and therefore on the cellular oxidative metabolism [96–98,233]. Secondly, histone chemical modifications either reduce or facilitate DNA access to RNA polymerases and to potentially genotoxic molecules [97,98,163,234,235]. By selectively protecting and regulating gene expression, histone chemical modifications contributed to the emergence of different cell types (i.e., multicellularity) containing the same genome but performing different metabolic activities (Figure 5E).

In summary, molecular innovations triggered by physicochemical constraints allow (directly or as side effects) the emergence of new properties at higher scales of life organization, as well as of complexity.

6.3. From Diversity to Complexity: Interplay between Germline Cell DNA and Somatic Cell Phenotype.

As depicted above, emergence through the course of evolution of molecular innovations (e.g., intracellular membranes, introns, and histones) impacting on cell and genome organization could have allowed an increase in the diversity of metabolic processes driven by one single genome (Figure 5E). As a consequence, genomes would have been exposed to an increasing diversity of potentially

genotoxic biochemical compounds. In this setting, it has been proposed that meiosis protected DNA from cell metabolic activities. This hypothesis, known as the "dirty work hypothesis" [236,237], corresponds to the fact that meiosis relies on homologous recombination, a mechanism that removes damaged (e.g., oxidized) nucleotides (see Box 3). In this model, meiosis is a process that removes damaged nucleotides while allowing the formation of haploid germ cells. Thanks to the formation of germ cells, the DNA transmitted across generations is no longer directly exposed to metabolic activities (i.e., the "dirty work") yet still "benefit" from these activities [236,237]. By considering that the phenotype that depends on the germ cell DNA corresponds to the production of somatic cells, the relationship between a genome and its phenotype depicted at the molecular level (see Section 4) still operates in multicellular organisms. Indeed, germ cell DNA (the genome) gives rise to somatic cells (the phenotype), those activities allow the integrity of the germline DNA molecule to be maintained. The germline DNA transmitted to the next generation after fertilization allows the same phenotype to be generated under a stable environment (Figure 5F); note here that the "same phenotype" should not be understood in a literal sense (see Discussion). Although speculative, the molecular mechanisms establishing a continuum between physiological and genetic adaptation (depicted in Section 5) could still operate in multicellular organisms, since specific germ cell DNA locations or their associated histones could be biochemically modified when somatic cells challenged by environmental fluctuations produce molecules (e.g., RNAs) targeting specific genomic locations of germ cell DNA (see Section 5, Figure 4F, G).

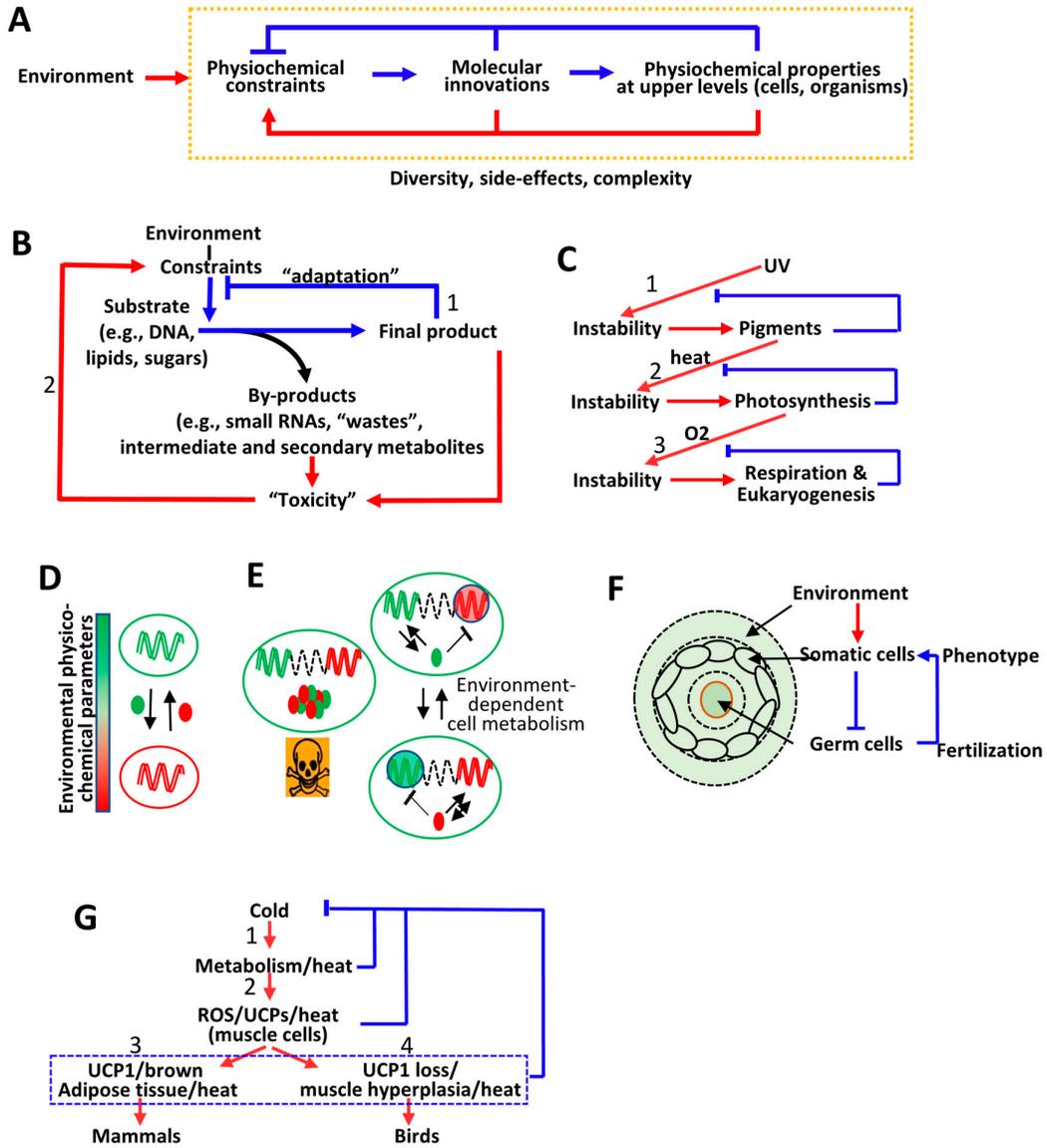


Figure 5. (A) The environment generates physicochemical constraints which, by destabilizing genomes, induce molecular innovations, allowing the emergence of new physicochemical properties at different scales of life organization. These molecular innovations and new properties are transmitted across generations if they release the initial constraints (blue lines) but can also generate new constraints (red lines), allowing the adaptation of the genome to its own activities and generating an evolutionary dynamic. (B) Any biochemical reaction leads to the synthesis of a final product (1) and "waste", by-products, or secondary metabolites. These compounds are not necessarily essential to the cell's survival, but they are the result of vital cellular activities. Therefore, these compounds are not "random products" even though they can have cellular toxic effects by interacting with cellular polymers (2), which can lead to mutations. (C) UV radiations induce genomic instability favoring genomes that produce pigments that, in turn, protect them from the initial constraint (1). Likewise, UV radiation-absorbing pigments induce genomic instability favoring genomes that produce photosynthetic reaction centers that, in turn, protect them from the initial constraint (2). Likewise, O₂ production by photosynthesis induces genomic instability favoring genomes that produce oxidases and cellular components that, in turn, protect them from the initial constraint (3). (D) Two unicellular organisms (green and red), which are adapted to different environment-dependent constraints, cooperate by exchanging various components in a fluctuating environment. (E) Not all enzymatic reactions generated from a genome can take place simultaneously (left panel). The selective

compaction of different regions of a genome according to the cellular metabolic state (and therefore its environment) through chemical modifications of histones protecting some genome parts and repressing their potentially toxic expression while allowing the expression of genes whose products contribute to maintain the cellular homeostasis. (F) Somatic cells “buffer” environment-dependent constraints, maintaining the stability of the genome from the germ cells. When “protected” by somatic cells (i.e., when the phenotype of the organism is adapted to its environment), germ cells give rise to gametes that generate after fertilization the same somatic cells, i.e., phenotype. (G) Cold induces physiological adaptation by activating cellular respiration that increases cellular heat production (1). However, an increase in cellular respiration increases ROS production, which can lead to activation of UCP proteins. This simultaneously balances the ROS genesis and increases heat production in muscle cells (2) or in brown adipose tissue (3). The loss of UCP1 in bird ancestor may have led to muscle hyperplasia for heat production with consequences on bird body plan (4).

The interplay between somatic cell phenotype and germ cell DNA stability can be illustrated by adaptation of multicellular organisms to cold, a process that can also illustrate several interesting relationships between the different scales of life organization. Indeed, cold can lead to genetic instability by decreasing the kinetics of (bio) chemical reactions, and this cold-induced genomic instability could be relaxed by increasing the activity of genes encoding enzymes involved in cellular respiration, and therefore increasing heat production [238]. However, increasing cellular respiration increases the production of ROS that induces genetic instability, and therefore can lead to increased activity of genes involved in the detoxification of ROS, such as those coding for uncoupling proteins (UCPs) [239,240] (Figure 5G, 1 and 2). Indeed, uncoupling proteins such as UCP1 mediate proton leaks across the inner mitochondrial membrane, which (i) mitigate ROS production and (ii) simultaneously lead to cellular heat production [239,240]. The UCP-dependent heat production has been proposed to contribute to the emergence of heat-producing muscle cells, and, next, the emergence of mammalian brown adipose cells that (i) express UCP1, (ii) derive from skeletal muscle progenitor cells, and (iii) play an important role in heat production in mammals [241,242] (Figure 5G, 2 and 3). Interestingly, it has been proposed that the loss of genes such as UCP1 in bird ancestors (due to yet unknown mechanisms) did not allow the emergence of brown adipose cells but instead led to hyperplasia of heat-producing muscle cells. Thermoregulation depending on muscle hyperplasia (vs. brown adipose cells) has been proposed to generate constraints during development, with consequences on the body plan organization of birds (vs. mammals) [243,244] (Figure 5G, 4).

To summarize, environment-dependent physicochemical parameters (e.g., cold temperatures) could create constraints on somatic and germ cells, trigger genetic instability, and lead to the emergence of new polymers (e.g., UCPs) and new cell types (e.g., muscle cells). Supporting this model, it has recently been shown that exposure of parents to cold induces epigenetic modifications in sperm with consequences on adipose tissue activity in the offspring [245]. Adaptation to cold also illustrates the interdependence between physicochemical properties at the molecular level (e.g., proton leakage) and physicochemical properties at the upper level of life organization (e.g., heat production by cells) that can relax the initiating constraints (Figure 5G). Adaptation to cold also illustrates that adaptation to environmental physicochemical parameters can generate side effects as well as diversity. For example, the emergence of specialized cells producing large amounts of energy that allow control over organism temperature would also allow the organism mobility to be improved, and different innovations (e.g., brown adipose tissue vs. muscle hyperplasia) have different consequences in terms of body plan. Thus, although first arising as a side-effect, emerging properties (e.g., mobility) would also be under the control of natural selection.

7. Conclusions

To summarize the proposed model, a genome in a stable environment generates a phenotype that maintains the stability of its originating genome, and both (genome and phenotype) are reproduced identically (Figure 6A, left panel). Obviously, the word “identical” should not be taken in the strictest sense, as sequence variations within genomes may not have major influence on the phenotype, and variations in phenotypes allow the same range of physicochemical constraints to be

relaxed. In other words, a range of genotypes correspond to a range of phenotypes, which can cover a range of environment-dependent physicochemical constraints (Figure 6B). However, outside a physiological range of physicochemical parameters, a genome generates a phenotype that no longer maintains the stability of its originating genome and instead triggers mutations, whose rate, nature, and location dependent on the initial constraints and the challenged phenotype. This process occurs until new genetic variants generate a phenotype that maintains the stability of its originating genome (Figure 6A, right panel). It is important to stress that it is not a question of overall genome stability but of stability of genomic regions (for example, regions hosting certain genes) that are challenged by environmental fluctuations.

