

Review

Living Organisms Author Their Read-Write Genomes in Evolution

James A. Shapiro 

Department of Biochemistry and Molecular Biology, University of Chicago GCIS W123B, 979 E. 57th Street, Chicago, IL 60637, USA; jsha@uchicago.edu; Tel.: +1-773-702-1625

Academic Editor: Andrés Moya

Received: 23 August 2017; Accepted: 28 November 2017; Published: 6 December 2017

Abstract: Evolutionary variations generating phenotypic adaptations and novel taxa resulted from complex cellular activities altering genome content and expression: (i) Symbiogenetic cell mergers producing the mitochondrion-bearing ancestor of eukaryotes and chloroplast-bearing ancestors of photosynthetic eukaryotes; (ii) interspecific hybridizations and genome doublings generating new species and adaptive radiations of higher plants and animals; and, (iii) interspecific horizontal DNA transfer encoding virtually all of the cellular functions between organisms and their viruses in all domains of life. Consequently, assuming that evolutionary processes occur in isolated genomes of individual species has become an unrealistic abstraction. Adaptive variations also involved natural genetic engineering of mobile DNA elements to rewire regulatory networks. In the most highly evolved organisms, biological complexity scales with “non-coding” DNA content more closely than with protein-coding capacity. Coincidentally, we have learned how so-called “non-coding” RNAs that are rich in repetitive mobile DNA sequences are key regulators of complex phenotypes. Both biotic and abiotic ecological challenges serve as triggers for episodes of elevated genome change. The intersections of cell activities, biosphere interactions, horizontal DNA transfers, and non-random Read-Write genome modifications by natural genetic engineering provide a rich molecular and biological foundation for understanding how ecological disruptions can stimulate productive, often abrupt, evolutionary transformations.

Keywords: genome rewriting; natural genetic engineering; symbiogenesis; holobiont; hybrid speciation; horizontal DNA transfer; mobile DNA elements; network rewiring; repetitive DNA formatting; ecological challenge

1. Introduction and Goals

Over the past 40 years, several books and numerous review articles have detailed the molecular mechanisms that cells utilize to alter their genomes [1–8]. These “natural genetic engineering” (NGE) processes are biochemical tools that living organisms possess to make adaptive use of their DNA databases as “Read-Write Genomes” [9]. Taking these mechanistic discoveries as established science, the goal of this review is to explore how bioinformatics has documented some of the ways that living organisms stimulate and benefit from NGE in the course of evolutionary change. Much of the relevant information will appear in lists and tables that have two objectives: (1) To make the primary literature accessible to the reader, and (2) to manifest how extensive the literature has become verifying that genome change in evolution results from a series of active biological processes, not from passive accidents. Previous reviews have summarized the outline of the basic arguments presented below [10–12], but this article presents each topic in greater depth and detail than earlier publications.

2. Parsing the Fundamental Question in Evolution: How Do Heritable Adaptive Novelties and New Groups of Organisms Arise?

What does it mean to say that living organisms author their read-write genomes? The answer encompasses a set of ideas about how and why genomes change in the course of evolution. The most basic idea is that evolutionary genome change results from biological activities, not from random accidents. The activities are both biochemical NGE functions that form new DNA sequences in genomes and cellular/organismal processes that lead to cell fusions, which alter genome content, and that regulate NGE. Considering genome change as an active biological process in these two ways makes it appropriate to consider the genome a “read-write” (RW) information storage component of the cell [9], different from the “read-only memory” (ROM) that is postulated by conventional evolutionary and molecular biological theories [13–15].

When Darwin first published his theory of evolutionary change involving modification with descent, he wanted to give pride of place as a causal mechanism to Natural Selection. Consequently, he had to reduce hereditary Variation to a minor and haphazard process. Thus, he wrote in Chapter 6, *Origin of Species*: “If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down. But I can find out no such case.” [16]. Over time, however, Darwin came to recognize cases where hereditary variation had its own creative power that he had initially ignored. In later editions of *Origin of Species*, he wrote about natural “sports” or “... variations which seem to us in our ignorance to arise spontaneously. It appears that I formerly underrated the frequency and value of these latter forms of variation, as leading to permanent modifications of structure independently of natural selection” (6th edition, Chapter XV, p. 395) [17]. Since Darwin’s day, classical and molecular genetics have uncovered a dizzying array of internal NGE mechanisms that organisms possess for generating hereditary variation in their genomes [1–8].

Given the new knowledge about how living organisms actively change their genomes, it is appropriate to raise the basic question of evolutionary biology: what are the sources of adaptive change and taxonomic variation in evolution? Natural selection operates after the fact, as a test for the biological value of changes that have already taken place. By definition, natural selection cannot explain the *origins* of novel inherited characteristics. A related question about the processes of evolutionary change is whether important hereditary variations are limited, as Darwin initially argued, to “numerous, successive, slight modifications”, or can they occur abruptly and involve major changes in phenotypic characters? The question about gradual versus saltatory change has come up repeatedly in the history of evolutionary studies, and a number of distinguished evolutionary biologists have argued for abrupt major changes by various means (Table 1).

Table 1. Key Scientists Advocating non-Gradualist Evolution in the 19th and 20th Centuries.

Evolutionist	Non-Gradual Evolutionary Process
William Bateson (1861–1926)	Discontinuous variation
Hugo de Vries (1848–1935)	Abrupt mutational variation
Konstantin Mereschkowsky (1855–1921)	Evolution by symbiogenesis
Ivan E. Wallin (1883–1969)	Evolution by symbiogenesis (“Symbiogenesis”)
Boris Mikhailovich Kozo-Polyansky, (1890–1957)	Evolution by symbiogenesis
George Gaylord Simpson (1902–1984)	Quantum evolution
Richard Goldschmidt (1878–1958)	“Hopeful monsters” formed by redirecting developmental programs
George Ledyard Stebbins (1906–2000)	Hybrid Speciation (“Cataclysmic Evolution”)
Niles Eldredge (b. 1943) and Stephen J. Gould (1941–2002)	Punctuated equilibrium
Lynn Margulis (1938–2011)	Evolution by symbiogenesis

A fully referenced version of this table is available online as Supplementary Table S1.

The evidence for the adaptive and genealogical importance of saltatory variations involving biological processes in evolutionary history is now overwhelming (Table 2). The symbiogenic merger

of two different cell types or the hybridization of two different species must necessarily produce an abruptly distinct organism with a novel combination of multiple traits. Moreover, it cannot be emphasized too strongly that much of evolutionary change occurs interactively in the biosphere, not isolated within the discrete genomes of species that are completely separated from each other. The recognition that different organisms are interconnected at the cellular and genomic levels and that novel lineages frequently do not owe their new hereditary constitution to a strictly vertical process of inheritance has been transformational in our understanding of how evolution occurs. The current deluge of genome sequence data reveals DNA exchanges between diverse lineages at all levels of taxonomic diversification (Section 5). Moreover, genomics has heightened our awareness that the biosphere is populated by organisms that metabolize, behave, reproduce, and evolve interactively, not in isolation, bringing many biologists to abandon the abstract idea of the single-genome species and embrace the concept of the “holobiont”, an integrated amalgam of smaller and larger species, which exhibits collective phenotypic properties (Section 3.3) [18–20].

Table 2. Selected Variation Processes that Lead Repeatedly and Abruptly to Adaptive and Taxonomic Novelties.

Symbiosis, Symbiogenesis and “Holobiont” Modifications by Cell Mergers (Section 3)
Interspecific hybridization (hybrid speciation) (Section 4)
Horizontal DNA transfers (Section 5)
Protein evolution by domain rearrangements and coding region innovation (Section 6)
Mobile DNA activity to restructure genomes, rewire developmental regulatory networks and form regulatory long non-coding RNA molecules (Section 7)

While most of this review will focus on molecular data, the conclusions are consistent with real time observations documenting unexpectedly rapid evolutionary changes among organisms in the wild. Among the most outstanding examples of such observations are those of Peter and Rosemary Grant’s four decades tracking Darwin’s finches *Geospiza* in the Galapagos Islands [21–23]. Their reporting of rapid modifications in *Geospiza* beaks following interspecific hybridization is particularly noteworthy because alteration of beak morphology was one of the paradigms that Darwin chose to exemplify the action of natural selection on minute, gradual phenotypic variations [16,22].

3. Biomath: One + One = One [24,25]; Ubiquitous Cell Mergers in Reproduction and Evolution (Reviewed in [26])

3.1. Symbiogenetic Origins of Eukaryotic Cells [26–32] and Their Photosynthetic Lineages

The ability of cells to merge, invade, and engulf one another is the basis of many basic biological processes that include endosymbiosis, sexual reproduction, phagocytosis, and pathogenesis [27]. Reviewing the genomic evidence, the highest-level taxonomic innovations that we can document involve interactions between distantly related organisms. In evolutionary history, molecular evidence clearly indicates that symbiogenesis underlies the origins of major eukaryotic lineages, as proposed by Mereschkowsky, Wallin, Kozo-Polyansky, and Margulis (Table 1). Symbiogenesis is the process by which an endosymbiotic microorganism loses its ability to reproduce outside the host cell and becomes an obligate intracellular organelle [33,34]. In addition, other simpler symbiotic events that do not involve the loss of autonomous reproduction contribute in multiple ways to eukaryotic evolution.

Although there is still uncertainty about whether the earliest eukaryotic ancestors possessed mitochondria [35], there is no doubt that the vast majority of extant eukaryotic lineages have oxidative mitochondria or related non-oxidative organelles, called hydrogenosomes and mitosomes [36–38]. Mitochondria contain DNA, carry out transcription and translation, and are clearly related in membrane, protein, and ribosome structure to α -proteobacteria [27,39–41]. Consequently, most (perhaps all) eukaryotic lineages originated in a symbiogenetic event involving an α -proteobacterium.

Photosynthetic eukaryotes have additional DNA-containing organelles that are labeled chloroplasts or plastids, which are related in their photosynthetic pathway, proteins, ribosomes, and membranes to photosynthetic bacteria [27,39,42–44]. Molecular analysis of photosynthetic eukaryotes distinguishes three major groups (Table 3):

- (1) lineages that arose from **primary** cyanobacteria endosymbiosis with a non-photosynthetic eukaryotic cell: *Archaeplastida* = red and green algae, *Glaucophytes* = algae containing peptidoglycans, and green plants = *Embryophyta*;
- (2) a separate lineage of photosynthetic amoebae *Paulinella chromatophora* that arose from a **primary** endosymbiosis with photosynthetic bacteria from the *Synechococcus-Prochloron* clade; and,
- (3) highly diverse photosynthetic lineages that arose from a **secondary** or even **tertiary** endosymbiosis of a photosynthetic eukaryotic cell with a non-photosynthetic eukaryote.

While primary plastid symbiogenesis creates a cell with three genome compartments (nucleus, mitochondrion, and plastid), a higher-order photosynthetic symbiogenesis creates a cell with four or more genome compartments (nucleus, mitochondrion, plastid, and the “nucleomorph” descended from the photosynthetic partner’s nucleus).

Table 3. Photosynthetic eukaryotic lineages resulting from symbiogenesis.

Taxonomic Group	Symbiogenetic Origin
<i>Archaeplastida</i>	
Green algae (<i>Chlorophyta</i>)	Primary cyanobacterial endosymbiosis
<i>Glaucophytes</i> (order <i>Chlorococcales</i>)	Primary cyanobacterial endosymbiosis
Red algae (<i>Rhodophyta</i>)	Primary cyanobacterial endosymbiosis
Land plants (<i>Embryophyta</i>)	Primary cyanobacterial endosymbiosis
Euglyphid amoeba <i>Paulinella chromatophora</i>	Primary cyanobacterial endosymbiosis (<i>Synechococcus-Prochloron</i> clade)
<i>Euglenids</i> (flagellated algae)	Secondary green algal endosymbiosis
<i>Chlorarachniophytes</i> (marine algae)	Secondary green algal endosymbiosis
<i>Chromalveolates</i> (multiple lineages including organisms responsible for a large fraction of atmospheric oxygen, such as brown algae, coccolithotrphs, cryptophytes and diatoms)	Secondary red algal endosymbiosis
<i>Dinoflagellates</i> (flagellated marine and fresh water protists)	Tertiary <i>chromalveolate</i> endosymbiosis or serial green or red alga endosymbioses
<i>Warnowiid</i> dinoflagellates with camera eye-like “ocelloids”	Ocelloid “cornea” formed by mitochondria and “retina” formed by red algae-derived plastids

A referenced version of this table is available online as Supplementary Table S3.

The organic products of the abrupt symbiogenetic evolutionary events listed in Table 3 are multiple and diverse lineages of photosynthetic organisms with a wide range of phenotypes that occupy virtually all ecologies exposed to light (and likely some not exposed to light). In at least one photosynthetic lineage, the *Warnowiid* dinoflagellates, serial symbiogenetic events have been linked to the morphogenesis of a light-collecting organelle (“ocelloid”) that bears a remarkable resemblance to the camera eyes in animals, with a cornea composed of mitochondria and a retinal body formed by plastids acquired from a red alga [45].

Genomic analysis of mitochondria and plastids make it clear that abrupt symbiogenetic events with major physiological consequences for cell structure and energy metabolism were foundational in establishing the phylogenetic bases of eukaryotic evolution. In addition, genome evolution continued after the initial symbiogenetic event. There is abundant evidence for subsequent DNA exchanges

from organelle to host nuclear genomes [46] and between mitochondrial and plastid genomes [47–51]. As a result of DNA transfers and rearrangements, coding content and physical organization diverge widely across taxonomic groups in plastid, as well as mitochondrial genomes [52,53]. Because of DNA loss from mitochondrial and plastid genomes and transfers to the nucleus, the large majorities of mitochondrial and plastid proteins are encoded by the host cell nuclear genome and are transported into the organelles. Moreover, while mitochondria and plastids contain nuclear-encoded proteins evolved from *α-proteobacteria* and *cyanobacteria* ancestors, they also contain proteins that derive phylogenetically from eukaryotes, from other bacterial lineages, and even from bacteriophages [48,54,55]. These findings highlight how deeply eukaryotic and organelle evolution reflect DNA exchanges across the biosphere.

3.2. Symbiosis as an Adaptive and Evolutionary Stimulus; Speciation by Endosymbiosis and Mating Incompatibility

As foreseen by Wallin, the combination of large eukaryotic hosts and their microbial symbionts (nowadays collectively labeled “the microbiome”) extend the adaptive capabilities of the resulting composite organism [56,57]. Well-known examples include the symbiotic acquisition of nitrogen fixation in legumes by *Rhizobium* root nodule formation [58,59], mycorrhizal fungi providing root functions for orchids, and other plants that are hampered in the ability to generate their own roots [60,61] and expansion of digestive and biosynthetic capacities conferred by microbial symbionts in the animal intestinal tract [62,63]. Symbiotic associations are particularly important in the broad range of animals that live on plant material and depend on associated microbes to digest cellulose and other phytopolymers [64,65]. There are even photosynthetic metazoa that are formed by secondary algal symbiosis, including corals [66], sea slugs [67], and salamanders [68]. In these animal symbioses, the endosymbiotic algae retain the capacity for autonomous reproduction.

Clearly, the establishment of these symbiotic relationships represented both a quantum leap in host adaptive potential, as well as an expansion of the symbiont’s ecological range and evolutionary potential [69–71]. DNA analysis has expanded our knowledge of symbiotic relationships and has led to increased recognition that associated microorganisms (collectively, the “microbiome”) of humans and other macroscopic organisms have powerful impacts on phenotypes that were previously ascribed only to the host. Some examples include:

- embryonic development and metabolic homeostasis [72–74];
- immunity of multicellular organisms [75,76], such as the Hawaiian squid *Euprymna scolopes*, host to the marine luminous bacterium *Vibrio fischeri* [77];
- susceptibility to infection by bacteria, parasites and viruses [78–80]; and,
- higher nervous system functions and behavior [81,82].

The fact that these complex multicellular phenotypes do not purely depend on the expression of the host genome means that we cannot account for their evolutionary trajectories only by genetic changes within a single organism.

One particular phenotype that is influenced by microbial symbiosis can have direct effects on taxonomic divergence in host evolution. In invertebrates, bacteria belonging to the genus *Wolbachia* are common intracellular endosymbionts [83,84]. *Wolbachia* colonize germline cells [85] and influence sexual differentiation with profound effects on mating [86]. In *Drosophila* species, mosquitos [87], parasitic wasps [88], and other arthropods, the infection of males with *Wolbachia* can lead to “cytoplasmic incompatibility” and sterility in mating with *Wolbachia*-free females [89,90]. The mating incompatibility that is generated by *Wolbachia* infection thus genetically isolates two populations from one species and provides a trigger for “speciation by symbiosis” [91,92]. Intriguingly, the intensity of cytoplasmic incompatibility in *Drosophila* is subject to control by viral bacteriophage functions expressed from a WO prophage integrated in the *Wolbachia* genome [93].

The few examples just cited illustrate how the phenotypic and reproductive effects of symbioses provide important accelerations to the evolution of eukaryotic host organisms. These inter-organismal biological influences typically result in taxonomic diversifications below the genus level. Nonetheless, they illustrate the principle that was stated at the beginning of this review. It is essential to think about the evolutionary process as occurring in a biosphere where distinct organisms are continually interacting, not as a purely endogenous process limited to the genome of an isolated species. Indeed, we now recognize that the organism of classical theory, evolving solely by internal changes, has become a largely abstract idealization, divorced from the real world of microbiomes, infections, and other biosphere interactions.

3.3. Holobiont Evolution: Lamarckian Acquisition and Inheritance of Novel Traits

An even more radical view of host-symbiont interactions has been to go beyond the concept of isolated species altogether and consider organisms as “holobionts”, each a consortium of distinct cell types transmitted across generations [18–20]. In many cases, like the examples in Section 3.2 above, holobionts comprise macroscopic eukaryotic hosts and smaller associated microorganisms, both prokaryotic and eukaryotic. Examples include corals [94], plants [95,96], tsetse flies [97], and termites (with their numerous eukaryotic intestinal protozoa [98,99]). Hereditary transmission of the symbiotic microorganisms typically occurs by the incorporation into germ cells or by reinfection of the newly formed zygote or embryo [86,100,101].

The holobiont concept is not limited to macroscopic hosts and their microbiomes, but also extends to consortia composed purely of different microorganisms, such as microbial mats [102], stromatolites [103], and microbial communities at deep-sea hydrothermal vents [104,105]. It is interesting to note that, the holobiont concept historically was applied unconsciously to composite organisms like lichens before their multi-species nature was recognized [106–109].

From the holobiont perspective, gain or loss of one or more members of the heritable consortium constitutes an evolutionary transition, often with major phenotypic consequences. Such evolution has been demonstrated experimentally [110]. Since heritable adaptive traits can be gained by infection, an intriguing feature of holobiont evolution is that it proceeds by the hereditary transmission of acquired characteristics. Aphid acquisition of resistance to parasitic wasps and other useful ecological traits are an example of this process [111,112]. In other words, holobiont evolution effectively constitutes a Lamarckian process based on well-documented biological mechanisms.

Frequently, as in *Wolbachia*-stimulated speciation by symbiosis, a change in holobiont composition constitutes the initial step in a series of evolutionary transitions. These transitions include DNA transfers within the holobiont, both between different microbial constituents and between symbiotic microbes and the host nuclear genome [113]. These DNA transfers are analogous to those from symbiogenetic organelles to the nuclear genome mentioned in Section 3.1 above. There are now numerous examples of symbiont to host genome DNA transfers [114]. Aphid genomes have acquired sequences from the bacterial endosymbiont *Buchnera* [115]. In the well-studied arthropod-*Wolbachia* system, *Wolbachia* sequences have entered the genomes of mosquitoes [116,117], tsetse flies (*Glossina*) [118], beetles [119,120], wasps, honeybees, ticks, and pathogenic filarial nematode worms [113], as well as numerous *Drosophila* species [113,121]. In *Drosophila ananassae*, for example, more than 2% of total nuclear DNA sequences come from *Wolbachia* endosymbionts and include the entire bacterial genome [122,123].

4. “Cataclysmic Evolution” by Interspecific Hybridization

The cytogenetic study of cultivated plants in the first half of the 20th Century uncovered the highly significant fact that many crop genomes are in fact hybrids, combinations of two genomes inherited from a pair of related species. In 1951, one of the principle cytogeneticists doing these studies, and a leading evolutionary theorist at that time, G. Ledyard Stebbins, published an article in *Scientific American* on such hybrid species [124]. Stebbins labeled the process of hybrid species

formation “Cataclysmic Evolution” and commented particularly on the fact that hybrids frequently have novel properties that are quite different from those of either parent:

“The remarkable fact about the wheat story is that the combination of chromosomes of a moderately useful plant, emmer wheat (*Triticum turgidum*), and those of a completely useless and noxious weed (goat grass, *Aegilops squarrosa*) produced the world’s most valuable crop plant (bread wheat, *Triticum aestivum*). This example should tell us that we cannot always predict in advance whether a particular hybrid will be worthless or a priceless new addition”. [124].

Contrary to the expectations of many conventional evolutionists, interspecific hybridization like that producing *Triticum aestivum* is not an exception but rather a common process that turns out to be a major factor in eukaryotic evolution [125].

4.1. Abundant Examples of Speciation and Adaptive Radiations by Interspecific Hybridization and Whole Genome Duplications (WGDs) in Plants and Animals

Stebbins explained in his 1951 article that, cataclysmic evolution actually involves two discrete biological activities: (1) mating between two related but distinct species (inter-specific hybridization), and (2) whole genome doubling (WGD) of the chromosomes in the resulting hybrid genome. WGD is essential to the accelerated generation of a hybrid species because eukaryotic chromosomes each need to pair with a homologous partner chromosome during the meiotic divisions that produce the haploid gametes needed for ongoing sexual reproduction. Experimental investigation indicates that hybridization itself, as well as other stresses and stimuli, trigger the genome replication events that are necessary for WGD [126,127]. The frequency of WGD events in the evolutionary histories of many eukaryotic lineages is indicative that inter-specific hybridization may have been a key factor in the origins of the new taxa [128]. Major taxonomic origins involving WGD include yeasts and fungi [129–131], ciliated protozoa [132], cereals [133], flowering plants [134], and crabs [135]. Vertebrate evolution, in particular, involved two successive WGD events [136,137], followed by further WGDs in fishes [138–140]. Intriguingly, WGD has been credited with the evolution of electric potential in fishes [141].

WGD is one feature of hybrid speciation that has captured the attention of geneticists that are interested in the establishment of novel regulatory networks in the course of evolution [142–146]. WGD provides at least two kinds of impetus for regulatory innovation. First, it doubles the content of loci encoding regulatory factors that are subject to modification and utilization in newly established mutant networks. Second, by duplicating the entire content of the genome, WGD assures that essential multi-locus functions can be maintained in one unmodified copy of the entire network, while individual components of the second copy are no longer essential, and thus are free for rewiring and evolutionary experimentation [147].

Abundant cytogenetic and genomic evidence has led to the conclusion that hybrid speciation has played a key role in the evolutionary diversification of a wide range of organisms (Table 4).

Table 4. Examples of Speciation and Adaptive Radiation Involving Interspecific Hybridization and Changes in Chromosome Number.

Fungi: <i>Saccharomyces</i> Yeast
Ciliated Protozoa
Plants: Tragopogon (<i>Asteraceae</i>), <i>Pinus densata</i> , <i>Primula</i> , Sunflowers (<i>Helianthus anomalus</i>), Irises (<i>Iris fulva</i> , <i>I. hexagona</i> , and <i>I. nelsonii</i>), <i>Nicotiana</i> (<i>Solanaceae</i>), <i>Orchidaceae</i> , <i>Brassica napus</i> , <i>Arabidopsis</i> , Potatoes (<i>Solanum stoloniferum</i> and <i>S. hjertingii</i>), Wheats (<i>Aegilops-Triticum</i> group), Cotton (<i>Gossypium</i>)
Animals: Tephritid fruitflies, Mosquitoes, Tiger Swallowtail Butterflies, <i>Heliconius</i> butterflies, Ants, Sculpins (<i>Cottus</i> sp., <i>Teleostei</i>), Sailfin silversides (<i>Teleostei</i>), East African Cichlids, Sparrows, Yellow-rumped (Audubon’s) warbler, Galapagos finches (<i>Geospiza</i>), Clymene dolphin (<i>Delphininae</i> (<i>Cetacea</i> , <i>Mammalia</i>)), Bats, Chinese hares (genus <i>Lepus</i>), Cats (<i>Felidae</i>), Colobine monkeys, Southern African baboons

A fully referenced copy of this table is available as Supplementary Table S4.

Two animal groups in Table 4 deserve special mention because they occupy key places in ongoing discussions about the nature of evolutionary processes: (1) The first one is *Geospiza*, Galapagos finches, whose beaks have served as primary paradigms for proponents of incremental gradualist evolution from Darwin on. Painstaking field study over four decades by Rosemary and Peter Grant [21], supplemented by genome sequencing [148–150], has clearly established the real time importance of introgressive hybridization in the adaptive evolution of beaks in this group. (2) The second paradigmatic group are the cichlid fishes of African lakes, which are considered exemplars of rapid vertebrate speciation and adaptive radiation [151,152]. Here too, hybrid speciation has proven to be critical to a remarkable phenomenon of accelerated and ongoing evolutionary diversification [153,154]. A further example of a role for hybridization in adaptive evolution comes from the generation of mimicry patterns by *Heliconius* butterflies [155–157].

The importance of introgressive hybridization between distinct populations in their evolutionary development is yet another factor that has led certain authors to point out that it is more appropriate to consider reticulate (net-like) over conventional branching (tree-like) models for inheritance by new species [158–160]. Altogether, we can see from the *Geospiza*, cichlid, and other cases like those cited in Table 4, that biological activity in the form of interspecific hybridization provides a key impetus to many (perhaps most?) paradigms of higher organism evolution [161–166]. So it is appropriate to look more closely at how such hybridization influences genome reorganization.

4.2. Genomic Consequences of Interspecific Hybridization

For practical reasons, chiefly ease of fertilization, experimental work on interspecific hybridization has been carried out more extensively in plants than in animals. Nonetheless, there is some work on experimental hybridization in certain animals, such as *Drosophila* and mice. Table 5 lists some of the genome expression and reorganization consequences that are observed in interspecific crosses of both plants and animals. The changes listed in Table 5 (ploidy, epigenetic alterations, new genome expression patterns, activation of mobile DNA elements and related genome rearrangements, karyotype modifications, and alteration of tandemly repeated DNA arrays) have all been associated with major novelties in adaptive phenotype and mating incompatibility.

Table 5. Genomic consequences of experimental interspecific hybridization in plants and animals [167].

Genome Effect
Changes in ploidy (mostly WGD)
Alteration of epigenetic modifications to the genome
Alterations in expression patterns across the genome
Activation and spread of mobile DNA elements
Genome restructuring involving mobile DNA elements
Changes in chromosome structure and karyotype
Alteration of tandem repetitive DNA arrays and centromeres

A fully referenced version of this table is available as Supplementary Table S5. Additional references are available on the internet at http://shapiro.bsd.uchicago.edu/Hybrid_dysgenesis_interspecific_hybridization.html and <http://shapiro.bsd.uchicago.edu/GenomicResponsesToChangesInPloidyInterspecificHybridization.html>.

It appears that a basic consequence of merging gametes from different species is to upset the established patterns of epigenetic regulation for the parental species, and that the breakdown in control then leads to the activation of multiple genome reorganization functions [168]. The activation of mobile DNA elements by interspecific hybridization is particularly relevant to the main argument of this review because these elements are primary biological agents for rewriting genome content [1–7].

4.3. The Special Genomic Impacts of Interspecific Hybridization on Evolutionary Innovation

In thinking about interspecific hybridization as a genome change process from the evolutionary perspective, it is helpful to point out how it differs from other mutagenic events that have been

considered in traditional evolutionary thinking. Within-species mutations are generally limited to one or a small number of genome locations. As a consequence, they typically have effects on one or a few organismal characters. That is why conventional evolutionary theory has long postulated species change to be a gradual, cumulative, multi-generational process. In contrast, mergers between the complete genomes of two distinct species involve all regions of those genomes, and therefore have the potential to affect a large number of phenotypic traits in a single generation. That is the reason some authors have denoted interspecific hybrids to be the “hopeful monsters” postulated by Richard Goldschmidt (Table 1) [153,169–173].

All in all, the major role of interspecific hybridization in stimulating taxonomic and adaptive divergence exemplifies the importance of purely biological functions (mating, cell fusion at fertilization, epigenetic modification, and triggering of NGE activities) as primary agents of evolutionary variation. It requires little imagination to see how ecological deterioration and mating population declines can lead to an increased incidence of interspecific matings, and thus accelerate the evolution of new species, some of which may be better adapted to the altered ecology (Section 8.1).

5. Widespread Horizontal DNA Sequence Mobility between Organisms

A separate form of rapid evolutionary variation involves genomic communication between different (and often quite distant) species. This is the phenomenon known as horizontal or lateral DNA transfer, frequently abbreviated as HGT (“horizontal gene transfer”) [174]. We have already discussed one aspect of this phenomenon in Section 3, comprising DNA transfers from symbiogenetic organelles and endosymbionts to the host nuclear genome. But, horizontal DNA transfers are more widespread than those cases and have been documented to occur between all domains of living organisms [175,176].

5.1. Distinct Modes of Intercellular DNA Transfer

We have known for over half a century that bacteria possess three distinct mechanisms of transferring DNA between cells: transformation by uptake of exogenous DNA, transduction by encapsidation of DNA in bacteriophage (viral) particles, and conjugation by direct cell-to-cell contact (Table 6) [177]. These mechanisms mediate extensive DNA transfers across different prokaryotic lineages, both bacterial and archaeal, and similar processes have also been found to facilitate DNA exchanges from bacteria to eukaryotes, and even among various eukaryotes (Table 6). The molecular biology of these DNA transfer and uptake modalities is well understood. Both conjugal DNA transfer and DNA uptake by bacteria that are competent for transformation is based on the ability of the cells to elaborate complex cell-surface structures labeled “type IV secretion systems” that facilitate macromolecular transport across the bacterial cell envelope (they are also used by pathogenic bacteria during infection to introduce “virulence factor” proteins into target mammalian cells) [178–181].

A limited number of examples of regular bacteria to eukaryote DNA exchange have been well known for several decades, such as the conjugal transfer of specific “T-DNA” from *Agrobacterium tumefaciens* to plant cells during crown gall tumor formation [182]. The *Agrobacterium*-plant T-DNA transfer system also became well known for its biotechnology application as a tool for the genetic engineering of plants [183,184]. Nonetheless, despite our knowledge of various bacterial-eukaryotic cell interactions, it has been rather surprising to discover the extent and variety of naturally-occurring DNA transfers that occur across the widest taxonomic distances (Table 6). It has also been surprising to find that certain processes that were thought to be exclusively prokaryotic, like transformation, also occur naturally in eukaryotes [185].

In addition to these natural processes, investigators routinely use various techniques to introduce DNA into eukaryotic cells [186], including the modalities listed at the bottom of Table 6. It is probable that intercellular DNA transfer inside membrane vesicles will turn out to be more ecologically and evolutionarily important than we currently know [187,188]. The use of DNA uptake by sperm cells has proved to be easy, and, like T-DNA, has found broad biotechnology applications in the

genetic engineering of animals [189,190]. Sperm-mediated DNA transfer illustrates the principle that horizontal transfer to multicellular eukaryotes encounters fewer obstacles at early stages of development, when genome isolation is less elaborated than in the mature organism [191].

Table 6. Modes of Horizontal DNA Transfer.

DNA Transfer Mode	Donor-Recipient Taxa
Liberation and uptake of extracellular DNA ("Transformation")	Archaea-archaea; Bacteria-bacteria; Yeast-yeast; <i>Stramenopile</i> red alga-alga; Plant-bacteria
Encapsulation in and delivery by a virus or virus-like particle ("Transduction")	Archaea-archaea; Bacteria-bacteria; Insect-insect; Bacteria-mammal; Mammal-mammal
Establishment of a DNA transport pore between two cell envelopes ("Conjugation")	Archaea-archaea; Bacteria-bacteria; Bacteria-yeast; Bacteria-fungi and mushrooms; Bacteria-plant; Bacteria-mammalian cells; Bacteria-diverse eukaryotes
Cell fusion	Archaea-archaea
Phagotrophism	Bacteria-protist
Additional DNA transfer modes within taxa:	
Protoplast fusion	Bacteria, Yeast, Fungi, Plants
Liposomal or membrane vesicle-mediated transfer	Archaea, Mammalian cells
Sperm-mediated DNA transfer	Insect larva, Mammals
Parasite- and endosymbiont-mediated transfer (inferred)	Red algae; Aphids, parasitoid wasps; a beetle; <i>Drosophila</i> ; Butterflies; Vertebrates; Reptile-mammal; Mammals

A fully referenced version of this table is available as Supplementary Table S6.

5.2. Lessons on Rapid Evolution from the Smallest Living Cells

To fully appreciate the evolutionary potential of horizontal DNA transfer, we need to look back to the middle of the 20th Century, when humanity undertook a global evolution experiment by initiating the widespread use of antibiotics to combat bacterial pathogens. This experiment quickly demonstrated unforeseen evolutionary capabilities as pathogenic bacteria rapidly acquired resistance to one antibiotic after another, eventually creating the current "superbug" crisis of pathogens that are resistant to all known antibiotics [192]. When extensive antibiotic therapy began in the 1940s–1950s, there was a well-elaborated theory of how bacteria would evolve resistance: bacterial cells would mutate and alter cell structures so that they were less susceptible to antibiotic action or the cells became less permeable to the antibiotic [177]. The acquisition of streptomycin resistance by bacterial ribosome protein alteration is an example of this process [193]. Increased levels of resistance would result from the accumulation of successive mutations, and this multi-step mutational process could easily be confirmed for many different antibiotics in the laboratory [177,194].

Although the step-by-step mutational path for evolution of antibiotic resistance has been found relevant in isolated disease situations [195], this process did not explain most bacterial resistance. In the 1960s, it was discovered that the major source of clinically significant antibiotic resistance evolved in natural bacterial populations by the acquisition and spread of transmissible antibiotic resistance determinants [196]. This previously unsuspected evolutionary mechanism differed from the predicted (and experimentally confirmed!) mutational process in two significant ways: (1) resistance resulted from the synthesis of proteins that either inactivated or expelled the antibiotic from the target bacterium [193,197], and (2) resistance was encoded on plasmids, independently replicating genetic determinants, or other mobile DNA elements that were readily transferred to sensitive bacteria [1,198–200].

While the initial focus of studies on horizontal DNA transfer in bacteria was antibiotic resistance, it has become evident that DNA elements encoding many other important adaptive phenotypes in both bacterial and archaeal prokaryotic cells are subject to intercellular transmission [201,202]. The transfers occur across large taxonomic distances within Archaea and Bacteria, and between the two

prokaryotic kingdoms [203,204]. The prokaryotic functions subject to horizontal transfer range widely: metabolism and translation [205], mutagenic DNA repair [206], nitrogen fixation [207], and complex phenotypes like symbiosis [207–209] and infectivity/virulence [210–213]. Given what we know about the ability of prokaryotic cells to mobilize DNA both within and between cells, there appear to be no limits on the traits that can be transferred from one prokaryote to another. Genomic evidence indicates that such transfers have contributed to the evolution of major new Archaeal taxa [214–216].

Rapid prokaryotic evolution is further enhanced by specialized structures in the genome that have molecular features endowing them with the propensity to accumulate tandem DNA modules that are encoding related functions, thereby facilitating their joint transfer from one host to another. These specialized structures go by a variety of names: integrons [217,218], integrative conjugative elements (ICEs) [219], genomic islands (GIs) [220], mobilizable genomic islands (MGIs) [221,222], and the recently discovered GI_{nts}, a new type of mobile genomic island widespread in bacteria but not archaea [200]. The modes of genome integration and intercellular transfer of these compound elements vary [223]. As their name indicates, ICEs encode their own excision, conjugative transfer, and integration functions, whereas the MGIs encode their own excision and integration functions, but parasitize the conjugation apparatus of plasmids or other helper elements for conjugal transfer. Integrons and certain GIs may integrate into self-replicating transmissible plasmids, while some GIs and MGIs use recombination activities from the recipient cell to insert into the genome following conjugal transfer.

The ability to build up and transfer defined DNA regions encoding multiple proteins relating to a single phenotype enables the virtually instantaneous acquisition of complex adaptations by intercellular horizontal DNA transfer. These composite elements accrete protein-coding cassettes by using NGE activities related either to bacteriophage site-specific recombinases [224] or to mobile DNA element transposases [223,225,226]. They can vary in size from elements encoding a few antibiotic resistance determinants, to much larger structures, such as the 126 kb “super-integron” segment of the *Vibrio cholera* genome encoding 214 proteins (including virulence/pathogenicity functions) [227], the 502 kb “ICEMIR7A” symbiosis island from *Mesorhizobium loti* [228], and the 674 kb PAIS_t pathogenicity island from *Streptomyces turgidiscabies* [229,230].

5.3. Horizontal DNA Transfer across Large Taxonomic Boundaries

The ability to receive DNA from different species provides numerous opportunities to acquire novel adaptive capabilities in all domains of life, not just prokaryotes (Table 7) [176,231–239]. The evidence for horizontal DNA transfer across large taxonomic distances comes from the phylogenetic analysis of genome sequence data [176]. A genetic locus or element is concluded to have entered an organism’s genome by horizontal transfer when sequence analysis finds the closest related examples in distant taxa, and the locus or element is absent from the genomes of taxa more closely related to the putative horizontal transfer recipient. For example, the phytopolymer digesting enzymes of plant pathogenic nematodes are most closely related to enzymes that are encoded by various bacteria and fungi than they are to enzymes with similar functions that are found in related nematode species [240]. Note that in this type of phylogenomic analysis, the recipient species can be identified with far more precision than the donor.

As Table 7 illustrates, there appear to be no absolute barriers to DNA transfers between organisms, and the genetic loci horizontally transferred frequently encode biochemical activities or functions adaptive for the lifestyle of the recipient organisms. This is clear in the example just cited and other cases, where the acquisition of microbial enzymes allows for animals to digest plant material as a food source and become herbivorous [240–242].

Table 7. Selected examples of inter-phylum adaptive horizontal DNA transfers based on genomic data.

Donor	Recipient	Function(s)
Prokaryote-Prokaryote		
Bacteria	Methanogenic Archaea	Aerobic metabolic activities
Bacteria	Thermophilic Archaea	Mesophilic metabolic activities
Prokaryote-Eukaryote microbe		
Bacteria	Plant pathogenic fungi	Extracellular proteins, interference with plant defense-response, degradation of plant cell walls, carbohydrate metabolism
Bacteria and Archaea	Red alga <i>Galdieria sulphuraria</i>	Growth in high temperature, toxic metal-rich, acidic environments
Proteobacteria, cyanobacteria and archaea	Diatom <i>Phaeodactylum tricornutum</i>	7.5% of genetic loci (784 loci); metabolic and biosynthetic activities
Bacteria, archaea	Rumen ciliates	Catabolism of complex carbohydrates
β , γ -Proteobacteria, Chlamydiae, other bacteria	Red alga <i>Cyanidioschyzon</i>	Biosynthetic activities
Rumen bacteria <i>Fibrobacter succinogenes</i>	Rumen fungus <i>Orpinomyces joyonii</i>	Endoglucanase (polysaccharide digestion)
Eukaryote microbe-Eukaryote microbe		
Fungi	Plant parasitic oomycetes	Plant cell wall digestion, resisting plant defenses
Algae, bacteria	Choanoflagellate (unicellular animal precursor)	405 genetic loci (4.4% nuclear genome); metabolic activities
Prokaryote-metazoa		
Bacteria	Sponge	Metabolic and biosynthetic activities
Bacteria	Sponge	Biomineralization
Bacteria	Starlet sea anemone	Biosynthetic activities
Bacteria	Hydra	Metabolic and biosynthetic activities
Bacteria	Plant parasitic nematodes	Cellulose and phytopolymer digestion
Bacteria	Stick and Leaf Insects	Pectinases
Bacteria	Phytophagous mites and <i>Lepidoptera</i>	Detoxification of plant defense molecules
Bacteria	Coffee berry borer beetle	Digestion of coffee storage polymer
Bacteria	Silkworm <i>Bombyx mori</i> and related <i>Lepidoptera</i>	Glycosyl hydrolase family, oxidoreductase family, and amino acid metabolism
Bacteria	Arthropods, echinoderms, and vertebrates, including platypus and opossum, but not in placental mammals	Glyoxylate cycle enzymes
Bacteria	Urochordate <i>Ciona intestinalis</i>	Cellulose synthase
Eukaryote microbe-metazoa		
Fungi	Plant parasitic nematodes	Cellulose and phytopolymer digestion
Fungi	Pea aphid, spider mite	Carotenoid pigment biosynthesis, cyanate metabolism
Algae	Tunicate <i>Ciona intestinalis</i>	Molecule transport, cellular regulation and methylation signaling
Microbe-plants		
Bacteria, Archaea, viruses, and sea anemones	Moss <i>Physcomitrella patens</i> (primitive land plant)	Xylem formation, plant defense, nitrogen recycling, plus biosynthesis of starch, polyamines, hormones and glutathione; water stress adaptation
Bacteria, fungi	Plants	Biosynthetic activities
Fungi	Plants (Bryophyte moss)	Metabolic activities
Plants-microbes		
Plants	Bacteria, fungi, amoebzoa	Expansins (plant cell-wall loosening proteins)
Plants	Fungi	Regulatory proteins
Plant-plant		
Parasitic plant (<i>Cuscuta</i> sp.)	Host plant (<i>Plantago</i> sp.)	Mitochondrial loci <i>atp1</i> , <i>atp6</i> and <i>matR</i>
<i>Rafflesia</i> host plant	Parasitic flowering plant <i>Rafflesia cantleyi</i>	Metabolic and mitochondrial activities
Grasses	Barley (<i>Hordeum</i>)	Ribosomal RNA (rDNA) sequences
Bryophyte hornworts	Fern	Novel chimeric photoreceptor—neochrome
Animal-animal		
Fish (herring or sea raven)	Rainbow smelt, <i>Osmerus mordax</i>	Type II anti-freeze protein

Table 7. Cont.

