Chapter 1

Genome Stability: An Evolutionary Perspective

I. Kovalchuk
University of Lethbridge, Lethbridge, AB, Canada

Chapter Outline

1. INTRODUCTION

Genome stability is a feature of every organism to preserve and faithfully transmit the genetic material from generation to generation or from one somatic cell to another. This includes an error-free replication of genetic material (DNA or RNA) and the repair of replication mistakes or damaged DNA/RNA (see the corresponding chapters in this book). In contrast, genome instability covers a broad range of topics referring to an increased rate of DNA damage and the associated mutations, the role of potential direct and indirect mutagens, the role of external and internal factors contributing or mitigating such instability, and so on.

In this chapter, we discuss genome stability/instability from the perspective of evolution, and specifically genome evolution. Genome evolution refers to changes in the genome structure or genome size over time. Usually, such changes are changes in the DNA (RNA) sequence that are passed on to the progeny and the accumulation/fixation of such genetic variants or their rejection in a population. Genome evolution is normally discussed in the context of evolution of new species, and thus it is closely associated with the appearance of new traits, new phenotypical and morphological features, and so on.

Considering that new traits and phenotypes may develop due to epigenetic changes, it is plausible to think that genome evolution in part also involves changes in genome structure due to epigenetic modifications. In contrast to genetic changes that include mutations in the DNA sequence, epigenetic changes involve heritable but potentially reversible changes in gene expression due to changes in DNA methylation and histone modifications, among others.

Hence, how does the genome actually evolve? What are driving forces in genome evolution? Does evolution have the spontaneous and random nature? Is it actually directed and driven by changes in the external and internal environment? In the following, we contrast Darwin's and Lamarck's views on evolution and discuss the mechanisms that impact rate of evolution and the crucial role of the environment, especially as far as evolution of genome is concerned.

Genome Stability. http://dx.doi.org/10.1016/B978-0-12-803309-8.00001-X

2. EVOLUTION THEORIES AND MY REFLECTION ON THEM

The first theory of evolution was proposed about 200 years ago by a prominent French biologist Jean-Baptiste Lamarck (1744–1829). Although he was the first who built a somewhat coherent theory of evolution, he initially referred to the process as "transformation," explaining that organisms transform as a result of "a new need that continues to make itself felt." Evolution, according to Lamarck, was also the process in which less complex species evolve into more complex ones. Lamarck's understanding of evolution was summarized by him in four major laws. He used these laws to explain the two forces that he believed comprise evolution; a force driving animals from simple to complex forms, and a force adapting animals to their local environments. He formulates:

> Law 1*: Life, by its own forces, continually tends to increase the volume of every body which possesses it and to enlarge the size of its parts up to a limit which it brings about itself.* Law 2*: The production of a new organ in an animal body results from the appearance of a new want or need, which continues to make itself felt, and from a new movement which this want gives birth to and maintains.* Law 3*: The development of the organs and their strength of action are constantly in proportion to the use of these organs.* Law 4*: All that has been acquired, impressed upon, or changed in the organization of individuals during the course of their life is preserved by generation and transmitted to the new individuals that come from those which have undergone those changes.*

Let's briefly discuss all four laws. *Law 1* suggests that in the process of evolution organisms become larger; by this Lamarck also meant—more complex. Although there is some evidence that in the process of evolution organisms increased in size, primarily on transition to multicellularity, there is no evidence that larger organisms are more complex in general. Also, there is no real correlation between the size and complexity of an organism; although we must admit that it is difficult to define how complex certain species are. However, there appears to be evidence that during the history of life, the complexity of organisms increased. Hence, the law may stand with limitations. *Law 2* postulates that an organism can change when there is a need or want. The need to change when the environment changes is understandable, but what Lamarck meant here was that organisms could drive their changes in the direction they want. This is definitely an overstatement, as there is no evidence that "desire" to change plays any role in evolution. Thus, it is difficult to defend this law, but we come back to it later on in the review. *Law 3* is a famous law of "use and disuse." There are a lot of interpretations of this law and a lot of debates around it. To me, it suggests that if a certain trait is constantly required (used) in a given environment, it becomes prominent and fixed in the progeny, whereas a trait that is not used (it may have been important in the previous environment), it slowly disappears upon transition from progeny to progeny. Although it seems that Lamarck links the law of "use and disuse" to the law of "want and need," the law in my opinion stands, as use or disuse of certain traits due to the presence or absence of environment that calls upon them may result in appearance or disappearance of this trait in evolution. *Law 4* is known as "inheritance of acquired characteristics." Lamarck suggested that every change acquired by an organism is passed on to the progeny. We have plenty of evidence that this is not true, but many articles, especially in the past 10–15 years, clearly demonstrate that the acquired characteristics can be inherited. We discuss this later on in the review. The law stands in part. Therefore, as it appears, no matter how simplistic and naive the postulates may have been, they have a lot of merit. Lamarck definitely deserves a credit for recognizing the environment as a shaping force of evolution.