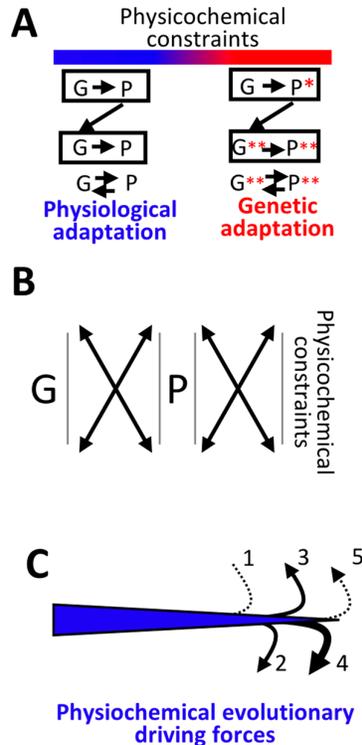


Figure 6. (A) A genome (G) generates a phenotype (P) that maintains the stability of its originating genome in a stable environment and both (genome and phenotype) are reproduced identically (left panel). In an unstable environment (corresponding to variations in physicochemical parameters above a physiological range), the genome generates a phenotype (P*) that no longer maintains the stability of its originating genome and instead triggers mutations whose rate, nature, and location dependent on the initial constraints and the phenotype (right panel). This process occurs until new genetic variants (G**) generate a phenotype (P**) that maintains the stability of its originating genome. (B) In a given environment, a genome (G) gives rise to a range of phenotypes (P), and similar phenotypes can correspond to a range of genomic sequences. Each phenotype can be adapted to a range of environmental physicochemical constraints. (C) Evolutionary driving forces rely on physicochemical processes whose probabilistic nature generates genetic and phenotypic diversity, as symbolized by arrows from 1 to 5. If a path (exemplified by Path 1) does not stabilize its originating genome, the genome and the corresponding phenotype will not be reproduced. While some paths might be neutral in terms of natural selection (Paths 2 and 3), some paths (Paths 4 and 5) could lead to the emergence of phenotypes that can be under natural selection.

In conclusion, while the notions of chance and natural selection are useful to highlight the fact that life has not been “created” by or for something, they cannot be considered as evolutionary driving forces. Instead, evolutionary driving forces correspond to environment-dependent

physicochemical constraints that challenge the phenotype and the underlying genome, and thereby direct their evolution. Such evolutionary driving forces cannot explain all the diversity of life for several reasons. The first is that mutation processes based on physicochemical processes are probabilistic. This means that diversity can still emerge from chance occurrences (Figure 6C). Secondly, mutations that relax the initial constraints allow the emergence of new characteristics or properties as side effects and contribute to the diversity of living organisms [246] (Figure 6C, 2 and 3). Finally, the emergent properties resulting from molecular innovations under the constraint of physicochemical parameters can also generate new properties that can confer certain advantages and disadvantages at the organismal level. This means that the probability of dissemination within a population of some mutations at the origin of molecular innovations can potentially be modulated by various phenomena, including natural selection (Figure 6C). Therefore, the notions of random mutations and natural selection are not evolutionary driving forces but contribute to life form diversity and act as a filter, respectively.

Funding: This research received no external funding.

Acknowledgments: I thank my colleagues, in particular M. Dutertre, C. Bourgeois, F. Mortreux, O. Gandrillon, N. Fontrodona, and M. Touillaud, for helpful discussion and a critical reading of the manuscript. I thank VA Raker for manuscript editing. I am grateful for the critical comments from the reviewers, which helped to improve the discussion.

Conflicts of Interest: The authors declare no conflict of interest.

Glossary

Chemical modifications Chemical modifications can change the physicochemical properties of monomers in nucleic and amino acid polymers (for example, a chemically modified nucleotide can “mimic” another nucleotide). Chemical modifications of nucleotides can, therefore, result in “transcription mutagenesis”, RNA editing, recoding during translation, and genetic mutations during replication. Similarly, modified amino acids within a protein (i.e., post-translational modifications) change the physicochemical properties of the modified proteins. As chemical modifications depend on the cellular metabolic activities that change under environmental fluctuations, this implies that environment-dependent metabolic activities can change the activities and nature (i.e., sequence) of gene products. Therefore, the cellular activities depend on genomic sequences and their chemical modifications in a manner that depends on extracellular and intracellular physicochemical parameters.

Composition bias The notion of composition bias relates to the fact that the global (or partial) composition of a polymer is different from that expected by chance. For example, the frequency of a mono-, di-, or trinucleotide can be higher than expected by chance. Nucleotide or amino acid composition biases correspond to specific physicochemical properties, as each monomer has specific physicochemical properties. The notion of composition bias is used in this manuscript in tight relation with the resulting and associated physicochemical properties.

Damage and error The word “error” is often used in biology. I do not use this term, as “error” refers to a deviation from a code. Codes are proposed by biologists to assemble a set of observations and to summarize the sum of knowledge at a time point. Therefore, deviation from a code can well be because an established code imperfectly describes the reality. Therefore, the word “damage” is preferred which can be used to highlight phenomena that challenge the physical integrity of biological objects (see “Genome and genetic stability” and “Mutational bias and signature”).

Emergence Emergence describes a process of coming into existence or the fact that a biological object (e.g., polymers) made of different elements (e.g., nucleotides) has physicochemical and structural properties that are more complex than the sum of the properties of the elements that composed them.

Genome and genetic stability Stability can refer to the physical integrity of components (e.g., genome stability) but also to maintenance over time of a particular sequence (e.g., genetic stability).

Both notions are related, as the physical instability of a genome can generate sequence variations and sequence variations are the product of physicochemical modifications of DNA.

Genome and phenotype Genome is a template that gives rise directly or indirectly to polymers (DNAs, RNAs, and proteins) in response to environment-dependent physicochemical parameters. Some of these polymers (proteins as gene products) can modify molecules supplied by the cellular environment and give rise to metabolites (“gene product products”). The phenotype corresponding to all the cellular or organism activities allows its originating template to be protected from environmental fluctuations and its integrity maintained. If not, the template changes (genome mutations) and gives rise to another phenotype, until it protects its originating genome against environmental fluctuations.

Mutational bias and signature Mutations depend on physicochemical laws. For example, a mutating agent has specific physicochemical properties, and therefore reacts with nucleotides and sequences which also have specific physicochemical properties. These reactions lead, therefore, to specific modifications of specific sequences. This corresponds to mutational bias and signature. Mutations are not “errors” since they are the consequences of physicochemical processes.

Box 1, Genome Physical Organization: From Gene Expression Regulation to Gene Product Functions

A DNA molecule is a template producing polymers (i.e., RNAs) in a manner that depends on its physical and structural properties at multiple scales. At the scale of a few nucleotides, DNA structure is determined by the arrangement of bases, the nature of which determines the depth and hydrophathy of the DNA grooves, as well as local electrostatic and electronic properties which collectively determine interactions between DNA and proteins such as transcription factors recruiting RNA polymerases [26]. At the scale of a dozen to hundreds of nucleotides, the nucleotide composition determines DNA thermodynamic and mechanical properties, such as resistance and response to topological stresses with consequences on transcription initiation and elongation [25,31,247]. Nucleotide composition-dependent mechanical properties, such as DNA bendability and rigidity, play a role in nucleosome positioning. The nucleosome frequency and the physical properties of the sequences between nucleosomes determines (i) the state of chromatin compaction and (ii) DNA/chromatin folding (e.g., the 3D organization) [248,249]. Chromatin organization and DNA folding play a major role in transcription regulation by allowing coordinated regulation of genes, as co-regulated genes in eukaryotes are parts of 3D structural units called topologically-associated domains (TADs) that correspond to regions of several hundred thousand of Kbps that fold and are nearby in the 3D nuclear space. An emerging model is that genes hosted by the same TAD share the same chromatin organization and depend on the same transcriptional regulators that form biomolecular condensates, allowing the simultaneous and efficient transcription of co-regulated genes [142–146]. As DNA folding relies on nucleotide composition-dependent physical properties that also determine the sets of interacting regulatory proteins, and as these features play a role in gene expression regulation, there is an overlap between the 1D (i.e., nucleotide composition bias) and 3D genome organization. Co-regulated genes, or genes within the same TADs, have similar nucleotide composition biases [250–253]. Of note, co-regulated genes coevolved toward the same nucleotide composition bias [252–254].

Interestingly, coding exons can share the same nucleotide composition bias as their flanking introns and their hosting genes [79,255]. As a consequence, mRNAs produced from co-regulated genes share the same nucleotide composition bias and are likely regulated by the same set of RNAs binding proteins, that interact with nucleotide composition biased sequences, in agreement with the concept of RNA regulons [50,253,256,257]. Furthermore, since the genetic code is not random, coding sequences sharing the same nucleotide composition bias encode for proteins having the same amino acid composition bias or the same physicochemical properties (see main text). As amino acid composition determines in turn the protein physicochemical properties, co-localized proteins or proteins contributing to the same cellular processes have similar amino acid composition biases [258,259].

In conclusion, composition bias determines, on the one hand, the location of genes in the 3D nuclear space and, on the other hand, gene product functions. Accordingly, genes contributing to the same cellular processes are in linear proximity in prokaryotes, and in 3D proximity in the nuclear space in eukaryotes [154–156,254,260–262]. Composition biases driving physicochemical properties of polymers, therefore, establish a straightforward link between production and function of gene products.

Box 2, The Genetic Code is not a Cipher but the Result of an Evolutionary Process Directed by Physicochemical Laws

While the genetic code is often described as a “frozen accident” and a cipher, early studies in the 1980s point out some physicochemical properties shared by amino acids and their cognate codons or anticodons. For example, the hydrophathy of amino acids correlates with the hydrophathy of the principal dinucleotides (excluding the wobble position) of their corresponding anticodons [44,56,57]. Along the same line, some protein physicochemical properties can be shared by their cognate mRNA. For example, the average hydrophobicity of a protein correlates with the average hydrophobicity of its cognate mRNA [256]. The physicochemical similarities between proteins and their cognate mRNAs could have played a role in the co-adaptation of these polymers to the same physicochemical parameters (see main text) but could have also been important for the physical interactions between proteins and their cognate RNAs [44,263,264]. Indeed, the ability of a protein to preferentially interact with its cognate RNA could have limited the free diffusion of proteins (see main text) and played a role in nascent protein folding [265]. Supporting a role for protein–RNA interaction in the evolution of the genetic code, (i) some amino acids preferentially interact with their corresponding codons or anti-codons; (ii) some RNA- and DNA-binding proteins bind preferentially to sequences owing to amino acid binding to their cognate codons; (iii) mRNAs enriched in a particular nucleobase (e.g., Gs) tend to encode proteins that interact with mRNAs made of the same nucleotide; and iv) many proteins, such as ribosomal proteins, bind to their own mRNAs [263–268].

The physical parameters of molecular dynamics regarding nucleic and amino acid polymerization could explain the triplet-based nature of the genetic code. Indeed, it is well-established that three-base codon structure of the genetic code contributes to translation efficiency and fidelity and molecular dynamics modeling suggests that charged particles (e.g., ribosomes) interacting with a polymer (e.g., an RNA) via electrostatic forces moves dynamically along the polymer in steps of three monomers [269]. Quite remarkably, there is now increasing evidence that triplets also play a major role in nucleic acid polymer properties and biogenesis which includes the following: (i) Triplets correspond to the width of the minor groove in a double-stranded nucleic acid polymer, and backbone atoms that are in proximity across the minor groove are separated by three nucleotides on the complementary strand [270,271]; (ii) a three-base periodicity has been observed outside coding sequences and provides, for example, specificity for the positioning of the transcription preinitiation complex [272]; (iii) codon bias affects transcription by affecting RNA folding, which favors transcription elongation by reducing pausing and RNA polymerase backtracking [69,273–275]; (iv) intra- and inter-trinucleotide stacking interactions contribute to stabilizing base pairing during the translation process but could have also played a role in replication early in evolution [76]. Collectively, these observations suggest that the three-base genetic code could have been constrained by physical parameters, allowing the simultaneous enhancement of RNA and protein biogenesis [269–271,276]. This would have been particularly important if both processes were physically coupled (see main text). In conclusion, physicochemical rules and parameters constrained the evolution of the genetic code that cannot be considered as a “frozen accident” but as an evolutionary process constrained by physicochemical laws.