Donor	Recipient	Function(s)
Eukaryote-prokaryote		
Eukaryotic cells (Impossible to identify donor because <i>Legionella</i> infects and grows in amoebae, protozoa, <i>Paramecium</i> , macrophages, and many other eukaryotic cells)	Bacteria <i>Legionella pneumophila</i>	Eukaryote-like regulatory “effector” proteins injected in the course of infecting eukaryotic cells
Virus-host cell		
Halovirus	<i>Halobacterium salinarum</i> (Archaea)	B-type DNA polymerase B1
T3/T7 family bacteriophage (bacterial virus)	Ancestral eukaryote nuclear and mitochondrial genomes	Mitochondrial replication and transcription activities
Double-stranded RNA viruses (<i>totiviruses</i> and <i>partitiviruses</i>)	Eukaryote nuclear genomes (plants, arthropods, fungi, nematodes, and protozoa)	Capsid protein and RNA-dependent RNA polymerases
Circular single-stranded DNA viruses (<i>geminiviruses</i> , <i>nanoviruses</i> and <i>circoviruses</i>)	Eukaryotic nuclear genomes (plants, fungi, animals and protists)	Replication initiation protein (Rep)-related sequences
Host cell-virus		
Bacteria	Temperate bacteriophages (capable of insertion into bacterial genome in repressed “prophage” state)	Host infectivity and virulence activities expressed in prophage state
Bacteria (primarily, endosymbionts and parasites), bacteriophages, protists, animals	Nucleocytoplasmic large DNA viruses (NCLDV), including <i>Poxviruses</i> and <i>Iridoviruses</i> infecting insects and vertebrates, <i>Mimiviruses</i> and <i>Phycodnaviruses</i> infecting amoebae, protists and algae	Diverse functions including rDNAs, nucleic acid metabolism, signaling and regulatory activities, defenses against apoptosis, immune response, growth factors, and resistance of cells to oxidative stress
Mammal	Influenza virus	28S rDNA insert in hemagglutinin sequence increases pathogenicity
γ -proteobacteria and insects (viral hosts)	Baculovirus	DNA ligase, ribonucleotide reductase 1, SNF2 global transactivator, inhibitor of apoptosis, chitinase, and UDP-glucosyltransferase
Marine phytoplankton <i>prasinophytes</i>	Prasinovirus	Glycosyltransferases, methyltransferases and amino acid synthesis enzymes

A more complete, detailed and fully referenced version of this table is available as Supplementary Table S7.

At the end of Table 7, there are several examples of horizontal transfers in both directions between viruses and various kinds of organisms. These cases were included to emphasize the point that the biosphere contains a large number of infectious agents that can be vectors for DNA transfer, and which may play a role in the origination of novel cell-based organisms [243,244]. The recently discovered “nucleocytoplasmic large DNA viruses” (NCLDVs) are particularly notable in this regard and have genomes in the megabase-pair range that readily acquire cellular DNA from all three major domains of life [245]. Many NCLDVs infect amoeba and other protists that are phagocytic, engulf other microorganisms, and also serve as hosts to bacteria like *Legionella*, *Helicobacter*, *Listeria*, and *Mycobacterium*, which also infect animal cells [246–248]. The amoeba-NCLDV-bacteria realm of the biosphere thus provides opportunities to mobilize DNA across multiple taxonomic barriers and, accordingly, has been characterized as a genomic “melting pot” [249,250].

In addition to DNA encoding specific adaptive traits, there is also abundant horizontal transfer of genome modification potential in the form of different classes of mobile DNA elements [251–253]. The horizontal transfer of mobile DNA is a further indicator of the range of genome connections between different organisms at all taxonomic levels: bacteria-eukaryotic microbes [254], metazoans-eukaryotic microbes [255], birds-nematodes [256], insect-mammal [257,258], mammal-insect [259], crustacean-insect [260], insect-insect [260,261], vertebrate-vertebrate [262,263], mammal-mammal [264,265], arthropod-plant [266], plant-plant [267,268], and vertebrates-invertebrates (and invertebrate viruses) [269–271].

All of the examples of horizontal DNA transfers presented above reinforce the two major themes that are articulated in Section 3. First of all, significant adaptive changes result from biological activities that mobilize DNA molecules between cells and across taxonomic boundaries. Secondly, the multiple genome connections that are documented across the biosphere add to the evidence showing how unrealistic has become the conventional view of evolution occurring solely by vertically inherited genome changes within an isolated species. Increasingly, we recognize that genome evolution results from a networked, or reticulate, process that connects multiple lineages [272].

6. Genome Writing by Natural Genetic Engineering—Protein Evolution by Natural Genetic Engineering, Exon Rearrangements and Exon Originations

Based on the central role of proteins in executing cellular biochemistry and determining adaptive organismal functions, a major focus of molecular evolution studies has been the dynamics of innovation in genomic protein-coding capacity. In Section 5, we saw how organisms could transfer the ability to synthesize different proteins across taxonomic boundaries. But, horizontal DNA transfer does not address fundamental questions of how living organisms originate and modify protein-coding sequences to confer novel or improved functionalities. While early theories of protein evolution were based on the sequential accumulation of individual amino acid changes, the last three decades have revolutionized our understanding of protein organization, protein coding in the genome, and cellular activities that reorganize and innovate protein coding DNA elements.

6.1. The Modular Domain-Based Structure of Proteins

Protein biochemistry and genomics have led to the recognition that the majority of proteins are not unitary functional entities, but are, instead, modular structures that are composed of distinct “domains,” each conferring a discrete aspect of overall protein activity [273–276]. For example, one domain may bind to a particular class of molecular substrate, another domain may contain the amino acids that execute the protein’s catalytic activity, a third domain may bind regulatory signaling molecules, and a fourth domain may connect the protein to other macromolecules in a multi-component complex at a defined subcellular location. In other words, each protein’s functionality is a combinatorial systems phenomenon reflecting the integrated properties of and interactions between its various domains.

Table 8 lists some of the most frequently cited protein domain search terms that are listed in the PubMed database (<https://www.ncbi.nlm.nih.gov/pubmed/>). While the kinase domain indicates a particular catalytic activity, it is significant that the other highly cited domains determine cellular localization or the specificity of interactions with other proteins, DNA, or lipids. The role of protein domains as combinatorial tools for molecular recognition, especially between proteins [277–279], makes them essential components in the elaboration of highly specific cell signaling [280,281] and regulatory networks [282,283]. Protein connectivity networks that are based on evolutionarily mobile domain modules exemplify the value of thinking about biological functionalities as information systems.

Table 8. Highly cited protein domain search terms in the PubMed database.

Domain	Function
Kinase domain	Catalytic (phosphorylation)
Transmembrane domain	Subcellular localization
PDZ domain	Protein-protein interaction module
SH3 domain	Proline-rich protein-protein interaction module
DNA binding domain	Molecular recognition
PAS domain	Sequence-specific DNA binding
WW domain	Protein-protein interaction module

Table 8. Cont.

Domain	Function
PH domain	Phosphatidylinositol binding in membranes
SET domain	Binding methylated DNA
SH2 domain	Binding proteins containing phosphor-tyrosine residues

A fully referenced version of this table is available as Supplementary Table S8.

Each domain has its own phylogeny, and homologous domains can occur in functionally different proteins in combination with other divergent domains [284]. There are protein databases dedicated to identifying individual domains rather than complete proteins [285,286], and it is common practice in contemporary comparative genomics to describe a coding region and its cognate protein product by its domain content. While many domains are shared across broad phylogenetic distances, patterns of multi-domain architectures are specific to each taxon [287,288].

The modular nature of proteins as domain combinations means that basic protein evolution questions must be posed at several levels:

- How does a particular domain combination vary as it is transmitted within and between taxonomic groups?
- How do different combinations of domains assemble?
- How does each separate domain originate and enter into its various protein contexts?

Question (i) indicates that proteins evolve by gaining or losing domains [289], which occurs most commonly at either end of the protein chain [290]. Gain goes by the term “domain accretion” [291], and documentation of this process in the course of eukaryotic protein evolution from yeast through nematodes and *Drosophila* to man was highlighted in Figures 42 and 45 of the original publication of the draft human genome sequence [292]. Domain accretion is perhaps the most common mechanism for proteins to acquire additional activities or other properties, such as protein-protein interaction (Table 8). Domain loss is less well documented, and apparently represents a mode of protein specialization.

Question (ii) indicates that novel protein functionalities can arise rapidly by the aggregation of domains in particular combinations and that these combinations are key features in the evolutionary history of various taxa, such as metazoa [293], vertebrates [294], and plants [295]. Naturally, domain-coding sequences are subject to horizontal transfer [296]. Question (ii) further implicates the operation of molecular processes that join discrete domain-coding elements into combinatorial genomic determinants for adaptive functions [297,298]. These combinations include domain repetitions, which play a special role in molecular pattern recognition by proteins [299,300]. Repetitive domains are more common in eukaryotes than prokaryotes [301], but are found in all three kingdoms of cellular life [302]. Section 6.2 will discuss the various molecular mechanisms that organisms possess to assemble and rearrange domain combinations, a process that is labeled “domain shuffling”, which was highlighted in Figures 42 and 45 of the original publication of the draft human genome sequence [292]. It goes without saying that both domain accretion and domain shuffling are active examples of biological RW genome authorship.

Question (iii) posits mechanisms capable of generating a complex genetic coding element that did not previously exist. This may be the deepest question of all and represents the highest degree of RW genome creativity in protein evolution. We will discuss some of the currently recognized *de novo* domain origination processes in Section 6.3.

6.2. Protein Evolution by Exon Shuffling and Exon Accumulation, Changes to Alternative Splicing Patterns, and Insertion of Reverse-Transcribed Coding Sequences [303]

A major advance in understanding protein domain rearrangements in evolution was the discovery from early DNA sequencing studies that protein coding sequences in many eukaryote genomes are not

continuous, but are rather composed of DNA segments that are expressed into partial protein chains (“exons”) separated by intervening non-coding DNA segments (“introns”) [304,305]. The segmented exons are stitched together into a continuous protein coding sequence prior to translation at the messenger RNA level by removal of the introns (“splicing”) [306]. Although the correspondence is not absolute, there is a high correlation between the boundaries of exons and of the encoded protein domains [307]. Thus, for organisms with discontinuous protein coding sequences, questions about domain behaviors in evolution can be reframed in terms of exon and intron behaviors at the DNA level. Virtually all of the studies on the processes underlying exon accretion, exon swapping, and exon origination have been done with those organisms. In our discussion, the term “genetic locus” replaces the more common “gene” to emphasize the composite nature of each coding region. As we shall see, a genetic locus may encode multiple protein products and thus not be a unitary genomic coding element, as the term “gene” generally implies.

A key feature of discontinuous protein coding is that mRNA splicing patterns are not necessarily fixed. The splicing apparatus can skip the splice sites (intron/exon boundaries) demarcating one or more exons in a particular coding region [308,309]. Consequently, alternative splicing of primary transcripts from a single locus can produce mRNAs containing different exon combinations encoding proteins with different domain architectures [310]. Alternative splicing thus enriches the protein repertoire [311], and it is intricately regulated in response to diverse ecological [312,313] and developmental signals [314,315].

Alternative splicing patterns evolve and confer different coding potentials on duplicated loci within a genome [316]. A recent paper ascribes a major role to alternative splicing in the evolution of phenotypic novelty [317]. Insertions of mobile DNA elements into introns frequently induce alternative splicing changes [318,319], and the acquisition of new introns has a similar effect [320]. As with other features of genome function, hybrid speciation and WGD events frequently lead to new patterns of alternative splicing, thereby diversifying the expression of duplicated copies of multiple genetic loci in a single episode of accelerated genome modification [321,322]. The effects of local duplications and WGDs on the evolution of domain architectures have been analyzed in yeast [323], *Drosophila* [324], and plants [325].

The enrichment of alternative splicing patterns is particularly characteristic of vertebrate evolution, which was apparently stimulated by the two successive WGD events at the origins of the vertebrate lineage [136]. These WGDs may explain why vertebrates display more variation in alternative splicing than do invertebrates [326]. A study of O₂-binding globin domains in various hemoglobin, myoglobin, and cytoglobin proteins that are encoded by vertebrate genomes implicated the two successive vertebrate WGDs in the evolution of a broad diversity of vertebrate oxygen transport systems [327]. In addition, various globin proteins have evolved domain fusions and globin domain repeats [328].

Domain rearrangements, protein fusions, and protein splits have been documented by genomic analysis in prokaryotes (“3.8 million potential fusions across 11,473 genomes”) [329], fungi [330], *Drosophila* [331], metazoa [332,333], and the human genome [334,335], as well as in the virosphere [336]. In addition to shuffling exons from cell genomes, the virosphere provides an external source of exons encoding protein domains that are not found in cellular organisms [337].

New domain/exon arrangements arise by a wide variety of natural genetic engineering (NGE) processes. DNA level exchanges can generate fused genetic loci combining exons to encode chimeric proteins [338]. Exclusively at the DNA level, NGE can involve non-allelic homologous recombination (NAHR) events between dispersed intronic repeat elements in the genome [339] or non-homologous recombination between introns [305,340]. Non-homologous intron-intron recombination can occur during repair of DS breaks by the well-studied process known as “non-homologous end-joining” (NHEJ) [8,341,342].

While we are accustomed to thinking of evolutionary change as a DNA level process, it is difficult to overstate the importance of reverse transcription-based events in eukaryotic evolution. Many novel domain arrangements have been found to originate at the RNA level and involve the insertion of

reverse-transcribed cDNA copies into the genome (“retroposition”) to generate loci encoding chimeric fusion proteins [343–354]. Retroposed coding sequences have been documented in nematodes [355], multicellular green algae *Volvox carteri* and *Chlamydomonas reinhardtii* [356], plants [357], silkworm [358], non-mammalian chordates [359], zebrafish [360], mammals [361], and primates [362,363]. Retrocopying is reported to have played a special role in the evolution of highly variable sex chromosomes [364]. Mammalian evolution has been impacted by episodic bursts of retrocopy formation [365]. One of these bursts occurred in the primate lineage that was leading up to the appearance of humans [366]. In primate genomes, there are approximately 17,500 retrocopies, of which, at least 3600 are expressed, many with “tissue-specific and even species-specific expression patterns” [367].

The innovative role of reverse transcribed cDNAs is further enhanced by the well-documented phenomenon of “trans-splicing”, which involves the joining of sequences from two distinct transcripts [368–379]. Genome insertion of cDNA reverse-transcribed from a trans-spliced RNA molecule generates a novel chimeric domain coding sequence [379,380]. The locus encoding the chimeric Jingwei protein in *Drosophila* apparently arose in this manner by adding exons to a processed retrocopy of the alcohol dehydrogenase (*Adh*) mRNA [381], and trans-splicing-generated protein rearrangements have been significant in vertebrate evolution [382].

In addition to the NGE processes outlined above, diverse mobile DNA elements, both DNA transposons and retrotransposons, play active roles in exon rearrangements (Table 9).

Table 9. Selected Reports of Exon Rearrangements and Retroposition by Mobile DNA element Activity.

Rearrangements	Taxa
Group II intron retrotransposition and exon shuffling	Yeast
Chimeric LINE-mediated retrogene formation	Rice blast fungus <i>Magnaporthe grisea</i>
DNA “splicing” by transposase-related excision functions	Ciliated protists
Exon shuffling by DNA transposons (Pack-MULE, Ac/Ds, CACTA, Helitron)	Plants, including beans and maize
LINE-mediated duplications	Dicots
Retrotransposon-mediated exon shuffling	Maize, <i>Medicago sativa</i>
Exon shuffling by DNA transposons (Helitron, FB transposon)	Lepidoptera, <i>Drosophila</i>
Retrotransduction by LINES and SINEs	<i>Drosophila</i>
Exon shuffling by retrotransposition	Mammals
Retrotransposon exon shuffling	Primates
TRIM5-Cyclophilin A (TRIMCyp) fusion by retrotransposition	Tree shrews and owl monkeys
Exon shuffling by retrotransposition	Humans
L1/LINE mediated retrotransduction	Humans
<i>Alu</i> or <i>SVA</i> SINE-mediated retrotransduction	Humans
Chimeric “retrogenes”, by LINE-mediated template switches at the RNA level	Humans

A more detailed and fully referenced version of this table is available as Supplementary Table S9.

6.3. Protein Evolution by Domain/Exon Origination

Exon shuffling and alternative splicing generate functional protein novelties by rearranging existing coding elements. While those Lego-like processes have played important roles in evolutionary variation, they do not tell us anything about how novel domain coding sequences arise in the first instance. This is a topic where our ignorance is much greater than our knowledge. Conventional gradualist evolutionary theories would postulate that the novel domains emerge by the accumulation of nucleotide substitutions converting an exon encoding one sequence of amino acids into an exon encoding an entirely different polypeptide. While there is abundant evidence that related exons often undergo minor sequence changes that modify the encoded amino acid composition, often in adaptive ways (e.g., producing changes in DNA- or protein-binding specificities), there is genomic evidence

indicating that other processes generate new exons more rapidly from both coding and non-coding precursor sequences.

The de novo origination of novel domains or complete novel protein-coding sequences is indicated in genome databases by the presence of so-called “orphan” coding sequences (also called ORFans, from ORF = “open reading frame” or uninterrupted polypeptide coding sequence) [383,384]. These are sequences that appear abruptly in the phylogenetic record without any apparent homology to sequences in ancestral lineages; consequently, they have a restricted taxonomic distribution. Orphan coding sequences were initially noted in the yeast genome, but have subsequently appeared in every class of sequenced genome (Table 10), and there is an ORFan database [385]. It is possible that many orphan sequences do not actually indicate de novo coding sequence originations, but instead appear phylogenetically isolated due to gaps in the sequence databases. In particular, unknown coding sequences may abruptly appear due to horizontal DNA transfer from an as yet unsequenced genome [386]. Nonetheless, there are many well-documented examples where the appearance of novel coding sequences can be traced back to their origins in the genome. The orphan examples in Table 10 indicate a variety of ways that coding and non-coding sequences can serve as precursors for novel exons:

- “Overprinting” of an existing exon by translating the sequence in a new reading frame [387];
- Cooption of the antisense strand from an expressed locus [388];
- Loss of a stop codon at the 3' end of a terminal exon, thereby extending the protein C terminus by continuing translation into previously non-coding sequence [389];
- Cooption of non-coding but transcribed sequences [390,391], especially transposable elements [392–394]; and,
- Retroposition [356,395,396], frequently from non-coding RNA [391,397], which may account for some ORFans arising from non-coding transcribed regions [390]. The novel coding potential of long non-coding RNAs has been cited as a general phenomenon [398], and we will see in Section 7 how abundant these RNAs are in eukaryotic genomes.

Table 10. Some Independently Reported Instances of Orphan Coding Sequences (“New Genes” and New Exons) Discovered in Sequenced Genomes.

Organism(s)	Characterization of Orphan Coding Sequences
Viruses (bacteriophages)	“Almost one-third of all ORFs in 1456 complete virus genomes correspond to ORFans ... 38.4% of phage ORFs have no homologs in other phages, and 30.1% have no homologs neither in the viral nor in the prokaryotic world ...”
Viruses	“ <i>de novo</i> genes ... in which an existing gene has been “overprinted” by a new open reading frame, a process that generates a new protein-coding gene overlapping the ancestral gene”
Prokaryotes (Archaea and Eubacteria)	20,000 orphan sequences: “... only 2.8% of all microbial ORFans have detectable homologs in viruses, while the percentage of non-ORFans with detectable homologs in viruses is 7.9%, a significantly higher figure.”
<i>Escherichia coli</i> O157:H7 (EHEC), <i>Escherichia coli</i> K12	“72 genes are taxonomically restricted and, therefore, appear to have evolved relatively recently de novo” ... “origin of a new gene through overprinting in <i>Escherichia coli</i> K12”
<i>Saccharomyces cerevisiae</i>	“BSC4 ... encoding a 132-amino-acid-long peptide ... no homologous ORF in ... closely related species ... Because the corresponding noncoding sequences in <i>S. paradoxus</i> , <i>S. mikatae</i> , and <i>S. bayanus</i> also transcribe, we propose that a new de novo protein-coding gene may have evolved from a previously expressed noncoding sequence.”
Yeast (<i>Saccharomyces cerevisiae</i>) and <i>Drosophila</i>	Protein C-termini: the co-option of short segments of noncoding sequence into the C-termini of existing proteins via the loss of a stop codon: “...54 examples of C-terminal extensions in <i>Saccharomyces</i> and 28 in <i>Drosophila</i> ... Four of the <i>Saccharomyces</i> C-terminal extensions (to ADH1, ARP8, TPM2, and PIS1) ... are predicted to lead to significant modification of a protein domain structure.”
Green multicellular algae <i>Chlamydomonas</i> and <i>Volvox carteri</i>	PHD domain added to condensin II by exonization of mobile DNA sequences; “141 retrogene candidates in total in both genomes, with their fraction being significantly higher in the multicellular <i>Volvox</i> .”

Table 10. Cont.

Organism(s)	Characterization of Orphan Coding Sequences
<i>Plasmodium vivax</i>	"... recent de novo origin of at least 13 protein-coding genes in the genome of <i>Plasmodium vivax</i> ... five of the genes identified in our analysis contain introns ... likely evolved from previously intergenic regions together with the coding sequences."
Nematode <i>Pristionchus pacificus</i>	"3818–7545 (39–76%) of orphan genes are under negative selection"
<i>Drosophila melanogaster</i>	"... a significant excess of retrogenes that originate from the X chromosome and retropose to autosomes; new genes retroposed from autosomes are scarce ... most of these X-derived autosomal retrogenes had evolved a testis expression pattern."
<i>Drosophila melanogaster</i>	"142 segregating and 106 fixed testis-expressed de novo genes in a population sample of <i>Drosophila melanogaster</i> ... appear to derive primarily from ancestral intergenic, unexpressed open reading frames (ORFs), with natural selection playing a significant role in their spread."
<i>Drosophila melanogaster</i>	"... six putatively protein-coding de novo genes ... two de novo genes emerged from novel long non-coding RNAs ... four other de novo genes evolved a translated open reading frame and transcription ... suggesting that nascent open reading frames (proto-ORFs) ... can contribute to the emergence of a new de novo gene"
Insects: arthropod genomes, focusing on seven recently sequenced ant genomes ... comparison between social <i>Hymenoptera</i> (ants and bees) and nonsocial <i>Diptera</i> (flies and mosquitoes) ...	"... between the two insect orders <i>Hymenoptera</i> and <i>Diptera</i> , orphan genes are more abundant and emerge more rapidly in <i>Hymenoptera</i> , in particular, in leaf-cutter ants. With respect to intragenomic localization, we find that ant orphan genes show little clustering ..."
Entelegyne spiders (Araneae, Entelegynae)	"... transcriptomes of six entelegyne spider species from three genera (<i>Cicurina trivisa</i> , <i>C. vibora</i> , <i>Habronattus signatus</i> , <i>H. ustulatus</i> , <i>Nesticus bishopi</i> , and <i>N. cooperi</i>) ... between ~550 and 1100 unique orphan genes were found in each genus."
Rodents	"75 murine genes (69 mouse genes and 6 rat genes) ... good evidence of de novo origin since the divergence of mouse and rat. Each of these genes is only found in either the mouse or rat lineages, with no candidate orthologs nor evidence for potentially-unannotated orthologs in the other lineage ... For 11 of the 75 candidate novel genes we could identify a mouse-specific mutation that led to the creation of the open reading frame (ORF) specifically in mouse ... A large number of them (51 out of 69 mouse genes and 3 out of 6 rat genes) also overlap with other genes, either within introns, or on the opposite strand."
Mouse and human	"... over 5000 new genes were integrated into the ancestral GGI {gene-gene interaction} networks of human and mouse"
Primates	"an unexpected important role of transposable elements in the formation of novel protein-coding genes in the genomes of primates."
Human and Chimpanzee	"... retrocopies of coding transcripts to generate proteins with novel N-terminal domains. Examples include thymopoietin beta (<i>TMPO</i>), eukaryotic translation initiation factor 3 subunit 5 (<i>EIF3F</i>), and the 5'-inverted retrocopy of small nuclear ribonucleoprotein polypeptide N (<i>SNRPN</i>).
Humans	"... human-specific de novo protein-coding gene, <i>FLJ33706</i> (alternative gene symbol <i>C20orf203</i>) ... originated from noncoding DNA sequences: insertion of repeat elements especially <i>Alu</i> contributed to the formation of the first coding exon and six standard splice junctions on the branch leading to humans and chimpanzees, and two subsequent substitutions in the human lineage escaped two stop codons and created an open reading frame of 194 amino acids."
Humans	"... 60 new protein-coding genes that originated de novo on the human lineage since divergence from the chimpanzee ... highest expression levels in the cerebral cortex and testes ..."
Humans	"24 hominoid-specific de novo protein-coding genes with precise origination timing in vertebrate phylogeny ... most of the hominoid-specific de novo protein-coding genes encoded polyadenylated non-coding RNAs in rhesus macaque or chimpanzee with a similar transcript structure and correlated tissue expression profile ..."
Humans	"... de novo origin of at least three human protein-coding genes since the divergence with chimp ... chimp, gorilla, gibbon, and macaque share the same disabling sequence difference, supporting the inference that the ancestral sequence was noncoding."
Humans	"... 426 different annotated young domains, totaling 995 domain occurrences, which represent about 12.3% of all human domains. We have observed that 61.3% of them arose in newly formed genes, while the remaining 38.7% are found combined with older domains ... Young domains are preferentially located at the N-terminus of the protein ..."
<i>Arabidopsis</i>	"... lineage-specific genes within the nuclear (1761 genes) and mitochondrial (28 genes) genomes are identified ... Almost a quarter of lineage-specific genes originate from non-lineage-specific paralogs, while the origins of ~10% of lineage-specific genes are partly derived from DNA exapted from transposable elements (twice the proportion observed for non-lineage-specific genes). Lineage-specific genes are also enriched in genes that have overlapping CDS, which is consistent with such novel genes arising from overprinting. Over half ... of the 958 lineage-specific genes (LSGs) ... in <i>Arabidopsis thaliana</i> have alignments to intergenic regions in <i>Arabidopsis lyrata</i> , consistent with either de novo origination or differential gene loss and retention ... LSGs are enriched for genes responsive to a wide range of abiotic stresses ..."

Table 10. Cont.

Organism(s)	Characterization of Orphan Coding Sequences
<i>Arabidopsis</i>	:“...new genes ... show a bias in expression to mature pollen ... Transposable elements are significantly enriched in the new genes ... high activity of transposable elements in the vegetative nucleus, compared with the germ cells, suggests that new genes ... generated in the vegetative nucleus in the mature pollen. We propose an “out of pollen” hypothesis for the origin of new genes in flowering plants.

A more complete and fully referenced version of this table is available as Supplementary Table S10.

The importance of mobile DNA elements in protein evolution was initially recognized when Nekrutenko and Li demonstrated the presence of numerous transposable element contributions to the coding sequences for many human proteins [399]. Since then, the discovery of mobile DNA sequences contributions to proteins has subsequently expanded (Table 11), especially in the human genome [400]. In addition to providing novel protein-coding exons, mobile element-derived non-coding exons can also modulate transcription and confer novel tissue-specific splicing patterns on human transcripts [401]. The relatively frequent exonization of segments from certain transposable elements results from the presence of cryptic splice sites that are embedded in their sequences [402]. But not all cases of exonization from a particular element are identical, and the same element can contribute different sequences to novel coding loci, as exemplified by the many cases of *Alu* exonization that are found in primate and human genomes [403–406].

Table 11. Origination of Novel Exons (“Exonization”) from Mobile DNA Elements.

Taxa	Mobile DNA Exonized
Green algae	Transposable elements (TEs)
Plants (coffee, rice, <i>Arabidopsis</i> , etc.)	<i>Ds</i> transposons and other transposable elements
Ancestral vertebrate	“... a more than 200-base-pair ultraconserved region, 100% identical in mammals, and 80% identical to the coelacanth SINE, contains a 31-amino-acid-residue alternatively spliced exon of the messenger RNA processing gene <i>PCBP2</i> ...”
Mammal	“Although ... not evolutionarily related, mammalian <i>TMPO</i> and <i>ZNF451</i> ... both code for splice isoforms that contain LAP2alpha domains ... related to the first ORF from a <i>DIRS1</i> -like retrotransposon ... domestication happened separately and resulted in proteins that combine retrotransposon and host protein domains. The alternative splicing of the retrotransposed sequence allowed the production of both the new and the untouched original isoforms ...”
Mammal, Primate	<i>MIR</i> retrotransposons
Mouse	<i>L1</i> retrotransposons; “antisense insertions results in an increased potential for exonization”
Rat	<i>L1</i> retrotransposon and ERV (endogenous retrovirus) in embryonically expressed <i>Rtdpoz-T1</i> and <i>-T2</i> locus
Human and mouse	“... exonization of transposed elements is biased towards the beginning of the coding sequence in both human and mouse genes ... cases of primate-specific <i>Alu</i> elements that depend on RNA editing for their exonization ...”
Primate	<i>Alu</i> exonization in <i>BCSIL</i> , RNA edited <i>Alu</i> element in human nuclear prelamins A recognition factor gene transcript
Primate	LINE retrotransposon in <i>ZRANB2</i> locus
Human	Anti-sense <i>Alu</i> SINEs; <i>Alu</i> -derived segments in two Bcl-family proteins.
Human	“Exons derived from <i>Alu</i> SINEs but also the exons from the TEs of other families were preferentially established in zinc finger (ZNF) genes.”

Table 11. Cont.

Taxa	Mobile DNA Exonized
Human	“Long Terminal Repeat (LTR) retrotransposons are associated with 1057 human genes (5.8%). In 256 cases LTR retrotransposons were observed in protein-coding regions, while 50 distinct protein coding exons in 45 genes were comprised exclusively of LTR RetroTransposon Sequence (LRTS) ... an alternatively spliced exon of the Interleukin 22 receptor, alpha 2 gene (<i>IL22RA2</i>) derived from a sequence of retrotransposon of the Mammalian apparent LTR retrotransposons (<i>MaLR</i>) family ... hypothesize that the recruitment of the part of LTR as a novel exon ... a result of a single mutation in the proto-splice site ...”
Human	“... human nuclear prelamina A recognition factor contains a primate-specific <i>Alu</i> -exon that exclusively depends on RNA editing for its exonization.”

A more complete, detailed and fully referenced version of this table is available as Supplementary Table S11.

The heavy bias towards transposable element exonization in the generation of novel protein coding-sequences is probably due, at least in part, to the overwhelming emphasis in genomic studies on the analysis of the human genome, where *Alu* elements, other SINEs (such as *SVA*), and LINE elements predominate in the retrotransposed fraction of the genome [292,344,407–409]. Nonetheless, the few examples that are documented in Tables 10 and 11 of transposable element exonizations in algae, plants, and non-primate vertebrates do hint at something special about transposable elements as sources of novel coding potential. A number of authors have speculated that mobile DNA elements play dedicated roles in providing rapid genomic variability during episodes of evolutionary change (e.g., [410,411]). This view is buttressed by the well-documented role that mobile DNA plays in rewiring genomic structural and regulatory networks as evolution produces increasingly complex multicellular organisms (Section 7).

7. Genome Writing by Natural Genetic Engineering: Mobile and Repetitive DNA Elements Actively Contributing to Genome Organization, Organismal Complexity and Genome Regulation

Our view of what genomes are has changed fundamentally since genetics and evolution first combined in the early 20th Century to form the Modern Synthesis [412]. At that time, geneticists could only examine genome function by the mutational analysis of phenotypes, recombinational mapping, and cytological examination of chromosome structure and behavior. The nature of genome analysis changed in the middle of the 20th Century, beginning with a series of key discoveries: DNA as the carrier of genetic information [413], the DNA double helix [414], regulation of DNA-encoded protein synthesis [415], and the existence of mobile DNA elements that were capable of altering genome structure and patterns of genome expression [416]. This section will review some consequences of these discoveries from a conceptual perspective leading to a detailed empirical discussion of the key roles that mobile repetitive DNA elements play in evolutionary variation.

7.1. Regulatory Studies Led to Recognizing the Syntactical Organization of Genomes

In the early days of molecular biology, there was a temptation to simplify genome functions to replication and protein coding: DNA makes more DNA and also RNA, RNA makes protein [13,14,417]. However, the reductionist view of genome action started to dissolve even as it was being formulated. Beginning with the study of how cells control protein synthesis in response to nutritional changes [415], it rapidly became clear in the 1960s and 1970s that there exist in bacteria, and in all organisms, elaborate multifactorial genomic codes that adaptively regulate the transcription of DNA sequences into RNA (e.g., [418,419]). As regulation studies extended to more complex eukaryotic organisms, the generic term “transcription factor” was adopted for proteins that interact with countless distinct binding sites to construct appropriate multimolecular complexes connecting coding sequences with signaling networks, and thereby controlling genome transcription [420,421].

With increasingly detailed studies of genome regulation in yeast and other eukaryotic organisms, the phenomenon of “epigenetic” control revealed further codes for organizing genomes into extended “euchromatic” and “heterochromatic” domains that were, respectively, accessible or inaccessible for transcription [422,423]. The application of sequencing-based technologies to genome function during the cell cycle and multicellular development uncovered yet more codes governing the three-dimensional organization of genomes during cellular differentiation [424–426]. From these, and many other lines of molecular biology research, it has become abundantly clear that every genome is a densely formatted molecular database, syntactically organized to respond appropriately to the dynamic conditions of cellular life.

7.2. Repetitive DNA Elements Provide Distributed Copies of Each Class of Regulatory Site

Contemporaneously with the early expansion of genome regulatory studies, purely physical studies of DNA extracted from different organisms unexpectedly revealed the presence of significant fractions of multi-copy repeated sequence components. These repeated sequences were present at higher concentrations than unique sequences, such as those encoding proteins. They were detected because single strands that were carrying repetitive DNA sequences renatured to a double helical configuration more rapidly following denaturation than single-copy sequences [427–429]. By closely following DNA renaturation kinetics, it was possible to estimate the fractional composition of repetitive sequences in DNA extracted from any particular organism. The fraction could be considerable. Advances in whole-genome sequencing technologies over the last two decades enabled far more precise measurements of total DNA content and the fractional contribution of different classes of DNA sequence components, in particular, genomes [430–432].

Special attention has been paid to repetitive DNA elements in genome sequencing studies [433–435]. Data on animal genomes from one database shows a wide variation in repetitive DNA content across taxa (Table 12). Repeats include tandem arrays of DNA at centromeres, telomeres, and loci encoding structural RNAs for ribosomes and nucleoli, but in almost all species, the vast majority of repetitive DNA consists of “interspersed” mobile DNA elements distributed throughout the genome (the fractions in parentheses in Table 12). In mammals, generally considered the most highly evolved animal group, at least one-third of the genome consists of repetitive DNA, and some primates (including humans), as well as other mammalian species have more than half of their genomes as repetitive sequences [408]. Other vertebrate genomes vary from a low of 1.2% in spotted gar, a primitive fish, to almost 38% in reptiles [436–440].

Table 12. Repetitive DNA Content of Some Annotated Animal Genomes (data source—<http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasetsAlt.html>).

Species	Genome Size	Fraction Total Repetitive DNA (Interspersed Mobile DNA)
Primates		
Human <i>Homo sapiens</i>	3,049,315,783 bp	52.58% (48.49%)
Orangutan <i>Pongo pygmaeus abelii</i>	3,093,543,172 bp	52.16% (48.79%)
Gorilla <i>Gorilla gorilla gorilla</i>	2,822,760,080 bp	49.43% (46.12%)
Gibbon <i>Nomascus leucogenys</i>	2,756,609,047 bp	51.96% (47.90%)
Chimp <i>Pan troglodytes</i>	2,902,338,967 bp	51.72% (48.77%)
Other Mammals		
Mouse <i>Mus musculus</i>	2,652,783,500 bp	45.03% (41.73%)
Cow <i>Bos taurus</i>	2,804,673,174 bp	49.38% (47.98%)
Killer whale <i>Orcinus orca</i>	2,249,582,112 bp	44.53% (43.23%)
Dolphin <i>Tursiops truncatus</i>	2,332,402,443 bp	44.00% (41.24%)
Squirrel <i>Spermophilus tridecemlineatus</i>	2,311,060,300 bp	36.30% (34.32%)
Elephant <i>Loxodonta africana</i>	3,118,565,340 bp	57.63% (56.38%)

Table 12. Cont.

Species	Genome Size	Fraction Total Repetitive DNA (Interspersed Mobile DNA)
Microbat <i>Myotis lucifugus</i>	1,966,419,868 bp	37.03% (35.51%)
Megabat <i>Pteropus vampyrus</i>	1,839,436,660 bp	35.82% (33.40%)
Birds		
Chicken <i>Gallus gallus</i>	1,032,854,810 bp	11.47% (9.74%)
Zebra finch <i>Taeniopygia guttata</i>	1,222,864,691 bp	8.45% (7.17%)
Mallard duck <i>Anas platyrhynchos</i>	1,069,972,754 bp	6.53% (4.69%)
Reptiles		
Lizard <i>Anolis carolinensis</i>	1,701,353,767 bp	36.01% (34.26%)
Alligator <i>Alligator mississippiensis</i>	2,098,626,832 bp	37.68% (36.96%)
Gharial crocodile <i>Gavialis gangeticus</i>	2,139,715,393 bp	37.55% (36.86%)
Painted turtle <i>Chrysemys picta bellii</i>	2,158,289,746 bp	30.86% (30.20%)
Amphibian		
Western clawed frog <i>Xenopus tropicalis</i>	1,365,936,747 bp	34.91% (31.88%)
Fish		
Coelacanth <i>Latimeria chalumnae</i>	2,183,592,768 bp	9.13% (6.82%)
Stickleback <i>Gasterosteus aculeatus</i>	446,627,861 bp	5.78% (3.23%)
Fugu <i>Takifugu rubripes</i>	350,961,831 bp	8.66% (5.92%)
Spotted gar <i>Lepisosteus oculatus</i>	869,414,359 bp	1.20% (0.00%)
Lower vertebrate		
Lamprey <i>Petromyzon marinus</i>	647,368,134 bp	35.46% (30.35%)
Chordate		
Lancelet amphioxus <i>Branchiostoma floridae</i>	480,418,582 bp	13.63% (12.35%)
Insects		
Bee <i>Apis mellifera</i>	231,030,884 bp	6.21% (0.24%)
Mosquito <i>Anopheles gambiae</i>	263,156,584 bp	14.10% (11.34%)
Fruit fly <i>Drosophila melanogaster</i>	162,367,812 bp	28.61% (20.44%)
Mollusca		
Sea hare mollusk <i>Aplysia californica</i>	619,228,092 bp	13.00% (4.86%)
Echinoderm		
Sea Urchin <i>Strongylocentrotus purpuratus</i>	809,969,717 bp	18.00% (13.82%)
Nematodes		
<i>Caenorhabditis briggsae</i>	105,451,667 bp	20.07% (16.02%)
<i>Caenorhabditis elegans</i>	100,286,070 bp	12.59% (10.31%)
Tunicate		
<i>Ciona intestinalis</i>	141,233,565 bp	16.57% (14.82%)

Repetitive DNA elements that do not encode proteins had no place in classical genetic and evolutionary theories, and so were labeled “junk” or “selfish” DNA by some prominent scientists [441,442]. Nonetheless, the potential significance of repetitive DNA elements for establishing distributed regulatory networks became evident to certain molecular biologists soon after their discovery by renaturation kinetics [429]. Today, the recognition of repetitive DNA elements as genome formatting entities occupies a key role in the analysis of sequence data and functional genome organization by large-scale genomics projects like ENCODE (Section 7.4) [443]. From a theoretical perspective, maximum computational efficiency (e.g., to minimize the total number of distinct DNA recognition proteins) requires that the diverse DNA elements that control genome behavior must function similarly at many different genomic locations [444–446]. Thus, numerous copies of many distinct classes of DNA sequence must be present throughout each genome [447,448]. Even an organism with one of the smallest known cellular genomes, *Mycoplasma genitalium*, contains repetitive DNA [449].