What about Charles Darwin's theory of evolution? There is no doubt that Darwin's work was influenced by Lamarck's work. Darwin summarized his voyage on the Beagle into the book *"On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life"* [1]. Darwin's theory of evolution suggested that organisms in a given population evolve through a process of natural selection. His main idea was that among individuals within a population there are always those ones who are fitter than others, and fitter organisms have higher chances to reproduce and leave more progeny, passing their heritable traits on to the next generation. As the environment changes due to the migration of a population of a given species or due to some unpredictable changes in the environment such as large temperature shifts, changes in water availability, or the appearance of competitors, a certain part of population that is more fit for the new environment adapts passing its genes/alleles on to the progeny, and eventually even may become a separate species. Seeing undisputable role of environment in the process of evolution, Darwin accepted a version of the inheritance of acquired characteristics proposed earlier by Lamarck. This is perhaps due to the fact that genetics and genes per se were not known to Darwin, and he had a difficult time coming up with a detailed mechanism for passing on traits from one generation to the next.

Evolution by Natural Selection proposed by Darwin did not transform much from the time he published his theory, but it was completed by additional knowledge of genes and genetics, thus transforming into the currently accepted *modern evolutionary synthesis* (*MES*) (also referred to as modern synthesis). One of the descriptions of MES was done by Futuyma. He writes:

> *The major tenets of the evolutionary synthesis, then, were that populations contain genetic variation that arises by random (i.e., not adaptively directed) mutation and recombination; that populations evolve by changes in gene frequency brought about by random genetic drift, gene flow, and especially natural selection; that most adaptive genetic variants have individually slight phenotypic*

effects so that phenotypic changes are gradual (although some alleles with discrete effects may be advantageous, as in certain color polymorphisms); that diversification comes about by speciation, which normally entails the gradual evolution of reproductive isolation among populations; and that these processes, continued for sufficiently long, give rise to changes of such great magnitude as to warrant the designation of higher taxonomic levels (genera, families, and so forth) [2].

First, whereas Darwin's theory of evolution suggested natural selection as the main (the only?) driving force, MES proposed several other possibilities, including genetic drifts and gene flow, although still giving the main emphasis to natural selection. *Second*, MES recognizes that traits are passed on from one generation to another in a discrete form of inheritance—genes. It further postulates that potential variants among individuals within the same species are due to the presence of multiple alleles of the same gene. *Third*, MES postulates that evolution is a gradual process, where small changes accumulate at the gene level becoming big changes leading to a speciation event. Therefore, macroevolution is, strictly speaking, multiple events of microevolution.

Understanding the process of evolution and what it involves still triggers many debates and disagreements that arise primarily from different perspectives, depending on whether you are a geneticist, a naturalist, a population biologist, an epigeneticist, or a paleontologist. It is still not clear whether evolution has gradual nature—whether macroevolution is a result of multiple steps of microevolution. Most of the paleontological findings suggest that speciation events occur rapidly. In 1972, the theory of punctuated equilibrium was put forward by Gould and Niles Eldredge; they proposed that evolutionary changes occur in relatively rapid spurts coincident with an increase in speciation rates [3]. They argued that such disruptions of equilibrium occur when a selective pressure is increased and when organisms adapted to a particular environment are no longer able to cope with the changing environment.

The appearance of new species may require several steps of microevolution. Since this is a continuous process, we hardly can trace the beginning and the end of the evolution process between, say, species A and B. Heterogeneity of individuals in a large population of certain species is already an initial step of microevolution, as a large pool of variations allows the population to acquire certain beneficial mutations that is favored in a certain environment much faster. *Species* are defined in many different ways. One definition is that they are a large group of individuals that carry similar phenotypic characteristics capable of interbreeding and giving rise to the fertile progeny. The ability to interbreed is the most important part of this definition. Therefore, until different populations of the same species are unable to interbreed, they are the same species, no matter how large the differences in their genetic material are. Is this actually true?