Box 3, Meiosis

Multicellularity that emerged several times in the course of evolution can be described as a way for cells containing the same genome to perform different sets of chemical reactions, therefore, protecting each other's by exchanging metabolites (i.e., “metabolic division of labor”) [277–280].

However, increasing the diversity of metabolite exchanges within a group of cells would have increased the probability of DNA damage events and could have favored the emergence of sexual reproduction. Indeed, meiosis, the cellular process involved in the production of haploid gametes, likely emerged over evolutionary time as a mechanism to erase DNA oxidation resulting from cellular activities [187–189]. Supporting this model, meiosis is associated in many species with homologous recombination, a process primarily involved in DNA repair. As mentioned in the main text, homologous recombination favors GC over AT nucleotides (i.e., the so-called GC-bias gene conversion) because of physical properties of DNA, for example, T (corresponding to an oxidized (met)C) flips outside the DNA molecule in T:G mismatches [110–114]. In so doing, homologous recombination limits the load of GC > AT mutations triggered by C oxidation. Further supporting a role of meiosis in erasing ROS-dependent DNA damages, oxidative stresses trigger meiotic homologous recombination (even in organisms such as male *Drosophila*, which normally do not perform meiotic homologous recombination), and meiotic recombination and crossovers do not occur randomly but in DNA regions localized between methylated nucleosomes that are less protected from ROS [187–189]. In addition, meiosis that results in the formation of haploid cells is essential to eliminate damaged alleles that can be otherwise masked in diploid cells [187–190]. A side effect of meiotic homologous recombination is that different parts of different DNA molecules are mixed, so that this process increases the genetic diversity without the need of de novo mutations.

References

1. Koonin, E.V. The Origin at 150: Is a new evolutionary synthesis in sight? *Trends Genet.* **2009**, *25*, 473–475, doi:10.1016/j.tig.2009.09.007.
2. Noble, D.; Jablonka, E.; Joyner, M.J.; Muller, G.B.; Omholt, S.W. Evolution evolves: Physiology returns to centre stage. *J. Physiol.* **2014**, *592*, 2237–2244, doi:10.1113/jphysiol.2014.273151.
3. Laland, K.; Uller, T.; Feldman, M.; Sterelny, K.; Muller, G.B.; Moczek, A.; Jablonka, E.; Odling-Smee, J.; Wray, G.A.; Hoekstra, H.E.; et al. Does evolutionary theory need a rethink? *Nature* **2014**, *514*, 161–164, doi:10.1038/514161a.
4. Lynch, M.; Ackerman, M.S.; Gout, J.F.; Long, H.; Sung, W.; Thomas, W.K.; Foster, P.L. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **2016**, *17*, 704–714, doi:10.1038/nrg.2016.104.
5. Tomkova, M.; Schuster-Bockler, B. DNA Modifications: Naturally More Error Prone? *Trends Genet.* **2018**, *34*, 627–638, doi:10.1016/j.tig.2018.04.005.
6. Makova, K.D.; Hardison, R.C. The effects of chromatin organization on variation in mutation rates in the genome. *Nat. Rev. Genet.* **2015**, *16*, 213–223, doi:10.1038/nrg3890.
7. Tomkova, M.; Tomek, J.; Kriaucionis, S.; Schuster-Bockler, B. Mutational signature distribution varies with DNA replication timing and strand asymmetry. *Genome Biol.* **2018**, *19*, 129, doi:10.1186/s13059-018-1509-y.
8. Boulikas, T. Evolutionary consequences of nonrandom damage and repair of chromatin domains. *J. Mol. Evol.* **1992**, *35*, 156–180, doi:10.1007/bf00183227.
9. Stamatoyannopoulos, J.A.; Adzhubei, I.; Thurman, R.E.; Kryukov, G.V.; Mirkin, S.M.; Sunyaev, S.R. Human mutation rate associated with DNA replication timing. *Nat. Genet.* **2009**, *41*, 393–395, doi:10.1038/ng.363.
10. Rahbari, R.; Wuster, A.; Lindsay, S.J.; Hardwick, R.J.; Alexandrov, L.B.; Turki, S.A.; Dominiczak, A.; Morris, A.; Porteous, D.; Smith, B.; et al. Timing, rates and spectra of human germline mutation. *Nat. Genet.* **2016**, *48*, 126–133, doi:10.1038/ng.3469.
11. Danchin, E.; Pocheville, A. Inheritance is where physiology meets evolution. *J. Physiol.* **2014**, *592*, 2307–2317, doi:10.1113/jphysiol.2014.272096.
12. Yona, A.H.; Frumkin, I.; Pilpel, Y. A relay race on the evolutionary adaptation spectrum. *Cell* **2015**, *163*, 549–559, doi:10.1016/j.cell.2015.10.005.
13. Noble, R.; Noble, D. Was the Watchmaker Blind? Or Was She One-Eyed? *Biology* **2017**, *6*, doi:10.3390/biology6040047.
14. Booker, T.R.; Jackson, B.C.; Keightley, P.D. Detecting positive selection in the genome. *BMC Biol.* **2017**, *15*, 98, doi:10.1186/s12915-017-0434-y.

15. Wideman, J.G.; Novick, A.; Munoz-Gomez, S.A.; Doolittle, W.F. Neutral evolution of cellular phenotypes. *Curr. Opin. Genet. Dev.* **2019**, *58–59*, 87–94, doi:10.1016/j.gde.2019.09.004.
16. Lanfear, R.; Kokko, H.; Eyre-Walker, A. Population size and the rate of evolution. *Trends Ecol. Evol.* **2014**, *29*, 33–41, doi:10.1016/j.tree.2013.09.009.
17. Barrett, R.D.; Hoekstra, H.E. Molecular spandrels: Tests of adaptation at the genetic level. *Nat. Rev. Genet.* **2011**, *12*, 767–780, doi:10.1038/nrg3015.
18. Darwin, C. *On the Origin of Species*; John Murray: London, UK, 1859.
19. Echave, J.; Wilke, C.O. Biophysical Models of Protein Evolution: Understanding the Patterns of Evolutionary Sequence Divergence. *Annu. Rev. Biophys.* **2017**, *46*, 85–103, doi:10.1146/annurev-biophys-070816-033819.
20. Reed, C.J.; Lewis, H.; Trejo, E.; Winston, V.; Evilia, C. Protein adaptations in archaeal extremophiles. *Archaea* **2013**, *2013*, 373275, doi:10.1155/2013/373275.
21. Harmel, R.; Fiedler, D. Features and regulation of non-enzymatic post-translational modifications. *Nat. Chem. Biol.* **2018**, *14*, 244–252, doi:10.1038/nchembio.2575.
22. Panda, A.; Ghosh, T.C. Prevalent structural disorder carries signature of prokaryotic adaptation to oxic atmosphere. *Gene* **2014**, *548*, 134–141, doi:10.1016/j.gene.2014.07.002.
23. Elser, J.J.; Acquisti, C.; Kumar, S. Stoichiogenomics: The evolutionary ecology of macromolecular elemental composition. *Trends Ecol. Evol.* **2011**, *26*, 38–44, doi:10.1016/j.tree.2010.10.006.
24. Bragg, J.G.; Thomas, D.; Baudouin-Cornu, P. Variation among species in proteomic sulphur content is related to environmental conditions. *Proc. Biol. Sci.* **2006**, *273*, 1293–1300, doi:10.1098/rspb.2005.3441.
25. Vinogradov, A.E. DNA helix: The importance of being GC-rich. *Nucleic Acids Res.* **2003**, *31*, 1838–1844, doi:10.1093/nar/gkg296.
26. Harteis, S.; Schneider, S. Making the bend: DNA tertiary structure and protein-DNA interactions. *Int. J. Mol. Sci.* **2014**, *15*, 12335–12363, doi:10.3390/ijms150712335.
27. Dans, P.D.; Faustino, I.; Battistini, F.; Zakrzewska, K.; Lavery, R.; Orozco, M. Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 11304–11320, doi:10.1093/nar/gku809.
28. Goncarenco, A.; Ma, B.G.; Berezovsky, I.N. Molecular mechanisms of adaptation emerging from the physics and evolution of nucleic acids and proteins. *Nucleic Acids Res.* **2014**, *42*, 2879–2892, doi:10.1093/nar/gkt1336.
29. Versteeg, R.; van Schaik, B.D.; van Batenburg, M.F.; Roos, M.; Monajemi, R.; Caron, H.; Bussemaker, H.J.; van Kampen, A.H. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.* **2003**, *13*, 1998–2004, doi:10.1101/gr.1649303.
30. Urrutia, A.O.; Hurst, L.D. The signature of selection mediated by expression on human genes. *Genome Res.* **2003**, *13*, 2260–2264, doi:10.1101/gr.641103.
31. Reymer, A.; Zakrzewska, K.; Lavery, R. Sequence-dependent response of DNA to torsional stress: A potential biological regulation mechanism. *Nucleic Acids Res.* **2018**, *46*, 1684–1694, doi:10.1093/nar/gkx1270.
32. Chen, K.; Zhao, B.S.; He, C. Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem. Biol.* **2016**, *23*, 74–85, doi:10.1016/j.chembiol.2015.11.007.
33. Olinski, R.; Gackowski, D.; Cooke, M.S. Endogenously generated DNA nucleobase modifications source, and significance as possible biomarkers of malignant transformation risk, and role in anticancer therapy. *Biochim. Biophys. Acta Rev. Cancer* **2018**, *1869*, 29–41, doi:10.1016/j.bbcan.2017.11.002.
34. Tubbs, A.; Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **2017**, *168*, 644–656, doi:10.1016/j.cell.2017.01.002.
35. Reichenberger, E.R.; Rosen, G.; Hershberg, U.; Hershberg, R. Prokaryotic nucleotide composition is shaped by both phylogeny and the environment. *Genome Biol. Evol.* **2015**, *7*, 1380–1389, doi:10.1093/gbe/evv063.
36. Seward, E.A.; Kelly, S. Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms. *Genome Biol.* **2016**, *17*, 226, doi:10.1186/s13059-016-1087-9.
37. Kelly, S. The Amount of Nitrogen Used for Photosynthesis Modulates Molecular Evolution in Plants. *Mol. Biol. Evol.* **2018**, *35*, 1616–1625, doi:10.1093/molbev/msy043.