By contrast with the large fractions of repetitive DNA in most eukaryotic genomes (Table 12), the protein-coding DNA content is often rather modest (under 1.5% in the human genome [450]). A semi-quantitative proxy for repetitive DNA involvement in evolutionary innovation turns out to be the fraction of the genome that is composed of non-protein-coding DNA. Unlike protein-coding

DNA sequences, which peak at 10^7 – 10^8 base-pairs per genome for all of the organisms sequenced, non-coding DNA sequences per genome continue to increase and scale with organismal complexity (measured by number of distinct cell types) up to levels as high as or greater than $2\text{--}3 \times 10^9$ base-pairs per genome [450,451]. This remarkable correlation means that the most complex organisms have the largest amounts of repetitive and non-coding DNA in their genomes. In other words, the organisms we assume to be the most highly evolved have apparently gained much of their complexity by accumulating non-coding DNA content rather than increasing the protein-coding capacity. This pattern indicates that non-coding DNA abundance contributes to the complexity of biological control circuits, both through well-documented roles as mobile transcriptional signals in the genome (Section 7.4), and, in other, more recently discovered, ways. As Section 7.5 explains, precisely such an unpredicted connection between mobile DNA repeats and biological control circuitry has appeared in the last two decades.

7.3. How Do Organisms Use Repetitive DNA for Genome Rewriting in Evolution? Dispersed Mobile DNA Elements

Conventional theories of evolutionary change based on the accumulation of random mutations due to copying errors in genome replication could not account for adaptive change in the repetitive DNA content of genomes. But, a strikingly appropriate (if at first poorly appreciated) solution was available in Barbara McClintock's discovery of mobile genetic elements in the maize genome [416,452]. These mobile components could transpose to novel locations, and thus accumulate throughout the genome [453]. Because they altered the patterns of expression from genetic loci where they were inserted, McClintock named her discoveries "controlling elements" [454]. She recognized that they could generate novel regulatory networks in the nucleus and pointed out parallels with the transcriptional control systems that were being studied in bacteria [455,456].

The discovery of analogous mobile DNA elements in bacteria, yeast, *Drosophila*, and eventually all of the species examined, made it clear by the last two decades of the 20th Century that living organisms possess the molecular tools to reformat their genomes by actively mobilizing repetitive DNA [1–6]. Any skepticism about the functionally adaptive role of mobile repetitive DNA elements has been answered by the accumulation of sequence data establishing connections between these elements and functional regulatory motifs in genomic DNA. The human genome is naturally the most intensively studied. In our cells, mobile genetic elements contribute to DNaseI-hypersensitive sites [457], transcription factor binding sites [458], enhancer elements [459], RNA splicing control [460], epigenetic control of transcription [461], and the establishment of the core human embryonic stem cell regulatory network [462].

There is a rapidly growing literature that is dedicated to the central role this class of mobile supposedly "non-coding" DNA plays in evolutionary variation in general [463], as well as in specific phylogenetic groups (Table 13). Repetitive DNA elements also play major roles in restructuring genome architecture. In part, this role results from providing many dispersed sites for homology-dependent recombination, but mobile DNA activity also generates rearrangements at non-homologous sites [3]. The role of a taxonomically-specific retotransposon was recently invoked to explain the rapid karyotype evolution distinguishing the Gibbon genome from other primates [464]. In other cases, speciation events have been attributed to mobile element activity: *Arabidopsis* [465], fish [466], *Anopheles gambiae* mosquitoes [467], and volcanic island radiations [468]. A number of contemporary evolutionary theorists have even come to ascribe a primary role to mobile DNA elements as drivers of evolutionary transitions [244,468–470].

7.4. Rewiring Transcriptional Regulatory Networks in Evolution of Complex Organisms

By virtue of their abundance and the capacity to move throughout the genome, mobile DNA elements have specific properties necessary for rewiring and innovating transcriptional regulatory networks in evolution. They are not tied to any specific phenotype and can reposition promoters,

enhancers, heterochromatin markers [471], insulators [472], splicing signals, and other *cis*-acting control elements that are embedded in their sequences [463,473,474]. It is noteworthy that biological triggers of speciation like interspecific hybridization and genome doubling activate mobile elements (Table 5) [475], as do a wide variety of ecological stressors [476] (<http://shapiro.bsd.uchicago.edu/StimuliDocumentedActivateNGE.html>). Consequently, network rewiring by mobile elements can connect to episodes of biosphere and ecological challenge. The literature contains a number of articles attributing evolutionary network innovation to mobile genetic elements in fungi, plants and animals (Table 13).

Table 13. Distributed genome network innovation attributed to mobile DNA elements.

Organisms	Phenotypes
18 Fungal Genomes	Whole-Genome Architecture and Transcriptional Profiles
Plants	Epigenetic Controls
Plants	C4 photosynthesis
Plants	Stress Response
Maize	Abiotic Stress Response
Maize	Helitron transposons reshuffle the transcriptome
Cotton	Fiber cell development
<i>Coffea</i>	Drought stress response
<i>Drosophila</i>	X chromosome dosage compensation
Mammals	Estrogen receptor network
Mammals	Pregnancy
Human	c-Myc regulatory subnetwork
Human	Core embryonic stem cell development

A fully referenced version of this table is available as Supplementary Table S13.

There is particularly compelling quantitative data on the roles of mobile DNA elements in evolution of certain well-studied mammalian regulatory networks [477–479]. A 2013 analysis of data from various human cell cultures found both phylogenetic and functional specificity of mobile DNA control regions: transposable elements contained 63% of the primate-specific sites for open chromatin (i.e., actively expressed regions), and fully 80% of the endogenous retrovirus (ERV) sequences in the human genome formed open chromatin in a cell type-specific manner, which was frequently associated with cell type-specific expression of neighboring genetic loci [480]. A bioinformatic approach to identify transcription factor binding sites found that approximately 110,000 are located in human ERVs or LTR retransposons [481]. Looking at specific stages and features of the mammalian life cycle provides a more detailed perspective on the roles of different types of mobile DNA in wiring regulatory networks, as well as giving an idea of how the literature has expanded in recent years, documenting the outcomes of natural genetic engineering in evolution.

7.4.1. Embryonic Stem Cells

A 2015 study of human embryonic stem cells (hESCs) identified fully 99.8% of the candidate human-specific transcription factor-binding sites within human-specific retrotransposable element-derived sequences, most notably LTR7/HERV-H, LTR5_Hs, and L1HS [459,482]. These observations confirmed the earlier reports that species-specific retrotransposons are (i) enriched in hESC-specific hypomethylated regions of the genome [483] and (ii) bind both transcriptional activators and repressors to shape hESC transcription [484,485]. A recent study analyzed the mouse-specific RLTR9 family of endogenous retroviruses (not present in rats), and found that

a significant fraction bound one and frequently several ESC-specific transcription factors, and experimentally verified that these mERVs provided an enhancer function in both synthetic constructs and for nearby ESC-specific loci [486]. When differentiated, mouse embryonic fibroblasts, human CD34(+) cells, or human primary hepatocytes are reprogrammed to become induced pluripotent stem cells (iPSCs), expression of all the endogenous retroelements is up-regulated so that their transcriptional profiles come closer to that of ESCs [487].

7.4.2. Early Embryonic Development

A 2017 analysis of 259 mouse embryonic cells at different stages from zygote to blastocyst reports enrichment of mobile DNA in expressed promoters and stage-specific utilization of different classes of mobile element [488]. At the 2–4 cell stage, LTR retrotransposons predominate and provide binding sites for homeobox transcription factors. For subsequent zygotic transcription upregulation, *B1* and *B2* SINE retrotransposons come into play as enhancer elements, paralleling the results for human *Alu* and bovine tRNA SINE elements in embryonic stem cells. A prior study similarly found that conversion of human embryonic stem cells (hESCs) to an epiblast-like state activates hundreds of blastocyst-specific hERV elements containing splice sites for linking LTR initiation transcripts to nearby exons for epiblast-specific functions [489].

7.4.3. Both Sides of the Fetal-Maternal Interface in Viviparous Reproduction

The *placental* tissue necessary for prenatal embryonic development depends on endogenous retroviruses for syncytial trophoblast development, as well as transcriptional programming. The essential “syncytin” proteins that fuse the cell membranes and provide an immune-protective barrier in the placenta are actually endogenous retroviral Env (envelope) proteins that are expressed from different ERV families in each mammalian order: “... the capture of *syncytin* or *syncytin*-like genes, sometimes as pairs, was found to have occurred independently from different endogenous retroviruses in diverse mammalian lineages, such as primates—including humans—muroids (rodents), leporids (rabbits), carnivores, caviids (S. American rodents), and ovis (sheep), between around 10 and 85 million years ago” [490]. At the transcriptional control level, in human placenta, there is hypomethylation of retrotransposon promoters for placenta-specific transcription [491]. Like *Env* exaptation for *syncytin* function, placental utilization of ERV promoters is also taxonomically-specific. In mouse, where detailed analysis of placental expressions patterns can be done, the species-specific ERV family RLTR13 contributes hundreds of enhancers for development of placental trophoblast stem cells (regulating over a third of all placenta-specific transcripts) [492]. Taxonomic specificity was so strong that a large majority of the mouse ERV enhancers were absent from rats.

The maternal or *uterine* side of viviparous reproduction displays a similar dependence on mobile DNA, but the elements are different. DNA transposons in the MER20 family provide enhancers for 200 uterine functions stimulated by progesterone and cAMP [493]. When all uterine-expressed Progesterone Receptor binding sites are tabulated, 1721 are found within Mammalian- or Eutherian-specific mobile DNA elements [494]. Although universally produced in the mammalian pituitary gland, certain mammalian orders have independently evolved the capacity to express the lactation-stimulating hormone prolactin from the uterus as well. Prolactin is not expressed from the uterus in rabbits, pigs, dogs, or armadillos, but it is in primates, mice, and elephants. Focusing on the promoter for uterine expression of decidual prolactin (dPRL) provides an interesting view of convergent mobile DNA behaviors in mammalian evolution. The human and spider monkey primate dPRL promoter contains a MER20 DNA transposon, as well as a MER39 LTR retrotransposon [495], while the mouse dPRL promoter derives from the MER77 LTR retrotransposon, and the elephant dPRL promoter originates from the lineage-specific LINE retrotransposon L1-2_LA [496].

These examples in mammalian viviparous reproduction of similar transcriptional wirings accomplished by non-homologous, species-specific families of mobile DNA elements exemplify an intriguing process of convergent evolution that remains to be explained. Do distinct mobile

elements come to execute functionally equivalent transcript rewiring in different mammalian orders purely by chance, or are the elements responding to some as-yet-to-be-defined regulatory process that guides the adaptive integration of newly established regulatory signals? It would be a useful exercise to calculate realistic probabilities for random insertions of mobile DNA elements into the appropriate genomic positions. Such calculations would allow for us to see whether that process constitutes a realistic basis for the repeated evolutionary accumulation of taxonomically-specific transcriptional circuits involving mobile DNA families. At the same time, we need to keep in mind that regulatory guidance of mobile DNA activity is not implausible because the targeting of mobile DNA insertions within genomes is a well-documented process (<http://shapiro.bsd.uchicago.edu/ExtraRefs.TargetingNaturalGeneticEngineeringInGenome.shtml>; http://shapiro.bsd.uchicago.edu/Targeting_retroviral_and_retrotransposon_insertions.html).

7.4.4. Brain and Nervous System Development

Since it was first proposed [497], several connections between SINE retrotransposons and mammalian neural development have been observed. A 2015 article reports the utilization of evolutionarily ancient MER130 SINEs to form an integrated enhancer network for embryonic development of mouse dorsal neocortex [498]. MER130 SINE elements are present in *Xenopus laevis* amphibians and reptiles (1135 copies in the green sea turtle genome) but not in fish, and so they arose close to the emergence of quadrupeds in vertebrate evolution. In the mouse, 23 of the total 90 MER130 copies in the genome act as experimentally verified developmental enhancers in the embryonic hindbrain, but not in the forebrain. The involvement of 24% of all genomic MER130 repeats in the dorsal brain development represents a 73-fold enrichment over random association with hindbrain-specific functions. The conclusion from this correlation is that the ancient MER130 elements were co-opted for this particular developmental function. That result is consistent with a 2016 survey analysis of mammalian neocortical enhancers that found a 30% overlap with mobile DNA repeats for eutherian-specific enhancers in more advanced mammals, as compared to less than a 6% overlap of older enhancers shared across all mammals [499].

A fascinating 2013 report links SINE elements to the epigenetic regulation of transcript elongation in the somatosensory cortex of young mice when they respond to novel enriched environmental (NEE) cues [500,501]. NEE stimulation modifies dendritic growth and synapse formation in brain neurons, providing a cellular basis for learning from experience [500]. The regulatory SINE elements were found located distal to NEE-induced promoters and embedded in silent chromatin that is subject to histone acetylation after NEE exposure. NEE-triggered acetylation relocates the SINE element chromatin to active transcription factories, and thereby facilitates the expression of NEE-induced proteins [472,500,502].

The role of SINE elements in the epigenetic control of neural transcript elongation in response to experience illustrates how versatile mobile DNA elements can be in regulating genome function, and how they can contribute to a higher nervous system phenotype like learning from experience [503,504]. A further connection between non-LTR retrotransposons and mammalian nervous system differentiation has been suggested to lie in activation of LINE-1 element retrotransposition. LINE-1 elements are silent in virtually all mammalian tissues, but they become active and retrotranspose to new locations in neurons [505]. This genome mobility creates somatic diversity within the nervous system and may prove adaptive by adding to neural complexity.

7.4.5. Innate Immunity

There are multiple connections between the mammalian immune systems and mobile DNA repeats. Although our topic here is the rewiring of transcriptional networks, it should be noted in passing that the ability of the adaptive immune system to cut and splice DNA to generate antibody diversity by VDJ joining of variable (V), diversity (D), and join (J) coding segments descended directly from the transposase activity of a DNA transposon belonging to the *Transib* superfamily, originating in

primitive metazoans [506,507]. While adaptive immunity in vertebrates uses mobile DNA functions to tailor antibodies specific to each particular invading organism or virus, the innate immune response in both plants and animals produces a generic response to infection that helps neutralize many different invaders [508,509].

In mammals, a key feature of innate immunity is a complex interferon (IFN)-dependent inflammatory response that attacks both invading pathogens and the host cells that they have infected to combat the spread of infection [510]. A recent study in human cells identified 962 MER41 family primate-specific endogenous retroviral elements as binding sites for IFN-triggered transcription factors [511]. Using CRISPR knockout technology, the authors directly demonstrated that four of these MER41 elements are essential for the expression of innate immunity proteins. In addition, luciferase reporter construct assays showed that MER41 and the LTRs of similar ERV elements from dog and cow drive IFN-inducible transcription in HeLa cells. Mice lack the MER41 ERV family, but the murid-specific endogenous γ -retrovirus RLTR30B also drives IFN-inducible transcription in the same assay system. Moreover, bioinformatic analysis of the mouse genome reveals a significant association of RLTR30B elements and functionally annotated immunity loci. So, it is difficult to escape the conclusion that these two distinct ERV families have contributed in a convergent manner to formatting the innate immune response networks in mice and humans.

7.5. Mobile DNA Elements Are Major Contributors to “Non-Coding” Regulatory RNA Molecules

It is only in the last couple of decades that molecular biologists have become broadly aware of a major class of genome regulatory molecules that are linked to mobile DNA elements. In 2003, the National Institutes of Health established the ENCODE project to analyse the informational content of rapidly growing human genome data. The name of the ENCODE project stands for Encyclopedia of DNA Elements [512–515]. The goal of ENCODE is to use modern genome sequence data to understand the manifold ways that DNA encodes biological properties, in the first instance of humans (based on the Human Genome Sequence) but more broadly of all forms of life. The name is well chosen because it avoids making assumptions about the nature of these DNA “elements” (as using a term like “gene” would have done). The neutral term “DNA element” came out of increasing experimental data showing that genomes do far more than replicate and encode proteins. In particular, molecular studies have documented a diverse and ever-growing repertoire of RNA molecules that do not encode proteins, but nonetheless contribute to determining cellular and organismal properties [516,517].

In 2007, ENCODE project researchers examined the expression of 1% of the human genome [518,519]. To their surprise, they found that 80% of the DNA sequences examined templated RNA transcripts, even though only <1.5% of the DNA actually coded for proteins [520]. This result posed a challenge to conventional views of genome content devoted primarily to encoding proteins. In 2012, the ENCODE project published a follow-up series of papers on the functionality of the full human genome [521]. The analysis confirmed the earlier results, showing that at least 76% of the human genome is transcribed into RNA, and expanded the repertoire of distinct RNA types that have been found [522,523]. As the news article reporting the 2012 analysis expressed in its title, “ENCODE project writes eulogy for junk DNA”, because it documented the active expression of most repetitive components of the human genome [443].

7.5.1. MicroRNAs

Among the first of the novel non-coding RNA products to be widely investigated in detail were short molecules (2–3 dozen nucleotides in length) that were grouped under the term “micro-RNAs”. MicroRNAs play key roles in guiding the epigenetic and transcriptional formatting of eukaryotic genomes, thereby affecting virtually every phenotypic character [524–526]. MicroRNAs provide target-specific guidance to genome formatting complexes by sequence complementarity with DNA or nascent RNAs. Examples of microRNA-regulated phenotypes include pluripotency [527], hematopoiesis [528], neural development [529,530], and DNA damage response [531] in mammals,

and heat stress response [532], innate immunity [533], and fruit development [534] in plants. Some microRNAs transfer intercellularly in membrane-bounded vesicles, where they serve as signaling molecules [535,536] and can affect properties, such as immune responsiveness [537].

As we might expect, the microRNA repertoire of each species is taxonomically specific [538–540], and thus is a key product of evolutionary variation [541–543]. A significant role for miRNAs was recently reported in sympatric evolution of naked mole rats [544]. Mobile and repetitive DNA elements encode or have contributed to the evolution of microRNAs in many different species [545–548]. In Vesper bats, for example, over 61% of taxon-specific microRNAs can be traced back to mobile elements, largely from bat-specific DNA transposons [549]. Mobile DNA-derived microRNAs arise from lineage-specific transposition events [550,551] that can occur in bursts of genome innovation [552]. From a functional point of view, it should be easy to appreciate that the presence of dispersed mobile DNA sequences in other classes of cellular transcripts, such as mRNAs, provides targets for recognition by complementary microRNAs that are derived from the same mobile element [553].

7.5.2. Long Non-Coding lncRNAs

Another major group of regulatory transcripts are called lncRNAs, for “long non-coding RNAs”, which are arbitrarily defined as molecules >200 nucleotides in length. Various lncRNA molecules have been linked empirically to the control and the expression of an extensive range of organismal phenotypes in eukaryotes, from the speed of metabolic induction in budding yeast to nervous system development in humans (Table 14). It is evident from the functions that are cited in Table 14 that some lncRNAs have rather specific effects on cell physiology or developmental processes (e.g., *RZE1* in *Cryptococcus neoformans*), while other lncRNAs participate in regulatory circuits common to many different cell types and organismal processes (e.g., *HOTAIR* and *ANRIL* in humans).

Table 14. Diverse Regulatory Functions Reported for Long Non-coding lncRNA Molecules.

Organism or Taxon	Function(s)
Budding yeast	Galactose metabolism, controls speed of transcriptional induction by galactose
Fission yeast	Phosphate-induced epigenetic silencing
Budding, fission yeast	Cellular responses to environmental changes
Fungal pathogen <i>Cryptococcus neoformans</i>	lncRNA <i>RZE1</i> regulates yeast-to-hypha transition
<i>Plasmodium falciparum</i>	1. Gametocyte development; 2. Telomere associated lncRNAs
Plants	Flower development and timing
<i>Arabidopsis</i>	1. Stress response lncRNAs; 2. lncRNA <i>TER</i> regulates telomerase activity
Tomato	Fruit ripening
<i>Drosophila</i>	1. Sex determination; 2. X chromosome dosage compensation
Tetrapods	Spermatogenesis, synaptic transmission, placenta development
Mammals	1. mRNA transcription; 2. Stem cell pluripotency; 3. Epigenetic chromatin formatting; 4. Placental membrane integrity; 5. Androgen receptor-regulated transcription; 6. Synaptic connectivity in brain
Marsupial (<i>Monodelphis domestica</i>)	Female X chromosome inactivation by repeat-rich lncRNA <i>Rsx</i>
Goat (<i>Capra hircus</i>)	Skin pigmentation

Table 14. Cont.

Organism or Taxon	Function(s)
Mouse	<ol style="list-style-type: none"> 1. Enhancer methylation and transcription control; 2. Diurnal metabolic regulation; 3. Cell reprogramming to pluripotent stem cells; 4. Imprinted <i>Igfr2</i> silencing; 5. Myogenesis (SINE-containing lncRNA acting on <i>STAU-1</i> mRNA decay); 6. lncRNA <i>Dum</i> regulates myogenic differentiation and muscle regeneration; 7. Translation of brain UCHL1 protein involved in preventing neurodegeneration, regulation by <i>SINEB2</i> recognition; 8. lncRNA <i>Evf2</i> modulates chromatin formation in forebrain development; 9. lncRNA-<i>HIT</i> mediates TGF beta-induced epithelial to mesenchymal transition in mammary epithelia and is essential for chondrogenic differentiation in the limb mesenchyme (cartilage formation); 10. lncRNA <i>Fendrr</i> regulates heart and body wall development; 11. Male germline development
Rat	Long-term potentiation of synaptic connectivity in adult brain development
Primate	<ol style="list-style-type: none"> 1. lncRNA <i>ANRIL</i> regulates expression of three cyclin-dependent kinase inhibitors and atherogenesis, <i>ANRIL Alu</i> exons acquired in primate lineage; 2. Primate-specific lncRNA <i>HPAT-5</i> required for stem cell pluripotency
Human	<ol style="list-style-type: none"> 1. 18,871,097 lncRNA-RNA base-pairings likely involved in processing, stability control and functions of 57,303 transcripts; 2. Centromere function; 3. Steroid receptor activation; 4. Endocrine regulation; 5. Rb and p53 signaling pathways; 6. lncRNA <i>RoR</i> is a p53 repressor in response to DNA damage and acts as microRNA sponge in transcription factor control; 7. lncRNA <i>LED</i> stimulates p53 activated transcription; 8. lncRNA <i>H19</i> modulates S-adenosylhomocysteine hydrolase and DNA methylation; 9. lncRNA <i>HOTAIR</i> regulates chromatin dynamics; 10. lncRNA <i>NBR2</i> regulates AMP-activated protein kinase under energy stress; 11. lncRNA <i>Xist</i> required for X inactivation; 12. lncRNA <i>XACT</i> in active X chromosome expression; 13. Apoptosis and lysosomal processes (<i>Alu</i> recognition by lncRNA <i>GAS5</i>); 14. <i>STAU-1</i> mRNA decay (stimulated by <i>Alu</i> pairing); 15. Cell cycle regulation by lncRNA <i>APTR</i>; 16. lncRNA <i>Firre</i> controls mRNA retention in the nucleus, nucleolar anchoring of inactive X; 17. lncRNA <i>ROR</i> regulates stem cell pluripotency; has sequences from >12 mobile elements, including a long <i>HERVH</i> 5' sequence characteristic of several pluripotent cell lncRNAs, acts as microRNA sponge; 18. Stem cell specificities; 19. Stem cell pluripotency and cancer cell proliferation, transcription from retroviral promoters; 20. Pluripotency and neuronal differentiation regulation of chromatin modifiers and transcription factors; 21. lncRNA-mediated regulation of the interferon response; 22. lncRNA <i>ANRIL</i> regulates inflammatory responses as a novel component of NF-kappaB pathway; 23. Innate and adaptive immune responses; 24. lncRNA <i>TINCR</i> regulates epidermal differentiation; 25. Neurodevelopment and brain function; 26. Chromatin modification, epigenetic regulation, alternative splicing, and translational control by <i>MALAT1</i>, <i>HOTAIR</i> and <i>TRE</i> lncRNAs represent important examples of lncRNA-mediated control of cell migration and invasion, epithelial-to-mesenchyme transition and metastasis

A more detailed and fully referenced version of this table is available as Supplementary Table S14.

The connections are strong between lncRNAs, rapid evolutionary change, and mobile repetitive DNA elements in tomatoes [554], as well as in mammals, where we have the most abundant information [555–557]. A 2012 analysis reported that 97% of human lncRNAs are primate-specific and found mobile DNA element sequences in 83% of 9241 human lncRNAs, where they constituted 42% of total lncRNA sequence [558]. Multiple reports have noted a strong correlation between lncRNA expression in stem cells and the presence in the corresponding genomic DNA of transcriptional regulatory signals from the HERVH family of human endogenous retroviruses [558,559]. HERVH sequences are present in the promoters of more than 100 lncRNAs, evidently providing transcriptional regulatory signals for pluripotent stem cell-specific expression [485,558,560]. A parallel stem cell-ERV association was found in mouse lncRNAs [561].

Because they are repetitive sequences, the mobile DNA-derived segments in lncRNAs are able to serve as recognition modules for regulatory interactions with other cellular RNAs or with genomic DNA regions that have complementary repeats. A few cases of this kind of recognition have been indicated in Table 14. *Alu* SINE-based lncRNA recognition to increase the synthesis of proteins from *Alu*-containing mRNAs has been called the “SINEUP” phenotype [562]. The role of mobile DNA sequences as functional domains of lncRNA molecules has similarly been assigned the acronym of RIDL, for Repeat Insertion Domains of lncRNA [563]. In human fibroblasts, *Alu* repeats in lncRNAs allow for them to serve as “sponges” to lower the content of *Alu*-derived micro-RNAs [564]. Some repetitive sequence interactions can be quite precise. It has been reported that a single point mutation in an *Alu* SINE retrotransposon sequence in a human lncRNA is pathogenic and leads to brainstem atrophy and death [564].

In addition to the relatively few species where functional studies have taken place (e.g., Table 14), reports of lncRNA repertoires have been published for diverse species, such as the fungus *Neurospora crassa* [565], the parasite *Trichomonas vaginalis* [566], the mosquito malaria vector *Anopheles gambiae* [567], the silkworm *Bombyx mori* [568], various plants [569,570], the early metazoan demosponge *Amphimedon queenslandica* [571], and Rainbow trout [572]. The abundance of lncRNA molecules in these various organisms has been interpreted as being indicative of the regulatory complexities involved in their reproduction and development. There are several articles reporting attempts to construct taxonomies of the different types of lncRNAs [573–576], but many years of transcriptomic and experimental studies almost certainly lay ahead before we have anything like a comprehensive classification and functional understanding of these versatile molecules.

8. Ecological Disruption and Read-Write Genome Modifications

As Eldredge and Gould noted over four decades ago, the paleontological record is characterized by episodes of relative evolutionary stasis that is punctuated by episodes of intense evolutionary variability and innovation [577] (Table 1). This pattern of “punctuated equilibrium” indicates that there is a connection between ecological stability or instability and evolutionary change. Now that we know about so many biological processes of genome rewriting, it is important to consider how ecological disruption relates to their activity. Although it has long been taboo in some evolutionary circles to assert that environmental influences can impact genome variability, we will see that there exists the beginning of a significant literature documenting precisely those impacts (<http://shapiro.bsd.uchicago.edu/ExtraRefs.CanGenomeChangeBeLinkedEcologicalDisruption.shtml>).

8.1. Ecological Change, Mating Population Decline and Interspecific Hybridization

Section 4 detailed the impact that interspecific hybridization has on speciation, WGDs, and the activation of mobile DNA elements. For hybrid speciation to occur, individuals of each parental species have to mate with individuals of the other species rather than their normal conspecific mating partners. It is easy to see that an adverse ecology and a consequent decline in the size of the mating population for one or both species will make cross-species mate choices more likely. This is precisely the pattern that has been observed in small island populations, where hybrid speciation

is unexpectedly common [468,578,579]. Continued observation of Darwin's finches *Geospiza* in the Galapagos Islands [580] has made it possible to link episodes of abrupt climate change to increased evolutionary variability [21,23].

Much of the variability in hybrid speciation comes from the well-documented activation of mobile DNA elements [468,475] (Table 5). Such hybrid activation can amplify the abundance of mobile DNA elements in the genome, as observed in *Drosophila* [581], and these newly dispersed elements in turn can lead to the formation of novel regulatory networks with the potential to generate adaptive phenotypic novelties (Section 7). The predicted accumulation of mobile DNA elements with adaptive radiation has occurred in Hox clusters of *Anolis* lizard genomes [582]. Similar roles for hybridization and mobile DNA activation have been invoked to provide "a more complete and satisfactory explanation for Darwin's 'abominable mystery': the spectacular success of the angiosperms" [583,584]. There is a report of exactly this hybridization-driven series of mobile DNA changes in the recently formed invasive cordgrass species *Spartina anglica* [585], and they have been invoked to account for the more general "genetic paradox" of invasive species [586]. A parallel role for interspecific hybridization, mobile DNA, and network rewiring (cf., Section 7.4) has also been suggested in primate evolution [587,588].

8.2. Regulated Biochemistry at the Basis of Point Mutations, Deletions, Translocations and Mutational "Storms"

It is useful to keep in mind that the first demonstrations of induced mutagenesis in bacteria documented quantitatively significant effects of environmental conditions, such as UV radiation, chemicals, and temperature [589]. Today we understand in exquisite detail the role of the SOS DNA damage-sensing control circuit and of SOS-regulated mutagenic trans-lesion DNA polymerases in producing the mutational responses to such environmental factors [590–592]. Y-family mutagenic DNA polymerases are present in both eukaryotes, as well as prokaryotes, where they are essential to both "spontaneous" and induced mutations [593] (http://shapiro.bsd.uchicago.edu/Translesion_mutator_polymerases.html). These mutagenic Y-family polymerases have also been implicated in chromosome breakages and complex genome rearrangements [594]. In other words, localized genetic change in response to DNA-damaging agents is fully as much an active biological process as the transposition or retrotransposition of mobile DNA elements.

In recent years, we have come to identify a variety of biochemical and cellular processes that mediate localized point mutations, deletions, and translocations, plus two newly recognized forms of clustered mutational changes frequently detected in cancer cell genomes: "kataegis" and "chromothripsis" (Table 15). Kataegis comes from the Greek word for thunder and designates a "shower" of point mutations that are spread over a contiguous region of the genome that can range from a few dozen bp to many kb in length [595]. Molecular characterization of the mutations in these kataegic showers indicates that they result from the action of the AID/APOBEC family of cytosine deaminases on exposed single-stranded domains [596]: mutations are predominantly (>70%) C-to-T transitions, as expected for C-to-U deamination products [597,598], and are largely fixed in the same DNA strand, a sign of catalytic processivity [599]. Chromothripsis means "chromosome shattering", and indicates cases where multiple chromosome fragments, usually from a single chromosome, are ligated together in a new arrangement with multiple transpositions, inversions, deletions, and duplications [600–602]. A variety of processes are thought to lead to the underlying multiple chromosome breakage and DS break repair events that are producing each chromothripsis occurrence (see Supplementary Table S15 for references), but we know that chromothripsis usually occurs with a single chromosome because breakage and repair processes can be seen to occur on individual chromosomes that are isolated in micronuclei compartments [603].

While being observed initially in studies of cancer cell genomes (Section 9.2), both kataegis and chromothripsis also occur in the tissue of healthy individuals [604,605]. Somatic hypermutation for antibody maturation takes place in activated B cells as a tightly regulated and targeted form of kataegis that are catalyzed by the AID cytosine deaminase [606]. A form of kataegis occurs in the yeast genome

at highly transcribed loci [607,608] and also at tRNA loci [609]. Chromothripsis affects multiple protein-coding loci in healthy individuals [604], and has been found to arise more frequently than previously thought in both gametogenesis and early human embryogenesis [601,610–612]. Obviously, germline episodes of either kataegis, chromothripsis, or both can contribute to major genome changes in evolutionary diversification and innovation [613,614]. The most important point from a conceptual perspective is that we now have experimentally established and mechanistically realistic processes for triggering multiple mutations “all at once” in a single cell division cycle. Intriguingly, chromothripsis can also result from L1 LINE-mediated retrotransposition and *Alu-Alu* non-allelic recombination [615]. Thus, ecological or cellular factors that activate L1 LINE elements are also able to trigger chromothripsis. These factors include gamma irradiation [616], benzyrene [617], oxidative stress [618], and heavy metals [619–621].

Table 15. Diverse Mutagenic Natural Genetic Engineering Outcomes.

Mutation Type	Biochemical Activity
Single nucleotide substitutions	Y-family mutagenic trans-lesion DNA polymerase; error-prone repair systems
Frameshifts	Y-family mutagenic trans-lesion DNA polymerase; error-prone repair systems
Deletions (bacteria)	Y-family mutagenic trans-lesion DNA polymerase; error-prone repair systems
Deletions and translocations (often with microhomologies)	Mre11, CtIP exonucleases; canonical or alternative non-homologous end-joining (NHEJ) complexes
Deletions and translocations	Non-allelic homologous recombination (NAHR) between mobile DNA repeats
Deletion	Elevated transcription, Topoisomerase I; <i>LINE-1</i> -mediated; non-allelic homologous recombination between dispersed mobile repeats
Translocations	Nonhomologous end joining or microhomology-mediated break-induced replication
Somatic hypermutation and kataegis: multiple clustered nucleotide substitutions	AID or APOBEC cytosine deaminase
Chromothripsis and complex chromosome segment insertions	Loss of p53-dependent checkpoints; Rad51 homologous recombination; NHEJ; premature chromosome condensation; segregation of chromosome breakage and repair into a micronucleus; L1-Mediated Retrotransposition and <i>Alu-Alu</i> NAHR (Not all these processes are involved in each chromothripsis event.)

A more detailed and fully referenced version of this table is available as Supplementary Table S15.

8.3. Diverse Ecological Impacts on Natural Genetic Engineering Functions

We have known for many decades that ecological disruption and stressors have direct effects on natural genetic engineering functions, as well as indirect effects by means of interspecific hybridization. The SOS response to DNA damage serves as a paradigm of a cellular sensing and regulatory system controlling mutagenic activities [591,622]. All organisms have DNA damage response (DDR) systems to correct errors that occur during genome replication or from a wide range of cellular insults that lead to DNA breakage, oxidation, or other chemical modifications [8,623]. An enormous literature documents how these systems are regulated and can operate in both error-free and error-prone modes (e.g., <http://shapiro.bsd.uchicago.edu/ExtraRefs.DNADamageRepairAndMutagenesis.shtml>), either to restore the original genome sequence or to generate novel genome sequences—and even novel genome configurations, as we have seen in the case of chromothripsis (Table 15).

Having well-studied models makes it easy to understand that many mutagenic DNA repair processes and other NGE systems are stress-induced or are otherwise sensitive to ecological changes (Table 16). The ecological factors that influence genome change fall into three broad categories:

1. organismal growth conditions (starvation, etc.), growth phase and cellular differentiation;
2. interactions with biomolecules, including antibiotics, hormones, nutrients, signals, extracellular products of pathogens (toxins, etc.), as well as biotic stresses, such as bacterial, fungal, and virus infection; and,

3. abiotic stresses, including heat, cold, drought, oxidizing agents, heavy metals, wounding, and even space travel.

Table 16. Ecological Factors that Induce Mutagenic DNA Repair or Modulate Natural Genetic Engineering (NGE) Responses.

Ecological Factors and NGE Effects	Affected Organisms
Growth conditions and cellular differentiation	
Stationary phase mutagenesis; anaerobic growth enhanced point mutations; aging colonies produce mutational hotspots and retromutations (8-oxo-guanosine, formed exclusively on the transcribed strand); adaptive selection-induced retromutations; nutrient-dependent mutability (Phosphorus/carbon limitation increase point mutations, iron/oxygen/carbon limitation increase <i>IS150</i> insertions, and phosphorus limitation increases indels)	<i>B. subtilis</i> and <i>E. coli</i>
Cystic Fibrosis lung growth hypermutability	<i>P. aeruginosa</i>
Adenine starvation stimulates <i>Ty1</i> retrotransposition; transcription induces APOBEC kataegis; Glucose- or phosphate-limited growth produced frequent genomic amplifications, rearrangements and novel retrotransposition; starvation leads to genome restructuring but has <2× effect on point mutation; nitrogen starvation increases copy number variations (CNVs)	Yeast <i>Saccharomyces cerevisiae</i>
Domestication leads to increase in repetitive DNA and retrotransposons	Maize
Early embryogenesis activates <i>mPing</i> DNA transposition	Rice
Plant regeneration activates chromovirus <i>LORE1</i> (ERV) retrotransposition	Model legume <i>Lotus japonicus</i>
Neural differentiation activates <i>L1</i> retrotransposition.	Rodents, humans
Aging induces retrotransposition (effect counter-acted by calorie restriction)	Mouse germline and somatic tissue
Early embryonic development displays a mutator state for copy number variation (CNV) of genomic duplications	Humans
Abiotic stresses	
UV irradiation stimulates hypermutation; oxidative stress induces DNA transposon non-allelic homologous recombination (NAHR)	<i>Pseudomonas aeruginosa</i> and <i>Burkholderia cenocepacia</i>
Copper induces expansion and contraction of <i>CUP1</i> arrays encoding copper-binding protein (copy number variation, CNV)	Budding yeast <i>Saccharomyces cerevisiae</i>
Heat shock, oxidative and copper sulphate stresses activate LTR-retrotransposons Pyret and MAGGY, DNA transposons Pot3, MINE, Mg-SINE, Grasshopper and MGLR3	Fungal pathogen <i>Magnaporthe oryzae</i>
Mild heat stress and UV activate <i>mariner-Mos1</i> transposition	<i>Drosophila simulans</i>
Uranium induces alternative NHEJ DSB repair processes	Embryonic zebrafish cells
“Two mechanisms . . . of cadmium mutagenicity: (i) induction of reactive oxygen species (ROS); and (ii) inhibition of DNA repair.”	Various mammals
Arsenic, vanadium, iron induce <i>VL30</i> retrotransposition	Mouse NIH3T3 cells
“Environmental stressors such as ionizing radiation (terrestrial, space, and UV-radiation), air pollution (including particulate matter {PM}-derived and gaseous), persistent organic pollutants, and metals” activate mobile DNA elements; mercury induces <i>LINE1</i> retrotransposition; low doses of NiCl ₂ and CdCl ₂ contributed to an increase in mutagenic deletions by <i>Alu-Alu</i> NAHR . . . cells exposed to arsenic trioxide preferentially repaired using the “error prone” non-homologous end joining (alt-NHEJ) while inhibiting repair by HR; Aluminum and low-level As ₂ O ₃ induce <i>LINE1</i> retrotransposition while copper treatment downregulated <i>L1</i> retrotransposition; exposure to CdCl ₂ and CdAc ₂ inhibits NHEJ and activates MRE11-dependent repair; Cold, heat, hypoxic, and oxidative stresses induce trinucleotide repeat mutagenesis	Human cells and tissues
Heat stress activates <i>ONSEN</i> , <i>COPIA</i> retrotransposition	<i>Brassicaceae</i> and <i>Arabidopsis</i>
Microsatellite mutation rate is significantly greater at 26 °C than at 18 °C	<i>C. elegans</i>

Table 16. Cont.

Ecological Factors and NGE Effects	Affected Organisms
Hyper salinity, stressed lineages accumulate ~100% more mutations, and these mutations exhibit a distinctive molecular mutational spectrum (specific increases in relative frequency of transversion and insertion/deletion (indel) mutations).	<i>A. thaliana</i>
Nitric oxide modulator, sodium nitroprusside induces <i>Tos17</i> LTR retrotransposition; laser irradiation stimulates DNA methylation changes and <i>mPing</i> DNA transposition	Rice
Fungicides boscalid (respiration inhibitor), iprodione (unclear mode of action), thiophanate methyl (inhibition of microtubulin synthesis) and azoxystrobin and pyraclostrobin (quinone outside inhibitors) raised mutation rates 1.7- to 60-fold compared to neutral conditions.	Plant pathogen <i>Sclerotinia sclerotiorum</i>
Biotic Stresses and Biomolecules	
Ethanol stress induces transient hypermutator state; food additive sepiolite stimulates antibiotic resistance plasmid transfer	<i>E. coli</i> and other bacteria
Joint action of LL-37 (antimicrobial peptide) and free iron induces mutagenesis	<i>P. aeruginosa</i>
Antibiotics induce SOS response and conjugal DNA transfer	<i>V. cholera</i> , <i>Pseudomonas aeruginosa</i>
Fluoroquinolone and norfloxacin antibiotics induced point mutations, <i>IS1</i> NAHR deletions, <i>IS5</i> NAHR duplications (but not transpositions)	<i>E. coli</i>
Beta-lactam antibiotics induced RpoS-dependent mutagenesis	<i>E. coli</i>
Subinhibitory concentrations of ciprofloxacin and vancomycin activate <i>IS256</i> transposition, induce SOS response	<i>Staphylococcus aureus</i>
Tigecycline induces hypermutation	<i>Acinetobacter baumannii</i>
Cationic antimicrobial peptide human cathelicidin LL-37 induces mutagenesis in CF lungs	<i>P. aeruginosa</i>
Canavanine proteotoxic stress induces mutagenesis	Yeast <i>S. cerevisiae</i>
Infection by the following bacteria can produces DNA damage, genome instabilities and alterations in DNA repair activities: <i>Chlamydia trachomatis</i> , <i>N. gonorrhoea</i> , <i>Helicobacter pylori</i> , <i>E. coli</i> , <i>Campylobacter jejuni</i> , <i>Haemophilus ducreyi</i> , <i>Actinobacillus actinomycetemcomitans</i> , <i>Shigella dysenteriae</i> , <i>Helicobacter cinaedi</i> , <i>Helicobacter hepaticus</i> , <i>Salmonella</i> species, <i>Shigella</i> strains, <i>Klebsiella pneumoniae</i> , <i>Enterobacter aerogenes</i> , <i>Citrobacter koseri</i> , <i>P. aeruginosa</i> , <i>Listeria monocytogenes</i>	Human cells
Infection by the following viruses can produce DNA damage, genome instabilities and alterations in DNA repair activities: human cytomegalovirus, Human T-cell lymphotropic virus 1 (HTLV-1) retrovirus, and Zika virus.	Human cells
<i>P. syringae</i> pathovar <i>tomato</i> infection induces DSB formation . . . abundance of infection-induced DSBs reduced by salicylic acid	<i>Arabidopsis</i>
Attack by the oomycete pathogen <i>Peronospora parasitica</i> stimulates somatic recombination	<i>Arabidopsis</i>
Tobacco mosaic virus (TMV) or oilseed rape mosaic virus (ORMV) tobacco leaf infection resulted in a systemic increase in homologous recombination (HR) . . . a similar phenomenon occurs in <i>Arabidopsis thaliana</i> plants infected with ORMV.	<i>Arabidopsis</i> , tobacco
<i>Bs1</i> Transposition detected in maize lines following barley stripe mosaic virus infection	<i>Zea mays</i>
Physiological stress, induced by climate change or invasion of new habitats, disrupts epigenetic regulation and activates mobile DNA elements	Diverse organisms

A more detailed and fully referenced version of this table is available as Supplementary Table S16.