The analysis of taxonomic trees and the actual ability of certain species to interbreed and give rise to a viable progeny gave surprising results. In animals, less than 40% of animal taxa and less than 70% plant taxa represent reproductively independent lineages [4]. These data suggest that the definition of species as those that are incapable of interbreeding is at least outdated.

3. THE ROLE OF SYMBIOSIS IN GENOME EVOLUTION

As species do not live in isolation and constantly interact with each other, it is not surprising that they have learned to coexist in *parasitic*, *commensalistic*, and *mutualistic* ways. Altogether, such coexistence is referred to as a *symbiotic interaction*, although some scientists still consider symbiosis as a mutualistic relationship. Symbiosis had a great influence on evolution, including genome evolution. The very appearance of eukaryotic organisms is believed to be the result of symbiotic relationships between aerobic and anaerobic bacteria. Chloroplasts and mitochondria as well as possibly peroxisomes, flagella, cilia, centrioles, and maybe even the nucleus itself may have been independent organisms at one point, coexisting inside of a larger bacterial ancestor. The endosymbiotic theory of the appearance of eukaryotic organisms is supported by several lines of evidence [5].

First of all, the mitochondrial and chloroplast genomes resemble the genomes of some of currently existing bacteria; sequences in mitochondria have a similarity to alphaproteobacteria, whereas sequences of chloroplasts resemble those of cyanobacteria [6]. *Second*, as expected, there are organisms alive today that are in symbiotic relationships and may be in transition to becoming a new form of multicellular organisms. Such organisms are called the living intermediates. For example, the giant multinucleated protist amoeba *Pelomyxa* uses aerobic bacteria instead of mitochondria for aerobic respiration [7]. Also, the organisms such as corals, clams, and some *Paramecium* species host algae cells. Each coral polyp has a zooxanthellae algae cell within itself that carries out photosynthesis. Some clam species have a special type of cells known as iridocytes; microscopic towers of algae cells grow under iridocytes, resembling the stack of grana in chloroplasts [8]. The role of iridocytes is to filter the wavelengths that the algae prefer. In both described cases, algae cells gain protection by sharing the product of photosynthesis with host cells.

3.1 Changes in the Structure of the Organellar Genome Over Time

The mitochondrial and chloroplast genomes have undergone substantial changes since the time they have been bacterial organisms, but they have retained a circular structure characteristic of prokaryotic organisms, and genes located in them in most cases lack introns, unlike the nuclear genomes. Most of the chloroplast genomes are about 150 kb in size represented as a single circular molecule, although in some rare cases, for example, in dinophytic algae, the chloroplast genome is distributed over as many as 40 small plasmids, each carrying only several genes [9]. Through the process of evolution, the chloroplast genomes lost many genes either due to gene elimination or gene transfer to the nuclear genome, the process known as an endosymbiotic gene transfer. As a result, on average, chloroplasts contain fewer than 100 genes, much less than the original number of genes (likely between 1000 and 5000) in endosymbiotic bacteria such as cyanobacteria. Plant nuclei contain many genes originating from chloroplasts, perhaps as many as 10–15%, with *Arabidopsis* having as much as 18% of all protein-coding genes stemming from the chloroplast [10].

Was the reduction of gene number in chloroplasts a gradual process, or was it a result of several massive rapid changes? Although there is no clear answer, it is believed that most of the genes were transferred early on [11]. However, one of the reports suggests that at least some endosymbiotic gene transfers may have occurred in land plants fairly recently. For example, the gene rpl22 encoding the chloroplast ribosomal protein CL22 is present in the chloroplast genome of all plants examined except legumes, while a functional copy of rpl22 is located in the nucleus of the legume pea [12], suggesting that the transfer occurred after speciation event in legumes.

Although it is commonly accepted that chloroplasts lost genes to the nucleus, there is very little evidence that they gained genes from the nucleus or from the environment. One such example includes the horizontal gene transfer (horizontal transfer, HT) of four genes between *Bacteroides* species and minicircles of plastid genomes of dinoflagellate species *Ceratium horridum* and *Pyrocystis lunula* [13].