38. Smarda, P.; Hejzman, M.; Brezinova, A.; Horova, L.; Steigerova, H.; Zedek, F.; Bures, P.; Hejzmanova, P.; Schellberg, J. Effect of phosphorus availability on the selection of species with different ploidy levels and genome sizes in a long-term grassland fertilization experiment. *New Phytol.* **2013**, *200*, 911–921, doi:10.1111/nph.12399.
39. Hunt, R.C.; Simhadri, V.L.; Iandoli, M.; Sauna, Z.E.; Kimchi-Sarfaty, C. Exposing synonymous mutations. *Trends Genet.* **2014**, *30*, 308–321, doi:10.1016/j.tig.2014.04.006.
40. Quintales, L.; Soriano, I.; Vazquez, E.; Segurado, M.; Antequera, F. A species-specific nucleosomal signature defines a periodic distribution of amino acids in proteins. *Open Biol.* **2015**, *5*, 140218, doi:10.1098/rsob.140218.
41. Babbitt, G.A.; Alawad, M.A.; Schulze, K.V.; Hudson, A.O. Synonymous codon bias and functional constraint on GC3-related DNA backbone dynamics in the prokaryotic nucleoid. *Nucleic Acids Res.* **2014**, *42*, 10915–10926, doi:10.1093/nar/gku811.
42. Bailey, S.F.; Hinz, A.; Kassen, R. Adaptive synonymous mutations in an experimentally evolved *Pseudomonas fluorescens* population. *Nat. Commun.* **2014**, *5*, 4076, doi:10.1038/ncomms5076.
43. Taylor, F.J.; Coates, D. The code within the codons. *Biosystems* **1989**, *22*, 177–187.
44. Woese, C.R.; Dugre, D.H.; Saxinger, W.C.; Dugre, S.A. The molecular basis for the genetic code. *Proc. Natl. Acad. Sci. USA* **1966**, *55*, 966–974, doi:10.1073/pnas.55.4.966.
45. Prilusky, J.; Bibi, E. Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6662–6666, doi:10.1073/pnas.0902029106.
46. Panda, A.; Podder, S.; Chakraborty, S.; Ghosh, T.C. GC-made protein disorder sheds new light on vertebrate evolution. *Genomics* **2014**, *104*, 530–537, doi:10.1016/j.ygeno.2014.09.003.
47. Brbic, M.; Warnecke, T.; Krisko, A.; Supek, F. Global Shifts in Genome and Proteome Composition Are Very Tightly Coupled. *Genome Biol. Evol.* **2015**, *7*, 1519–1532, doi:10.1093/gbe/evv088.
48. Goncarenco, A.; Berezovsky, I.N. The fundamental tradeoff in genomes and proteomes of prokaryotes established by the genetic code, codon entropy, and physics of nucleic acids and proteins. *Biol. Direct* **2014**, *9*, 29, doi:10.1186/s13062-014-0029-2.
49. Warnecke, T.; Weber, C.C.; Hurst, L.D. Why there is more to protein evolution than protein function: Splicing, nucleosomes and dual-coding sequence. *Biochem. Soc. Trans.* **2009**, *37*, 756–761, doi:10.1042/BST0370756.
50. Fontrodona, N.; Aube, F.; Claude, J.B.; Polveche, H.; Lemaire, S.; Tranchevent, L.C.; Modolo, L.; Mortreux, F.; Bourgeois, C.F.; Auboeuf, D. Interplay between coding and exonic splicing regulatory sequences. *Genome Res.* **2019**, *29*, 711–722, doi:10.1101/gr.241315.118.
51. Faure, G.; Ogurtsov, A.Y.; Shabalina, S.A.; Koonin, E.V. Adaptation of mRNA structure to control protein folding. *RNA Biol.* **2017**, *14*, 1649–1654, doi:10.1080/15476286.2017.1349047.
52. Brunak, S.; Engelbrecht, J. Protein structure and the sequential structure of mRNA: Alpha-helix and beta-sheet signals at the nucleotide level. *Proteins* **1996**, *25*, 237–252, doi:10.1002/(SICI)1097-0134(199606)25:2<237::AID-PROT9>3.0.CO;2-E.
53. Trifonov, E.N.; Volkovich, Z.; Frenkel, Z.M. Multiple levels of meaning in DNA sequences, and one more. *Ann. N. Y. Acad. Sci.* **2012**, *1267*, 35–38, doi:10.1111/j.1749-6632.2012.06589.x.
54. Ponce de Leon, M.; de Miranda, A.B.; Alvarez-Valin, F.; Carels, N. The Purine Bias of Coding Sequences is Determined by Physicochemical Constraints on Proteins. *Bioinform. Biol. Insights* **2014**, *8*, 93–108, doi:10.4137/BBL.S13161.
55. Tagami, S.; Attwater, J.; Holliger, P. Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat. Chem.* **2017**, *9*, 325–332, doi:10.1038/nchem.2739.
56. Kun, A.; Radvanyi, A. The evolution of the genetic code: Impasses and challenges. *Biosystems* **2018**, *164*, 217–225, doi:10.1016/j.biosystems.2017.10.006.
57. Koonin, E.V.; Novozhilov, A.S. Origin and Evolution of the Universal Genetic Code. *Annu. Rev. Genet.* **2017**, *51*, 45–62, doi:10.1146/annurev-genet-120116-024713.
58. Gulik, P.T. On the Origin of Sequence. *Life* **2015**, *5*, 1629–1637, doi:10.3390/life5041629.
59. Grosjean, H.; Westhof, E. An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res.* **2016**, *44*, 8020–8040, doi:10.1093/nar/gkw608.
60. Szostak, J.W. On the origin of life. *Medicina* **2016**, *76*, 199–203.

61. Usui, K.; Ichihashi, N.; Yomo, T. A design principle for a single-stranded RNA genome that replicates with less double-strand formation. *Nucleic Acids Res.* **2015**, *43*, 8033–8043, doi:10.1093/nar/gkv742.
62. Bansho, Y.; Ichihashi, N.; Kazuta, Y.; Matsuura, T.; Suzuki, H.; Yomo, T. Importance of parasite RNA species repression for prolonged translation-coupled RNA self-replication. *Chem. Biol.* **2012**, *19*, 478–487, doi:10.1016/j.chembiol.2012.01.019.
63. Eigen, M.; Biebricher, C.K.; Gebinoga, M.; Gardiner, W.C. The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. *Biochemistry* **1991**, *30*, 11005–11018, doi:10.1021/bi00110a001.
64. Carter, C.W., Jr.; Wills, P.R. Interdependence, Reflexivity, Fidelity, Impedance Matching, and the Evolution of Genetic Coding. *Mol. Biol. Evol.* **2018**, *35*, 269–286, doi:10.1093/molbev/msx265.
65. Saad, N. A ribonucleopeptide world at the origin of life: Co-evolution of RNA. *J. Syst. Evol.* **2018**, *56*, 1–13, doi:10.1111/jse.12287.
66. Francis, B.R. The Hypothesis that the Genetic Code Originated in Coupled Synthesis of Proteins and the Evolutionary Predecessors of Nucleic Acids in Primitive Cells. *Life* **2015**, *5*, 467–505, doi:10.3390/life5010467.
67. Gounaris, Y. An evolutionary theory based on a protein-mRNA co-synthesis hypothesis. *J. Biol. Res. Thessalon.* **2011**, *15*, 3–16.
68. Gordon, K.H. Were RNA replication and translation directly coupled in the RNA (+protein?) World? *J. Theor. Biol.* **1995**, *173*, 179–193, doi:10.1006/jtbi.1995.0054.
69. Zamft, B.; Bintu, L.; Ishibashi, T.; Bustamante, C. Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8948–8953, doi:10.1073/pnas.1205063109.
70. Chen, X.; Yang, J.R.; Zhang, J. Nascent RNA folding mitigates transcription-associated mutagenesis. *Genome Res.* **2016**, *26*, 50–59, doi:10.1101/gr.195164.115.
71. Proshkin, S.; Rahmouni, A.R.; Mironov, A.; Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **2010**, *328*, 504–508, doi:10.1126/science.1184939.
72. Auboeuf, D. Alternative mRNA processing sites decrease genetic variability while increasing functional diversity. *Transcription* **2018**, *9*, 75–87, doi:10.1080/21541264.2017.1373891.
73. Morgens, D.W. The protein invasion: A broad review on the origin of the translational system. *J. Mol. Evol.* **2013**, *77*, 185–196, doi:10.1007/s00239-013-9592-x.
74. Altstein, A.D. The progene hypothesis: The nucleoprotein world and how life began. *Biol. Direct* **2015**, *10*, 67, doi:10.1186/s13062-015-0096-z.
75. Sutherland, J.D.; Blackburn, J.M. Killing two birds with one stone: A chemically plausible scheme for linked nucleic acid replication and coded peptide synthesis. *Chem. Biol.* **1997**, *4*, 481–488, doi:10.1016/s1074-5521(97)90318-5.
76. Attwater, J.; Raguram, A.; Morgunov, A.S.; Gianni, E.; Holliger, P. Ribozyme-catalysed RNA synthesis using triplet building blocks. *Elife* **2018**, *7*, doi:10.7554/eLife.35255.
77. Maizels, N.; Weiner, A.M. Phylogeny from function: Evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6729–6734, doi:10.1073/pnas.91.15.6729.
78. Zeldovich, K.B.; Berezovsky, I.N.; Shakhnovich, E.I. Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Comput. Biol.* **2007**, *3*, e5, doi:10.1371/journal.pcbi.0030005.
79. Lemaire, S.; Fontrodona, N.; Aube, F.; Claude, J.B.; Polveche, H.; Modolo, L.; Bourgeois, C.F.; Mortreux, F.; Auboeuf, D. Characterizing the interplay between gene nucleotide composition bias and splicing. *Genome Biol.* **2019**, *20*, 259, doi:10.1186/s13059-019-1869-y.
80. Granold, M.; Hajjeva, P.; Tosa, M.I.; Irimie, F.D.; Moosmann, B. Modern diversification of the amino acid repertoire driven by oxygen. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 41–46, doi:10.1073/pnas.1717100115.
81. Chowdhury, K.; Kumar, S.; Sharma, T.; Sharma, A.; Bhagat, M.; Kamai, A.; Ford, B.M.; Asthana, S.; Mandal, C.C. Presence of a consensus DNA motif at nearby DNA sequence of the mutation susceptible CG nucleotides. *Gene* **2018**, *639*, 85–95, doi:10.1016/j.gene.2017.10.001.
82. Szpiech, Z.A.; Strauli, N.B.; White, K.A.; Ruiz, D.G.; Jacobson, M.P.; Barber, D.L.; Hernandez, R.D. Prominent features of the amino acid mutation landscape in cancer. *PLoS ONE* **2017**, *12*, e0183273, doi:10.1371/journal.pone.0183273.