Looking at the results that are summarized in Table 16 and those reported online in 2011 (<http://shapiro.bsd.uchicago.edu/StimuliDocumentedActivateNGE.html>), it is significant to note the many cases where genome change is triggered by biosphere interactions. These include the inter-species effects of biological molecules (e.g., antibiotics on bacteria or bacterial toxins on plants and animals) and direct interactions between different organisms (e.g., the many different bacterial infections that induce DNA damage and mutagenic responses in human cells, viral infections that stimulate mutagenic genome repair or mobile DNA activity in plants and animals). There are two major take-home lessons from these initial data about the ecological effects on active genome rewriting: (1) They reinforce the idea that it does not make sense to think of evolution as a process that only involves the isolated

genome of one species at a time, and (2) they make it clear that ecological disruptions both stimulate and influence the nature of evolutionary genome change in complex ways [624–631].

An instructive example of the danger of ignoring ecological challenge as a stimulant for genome change in evolution studies is the case of a citrate-utilizing strain emerging unexpectedly in *E. coli* cultures after years of growth under standard laboratory conditions [632]. The appearance of this strain, which activated the expression of a citrate transporter (CitT) by a non-homologous transcriptional fusion, was originally claimed to be so unusual as to be considered equivalent to a paleontological speciation event [633,634]. However, when *E. coli* was subjected to direct selection for citrate utilization, and experienced the NGE triggering effect of aerobic starvation [635], the same class of CitT activating fusion was obtained repeatedly in each of a series of replicate cultures within a matter of days [632,636,637].

9. Further Reflections on Genome Rewriting by NGE as a Core Biological Capability

Certainly, one of the most basic facts of life for all organisms is that their reproduction and evolution take place in a highly dynamic environment. Ecology and biosphere interactions are subject to constant change. The unceasing flux in the conditions of life means that survival requires constant adaptation and change on the part of each organism. In the short term, adaptive variability operates through the action of physiological and behavioral control networks, as well as opportunistic symbiotic associations. Over the longer term, however, adaptation requires more fundamental changes in organismal structure and genome composition. That is what we mean by the term “evolution”. As a consequence of this evolutionary fact of life, the active biological processes outlined in the preceding sections of this review—symbiogenetic cell fusions, intercellular DNA transfers, and natural genetic engineering functions—must be considered as core biological capabilities. In this final section of the review, it is appropriate to consider lessons to be learned from cases where active genome rewriting occurs outside the evolutionary context and to pose one fundamental unanswered question that we need to address in future research.

9.1. Natural Genetic Engineering as Part of the Normal Life Cycle

A further appreciation of cell capacities for genome rewriting is apparent in the range of organisms that have evolved dedicated DNA modification capabilities to meet special needs of their normal life cycles. These dedicated NGE activities typically display a high degree of specificity within the genome, and thus exemplify mechanisms that living cells possess for targeting DNA rewriting to sequences with particular functional significance. We may consider these life cycle systems as examples of cellular virtuosity in genome modification:

9.1.1. Diversity-Generating Retroelements (DGRs)

DGRs are targeted DNA mutagenesis operators that are found in bacterial and bacteriophage genomes that promote rapid and repeated diversification of a specific region in a protein coding sequence [638,639]. DGR diversification utilizes a highly mutagenic reverse transcriptase to generate novel DNA segments for the incorporation into a specific region of the target coding sequence (“mutagenic homing”). Mutagenic homing allows for bacteriophages to alter their attachment specificity for host cell receptors and bacteria to diversify their surface proteins.

9.1.2. Bacterial Phase Variation

Bacterial phase variation [640] signifies the ability of bacteria to alternate between different protein expression states [641–643]. The action of site-specific recombinases to invert promoter elements relative to the transcribed coding sequence reversibly activates or inhibits protein expression [644–646]. Reversible inactivation of expression also occurs by the insertion and excision of DNA transposons (insertion sequences or IS elements) at the target locus [647], as well as the expansion and contraction of simple sequence repeats (SSRs) that change reading frame [648,649].

9.1.3. Bacterial Antigenic Variation

Bacterial antigenic variation signifies the ability of bacteria to change the structure or identity of surface proteins [641–643]. The molecular mechanisms adapted for changing surface proteins include site-specific recombination to invert parts of the coding sequence (“shufflons”) [650] and segmental coding sequence replacement from silent cassettes by targeted homologous recombination [651–653].

9.1.4. CRISPR Systems for Adaptive Immunity

Both archaea and eubacteria possess the famous CRISPR systems for adaptive immunity against the entrance of bacteriophage or plasmid DNA. Upon infection, the CRISPR defense involves rapid target sequence acquisition [654,655] (adaptation) into a specialized genomic expression site (the CRISPR array) to enable RNA-directed cleavage of complementary invading DNA (immunity) [656–658]. Separate dedicated DNA cleavage-ligation complexes are involved in spacer integration into genomic CRISPR arrays [659], on the one hand, and RNA-targeted attack of invading target DNA, on the other [660,661]. The adaptation activities appear to have evolved from a family of mobile DNA elements called “casposons” [662,663].

9.1.5. Prokaryotic DNA-Targeted Adaptive Immune Defense

Some archaea and bacteria possess a distinct adaptive immune defense against incoming plasmid DNA that is targeted by acquired DNA segments that are bound to a prokaryotic Argonaute (Ago) protein [664–666]. Little is known yet about how the targeting DNA is acquired. It should be noted that the prokaryotic Ago is the ancestor of eukaryotic Argonaute family proteins that participate in piRNA-targeted defense against invading mobile DNA elements, as well as siRNA regulation of genome expression [667].

9.1.6. Prokaryotic Systems for Aggregating Coding Sequence Cassettes

As previously described in Section 5.2, both archaea and bacteria possess multiple complex genomic elements that utilize site-specific recombinases or transposases for aggregating coding sequence cassettes that are related to a particular adaptive phenotype, such as antibiotic resistance, pathogenicity, or symbiosis in plasmids, transposons, integrons, genomic islands, and integrative and conjugative elements.

9.1.7. Yeast Mating-Type Switching

In certain yeasts, haploid cells execute tightly regulated switches from one cell mating-type (MAT) to the complementary cell type by DNA rearrangements [668]. (i) In the more primitive mating-type switches, a DNA segment carrying the two distinct MAT loci is flanked by inverted repeat (IR) sequences and borders an epigenetically silenced region of the genome. As a consequence of this arrangement, recombination between the IRs inverts the segment and changes which MAT locus is silenced and which expressed [669]. (ii) In the more highly evolved process, the genome contains one expressed MAT locus determining cell type, and two silent cassettes that contain protein sequences that determine each of the different mating types. A targeted process of directional recombination switches mating type by replacing DNA at the functional MAT locus with sequences from the silent cassette of the opposite type. The molecular details of the cassette replacement mechanism differ significantly in budding yeast and fission yeast, and thus appear to have evolved independently [668].

9.1.8. Trypanosome Antigenic Variation

Trypanosomes and certain other eukaryotic pathogens undergo antigenic variation by introducing sequences from silent DNA cassettes into genome expression sites [670–672]. Variation typically only affects a segment of the expressed surface protein (“segmental gene conversion”) [673], and variability is enhanced by the presence in the genome of arrays of multiple silent cassette sequences (~1000

or more in some trypanosomes). The process of segmental gene conversion involves homologous recombination at embedded repeats and starts with targeted DNA breaks.

9.1.9. Ciliate Macronucleus Genome Restructuring

The most prolific natural genetic engineers in the normal life cycle are the ciliates, also known as ciliated protozoa [674–677]. After each mating cycle, these remarkable unicellular organisms completely restructure their typically eukaryotic germ-line genomes into a distinct somatic genome organized as numerous multi-copy mini-chromosomes that are contained within a large transcriptionally active “macronucleus”. The generation of the macronucleus genome involves programmatic RNA-directed cleavage of the germ-line DNA and the elimination of germ-line specific DNA, followed by a distinct process of RNA-directed reassembly of the remaining DNA segments into functional mini-chromosomes that are capped with telomeres at each end. The exact details of germ-line DNA loss and mini-chromosome construction differ between ciliate taxa, but in the well-studied case of *Oxytricha*, at least 90% of germ-line DNA is eliminated, and over 200,000 remaining micronucleus fragments are assembled into ~16,000 intact single coding sequence mini-chromosomes. The fragments are often joined together in a different order from that which is present in germ-line DNA, so the macronucleus arrangement is designated as “scrambled”. During assembly of the ~16,000 scrambled *Oxytricha* mini-chromosomes [678], there are both error correction processes [679] as well as the occasional generation of novel coding sequence combinations [680]. The ability of *Oxytricha* and other ciliates to rapidly generate so many DNA cleavage and rejoining events and properly order the resulting constructs establishes a high standard for cellular control of complex NGE processes.

9.1.10. Mammalian Adaptive Immune System Rearrangements

An outstanding series of highly targeted and integrated NGE processes in response to biological signals takes place in cells of the mammalian adaptive immune system. Mammalian B lymphocytes produce, refine, and diversify a virtually limitless repertoire of highly specific defense molecules to detect and counter invading pathogens by constructing and modifying antibody heavy (H) and light (L) chain coding sequences [681]. The first step is to generate a limitless array of diverse antibody binding specificities. B cells use a domesticated transposase protein in a spatially organized process to cleave stochastically chosen cassettes from genomic arrays, then to add untemplated nucleotides to some of them, and finally to use classical NHEJ repair activity to join the appropriately cleaved cassettes into novel coding sequences, determining the variable (V) region of one H and one L chain per lymphocyte (“V(D)J joining”) [507,682,683]. Paired V_H and V_L domains form an extremely diverse population of combinatorial antigen binding sites on the B cell surfaces. When the appropriate combination binds an invader antigen, the encoding B cell is “activated” to multiply and undergo a further sequence of specific DNA changes involving the AID activation-induced cytosine deaminase. The DNA sequence encoding the V region (but not the constant or C region) of each chain is subjected to a kataegis-like somatic hypermutation “storm”, which is precisely targeted by transcriptional signals [684,685]. As some of the hypermutated antibodies bind the target antigen with a higher affinity, the activated B cells are further stimulated to multiply and undergo a “class switch recombination” (CSR) process that replaces the exon encoding the C region of the H chain and thereby targets the higher affinity antibody to the appropriate region of the body for each particular infection without altering its binding specificity. C region exon choice is determined by transcription signals determining the sites of AID-dependent DSB and NHEJ nonhomologous recombination events joining V region exons to new C region exons [686,687]. The highly orchestrated NGE processes that are involved in antibody formation and maturation are especially noteworthy because they illustrate how living cells are able to combine great precision in DNA modifications with extraordinary diversification of outcomes for a well-defined adaptive purpose.

9.2. Lessons on the Real Time Potential of Natural Genetic Engineering from Cancer Genomes

Cancer provides a model evolutionary system, where we can observe the possible extent of genome changes that are occurring in real time (the months or years it takes for a tumor population to develop). Genome changes are usually linked to tumor progression (the appearance of more aggressive and rapidly proliferating cancer cells) and consequently are of great interest to clinicians, who try to combat the disease. For our purposes, however, we are interested in cancer as a demonstration of how widely and rapidly complex genome changes have been documented to occur. It is broadly accepted that cancer cells are destabilized in maintenance of genome stability [688], and the exceptional tumor state can serve as a paradigm for genome destabilization that is triggered by ecological disruption in periods of evolutionary change [689]. It is no coincidence that many of the ecological triggers for NGE activity in Table 16 are also carcinogenic (e.g., genome-destabilizing microbial infections) [690–692].

Cancer cells reproducibly display many forms of natural genetic engineering discussed earlier in this review (Table 17). This is solid evidence for the repeated occurrence of rapid genome change by NGE in real time under the regulatory context of tumor development. The actions of RAG transposase and AID cytosine deaminase in immune system tumors demonstrate the adaptive utilization of available NGE resources in tumor evolution [693–695].

Table 17. Different Genome Changes Observed in Cancer Cells.

Stress-Induced Mutagenic Activity
Hypermutable following loss of replication proofreading functions
Massive genome rearrangements (“karyotype chaos”)
Homology-independent rearrangements (NHEJ)
Retrotransposon activation
Non-canonical termination of homologous recombination
Kataegis and somatic hypermutation
Cytosine deaminase-dependent chromosome translocation
Chromothripsis
Chromothripsis linked to oncogene amplification
Complex insertion-deletion mutations (indels)
Tandem duplications as well as formation of “amplicons” with rearranged and amplified chromosomal segments, <i>a.k.a.</i> copy number variations (CNVs)
Formation of amplified circular extrachromosomal DNAs
Processed pseudogene formation
L1 retrotransposition
Extensive L1 retrotransduction of non-repetitive DNA
Transfer of mitochondrial DNA into nuclear genome
RAG transposase/recombinase-mediated chromosome rearrangement in immune system tumors
Somatic hypermutation involving a reverse transcriptase-based mutator activity

A fully referenced version of this table is available as Supplementary Table S17. (See also <http://shapiro.bsd.uchicago.edu/Cancer%20Genome%20Changes.pdf>).

9.3. What Factors May Bias Genome Rewriting to Generate Selectively Positive Outcomes?

The preceding discussion lays out only a small selection of the massive experimental and genomic data that show that genome change in evolution is invariably a product of cellular and biochemical action. From single nucleotide changes involving error-prone repair functions (Table 15), through to the formation of distributed regulatory networks by mobile DNA formatting (Section 7.4), to whole genome restructuring that is triggered by interspecific hybridization events, non-random biological action is fundamental to genome innovation. Since cellular and biochemical activities are regulated and sensitive to multiple biotic and abiotic ecological inputs (Table 16), the recognition of biological agency in genome rewriting enables us to pose a question that has long been considered taboo in orthodox discussions of evolution: What factors may bias genome rewriting to generate selectively positive outcomes? We know unambiguously that such bias is at work in the somatic genome targeting processes described in Section 9.1. Why not in the course of evolution as well?

A useful way to address the global question of bias towards selective utility in NGE evolution processes is to separate out a series of easier-to-answer subordinate questions. If we can answer those affirmatively, then there is a solid basis for pursuing a scientific approach to adaptive guidance in evolutionary NGE genome variability.

9.3.1. Do Living Organisms Possess NGE Operators of Clear Evolutionary Utility?

The answer to this question is unambiguously yes. The NGE operators range from horizontal DNA transfer systems (Tables 6 and 7) to the DNA-based and RNA-based mutational operators that play key roles in protein evolution (Tables 9 and 10) to mobile DNA elements that rewire transcriptional regulatory networks (Table 13 and Section 7.4). Additional NGE processes that contribute to taxonomic and adaptive innovation in evolution include large-scale genome rearrangements, as documented in Tables 5, 15 and 17.

9.3.2. Can the Evolutionarily Useful NGE Operators Be Regulated and/or Targeted in Ways that Could Enhance Their Adaptive Utility?

Here again, the answer is clearly yes. We see both regulation and targeting at work in adaptively useful ways in the various somatic NGE systems described in the preceding Section 9.1. Diverse ecological triggers for NGE activity are presented in Section 8 and Table 16, and posted online (<http://shapiro.bsd.uchicago.edu/StimuliDocumentedActivateNGE.html>). There is also extensive documentation of targeting mechanisms at work directing mobile DNA element insertions to specific genome sites or regions in organisms that range from bacteria to plants and animals [696–699], (<http://shapiro.bsd.uchicago.edu/ExtraRefs.TargetingNaturalGeneticEngineeringInGenome.shtml>; http://shapiro.bsd.uchicago.edu/Targeting_retroviral_and_retrotransposon_insertions.html; http://shapiro.bsd.uchicago.edu/P_factor_homing.html). Targeting and activation are particularly relevant to the larger question of positive bias in evolution when they operate via a global genome control process like epigenetic genome modification as, for example, in mouse embryonic stem cells [696].

9.3.3. Are There Generic Features of Cellular Genome Function That Can Favor Evolutionarily Adaptive NGE Outcomes?

Once again, the answer is positive, but in this case, we need to make a distinction between genome features that are already described as having contributed to evolutionary success and other features whose potential contributions to evolutionary innovation remain to be explored further.

The aspects of NGE activity on cellular genomes that facilitate the positive adaptations detailed above include active nucleotide sequence modification (Table 15), the capacity for both homologous and non-homologous DNA exchanges, interactions with viral genomes, the presence of mobile DNA elements, and the formation of genome duplications [325,700–707]. In prokaryotes, the most numerous organisms in the biosphere, mobile DNA, includes transmissible plasmids and other conjugative elements. But, intercellular DNA transfers are by no means restricted to prokaryotes because all organisms can exchange nucleic acids that are encapsidated in viral particles, and in some cases, take up DNA directly from the environment, in lipid vesicles (exosomes), or by direct contact with other cells, such as endosymbionts (Table 6). In eukaryotes (but not exclusively), mobile DNA includes reverse-transcribed cDNA retrocopies of processed cell RNAs, which are important contributors to protein evolution (Table 9). The non-random nature of certain NGE operations, in particular the movement of defined DNA transposon and retrotransposon elements, favor their adaptive utility in mobilizing transcription and other signals to new genome locations. As stated in the title of this section, producing genome change to generate adaptive novelties is clearly “a core biological capacity”.

In addition to those features of genome structure and function that enable adaptive DNA rewriting, there are higher-order systemic features that may work informatically to bias genome change towards adaptively useful outcomes. A number of authors from the Evo-Devo school

have pointed to modularity and repetition in control networks as features that favor successful innovation in developmental and morphogenetic processes [708–710]. Some have even cited Richard Goldschmidt, considered by some as the “father” of Evo-Devo thinking (Table 1), and pointed out that organisms bearing morphogenetic innovations may be thought of as Goldschmidt’s “hopeful monsters” [170,172,173].

A further generic feature of genome structure and action that could help bias NGE towards useful outcomes is that genome expression adapted to environmental conditions is critically dependent upon the physical and chemical status of cellular DNA. Chemical modifications like nucleotide methylation, DNA binding to specific packaging, replication, transcription and regulatory proteins, organization of the genome into functionally distinct epigenetic domains, and three-dimensional alignment of sometimes distant DNA sites by interactions with protein and lncRNA networks are all information-rich features of genome structure, which have the potential to target and coordinate NGE activities [711–714].

9.3.4. Are There Feasible Experimental Approaches to Demonstrating and Dissecting Selectively Advantageous Bias in Complex Evolutionary NGE Events?

This is a critical question for evolution research in the coming decades. There have been numerous studies of how organisms acquire one or a few individual traits, such as the *E. coli* citrate utilization case described above (Section 8.3, [636,637]), but (as far as I am aware) it still remains to be demonstrated that a truly complex and adaptive “hopeful monster” innovation can be generated *de novo* in real time so that successful cases can be subjected to genomic analysis to determine the underlying NGE operations. The approach is similar to that employed in real time adaptive evolution of microbial strains for biotechnology applications [715]. The “hopeful monster” class might include complex innovations like (i) origination of a new multi-step catabolic or biosynthetic pathway in a microorganism, (ii) morphogenesis of a useful appendage or organismal structure on a plant or animal that is not simply a modification of a pre-existing feature, or (iii) elaboration of a functional signaling system that is based on a novel signal molecule, receptor, and receptor-linked signal transduction cascade.

The major challenge to observing complex evolutionary events in real time is the elaboration of a suitable selection scheme. In the case of microbial pathways, this is achievable by utilizing a novel selective substrate (i.e., one for which no catabolic pathway exists) as the source of an essential growth requirement (C, N, etc.), or starting with an organism that is completely lacking the ability to synthesize structures required for a selectable property, such as motility in fluid medium or across a solid surface.

If the selection proves unexpectedly easy, then it is likely that silent precursor coding elements already existed in the genome, and sequence analysis will reveal their reproducible activation in independently selected clones. However, if the selection proves difficult and requires prolonged incubation (perhaps several weeks or months for bacterial systems, cf., [716]), then it may well be that a complex NGE process was at work. Genome sequence analysis will reveal how many distinct steps were required, as well as how reproducible those steps proved to be in independently selected isolates. Repeatedly finding certain DNA rearrangements in the adaptation process (e.g., the movement of mobile DNA elements or repositioning of protein-coding sequences) would be evidence for genome features that bias changes towards adaptive success. The confirmation of such a conclusion is possible by removing or altering the putative biasing feature and measuring the resulting impact on successful responses to selection. Fortunately, contemporary CRISPR-based technologies make all manner of genome alterations feasible on virtually any scale in all organisms [717]. In some cases, horizontal DNA transfer may be required for a complex selection to succeed. If so, then the starting culture would have to contain two or more species, and that requirement itself would be a significant result, which is consistent with genomic observations on adaptive sequence acquisitions across taxonomic boundaries (Section 5).

It is beyond the competence of a bacterial geneticist to speculate on selective regimes for truly complex adaptations in multicellular organisms. Nonetheless, two aspects of successful “hopeful monster” hunts in plants and animals can be predicted based on the biological genome rewriting activities that are discussed above. (1) The first is that success is most likely to depend upon the stimulating effect of interspecific hybridization for providing greater starting sequence variability and generating increased NGE activity prior to selection (Table 5). This effect of interspecific hybridization has been observed in *Saccharomyces* yeasts, where interspecific hybrids reproducibly evolve a chromosome translocation that confers adaptation to ammonia limitation [718]. (2) The second is that plants and animals all exist as holobionts, and it is probable that complex evolutionary success will involve microbial NGE triggering (Table 16) or horizontal DNA transfer (Section 5). Thus, it may turn out that success in plant or animal “hopeful monster” hunts may be dependent upon the presence of certain symbiotic or even pathogenic microbes in the experimental material. Clearly, the discovery that experimental plant or animal evolution is sensitive to microbial stimulation would be a highly significant result, demonstrating the holobiont principle [18,719]. Tests for microbial stimulation could include viruses, which have been found to be significant contributors to the evolution of species, like our own [720].

9.4. Conclusions

Hopefully, the preceding discussion will make it clear that there are well-defined empirical approaches to addressing key questions about what active biological processes are operational in real-time evolution of complex adaptive traits. The fact that evolution experiments may well involve interspecific hybridization, mixed cultures, and microbial infections of test organisms only serves to highlight how far contemporary thinking about hereditary change has advanced from the sterile abstractions of the past two centuries. (The word “sterile” is used here in both usual meanings: (i) devoid of contaminating organisms, and (ii) unable to lead to a productive conclusion.) The experimental and conceptual palette that we have today for probing and depicting the nature of biological variation is far richer than the information that was available in the 20th Century. Integrating powerful genome analysis and modification technologies with imaginative experimental design is certain to provide us with a surprising picture of the many ways living organisms change themselves in the course of evolution.

Supplementary Materials: The following are available online at www.mdpi.com/2079-7737/6/4/42/s1, Supplementary Table S1: Key Scientists Advocating non-Gradualist Evolution in the 19th and 20th Centuries, Table S2: No supplementary table needed (To keep table numbers to be the same in the text and Supplementary Material, the author added Table S2), Table S3: Photosynthetic eukaryotic lineages resulting from symbiogenesis, Table S4: Selected Examples of Speciation and Adaptive Radiation Involving Interspecific Hybridization and Changes in Chromosome Number, Table S5: Genomic consequences of experimental interspecific hybridization in plants and animals, Table S6: Modes of Horizontal DNA Transfer, Table S7: Selected examples of inter-phylum adaptive horizontal DNA transfers based on genomic data, Table S8: Highly cited protein domain search terms in the PubMed database, Supplementary Table S9: Selected Reports of Exon Rearrangements and Retroposition by Mobile DNA element Activity, Table S10: Some Reported Instances of Orphan Coding Sequences (“New genes” and New exons) Discovered in Sequenced Genomes, Table S11: Origination of Novel Exons from Mobile DNA Elements, Table S12, No supplementary table needed (To keep table numbers to be the same in the text and Supplementary Material, the author added Table S12), Table S13: Distributed genome network innovation attributed to mobile DNA elements, Table S14: Diverse Regulatory Functions Reported for Long Non-coding lncRNA molecules, Table S15: Diverse Mutagenic Natural Genetic Engineering Activities, Table S16: Ecological Factors that Induce Mutagenic DNA Repair or Modulate NGE Responses, Table S17: Different Genome Changes Observed in Cancer Cells.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

Ac	long interspersed nucleotide element
AID	activation-induced deaminase
AP	apurinic
BER	base excision repair
BFB	breakage-fusion-bridge cycle
CDS	coding sequence
CDT	cytotoxic distending toxin
CNV	copy number variation
CRISPR	clustered regularly interspaced short palindromic repeats
DDR	DNA damage response
DGR	diversity-generating retroelement
dPRL	decidual prolactin
Ds	Dissociator transposon
DSB	double-strand break
ENV	envelope protein (retrovirus)
ERV	endogenous retrovirus
ESC	embryonic stem cell
HERV/hERV	human ERV
HGT	horizontal gene transfer
HR	homologous recombination
Indel	insertion plus deletion
IPSC	induced pluripotent stem cell
IR	inverted repeat
lincRNA	long intergenic non-coding RNA
LINE	long interspersed nucleotide element
lncRNA	long non-coding RNA
LSG	lineage-specific gene
LTRs	long terminal repeats
MaLR	mammalian apparent LTR-retrotransposon
MER	mammalian endogenous repeat
mERV	mouse endogenous retrovirus
MIR	mammalian-wide interspersed repeat
miRNA	microRNA
MMR	mutagenic mismatch repair
mRNA	messenger RNA
MT	mouse transposon
NAHR	non-allelic homologous recombination
NCLDV	nucleocytoplasmic large DNA virus
NEE	novel enriched environment
NER	nucleotide excision repair
NGE	natural genetic engineering
NHEJ	non-homologous end-joining
ORF	open reading frame
ORR	origin region repeat
RAG	recombination-activating gene, transposase activity needed for V(D)J recombination in adaptive immunity
RIDL	repeat insertion domains of lncRNA
RLEs	retrovirus-like elements
ROS	reactive oxygen species
SINE	short interspersed nucleotide element
SOS	DNA damage signal-inducible repair response in bacteria

SSB	single-strand break
SSR	simple sequence repeat
SVA	a hominid retrotransposon containing SINE, VNTR and <i>Alu</i> components
TE	transposable element
WGD	whole genome duplication

References and Notes

- Bukhari, A.I.; Shapiro, J.A.; Adhya, S.L. *DNA Insertion Elements, Plasmids and Episomes*; Cold Spring Harbor Press: Cold Spring Harbor, NY, USA, 1977.
- Shapiro, J.A. *Mobile Genetic Elements*; Academic Press: New York, NY, USA, 1983.
- Shapiro, J.A. *Evolution: A View from the 21st Century*; FT Press Science: Upper Saddle River, NJ, USA, 2011; p. 272.
- Berg, D.E.; Howe, M.M. (Eds.) *Mobile DNA*; American Society for Microbiology: Washington, DC, USA, 1989; p. 972.
- Craig, N.; Craigie, R.; Gellert, M.; Lambowitz, A.M. *Mobile DNA II*; American Society for Microbiology Press: Washington, DC, USA, 2002.
- Craig, N.L.; Chandler, M.; Gellert, M.; Lambowitz, A.L.; Rice, P.A.; Sandmeyer, S.B. (Eds.) *Mobile DNA III*; American Society for Microbiology: Washington, DC, USA, 2015.
- Kazazian, H.H. *Mobile DNA: Finding Treasure in Junk*; FT Press Science: Upper Saddle River, NJ, USA, 2011.
- Haber, J.E. *Genome Stability: DNA Repair and Recombination*, 1st ed.; Garland Scientific: New York, NY, USA, 2013.
- Shapiro, J.A. How life changes itself: The Read-Write (RW) genome. *Phys. Life Rev.* **2013**, *10*, 287–323. [[CrossRef](#)] [[PubMed](#)]
- Shapiro, J.A. Exploring the read-write genome: Mobile DNA and mammalian adaptation. *Crit. Rev. Biochem. Mol. Biol.* **2017**, *52*, 1–17. [[CrossRef](#)] [[PubMed](#)]
- Shapiro, J.A. Nothing in Evolution Makes Sense Except in the Light of Genomics: Read-Write Genome Evolution as an Active Biological Process. *Biology* **2016**, *5*, 27. [[CrossRef](#)] [[PubMed](#)]
- Shapiro, J.A. Biological action in Read-Write genome evolution. *Interface Focus* **2017**, *7*, 20160115. [[CrossRef](#)] [[PubMed](#)]
- Crick, F. On protein synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–163. [[PubMed](#)]
- Crick, F. Central dogma of molecular biology. *Nature* **1970**, *227*, 561–563. [[CrossRef](#)] [[PubMed](#)]
- Brenner, S. Turing centenary: Life's code script. *Nature* **2012**, *482*, 461. [[CrossRef](#)] [[PubMed](#)]
- Darwin, C. *Origin of Species*; John Russel: London, UK, 1859.
- Darwin, C. *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 6th ed.; John Murray: London, UK, 1872.
- Richardson, L.A. Evolving as a holobiont. *PLoS Biol.* **2017**, *15*, e2002168. [[CrossRef](#)] [[PubMed](#)]
- Zrzavy, J.; Skala, Z. Holobionts, hybrids, and cladistic classification. *Biosystems* **1993**, *31*, 127–130, discussion 130–133. [[CrossRef](#)]
- Rosenberg, E.; Sharon, G.; Zilber-Rosenberg, I. The hologenome theory of evolution contains Lamarckian aspects within a Darwinian framework. *Environ. Microbiol.* **2009**, *11*, 2959–2962. [[CrossRef](#)] [[PubMed](#)]
- Grant, B.R.; Grant, P.R. Watching speciation in action. *Science* **2017**, *355*, 910–911. [[CrossRef](#)] [[PubMed](#)]
- Grant, P.R.; Grant, B.R. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **2002**, *296*, 707–711. [[CrossRef](#)] [[PubMed](#)]
- Grant, B.R.; Grant, P.R. Evolution of Darwin's finches caused by a rare climatic event. *Proc. R. Soc. B* **1993**, *251*, 111–117. [[CrossRef](#)]
- Margulis, L. Symbiogenesis. A new principle of evolution rediscovery of Boris Mikhaylovich Kozo-Polyansky (1890–1957). *Paleontol. J.* **2010**, *44*, 1525–1539. [[CrossRef](#)]
- Kozo-Polyansky, B.M. *Symbiogenesis: A New Principle of Evolution (1924)*; Harvard University Press: Cambridge, MA, USA, 2010.
- Archibald, J. *One Plus One Equals One: Symbiosis and the Evolution of Complex Life*; Oxford University Press: Oxford, UK, 2014.
- Harold, F.M. *In Search of Cell History: The Evolution of Life's Building Blocks*; University Of Chicago Press: Chicago, IL, USA, 2014.

28. Knoll, A.H.; Javaux, E.J.; Hewitt, D.; Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2006**, *361*, 1023–1038. [[CrossRef](#)] [[PubMed](#)]
29. Allwood, A.C.; Grotzinger, J.P.; Knoll, A.H.; Burch, I.W.; Anderson, M.S.; Coleman, M.L.; Kanik, I. Controls on development and diversity of Early Archean stromatolites. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9548–9555. [[CrossRef](#)] [[PubMed](#)]
30. Margulis, L. *Origin of Eukaryotic Cells*; Yale University Press: New Haven, CT, USA, 1970.
31. Margulis, L. The origin of plant and animal cells. *Am. Sci.* **1971**, *59*, 230–235. [[PubMed](#)]
32. One of the great mysteries of evolution is how eukaryotic organisms suddenly appeared in the paleontological record in the Proterozoic era between 1.6 to 2.1 billion years ago [28], about 1.3–1.8 billion years after the oldest prokaryotic fossils [29]. Eukaryotic cells contain many cell structures and systems that are not present in prokaryotes [27]. These include the nucleus, endoplasmic reticulum, lysosomes, vacuoles, centrioles, kinetoplasts, cytoskeleton, and eukaryotic flagella that are quite distinct from prokaryotic flagella. How such structures evolved without any evidence for intermediate cellular precursors is a major unsolved problem. As the text shows, there is compelling DNA evidence that two eukaryotic organelles, the mitochondrion and the plastid, arose through bacterial and cyanobacterial endosymbiogenesis. Lynn Margulis has built upon these cases and direct observations of close associations between eukaryotic and prokaryotic cells to put forward the serial endosymbiotic theory [30,31]. This theory hypothesizes that many (perhaps all) characteristic eukaryotic cell structures evolved from various endosymbiotic prokaryotic ancestors. Unfortunately, most eukaryotic cell organelles lack DNA and a protein synthesis apparatus; so direct evidence for connection to a prokaryotic ancestry is lacking. However, if there is merit to the serial endosymbiotic theory, then active cell mergers would have played an even greater role in early eukaryotic cell evolution than we are currently able to substantiate.
33. Dyall, S.D.; Brown, M.T.; Johnson, P.J. Ancient invasions: From endosymbionts to organelles. *Science* **2004**, *304*, 253–257. [[CrossRef](#)] [[PubMed](#)]
34. Bhattacharya, D.; Archibald, J.M.; Weber, A.P.; Reyes-Prieto, A. How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays* **2007**, *29*, 1239–1246. [[CrossRef](#)] [[PubMed](#)]
35. Williams, T.A.; Embley, T.M. Changing ideas about eukaryotic origins. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2015**, *370*, 20140318. [[CrossRef](#)] [[PubMed](#)]
36. Embley, T.M. Multiple secondary origins of the anaerobic lifestyle in eukaryotes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2006**, *361*, 1055–1067. [[CrossRef](#)] [[PubMed](#)]
37. Van der Giezen, M. Hydrogenosomes and mitosomes: Conservation and evolution of functions. *J. Eukaryot. Microbiol.* **2009**, *56*, 221–231. [[CrossRef](#)] [[PubMed](#)]
38. Makiuchi, T.; Nozaki, T. Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa. *Biochimie* **2014**, *100*, 3–17. [[CrossRef](#)] [[PubMed](#)]
39. Embley, T.M.; Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **2006**, *440*, 623–630. [[CrossRef](#)] [[PubMed](#)]
40. Esser, C.; Ahmadinejad, N.; Wiegand, C.; Rotte, C.; Sebastiani, F.; Gelius-Dietrich, G.; Henze, K.; Kretschmann, E.; Richly, E.; Leister, D.; et al. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **2004**, *21*, 1643–1660. [[CrossRef](#)] [[PubMed](#)]
41. Thrash, J.C.; Boyd, A.; Huggett, M.J.; Grote, J.; Carini, P.; Yoder, R.J.; Robbertse, B.; Spatafora, J.W.; Rappe, M.S.; Giovannoni, S.J. Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci. Rep.* **2011**, *1*, 13. [[CrossRef](#)] [[PubMed](#)]
42. Keeling, P.J. The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2010**, *365*, 729–748. [[CrossRef](#)] [[PubMed](#)]
43. Reyes-Prieto, A.; Yoon, H.S.; Moustafa, A.; Yang, E.C.; Andersen, R.A.; Boo, S.M.; Nakayama, T.; Ishida, K.; Bhattacharya, D. Differential gene retention in plastids of common recent origin. *Mol. Biol. Evol.* **2010**, *27*, 1530–1537. [[CrossRef](#)] [[PubMed](#)]
44. McFadden, G.I. Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016105. [[CrossRef](#)] [[PubMed](#)]
45. Gavelis, G.S.; Hayakawa, S.; White Iii, R.A.; Gojobori, T.; Suttle, C.A.; Keeling, P.J.; Leander, B.S. Eye-like ocelloids are built from different endosymbiotically acquired components. *Nature* **2015**, *523*, 204–207. [[CrossRef](#)] [[PubMed](#)]

46. Adams, K.L.; Palmer, J.D. Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* **2003**, *29*, 380–395. [[CrossRef](#)]
47. Gandini, C.L.; Sanchez-Puerta, M.V. Foreign Plastid Sequences in Plant Mitochondria are Frequently Acquired Via Mitochondrion-to-Mitochondrion Horizontal Transfer. *Sci. Rep.* **2017**, *7*, 43402. [[CrossRef](#)] [[PubMed](#)]
48. Shutt, T.E.; Gray, M.W. Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet.* **2006**, *22*, 90–95. [[CrossRef](#)] [[PubMed](#)]
49. Lang, B.F.; Burger, G.; O’Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sankoff, D.; Turmel, M.; Gray, M.W. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **1997**, *387*, 493–497. [[CrossRef](#)] [[PubMed](#)]
50. Boussau, B.; Karlberg, E.O.; Frank, A.C.; Legault, B.A.; Andersson, S.G. Computational inference of scenarios for alpha-proteobacterial genome evolution. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9722–9727. [[CrossRef](#)] [[PubMed](#)]
51. “The size and structure of the mitochondrial genome varies greatly in different eukaryotes. It is highly reduced in gene content compared with its bacterial ancestor. For example, “there are 97 genes in the most gene-rich mitochondrial genome (the heterotrophic flagellate *Reclinomonas americana* [49]) compared with 835 genes in the smallest α -proteobacterial genome sequenced so far (*Rickettsia prowazekii* [50])” [48].
52. Gray, M.W. Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a011403. [[CrossRef](#)] [[PubMed](#)]
53. Howe, C.J.; Nisbet, R.E.; Barbrook, A.C. The remarkable chloroplast genome of dinoflagellates. *J. Exp. Bot.* **2008**, *59*, 1035–1045. [[CrossRef](#)] [[PubMed](#)]
54. Szklarczyk, R.; Huynen, M.A. Mosaic origin of the mitochondrial proteome. *Proteomics* **2010**, *10*, 4012–4024. [[CrossRef](#)] [[PubMed](#)]
55. Gray, M.W. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10133–10138. [[CrossRef](#)] [[PubMed](#)]
56. Wallin, I.E. *Symbiogenesis and the Origin of Species*; Williams & Wilkins: Baltimore, MD, USA, 1927.
57. Zook, D. Symbiosis—Evolution’s Co-Author. In *Reticulate Evolution, Interdisciplinary Evolution Research 3*; Gontier, N., Ed.; Springer: Heidelberg, Germany, 2015.
58. Crespi, M.; Frugier, F. De novo organ formation from differentiated cells: Root nodule organogenesis. *Sci. Signal.* **2008**, *1*, re11. [[CrossRef](#)] [[PubMed](#)]
59. Masson-Boivin, C.; Giraud, E.; Perret, X.; Batut, J. Establishing nitrogen-fixing symbiosis with legumes: How many rhizobium recipes? *Trends Microbiol.* **2009**, *17*, 458–466. [[CrossRef](#)] [[PubMed](#)]
60. Finlay, R.D. Ecological aspects of mycorrhizal symbiosis: With special emphasis on the functional diversity of interactions involving the extraradical mycelium. *J. Exp. Bot.* **2008**, *59*, 1115–1126. [[CrossRef](#)] [[PubMed](#)]
61. Bonfante, P.; Anca, I.A. Plants, mycorrhizal fungi, and bacteria: A network of interactions. *Annu. Rev. Microbiol.* **2009**, *63*, 363–383. [[CrossRef](#)] [[PubMed](#)]
62. Pais, R.; Lohs, C.; Wu, Y.; Wang, J.; Aksoy, S. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Appl. Environ. Microbiol.* **2008**, *74*, 5965–5974. [[CrossRef](#)] [[PubMed](#)]
63. Leser, T.D.; Molbak, L. Better living through microbial action: The benefits of the mammalian gastrointestinal microbiota on the host. *Environ. Microbiol.* **2009**, *11*, 2194–2206. [[CrossRef](#)] [[PubMed](#)]
64. Kudo, T. Termite-microbe symbiotic system and its efficient degradation of lignocellulose. *BioSci. Biotechnol. Biochem.* **2009**, *73*, 2561–2567. [[CrossRef](#)] [[PubMed](#)]
65. Suen, G.; Scott, J.J.; Aylward, F.O.; Adams, S.M.; Tringe, S.G.; Pinto-Tomas, A.A.; Foster, C.E.; Pauly, M.; Weimer, P.J.; Barry, K.W.; et al. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet.* **2010**, *6*, e1001129. [[CrossRef](#)] [[PubMed](#)]
66. Blackall, L.L.; Wilson, B.; van Oppen, M.J. Coral—the world’s most diverse symbiotic ecosystem. *Mol. Ecol.* **2015**, *24*, 5330–5347. [[CrossRef](#)] [[PubMed](#)]
67. Rumpho, M.E.; Pelletreau, K.N.; Moustafa, A.; Bhattacharya, D. The making of a photosynthetic animal. *J. Exp. Biol.* **2011**, *214*, 303–311. [[CrossRef](#)] [[PubMed](#)]
68. Kerney, R.; Kim, E.; Hangarter, R.P.; Heiss, A.A.; Bishop, C.D.; Hall, B.K. Intracellular invasion of green algae in a salamander host. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6497–6502. [[CrossRef](#)] [[PubMed](#)]