Mitochondrial genomes are much more complex and diverse as compared to plastid genomes. They can be either circular as in most multicellular animals, linear as in fungi, protozoa, and algae, or a combination of circular and linear chromosomes as in many plants. While animal mitochondrial genomes are fairly small, less than 20 kb in size, plant genomes are much larger, between 200 kb and 2000 kb. In animals, most of the genes are intronless, with coding regions representing over 90%, whereas in plants, many genes contain introns and only 10% of DNA represent coding regions. Cucumber has one of the largest mitochondrial genomes (~1700 kb), and unlike most plant mitochondrial genomes, it has three circular chromosomes instead of one [14]. The smaller size of mitochondrial genomes in animals compared to plants can be due to the higher rate of endosymbiotic gene transfer in animals versus plants [15].

One of the theories explaining the relative compactness of organelle genomes is an advantage in the replication process. Smaller genomes replicate faster [15], allowing the organelle to divide faster and therefore to be overrepresented in the cytoplasm. Thus, organelles with smaller genomes may have had an advantage in a natural selection process.

3.2 Mutation Rates in Organellar Genomes and Adaptive Evolution

Do the organellar genomes still evolve? Do they evolve at the same rate in different species? In part this can be answered by the analysis of mutation rate in different organellar genomes. Plastid-bearing eukaryotes typically have lower rates of silent substitutions in plastid genomes versus mitochondrial genomes [16]; however, it is generally accepted that for land plants, the plastid and nuclear genomes have up to a 10-fold greater mutation rate than the mitochondrial genome. At the same time, several plants have higher point mutation rates in mitochondria than in plastid genomes [17,18].

In animals, the mitochondrial genome mutation rate is much higher than that in the nuclear genome [19]. In addition, the mutation rate in mtDNA varies greatly among different animals. Two major theories of such variations are the generation time and metabolic rate hypotheses. The generation time hypothesis suggests that short-lived species have a higher number of replication errors per a certain time unit (eg, a year) due to a higher number of DNA replication rounds. Obviously, this model only explains the difference in replication-dependent mutations, assuming that the replication error is constant across species [20]. The metabolic rate hypothesis expresses that the mitochondrial mutation rate is a reflection of metabolic rates and the level of produced free radicals, which are different in animals of different body mass [21]. This model may be applicable to all type of mutations, regardless of the replication rate.

The detailed analysis of nucleotide substitution rates at the gene encoding cytochrome *b* across 1696 mammalian species revealed a two orders of magnitude variation between the tested lineages [20]. It was found that cytochrome *b* third codon positions are renewed every 1–2 Myr in the fastest evolving mammals and over 100 Myr in slow-evolving ones. The authors further demonstrated that the generation time and metabolic rate hypothesis could not fully explain the data. It was suggested that mitochondrial mutation rates decrease in long-lived species, which is in agreement with the mitochondrial

theory of aging, which suggests that organisms age in part due to mutations accumulating in mitochondrial [20]. This finding also suggests that mitochondrial mutation rates may have the partial adaptive nature. A more substantial proof of the adaptive nature of mutations in mtDNA comes from the work of James et al. [22]. The authors used a pairwise comparison and computer algorithms for the mitochondrial genome sequence data from 500 animals and found the evidence that mitochondria generally experience a substantial level of adaptive evolution. In addition, they found some weak evidence that the level of adaptive evolution in mitochondria correlates with the *effective population size* (Ne).

3.3 Symbiotic Interactions Between Viruses, Prokaryotes, and Eukaryotes: The Role of Transposable Elements

Symbiotic interactions between viruses and prokaryotic or eukaryotic cells as well as between prokaryotes and eukaryotes continue to shape the genomes of modern species.