83. Tsuber, V.; Kadamov, Y.; Brautigam, L.; Berglund, U.W.; Helleday, T. Mutations in Cancer Cause Gain of Cysteine, Histidine, and Tryptophan at the Expense of a Net Loss of Arginine on the Proteome Level. *Biomolecules* **2017**, *7*, doi:10.3390/biom7030049.
84. Son, H.; Kang, H.; Kim, H.S.; Kim, S. Somatic mutation driven codon transition bias in human cancer. *Sci. Rep.* **2017**, *7*, 14204, doi:10.1038/s41598-017-14543-1.
85. Tan, H.; Bao, J.; Zhou, X. Genome-wide mutational spectra analysis reveals significant cancer-specific heterogeneity. *Sci. Rep.* **2015**, *5*, 12566, doi:10.1038/srep12566.
86. Suarez-Villagran, M.Y.; Azevedo, R.B.R.; Miller, J.H., Jr. Influence of Electron-Holes on DNA Sequence-Specific Mutation Rates. *Genome Biol. Evol.* **2018**, *10*, 1039–1047, doi:10.1093/gbe/evy060.
87. Bender, A.; Hajieva, P.; Moosmann, B. Adaptive antioxidant methionine accumulation in respiratory chain complexes explains the use of a deviant genetic code in mitochondria. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16496–16501, doi:10.1073/pnas.0802779105.
88. Vetsigian, K.; Woese, C.; Goldenfeld, N. Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10696–10701, doi:10.1073/pnas.0603780103.
89. Hlevnjak, M.; Polyansky, A.A.; Zagrovic, B. Sequence signatures of direct complementarity between mRNAs and cognate proteins on multiple levels. *Nucleic Acids Res.* **2012**, *40*, 8874–8882, doi:10.1093/nar/gks679.
90. Polyansky, A.A.; Zagrovic, B. Evidence of direct complementary interactions between messenger RNAs and their cognate proteins. *Nucleic Acids Res.* **2013**, *41*, 8434–8443, doi:10.1093/nar/gkt618.
91. Zagrovic, B.; Bartonek, L.; Polyansky, A.A. RNA-protein interactions in an unstructured context. *FEBS Lett.* **2018**, *592*, 2901–2916, doi:10.1002/1873-3468.13116.
92. Cadena-Nava, R.D.; Comas-Garcia, M.; Garmann, R.F.; Rao, A.L.; Knobler, C.M.; Gelbart, W.M. Self-assembly of viral capsid protein and RNA molecules of different sizes: Requirement for a specific high protein/RNA mass ratio. *J. Virol.* **2012**, *86*, 3318–3326, doi:10.1128/JVI.06566-11.
93. Wong, J.T.; Ng, S.K.; Mat, W.K.; Hu, T.; Xue, H. Coevolution Theory of the Genetic Code at Age Forty: Pathway to Translation and Synthetic Life. *Life* **2016**, *6*, 12, doi:10.3390/life6010012.
94. Di Giulio, M. The aminoacyl-tRNA synthetases had only a marginal role in the origin of the organization of the genetic code: Evidence in favor of the coevolution theory. *J. Theor. Biol.* **2017**, *432*, 14–24, doi:10.1016/j.jtbi.2017.08.005.
95. Copley, S.D.; Smith, E.; Morowitz, H.J. A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 4442–4447, doi:10.1073/pnas.0501049102.
96. van der Knaap, J.A.; Verrijzer, C.P. Undercover: Gene control by metabolites and metabolic enzymes. *Genes Dev.* **2016**, *30*, 2345–2369, doi:10.1101/gad.289140.116.
97. Schvartzman, J.M.; Thompson, C.B.; Finley, L.W.S. Metabolic regulation of chromatin modifications and gene expression. *J. Cell Biol.* **2018**, *217*, 2247–2259, doi:10.1083/jcb.201803061.
98. Reid, M.A.; Dai, Z.; Locasale, J.W. The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nat. Cell Biol.* **2017**, *19*, 1298–1306, doi:10.1038/ncb3629.
99. Varela, F.G.; Maturana, H.R.; Uribe, R. Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems* **1974**, *5*, 187–196, doi:10.1016/0303-2647(74)90031-8.
100. Maturana, H.; Varela, F. *The Tree of Knowledge*; Shambhala Publications: Boston, MA, USA, 1998.
101. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523, doi:10.1007/bf00623322.
102. Eigen, M.; Schuster, P. Stages of emerging life—Five principles of early organization. *J. Mol. Evol.* **1982**, *19*, 47–61, doi:10.1007/bf02100223.
103. Cairns, J.; Overbaugh, J.; Miller, S. The origin of mutants. *Nature* **1988**, *335*, 142–145, doi:10.1038/335142a0.
104. Wright, B.E. A biochemical mechanism for nonrandom mutations and evolution. *J. Bacteriol.* **2000**, *182*, 2993–3001, doi:10.1128/jb.182.11.2993-3001.2000.
105. Correa, R.; Thornton, P.C.; Rosenberg, S.M.; Hastings, P.J. Oxygen and RNA in stress-induced mutation. *Curr. Genet.* **2018**, *64*, 769–776, doi:10.1007/s00294-017-0801-9.
106. Sebastian, R.; Oberdoerffer, P. Transcription-associated events affecting genomic integrity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*, doi:10.1098/rstb.2016.0288.
107. Wang, G.; Vasquez, K.M. Effects of Replication and Transcription on DNA Structure-Related Genetic Instability. *Genes* **2017**, *8*, 17, doi:10.3390/genes8010017.

108. Merrikh, H. Spatial and Temporal Control of Evolution through Replication-Transcription Conflicts. *Trends Microbiol.* **2017**, *25*, 515–521, doi:10.1016/j.tim.2017.01.008.
109. Chen, Y.H.; Keegan, S.; Kahli, M.; Tonzi, P.; Fenyó, D.; Huang, T.T.; Smith, D.J. Transcription shapes DNA replication initiation and termination in human cells. *Nat. Struct. Mol. Biol.* **2019**, *26*, 67–77, doi:10.1038/s41594-018-0171-0.
110. Duret, L.; Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genom. Hum. Genet.* **2009**, *10*, 285–311, doi:10.1146/annurev-genom-082908-150001.
111. Long, H.; Sung, W.; Kucukyildirim, S.; Williams, E.; Miller, S.F.; Guo, W.; Patterson, C.; Gregory, C.; Strauss, C.; Stone, C.; et al. Evolutionary determinants of genome-wide nucleotide composition. *Nat. Ecol. Evol.* **2018**, *2*, 237–240, doi:10.1038/s41559-017-0425-y.
112. Pozzoli, U.; Menozzi, G.; Fumagalli, M.; Cereda, M.; Comi, G.P.; Cagliani, R.; Bresolin, N.; Sironi, M. Both selective and neutral processes drive GC content evolution in the human genome. *BMC Evol. Biol.* **2008**, *8*, 99, doi:10.1186/1471-2148-8-99.
113. Kudla, G.; Helwak, A.; Lipinski, L. Gene conversion and GC-content evolution in mammalian Hsp70. *Mol. Biol. Evol.* **2004**, *21*, 1438–1444, doi:10.1093/molbev/msh146.
114. Yin, Y.; Yang, L.; Zheng, G.; Gu, C.; Yi, C.; He, C.; Gao, Y.Q.; Zhao, X.S. Dynamics of spontaneous flipping of a mismatched base in DNA duplex. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8043–8048, doi:10.1073/pnas.1400667111.
115. Sankar, T.S.; Wastuwidyaningtyas, B.D.; Dong, Y.; Lewis, S.A.; Wang, J.D. The nature of mutations induced by replication-transcription collisions. *Nature* **2016**, *535*, 178–181, doi:10.1038/nature18316.
116. Sueoka, N. Wide intra-genomic G+C heterogeneity in human and chicken is mainly due to strand-symmetric directional mutation pressures: dGTP-oxidation and symmetric cytosine-deamination hypotheses. *Gene* **2002**, *300*, 141–154, doi:10.1016/s0378-1119(02)01046-6.
117. Chen, C.L.; Rappailles, A.; Duquenne, L.; Huvet, M.; Guilbaud, G.; Farinelli, L.; Audit, B.; d'Aubenton-Carafa, Y.; Arneodo, A.; Hyrien, O.; et al. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res.* **2010**, *20*, 447–457, doi:10.1101/gr.098947.109.
118. Kenigsberg, E.; Yehuda, Y.; Marjavaara, L.; Keszthelyi, A.; Chabes, A.; Tanay, A.; Simon, I. The mutation spectrum in genomic late replication domains shapes mammalian GC content. *Nucleic Acids Res.* **2016**, *44*, 4222–4232, doi:10.1093/nar/gkw268.
119. Gul, I.S.; Staal, J.; Hulpiau, P.; De Keuckelaere, E.; Kamm, K.; Deroo, T.; Sanders, E.; Staes, K.; Driège, Y.; Saey, Y.; et al. GC Content of Early Metazoan Genes and Its Impact on Gene Expression Levels in Mammalian Cell Lines. *Genome Biol. Evol.* **2018**, *10*, 909–917, doi:10.1093/gbe/evy040.
120. Rao, Y.S.; Wang, Z.F.; Chai, X.W.; Wu, G.Z.; Zhou, M.; Nie, Q.H.; Zhang, X.Q. Selection for the compactness of highly expressed genes in *Gallus gallus*. *Biol. Direct* **2010**, *5*, 35, doi:10.1186/1745-6150-5-35.
121. Tilgner, H.; Knowles, D.G.; Johnson, R.; Davis, C.A.; Chakraborty, S.; Djebali, S.; Curado, J.; Snyder, M.; Gingeras, T.R.; Guigo, R. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **2012**, *22*, 1616–1625, doi:10.1101/gr.134445.111.
122. Quandt, E.M.; Traverse, C.C.; Ochman, H. Local genic base composition impacts protein production and cellular fitness. *PeerJ* **2018**, *6*, e4286, doi:10.7717/peerj.4286.
123. Gorochoowski, T.E.; Ignatova, Z.; Bovenberg, R.A.; Roubos, J.A. Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* **2015**, *43*, 3022–3032, doi:10.1093/nar/gkv199.
124. Kudla, G.; Murray, A.W.; Tollervey, D.; Plotkin, J.B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **2009**, *324*, 255–258, doi:10.1126/science.1170160.
125. Rao, Y.; Wang, Z.; Chai, X.; Nie, Q.; Zhang, X. Hydrophobicity and aromaticity are primary factors shaping variation in amino acid usage of chicken proteome. *PLoS ONE* **2014**, *9*, e110381, doi:10.1371/journal.pone.0110381.
126. Du, M.Z.; Zhang, C.; Wang, H.; Liu, S.; Wei, W.; Guo, F.B. The GC Content as a Main Factor Shaping the Amino Acid Usage During Bacterial Evolution Process. *Front. Microbiol.* **2018**, *9*, 2948, doi:10.3389/fmicb.2018.02948.
127. Gao, N.; Lu, G.; Lercher, M.J.; Chen, W.H. Selection for energy efficiency drives strand-biased gene distribution in prokaryotes. *Sci. Rep.* **2017**, *7*, 10572, doi:10.1038/s41598-017-11159-3.