69. Angelard, C.; Tanner, C.J.; Fontanillas, P.; Niculita-Hirzel, H.; Masclaux, F.; Sanders, I.R. Rapid genotypic change and plasticity in arbuscular mycorrhizal fungi is caused by a host shift and enhanced by segregation. *ISME J.* **2014**, *8*, 284–294. [[CrossRef](#)] [[PubMed](#)]
70. Torres-Cortes, G.; Ghignone, S.; Bonfante, P.; Schussler, A. Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: Transkingdom gene transfer in an ancient mycoplasma-fungus association. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 7785–7790. [[CrossRef](#)] [[PubMed](#)]
71. Symbiotic associations can have more than purely physiological consequences. For example, the shift of a mycorrhizal partner to a novel plant has been reported to trigger rapid change in the fungal genome [69]. It is also interesting to note that mycorrhizal fungi harbor their own prokaryotic endosymbionts [70].
72. Shin, S.C.; Kim, S.H.; You, H.; Kim, B.; Kim, A.C.; Lee, K.A.; Yoon, J.H.; Ryu, J.H.; Lee, W.J. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* **2011**, *334*, 670–674. [[CrossRef](#)] [[PubMed](#)]
73. Coon, K.L.; Valzania, L.; McKinney, D.A.; Vogel, K.J.; Brown, M.R.; Strand, M.R. Bacteria-mediated hypoxia functions as a signal for mosquito development. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E5362–E5369. [[CrossRef](#)] [[PubMed](#)]
74. Benoit, J.B.; Vigneron, A.; Broderick, N.A.; Wu, Y.; Sun, J.S.; Carlson, J.R.; Aksoy, S.; Weiss, B.L. Symbiont-induced odorant binding proteins mediate insect host hematopoiesis. *eLife* **2017**, *6*, e19535. [[CrossRef](#)] [[PubMed](#)]
75. Round, J.L.; Mazmanian, S.K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **2009**, *9*, 313–323. [[CrossRef](#)] [[PubMed](#)]
76. Molloy, M.J.; Bouladoux, N.; Belkaid, Y. Intestinal microbiota: Shaping local and systemic immune responses. *Semin. Immunol.* **2012**, *24*, 58–66. [[CrossRef](#)] [[PubMed](#)]
77. McFall-Ngai, M.; Heath-Heckman, E.A.; Gillette, A.A.; Peyer, S.M.; Harvie, E.A. The secret languages of coevolved symbioses: Insights from the *Euprymna scolopes-Vibrio fischeri* symbiosis. *Semin. Immunol.* **2012**, *24*, 3–8. [[CrossRef](#)] [[PubMed](#)]
78. Bian, G.; Joshi, D.; Dong, Y.; Lu, P.; Zhou, G.; Pan, X.; Xu, Y.; Dimopoulos, G.; Xi, Z. *Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to Plasmodium infection. *Science* **2013**, *340*, 748–751. [[CrossRef](#)] [[PubMed](#)]
79. Moreira, L.A.; Iturbe-Ormaetxe, I.; Jeffery, J.A.; Lu, G.; Pyke, A.T.; Hedges, L.M.; Rocha, B.C.; Hall-Mendelin, S.; Day, A.; Riegler, M.; et al. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and Plasmodium. *Cell* **2009**, *139*, 1268–1278. [[CrossRef](#)] [[PubMed](#)]
80. Hedges, L.M.; Brownlie, J.C.; O'Neill, S.L.; Johnson, K.N. *Wolbachia* and virus protection in insects. *Science* **2008**, *322*, 702. [[CrossRef](#)] [[PubMed](#)]
81. Collins, S.M.; Surette, M.; Bercik, P. The interplay between the intestinal microbiota and the brain. *Nat. Rev. Microbiol.* **2012**, *10*, 735–742. [[CrossRef](#)] [[PubMed](#)]
82. O'Mahony, S.M.; Clarke, G.; Borre, Y.E.; Dinan, T.G.; Cryan, J.F. Serotonin, tryptophan metabolism and the brain-gut-microbiome axis. *Behav. Brain Res.* **2015**, *277*, 32–48. [[CrossRef](#)] [[PubMed](#)]
83. Werren, J.H.; Baldo, L.; Clark, M.E. *Wolbachia*: Master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **2008**, *6*, 741–751. [[CrossRef](#)] [[PubMed](#)]
84. Bordenstein, S.R.; Marshall, M.L.; Fry, A.J.; Kim, U.; Wernegreen, J.J. The tripartite associations between bacteriophage, *Wolbachia*, and arthropods. *PLoS Pathog.* **2006**, *2*, e43.
85. Serbus, L.R.; Sullivan, W. A cellular basis for *Wolbachia* recruitment to the host germline. *PLoS Pathog.* **2007**, *3*, e190. [[CrossRef](#)] [[PubMed](#)]
86. Stouthamer, R.; Breeuwer, J.A.; Hurst, G.D. *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* **1999**, *53*, 71–102. [[CrossRef](#)] [[PubMed](#)]
87. Yen, J.H.; Barr, A.R. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature* **1971**, *232*, 657–658. [[CrossRef](#)] [[PubMed](#)]
88. White, J.A.; Kelly, S.E.; Perlman, S.J.; Hunter, M.S. Cytoplasmic incompatibility in the parasitic wasp *Encarsia inaron*: Disentangling the roles of *Cardinium* and *Wolbachia* symbionts. *Heredity* **2009**, *102*, 483–489. [[CrossRef](#)] [[PubMed](#)]
89. Harris, H.L.; Braig, H.R. Sperm chromatin remodelling and *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila*. *Biochem. Cell Biol.* **2003**, *81*, 229–240. [[CrossRef](#)] [[PubMed](#)]

90. Dobson, S.L. Evolution of *Wolbachia* cytoplasmic incompatibility types. *Evolution* **2004**, *58*, 2156–2166. [[CrossRef](#)] [[PubMed](#)]
91. Kutschera, U.; Niklas, K.J. Macroevolution via secondary endosymbiosis: A Neo-Goldschmidtian view of unicellular hopeful monsters and Darwin's primordial intermediate form. *Theory BioSci.* **2008**, *127*, 277–289. [[CrossRef](#)] [[PubMed](#)]
92. Shropshire, J.D.; Bordenstein, S.R. Speciation by Symbiosis: The Microbiome and Behavior. *MBio* **2016**, *7*, e01785. [[CrossRef](#)] [[PubMed](#)]
93. LePage, D.P.; Metcalf, J.A.; Bordenstein, S.R.; On, J.; Perlmutter, J.I.; Shropshire, J.D.; Layton, E.M.; Funkhouser-Jones, L.J.; Beckmann, J.F.; Bordenstein, S.R. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* **2017**, *543*, 243–247. [[CrossRef](#)] [[PubMed](#)]
94. Thompson, J.R.; Rivera, H.E.; Closek, C.J.; Medina, M. Microbes in the coral holobiont: Partners through evolution, development, and ecological interactions. *Front. Cell. Infect. Microbiol.* **2014**, *4*, 176. [[CrossRef](#)] [[PubMed](#)]
95. Vandenkoornhuysen, P.; Quaiser, A.; Duhamel, M.; Le Van, A.; Dufresne, A. The importance of the microbiome of the plant holobiont. *New Phytol.* **2015**, *206*, 1196–1206. [[CrossRef](#)] [[PubMed](#)]
96. Brooks, A.W.; Kohl, K.D.; Brucker, R.M.; van Opstal, E.J.; Bordenstein, S.R. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLoS Biol.* **2016**, *14*, e2000225. [[CrossRef](#)] [[PubMed](#)]
97. Snyder, A.K.; Rio, R.V. Interwoven biology of the tsetse holobiont. *J. Bacteriol.* **2013**, *195*, 4322–4330. [[CrossRef](#)] [[PubMed](#)]
98. Poulsen, M.; Hu, H.; Li, C.; Chen, Z.; Xu, L.; Otani, S.; Nygaard, S.; Nobre, T.; Klaubauf, S.; Schindler, P.M.; et al. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14500–14505. [[CrossRef](#)] [[PubMed](#)]
99. Husseneder, C. Symbiosis in subterranean termites: A review of insights from molecular studies. *Environ. Entomol.* **2010**, *39*, 378–388. [[CrossRef](#)] [[PubMed](#)]
100. Balmand, S.; Lohs, C.; Aksoy, S.; Heddi, A. Tissue distribution and transmission routes for the tsetse fly endosymbionts. *J. Invertebr. Pathol.* **2013**, *112*, S116–S122. [[CrossRef](#)] [[PubMed](#)]
101. Gilbert, S.F. A holobiont birth narrative: The epigenetic transmission of the human microbiome. *Front. Genet.* **2014**, *5*, 282. [[CrossRef](#)] [[PubMed](#)]
102. Allen, M.A.; Goh, F.; Burns, B.P.; Neilan, B.A. Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. *Geobiology* **2009**, *7*, 82–96. [[CrossRef](#)] [[PubMed](#)]
103. Goh, F.; Allen, M.A.; Leuko, S.; Kawaguchi, T.; Decho, A.W.; Burns, B.P.; Neilan, B.A. Determining the specific microbial populations and their spatial distribution within the stromatolite ecosystem of Shark Bay. *ISME J.* **2009**, *3*, 383–396. [[CrossRef](#)] [[PubMed](#)]
104. Anderson, R.E.; Sogin, M.L.; Baross, J.A. Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics. *PLoS ONE* **2014**, *9*, e109696. [[CrossRef](#)] [[PubMed](#)]
105. McCliment, E.A.; Voglesonger, K.M.; O'Day, P.A.; Dunn, E.E.; Holloway, J.R.; Cary, S.C. Colonization of nascent, deep-sea hydrothermal vents by a novel Archaeal and Nanoarchaeal assemblage. *Environ. Microbiol.* **2006**, *8*, 114–125. [[CrossRef](#)] [[PubMed](#)]
106. Ahmadjian, V. Lichens. *Annu. Rev. Microbiol.* **1965**, *19*, 1–20. [[CrossRef](#)] [[PubMed](#)]
107. Grube, M.; Cardinale, M.; de Castro, J.V., Jr.; Muller, H.; Berg, G. Species-specific structural and functional diversity of bacterial communities in lichen symbioses. *ISME J.* **2009**, *3*, 1105–1115. [[CrossRef](#)] [[PubMed](#)]
108. Lucking, R.; Dal-Forno, M.; Sikaroodi, M.; Gillevet, P.M.; Bungartz, F.; Moncada, B.; Yanez-Ayabaca, A.; Chaves, J.L.; Coca, L.F.; Lawrey, J.D. A single macrolichen constitutes hundreds of unrecognized species. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11091–11096. [[CrossRef](#)] [[PubMed](#)]
109. Honegger, R. Simon Schwendener (1829–1919) and the dual hypothesis in lichens. *Bryologist* **2000**, *103*, 307–313. [[CrossRef](#)]
110. Kohl, K.D.; Sadowska, E.T.; Rudolf, A.M.; Dearing, M.D.; Koteja, P. Experimental Evolution on a Wild Mammal Species Results in Modifications of Gut Microbial Communities. *Front. Microbiol.* **2016**, *7*, 634. [[CrossRef](#)] [[PubMed](#)]

111. Oliver, K.M.; Degnan, P.H.; Burke, G.R.; Moran, N.A. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu. Rev. Entomol.* **2010**, *55*, 247–266. [[CrossRef](#)] [[PubMed](#)]
112. Oliver, K.M.; Russell, J.A.; Moran, N.A.; Hunter, M.S. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1803–1807. [[CrossRef](#)] [[PubMed](#)]
113. Dunning Hotopp, J.C.; Clark, M.E.; Oliveira, D.C.; Foster, J.M.; Fischer, P.; Munoz Torres, M.C.; Giebel, J.D.; Kumar, N.; Ishmael, N.; Wang, S.; et al. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **2007**, *317*, 1753–1756. [[CrossRef](#)] [[PubMed](#)]
114. Timmis, J.N.; Ayliffe, M.A.; Huang, C.Y.; Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **2004**, *5*, 123–135. [[CrossRef](#)] [[PubMed](#)]
115. Nikoh, N.; Nakabachi, A. Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biol.* **2009**, *7*, 12. [[CrossRef](#)] [[PubMed](#)]
116. Woolfit, M.; Iturbe-Ormaetxe, I.; McGraw, E.A.; O'Neill, S.L. An ancient horizontal gene transfer between mosquito and the endosymbiotic bacterium *Wolbachia pipientis*. *Mol. Biol. Evol.* **2009**, *26*, 367–374. [[CrossRef](#)] [[PubMed](#)]
117. Klasson, L.; Kambris, Z.; Cook, P.E.; Walker, T.; Sinkins, S.P. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. *BMC Genom.* **2009**, *10*, 33. [[CrossRef](#)] [[PubMed](#)]
118. Brelsfoard, C.; Tsiamis, G.; Falchetto, M.; Gomulski, L.M.; Telleria, E.; Alam, U.; Doudoumis, V.; Scolari, F.; Benoit, J.B.; Swain, M.; et al. Presence of Extensive *Wolbachia* Symbiont Insertions Discovered in the Genome of Its Host *Glossina morsitans morsitans*. *PLoS Negl. Trop. Dis.* **2014**, *8*, e2728. [[CrossRef](#)] [[PubMed](#)]
119. Aikawa, T.; Anbutsu, H.; Nikoh, N.; Kikuchi, T.; Shibata, F.; Fukatsu, T. Longicorn beetle that vectors pinewood nematode carries many *Wolbachia* genes on an autosome. *Proc. Biol. Sci.* **2009**, *276*, 3791–3798. [[CrossRef](#)] [[PubMed](#)]
120. Nikoh, N.; Tanaka, K.; Shibata, F.; Kondo, N.; Hizume, M.; Shimada, M.; Fukatsu, T. *Wolbachia* genome integrated in an insect chromosome: Evolution and fate of laterally transferred endosymbiont genes. *Genome Res.* **2008**, *18*, 272–280. [[CrossRef](#)] [[PubMed](#)]
121. Salzberg, S.L.; Hotopp, J.C.; Delcher, A.L.; Pop, M.; Smith, D.R.; Eisen, M.B.; Nelson, W.C. Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol.* **2005**, *6*, R23. [[CrossRef](#)] [[PubMed](#)]
122. Klasson, L.; Kumar, N.; Bromley, R.; Sieber, K.; Flowers, M.; Ott, S.H.; Tallon, L.J.; Andersson, S.G.; Dunning Hotopp, J.C. Extensive duplication of the *Wolbachia* DNA in chromosome four of *Drosophila ananassae*. *BMC Genom.* **2014**, *15*, 1097. [[CrossRef](#)] [[PubMed](#)]
123. Choi, J.Y.; Bubnell, J.E.; Aquadro, C.F. Population Genomics of Infectious and Integrated *Wolbachia pipientis* Genomes in *Drosophila ananassae*. *Genome Biol. Evol.* **2015**, *7*, 2362–2382. [[CrossRef](#)] [[PubMed](#)]
124. Stebbins, J.G.L. Cataclysmic Evolution. *Sci. Am.* **1951**, *184*, 54–59. [[CrossRef](#)]
125. Pennisi, E. Shaking up the Tree of Life. *Science* **2016**, *354*, 817–821. [[CrossRef](#)] [[PubMed](#)]
126. Hegarty, M.J.; Barker, G.L.; Wilson, I.D.; Abbott, R.J.; Edwards, K.J.; Hiscock, S.J. Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. *Curr. Biol.* **2006**, *16*, 1652–1659. [[CrossRef](#)] [[PubMed](#)]
127. Tayale, A.; Parisod, C. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* **2013**, *140*, 79–96. [[CrossRef](#)] [[PubMed](#)]
128. Jaillon, O.; Aury, J.M.; Wincker, P. “Changing by doubling”, the impact of Whole Genome Duplications in the evolution of eukaryotes. *C. R. Biol.* **2009**, *332*, 241–253. [[CrossRef](#)] [[PubMed](#)]
129. Marcet-Houben, M.; Gabaldon, T. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker’s Yeast Lineage. *PLoS Biol.* **2015**, *13*, e1002220. [[CrossRef](#)] [[PubMed](#)]
130. Wolfe, K.H. Origin of the Yeast Whole-Genome Duplication. *PLoS Biol.* **2015**, *13*, e1002221. [[CrossRef](#)] [[PubMed](#)]
131. Albertin, W.; Marullo, P. Polyploidy in fungi: Evolution after whole-genome duplication. *Proc. Biol. Sci.* **2012**, *279*, 2497–2509. [[CrossRef](#)] [[PubMed](#)]
132. Aury, J.M. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **2006**, *444*, 171–178. [[CrossRef](#)] [[PubMed](#)]
133. Tian, C.G.; Xiong, Y.Q.; Liu, T.Y.; Sun, S.H.; Chen, L.B.; Chen, M.S. Evidence for an ancient whole-genome duplication event in rice and other cereals. *Yi Chuan Xue Bao* **2005**, *32*, 519–527. [[PubMed](#)]

134. Jiao, Y.; Wickett, N.J.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, L.P.; Hu, Y.; Liang, H.; Soltis, P.S.; et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **2011**, *473*, 97–100. [[CrossRef](#)] [[PubMed](#)]
135. Nossa, C.W.; Havlak, P.; Yue, J.X.; Lv, J.; Vincent, K.Y.; Brockmann, H.J.; Putnam, N.H. Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *Gigascience* **2014**, *3*, 9. [[CrossRef](#)] [[PubMed](#)]
136. Dehal, P.; Boore, J.L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **2005**, *3*, e314. [[CrossRef](#)] [[PubMed](#)]
137. Hufton, A.L.; Groth, D.; Vingron, M.; Lehrach, H.; Poustka, A.J.; Panopoulou, G. Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. *Genome Res.* **2008**, *18*, 1582–1591. [[CrossRef](#)] [[PubMed](#)]
138. Taylor, J.S.; Braasch, I.; Frickey, T.; Meyer, A.; Van de Peer, Y. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* **2003**, *13*, 382–390. [[CrossRef](#)] [[PubMed](#)]
139. Hoegg, S.; Brinkmann, H.; Taylor, J.S.; Meyer, A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* **2004**, *59*, 190–203. [[CrossRef](#)] [[PubMed](#)]
140. Glasauer, S.M.; Neuhauss, S.C. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genom.* **2014**, *289*, 1045–1060. [[CrossRef](#)] [[PubMed](#)]
141. Brodie, E.D., III. How an ancient genome duplication electrified modern fish. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21953–21954. [[CrossRef](#)] [[PubMed](#)]
142. Veron, A.S.; Kaufmann, K.; Bornberg-Bauer, E. Evidence of interaction network evolution by whole-genome duplications: A case study in MADS-box proteins. *Mol. Biol. Evol.* **2007**, *24*, 670–678. [[CrossRef](#)] [[PubMed](#)]
143. Kassahn, K.S.; Dang, V.T.; Wilkins, S.J.; Perkins, A.C.; Ragan, M.A. Evolution of gene function and regulatory control after whole-genome duplication: Comparative analyses in vertebrates. *Genome Res.* **2009**, *19*, 1404–1418. [[CrossRef](#)] [[PubMed](#)]
144. Fusco, D.; Grassi, L.; Bassetti, B.; Caselle, M.; Cosentino Lagomarsino, M. Ordered structure of the transcription network inherited from the yeast whole-genome duplication. *BMC Syst. Biol.* **2010**, *4*, 77. [[CrossRef](#)] [[PubMed](#)]
145. Conant, G.C. Rapid reorganization of the transcriptional regulatory network after genome duplication in yeast. *Proc. Biol. Sci.* **2010**, *277*, 869–876. [[CrossRef](#)] [[PubMed](#)]
146. Zhu, Y.; Lin, Z.; Nakhleh, L. Evolution after whole-genome duplication: A network perspective. *G3 (Bethesda)* **2013**, *3*, 2049–2057. [[CrossRef](#)] [[PubMed](#)]
147. De Smet, R.; Van de Peer, Y. Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr. Opin. Plant Biol.* **2012**, *15*, 168–176. [[CrossRef](#)] [[PubMed](#)]
148. Lamichhaney, S.; Berglund, J.; Almen, M.S.; Maqbool, K.; Grabherr, M.; Martinez-Barrio, A.; Promerova, M.; Rubin, C.J.; Wang, C.; Zamani, N.; et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **2015**, *518*, 371–375. [[CrossRef](#)] [[PubMed](#)]
149. Almen, M.S.; Lamichhaney, S.; Berglund, J.; Grant, B.R.; Grant, P.R.; Webster, M.T.; Andersson, L. Adaptive radiation of Darwin's finches revisited using whole genome sequencing. *Bioessays* **2016**, *38*, 14–20. [[CrossRef](#)] [[PubMed](#)]
150. Palmer, D.H.; Kronforst, M.R. Divergence and gene flow among Darwin's finches: A genome-wide view of adaptive radiation driven by interspecies allele sharing. *Bioessays* **2015**, *37*, 968–974. [[CrossRef](#)] [[PubMed](#)]
151. Kocher, T.D. Adaptive evolution and explosive speciation: The cichlid fish model. *Nat. Rev. Genet.* **2004**, *5*, 288–298. [[CrossRef](#)] [[PubMed](#)]
152. Fan, S.; Meyer, A. Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Front. Genet.* **2014**, *5*, 163. [[CrossRef](#)] [[PubMed](#)]
153. Svensson, O.; Smith, A.; Garcia-Alonso, J.; van Oosterhout, C. Hybridization generates a hopeful monster: A hermaphroditic selfing cichlid. *R. Soc. Open Sci.* **2016**, *3*, 150684. [[CrossRef](#)] [[PubMed](#)]
154. Brawand, D.; Wagner, C.E.; Li, Y.I.; Malinsky, M.; Keller, I.; Fan, S.; Simakov, O.; Ng, A.Y.; Lim, Z.W.; Bezault, E.; et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **2014**, *513*, 375–381. [[CrossRef](#)] [[PubMed](#)]
155. Heliconius, G.C. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **2012**, *487*, 94–98.

156. Pardo-Diaz, C.; Salazar, C.; Baxter, S.W.; Merot, C.; Figueiredo-Ready, W.; Joron, M.; McMillan, W.O.; Jiggins, C.D. Adaptive Introgression across Species Boundaries in *Heliconius* Butterflies. *PLoS Genet.* **2012**, *8*, e1002752. [[CrossRef](#)] [[PubMed](#)]
157. Brower, A.V. Introgression of wing pattern alleles and speciation via homoploid hybridization in *Heliconius* butterflies: A review of evidence from the genome. *Proc. Biol. Sci.* **2013**, *280*, 20122302. [[CrossRef](#)] [[PubMed](#)]
158. Edwards, S.V.; Potter, S.; Schmitt, C.J.; Bragg, J.G.; Moritz, C. Reticulation, divergence, and the phylogeography-phylogenetics continuum. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8025–8032. [[CrossRef](#)] [[PubMed](#)]
159. Wen, D.; Yu, Y.; Hahn, M.W.; Nakhleh, L. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Mol. Ecol.* **2016**, *25*, 2361–2372. [[CrossRef](#)] [[PubMed](#)]
160. Crawford, J.E.; Riehle, M.M.; Guelbeogo, W.M.; Gnome, A.; Sagnon, N.; Vernick, K.D.; Nielsen, R.; Lazzaro, B.P. Reticulate Speciation and Barriers to Introgression in the *Anopheles gambiae* Species Complex. *Genome Biol. Evol.* **2015**, *7*, 3116–3131. [[CrossRef](#)] [[PubMed](#)]
161. Lotsy, J.P. *Evolution by Means of Hybridization*; Martinus Nijhoff: The Hague, The Netherlands, 1916.
162. Barton, N.H. The role of hybridization in evolution. *Mol. Ecol.* **2001**, *10*, 551–568. [[CrossRef](#)] [[PubMed](#)]
163. Anderson, E.; Stebbins, G.L., Jr. Hybridization as an evolutionary stimulus. *Evolution* **1954**, *8*, 378–388. [[CrossRef](#)]
164. Arnold, M.L. Transfer and origin of adaptations through natural hybridization: Were Anderson and Stebbins right? *Plant Cell* **2004**, *16*, 562–570. [[CrossRef](#)] [[PubMed](#)]
165. Payseur, B.A.; Rieseberg, L.H. A genomic perspective on hybridization and speciation. *Mol. Ecol.* **2016**, *25*, 2337–2360. [[CrossRef](#)] [[PubMed](#)]
166. Abbott, R.; Albach, D.; Ansell, S.; Arntzen, J.W.; Baird, S.J.; Bierne, N.; Boughman, J.; Brelsford, A.; Buerkle, C.A.; Buggs, R.; et al. Hybridization and speciation. *J. Evol. Biol.* **2013**, *26*, 229–246. [[CrossRef](#)] [[PubMed](#)]
167. Parisod, C.; Alix, K.; Just, J.; Petit, M.; Sarilar, V.; Mhiri, C.; Ainouche, M.; Chalhouh, B.; Grandbastien, M.A. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol.* **2010**, *186*, 37–45. [[CrossRef](#)] [[PubMed](#)]
168. Shapiro, J.A. Epigenetic control of mobile DNA as an interface between experience and genome change. *Front. Genet.* **2014**, *5*, 87. [[CrossRef](#)] [[PubMed](#)]
169. Goldschmidt, R. *The Material Basis of Evolution, Reissued (The Silliman Memorial Lectures Series)*, 1982; Yale University Press: New Haven, CT, USA, 1940.
170. Dittrich-Reed, D.R.; Fitzpatrick, B.M. Transgressive Hybrids as Hopeful Monsters. *Evol. Biol.* **2013**, *40*, 310–315. [[CrossRef](#)] [[PubMed](#)]
171. Dietrich, M.R. Richard Goldschmidt: Hopeful monsters and other ‘heresies’. *Nat. Rev. Genet.* **2003**, *4*, 68–74. [[CrossRef](#)] [[PubMed](#)]
172. Theissen, G. Saltational evolution: Hopeful monsters are here to stay. *Theory BioSci.* **2009**, *128*, 43–51. [[CrossRef](#)] [[PubMed](#)]
173. Theissen, G. The proper place of hopeful monsters in evolutionary biology. *Theory BioSci.* **2006**, *124*, 349–369. [[CrossRef](#)] [[PubMed](#)]
174. Syvanen, M. Evolutionary implications of horizontal gene transfer. *Annu. Rev. Genet.* **2012**, *46*, 341–358. [[CrossRef](#)] [[PubMed](#)]
175. Fournier, G.P.; Huang, J.; Gogarten, J.P. Horizontal gene transfer from extinct and extant lineages: Biological innovation and the coral of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2009**, *364*, 2229–2239. [[CrossRef](#)] [[PubMed](#)]
176. Soucy, S.M.; Huang, J.; Gogarten, J.P. Horizontal gene transfer: Building the web of life. *Nat. Rev. Genet.* **2015**, *16*, 472–482. [[CrossRef](#)] [[PubMed](#)]
177. Hayes, W. *The Genetics of Bacteria and their Viruses*, 2nd ed.; Blackwell: London, UK, 1968.
178. Alvarez-Martinez, C.E.; Christie, P.J. Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* **2009**, *73*, 775–808. [[CrossRef](#)] [[PubMed](#)]
179. Cascales, E.; Christie, P.J. The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **2003**, *1*, 137–149. [[CrossRef](#)] [[PubMed](#)]

180. Averhoff, B.; Friedrich, A. Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. *Arch. Microbiol.* **2003**, *180*, 385–393. [[CrossRef](#)] [[PubMed](#)]
181. Juhas, M.; Crook, D.W.; Hood, D.W. Type IV secretion systems: Tools of bacterial horizontal gene transfer and virulence. *Cell. Microbiol.* **2008**, *10*, 2377–2386. [[CrossRef](#)] [[PubMed](#)]
182. Zupan, J.R.; Zambryski, P. Transfer of T-DNA from *Agrobacterium* to the plant cell. *Plant Physiol.* **1995**, *107*, 1041–1047. [[CrossRef](#)] [[PubMed](#)]
183. Chilton, M.D. A Vector for Introducing New Genes into Plants. *Sci. Am.* **1983**, *248*, 50–59. [[CrossRef](#)]
184. Chilton, M.D. *Agrobacterium*. A memoir. *Plant Physiol.* **2001**, *125*, 9–14. [[CrossRef](#)] [[PubMed](#)]
185. Mitrikeski, P.T. Ecologically driven competence for exogenous DNA uptake in yeast. *Curr. Microbiol.* **2015**, *70*, 883–893. [[CrossRef](#)] [[PubMed](#)]
186. Koontz, L. Explanatory chapter: Introducing exogenous DNA into cells. *Methods Enzymol.* **2013**, *529*, 29–34. [[PubMed](#)]
187. Cai, J.; Han, Y.; Ren, H.; Chen, C.; He, D.; Zhou, L.; Eisner, G.M.; Asico, L.D.; Jose, P.A.; Zeng, C. Extracellular vesicle-mediated transfer of donor genomic DNA to recipient cells is a novel mechanism for genetic influence between cells. *J. Mol. Cell. Biol.* **2013**, *5*, 227–238. [[CrossRef](#)] [[PubMed](#)]
188. Fischer, S.; Cornils, K.; Speiseder, T.; Badbaran, A.; Reimer, R.; Indenbirken, D.; Grundhoff, A.; Brunswig-Spickenheier, B.; Alawi, M.; Lange, C. Indication of Horizontal DNA Gene Transfer by Extracellular Vesicles. *PLoS ONE* **2016**, *11*, e0163665. [[CrossRef](#)] [[PubMed](#)]
189. Zhao, Y.; Yu, M.; Wang, L.; Li, Y.; Fan, J.; Yang, Q.; Jin, Y. Spontaneous uptake of exogenous DNA by goat spermatozoa and selection of donor bucks for sperm-mediated gene transfer. *Mol. Biol. Rep.* **2012**, *39*, 2659–2664. [[CrossRef](#)] [[PubMed](#)]
190. Arias, M.E.; Sanchez-Villalba, E.; Delgado, A.; Felmer, R. Effect of transfection and co-incubation of bovine sperm with exogenous DNA on sperm quality and functional parameters for its use in sperm-mediated gene transfer. *Zygote* **2017**, *25*, 85–97. [[CrossRef](#)] [[PubMed](#)]
191. Huang, J. Horizontal gene transfer in eukaryotes: The weak-link model. *Bioessays* **2013**, *35*, 868–875. [[CrossRef](#)] [[PubMed](#)]
192. Adegoke, A.A.; Faleye, A.C.; Singh, G.; Stenstrom, T.A. Antibiotic Resistant Superbugs: Assessment of the Interrelationship of Occurrence in Clinical Settings and Environmental Niches. *Molecules* **2016**, *22*, 29. [[CrossRef](#)] [[PubMed](#)]
193. Mazel, D.; Davies, J. Antibiotic resistance in microbes. *Cell. Mol. Life Sci.* **1999**, *56*, 742–754. [[CrossRef](#)] [[PubMed](#)]
194. Toprak, E.; Veres, A.; Michel, J.B.; Chait, R.; Hartl, D.L.; Kishony, R. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **2011**, *44*, 101–105. [[CrossRef](#)] [[PubMed](#)]
195. Das, R.; Gupta, P.; Singh, P.; Chauhan, D.S.; Katoch, K.; Katoch, V.M. Association of mutations in rpsL gene with high degree of streptomycin resistance in clinical isolates of *Mycobacterium tuberculosis* in India. *Indian J. Med. Res.* **2009**, *129*, 108–110. [[PubMed](#)]
196. Watanabe, T.; Fukasawa, T. Episome-mediated transfer of drug resistance in Enterobacteriaceae. III. Transduction of resistance factors. *J. Bacteriol.* **1961**, *82*, 202–209. [[PubMed](#)]
197. Morar, M.; Wright, G.D. The genomic enzymology of antibiotic resistance. *Annu. Rev. Genet.* **2010**, *44*, 25–51. [[CrossRef](#)] [[PubMed](#)]
198. Ravi, A.; Avershina, E.; Ludvigsen, J.; L’Abee-Lund, T.M.; Rudi, K. Integrons in the intestinal microbiota as reservoirs for transmission of antibiotic resistance genes. *Pathogens* **2014**, *3*, 238–248. [[CrossRef](#)] [[PubMed](#)]
199. Andam, C.P.; Fournier, G.P.; Gogarten, J.P. Multilevel populations and the evolution of antibiotic resistance through horizontal gene transfer. *FEMS Microbiol. Rev.* **2011**, *35*, 756–767. [[CrossRef](#)] [[PubMed](#)]
200. Bardaji, L.; Echeverria, M.; Rodriguez-Palenzuela, P.; Martinez-Garcia, P.M.; Murillo, J. Four genes essential for recombination define GI_{nts}, a new type of mobile genomic island widespread in bacteria. *Sci. Rep.* **2017**, *7*, 46254. [[CrossRef](#)] [[PubMed](#)]
201. Koonin, E.V.; Makarova, K.S.; Aravind, L. Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu. Rev. Microbiol.* **2001**, *55*, 709–742. [[CrossRef](#)] [[PubMed](#)]
202. Koonin, E.V. Horizontal gene transfer: Essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Research* **2016**, *5*. [[CrossRef](#)] [[PubMed](#)]
203. Aravind, L.; Tatusov, R.L.; Wolf, Y.I.; Walker, D.R.; Koonin, E.V. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* **1998**, *14*, 442–444. [[CrossRef](#)]

204. Dodsworth, J.A.; Li, L.; Wei, S.; Hedlund, B.P.; Leigh, J.A.; de Figueiredo, P. Inter-domain conjugal transfer of DNA from Bacteria to Archaea. *Appl. Environ. Microbiol.* **2010**, *76*, 5644–5647. [[CrossRef](#)] [[PubMed](#)]
205. Kanhere, A.; Vingron, M. Horizontal Gene Transfers in prokaryotes show differential preferences for metabolic and translational genes. *BMC Evol. Biol.* **2009**, *9*, 9. [[CrossRef](#)] [[PubMed](#)]
206. Little, C.A.; Tweats, D.J.; Pinney, R.J. Relevance of plasmid pKM101-mediated mutagenicity in bacteria to genotoxicity in mammalian cells. *Mutagenesis* **1989**, *4*, 371–376. [[CrossRef](#)] [[PubMed](#)]
207. Okubo, T.; Piromyong, P.; Tittabutr, P.; Teaumroong, N.; Minamisawa, K. Origin and Evolution of Nitrogen Fixation Genes on Symbiosis Islands and Plasmid in *Bradyrhizobium*. *Microbes Environ.* **2016**, *31*, 260–267. [[CrossRef](#)] [[PubMed](#)]
208. Hoffmeister, M.; Martin, W. Interspecific evolution: Microbial symbiosis, endosymbiosis and gene transfer. *Environ. Microbiol.* **2003**, *5*, 641–649. [[CrossRef](#)] [[PubMed](#)]
209. Nakatsukasa, H.; Uchiumi, T.; Kucho, K.; Suzuki, A.; Higashi, S.; Abe, M. Transposon mediation allows a symbiotic plasmid of *Rhizobium leguminosarum* bv. trifolii to become a symbiosis island in *Agrobacterium* and *Rhizobium*. *J. Gen. Appl. Microbiol.* **2008**, *54*, 107–118. [[CrossRef](#)] [[PubMed](#)]
210. Clewell, D.B. Properties of *Enterococcus faecalis* plasmid pAD1, a member of a widely disseminated family of pheromone-responding, conjugative, virulence elements encoding cytolysin. *Plasmid* **2007**, *58*, 205–227. [[CrossRef](#)] [[PubMed](#)]
211. Moreno Switt, A.I.; den Bakker, H.C.; Cummings, C.A.; Rodriguez-Rivera, L.D.; Govoni, G.; Raneiri, M.L.; Degoricija, L.; Brown, S.; Hoelzer, K.; Peters, J.E.; et al. Identification and characterization of novel Salmonella mobile elements involved in the dissemination of genes linked to virulence and transmission. *PLoS ONE* **2012**, *7*, e41247. [[CrossRef](#)] [[PubMed](#)]
212. Penades, J.R.; Chen, J.; Quiles-Puchalt, N.; Carpena, N.; Novick, R.P. Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* **2015**, *23*, 171–178. [[CrossRef](#)] [[PubMed](#)]
213. Cornelis, G.R.; Boland, A.; Boyd, A.P.; Geuijen, C.; Iriarte, M.; Neyt, C.; Sory, M.P.; Stainier, I. The virulence plasmid of Yersinia, an antihost genome. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1315–1352. [[PubMed](#)]
214. Nelson-Sathi, S.; Dagan, T.; Landan, G.; Janssen, A.; Steel, M.; McInerney, J.O.; Deppenmeier, U.; Martin, W.F. Acquisition of 1000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20537–20542. [[CrossRef](#)] [[PubMed](#)]
215. Lopez-Garcia, P.; Zivanovic, Y.; Deschamps, P.; Moreira, D. Bacterial gene import and mesophilic adaptation in archaea. *Nat. Rev. Microbiol.* **2015**, *13*, 447–456. [[CrossRef](#)] [[PubMed](#)]
216. Deschamps, P.; Zivanovic, Y.; Moreira, D.; Rodriguez-Valera, F.; Lopez-Garcia, P. Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic thaumarchaeota and euryarchaeota. *Genome Biol. Evol.* **2014**, *6*, 1549–1563. [[CrossRef](#)] [[PubMed](#)]
217. Hall, R.M.; Stokes, H.W. Integrons: Novel DNA elements which capture genes by site-specific recombination. *Genetica* **1993**, *90*, 115–132. [[CrossRef](#)] [[PubMed](#)]
218. Hall, R.M. Integrons and gene cassettes: Hotspots of diversity in bacterial genomes. *Ann. N. Y. Acad. Sci.* **2012**, *1267*, 71–78. [[CrossRef](#)] [[PubMed](#)]
219. Johnson, C.M.; Grossman, A.D. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu. Rev. Genet.* **2015**, *49*, 577–601. [[CrossRef](#)] [[PubMed](#)]
220. Juhas, M.; van der Meer, J.R.; Gaillard, M.; Harding, R.M.; Hood, D.W.; Crook, D.W. Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **2009**, *33*, 376–393. [[CrossRef](#)] [[PubMed](#)]
221. Daccord, A.; Ceccarelli, D.; Rodrigue, S.; Burrus, V. Comparative analysis of mobilizable genomic islands. *J. Bacteriol.* **2013**, *195*, 606–614. [[CrossRef](#)] [[PubMed](#)]
222. Carraro, N.; Rivard, N.; Burrus, V.; Ceccarelli, D. Mobilizable genomic islands, different strategies for the dissemination of multidrug resistance and other adaptive traits. *Mob. Genet. Elem.* **2017**, *7*, 1–6. [[CrossRef](#)] [[PubMed](#)]
223. Bellanger, X.; Payot, S.; Leblond-Bourget, N.; Guedon, G. Conjugative and mobilizable genomic islands in bacteria: Evolution and diversity. *FEMS Microbiol. Rev.* **2014**, *38*, 720–760. [[CrossRef](#)] [[PubMed](#)]
224. Groth, A.C.; Calos, M.P. Phage integrases: Biology and applications. *J. Mol. Biol.* **2004**, *335*, 667–678. [[CrossRef](#)] [[PubMed](#)]
225. Nesmelova, I.V.; Hackett, P.B. DDE transposases: Structural similarity and diversity. *Adv. Drug Deliv. Rev.* **2010**, *62*, 1187–1195. [[CrossRef](#)] [[PubMed](#)]