Horizontal transfer (HT)—the transfer of genetic material between asexually reproducing organisms—is one of the mechanisms shaping the genomes of various species. This is a common mechanism among prokaryotes that contains passive (transduction, transformation) and active (conjugation, transformation) ways of HT. In contrast, HT is relatively rare in eukaryotes, partially due to the presence of barriers such as the presence of a nuclear envelope and the separation of gametic and somatic cells in multicellular eukaryotes. Despite this fact, many HT events have been recorded, most of them giving a selective adaptive advantage [11]. HT events can include the transfer of nongenic regions, genes, and transposable elements (TEs). HT of TE has been documented to occur across multiple phylogenetic levels both within prokaryotes and eukaryotes (reviewed in Ref. [23]). In contrast to HT of genes or TEs within prokaryotic or eukaryotic species, HT between the domains of life is less common for both genes and TEs. Moreover, interdomain HT of genes is substantially more common than interdomain HT of TEs. Hundreds of cases of prokaryote-to-eukaryote gene HTs have been characterized (reviewed in Ref. [23]). Gilbert and Cordaux carried out a comprehensive search for various groups of prokaryotic insertion sequences in 430 eukaryotic genomes [23]. They have identified 80 sequences integrated in the genomes of 14 different eukaryotes, all belonging to four distinct phyla (Amoebozoa, Ascomycetes, Basidiomycetes, and Stramenopiles). Further analysis revealed that these insertions were relatively recent events.

TEs have been initially discovered in 1948 by Barbara McClintock, a Nobel Prize winner in 1983. From the time of their discovery in maize to current days, they continue to be a puzzle. It is still unclear whether they are parasitic elements or symbionts of prokaryotic and eukaryotic species (perhaps at least some of them). Regardless of a clear definition, TEs have a powerful effect on genomes and evolution. TEs are capable of shaping host genomes through insertions, deletions, and recombination.

They can be broadly divided into *transposons*, elements capable of excision and reinsertion (conservative transposition) or elements capable of making a DNA copy first (replicative transposition) and *retrotransposons*, elements that amplify in the genome through RNA intermediates and reverse transcription followed by reinsertion. TEs are abundant among many species in various domains of life. Eubacteria, for example, may contain up to 20% of transposons [24], although many species have a very small number of transposons, and most of them are believed to be recent insertions [25]. Archaea are also believed to be similar in this matter [26]. In eukaryotes, especially in multicellular eukaryotes, TEs occupy a much larger genomic area: 45% in humans [27] and up to 85% in maize [28].

If TEs behave as any other genetic element, they have to evolve as any other genomic region under a certain selection pressure. If they are deleterious to the genome (species), they should gradually disappear, whereas if they are beneficial, they should be fixed. The fact that TEs exist in abundance in nearly all lineages suggests that transposons are likely to be under a positive selection pressure. Lineages that have a large number of TEs activated in response to stress may have a great advantage over those that either do not have them or have mobile elements that are irresponsive to environmental pressures. The mobilization of TEs may result in beneficial mutations, no matter how rare they are. These rare advantageous mutations may balance the fitness cost associated with maintaining and propagating TEs. Despite certain differences in the TE structure and the mode of activation in prokaryotes and eukaryotes, TEs have the same properties, and nearly in all cases, they are regulated by environmental factors.

Does the rate of occurrence of TEs depend on the genome size or reproduction mode? It was shown that TEs accumulate much faster in species with a low population size such as multicellular eukaryotes as compared to prokaryotes that have a large population size [29,30]. A larger population size may allow a more efficient elimination of slightly deleterious insertions by natural selection (see later), whereas a smaller population size may allow to retain them longer. The work by Startek et al. suggests that TE accumulation is also more likely in asexual eukaryotic populations that are under the constant environmental pressure than in populations living in the normal stable environment [31]. Similarly, it was suggested that in bacteria, the genomic TE content might also be influenced by stress [24].

The jury is still out whether TEs are simple cellular parasites (like viruses) or they are symbionts. The fact that cells have developed various strategies to suppress the TE activity and that during the process of evolution the majority of transposons have lost the ability to transpose strongly suggests that TEs are DNA parasites. On the other hand, TEs may very likely be symbionts of cellular organisms because they give them the advantage of fitness in the time of adverse environmental pressures, while being propagated themselves.

4. FIXATION OF A MUTANT ALLELE IN A POPULATION

The importance of the maintenance of genome integrity is reflected by the fact that mutations (changes in DNA/RNA sequence) are very rare, and they are often in the range of 10^{-9}–10^{-10} per single nucleotide per cell division. Many occurring mutations are silent, that is they do not result in the change of the phenotype (trait) because they are either *synonymous* resulting in no changes in amino acid composition of the encoded protein (due to the degenerate nature of the genetic code) or *nonsynonymous* resulting in changes to amino acid with similar properties (charge, binding capacity, and so on). Considering a low mutation rate and the fact that only a fraction of mutations changes certain inherited characteristics, 1 in 10,000,000 individuals (10^{-7}–10^{-8}) in the population would have a phenotype change due to a mutation in a defined gene of an average size of 1000 coding nucleotides. Any such mutation has a chance to be lost or fixed in the population. According to population genetics, the fixation of a mutation (or an allele) results from an increase of the frequency of such mutation from being present only in a few individuals to occurring in all individuals in the population carrying such mutation (an occurrence of 100% or near 100% in the population). Mutations in the population that spread through random genetic drift and without a positive selection pressure (the neutral theory of evolution) may have a very low chance of fixation in the population.