128. Hull, R.M.; Cruz, C.; Jack, C.V.; Houseley, J. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biol.* **2017**, *15*, e2001333, doi:10.1371/journal.pbio.2001333.
129. Ambriz-Avina, V.; Yasbin, R.E.; Robleto, E.A.; Pedraza-Reyes, M. Role of Base Excision Repair (BER) in Transcription-associated Mutagenesis of Nutritionally Stressed Nongrowing *Bacillus subtilis* Cell Subpopulations. *Curr. Microbiol.* **2016**, *73*, 721–726, doi:10.1007/s00284-016-1122-9.
130. Shewaramani, S.; Finn, T.J.; Leahy, S.C.; Kassen, R.; Rainey, P.B.; Moon, C.D. Anaerobically Grown *Escherichia coli* Has an Enhanced Mutation Rate and Distinct Mutational Spectra. *PLoS Genet.* **2017**, *13*, e1006570, doi:10.1371/journal.pgen.1006570.
131. Liu, H.; Zhang, J. Yeast Spontaneous Mutation Rate and Spectrum Vary with Environment. *Curr. Biol.* **2019**, *29*, 1584–1591.e3, doi:10.1016/j.cub.2019.03.054.
132. Maharjan, R.P.; Ferenci, T. A shifting mutational landscape in 6 nutritional states: Stress-induced mutagenesis as a series of distinct stress input-mutation output relationships. *PLoS Biol.* **2017**, *15*, e2001477, doi:10.1371/journal.pbio.2001477.
133. Chu, X.L.; Zhang, B.W.; Zhang, Q.G.; Zhu, B.R.; Lin, K.; Zhang, D.Y. Temperature responses of mutation rate and mutational spectrum in an *Escherichia coli* strain and the correlation with metabolic rate. *BMC Evol. Biol.* **2018**, *18*, 126, doi:10.1186/s12862-018-1252-8.
134. Matsuba, C.; Ostrow, D.G.; Salomon, M.P.; Tolani, A.; Baer, C.F. Temperature, stress and spontaneous mutation in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *Biol. Lett.* **2013**, *9*, 20120334, doi:10.1098/rsbl.2012.0334.
135. Rogozin, I.B.; Pavlov, Y.I.; Goncarencu, A.; De, S.; Lada, A.G.; Poliakov, E.; Panchenko, A.R.; Cooper, D.N. Mutational signatures and mutable motifs in cancer genomes. *Brief. Bioinform.* **2018**, *19*, 1085–1101, doi:10.1093/bib/bbx049.
136. White, K.A.; Ruiz, D.G.; Szpiech, Z.A.; Strauli, N.B.; Hernandez, R.D.; Jacobson, M.P.; Barber, D.L. Cancer-associated arginine-to-histidine mutations confer a gain in pH sensing to mutant proteins. *Sci. Signal.* **2017**, *10*, doi:10.1126/scisignal.aam9931.
137. Saier, M.H., Jr.; Kukita, C.; Zhang, Z. Transposon-mediated directed mutation in bacteria and eukaryotes. *Front. Biosci. (Landmark Ed.)* **2017**, *22*, 1458–1468.
138. Newman, A.G.; Bessa, P.; Tarabykin, V.; Singh, P.B. Activity-DEPENDent Transposition. *EMBO Rep.* **2017**, *18*, 346–348, doi:10.15252/embr.201643797.
139. Vandecraen, J.; Monsieurs, P.; Mergeay, M.; Leys, N.; Aertsen, A.; Van Houdt, R. Zinc-Induced Transposition of Insertion Sequence Elements Contributes to Increased Adaptability of *Cupriavidus metallidurans*. *Front. Microbiol.* **2016**, *7*, 359, doi:10.3389/fmicb.2016.00359.
140. Grandbastien, M.A. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochim. Biophys. Acta* **2015**, *1849*, 403–416, doi:10.1016/j.bbagr.2014.07.017.
141. Miousse, I.R.; Chalbot, M.C.; Lumen, A.; Ferguson, A.; Kavouras, I.G.; Koturbash, I. Response of transposable elements to environmental stressors. *Mutat. Res. Rev. Mutat. Res.* **2015**, *765*, 19–39, doi:10.1016/j.mrrev.2015.05.003.
142. Rada-Iglesias, A.; Grosveld, F.G.; Papantonis, A. Forces driving the three-dimensional folding of eukaryotic genomes. *Mol. Syst. Biol.* **2018**, *14*, e8214, doi:10.15252/msb.20188214.
143. Meyer, S.; Reverchon, S.; Nasser, W.; Muskhelishvili, G. Chromosomal organization of transcription: In a nutshell. *Curr. Genet.* **2018**, *64*, 555–565, doi:10.1007/s00294-017-0785-5.
144. Lin, Y.H.; Forman-Kay, J.D.; Chan, H.S. Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. *Biochemistry* **2018**, *57*, 2499–2508, doi:10.1021/acs.biochem.8b00058.
145. Erdel, F.; Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophys. J.* **2018**, *114*, 2262–2270, doi:10.1016/j.bpj.2018.03.011.
146. Rieder, D.; Trajanoski, Z.; McNally, J.G. Transcription factories. *Front. Genet.* **2012**, *3*, 221, doi:10.3389/fgene.2012.00221.
147. Gu, Z.; Jin, K.; Crabbe, M.J.C.; Zhang, Y.; Liu, X.; Huang, Y.; Hua, M.; Nan, P.; Zhang, Z.; Zhong, Y. Enrichment analysis of Alu elements with different spatial chromatin proximity in the human genome. *Protein Cell* **2016**, *7*, 250–266, doi:10.1007/s13238-015-0240-7.
148. Sundaram, V.; Cheng, Y.; Ma, Z.; Li, D.; Xing, X.; Edge, P.; Snyder, M.P.; Wang, T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **2014**, *24*, 1963–1976, doi:10.1101/gr.168872.113.

149. Puc, J.; Aggarwal, A.K.; Rosenfeld, M.G. Physiological functions of programmed DNA breaks in signal-induced transcription. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 471–476, doi:10.1038/nrm.2017.43.
150. Kaiser, V.B.; Semple, C.A. Chromatin loop anchors are associated with genome instability in cancer and recombination hotspots in the germline. *Genome Biol.* **2018**, *19*, 101, doi:10.1186/s13059-018-1483-4.
151. Schwer, B.; Wei, P.C.; Chang, A.N.; Kao, J.; Du, Z.; Meyers, R.M.; Alt, F.W. Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2258–2263, doi:10.1073/pnas.1525564113.
152. Roychowdhury, T.; Abyzov, A. Chromatin organization modulates the origin of heritable structural variations in human genome. *Nucleic Acids Res.* **2019**, *47*, 2766–2777, doi:10.1093/nar/gkz103.
153. Mellor, J.; Woloszczuk, R.; Howe, F.S. The Interleaved Genome. *Trends Genet.* **2016**, *32*, 57–71, doi:10.1016/j.tig.2015.10.006.
154. Hurst, L.D.; Pal, C.; Lercher, M.J. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **2004**, *5*, 299–310, doi:10.1038/nrg1319.
155. Yin, Y.; Zhang, H.; Olman, V.; Xu, Y. Genomic arrangement of bacterial operons is constrained by biological pathways encoded in the genome. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 6310–6315, doi:10.1073/pnas.0911237107.
156. Nutzmans, H.W.; Huang, A.; Osbourn, A. Plant metabolic clusters—From genetics to genomics. *New Phytol.* **2016**, *211*, 771–789, doi:10.1111/nph.13981.
157. Gordon, A.J.; Satory, D.; Halliday, J.A.; Herman, C. Lost in transcription: Transient errors in information transfer. *Curr. Opin. Microbiol.* **2015**, *24*, 80–87, doi:10.1016/j.mib.2015.01.010.
158. Bradley, C.C.; Gordon, A.J.E.; Halliday, J.A.; Herman, C. Transcription fidelity: New paradigms in epigenetic inheritance, genome instability and disease. *DNA Repair* **2019**, *81*, 102652, doi:10.1016/j.dnarep.2019.102652.
159. Reid-Bayliss, K.S.; Loeb, L.A. Accurate RNA consensus sequencing for high-fidelity detection of transcriptional mutagenesis-induced epimutations. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9415–9420, doi:10.1073/pnas.1709166114.
160. Xu, L.; Wang, W.; Chong, J.; Shin, J.H.; Xu, J.; Wang, D. RNA polymerase II transcriptional fidelity control and its functional interplay with DNA modifications. *Crit. Rev. Biochem. Mol. Biol.* **2015**, *50*, 503–519, doi:10.3109/10409238.2015.1087960.
161. Morreall, J.; Kim, A.; Liu, Y.; Degtyareva, N.; Weiss, B.; Doetsch, P.W. Evidence for Retromutagenesis as a Mechanism for Adaptive Mutation in Escherichia coli. *PLoS Genet.* **2015**, *11*, e1005477, doi:10.1371/journal.pgen.1005477.
162. Sekowska, A.; Wendel, S.; Fischer, E.C.; Norholm, M.H.H.; Danchin, A. Generation of mutation hotspots in ageing bacterial colonies. *Sci. Rep.* **2016**, *6*, 2, doi:10.1038/s41598-016-0005-4.
163. Williamson, A.K.; Zhu, Z.; Yuan, Z.M. Epigenetic mechanisms behind cellular sensitivity to DNA damage. *Cell Stress* **2018**, *2*, 176–180, doi:10.15698/cst2018.07.145.
164. Gonzalez-Perez, A.; Sabarinathan, R.; Lopez-Bigas, N. Local Determinants of the Mutational Landscape of the Human Genome. *Cell* **2019**, *177*, 101–114, doi:10.1016/j.cell.2019.02.051.
165. Elfman, J.; Li, H. Chimeric RNA in Cancer and Stem Cell Differentiation. *Stem Cells Int.* **2018**, *2018*, 3178789, doi:10.1155/2018/3178789.
166. Keskin, H.; Shen, Y.; Huang, F.; Patel, M.; Yang, T.; Ashley, K.; Mazin, A.V.; Storici, F. Transcript-RNA-templated DNA recombination and repair. *Nature* **2014**, *515*, 436–439, doi:10.1038/nature13682.
167. Yang, Y.G.; Qi, Y. RNA-directed repair of DNA double-strand breaks. *DNA Repair* **2015**, *32*, 82–85, doi:10.1016/j.dnarep.2015.04.017.
168. Shapiro, J.A. Living Organisms Author Their Read-Write Genomes in Evolution. *Biology* **2017**, *6*, 42, doi:10.3390/biology6040042.
169. Khanduja, J.S.; Calvo, I.A.; Joh, R.I.; Hill, I.T.; Motamedi, M. Nuclear Noncoding RNAs and Genome Stability. *Mol. Cell* **2016**, *63*, 7–20, doi:10.1016/j.molcel.2016.06.011.
170. Auboeuf, D. Putative RNA-Directed Adaptive Mutations in Cancer Evolution. *Transcription* **2016**, *7*, 164–187, doi:10.1080/21541264.2016.1221491.
171. Auboeuf, D. Genome evolution is driven by gene expression-generated biophysical constraints through RNA-directed genetic variation: A hypothesis. *Bioessays* **2017**, *39*, doi:10.1002/bies.201700069.

172. Ibrahim, F.; Maragkakis, M.; Alexiou, P.; Mourelatos, Z. Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis. *Nat. Struct. Mol. Biol.* **2018**, *25*, 302–310, doi:10.1038/s41594-018-0042-8.
173. Ikeuchi, K.; Izawa, T.; Inada, T. Recent Progress on the Molecular Mechanism of Quality Controls Induced by Ribosome Stalling. *Front. Genet.* **2018**, *9*, 743, doi:10.3389/fgene.2018.00743.
174. Karimi, J.; Goodarzi, M.T.; Tavilani, H.; Khodadadi, I.; Amiri, I. Increased receptor for advanced glycation end products in spermatozoa of diabetic men and its association with sperm nuclear DNA fragmentation. *Andrologia* **2012**, *44*, 280–286, doi:10.1111/j.1439-0272.2011.01178.x.
175. Sies, H. Strategies of antioxidant defense. *Eur. J. Biochem.* **1993**, *215*, 213–219, doi:10.1111/j.1432-1033.1993.tb18025.x.
176. Sales, V.M.; Ferguson-Smith, A.C.; Patti, M.E. Epigenetic Mechanisms of Transmission of Metabolic Disease across Generations. *Cell Metab.* **2017**, *25*, 559–571, doi:10.1016/j.cmet.2017.02.016.
177. Vanhees, K.; Vonhogen, I.G.; van Schooten, F.J.; Godschalk, R.W. You are what you eat, and so are your children: The impact of micronutrients on the epigenetic programming of offspring. *Cell. Mol. Life Sci.* **2014**, *71*, 271–285, doi:10.1007/s00018-013-1427-9.
178. da Silveira, J.C.; de Avila, A.; Garrett, H.L.; Bruemmer, J.E.; Winger, Q.A.; Bouma, G.J. Cell-secreted vesicles containing microRNAs as regulators of gamete maturation. *J. Endocrinol.* **2018**, *236*, R15–R27, doi:10.1530/JOE-17-0200.
179. Chen, Q.; Yan, M.; Cao, Z.; Li, X.; Zhang, Y.; Shi, J.; Feng, G.H.; Peng, H.; Zhang, X.; Zhang, Y.; et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **2016**, *351*, 397–400, doi:10.1126/science.aad7977.
180. Chen, Q.; Yan, W.; Duan, E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat. Rev. Genet.* **2016**, *17*, 733–743, doi:10.1038/nrg.2016.106.
181. Danchin, E.; Pocheville, A.; Huneman, P. Early in life effects and heredity: Reconciling neo-Darwinism with neo-Lamarckism under the banner of the inclusive evolutionary synthesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2019**, *374*, 20180113, doi:10.1098/rstb.2018.0113.
182. Klosin, A.; Lehner, B. Mechanisms, timescales and principles of trans-generational epigenetic inheritance in animals. *Curr. Opin. Genet. Dev.* **2016**, *36*, 41–49, doi:10.1016/j.gde.2016.04.001.
183. Horsthemke, B. A critical view on transgenerational epigenetic inheritance in humans. *Nat. Commun.* **2018**, *9*, 2973, doi:10.1038/s41467-018-05445-5.
184. Monk, D.; Mackay, D.J.G.; Eggermann, T.; Maher, E.R.; Riccio, A. Genomic imprinting disorders: Lessons on how genome, epigenome and environment interact. *Nat. Rev. Genet.* **2019**, *20*, 235–248, doi:10.1038/s41576-018-0092-0.
185. Zhang, Y.; Zhang, X.; Shi, J.; Tuorto, F.; Li, X.; Liu, Y.; Liebers, R.; Zhang, L.; Qu, Y.; Qian, J.; et al. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat. Cell. Biol.* **2018**, *20*, 535–540, doi:10.1038/s41556-018-0087-2.
186. Hollick, J.B. Paramutation and related phenomena in diverse species. *Nat. Rev. Genet.* **2017**, *18*, 5–23, doi:10.1038/nrg.2016.115.
187. Bernstein, H.; Bernstein, C.; Michod, R.E. Meiosis as an Evolutionary Adaptation for DNA Repair. In *DNA Repair*; Inna Kruman, IntechOpen 2011; doi:10.5772/25117.
188. Horandl, E.; Speijer, D. How oxygen gave rise to eukaryotic sex. *Proc. Biol. Sci.* **2018**, *285*, doi:10.1098/rspb.2017.2706.
189. Poljsak, B.; Milisav, I.; Lampe, T.; Ostan, I. Reproductive benefit of oxidative damage: An oxidative stress “malevolence”? *Oxidative Med. Cell. Longev.* **2011**, *2011*, 760978, doi:10.1155/2011/760978.
190. Immler, S.; Otto, S.P. The Evolutionary Consequences of Selection at the Haploid Gametic Stage. *Am. Nat.* **2018**, *192*, 241–249, doi:10.1086/698483.
191. Fishman, L.; McIntosh, M. Standard Deviations: The Biological Bases of Transmission Ratio Distortion. *Annu. Rev. Genet.* **2019**, *53*, 347–372, doi:10.1146/annurev-genet-112618-043905.
192. Tock, A.J.; Henderson, I.R. Hotspots for Initiation of Meiotic Recombination. *Front. Genet.* **2018**, *9*, 521, doi:10.3389/fgene.2018.00521.
193. Brachet, E.; Sommermeyer, V.; Borde, V. Interplay between modifications of chromatin and meiotic recombination hotspots. *Biol. Cell* **2012**, *104*, 51–69, doi:10.1111/boc.201100113.
194. Acuna-Hidalgo, R.; Veltman, J.A.; Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* **2016**, *17*, 241, doi:10.1186/s13059-016-1110-1.