226. Guerillot, R.; Siguier, P.; Gourbeyre, E.; Chandler, M.; Glaser, P. The diversity of prokaryotic DDE transposases of the mutator superfamily, insertion specificity, and association with conjugation machineries. *Genome Biol. Evol.* **2014**, *6*, 260–272. [[CrossRef](#)] [[PubMed](#)]
227. Rowe-Magnus, D.A.; Guérout, A.M.; Mazel, D. Super-integrins. *Res. Microbiol.* **1999**, *150*, 641–651. [[CrossRef](#)]
228. Ramsay, J.P.; Sullivan, J.T.; Stuart, G.S.; Lamont, I.L.; Ronson, C.W. Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol. Microbiol.* **2006**, *62*, 723–734. [[CrossRef](#)] [[PubMed](#)]
229. Kers, J.A.; Cameron, K.D.; Joshi, M.V.; Bukhalid, R.A.; Morello, J.E.; Wach, M.J.; Gibson, D.M.; Loria, R. A large, mobile pathogenicity island confers plant pathogenicity on *Streptomyces* species. *Mol. Microbiol.* **2005**, *55*, 1025–1033. [[CrossRef](#)] [[PubMed](#)]
230. Huguet-Tapia, J.C.; Badger, J.H.; Loria, R.; Pettis, G.S. *Streptomyces turgidiscabies* Car8 contains a modular pathogenicity island that shares virulence genes with other actinobacterial plant pathogens. *Plasmid* **2011**, *65*, 118–124. [[CrossRef](#)] [[PubMed](#)]
231. Baptiste, E. *Les Gènes Voyageurs: L'odyssée de L'évolution*; Belin: Paris, France, 2013.
232. Crisp, A.; Boschetti, C.; Perry, M.; Tunnacliffe, A.; Micklem, G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **2015**, *16*, 50. [[CrossRef](#)] [[PubMed](#)]
233. Hirt, R.P.; Alsmark, C.; Embley, T.M. Lateral gene transfers and the origins of the eukaryote proteome: A view from microbial parasites. *Curr. Opin. Microbiol.* **2015**, *23*, 155–162. [[CrossRef](#)] [[PubMed](#)]
234. Gao, C.; Ren, X.; Mason, A.S.; Liu, H.; Xiao, M.; Li, J.; Fu, D. Horizontal gene transfer in plants. *Funct. Integr. Genom.* **2014**, *14*, 23–29. [[CrossRef](#)] [[PubMed](#)]
235. Andersson, J.O. Gene transfer and diversification of microbial eukaryotes. *Annu. Rev. Microbiol.* **2009**, *63*, 177–193. [[CrossRef](#)] [[PubMed](#)]
236. Boto, L. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc. Biol. Sci.* **2014**, *281*, 20132450. [[CrossRef](#)] [[PubMed](#)]
237. Daubin, V.; Szollosi, G.J. Horizontal Gene Transfer and the History of Life. *Cold Spring Harb. Perspect. Biol.* **2016**, *8*, a018036. [[CrossRef](#)] [[PubMed](#)]
238. Grassi, L.; Grilli, J.; Lagomarsino, M.C. Large-scale dynamics of horizontal transfers. *Mob. Genet. Elem.* **2012**, *2*, 163–167. [[CrossRef](#)] [[PubMed](#)]
239. Grilli, J.; Romano, M.; Bassetti, F.; Cosentino Lagomarsino, M. Cross-species gene-family fluctuations reveal the dynamics of horizontal transfers. *Nucleic Acids Res.* **2014**, *42*, 6850–6860. [[CrossRef](#)] [[PubMed](#)]
240. Haegeman, A.; Jones, J.T.; Danchin, E.G. Horizontal gene transfer in nematodes: A catalyst for plant parasitism? *Mol. Plant-Microbe Interact.* **2011**, *24*, 879–887. [[CrossRef](#)] [[PubMed](#)]
241. Wybouw, N.; Pauchet, Y.; Heckel, D.G.; Van Leeuwen, T. Horizontal Gene Transfer Contributes to the Evolution of Arthropod Herbivory. *Genome Biol. Evol.* **2016**, *8*, 1785–1801. [[CrossRef](#)] [[PubMed](#)]
242. McKenna, D.D.; Scully, E.D.; Pauchet, Y.; Hoover, K.; Kirsch, R.; Geib, S.M.; Mitchell, R.F.; Waterhouse, R.M.; Ahn, S.J.; Arsala, D.; et al. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biol.* **2016**, *17*, 227. [[CrossRef](#)] [[PubMed](#)]
243. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **2006**, *117*, 5–16. [[CrossRef](#)] [[PubMed](#)]
244. Koonin, E.V. Viruses and mobile elements as drivers of evolutionary transitions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2016**, *371*. [[CrossRef](#)] [[PubMed](#)]
245. Filee, J. Route of NCLDV evolution: The genomic accordion. *Curr. Opin. Virol.* **2013**, *3*, 595–599. [[CrossRef](#)] [[PubMed](#)]
246. Gomez-Valero, L.; Buchrieser, C. Genome dynamics in *Legionella*: The basis of versatility and adaptation to intracellular replication. *Cold Spring Harb. Perspect. Med.* **2013**, *3*. [[CrossRef](#)] [[PubMed](#)]
247. Hoffmann, C.; Harrison, C.F.; Hilbi, H. The natural alternative: Protozoa as cellular models for *Legionella* infection. *Cell. Microbiol.* **2014**, *16*, 15–26. [[CrossRef](#)] [[PubMed](#)]
248. Yousuf, F.A.; Siddiqui, R.; Khan, N.A. *Acanthamoeba castellanii* of the T4 genotype is a potential environmental host for *Enterobacter aerogenes* and *Aeromonas hydrophila*. *Parasit Vectors* **2013**, *6*, 169. [[CrossRef](#)] [[PubMed](#)]
249. Moliner, C.; Fournier, P.E.; Raoult, D. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol. Rev.* **2010**, *34*, 281–294. [[CrossRef](#)] [[PubMed](#)]

250. Boyer, M.; Yutin, N.; Pagnier, I.; Barrassi, L.; Fournous, G.; Espinosa, L.; Robert, C.; Azza, S.; Sun, S.; Rossmann, M.G.; et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21848–21853. [[CrossRef](#)] [[PubMed](#)]
251. Schaack, S.; Gilbert, C.; Feschotte, C. Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **2010**, *25*, 537–546. [[CrossRef](#)] [[PubMed](#)]
252. Panaud, O. Horizontal transfers of transposable elements in eukaryotes: The flying genes. *C. R. Biol.* **2016**, *339*, 296–299. [[CrossRef](#)] [[PubMed](#)]
253. Walsh, A.M.; Kortschak, R.D.; Gardner, M.G.; Bertozzi, T.; Adelson, D.L. Widespread horizontal transfer of retrotransposons. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1012–1016. [[CrossRef](#)] [[PubMed](#)]
254. Gilbert, C.; Cordaux, R. Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes. *Genome Biol. Evol.* **2013**, *5*, 822–832. [[CrossRef](#)] [[PubMed](#)]
255. Parisot, N.; Pelin, A.; Gasc, C.; Polonais, V.; Belkorchia, A.; Panek, J.; El Alaoui, H.; Biron, D.G.; Brasset, E.; Vaury, C.; et al. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. *Genome Biol. Evol.* **2014**, *6*, 2289–2300. [[CrossRef](#)] [[PubMed](#)]
256. Suh, A.; Witt, C.C.; Menger, J.; Sadanandan, K.R.; Podsiadlowski, L.; Gerth, M.; Weigert, A.; McGuire, J.A.; Mudge, J.; Edwards, S.V.; et al. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. *Nat. Commun.* **2016**, *7*, 11396. [[CrossRef](#)] [[PubMed](#)]
257. Gilbert, C.; Schaack, S.; Pace, J.K., II; Brindley, P.J.; Feschotte, C. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* **2010**, *464*, 1347–1350. [[CrossRef](#)] [[PubMed](#)]
258. Tang, Z.; Zhang, H.H.; Huang, K.; Zhang, X.G.; Han, M.J.; Zhang, Z. Repeated horizontal transfers of four DNA transposons in invertebrates and bats. *Mob. DNA* **2015**, *6*, 3. [[CrossRef](#)] [[PubMed](#)]
259. Oliveira, S.G.; Bao, W.; Martins, C.; Jurka, J. Horizontal transfers of Mariner transposons between mammals and insects. *Mob. DNA* **2012**, *3*, 14. [[CrossRef](#)] [[PubMed](#)]
260. Dupeyron, M.; Leclercq, S.; Cerveau, N.; Bouchon, D.; Gilbert, C. Horizontal transfer of transposons between and within crustaceans and insects. *Mob. DNA* **2014**, *5*, 4. [[CrossRef](#)] [[PubMed](#)]
261. Carareto, C.M. Tropical Africa as a cradle for horizontal transfers of transposable elements between species of the genera *Drosophila* and *Zaprionus*. *Mob. Genet. Elem.* **2011**, *1*, 179–186. [[CrossRef](#)] [[PubMed](#)]
262. Ivancevic, A.M.; Walsh, A.M.; Kortschak, R.D.; Adelson, D.L. Jumping the fine LINE between species: Horizontal transfer of transposable elements in animals catalyses genome evolution. *Bioessays* **2013**, *35*, 1071–1082. [[CrossRef](#)] [[PubMed](#)]
263. Kordis, D.; Gubensek, F. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10704–10709. [[CrossRef](#)] [[PubMed](#)]
264. Novick, P.; Smith, J.; Ray, D.; Boissinot, S. Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. *Gene* **2010**, *449*, 85–94. [[CrossRef](#)] [[PubMed](#)]
265. Pace, J.K., II; Gilbert, C.; Clark, M.S.; Feschotte, C. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17023–17028. [[CrossRef](#)] [[PubMed](#)]
266. Lin, X.; Faridi, N.; Casola, C. An Ancient Transkingdom Horizontal Transfer of Penelope-Like Retroelements from Arthropods to Conifers. *Genome Biol. Evol.* **2016**, *8*, 1252–1266. [[PubMed](#)]
267. Roulin, A.; Piegu, B.; Fortune, P.M.; Sabot, F.; D’Hont, A.; Manicacci, D.; Panaud, O. Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae. *BMC Evol. Biol.* **2009**, *9*, 58. [[CrossRef](#)] [[PubMed](#)]
268. El Baidouri, M.; Carpentier, M.C.; Cooke, R.; Gao, D.; Lasserre, E.; Llauro, C.; Mirouze, M.; Picault, N.; Jackson, S.A.; Panaud, O. Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Res.* **2014**, *24*, 831. [[CrossRef](#)] [[PubMed](#)]
269. Thomas, J.; Schaack, S.; Pritham, E.J. Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol. Evol.* **2010**, *2*, 656–664. [[CrossRef](#)] [[PubMed](#)]
270. Zhang, H.H.; Feschotte, C.; Han, M.J.; Zhang, Z. Recurrent horizontal transfers of Chapaev transposons in diverse invertebrate and vertebrate animals. *Genome Biol. Evol.* **2014**, *6*, 1375–1386. [[CrossRef](#)] [[PubMed](#)]
271. Gilbert, C.; Chateigner, A.; Ernenwein, L.; Barbe, V.; Bezier, A.; Herniou, E.A.; Cordaux, R. Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nat. Commun.* **2014**, *5*, 3348. [[CrossRef](#)] [[PubMed](#)]

272. Gontier, N. (Ed.) *Reticulate Evolution: Symbiogenesis, Lateral Gene Transfer, Hybridization and Infectious Heredity*; Springer International Publishing AG: Cham, Switzerland, 2015.
273. Doolittle, R.F. The multiplicity of domains in proteins. *Annu. Rev. Biochem.* **1995**, *64*, 287–314. [[CrossRef](#)] [[PubMed](#)]
274. Bornberg-Bauer, E.; Beaussart, F.; Kummerfeld, S.K.; Teichmann, S.A.; Weiner, J., III. The evolution of domain arrangements in proteins and interaction networks. *Cell. Mol. Life Sci.* **2005**, *62*, 435–445. [[CrossRef](#)] [[PubMed](#)]
275. Wang, M.; Caetano-Anolles, G. The evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. *Structure* **2009**, *17*, 66–78. [[CrossRef](#)] [[PubMed](#)]
276. Itoh, M.; Nacher, J.C.; Kuma, K.; Goto, S.; Kanehisa, M. Evolutionary history and functional implications of protein domains and their combinations in eukaryotes. *Genome Biol.* **2007**, *8*, R121. [[CrossRef](#)] [[PubMed](#)]
277. Del Sol, A.; Carbonell, P. The modular organization of domain structures: Insights into protein-protein binding. *PLoS Comput. Biol.* **2007**, *3*, e239. [[CrossRef](#)] [[PubMed](#)]
278. Xia, K.; Fu, Z.; Hou, L.; Han, J.D. Impacts of protein-protein interaction domains on organism and network complexity. *Genome Res.* **2008**, *18*, 1500–1508. [[CrossRef](#)] [[PubMed](#)]
279. Schuster-Bockler, B.; Bateman, A. Reuse of structural domain-domain interactions in protein networks. *BMC Bioinform.* **2007**, *8*, 259. [[CrossRef](#)] [[PubMed](#)]
280. Peisajovich, S.G.; Garbarino, J.E.; Wei, P.; Lim, W.A. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science* **2010**, *328*, 368–372. [[CrossRef](#)] [[PubMed](#)]
281. Kuriyan, J.; Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 259–288. [[CrossRef](#)] [[PubMed](#)]
282. Mayer, B.J. The discovery of modular binding domains: Building blocks of cell signalling. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 691–698. [[CrossRef](#)] [[PubMed](#)]
283. Bhattacharyya, R.P.; Remenyi, A.; Yeh, B.J.; Lim, W.A. Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **2006**, *75*, 655–680. [[CrossRef](#)] [[PubMed](#)]
284. Yang, S.; Doolittle, R.F.; Bourne, P.E. Phylogeny determined by protein domain content. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 373–378. [[CrossRef](#)] [[PubMed](#)]
285. Bateman, A.; Coin, L.; Durbin, R.; Finn, R.D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E.L.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2004**, *32*, D138–D141. [[CrossRef](#)] [[PubMed](#)]
286. Gonzalez, M.W.; Pearson, W.R. RefProtDom: A Protein Database with Improved Domain Boundaries and Homology Relationships. *Bioinformatics* **2010**, *26*, 2361–2362. [[CrossRef](#)] [[PubMed](#)]
287. Levitt, M. Nature of the protein universe. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11079–11084. [[CrossRef](#)] [[PubMed](#)]
288. Yang, S.; Bourne, P.E. The evolutionary history of protein domains viewed by species phylogeny. *PLoS ONE* **2009**, *4*, e8378. [[CrossRef](#)] [[PubMed](#)]
289. Nasir, A.; Kim, K.M.; Caetano-Anolles, G. Global patterns of protein domain gain and loss in superkingdoms. *PLoS Comput. Biol.* **2014**, *10*, e1003452. [[CrossRef](#)] [[PubMed](#)]
290. Buljan, M.; Bateman, A. The evolution of protein domain families. *Biochem. Soc. Trans.* **2009**, *37*, 751–755. [[CrossRef](#)] [[PubMed](#)]
291. Eichler, E.E. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.* **2001**, *17*, 661–669. [[CrossRef](#)]
292. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)] [[PubMed](#)]
293. Ekman, D.; Bjorklund, A.K.; Elofsson, A. Quantification of the elevated rate of domain rearrangements in metazoa. *J. Mol. Biol.* **2007**, *372*, 1337–1348. [[CrossRef](#)] [[PubMed](#)]
294. Kawashima, T.; Kawashima, S.; Tanaka, C.; Murai, M.; Yoneda, M.; Putnam, N.H.; Rokhsar, D.S.; Kanehisa, M.; Satoh, N.; Wada, H. Domain shuffling and the evolution of vertebrates. *Genome Res.* **2009**, *19*, 1393–1403. [[CrossRef](#)] [[PubMed](#)]
295. Zhang, X.C.; Wang, Z.; Zhang, X.; Le, M.H.; Sun, J.; Xu, D.; Cheng, J.; Stacey, G. Evolutionary dynamics of protein domain architecture in plants. *BMC Evol. Biol.* **2012**, *12*, 6. [[CrossRef](#)] [[PubMed](#)]

296. Chan, C.X.; Darling, A.E.; Beiko, R.G.; Ragan, M.A. Are protein domains modules of lateral genetic transfer? *PLoS ONE* **2009**, *4*, e4524. [[CrossRef](#)] [[PubMed](#)]
297. Basu, M.K.; Poliakov, E.; Rogozin, I.B. Domain mobility in proteins: Functional and evolutionary implications. *Brief. Bioinform.* **2009**, *10*, 205–216. [[CrossRef](#)] [[PubMed](#)]
298. Moore, A.D.; Bjorklund, A.K.; Ekman, D.; Bornberg-Bauer, E.; Elofsson, A. Arrangements in the modular evolution of proteins. *Trends Biochem. Sci.* **2008**, *33*, 444–451. [[CrossRef](#)] [[PubMed](#)]
299. Zhang, Q.; Zmasek, C.M.; Godzik, A. Domain architecture evolution of pattern-recognition receptors. *Immunogenetics* **2010**, *62*, 263–272. [[CrossRef](#)] [[PubMed](#)]
300. Wu, D.; Rastinejad, F. Structural characterization of mammalian bHLH-PAS transcription factors. *Curr. Opin. Struct. Biol.* **2016**, *43*, 1–9. [[CrossRef](#)] [[PubMed](#)]
301. Bjorklund, A.K.; Ekman, D.; Elofsson, A. Expansion of protein domain repeats. *PLoS Comput. Biol.* **2006**, *2*, e114. [[CrossRef](#)] [[PubMed](#)]
302. Jernigan, K.K.; Bordenstein, S.R. Tandem-repeat protein domains across the tree of life. *PeerJ* **2015**, *3*, e732. [[CrossRef](#)] [[PubMed](#)]
303. Franca, G.S.; Cancherini, D.V.; de Souza, S.J. Evolutionary history of exon shuffling. *Genetica* **2012**, *140*, 249–257. [[CrossRef](#)] [[PubMed](#)]
304. Gilbert, W. Why genes in pieces? *Nature* **1978**, *271*, 501. [[CrossRef](#)] [[PubMed](#)]
305. Patthy, L. Modular assembly of genes and the evolution of new functions. *Genetica* **2003**, *118*, 217–231. [[CrossRef](#)] [[PubMed](#)]
306. Sharp, P.A. Split genes and RNA splicing. *Cell* **1994**, *77*, 805–815. [[CrossRef](#)]
307. Liu, M.; Grigoriev, A. Protein domains correlate strongly with exons in multiple eukaryotic genomes—evidence of exon shuffling? *Trends Genet.* **2004**, *20*, 399–403. [[CrossRef](#)] [[PubMed](#)]
308. Kornblihtt, A.R. Coupling transcription and alternative splicing. *Adv. Exp. Med. Biol.* **2007**, *623*, 175–189. [[PubMed](#)]
309. Xing, Y.; Lee, C. Alternative splicing and RNA selection pressure—evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.* **2006**, *7*, 499–509. [[CrossRef](#)] [[PubMed](#)]
310. Koralewski, T.E.; Krutovsky, K.V. Evolution of exon-intron structure and alternative splicing. *PLoS ONE* **2011**, *6*, e18055. [[CrossRef](#)] [[PubMed](#)]
311. Yang, X.; Coulombe-Huntington, J.; Kang, S.; Sheynkman, G.M.; Hao, T.; Richardson, A.; Sun, S.; Yang, F.; Shen, Y.A.; Murray, R.R.; et al. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **2016**, *164*, 805–817. [[CrossRef](#)] [[PubMed](#)]
312. Mastrangelo, A.M.; Marone, D.; Laido, G.; De Leonardis, A.M.; De Vita, P. Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* **2012**, *185–186*, 40–49. [[CrossRef](#)] [[PubMed](#)]
313. Seo, P.J.; Park, M.J.; Park, C.M. Alternative splicing of transcription factors in plant responses to low temperature stress: Mechanisms and functions. *Planta* **2013**, *237*, 1415–1424. [[CrossRef](#)] [[PubMed](#)]
314. Wang, H.; You, C.; Chang, F.; Wang, Y.; Wang, L.; Qi, J.; Ma, H. Alternative splicing during Arabidopsis flower development results in constitutive and stage-regulated isoforms. *Front. Genet.* **2014**, *5*, 25. [[CrossRef](#)] [[PubMed](#)]
315. Staiger, D.; Brown, J.W. Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* **2013**, *25*, 3640–3656. [[CrossRef](#)] [[PubMed](#)]
316. Su, Z.; Gu, X. Revisit on the evolutionary relationship between alternative splicing and gene duplication. *Gene* **2012**. [[CrossRef](#)] [[PubMed](#)]
317. Bush, S.J.; Chen, L.; Tovar-Corona, J.M.; Urrutia, A.O. Alternative splicing and the evolution of phenotypic novelty. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*. [[CrossRef](#)] [[PubMed](#)]
318. Varagona, M.J.; Purugganan, M.; Wessler, S.R. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* **1992**, *4*, 811–820. [[CrossRef](#)] [[PubMed](#)]
319. Lev-Maor, G.; Ram, O.; Kim, E.; Sela, N.; Goren, A.; Levanon, E.Y.; Ast, G. Intronic Alus influence alternative splicing. *PLoS Genet.* **2008**, *4*, e1000204. [[CrossRef](#)] [[PubMed](#)]
320. Tarrio, R.; Ayala, F.J.; Rodriguez-Trelles, F. Alternative splicing: A missing piece in the puzzle of intron gain. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7223–7228. [[CrossRef](#)] [[PubMed](#)]
321. Zhou, R.; Moshgabadi, N.; Adams, K.L. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16122–16127. [[CrossRef](#)] [[PubMed](#)]

322. Zhang, P.G.; Huang, S.Z.; Pin, A.L.; Adams, K.L. Extensive divergence in alternative splicing patterns after gene and genome duplication during the evolutionary history of Arabidopsis. *Mol. Biol. Evol.* **2010**, *27*, 1686–1697. [[CrossRef](#)] [[PubMed](#)]
323. Grassi, L.; Fusco, D.; Sellerio, A.; Cora, D.; Bassetti, B.; Caselle, M.; Lagomarsino, M.C. Identity and divergence of protein domain architectures after the yeast whole-genome duplication event. *Mol. Biosyst.* **2010**, *6*, 2305–2315. [[CrossRef](#)] [[PubMed](#)]
324. Wu, Y.C.; Rasmussen, M.D.; Kellis, M. Evolution at the subgene level: Domain rearrangements in the Drosophila phylogeny. *Mol. Biol. Evol.* **2012**, *29*, 689–705. [[CrossRef](#)] [[PubMed](#)]
325. Carretero-Paulet, L.; Fares, M.A. Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. *Mol. Biol. Evol.* **2012**, *29*, 3541–3551. [[CrossRef](#)] [[PubMed](#)]
326. Kim, E.; Magen, A.; Ast, G. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* **2007**, *35*, 125–131. [[CrossRef](#)] [[PubMed](#)]
327. Hoffmann, F.G.; Opazo, J.C.; Storz, J.F. Whole-Genome Duplications Spurred the Functional Diversification of the Globin Gene Superfamily in Vertebrates. *Mol. Biol. Evol.* **2011**, *29*, 303–312. [[CrossRef](#)] [[PubMed](#)]
328. Hade, M.D.; Kaur, J.; Chakraborti, P.K.; Dikshit, K.L. Multidomain truncated hemoglobins: New members of the globin family exhibiting tandem repeats of globin units and domain fusion. *IUBMB Life* **2017**, *69*, 479–488. [[CrossRef](#)] [[PubMed](#)]
329. Henry, C.S.; Lerma-Ortiz, C.; Gerdes, S.Y.; Mullen, J.D.; Colasanti, R.; Zhukov, A.; Frelin, O.; Thiaville, J.J.; Zallot, R.; Niehaus, T.D.; et al. Systematic identification and analysis of frequent gene fusion events in metabolic pathways. *BMC Genom.* **2016**, *17*, 473. [[CrossRef](#)] [[PubMed](#)]
330. Durrens, P.; Nikolski, M.; Sherman, D. Fusion and fission of genes define a metric between fungal genomes. *PLoS Comput. Biol.* **2008**, *4*, e1000200. [[CrossRef](#)] [[PubMed](#)]
331. Rogers, R.L.; Hartl, D.L. Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. *Mol. Biol. Evol.* **2012**, *29*, 517–529. [[CrossRef](#)] [[PubMed](#)]
332. Marsh, J.A.; Teichmann, S.A. How do proteins gain new domains? *Genome Biol.* **2010**, *11*, 126. [[CrossRef](#)] [[PubMed](#)]
333. Cromar, G.; Wong, K.C.; Loughran, N.; On, T.; Song, H.; Xiong, X.; Zhang, Z.; Parkinson, J. New tricks for “old” domains: How novel architectures and promiscuous hubs contributed to the organization and evolution of the ECM. *Genome Biol. Evol.* **2014**, *6*, 2897–2917. [[CrossRef](#)] [[PubMed](#)]
334. Bailey, J.A.; Eichler, E.E. Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* **2006**, *7*, 552–564. [[CrossRef](#)] [[PubMed](#)]
335. Kaessmann, H.; Zollner, S.; Nekrutenko, A.; Li, W.H. Signatures of domain shuffling in the human genome. *Genome Res.* **2002**, *12*, 1642–1650. [[CrossRef](#)] [[PubMed](#)]
336. Jachiet, P.A.; Colson, P.; Lopez, P.; Baptiste, E. Extensive gene remodeling in the viral world: New evidence for nongradual evolution in the mobilome network. *Genome Biol. Evol.* **2014**, *6*, 2195–2205. [[CrossRef](#)] [[PubMed](#)]
337. Abroi, A. A protein domain-based view of the virosphere-host relationship. *Biochimie* **2015**, *119*, 231–243. [[CrossRef](#)] [[PubMed](#)]
338. Arguello, J.R.; Fan, C.; Wang, W.; Long, M. Origination of chimeric genes through DNA-level recombination. *Genome Dyn.* **2007**, *3*, 131–146. [[PubMed](#)]
339. Yang, S.; Arguello, J.R.; Li, X.; Ding, Y.; Zhou, Q.; Chen, Y.; Zhang, Y.; Zhao, R.; Brunet, F.; Peng, L.; et al. Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLoS Genet.* **2008**, *4*, e3. [[CrossRef](#)] [[PubMed](#)]
340. Van Rijk, A.; Bloemendal, H. Molecular mechanisms of exon shuffling: Illegitimate recombination. *Genetica* **2003**, *118*, 245–249. [[CrossRef](#)] [[PubMed](#)]
341. Bennardo, N.; Cheng, A.; Huang, N.; Stark, J.M. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet.* **2008**, *4*, e1000110. [[CrossRef](#)] [[PubMed](#)]
342. Lieber, M.R.; Gu, J.; Lu, H.; Shimazaki, N.; Tsai, A.G. Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. *Subcell. Biochem.* **2010**, *50*, 279–296. [[PubMed](#)]
343. Fu, B.; Chen, M.; Zou, M.; Long, M.; He, S. The rapid generation of chimerical genes expanding protein diversity in zebrafish. *BMC Genom.* **2010**, *11*, 657. [[CrossRef](#)] [[PubMed](#)]

344. Damert, A.; Raiz, J.; Horn, A.V.; Lower, J.; Wang, H.; Xing, J.; Batzer, M.A.; Lower, R.; Schumann, G.G. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res.* **2009**, *19*, 1992–2008. [[CrossRef](#)] [[PubMed](#)]
345. Gladyshev, E.A.; Arkipova, I.R. A widespread class of reverse transcriptase-related cellular genes. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20311–20316. [[CrossRef](#)] [[PubMed](#)]
346. Kazazian, H.H., Jr. Genetics. L1 retrotransposons shape the mammalian genome. *Science* **2000**, *289*, 1152–1153. [[CrossRef](#)] [[PubMed](#)]
347. Garcia-Perez, J.L.; Doucet, A.J.; Bucheton, A.; Moran, J.V.; Gilbert, N. Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res.* **2007**, *17*, 602–611. [[CrossRef](#)] [[PubMed](#)]
348. Beck, C.R.; Collier, P.; Macfarlane, C.; Malig, M.; Kidd, J.M.; Eichler, E.E.; Badge, R.M.; Moran, J.V. LINE-1 retrotransposition activity in human genomes. *Cell* **2010**, *141*, 1159–1170. [[CrossRef](#)] [[PubMed](#)]
349. Richardson, S.R.; Doucet, A.J.; Kopera, H.C.; Moldovan, J.B.; Garcia-Perez, J.L.; Moran, J.V. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiol. Spectr.* **2015**, *3*. [[CrossRef](#)] [[PubMed](#)]
350. Moran, J.V.; DeBerardinis, R.J.; Kazazian, H.H., Jr. Exon shuffling by L1 retrotransposition. *Science* **1999**, *283*, 1530–1534. [[CrossRef](#)] [[PubMed](#)]
351. Song, M.; Boissinot, S. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* **2007**, *390*, 206–213. [[CrossRef](#)] [[PubMed](#)]
352. Tubio, J.M.; Li, Y.; Ju, Y.S.; Martincorena, I.; Cooke, S.L.; Tojo, M.; Gundem, G.; Pipinikas, C.P.; Zamora, J.; Raine, K.; et al. Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **2014**, *345*, 1251343. [[CrossRef](#)] [[PubMed](#)]
353. Richardson, S.R.; Salvador-Palomeque, C.; Faulkner, G.J. Diversity through duplication: Whole-genome sequencing reveals novel gene retrocopies in the human population. *Bioessays* **2014**, *36*, 475–481. [[CrossRef](#)] [[PubMed](#)]
354. Cell-encoded reverse transcriptases capable of copying RNA into cDNA are found in both prokaryotes and eukaryotes [345]. In eukaryotes, LINE retrotransposons encode the biochemical activities necessary for reverse transcription and genome insertion of all the classes of nuclear RNA-templated cDNAs [346–349]. Moreover, mobilization of both 3' and 5' adjacent exons by both LINE and SINE retrotransposons (“retrotransduction”) are well-documented phenomena [344,350,351]. “Extensive” LINE-mediated retrotransduction has been observed to occur in real time in cancer cells [352], and somatic retrocopies have been detected in tissue samples from the human population [353].
355. Zou, M.; Wang, G.; He, S. Evolutionary patterns of RNA-based gene duplicates in *Caenorhabditis* nematodes coincide with their genomic features. *BMC Res. Notes* **2012**, *5*, 398. [[CrossRef](#)] [[PubMed](#)]
356. Jakalski, M.; Takeshita, K.; Deblieck, M.; Koyanagi, K.O.; Makalowska, I.; Watanabe, H.; Makalowski, W. Comparative genomic analysis of retrogene repertoire in two green algae *Volvox carteri* and *Chlamydomonas reinhardtii*. *Biol. Direct* **2016**, *11*, 35. [[CrossRef](#)] [[PubMed](#)]
357. Wang, W.; Zheng, H.; Fan, C.; Li, J.; Shi, J.; Cai, Z.; Zhang, G.; Liu, D.; Zhang, J.; Vang, S.; et al. High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* **2006**, *18*, 1791–1802. [[CrossRef](#)] [[PubMed](#)]
358. Wang, J.; Long, M.; Vibranovski, M.D. Retrogenes moved out of the z chromosome in the silkworm. *J. Mol. Evol.* **2012**, *74*, 113–126. [[CrossRef](#)] [[PubMed](#)]
359. Chen, M.; Zou, M.; Fu, B.; Li, X.; Vibranovski, M.D.; Gan, X.; Wang, D.; Wang, W.; Long, M.; He, S. Evolutionary Patterns of RNA-Based Duplication in Non-Mammalian Chordates. *PLoS ONE* **2011**, *6*, e21466. [[CrossRef](#)] [[PubMed](#)]
360. Zhong, Z.; Yang, L.; Zhang, Y.E.; Xue, Y.; He, S. Correlated expression of retrocopies and parental genes in zebrafish. *Mol. Genet. Genom.* **2016**, *291*, 723–737. [[CrossRef](#)] [[PubMed](#)]
361. Carelli, F.N.; Hayakawa, T.; Go, Y.; Imai, H.; Warnefors, M.; Kaessmann, H. The life history of retrocopies illuminates the evolution of new mammalian genes. *Genome Res.* **2016**, *26*, 301–314. [[CrossRef](#)] [[PubMed](#)]
362. Vinckenbosch, N.; Dupanloup, I.; Kaessmann, H. Evolutionary fate of retroposed gene copies in the human genome. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3220–3225. [[CrossRef](#)] [[PubMed](#)]
363. Baertsch, R.; Diekhans, M.; Kent, W.J.; Haussler, D.; Brosius, J. Retrocopy contributions to the evolution of the human genome. *BMC Genom.* **2008**, *9*, 466. [[CrossRef](#)] [[PubMed](#)]

364. McLysaght, A. Evolutionary steps of sex chromosomes are reflected in retrogenes. *Trends Genet.* **2008**, *24*, 478–481. [[CrossRef](#)] [[PubMed](#)]
365. Pan, D.; Zhang, L. Burst of young retrogenes and independent retrogene formation in mammals. *PLoS ONE* **2009**, *4*, e5040. [[CrossRef](#)] [[PubMed](#)]
366. Marques, A.C.; Dupanloup, I.; Vinckenbosch, N.; Reymond, A.; Kaessmann, H. Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol.* **2005**, *3*, e357. [[CrossRef](#)] [[PubMed](#)]
367. Navarro, F.C.; Galante, P.A. A Genome-Wide Landscape of Retrocopies in Primate Genomes. *Genome Biol. Evol.* **2015**, *7*, 2265–2275. [[CrossRef](#)] [[PubMed](#)]
368. Yang, Y.; Walsh, C.E. Spliceosome-mediated RNA trans-splicing. *Mol. Ther.* **2005**, *12*, 1006–1012. [[CrossRef](#)] [[PubMed](#)]
369. Lasda, E.L.; Blumenthal, T. Trans-splicing. *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 417–434. [[CrossRef](#)] [[PubMed](#)]
370. Douris, V.; Telford, M.J.; Averof, M. Evidence for multiple independent origins of trans-splicing in Metazoa. *Mol. Biol. Evol.* **2010**, *27*, 684–693. [[CrossRef](#)] [[PubMed](#)]
371. Jividen, K.; Li, H. Chimeric RNAs generated by intergenic splicing in normal and cancer cells. *Genes Chromosomes Cancer* **2014**, *53*, 963–971. [[CrossRef](#)] [[PubMed](#)]
372. Okonechnikov, K.; Imai-Matsushima, A.; Paul, L.; Seitz, A.; Meyer, T.F.; Garcia-Alcalde, F. InFusion: Advancing Discovery of Fusion Genes and Chimeric Transcripts from Deep RNA-Sequencing Data. *PLoS ONE* **2016**, *11*, e0167417. [[CrossRef](#)] [[PubMed](#)]
373. Kozlov, A.P. Expression of evolutionarily novel genes in tumors. *Infect. Agent Cancer* **2016**, *11*, 34. [[CrossRef](#)] [[PubMed](#)]
374. Kumar-Sinha, C.; Kalyana-Sundaram, S.; Chinnaiyan, A.M. Landscape of gene fusions in epithelial cancers: Seq and ye shall find. *Genome Med.* **2015**, *7*, 129. [[CrossRef](#)] [[PubMed](#)]
375. Annala, M.J.; Parker, B.C.; Zhang, W.; Nykter, M. Fusion genes and their discovery using high throughput sequencing. *Cancer Lett.* **2013**, *340*, 192–200. [[CrossRef](#)] [[PubMed](#)]
376. Jia, Y.; Xie, Z.; Li, H. Intergenically Spliced Chimeric RNAs in Cancer. *Trends Cancer* **2016**, *2*, 475–484. [[CrossRef](#)] [[PubMed](#)]
377. Seki, Y.; Mizukami, T.; Kohno, T. Molecular Process Producing Oncogene Fusion in Lung Cancer Cells by Illegitimate Repair of DNA Double-Strand Breaks. *Biomolecules* **2015**, *5*, 2464–2476. [[CrossRef](#)] [[PubMed](#)]
378. Lawson, A.R.; Hindley, G.F.; Forshew, T.; Tatevossian, R.G.; Jamie, G.A.; Kelly, G.P.; Neale, G.A.; Ma, J.; Jones, T.A.; Ellison, D.W.; et al. RAF gene fusion breakpoints in pediatric brain tumors are characterized by significant enrichment of sequence microhomology. *Genome Res.* **2011**, *21*, 505–514. [[CrossRef](#)] [[PubMed](#)]
379. Advanced sequencing techniques document the abundance of recombinant RNA molecules in cellular transcriptomes [371,372]. The generation of DNA and RNA encoding novel chimeric protein fusions occurs commonly in real time in tumor cells [373–375]. RNA trans-splicing appears to be involved in some cases [371,376], while microhomologies, DS breaks and NHEJ may be responsible in others [377,378].
380. Frenkel-Morgenstern, M.; Valencia, A. Novel domain combinations in proteins encoded by chimeric transcripts. *Bioinformatics* **2012**, *28*, i67–i74. [[CrossRef](#)] [[PubMed](#)]
381. Long, M.; Langley, C.H. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* **1993**, *260*, 91–95. [[CrossRef](#)] [[PubMed](#)]
382. Lei, Q.; Li, C.; Zuo, Z.; Huang, C.; Cheng, H.; Zhou, R. Evolutionary Insights into RNA trans-Splicing in Vertebrates. *Genome Biol. Evol.* **2016**, *8*, 562–577. [[CrossRef](#)] [[PubMed](#)]
383. Fischer, D.; Eisenberg, D. Finding families for genomic ORFans. *Bioinformatics* **1999**, *15*, 759–762. [[CrossRef](#)] [[PubMed](#)]
384. Tautz, D.; Domazet-Loso, T. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **2011**, *12*, 692–702. [[CrossRef](#)] [[PubMed](#)]
385. Siew, N.; Azaria, Y.; Fischer, D. The ORFanage: An ORFan database. *Nucleic Acids Res.* **2004**, *32*, D281–D283. [[CrossRef](#)] [[PubMed](#)]
386. Yin, Y.; Fischer, D. On the origin of microbial ORFans: Quantifying the strength of the evidence for viral lateral transfer. *BMC Evol. Biol.* **2006**, *6*, 63. [[CrossRef](#)] [[PubMed](#)]
387. Sabath, N.; Wagner, A.; Karlin, D. Evolution of viral proteins originated de novo by overprinting. *Mol. Biol. Evol.* **2012**, *29*, 3767–3780. [[CrossRef](#)] [[PubMed](#)]
388. Murphy, D.N.; McLysaght, A. De novo origin of protein-coding genes in murine rodents. *PLoS ONE* **2012**, *7*, e48650. [[CrossRef](#)] [[PubMed](#)]

389. Andreatta, M.E.; Levine, J.A.; Foy, S.G.; Guzman, L.D.; Kosinski, L.J.; Cordes, M.H.; Masel, J. The Recent De Novo Origin of Protein C-Termini. *Genome Biol. Evol.* **2015**, *7*, 1686–1701. [[CrossRef](#)] [[PubMed](#)]
390. Cai, J.; Zhao, R.; Jiang, H.; Wang, W. De novo origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* **2008**, *179*, 487–496. [[CrossRef](#)] [[PubMed](#)]
391. Xie, C.; Zhang, Y.E.; Chen, J.Y.; Liu, C.J.; Zhou, W.Z.; Li, Y.; Zhang, M.; Zhang, R.; Wei, L.; Li, C.Y. Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. *PLoS Genet.* **2012**, *8*, e1002942. [[CrossRef](#)] [[PubMed](#)]
392. Toll-Riera, M.; Castelo, R.; Bellora, N.; Alba, M.M. Evolution of primate orphan proteins. *Biochem. Soc. Trans.* **2009**, *37*, 778–782. [[CrossRef](#)] [[PubMed](#)]
393. Casola, C.; Lawing, A.M.; Betran, E.; Feschotte, C. PIF-like transposons are common in drosophila and have been repeatedly domesticated to generate new host genes. *Mol. Biol. Evol.* **2007**, *24*, 1872–1888. [[CrossRef](#)] [[PubMed](#)]
394. Philippsen, G.S.; Avaca-Crusca, J.S.; Araujo, A.P.; DeMarco, R. Distribution patterns and impact of transposable elements in genes of green algae. *Gene* **2016**, *594*, 151–159. [[CrossRef](#)] [[PubMed](#)]
395. Zhou, Q.; Zhang, G.; Zhang, Y.; Xu, S.; Zhao, R.; Zhan, Z.; Li, X.; Ding, Y.; Yang, S.; Wang, W. On the origin of new genes in *Drosophila*. *Genome Res.* **2008**, *18*, 1446–1455. [[CrossRef](#)] [[PubMed](#)]
396. Kojima, K.K.; Okada, N. mRNA retrotransposition coupled with 5' inversion as a possible source of new genes. *Mol. Biol. Evol.* **2009**, *26*, 1405–1420. [[CrossRef](#)] [[PubMed](#)]
397. Reinhardt, J.A.; Wanjiru, B.M.; Brant, A.T.; Saelao, P.; Begun, D.J.; Jones, C.D. De novo ORFs in *Drosophila* are important to organismal fitness and evolved rapidly from previously non-coding sequences. *PLoS Genet.* **2013**, *9*, e1003860. [[CrossRef](#)] [[PubMed](#)]
398. Ruiz-Orera, J.; Messeguer, X.; Subirana, J.A.; Alba, M.M. Long non-coding RNAs as a source of new peptides. *eLife* **2014**, *3*, e03523. [[CrossRef](#)] [[PubMed](#)]
399. Nekrutenko, A.; Li, W.H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **2001**, *17*, 619–621. [[CrossRef](#)]
400. Britten, R.J. Transposable elements have contributed to thousands of human proteins. *Proc. Nat. Acad. Sci. USA* **2006**, *103*, 1798–1803. [[CrossRef](#)] [[PubMed](#)]
401. Kaer, K.; Branovets, J.; Hallikma, A.; Nigumann, P.; Speek, M. Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. *PLoS ONE* **2011**, *6*, e26099. [[CrossRef](#)] [[PubMed](#)]
402. Vorechovsky, I. Transposable elements in disease-associated cryptic exons. *Hum. Genet.* **2010**, *127*, 135–154. [[CrossRef](#)] [[PubMed](#)]
403. Krull, M.; Brosius, J.; Schmitz, J. Alu-SINE exonization: En route to protein-coding function. *Mol. Biol. Evol.* **2005**, *22*, 1702–1711. [[CrossRef](#)] [[PubMed](#)]
404. Moller-Krull, M.; Zemann, A.; Roos, C.; Brosius, J.; Schmitz, J. Beyond DNA: RNA editing and steps toward Alu exonization in primates. *J. Mol. Biol.* **2008**, *382*, 601–609. [[CrossRef](#)] [[PubMed](#)]
405. Sela, N.; Mersch, B.; Hotz-Wagenblatt, A.; Ast, G. Characteristics of transposable element exonization within human and mouse. *PLoS ONE* **2010**, *5*, e10907. [[CrossRef](#)] [[PubMed](#)]
406. Lev-Maor, G.; Sorek, R.; Levanon, E.Y.; Paz, N.; Eisenberg, E.; Ast, G. RNA-editing-mediated exon evolution. *Genome Biol.* **2007**, *8*, R29. [[CrossRef](#)] [[PubMed](#)]
407. Huang, C.R.; Schneider, A.M.; Lu, Y.; Niranjana, T.; Shen, P.; Robinson, M.A.; Steranka, J.P.; Valle, D.; Civin, C.I.; Wang, T.; et al. Mobile interspersed repeats are major structural variants in the human genome. *Cell* **2010**, *141*, 1171–1182. [[CrossRef](#)] [[PubMed](#)]
408. De Koning, A.P.; Gu, W.; Castoe, T.A.; Batzer, M.A.; Pollock, D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **2011**, *7*, e1002384. [[CrossRef](#)] [[PubMed](#)]
409. Cordaux, R.; Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **2009**, *10*, 691–703. [[CrossRef](#)] [[PubMed](#)]
410. Dridi, S. Alu mobile elements: From junk DNA to genomic gems. *Scientifica (Cairo)* **2012**, *2012*, 545328. [[PubMed](#)]
411. Ponicsan, S.L.; Kugel, J.F.; Goodrich, J.A. Genomic gems: SINE RNAs regulate mRNA production. *Curr. Opin. Genet. Dev.* **2010**, *20*, 149–155. [[CrossRef](#)] [[PubMed](#)]
412. Huxley, J. *Evolution: The Modern Synthesis*; Allen & Unwin: London, UK, 1942.