195. Goldmann, J.M.; Wong, W.S.; Pinelli, M.; Farrah, T.; Bodian, D.; Stittrich, A.B.; Glusman, G.; Vissers, L.E.; Hoischen, A.; Roach, J.C.; et al. Parent-of-origin-specific signatures of de novo mutations. *Nat. Genet.* **2016**, *48*, 935–939, doi:10.1038/ng.3597.
196. Michaelson, J.J.; Shi, Y.; Gujral, M.; Zheng, H.; Malhotra, D.; Jin, X.; Jian, M.; Liu, G.; Greer, D.; Bhandari, A.; et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* **2012**, *151*, 1431–1442, doi:10.1016/j.cell.2012.11.019.
197. Skinner, M.K.; Guerrero-Bosagna, C.; Haque, M.M. Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations. *Epigenetics* **2015**, *10*, 762–771, doi:10.1080/15592294.2015.1062207.
198. Guerrero-Bosagna, C.; Morisson, M.; Liaubet, L.; Rodenburg, T.B.; de Haas, E.N.; Kostal, L.; Pitel, F. Transgenerational epigenetic inheritance in birds. *Environ. Epigenet.* **2018**, *4*, dvy008, doi:10.1093/eep/dvy008.
199. Wurdinger, T.; Gatsen, N.N.; Balaj, L.; Kaur, B.; Breakefield, X.O.; Pegtel, D.M. Extracellular vesicles and their convergence with viral pathways. *Adv. Virol.* **2012**, *2012*, 767694, doi:10.1155/2012/767694.
200. Koonin, E.V. Evolution of RNA- and DNA-guided antiviral defense systems in prokaryotes and eukaryotes: Common ancestry vs convergence. *Biol. Direct* **2017**, *12*, 5, doi:10.1186/s13062-017-0177-2.
201. Durdevic, Z.; Schaefer, M. Dnmt2 methyltransferases and immunity: An ancient overlooked connection between nucleotide modification and host defense? *Bioessays* **2013**, *35*, 1044–1049, doi:10.1002/bies.201300088.
202. Rechavi, O. Guest list or black list: Heritable small RNAs as immunogenic memories. *Trends Cell Biol.* **2014**, *24*, 212–220, doi:10.1016/j.tcb.2013.10.003.
203. Zheng, Y.; Lorenzo, C.; Beal, P.A. DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. *Nucleic Acids Res.* **2017**, *45*, 3369–3377, doi:10.1093/nar/gkx050.
204. Zhang, X.; Cozen, A.E.; Liu, Y.; Chen, Q.; Lowe, T.M. Small RNA Modifications: Integral to Function and Disease. *Trends Mol. Med.* **2016**, *22*, 1025–1034, doi:10.1016/j.molmed.2016.10.009.
205. Linster, C.L.; Van Schaftingen, E.; Hanson, A.D. Metabolite damage and its repair or pre-emption. *Nat. Chem. Biol.* **2013**, *9*, 72–80, doi:10.1038/nchembio.1141.
206. Mulikdjanian, A.Y.; Junge, W. On the origin of photosynthesis as inferred from sequence analysis. *Photosynth. Res.* **1997**, *51*, 27–42.
207. Wolstencroft, R.D.; Raven, J.A. Photosynthesis: Likelihood of Occurrence and Possibility of Detection on Earth-like Planets. *Icarus* **2002**, *157*, 535–548.
208. Michaelian, K.; Simeonov, A. Fundamental molecules of life are pigments which arose and co-evolved as a response to the thermodynamic imperative of dissipating the prevailing solar spectrum *Biogeosciences* **2015**, *12*, 4913–4937.
209. Degli Esposti, M.; Mentel, M.; Martin, W.; Sousa, F.L. Oxygen Reductases in Alphaproteobacterial Genomes: Physiological Evolution From Low to High Oxygen Environments. *Front. Microbiol.* **2019**, *10*, 499, doi:10.3389/fmicb.2019.00499.
210. Muller, M.; Mentel, M.; van Hellemond, J.J.; Henze, K.; Woehle, C.; Gould, S.B.; Yu, R.Y.; van der Giezen, M.; Tielens, A.G.; Martin, W.F. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 444–495, doi:10.1128/MMBR.05024-11.
211. Forte, E.; Borisov, V.B.; Falabella, M.; Colaco, H.G.; Tinajero-Trejo, M.; Poole, R.K.; Vicente, J.B.; Sarti, P.; Giuffrè, A. The Terminal Oxidase Cytochrome bd Promotes Sulfide-resistant Bacterial Respiration and Growth. *Sci. Rep.* **2016**, *6*, 23788, doi:10.1038/srep23788.
212. Margulis, L.; Chapman, M.; Guerrero, R.; Hall, J. The last eukaryotic common ancestor (LECA): Acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13080–13085, doi:10.1073/pnas.0604985103.
213. Kurland, C.G.; Andersson, S.G. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 786–820, doi:10.1128/mmb.64.4.786-820.2000.
214. Speijer, D. Alternating terminal electron-acceptors at the basis of symbiogenesis: How oxygen ignited eukaryotic evolution. *Bioessays* **2017**, *39*, doi:10.1002/bies.201600174.
215. Raymond, J.; Segre, D. The effect of oxygen on biochemical networks and the evolution of complex life. *Science* **2006**, *311*, 1764–1767, doi:10.1126/science.1118439.

216. Jiang, Y.Y.; Kong, D.X.; Qin, T.; Li, X.; Caetano-Anolles, G.; Zhang, H.Y. The impact of oxygen on metabolic evolution: A chemoinformatic investigation. *PLoS Comput. Biol.* **2012**, *8*, e1002426, doi:10.1371/journal.pcbi.1002426.
217. Desmond, E.; Gribaldo, S. Phylogenomics of sterol synthesis: Insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.* **2009**, *1*, 364–381, doi:10.1093/gbe/evp036.
218. Zhang, X.; Barraza, K.M.; Beauchamp, J.L. Cholesterol provides nonsacrificial protection of membrane lipids from chemical damage at air-water interface. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3255–3260, doi:10.1073/pnas.1722323115.
219. Galea, A.M.; Brown, A.J. Special relationship between sterols and oxygen: Were sterols an adaptation to aerobic life? *Free Radic. Biol. Med.* **2009**, *47*, 880–889, doi:10.1016/j.freeradbiomed.2009.06.027.
220. Deng, Y.; Almsharqi, Z.A. Evolution of cubic membranes as antioxidant defence system. *Interface Focus* **2015**, *5*, 20150012, doi:10.1098/rsfs.2015.0012.
221. Lambowitz, A.M.; Belfort, M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. *Microbiol. Spectr.* **2015**, *3*, MDNA3-0050-2014, doi:10.1128/microbiolspec.MDNA3-0050-2014.
222. de Lange, T. A loopy view of telomere evolution. *Front. Genet.* **2015**, *6*, 321, doi:10.3389/fgene.2015.00321.
223. Coros, C.J.; Piazza, C.L.; Chalamcharla, V.R.; Belfort, M. A mutant screen reveals RNase E as a silencer of group II intron retromobility in *Escherichia coli*. *RNA* **2008**, *14*, 2634–2644, doi:10.1261/rna.1247608.
224. Belfort, M. Mobile self-splicing introns and inteins as environmental sensors. *Curr. Opin. Microbiol.* **2017**, *38*, 51–58, doi:10.1016/j.mib.2017.04.003.
225. Friedman, K.; Heller, A. On the Non-Uniform Distribution of Guanine in Introns of Human Genes: Possible Protection of Exons against Oxidation by Proximal Intron Poly-G Sequences. *J. Phys. Chem. B* **2001**, *105*, 11859–11865.
226. Amit, M.; Donyo, M.; Hollander, D.; Goren, A.; Kim, E.; Gelfman, S.; Lev-Maor, G.; Burstein, D.; Schwartz, S.; Postolsky, B.; et al. Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* **2012**, *1*, 543–556, doi:10.1016/j.celrep.2012.03.013.
227. Enright, H.; Miller, W.J.; Hays, R.; Floyd, R.A.; Hebbel, R.P. Preferential targeting of oxidative base damage to internucleosomal DNA. *Carcinogenesis* **1996**, *17*, 1175–1177, doi:10.1093/carcin/17.5.1175.
228. Colangeli, R.; Haq, A.; Arcus, V.L.; Summers, E.; Magliozzo, R.S.; McBride, A.; Mitra, A.K.; Radjainia, M.; Khajo, A.; Jacobs, W.R., Jr.; et al. The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4414–4418, doi:10.1073/pnas.0810126106.
229. Speijer, D. Birth of the eukaryotes by a set of reactive innovations: New insights force us to relinquish gradual models. *Bioessays* **2015**, *37*, 1268–1276, doi:10.1002/bies.201500107.
230. Ljungman, M.; Hanawalt, P.C. Efficient protection against oxidative DNA damage in chromatin. *Mol. Carcinog.* **1992**, *5*, 264–269.
231. Cannan, W.J.; Tsang, B.P.; Wallace, S.S.; Pederson, D.S. Nucleosomes suppress the formation of double-strand DNA breaks during attempted base excision repair of clustered oxidative damages. *J. Biol. Chem.* **2014**, *289*, 19881–19893, doi:10.1074/jbc.M114.571588.
232. D'Souza, G.; Shitut, S.; Preussger, D.; Yousif, G.; Waschina, S.; Kost, C. Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat. Prod. Rep.* **2018**, *35*, 455–488, doi:10.1039/c8np00009c.
233. Jeltsch, A. Oxygen, epigenetic signaling, and the evolution of early life. *Trends Biochem. Sci.* **2013**, *38*, 172–176, doi:10.1016/j.tibs.2013.02.001.
234. Drinnenberg, I.A.; Berger, F.; Elsasser, S.J.; Andersen, P.R.; Ausio, J.; Bickmore, W.A.; Blackwell, A.R.; Erwin, D.H.; Gahan, J.M.; Gaut, B.S.; et al. EvoChromo: Towards a synthesis of chromatin biology and evolution. *Development* **2019**, *146*, doi:10.1242/dev.178962.
235. Aravind, L.; Burroughs, A.M.; Zhang, D.; Iyer, L.M. Protein and DNA modifications: Evolutionary imprints of bacterial biochemical diversification and geochemistry on the provenance of eukaryotic epigenetics. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016063, doi:10.1101/cshperspect.a016063.
236. Wahl, M.E.; Murray, A.W. Multicellularity makes somatic differentiation evolutionarily stable. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8362–8367, doi:10.1073/pnas.1608278113.
237. Goldsby, H.J.; Knoester, D.B.; Ofria, C.; Kerr, B. The evolutionary origin of somatic cells under the dirty work hypothesis. *PLoS Biol.* **2014**, *12*, e1001858, doi:10.1371/journal.pbio.1001858.
238. Blagojevic, D.P.; Grubor-Lajsic, G.N.; Spasic, M.B. Cold defence responses: The role of oxidative stress. *Front. Biosci. (Schol. Ed.)* **2011**, *3*, 416–427.