413. Avery, O.T.; MacLeod, C.M.; McCarty, M. Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *J. Exp. Med.* **1944**, *79*, 137–158. [[CrossRef](#)] [[PubMed](#)]
414. Watson, J.D.; Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737–738. [[CrossRef](#)] [[PubMed](#)]
415. Jacob, F.; Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **1961**, *3*, 318–356. [[CrossRef](#)]
416. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 344–355. [[CrossRef](#)] [[PubMed](#)]
417. Watson, J.D.; Crick, F.H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **1953**, *171*, 964–967. [[CrossRef](#)] [[PubMed](#)]
418. Collado-Vides, J. A linguistic representation of the regulation of transcription initiation. I. An ordered array of complex symbols with distinctive features. *Biosystems* **1993**, *29*, 87–104. [[CrossRef](#)]
419. Andersson, R.; Gebhard, C.; Miguel-Escalada, I.; Hoof, I.; Bornholdt, J.; Boyd, M.; Chen, Y.; Zhao, X.; Schmidl, C.; Suzuki, T.; et al. An atlas of active enhancers across human cell types and tissues. *Nature* **2014**, *507*, 455–461. [[CrossRef](#)] [[PubMed](#)]
420. Bondos, S.E.; Tan, X.X. Combinatorial transcriptional regulation: The interaction of transcription factors and cell signaling molecules with homeodomain proteins in *Drosophila* development. *Crit. Rev. Eukaryot. Gene Expr.* **2001**, *11*, 145–171. [[CrossRef](#)] [[PubMed](#)]
421. Slattery, M.; Zhou, T.; Yang, L.; Dantas Machado, A.C.; Gordan, R.; Rohs, R. Absence of a simple code: How transcription factors read the genome. *Trends Biochem. Sci.* **2014**, *39*, 381–399. [[CrossRef](#)] [[PubMed](#)]
422. Allis, C.D.; Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **2016**, *17*, 487–500. [[CrossRef](#)] [[PubMed](#)]
423. Thurman, R.E.; Rynes, E.; Humbert, R.; Vierstra, J.; Maurano, M.T.; Haugen, E.; Sheffield, N.C.; Stergachis, A.B.; Wang, H.; Vernot, B.; et al. The accessible chromatin landscape of the human genome. *Nature* **2012**, *489*, 75–82. [[CrossRef](#)] [[PubMed](#)]
424. Dekker, J.; Rippe, K.; Dekker, M.; Kleckner, N. Capturing chromosome conformation. *Science* **2002**, *295*, 1306–1311. [[CrossRef](#)] [[PubMed](#)]
425. Cao, R.; Cheng, J. Deciphering the association between gene function and spatial gene-gene interactions in 3D human genome conformation. *BMC Genom.* **2015**, *16*, 880. [[CrossRef](#)] [[PubMed](#)]
426. Phillips-Cremins, J.E.; Sauria, M.E.; Sanyal, A.; Gerasimova, T.I.; Lajoie, B.R.; Bell, J.S.; Ong, C.T.; Hookway, T.A.; Guo, C.; Sun, Y.; et al. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **2013**, *153*, 1281–1295. [[CrossRef](#)] [[PubMed](#)]
427. Britten, R.; Kohne, D.E. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **1968**, *161*, 529–540. [[CrossRef](#)] [[PubMed](#)]
428. Britten, R.J.; Kohne, D.E. Repeated Segments of DNA. *Sci. Am.* **1970**, *222*, 24–31. [[CrossRef](#)] [[PubMed](#)]
429. Britten, R.J.; Davidson, E.H. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **1971**, *46*, 111–138. [[CrossRef](#)] [[PubMed](#)]
430. Kodama, Y.; Shumway, M.; Leinonen, R. The Sequence Read Archive: Explosive growth of sequencing data. *Nucleic Acids Res.* **2012**, *40*, D54–D56. [[CrossRef](#)] [[PubMed](#)]
431. Benson, D.A.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2015**, *43*, D30–D35. [[CrossRef](#)] [[PubMed](#)]
432. Gregory, T.R.; Nicol, J.A.; Tamm, H.; Kullman, B.; Kullman, K.; Leitch, I.J.; Murray, B.G.; Kapraun, D.F.; Greilhuber, J.; Bennett, M.D. Eukaryotic genome size databases. *Nucleic Acids Res.* **2007**, *35*, D332–D338. [[CrossRef](#)] [[PubMed](#)]
433. Hubley, R.; Finn, R.D.; Clements, J.; Eddy, S.R.; Jones, T.A.; Bao, W.; Smit, A.F.; Wheeler, T.J. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **2016**, *44*, D81–D89. [[CrossRef](#)] [[PubMed](#)]
434. Jurka, J.; Kapitonov, V.V.; Kohany, O.; Jurka, M.V. Repetitive sequences in complex genomes: Structure and evolution. *Annu. Rev. Genom. Hum. Genet.* **2007**, *8*, 241–259. [[CrossRef](#)] [[PubMed](#)]
435. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [[CrossRef](#)] [[PubMed](#)]

436. Sun, C.; Shepard, D.B.; Chong, R.A.; Lopez Arriaza, J.; Hall, K.; Castoe, T.A.; Feschotte, C.; Pollock, D.D.; Mueller, R.L. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **2012**, *4*, 168–183. [[CrossRef](#)] [[PubMed](#)]
437. Macas, J.; Neumann, P.; Navratilova, A. Repetitive DNA in the pea (*Pisum sativum* L.) genome: Comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genom.* **2007**, *8*, 427. [[CrossRef](#)] [[PubMed](#)]
438. Xia, E.H.; Zhang, H.B.; Sheng, J.; Li, K.; Zhang, Q.J.; Kim, C.; Zhang, Y.; Liu, Y.; Zhu, T.; Li, W.; et al. The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis. *Mol. Plant* **2017**, *10*, 866–877. [[CrossRef](#)] [[PubMed](#)]
439. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A.; et al. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **2009**, *326*, 1112–1115. [[CrossRef](#)] [[PubMed](#)]
440. There is by no means a strict correlation of repetitive DNA content with either genome size or taxonomic status. Birds (1–1.2 GB total DNA) are relatively low in repeats (6–12%), while the more primitive vertebrate lamprey (0.65 GB total DNA) has 36% repeats, and a chordate lancelet (0.48 GB) has 13.6%. Insects (0.16–0.26 GB) range from 6–28% repeats, and a relatively simple tunicate (*C. intestinalis*, 0.14 GB total DNA) has over 16% repeats. Among vertebrates, certain groups are exceptional in genome content. Salamander genomes are very large and range from ~14 to ~120 Gb total DNA, but, like other vertebrates, these giant genomes contain 24–47% interspersed repetitive mobile DNA [436]. Plants have relatively large genomes, comparable in size to those of mammals (2–3 GB total DNA), generally with comparable repetitive DNA contents (~40% in peas [437]). But there are exceptions among plant genomes. Among the exceptions are the recently sequenced 3.02-GB genome of the cultivated tea tree *Camellia sinensis* (80.9% repetitive DNA) [438] and the 2.3 GB Maize strain B73 genome (>84% mobile DNA elements) [439].
441. Ohno, S. *So Much “Junk DNA” in Our Genome*; Smith, H.H., Ed.; Gordon and Breach: New York, NY, USA, 1972; pp. 366–370.
442. Orgel, L.E.; Crick, F.H. Selfish DNA: The ultimate parasite. *Nature* **1980**, *284*, 604–607. [[CrossRef](#)] [[PubMed](#)]
443. Pennisi, E. ENCODE project writes eulogy for junk DNA. *Science* **2012**, *337*, 1159. [[CrossRef](#)] [[PubMed](#)]
444. Buchler, N.E.; Gerland, U.; Hwa, T. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5136–5141. [[CrossRef](#)] [[PubMed](#)]
445. Zinzen, R.P.; Girardot, C.; Gagneur, J.; Braun, M.; Furlong, E.E. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **2009**, *462*, 65–70. [[CrossRef](#)] [[PubMed](#)]
446. Guet, C.C.; Elowitz, M.B.; Hsing, W.; Leibler, S. Combinatorial synthesis of genetic networks. *Science* **2002**, *296*, 1466–1470. [[CrossRef](#)] [[PubMed](#)]
447. Shapiro, J.A.; von Sternberg, R. Why repetitive DNA is essential to genome function. *Biol. Revs. Camb. Philos. Soc.* **2005**, *80*, 227–250. [[CrossRef](#)]
448. Shapiro, J.A. A 21st century view of evolution: Genome system architecture, repetitive DNA, and natural genetic engineering. *Gene* **2005**, *345*, 91–100. [[CrossRef](#)] [[PubMed](#)]
449. Peterson, S.N.; Bailey, C.C.; Jensen, J.S.; Borre, M.B.; King, E.S.; Bott, K.F.; Hutchison, C.A., III. Characterization of repetitive DNA in the *Mycoplasma genitalium* genome: Possible role in the generation of antigenic variation. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 11829–11833. [[CrossRef](#)] [[PubMed](#)]
450. Taft, R.J.; Pheasant, M.; Mattick, J.S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **2007**, *29*, 288–299. [[CrossRef](#)] [[PubMed](#)]
451. Liu, G.; Mattick, J.S.; Taft, R.J. A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle* **2013**, *12*, 2061–2072. [[CrossRef](#)] [[PubMed](#)]
452. McClintock, B. Induction of Instability at Selected Loci in Maize. *Genetics* **1953**, *38*, 579–599. [[PubMed](#)]
453. McClintock, B. *Discovery and Characterization of Transposable Elements: The Collected Papers of Barbara McClintock*; Garland: New York, NY, USA, 1987.
454. McClintock, B. Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **1952**, *21*, 197–216. [[CrossRef](#)]
455. McClintock, B. Some parallels between gene control systems in maize and in bacteria. *Am. Nat.* **1961**, *95*, 265–277. [[CrossRef](#)]
456. McClintock, B. Intranuclear systems controlling gene action and mutation. *Brookhaven Symp. Biol.* **1956**, 58–74.

457. Marino-Ramirez, L.; Jordan, I.K. Transposable element derived DNaseI-hypersensitive sites in the human genome. *Biol. Direct* **2006**, *1*, 20. [[CrossRef](#)] [[PubMed](#)]
458. Jordan, I.K.; Rogozin, I.B.; Glazko, G.V.; Koonin, E.V. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* **2003**, *19*, 68–72. [[CrossRef](#)]
459. Glinsky, G.V. Transposable Elements and DNA Methylation Create in Embryonic Stem Cells Human-Specific Regulatory Sequences Associated with Distal Enhancers and Noncoding RNAs. *Genome Biol. Evol.* **2015**, *7*, 1432–1454. [[CrossRef](#)] [[PubMed](#)]
460. Kelley, D.R.; Hendrickson, D.G.; Tenen, D.; Rinn, J.L. Transposable elements modulate human RNA abundance and splicing via specific RNA-protein interactions. *Genome Biol.* **2014**, *15*, 537. [[CrossRef](#)] [[PubMed](#)]
461. Huda, A.; Bushel, P.R. Widespread Exonization of Transposable Elements in Human Coding Sequences is Associated with Epigenetic Regulation of Transcription. *Transcr. Open Access* **2013**, *1*. [[CrossRef](#)]
462. Kunarso, G.; Chia, N.Y.; Jeyakani, J.; Hwang, C.; Lu, X.; Chan, Y.S.; Ng, H.H.; Bourque, G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **2010**, *42*, 631–634. [[CrossRef](#)] [[PubMed](#)]
463. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [[CrossRef](#)] [[PubMed](#)]
464. Carbone, L.; Harris, R.A.; Gnerre, S.; Veeramah, K.R.; Lorente-Galdos, B.; Huddleston, J.; Meyer, T.J.; Herrero, J.; Roos, C.; Aken, B.; et al. Gibbon genome and the fast karyotype evolution of small apes. *Nature* **2014**, *513*, 195–201. [[CrossRef](#)] [[PubMed](#)]
465. Hollister, J.D.; Smith, L.M.; Guo, Y.L.; Ott, F.; Weigel, D.; Gaut, B.S. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2322–2327. [[CrossRef](#)] [[PubMed](#)]
466. Symonova, R.; Majtanova, Z.; Sember, A.; Staaks, G.B.; Bohlen, J.; Freyhof, J.; Rabova, M.; Rab, P. Genome differentiation in a species pair of coregonine fishes: An extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol. Biol.* **2013**, *13*, 42. [[CrossRef](#)] [[PubMed](#)]
467. Santolamazza, F.; Mancini, E.; Simard, F.; Qi, Y.; Tu, Z.; della Torre, A. Insertion polymorphisms of SINE200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. *Malar. J.* **2008**, *7*, 163. [[CrossRef](#)] [[PubMed](#)]
468. Craddock, E.M. Profuse evolutionary diversification and speciation on volcanic islands: Transposon instability and amplification bursts explain the genetic paradox. *Biol. Direct* **2016**, *11*, 44. [[CrossRef](#)] [[PubMed](#)]
469. Kazazian, H.H., Jr. Mobile elements: Drivers of genome evolution. *Science* **2004**, *303*, 1626–1632. [[CrossRef](#)] [[PubMed](#)]
470. Deininger, P.L.; Moran, J.V.; Batzer, M.A.; Kazazian, H.H., Jr. Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* **2003**, *13*, 651–658. [[CrossRef](#)] [[PubMed](#)]
471. Lippman, Z.; Gendrel, A.V.; Black, M.; Vaughn, M.W.; Dedhia, N.; McCombie, W.R.; Lavine, K.; Mittal, V.; May, B.; Kasschau, K.D.; et al. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **2004**, *430*, 471–476. [[CrossRef](#)] [[PubMed](#)]
472. Roman, A.C.; Gonzalez-Rico, F.J.; Fernandez-Salguero, P.M. B1-SINE retrotransposons: Establishing genomic insulatory networks. *Mob. Genet. Elem.* **2011**, *1*, 66–70. [[CrossRef](#)] [[PubMed](#)]
473. Xiao, X.; Wang, Z.; Jang, M.; Burge, C.B. Coevolutionary networks of splicing cis-regulatory elements. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18583–18588. [[CrossRef](#)] [[PubMed](#)]
474. 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. [[CrossRef](#)] [[PubMed](#)]
475. Guerreiro, M.P. Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in *Drosophila*. *Mob. Genet. Elem.* **2014**, *4*, e34394. [[CrossRef](#)] [[PubMed](#)]
476. 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. [[CrossRef](#)] [[PubMed](#)]
477. Gerdes, P.; Richardson, S.R.; Mager, D.L.; Faulkner, G.J. Transposable elements in the mammalian embryo: Pioneers surviving through stealth and service. *Genome Biol.* **2016**, *17*, 100. [[CrossRef](#)] [[PubMed](#)]

478. Goke, J.; Ng, H.H. CTRL+INSERT: Retrotransposons and their contribution to regulation and innovation of the transcriptome. *EMBO Rep.* **2016**, *17*, 1131–1144. [[CrossRef](#)] [[PubMed](#)]
479. Thompson, P.J.; Macfarlan, T.S.; Lorincz, M.C. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Mol. Cell* **2016**, *62*, 766–776. [[CrossRef](#)] [[PubMed](#)]
480. Jacques, P.E.; Jeyakani, J.; Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **2013**, *9*, e1003504. [[CrossRef](#)] [[PubMed](#)]
481. Garazha, A.; Ivanova, A.; Suntsova, M.; Malakhova, G.; Roumiantsev, S.; Zhavoronkov, A.; Buzdin, A. New bioinformatic tool for quick identification of functionally relevant endogenous retroviral inserts in human genome. *Cell Cycle* **2015**, *14*, 1476–1484. [[CrossRef](#)] [[PubMed](#)]
482. “A total of 4094 candidate HSRL (human-specific regulatory loci) display selective and site-specific binding of critical regulators (NANOG {Nanog homeobox}, POU5F1 {POU class 5 homeobox 1}, CCCTC-binding factor {CTCF}, Lamin B1), and are preferentially located within the matrix of transcriptionally active DNA segments that are hypermethylated in hESC. hESC-specific NANOG-binding sites are enriched near the protein-coding genes regulating brain size, pluripotency long noncoding RNAs, hESC enhancers, and 5-hydroxymethylcytosine-harboring regions immediately adjacent to binding sites. Sequences of only 4.3% of hESC-specific NANOG-binding sites are present in Neanderthals’ genome, suggesting that a majority of these regulatory elements emerged in Modern Humans. Comparisons of estimated creation rates of novel TF-binding sites revealed that there was a 49.7-fold acceleration of creation rates of NANOG-binding sites in genomes of Chimpanzees when compared with the mouse genomes and further 5.7-fold acceleration in genomes of Modern Humans compared with the Chimpanzees genomes” [[459](#)].
483. Muramoto, H.; Yagi, S.; Hirabayashi, K.; Sato, S.; Ohgane, J.; Tanaka, S.; Shiota, K. Enrichment of short interspersed transposable elements to embryonic stem cell-specific hypomethylated gene regions. *Genes Cells* **2010**, *15*, 855–865. [[CrossRef](#)] [[PubMed](#)]
484. Castro-Diaz, N.; Ecco, G.; Coluccio, A.; Kapopoulou, A.; Yazdanpanah, B.; Friedli, M.; Duc, J.; Jang, S.M.; Turelli, P.; Trono, D. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **2014**, *28*, 1397–1409. [[CrossRef](#)] [[PubMed](#)]
485. Robbez-Masson, L.; Rowe, H.M. Retrotransposons shape species-specific embryonic stem cell gene expression. *Retrovirology* **2015**, *12*, 45. [[CrossRef](#)] [[PubMed](#)]
486. Sundaram, V.; Choudhary, M.N.; Pehrsson, E.; Xing, X.; Fiore, C.; Pandey, M.; Maricque, B.; Udawatta, M.; Ngo, D.; Chen, Y.; et al. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat. Commun.* **2017**, *8*, 14550. [[CrossRef](#)] [[PubMed](#)]
487. Friedli, M.; Turelli, P.; Kapopoulou, A.; Rauwel, B.; Castro-Diaz, N.; Rowe, H.M.; Ecco, G.; Unzu, C.; Planet, E.; Lombardo, A.; et al. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res.* **2014**, *24*, 1251–1259. [[CrossRef](#)] [[PubMed](#)]
488. Ge, S.X. Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development. *BMC Genom.* **2017**, *18*, 200. [[CrossRef](#)] [[PubMed](#)]
489. Goke, J.; Lu, X.; Chan, Y.S.; Ng, H.H.; Ly, L.H.; Sachs, F.; Szczerbinska, I. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **2015**, *16*, 135–141. [[CrossRef](#)] [[PubMed](#)]
490. Dupressoir, A.; Lavalie, C.; Heidmann, T. From ancestral infectious retroviruses to bona fide cellular genes: Role of the captured syncytins in placentation. *Placenta* **2012**, *33*, 663–671. [[CrossRef](#)] [[PubMed](#)]
491. Macaulay, E.C.; Weeks, R.J.; Andrews, S.; Morison, I.M. Hypomethylation of functional retrotransposon-derived genes in the human placenta. *Mamm. Genome* **2011**, *22*, 722–735. [[CrossRef](#)] [[PubMed](#)]
492. Chuong, E.B.; Rumi, M.A.; Soares, M.J.; Baker, J.C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **2013**, *45*, 325–329. [[CrossRef](#)] [[PubMed](#)]
493. Lynch, V.J.; Leclerc, R.D.; May, G.; Wagner, G.P. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **2011**, *43*, 1154–1159. [[CrossRef](#)] [[PubMed](#)]
494. Lynch, V.J.; Nnamani, M.C.; Kapusta, A.; Brayer, K.; Plaza, S.L.; Mazur, E.C.; Emera, D.; Sheikh, S.Z.; Grutzner, F.; Bauersachs, S.; et al. Ancient Transposable Elements Transformed the Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian Pregnancy. *Cell Rep.* **2015**, *10*, 551–561. [[CrossRef](#)] [[PubMed](#)]
495. Emera, D.; Wagner, G.P. Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11246–11251. [[CrossRef](#)] [[PubMed](#)]

496. Emera, D.; Casola, C.; Lynch, V.J.; Wildman, D.E.; Agnew, D.; Wagner, G.P. Convergent evolution of endometrial prolactin expression in primates, mice, and elephants through the independent recruitment of transposable elements. *Mol. Biol. Evol.* **2012**, *29*, 239–247. [[CrossRef](#)] [[PubMed](#)]
497. Sasaki, T.; Nishihara, H.; Hirakawa, M.; Fujimura, K.; Tanaka, M.; Kokubo, N.; Kimura-Yoshida, C.; Matsuo, I.; Sumiyama, K.; Saitou, N.; et al. Possible involvement of SINEs in mammalian-specific brain formation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4220–4225. [[CrossRef](#)] [[PubMed](#)]
498. Notwell, J.H.; Chung, T.; Heavner, W.; Bejerano, G. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat. Commun.* **2015**, *6*, 6644. [[CrossRef](#)] [[PubMed](#)]
499. Emera, D.; Yin, J.; Reilly, S.K.; Gockley, J.; Noonan, J.P. Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2617–E2626. [[CrossRef](#)] [[PubMed](#)]
500. Crepaldi, L.; Policarpi, C.; Coatti, A.; Sherlock, W.T.; Jongbloets, B.C.; Down, T.A.; Riccio, A. Binding of TFIIIC to sine elements controls the relocation of activity-dependent neuronal genes to transcription factories. *PLoS Genet.* **2013**, *9*, e1003699. [[CrossRef](#)] [[PubMed](#)]
501. Alexander, J.M.; Lomvardas, S. Nuclear architecture as an epigenetic regulator of neural development and function. *Neuroscience* **2014**, *264*, 39–50. [[CrossRef](#)] [[PubMed](#)]
502. Under normal circumstances, acetylation of SINE heterochromatin is inhibited by the transcription factor TFIIIC. TFIIIC binding to SINE elements is well-documented and contributes to their role as epigenetic insulator elements [472]. Silencing of TFIIIC expression in non-stimulated neurons mimics NEE exposure and chronic neural depolarization, leads to uncontrolled SINE acetylation and transcription, and induces a dramatic increase in dendrite length and branching [500].
503. Watson, L.A.; Tsai, L.H. In the loop: How chromatin topology links genome structure to function in mechanisms underlying learning and memory. *Curr. Opin. NeuroBiol.* **2017**, *43*, 48–55. [[CrossRef](#)] [[PubMed](#)]
504. Kim, S.; Kaang, B.K. Epigenetic regulation and chromatin remodeling in learning and memory. *Exp. Mol. Med.* **2017**, *49*, e281. [[CrossRef](#)] [[PubMed](#)]
505. Thomas, C.A.; Paquola, A.C.; Muotri, A.R. LINE-1 retrotransposition in the nervous system. *Annu. Rev. Cell Dev. Biol.* **2012**, *28*, 555–573. [[CrossRef](#)] [[PubMed](#)]
506. Kapitonov, V.V.; Koonin, E.V. Evolution of the RAG1-RAG2 locus: Both proteins came from the same transposon. *Biol. Direct* **2015**, *10*, 20. [[CrossRef](#)] [[PubMed](#)]
507. Huang, S.; Tao, X.; Yuan, S.; Zhang, Y.; Li, P.; Beilinson, H.A.; Zhang, Y.; Yu, W.; Pontarotti, P.; Escrava, H.; et al. Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* **2016**, *166*, 102–114. [[CrossRef](#)] [[PubMed](#)]
508. Zak, D.E.; Aderem, A. Systems biology of innate immunity. *Immunol. Rev.* **2009**, *227*, 264–282. [[CrossRef](#)] [[PubMed](#)]
509. Nurnberger, T.; Brunner, F.; Kemmerling, B.; Piater, L. Innate immunity in plants and animals: Striking similarities and obvious differences. *Immunol. Rev.* **2004**, *198*, 249–266. [[CrossRef](#)] [[PubMed](#)]
510. Platanias, L.C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **2005**, *5*, 375–386. [[CrossRef](#)] [[PubMed](#)]
511. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **2016**, *351*, 1083–1087. [[CrossRef](#)] [[PubMed](#)]
512. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.
513. Ecker, J.R.; Bickmore, W.A.; Barroso, I.; Pritchard, J.K.; Gilad, Y.; Segal, E. Genomics: ENCODE explained. *Nature* **2012**, *489*, 52–55. [[CrossRef](#)] [[PubMed](#)]
514. Birney, E. The making of ENCODE: Lessons for big-data projects. *Nature* **2012**, *489*, 49–51. [[CrossRef](#)] [[PubMed](#)]
515. Skipper, M.; Dhand, R.; Campbell, P. Presenting ENCODE. *Nature* **2012**, *489*, 45. [[CrossRef](#)] [[PubMed](#)]
516. Glinsky, G.V. Phenotype-defining functions of multiple non-coding RNA pathways. *Cell Cycle* **2008**, *7*, 1630–1639. [[CrossRef](#)] [[PubMed](#)]
517. Mattick, J.S. Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Rep.* **2001**, *2*, 986–991. [[CrossRef](#)] [[PubMed](#)]
518. Birney, E.; Stamatoyannopoulos, J.A.; Dutta, A.; Guigo, R.; Gingeras, T.R.; Margulies, E.H.; Weng, Z.; Snyder, M.; Dermitzakis, E.T.; Thurman, R.E.; et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **2007**, *447*, 799–816. [[CrossRef](#)] [[PubMed](#)]

519. Pennisi, E. DNA study forces rethink of what it means to be a gene. *Science* **2007**, *316*, 1556–1557. [[CrossRef](#)] [[PubMed](#)]
520. Kapranov, P.; Cheng, J.; Dike, S.; Nix, D.A.; Duttagupta, R.; Willingham, A.T.; Stadler, P.F.; Hertel, J.; Hackermuller, J.; Hofacker, I.L.; et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **2007**, *316*, 1484–1488. [[CrossRef](#)] [[PubMed](#)]
521. Djebali, S.; Davis, C.A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; et al. Landscape of transcription in human cells. *Nature* **2012**, *489*, 101–108. [[CrossRef](#)] [[PubMed](#)]
522. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **2012**, *22*, 1775–1789. [[CrossRef](#)] [[PubMed](#)]
523. Cheng, C.; Alexander, R.; Min, R.; Leng, J.; Yip, K.Y.; Rozowsky, J.; Yan, K.K.; Dong, X.; Djebali, S.; Ruan, Y.; et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res.* **2012**, *22*, 1658–1667. [[CrossRef](#)] [[PubMed](#)]
524. Guo, Y.; Alexander, K.; Clark, A.G.; Grimson, A.; Yu, H. Integrated network analysis reveals distinct regulatory roles of transcription factors and microRNAs. *RNA* **2016**, *22*, 1663–1672. [[CrossRef](#)] [[PubMed](#)]
525. He, L.; Hannon, G.J. MicroRNAs: Small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **2004**, *5*, 522–531. [[CrossRef](#)] [[PubMed](#)]
526. Bartel, D.P. MicroRNAs: Target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233. [[CrossRef](#)] [[PubMed](#)]
527. Chen, T.; Hao, Y.J.; Zhang, Y.; Li, M.M.; Wang, M.; Han, W.; Wu, Y.; Lv, Y.; Hao, J.; Wang, L.; et al. m⁶A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell Stem Cell* **2015**, *16*, 289–301. [[CrossRef](#)] [[PubMed](#)]
528. Chen, C.Z.; Li, L.; Lodish, H.F.; Bartel, D.P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **2004**, *303*, 83–86. [[CrossRef](#)] [[PubMed](#)]
529. Chen, T.; Xiang, J.F.; Zhu, S.; Chen, S.; Yin, Q.F.; Zhang, X.O.; Zhang, J.; Feng, H.; Dong, R.; Li, X.J.; et al. ADAR1 is required for differentiation and neural induction by regulating microRNA processing in a catalytically independent manner. *Cell Res.* **2015**, *25*, 459–476. [[CrossRef](#)] [[PubMed](#)]
530. Sun, A.X.; Crabtree, G.R.; Yoo, A.S. MicroRNAs: Regulators of neuronal fate. *Curr. Opin. Cell Biol.* **2013**, *25*, 215–221. [[CrossRef](#)] [[PubMed](#)]
531. Wouters, M.D.; van Gent, D.C.; Hoeijmakers, J.H.; Pothof, J. MicroRNAs, the DNA damage response and cancer. *Mutat. Res.* **2011**, *717*, 54–66. [[CrossRef](#)] [[PubMed](#)]
532. Kruszka, K.; Pacak, A.; Swida-Barteczka, A.; Nuc, P.; Alaba, S.; Wroblewska, Z.; Karlowski, W.; Jarmolowski, A.; Szweykowska-Kulinska, Z. Transcriptionally and post-transcriptionally regulated microRNAs in heat stress response in barley. *J. Exp. Bot.* **2014**, *65*, 6123–6135. [[CrossRef](#)] [[PubMed](#)]
533. Li, F.; Pignatta, D.; Bendix, C.; Brunkard, J.O.; Cohn, M.M.; Tung, J.; Sun, H.; Kumar, P.; Baker, B. MicroRNA regulation of plant innate immune receptors. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1790–1795. [[CrossRef](#)] [[PubMed](#)]
534. Ripoll, J.J.; Bailey, L.J.; Mai, Q.-A.; Wu, S.L.; Hon, C.T.; Chapman, E.J.; Ditta, G.S.; Estelle, M.; Yanofsky, M.F. microRNA regulation of fruit growth. *Nat. Plants* **2015**, *1*, 15036. [[CrossRef](#)] [[PubMed](#)]
535. Boon, R.A.; Vickers, K.C. Intercellular transport of microRNAs. *Arterioscler. Thromb. Vasc. Biol.* **2013**, *33*, 186–192. [[CrossRef](#)] [[PubMed](#)]
536. Ajit, S.K. Circulating microRNAs as biomarkers, therapeutic targets, and signaling molecules. *Sensors* **2012**, *12*, 3359–3369. [[CrossRef](#)] [[PubMed](#)]
537. Fernandez-Messina, L.; Gutierrez-Vazquez, C.; Rivas-Garcia, E.; Sanchez-Madrid, F.; de la Fuente, H. Immunomodulatory role of microRNAs transferred by extracellular vesicles. *Biol. Cell* **2015**, *107*, 61–77. [[CrossRef](#)] [[PubMed](#)]
538. Khajuria, C.; Williams, C.E.; El Bouhssini, M.; Whitworth, R.J.; Richards, S.; Stuart, J.J.; Chen, M.S. Deep sequencing and genome-wide analysis reveals the expansion of MicroRNA genes in the gall midge *Mayetiola destructor*. *BMC Genom.* **2013**, *14*, 187. [[CrossRef](#)] [[PubMed](#)]
539. Londin, E.; Loher, P.; Telonis, A.G.; Quann, K.; Clark, P.; Jing, Y.; Hatzimichael, E.; Kirino, Y.; Honda, S.; Lally, M.; et al. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1106–E1115. [[CrossRef](#)] [[PubMed](#)]

540. Du, Z.Q.; Yang, C.X.; Rothschild, M.F.; Ross, J.W. Novel microRNA families expanded in the human genome. *BMC Genom.* **2013**, *14*, 98. [[CrossRef](#)] [[PubMed](#)]
541. Taylor, R.S.; Tarver, J.E.; Hiscock, S.J.; Donoghue, P.C. Evolutionary history of plant microRNAs. *Trends Plant. Sci.* **2014**, *19*, 175–182. [[CrossRef](#)] [[PubMed](#)]
542. Loh, Y.H.; Yi, S.V.; Streelman, J.T. Evolution of microRNAs and the diversification of species. *Genome Biol. Evol.* **2011**, *3*, 55–65. [[CrossRef](#)] [[PubMed](#)]
543. Berezikov, E. Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.* **2011**, *12*, 846–860. [[CrossRef](#)] [[PubMed](#)]
544. Li, K.; Wang, L.; Knisbacher, B.A.; Xu, Q.; Levanon, E.Y.; Wang, H.; Frenkel-Morgenstern, M.; Tagore, S.; Fang, X.; Bazak, L.; et al. Transcriptome, genetic editing, and microRNA divergence substantiate sympatric speciation of blind mole rat, Spalax. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7584–7589. [[CrossRef](#)] [[PubMed](#)]
545. Gim, J.A.; Ha, H.S.; Ahn, K.; Kim, D.S.; Kim, H.S. Genome-Wide Identification and Classification of MicroRNAs Derived from Repetitive Elements. *Genom. Inform.* **2014**, *12*, 261–267. [[CrossRef](#)] [[PubMed](#)]
546. Piriyaopongsa, J.; Marino-Ramirez, L.; Jordan, I.K. Origin and evolution of human microRNAs from transposable elements. *Genetics* **2007**, *176*, 1323–1337. [[CrossRef](#)] [[PubMed](#)]
547. Qin, S.; Jin, P.; Zhou, X.; Chen, L.; Ma, F. The Role of Transposable Elements in the Origin and Evolution of MicroRNAs in Human. *PLoS ONE* **2015**, *10*, e0131365. [[CrossRef](#)] [[PubMed](#)]
548. Roberts, J.T.; Cardin, S.E.; Borchert, G.M. Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences. *Mob. Genet. Elem.* **2014**, *4*, e29255. [[CrossRef](#)] [[PubMed](#)]
549. Platt, R.N., II; Vandeweghe, M.W.; Kern, C.; Schmidt, C.J.; Hoffmann, F.G.; Ray, D.A. Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. *Mol. Biol. Evol.* **2014**, *31*, 1536–1545. [[CrossRef](#)] [[PubMed](#)]
550. Spengler, R.M.; Oakley, C.K.; Davidson, B.L. Functional microRNAs and target sites are created by lineage-specific transposition. *Hum. Mol. Genet.* **2014**, *23*, 1783–1793. [[CrossRef](#)] [[PubMed](#)]
551. Devor, E.J.; Peek, A.S.; Lanier, W.; Samollow, P.B. Marsupial-specific microRNAs evolved from marsupial-specific transposable elements. *Gene* **2009**, *448*, 187–191. [[CrossRef](#)] [[PubMed](#)]
552. Hoffmann, F.G.; McGuire, L.P.; Counterman, B.A.; Ray, D.A. Transposable elements and small RNAs: Genomic fuel for species diversity. *Mob. Genet. Elem.* **2015**, *5*, 63–66. [[CrossRef](#)] [[PubMed](#)]
553. Smalheiser, N.R.; Torvik, V.I. Alu elements within human mRNAs are probable microRNA targets. *Trends Genet.* **2006**, *22*, 532–536. [[CrossRef](#)] [[PubMed](#)]
554. Wang, X.; Ai, G.; Zhang, C.; Cui, L.; Wang, J.; Li, H.; Zhang, J.; Ye, Z. Expression and diversification analysis reveals transposable elements play important roles in the origin of Lycopersicon-specific lncRNAs in tomato. *New Phytol.* **2016**, *209*, 1442–1455. [[CrossRef](#)] [[PubMed](#)]
555. Kapusta, A.; Feschotte, C. Volatile evolution of long noncoding RNA repertoires: Mechanisms and biological implications. *Trends Genet.* **2014**, *30*, 439–452. [[CrossRef](#)] [[PubMed](#)]
556. Hutchins, A.P.; Pei, D. Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs. *Sci. Bull.* **2015**, *60*, 1722–1733. [[CrossRef](#)] [[PubMed](#)]
557. Kannan, S.; Chernikova, D.; Rogozin, I.B.; Poliakov, E.; Managadze, D.; Koonin, E.V.; Milanese, L. Transposable Element Insertions in Long Intergenic Non-Coding RNA Genes. *Front. Bioeng. Biotechnol.* **2015**, *3*, 71. [[CrossRef](#)] [[PubMed](#)]
558. Kelley, D.; Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **2012**, *13*, R107. [[CrossRef](#)] [[PubMed](#)]
559. St Laurent, G., III; Shtokalo, D.; Dong, B.; Tackett, M.R.; Fan, X.; Lazorthes, S.; Nicolas, E.; Sang, N.; Triche, T.J.; McCaffrey, T.A.; et al. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol.* **2013**, *14*, R73. [[CrossRef](#)] [[PubMed](#)]
560. Durruthy-Durruthy, J.; Sebastiano, V.; Wossidlo, M.; Cepeda, D.; Cui, J.; Grow, E.J.; Davila, J.; Mall, M.; Wong, W.H.; Wysocka, J.; et al. The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. *Nat. Genet.* **2016**, *48*, 44–52. [[CrossRef](#)] [[PubMed](#)]
561. Fort, A.; Hashimoto, K.; Yamada, D.; Salimullah, M.; Keya, C.A.; Saxena, A.; Bonetti, A.; Voineagu, I.; Bertin, N.; Kratz, A.; et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* **2014**, *46*, 558–566. [[CrossRef](#)] [[PubMed](#)]

562. Zucchelli, S.; Fasolo, F.; Russo, R.; Cimatti, L.; Patrucco, L.; Takahashi, H.; Jones, M.H.; Santoro, C.; Sblattero, D.; Cotella, D.; et al. SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. *Front. Cell. NeuroSci.* **2015**, *9*, 174. [[CrossRef](#)] [[PubMed](#)]
563. Johnson, R.; Guigo, R. The RIDL hypothesis: Transposable elements as functional domains of long noncoding RNAs. *RNA* **2014**, *20*, 959–976. [[CrossRef](#)] [[PubMed](#)]
564. Hadjiargyrou, M.; Delihias, N. The Intertwining of Transposable Elements and Non-Coding RNAs. *Int. J. Mol. Sci.* **2013**, *14*, 13307–13328. [[CrossRef](#)] [[PubMed](#)]
565. Arthanari, Y.; Heintzen, C.; Griffiths-Jones, S.; Crosthwaite, S.K. Natural Antisense Transcripts and Long Non-Coding RNA in *Neurospora crassa*. *PLoS ONE* **2014**, *9*, e91353. [[CrossRef](#)] [[PubMed](#)]
566. Woehle, C.; Kusdian, G.; Radine, C.; Gaur, D.; Landan, G.; Gould, S.B. The parasite *Trichomonas vaginalis* expresses thousands of pseudogenes and long non-coding RNAs independently from functional neighbouring genes. *BMC Genom.* **2014**, *15*, 906. [[CrossRef](#)] [[PubMed](#)]
567. Jenkins, A.M.; Waterhouse, R.M.; Muskavitch, M.A. Long non-coding RNA discovery across the genus *Anopheles* reveals conserved secondary structures within and beyond the Gambiae complex. *BMC Genom.* **2015**, *16*, 337. [[CrossRef](#)] [[PubMed](#)]
568. Wu, Y.; Cheng, T.; Liu, C.; Liu, D.; Zhang, Q.; Long, R.; Zhao, P.; Xia, Q. Systematic Identification and Characterization of Long Non-Coding RNAs in the Silkworm, *Bombyx mori*. *PLoS ONE* **2016**, *11*, e0147147. [[CrossRef](#)] [[PubMed](#)]
569. Szczesniak, M.W.; Rosikiewicz, W.; Makalowska, I. CANTATAdb: A Collection of Plant Long Non-Coding RNAs. *Plant Cell Physiol.* **2016**, *57*, e8. [[CrossRef](#)] [[PubMed](#)]
570. Zhu, Q.H.; Wang, M.B. Molecular Functions of Long Non-Coding RNAs in Plants. *Genes* **2012**, *3*, 176–190. [[CrossRef](#)] [[PubMed](#)]
571. Gaiti, F.; Fernandez-Valverde, S.L.; Nakanishi, N.; Calcino, A.D.; Yanai, I.; Tanurdzic, M.; Degnan, B.M. Dynamic and Widespread lncRNA Expression in a Sponge and the Origin of Animal Complexity. *Mol. Biol. Evol.* **2015**, *32*, 2367–2382. [[CrossRef](#)] [[PubMed](#)]
572. Al-Tobasei, R.; Paneru, B.; Salem, M. Genome-Wide Discovery of Long Non-Coding RNAs in Rainbow Trout. *PLoS ONE* **2016**, *11*, e0148940. [[CrossRef](#)] [[PubMed](#)]
573. Ma, L.; Bajic, V.B.; Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol.* **2013**, *10*, 925–933. [[CrossRef](#)] [[PubMed](#)]
574. Hon, C.C.; Ramilowski, J.A.; Harshbarger, J.; Bertin, N.; Rackham, O.J.; Gough, J.; Denisenko, E.; Schmeier, S.; Poulsen, T.M.; Severin, J.; et al. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **2017**, *543*, 199–204. [[CrossRef](#)] [[PubMed](#)]
575. Ernst, C.; Morton, C.C. Identification and function of long non-coding RNA. *Front. Cell. NeuroSci.* **2013**, *7*, 168. [[CrossRef](#)] [[PubMed](#)]
576. St Laurent, G.; Wahlestedt, C.; Kapranov, P. The Landscape of long noncoding RNA classification. *Trends Genet.* **2015**, *31*, 239–251. [[CrossRef](#)] [[PubMed](#)]
577. Eldredge, N.; Gould, S.J. Punctuated equilibria: An alternative to phyletic gradualism. In *Models in Paleobiology*; Schopf, T.J.M., Ed.; Freeman, Cooper and Company: San Francisco, CA, USA, 1972; pp. 82–115.
578. Grant, B.R.; Grant, P.R. Fission and fusion of Darwin's finches populations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2008**, *363*, 2821–2829. [[CrossRef](#)] [[PubMed](#)]
579. Grant, P.; Grant, B.R. Hybridization of bird species. *Science* **1992**, *256*, 193–197. [[CrossRef](#)] [[PubMed](#)]
580. Grant, P.R.; Grant, B.R. *40 Years of Evolution. Darwin's Finches on Daphne Major Island*; Princeton University Press: Princeton, NJ, USA, 2014.
581. Romero-Soriano, V.; Burlet, N.; Vela, D.; Fontdevila, A.; Vieira, C.; Garcia Guerreiro, M.P. *Drosophila* Females Undergo Genome Expansion after Interspecific Hybridization. *Genome Biol. Evol.* **2016**, *8*, 556–561. [[CrossRef](#)] [[PubMed](#)]
582. Feiner, N. Accumulation of transposable elements in Hox gene clusters during adaptive radiation of *Anolis lizards*. *Proc. Biol. Sci.* **2016**, *283*. [[CrossRef](#)] [[PubMed](#)]
583. Oliver, K.R.; McComb, J.A.; Greene, W.K. Transposable elements: Powerful contributors to angiosperm evolution and diversity. *Genome Biol. Evol.* **2013**, *5*, 1886–1901. [[CrossRef](#)] [[PubMed](#)]
584. Lisch, D. How important are transposons for plant evolution? *Nat. Rev. Genet.* **2013**, *14*, 49–61. [[CrossRef](#)] [[PubMed](#)]

585. Parisod, C.; Salmon, A.; Zerjal, T.; Tenaillon, M.; Grandbastien, M.A.; Ainouche, M. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytol.* **2009**, *184*, 1003–1015. [[CrossRef](#)] [[PubMed](#)]
586. Stapley, J.; Santure, A.W.; Dennis, S.R. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Mol. Ecol.* **2015**, *24*, 2241–2252. [[CrossRef](#)] [[PubMed](#)]
587. Arnold, M.L.; Meyer, A. Natural hybridization in primates: One evolutionary mechanism. *Zoology* **2006**, *109*, 261–276. [[CrossRef](#)] [[PubMed](#)]
588. Oliver, K.R.; Greene, W.K. Mobile DNA and the TE-Thrust hypothesis: Supporting evidence from the primates. *Mob. DNA* **2011**, *2*, 8. [[CrossRef](#)] [[PubMed](#)]
589. Witkin, E.M. Radiation-induced mutations and their repair. *Science* **1966**, *152*, 1345–1353. [[CrossRef](#)] [[PubMed](#)]
590. Napolitano, R.; Janel-Bintz, R.; Wagner, J.; Fuchs, R.P. All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. *EMBO J.* **2000**, *19*, 6259–6265. [[CrossRef](#)] [[PubMed](#)]
591. Sutton, M.D.; Smith, B.T.; Godoy, V.G.; Walker, G.C. The SOS response: Recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu. Rev. Genet.* **2000**, *34*, 479–497. [[CrossRef](#)] [[PubMed](#)]
592. Goodman, M.F. Purposeful mutations. *Nature* **1998**, *395*, 221–223. [[CrossRef](#)] [[PubMed](#)]
593. Goodman, M. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Ann. Rev. Biochem.* **2002**, *71*, 17–50. [[CrossRef](#)] [[PubMed](#)]
594. Sakofsky, C.J.; Ayyar, S.; Deem, A.K.; Chung, W.H.; Ira, G.; Malkova, A. Translesion Polymerases Drive Microhomology-Mediated Break-Induced Replication Leading to Complex Chromosomal Rearrangements. *Mol. Cell* **2015**, *60*, 860–872. [[CrossRef](#)] [[PubMed](#)]
595. Lada, A.G.; Dhar, A.; Boissy, R.J.; Hirano, M.; Rubel, A.A.; Rogozin, I.B.; Pavlov, Y.I. AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biol. Direct* **2012**, *7*, 47. [[CrossRef](#)] [[PubMed](#)]
596. Taylor, B.J.; Nik-Zainal, S.; Wu, Y.L.; Stebbings, L.A.; Raine, K.; Campbell, P.J.; Rada, C.; Stratton, M.R.; Neuberger, M.S. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* **2013**, *2*, e00534. [[CrossRef](#)] [[PubMed](#)]
597. Jacobs, A.L.; Schar, P. DNA glycosylases: In DNA repair and beyond. *Chromosoma* **2012**, *121*, 1–20. [[CrossRef](#)] [[PubMed](#)]
598. The nucleotide substitution mutations that are induced by cytosine deaminase are not exclusively C-to-T transitions because some of the uracil bases that are produced by deamination are removed from the affected DNA strand by uracil-N-glycosylase activity and are replaced by any of the four DNA bases, producing other transition and transversion substitutions, or erasing mutagenesis altogether [597].
599. Casellas, R.; Basu, U.; Yewdell, W.T.; Chaudhuri, J.; Robbiani, D.F.; Di Noia, J.M. Mutations, kataegis and translocations in B cells: Understanding AID promiscuous activity. *Nat. Rev. Immunol.* **2016**, *16*, 164–176. [[CrossRef](#)] [[PubMed](#)]
600. Leibowitz, M.L.; Zhang, C.Z.; Pellman, D. Chromothripsis: A New Mechanism for Rapid Karyotype Evolution. *Annu. Rev. Genet.* **2015**, *49*, 183–211. [[CrossRef](#)] [[PubMed](#)]
601. Poot, M. Of Simple and Complex Genome Rearrangements, Chromothripsis, Chromoanasythesis, and Chromosome Chaos. *Mol. Syndromol.* **2017**, *8*, 115–117. [[CrossRef](#)] [[PubMed](#)]
602. Liu, P.; Erez, A.; Nagamani, S.C.; Dhar, S.U.; Kolodziejska, K.E.; Dharmadhikari, A.V.; Cooper, M.L.; Wiszniewska, J.; Zhang, F.; Withers, M.A.; et al. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* **2011**, *146*, 889–903. [[CrossRef](#)] [[PubMed](#)]
603. Zhang, C.Z.; Spektor, A.; Cornils, H.; Francis, J.M.; Jackson, E.K.; Liu, S.; Meyerson, M.; Pellman, D. Chromothripsis from DNA damage in micronuclei. *Nature* **2015**, *522*, 179–184. [[CrossRef](#)] [[PubMed](#)]
604. De Pagter, M.S.; van Roosmalen, M.J.; Baas, A.F.; Renkens, I.; Duran, K.J.; van Binsbergen, E.; Tavakoli-Yaraki, M.; Hochstenbach, R.; van der Veken, L.T.; Cuppen, E.; et al. Chromothripsis in Healthy Individuals Affects Multiple Protein-Coding Genes and Can Result in Severe Congenital Abnormalities in Offspring. *Am. J. Hum. Genet.* **2015**. [[CrossRef](#)]
605. Goodman, M.F. Better living with hyper-mutation. *Environ. Mol. Mutagen.* **2016**, *57*, 421–434. [[CrossRef](#)] [[PubMed](#)]
606. Teng, G.; Papavasiliou, F.N. Immunoglobulin somatic hypermutation. *Annu. Rev. Genet.* **2007**, *41*, 107–120. [[CrossRef](#)] [[PubMed](#)]

607. Lada, A.G.; Stepchenkova, E.I.; Waisertreiger, I.S.; Noskov, V.N.; Dhar, A.; Eudy, J.D.; Boissy, R.J.; Hirano, M.; Rogozin, I.B.; Pavlov, Y.I. Genome-wide mutation avalanches induced in diploid yeast cells by a base analog or an APOBEC deaminase. *PLoS Genet.* **2013**, *9*, e1003736. [[CrossRef](#)] [[PubMed](#)]
608. Lada, A.G.; Kliver, S.F.; Dhar, A.; Plev, D.E.; Masharsky, A.E.; Rogozin, I.B.; Pavlov, Y.I. Disruption of Transcriptional Coactivator Sub1 Leads to Genome-Wide Re-distribution of Clustered Mutations Induced by APOBEC in Active Yeast Genes. *PLoS Genet.* **2015**, *11*, e1005217. [[CrossRef](#)] [[PubMed](#)]
609. Saini, N.; Roberts, S.A.; Sterling, J.F.; Malc, E.P.; Mieczkowski, P.A.; Gordenin, D.A. APOBEC3B cytidine deaminase targets the non-transcribed strand of tRNA genes in yeast. *DNA Repair (Amst)* **2017**, *53*, 4–14. [[CrossRef](#)] [[PubMed](#)]
610. Pellestor, F.; Gatinois, V.; Puechberty, J.; Genevieve, D.; Lefort, G. Chromothripsis: Potential origin in gametogenesis and preimplantation cell divisions. A review. *Fertil. Steril.* **2014**, *102*, 1785–1796. [[CrossRef](#)] [[PubMed](#)]
611. Weckselblatt, B.; Hermetz, K.E.; Rudd, M.K. Unbalanced translocations arise from diverse mutational mechanisms including chromothripsis. *Genome Res.* **2015**, *25*, 937–947. [[CrossRef](#)] [[PubMed](#)]
612. Middelkamp, S.; van Heesch, S.; Braat, A.K.; de Ligt, J.; van Iterson, M.; Simonis, M.; van Roosmalen, M.J.; Kelder, M.J.; Kruisselbrink, E.; Hochstenbach, R.; et al. Molecular dissection of germline chromothripsis in a developmental context using patient-derived iPSCs. *Genome Med.* **2017**, *9*, 9. [[CrossRef](#)] [[PubMed](#)]
613. Zhang, C.Z.; Leibowitz, M.L.; Pellman, D. Chromothripsis and beyond: Rapid genome evolution from complex chromosomal rearrangements. *Genes Dev.* **2013**, *27*, 2513–2530. [[CrossRef](#)] [[PubMed](#)]
614. Fukami, M.; Shima, H.; Suzuki, E.; Ogata, T.; Matsubara, K.; Kamimaki, T. Catastrophic Cellular Events Leading to Complex Chromosomal Rearrangements in the Germline. *Clin. Genet.* **2016**, *91*, 653–660. [[CrossRef](#)] [[PubMed](#)]
615. Nazaryan-Petersen, L.; Bertelsen, B.; Bak, M.; Jonson, L.; Tommerup, N.; Hancks, D.C.; Tumer, Z. Germline Chromothripsis Driven by L1-Mediated Retrotransposition and Alu/Alu Homologous Recombination. *Hum. Mutat.* **2016**, *37*, 385–395. [[CrossRef](#)] [[PubMed](#)]
616. Farkash, E.A.; Kao, G.D.; Horman, S.R.; Prak, E.T. Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Res.* **2006**, *34*, 1196–1204. [[CrossRef](#)] [[PubMed](#)]
617. Stribinskis, V.; Ramos, K.S. Activation of human long interspersed nuclear element 1 retrotransposition by benzo(a)pyrene, an ubiquitous environmental carcinogen. *Cancer Res.* **2006**, *66*, 2616–2620. [[CrossRef](#)] [[PubMed](#)]
618. Teneng, I.; Stribinskis, V.; Ramos, K.S. Context-specific regulation of LINE-1. *Genes Cells* **2007**, *12*, 1101–1110. [[CrossRef](#)] [[PubMed](#)]
619. Kale, S.P.; Moore, L.; Deininger, P.L.; Roy-Engel, A.M. Heavy metals stimulate human LINE-1 retrotransposition. *Int. J. Environ. Res. Public Health* **2005**, *2*, 14–23. [[CrossRef](#)] [[PubMed](#)]
620. Kale, S.P.; Carmichael, M.C.; Harris, K.; Roy-Engel, A.M. The L1 retrotranspositional stimulation by particulate and soluble cadmium exposure is independent of the generation of DNA breaks. *Int. J. Environ. Res. Public Health* **2006**, *3*, 121–128. [[CrossRef](#)] [[PubMed](#)]
621. El-Sawy, M.; Kale, S.P.; Dugan, C.; Nguyen, T.Q.; Belancio, V.; Bruch, H.; Roy-Engel, A.M.; Deininger, P.L. Nickel stimulates L1 retrotransposition by a post-transcriptional mechanism. *J. Mol. Biol.* **2005**, *354*, 246–257. [[CrossRef](#)] [[PubMed](#)]
622. Al Mamun, A.A.; Lombardo, M.J.; Shee, C.; Lisewski, A.M.; Gonzalez, C.; Lin, D.; Nehring, R.B.; Saint-Ruf, C.; Gibson, J.L.; Frisch, R.L.; et al. Identity and function of a large gene network underlying mutagenic repair of DNA breaks. *Science* **2012**, *338*, 1344–1348. [[CrossRef](#)] [[PubMed](#)]
623. Giglia-Mari, G.; Zotter, A.; Vermeulen, W. DNA damage response. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a000745. [[CrossRef](#)] [[PubMed](#)]
624. Lamb, B.C.; Mandaokar, S.; Bahsoun, B.; Grishkan, I.; Nevo, E. Differences in spontaneous mutation frequencies as a function of environmental stress in soil fungi at “Evolution Canyon,” Israel. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5792–5796. [[CrossRef](#)] [[PubMed](#)]
625. Li, Y.C.; Fahima, T.; Roder, M.S.; Kirzhner, V.M.; Beiles, A.; Korol, A.B.; Nevo, E. Genetic effects on microsatellite diversity in wild emmer wheat (*Triticum dicoccoides*) at the Yehudiyya microsite, Israel. *Heredity (Edinb)* **2003**, *90*, 150–156. [[CrossRef](#)] [[PubMed](#)]

626. Li, Y.C.; Roder, M.S.; Fahima, T.; Kirzhner, V.M.; Beiles, A.; Korol, A.B.; Nevo, E. Climatic effects on microsatellite diversity in wild emmer wheat (*Triticum dicoccoides*) at the Yehudiyya microsite, Israel. *Heredity (Edinb)* **2002**, *89*, 127–132. [[CrossRef](#)] [[PubMed](#)]
627. Hübner, S.; Rashkovetsky, E.; Kim, Y.B.; Oh, J.H.; Michalak, K.; Weiner, D.; Korol, A.B.; Nevo, E.; Michalak, P. Genome differentiation of *Drosophila melanogaster* from a microclimate contrast in Evolution Canyon, Israel. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 21059–21064. [[CrossRef](#)] [[PubMed](#)]
628. Beiles, A.; Raz, S.; Ben-Abu, Y.; Nevo, E. Putative adaptive inter-slope divergence of transposon frequency in fruit flies (*Drosophila melanogaster*) at “Evolution Canyon”, Mount Carmel, Israel. *Biol. Direct* **2015**, *10*, 58. [[CrossRef](#)] [[PubMed](#)]
629. Kim, Y.B.; Oh, J.H.; McIver, L.J.; Rashkovetsky, E.; Michalak, K.; Garner, H.R.; Kang, L.; Nevo, E.; Korol, A.B.; Michalak, P. Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10630–10635. [[CrossRef](#)] [[PubMed](#)]
630. Kalendar, R.; Tanskanen, J.; Immonen, S.; Nevo, E.; Schulman, A.H. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6603–6607. [[CrossRef](#)] [[PubMed](#)]
631. A good illustration of the interplay between ecology and NGE functions emerges from the divergences that are observed in several genome parameters between climatically distinct zones at the Evolution Canyon (Yehudiyya) microsite in Israel: Spontaneous mutation in soil fungi [624], microsatellite expansion in wild emmer wheat [625,626], transposition in *Drosophila* [627–629], and retrotransposition in wild barley [630].
632. Blount, Z.D.; Barrick, J.E.; Davidson, C.J.; Lenski, R.E. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **2012**, *489*, 513–518. [[CrossRef](#)] [[PubMed](#)]
633. Hendrickson, H.; Rainey, P.B. Evolution: How the unicorn got its horn. *Nature* **2012**, *489*, 504–505. [[CrossRef](#)] [[PubMed](#)]
634. Kahrstrom, C.T. Bacterial evolution: Decoding fossil records. *Nat. Rev. Genet.* **2012**, *13*, 757. [[CrossRef](#)] [[PubMed](#)]
635. Maenhaut-Michel, G.; Shapiro, J.A. The roles of starvation and selective substrates in the emergence of araB-lacZ fusion clones. *EMBO J.* **1994**, *13*, 5229–5239. [[PubMed](#)]
636. Van Hofwegen, D.J.; Hovde, C.J.; Minnich, S.A. Rapid Evolution of Citrate Utilization by *Escherichia coli* by Direct Selection Requires citT and dctA. *J. Bacteriol.* **2016**, *198*, 1022–1034. [[CrossRef](#)] [[PubMed](#)]
637. To quote the authors of the direct selection experiment: “Long-term evolution experiments (LTEE) that were performed by Blount et al. [632] found a single aerobic, citrate-utilizing *E. coli* strain after 33,000 generations (15 years). This was interpreted as a speciation event. Here, we show why it probably was not a speciation event. Using similar media, 46 independent citrate-utilizing mutants were isolated in as few as 12 to 100 generations. Genomic DNA sequencing revealed an amplification of the citT and dctA loci and DNA rearrangements to capture a promoter to express CitT, aerobically. These are members of the same class of mutations identified by the LTEE. We conclude that the rarity of the LTEE mutant was an artifact of the experimental conditions and not a unique evolutionary event. No new genetic information (novel gene function) evolved” [636].
638. Guo, H.; Arambula, D.; Ghosh, P.; Miller, J.F. Diversity-generating Retroelements in Phage and Bacterial Genomes. *Microbiol. Spectr.* **2014**, *2*. [[CrossRef](#)] [[PubMed](#)]
639. Ye, Y. Identification of diversity-generating retroelements in human microbiomes. *Int. J. Mol. Sci.* **2014**, *15*, 14234–14246. [[CrossRef](#)] [[PubMed](#)]
640. Sometimes the terms phase variation and antigenic variation are used interchangeably. The distinction that is made here is intended to clarify the effects on protein expression and set them apart from effects on protein structure.
641. Patel, S. Drivers of bacterial genomes plasticity and roles they play in pathogen virulence, persistence and drug resistance. *Infect. Genet. Evol.* **2016**, *45*, 151–164. [[CrossRef](#)] [[PubMed](#)]
642. Foley, J. Mini-review: Strategies for Variation and Evolution of Bacterial Antigens. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 407–416. [[CrossRef](#)] [[PubMed](#)]
643. Wisniewski-Dye, F.; Vial, L. Phase and antigenic variation mediated by genome modifications. *Antonie Van Leeuwenhoek* **2008**, *94*, 493–515. [[CrossRef](#)] [[PubMed](#)]

644. Kutsukake, K.; Nakashima, H.; Tominaga, A.; Abo, T. Two DNA invertases contribute to flagellar phase variation in *Salmonella enterica* serovar *Typhimurium* strain LT2. *J. Bacteriol.* **2006**, *188*, 950–957. [[CrossRef](#)] [[PubMed](#)]
645. Horino, A.; Kenri, T.; Sasaki, Y.; Okamura, N.; Sasaki, T. Identification of a site-specific tyrosine recombinase that mediates promoter inversions of phase-variable *mpl* lipoprotein genes in *Mycoplasma penetrans*. *Microbiology* **2009**, *155*, 1241–1249. [[CrossRef](#)] [[PubMed](#)]
646. Anjuwon-Foster, B.R.; Tamayo, R. A genetic switch controls the production of flagella and toxins in *Clostridium difficile*. *PLoS Genet.* **2017**, *13*, e1006701. [[CrossRef](#)] [[PubMed](#)]
647. Valle, J.; Vergara-Irigaray, M.; Merino, N.; Penades, J.R.; Lasa, I. sigmaB regulates IS256-mediated *Staphylococcus aureus* biofilm phenotypic variation. *J. Bacteriol.* **2007**, *189*, 2886–2896. [[CrossRef](#)] [[PubMed](#)]
648. Brooks, J.L.; Jefferson, K.K. Phase variation of poly-N-acetylglucosamine expression in *Staphylococcus aureus*. *PLoS Pathog.* **2014**, *10*, e1004292. [[CrossRef](#)] [[PubMed](#)]
649. Bayliss, C.D.; Palmer, M.E. Evolution of simple sequence repeat-mediated phase variation in bacterial genomes. *Ann. N. Y. Acad. Sci.* **2012**, *1267*, 39–44. [[CrossRef](#)] [[PubMed](#)]
650. Komano, T. Shufflons: Multiple inversion systems and integrons. *Annu. Rev. Genet.* **1999**, *33*, 171–191. [[CrossRef](#)] [[PubMed](#)]
651. Hanada, K.; Yamaoka, Y. Genetic battle between *Helicobacter pylori* and humans. The mechanism underlying homologous recombination in bacteria, which can infect human cells. *Microbes Infect.* **2014**, *16*, 833–839. [[CrossRef](#)] [[PubMed](#)]
652. Cahoon, L.A.; Seifert, H.S. Transcription of a cis-acting, noncoding, small RNA is required for pilin antigenic variation in *Neisseria gonorrhoeae*. *PLoS Pathog.* **2013**, *9*, e1003074. [[CrossRef](#)] [[PubMed](#)]
653. Vink, C.; Rudenko, G.; Seifert, H.S. Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiol. Rev.* **2012**, *36*, 917–948. [[CrossRef](#)] [[PubMed](#)]
654. Modell, J.W.; Jiang, W.; Marraffini, L.A. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **2017**, *544*, 101–104. [[CrossRef](#)] [[PubMed](#)]
655. Viral target DNA appears to be acquired during the injection process itself, before the virus has a chance to reproduce [[654](#)].
656. Marraffini, L.A. CRISPR-Cas immunity in prokaryotes. *Nature* **2015**, *526*, 55–61. [[CrossRef](#)] [[PubMed](#)]
657. Nunez, J.K.; Lee, A.S.; Engelman, A.; Doudna, J.A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **2015**. [[CrossRef](#)]
658. Vestergaard, G.; Garrett, R.A.; Shah, S.A. CRISPR adaptive immune systems of Archaea. *RNA Biol.* **2014**, *11*, 156–167. [[CrossRef](#)] [[PubMed](#)]
659. Rollie, C.; Schneider, S.; Brinkmann, A.S.; Bolt, E.L.; White, M.F. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife* **2015**, *4*. [[CrossRef](#)] [[PubMed](#)]
660. Sternberg, S.H.; LaFrance, B.; Kaplan, M.; Doudna, J.A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **2015**, *527*, 110–113. [[CrossRef](#)] [[PubMed](#)]
661. Barrangou, R. RNA events. Cas9 targeting and the CRISPR revolution. *Science* **2014**, *344*, 707–708. [[CrossRef](#)] [[PubMed](#)]
662. Krupovic, M.; Beguin, P.; Koonin, E.V. Casposons: Mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* **2017**, *38*, 36–43. [[CrossRef](#)] [[PubMed](#)]
663. Krupovic, M.; Koonin, E.V. Self-synthesizing transposons: Unexpected key players in the evolution of viruses and defense systems. *Curr. Opin. Microbiol.* **2016**, *31*, 25–33. [[CrossRef](#)] [[PubMed](#)]
664. Sheng, G.; Zhao, H.; Wang, J.; Rao, Y.; Tian, W.; Swarts, D.C.; van der Oost, J.; Patel, D.J.; Wang, Y. Structure-based cleavage mechanism of *Thermus thermophilus* Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 652–657. [[CrossRef](#)] [[PubMed](#)]
665. Swarts, D.C.; Jore, M.M.; Westra, E.R.; Zhu, Y.; Janssen, J.H.; Snijders, A.P.; Wang, Y.; Patel, D.J.; Berenguer, J.; Brouns, S.J.; et al. DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **2014**. [[CrossRef](#)] [[PubMed](#)]
666. Swarts, D.C.; Hegge, J.W.; Hinojo, I.; Shiimori, M.; Ellis, M.A.; Dumrongkulraksa, J.; Terns, R.M.; Terns, M.P.; van der Oost, J. Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Res.* **2015**, *43*, 5120–5129. [[CrossRef](#)] [[PubMed](#)]
667. 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. [[CrossRef](#)] [[PubMed](#)]

668. Hanson, S.J.; Wolfe, K.H. An Evolutionary Perspective on Yeast Mating-Type Switching. *Genetics* **2017**, *206*, 9–32. [[CrossRef](#)] [[PubMed](#)]
669. Hanson, S.J.; Byrne, K.P.; Wolfe, K.H. Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus *Saccharomyces cerevisiae* system. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4851–E4858. [[CrossRef](#)] [[PubMed](#)]
670. Devlin, R.; Marques, C.A.; McCulloch, R. Does DNA replication direct locus-specific recombination during host immune evasion by antigenic variation in the African trypanosome? *Curr. Genet.* **2017**, *63*, 441–449. [[CrossRef](#)] [[PubMed](#)]
671. McCulloch, R.; Morrison, L.J.; Hall, J.P. DNA Recombination Strategies During Antigenic Variation in the African Trypanosome. *Microbiol. Spectr.* **2015**, *3*. [[CrossRef](#)] [[PubMed](#)]
672. Horn, D. Antigenic variation in African trypanosomes. *Mol. Biochem. Parasitol.* **2014**, *195*, 123–129. [[CrossRef](#)] [[PubMed](#)]
673. Marcello, L.; Barry, J.D. Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res.* **2007**, *17*, 1344–1352. [[CrossRef](#)] [[PubMed](#)]
674. Bracht, J.R.; Fang, W.; Goldman, A.D.; Dolzhenko, E.; Stein, E.M.; Landweber, L.F. Genomes on the edge: Programmed genome instability in ciliates. *Cell* **2013**, *152*, 406–416. [[CrossRef](#)] [[PubMed](#)]
675. Mochizuki, K. Developmentally programmed, RNA-directed genome rearrangement in *Tetrahymena*. *Dev. Growth Differ.* **2012**, *54*, 108–119. [[CrossRef](#)] [[PubMed](#)]
676. Nowacki, M.; Shetty, K.; Landweber, L.F. RNA-Mediated Epigenetic Programming of Genome Rearrangements. *Annu. Rev. Genom. Hum. Genet.* **2011**, *12*, 367–389. [[CrossRef](#)] [[PubMed](#)]
677. Yerlici, V.T.; Landweber, L.F. Programmed Genome Rearrangements in the Ciliate *Oxytricha*. *Microbiol. Spectr.* **2014**, *2*. [[CrossRef](#)] [[PubMed](#)]
678. Swart, E.C.; Bracht, J.R.; Magrini, V.; Minx, P.; Chen, X.; Zhou, Y.; Khurana, J.S.; Goldman, A.D.; Nowacki, M.; Schotanus, K.; et al. The *Oxytricha trifallax* macronuclear genome: A complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol.* **2013**, *11*, e1001473. [[CrossRef](#)] [[PubMed](#)]
679. Mollenbeck, M.; Zhou, Y.; Cavalcanti, A.R.; Jonsson, F.; Higgins, B.P.; Chang, W.J.; Juranek, S.; Doak, T.G.; Rozenberg, G.; Lipps, H.J.; et al. The pathway to detangle a scrambled gene. *PLoS ONE* **2008**, *3*, e2330. [[CrossRef](#)] [[PubMed](#)]
680. Chen, X.; Jung, S.; Beh, L.Y.; Eddy, S.R.; Landweber, L.F. Combinatorial DNA Rearrangement Facilitates the Origin of New Genes in Ciliates. *Genome Biol. Evol.* **2015**, *7*, 2859–2870. [[CrossRef](#)] [[PubMed](#)]
681. Alt, F.W.; Zhang, Y.; Meng, F.L.; Guo, C.; Schwer, B. Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* **2013**, *152*, 417–429. [[CrossRef](#)] [[PubMed](#)]
682. Schatz, D.G.; Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* **2011**, *11*, 251–263. [[CrossRef](#)] [[PubMed](#)]
683. Chaumeil, J.; Micsinai, M.; Ntziachristos, P.; Deriano, L.; Wang, J.M.; Ji, Y.; Nora, E.P.; Rodesch, M.J.; Jeddeloh, J.A.; Aifantis, I.; et al. Higher-order looping and nuclear organization of Tcra facilitate targeted rag cleavage and regulated rearrangement in recombination centers. *Cell Rep.* **2013**, *3*, 359–370. [[CrossRef](#)] [[PubMed](#)]
684. Yang, S.Y.; Fugmann, S.D.; Schatz, D.G. Control of gene conversion and somatic hypermutation by immunoglobulin promoter and enhancer sequences. *J. Exp. Med.* **2006**, *203*, 2919–2928. [[CrossRef](#)] [[PubMed](#)]
685. Buerstedde, J.M.; Alinikula, J.; Arakawa, H.; McDonald, J.J.; Schatz, D.G. Targeting Of Somatic Hypermutation By immunoglobulin Enhancer And Enhancer-Like Sequences. *PLoS Biol.* **2014**, *12*, e1001831. [[CrossRef](#)] [[PubMed](#)]
686. Longereich, S.; Basu, U.; Alt, F.; Storb, U. AID in somatic hypermutation and class switch recombination. *Curr. Opin. Immunol.* **2006**, *18*, 164–174. [[CrossRef](#)] [[PubMed](#)]
687. Chaudhuri, J.; Basu, U.; Zarrin, A.; Yan, C.; Franco, S.; Perlot, T.; Vuong, B.; Wang, J.; Phan, R.T.; Datta, A.; et al. Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Adv. Immunol.* **2007**, *94*, 157–214. [[PubMed](#)]
688. Kass, E.M.; Moynahan, M.E.; Jasin, M. When Genome Maintenance Goes Badly Awry. *Mol. Cell* **2016**, *62*, 777–787. [[CrossRef](#)] [[PubMed](#)]
689. Markowitz, F. A saltationist theory of cancer evolution. *Nat. Genet.* **2016**, *48*, 1102–1103. [[CrossRef](#)] [[PubMed](#)]

690. Touati, E. When bacteria become mutagenic and carcinogenic: Lessons from *H. pylori*. *Mutat. Res.* **2010**, *703*, 66–70. [[CrossRef](#)] [[PubMed](#)]
691. Chumduri, C.; Gurumurthy, R.K.; Zietlow, R.; Meyer, T.F. Subversion of host genome integrity by bacterial pathogens. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 659–673. [[CrossRef](#)] [[PubMed](#)]
692. Kovalchuk, O.; Walz, P.; Kovalchuk, I. Does bacterial infection cause genome instability and cancer in the host cell? *Mutat. Res.* **2014**, *761*, 1–14. [[CrossRef](#)] [[PubMed](#)]
693. Kasar, S.; Kim, J.; Improgo, R.; Tiao, G.; Polak, P.; Haradhvala, N.; Lawrence, M.S.; Kiezun, A.; Fernandes, S.M.; Bahl, S.; et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat. Commun.* **2015**, *6*, 8866. [[CrossRef](#)] [[PubMed](#)]
694. Pettersen, H.S.; Galashevskaya, A.; Doseth, B.; Sousa, M.M.; Sarno, A.; Visnes, T.; Aas, P.A.; Liabakk, N.B.; Slupphaug, G.; Saetrom, P.; et al. AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. *DNA Repair (Amst)* **2015**, *25*, 60–71. [[CrossRef](#)] [[PubMed](#)]
695. Gu, X.; Shivarov, V.; Strout, M.P. The role of activation-induced cytidine deaminase in lymphomagenesis. *Curr. Opin. Hematol.* **2012**, *19*, 292–298. [[CrossRef](#)] [[PubMed](#)]
696. Yoshida, J.; Akagi, K.; Misawa, R.; Kokubu, C.; Takeda, J.; Horie, K. Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Sci. Rep.* **2017**, *7*, 43613. [[CrossRef](#)] [[PubMed](#)]
697. Saier, M.H., Jr.; Kukita, C.; Zhang, Z. Transposon-mediated directed mutation in bacteria and eukaryotes. *Front. BioSci. (Landmark Ed.)* **2017**, *22*, 1458–1468. [[CrossRef](#)] [[PubMed](#)]
698. Zhang, Z.; Saier, M.H., Jr. Transposon-mediated adaptive and directed mutations and their potential evolutionary benefits. *J. Mol. Microbiol. Biotechnol.* **2011**, *21*, 59–70. [[CrossRef](#)] [[PubMed](#)]
699. Baller, J.A.; Gao, J.; Stamenova, R.; Curcio, M.J.; Voytas, D.F. A nucleosomal surface defines an integration hotspot for the *Saccharomyces cerevisiae* Ty1 retrotransposon. *Genome Res.* **2012**, *22*, 704–713. [[CrossRef](#)] [[PubMed](#)]
700. Ohno, S. *Evolution by Gene Duplication*; George Allen and Unwin: London, UK, 1970.
701. Toll-Riera, M.; San Millan, A.; Wagner, A.; MacLean, R.C. The Genomic Basis of Evolutionary Innovation in *Pseudomonas aeruginosa*. *PLoS Genet.* **2016**, *12*, e1006005. [[CrossRef](#)] [[PubMed](#)]
702. Keane, O.M.; Toft, C.; Carretero-Paulet, L.; Jones, G.W.; Fares, M.A. Preservation of genetic and regulatory robustness in ancient gene duplicates of *Saccharomyces cerevisiae*. *Genome Res.* **2014**, *24*, 1830–1841. [[CrossRef](#)] [[PubMed](#)]
703. Mattenberger, F.; Sabater-Munoz, B.; Toft, C.; Fares, M.A. The Phenotypic Plasticity of Duplicated Genes in *Saccharomyces cerevisiae* and the Origin of Adaptations. *G3 (Bethesda)* **2017**, *7*, 63–75. [[CrossRef](#)] [[PubMed](#)]
704. Magadum, S.; Banerjee, U.; Murugan, P.; Gangapur, D.; Ravikesavan, R. Gene duplication as a major force in evolution. *J. Genet.* **2013**, *92*, 155–161. [[CrossRef](#)] [[PubMed](#)]
705. Castagnone-Sereno, P.; Danchin, E.G. Parasitic success without sex—The nematode experience. *J. Evol. Biol.* **2014**, *27*, 1323–1333. [[CrossRef](#)] [[PubMed](#)]
706. Blanc-Mathieu, R.; Perfus-Barbeoch, L.; Aury, J.M.; Da Rocha, M.; Gouzy, J.; Sallet, E.; Martin-Jimenez, C.; Bailly-Bechet, M.; Castagnone-Sereno, P.; Flot, J.F.; et al. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genet.* **2017**, *13*, e1006777. [[CrossRef](#)] [[PubMed](#)]
707. Duplications have been well-recognized evolutionary facilitators for over four decades [700]. Some papers in the last few years emphasize the evolutionary importance of duplications in bacteria [701], yeast [702,703], and plants [325,704], while whole genome duplications (WGDs) have been cited as facilitators of diversity in plant-parasitic nematodes that reproduce without sex [705,706].
708. Muller, G.B. Evo-devo: Extending the evolutionary synthesis. *Nat. Rev. Genet.* **2007**, *8*, 943–949. [[CrossRef](#)] [[PubMed](#)]
709. Kouvaris, K.; Clune, J.; Kounios, L.; Brede, M.; Watson, R.A. How evolution learns to generalise: Using the principles of learning theory to understand the evolution of developmental organisation. *PLoS Comput. Biol.* **2017**, *13*, e1005358. [[CrossRef](#)] [[PubMed](#)]
710. Gilbert, S.F.; Bosch, T.C.; Ledon-Rettig, C. Eco-Evo-Devo: Developmental symbiosis and developmental plasticity as evolutionary agents. *Nat. Rev. Genet.* **2015**, *16*, 611–622. [[CrossRef](#)] [[PubMed](#)]

711. 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. [[CrossRef](#)] [[PubMed](#)]
712. Fares, M.A.; Sabater-Munoz, B.; Toft, C. Genome Mutational and Transcriptional Hotspots Are Traps for Duplicated Genes and Sources of Adaptations. *Genome Biol. Evol.* **2017**, *9*, 1229–1240. [[CrossRef](#)] [[PubMed](#)]
713. Steinrueck, M.; Guet, C.C. Complex chromosomal neighborhood effects determine the adaptive potential of a gene under selection. *eLife* **2017**, *6*. [[CrossRef](#)] [[PubMed](#)]
714. Detailed studies of individual traits provide insights into how intricately genomes may be structured for adaptive mutation. A very recent paper on copy number variation (CNV) of the CUP1 copper resistance locus in *Saccharomyces* budding yeast, for example, highlights how ancestral CUP1 repeat arrays, chromatin formatting (H3K56 acetylation), appropriately positioned replication pause sites and bidirectional promoters are all conditions for copper-induced, DSB recombination-dependent CNVs to emerge under the experimental conditions that are imposed [711]. Similar genome architectural features are observed in other *Saccharomyces* adaptive changes that involve duplications [712]. Another recent paper in *E. coli* demonstrates that genome location has a significant effect on contributions to fitness of adaptive mutations to antibiotic resistance, as well as on the spectrum of mutations that are obtained following selection [713]. In this *E. coli* example, adaptive mutations occur by nucleotide substitutions, different activating IS element (DNA transposon) insertions, or NAHR between flanking IS5 DNA transposons to duplicate the resistance locus.
715. Reyes, L.H.; Winkler, J.; Kao, K.C. Visualizing evolution in real-time method for strain engineering. *Front. Microbiol.* **2012**, *3*, 198. [[CrossRef](#)] [[PubMed](#)]
716. Shapiro, J.A. Observations on the formation of clones containing araB-lacZ cistron fusions. *Mol. Gen. Genet.* **1984**, *194*, 79–90. [[CrossRef](#)] [[PubMed](#)]
717. Komor, A.C.; Badran, A.H.; Liu, D.R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **2017**, *168*, 20–36. [[CrossRef](#)] [[PubMed](#)]
718. Dunn, B.; Paulish, T.; Stanbery, A.; Piotrowski, J.; Koniges, G.; Kroll, E.; Louis, E.J.; Liti, G.; Sherlock, G.; Rosenzweig, F. Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet.* **2013**, *9*, e1003366. [[CrossRef](#)] [[PubMed](#)]
719. Salvucci, E. Microbiome, holobiont and the net of life. *Crit. Rev. Microbiol.* **2014**, *42*, 485–494. [[CrossRef](#)] [[PubMed](#)]
720. Ryan, F.P. Viral symbiosis and the holobiontic nature of the human genome. *APMIS* **2016**, *124*, 11–19. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).