239. Speijer, D. Being right on Q: Shaping eukaryotic evolution. *Biochem. J.* **2016**, *473*, 4103–4127, doi:10.1042/BCJ20160647.
240. Oelkrug, R.; Goetze, N.; Meyer, C.W.; Jastroch, M. Antioxidant properties of UCP1 are evolutionarily conserved in mammals and buffer mitochondrial reactive oxygen species. *Free Radic. Biol. Med.* **2014**, *77*, 210–216, doi:10.1016/j.freeradbiomed.2014.09.004.
241. Rowland, L.A.; Bal, N.C.; Periasamy, M. The role of skeletal-muscle-based thermogenic mechanisms in vertebrate endothermy. *Biol. Rev. Camb. Philos. Soc.* **2015**, *90*, 1279–1297, doi:10.1111/brv.12157.
242. Nowack, J.; Giroud, S.; Arnold, W.; Ruf, T. Muscle Non-shivering Thermogenesis and Its Role in the Evolution of Endothermy. *Front. Physiol.* **2017**, *8*, 889, doi:10.3389/fphys.2017.00889.
243. Newman, S.A. Form and function remixed: Developmental physiology in the evolution of vertebrate body plans. *J. Physiol.* **2014**, *592*, 2403–2412, doi:10.1113/jphysiol.2014.271437.
244. Newman, S.A.; Mezentseva, N.V.; Badyaev, A.V. Gene loss, thermogenesis, and the origin of birds. *Ann. N. Y. Acad. Sci.* **2013**, *1289*, 36–47, doi:10.1111/nyas.12090.
245. Sun, W.; Dong, H.; Becker, A.S.; Dapito, D.H.; Modica, S.; Grandl, G.; Opitz, L.; Efthymiou, V.; Straub, L.G.; Sarker, G.; et al. Cold-induced epigenetic programming of the sperm enhances brown adipose tissue activity in the offspring. *Nat. Med.* **2018**, *24*, 1372–1383, doi:10.1038/s41591-018-0102-y.
246. Koonin, E.V. Splendor and misery of adaptation, or the importance of neutral null for understanding evolution. *BMC Biol.* **2016**, *14*, 114, doi:10.1186/s12915-016-0338-2.
247. Il'icheva, I.A.; Khodikov, M.V.; Poptsova, M.S.; Nechipurenko, D.Y.; Nechipurenko, Y.D.; Grokhovskiy, S.L. Structural features of DNA that determine RNA polymerase II core promoter. *BMC Genom.* **2016**, *17*, 973, doi:10.1186/s12864-016-3292-z.
248. Todolli, S.; Perez, P.J.; Clauvelin, N.; Olson, W.K. Contributions of Sequence to the Higher-Order Structures of DNA. *Biophys. J.* **2017**, *112*, 416–426, doi:10.1016/j.bpj.2016.11.017.
249. Travers, A.; Muskhelishvili, G. DNA structure and function. *FEBS J.* **2015**, *282*, 2279–2295, doi:10.1111/febs.13307.
250. Ramirez, F.; Bhardwaj, V.; Arrigoni, L.; Lam, K.C.; Gruning, B.A.; Villaveces, J.; Habermann, B.; Akhtar, A.; Manke, T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **2018**, *9*, 189, doi:10.1038/s41467-017-02525-w.
251. Jabbari, K.; Bernardi, G. An Isochore Framework Underlies Chromatin Architecture. *PLoS ONE* **2017**, *12*, e0168023, doi:10.1371/journal.pone.0168023.
252. Lian, S.; Liu, T.; Jing, S.; Yuan, H.; Zhang, Z.; Cheng, L. Intrachromosomal colocalization strengthens co-expression, co-modification and evolutionary conservation of neighboring genes. *BMC Genom.* **2018**, *19*, 455, doi:10.1186/s12864-018-4844-1.
253. Bessiere, C.; Taha, M.; Petitprez, F.; Vandel, J.; Marin, J.M.; Brehelin, L.; Lebre, S.; Lecellier, C.H. Probing instructions for expression regulation in gene nucleotide compositions. *PLoS Comput. Biol.* **2018**, *14*, e1005921, doi:10.1371/journal.pcbi.1005921.
254. Yin, H.; Wang, G.; Ma, L.; Yi, S.V.; Zhang, Z. What Signatures Dominantly Associate with Gene Age? *Genome Biol. Evol.* **2016**, *8*, 3083–3089, doi:10.1093/gbe/evw216.
255. Fuertes, M.A.; Rodrigo, J.R.; Alonso, C. Do Intron and Coding Sequences of Some Human-Mouse Orthologs Evolve as a Single Unit? *J. Mol. Evol.* **2016**, *82*, 247–250, doi:10.1007/s00239-016-9746-8.
256. Polyansky, A.A.; Hlevnjak, M.; Zagrovic, B. Analogue encoding of physicochemical properties of proteins in their cognate messenger RNAs. *Nat. Commun.* **2013**, *4*, 2784, doi:10.1038/ncomms3784.
257. Keene, J.D. RNA regulons: Coordination of post-transcriptional events. *Nat. Rev. Genet.* **2007**, *8*, 533–543, doi:10.1038/nrg2111.
258. Cascarina, S.M.; Ross, E.D. Proteome-scale relationships between local amino acid composition and protein fates and functions. *PLoS Comput. Biol.* **2018**, *14*, e1006256, doi:10.1371/journal.pcbi.1006256.
259. Wang, T.; Tang, H. The physical characteristics of human proteins in different biological functions. *PLoS ONE* **2017**, *12*, e0176234, doi:10.1371/journal.pone.0176234.
260. Karathia, H.; Kingsford, C.; Girvan, M.; Hannenhalli, S. A pathway-centric view of spatial proximity in the 3D nucleome across cell lines. *Sci. Rep.* **2016**, *6*, 39279, doi:10.1038/srep39279.
261. Paz, A.; Frenkel, S.; Snir, S.; Kirzhner, V.; Korol, A.B. Implications of human genome structural heterogeneity: Functionally related genes tend to reside in organizationally similar genomic regions. *BMC Genom.* **2014**, *15*, 252, doi:10.1186/1471-2164-15-252.

262. Tsochatzidou, M.; Malliarou, M.; Papanikolaou, N.; Roca, J.; Nikolaou, C. Genome urbanization: Clusters of topologically co-regulated genes delineate functional compartments in the genome of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **2017**, *45*, 5818–5828, doi:10.1093/nar/gkx198.
263. Hlevnjak, M.; Zagrovic, B. Malleable nature of mRNA-protein compositional complementarity and its functional significance. *Nucleic Acids Res.* **2015**, *43*, 3012–3021, doi:10.1093/nar/gkv166.
264. Nahalka, J. Protein-RNA recognition: Cracking the code. *J. Theor. Biol.* **2014**, *343*, 9–15, doi:10.1016/j.jtbi.2013.11.006.
265. Biro, J.C. Coding nucleic acids are chaperons for protein folding: A novel theory of protein folding. *Gene* **2013**, *515*, 249–257, doi:10.1016/j.gene.2012.12.048.
266. Yarus, M. The Genetic Code and RNA-Amino Acid Affinities. *Life* **2017**, *7*, 13, doi:10.3390/life7020013.
267. de Ruyter, A.; Zagrovic, B. Absolute binding-free energies between standard RNA/DNA nucleobases and amino-acid sidechain analogs in different environments. *Nucleic Acids Res.* **2015**, *43*, 708–718, doi:10.1093/nar/gku1344.
268. Root-Bernstein, R.; Root-Bernstein, M. The ribosome as a missing link in prebiotic evolution II: Ribosomes encode ribosomal proteins that bind to common regions of their own mRNAs and rRNAs. *J. Theor. Biol.* **2016**, *397*, 115–127, doi:10.1016/j.jtbi.2016.02.030.
269. Aldana-Gonzalez, M.; Cocho, G.; Larralde, H.; Martinez-Mekler, G. Translocation properties of primitive molecular machines and their relevance to the structure of the genetic code. *J. Theor. Biol.* **2003**, *220*, 27–45, doi:10.1006/jtbi.2003.3108.
270. Babbitt, G.A.; Coppola, E.E.; Mortensen, J.S.; Ekeren, P.X.; Viola, C.; Goldblatt, D.; Hudson, A.O. Triplet-Based Codon Organization Optimizes the Impact of Synonymous Mutation on Nucleic Acid Molecular Dynamics. *J. Mol. Evol.* **2018**, *86*, 91–102, doi:10.1007/s00239-018-9828-x.
271. Taghavi, A.; van der Schoot, P.; Berryman, J.T. DNA partitions into triplets under tension in the presence of organic cations, with sequence evolutionary age predicting the stability of the triplet phase. *Q. Rev. Biophys.* **2017**, *50*, e15, doi:10.1017/S0033583517000130.
272. Goldshtein, M.; Lukatsky, D.B. Specificity-Determining DNA Triplet Code for Positioning of Human Preinitiation Complex. *Biophys. J.* **2017**, *112*, 2047–2050, doi:10.1016/j.bpj.2017.04.023.
273. Lukacisin, M.; Landon, M.; Jajoo, R. Sequence-specific thermodynamic properties of nucleic acids influence both transcriptional pausing and backtracking in yeast. *PLoS ONE* **2017**, *12*, e0174066, doi:10.1371/journal.pone.0174066.
274. Trotta, E. Selection on codon bias in yeast: A transcriptional hypothesis. *Nucleic Acids Res.* **2013**, *41*, 9382–9395, doi:10.1093/nar/gkt740.
275. Dai, Z.; Dai, X. Gene expression divergence is coupled to evolution of DNA structure in coding regions. *PLoS Comput. Biol.* **2011**, *7*, e1002275, doi:10.1371/journal.pcbi.1002275.
276. Bosaeus, N.; Reymer, A.; Beke-Somfai, T.; Brown, T.; Takahashi, M.; Wittung-Stafshede, P.; Rocha, S.; Norden, B. A stretched conformation of DNA with a biological role? *Q. Rev. Biophys.* **2017**, *50*, e11, doi:10.1017/S0033583517000099.
277. Rokas, A. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu. Rev. Genet.* **2008**, *42*, 235–251, doi:10.1146/annurev.genet.42.110807.091513.
278. Michod, R.E. Evolution of individuality during the transition from unicellular to multicellular life. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (Suppl. 1), 8613–8618, doi:10.1073/pnas.0701489104.
279. Kaiser, D. Building a multicellular organism. *Annu. Rev. Genet.* **2001**, *35*, 103–123, doi:10.1146/annurev.genet.35.102401.090145.
280. Kupiec, J.J. A Darwinian theory for the origin of cellular differentiation. *Mol. Gen. Genet.* **1997**, *255*, 201–208, doi:10.1007/s004380050490.