Let us assume that we have a population of 1000 individuals. Then, a mutation occurring in one of two alleles of these 1000 individuals will arise with a frequency of ½×1000 (1/2000). In a genetic drift, the probability that this allele will be fixed by chance is equal to its frequency in the population, ½×1000 or 5×10^{-4}. If the population is larger, let's say 1,000,000 individuals, the probability of fixing a single nucleotide mutation is ½×1,000,000 or 5×10^{-7}. With the neutral selection pressure, the rate of mutation fixation equals the rate of the introduction of such mutations. Since the fixation depends on the size of the population, the fixation of alleles in the population can only have a realistic chance either in a very small population or at a high initial frequency of occurrence of such alleles in the population. Genome position may have a substantial influence on the frequency of fixation; when mutation is positioned in the region with the high initial frequency of crossing over, the chance of fixation increases due to the higher frequency of homozygotization of de novo mutations [32]. According to Monte Carlo simulation performed by Kimura, a rare mutant gene in the effective population Ne can be fixed in 4Ne generations while being under the neutral selection pressure [33,34]. Since the population size also has a tendency to change, the decreasing population would lead to a decrease in the time required for the fixation, and the increasing population would increase the fixation time.

Selection is also an extremely important factor contributing to the fixation time of a mutation. In a deterministic model, an initially rare beneficial mutation occurring in a relatively small population will increase in frequency in each generation reaching fixation [35]. The more beneficial is the mutation, the higher selection coefficient for fixation it will have. Natural fluctuations in the frequency of the occurrence of this allele, known as *genetic drift*, especially in a larger population, may however lead to the extinction of such allele, regardless of how beneficial it is. An increase of the frequency of a beneficial allele to a certain meaningful number (10–20%) typically allows it to get fixed, and the time to fixation can be estimated using a deterministic model. In the nature, populations are in a constant flux, and the growth or decline in the population size may have a dramatic effect on beneficial mutations. Selection coefficients may be more effective for the growing population and less effective for the declining one. In the case of deleterious mutations, the effect may be opposite, selection coefficients may be more effective for the declining population and less effective for the growing population. This means that when populations grow, beneficial mutations have more chances to get fixed, while deleterious mutations will be lost. When populations decline in size, beneficial mutations are lost more frequently, while deleterious mutations are fixed more frequently.

A complete fixation of neutral or deleterious alleles is difficult to demonstrate in a research experiment due to the length of time required to achieve it when a de novo mutation with a low frequency is considered. Most of the examples of fixation of mutations come from studies in bacterial populations propagated under the selective pressure; however, advances in sequencing technology during 2010s allowed performing the so-called "evolve-and-resequence" experiments where the fixation of certain mutations is demonstrated through sequencing and mainly for point mutations. Several experimental evolution studies were performed in *Drosophila*, demonstrating the near fixation of mutations under the selective pressure [36–38].

Experiments since 2000, using partial outcrossing in worms for 50 generations also demonstrated the near fixation (near 90%) of several de novo mutations developed in a population with known fixed deleterious mutations [39]. The deleterious mutations were introduced in genes rendering sex determination temperature sensitive. Ten founder populations were evolved in intermediate temperatures for 50 generations. In these conditions, the temperature-sensitive mutants had up to a 75% reduction in fecundity and a 50% reduction in fertility. After 50 generations of experimental evolution, fertility rates have recovered in some populations almost to the wild-type levels, likely due to the selection of alleles that allowed the suppression of deleterious mutant phenotypes. Simulations performed by the author demonstrated that the fixation of de novo mutations in several populations could be due to low rates of outcrossing and a high selection pressure existing in laboratory conditions. It is interesting that a deleterious mutation can survive the selection pressure and get fixed through the co-selection with compensatory alleles (which are also favored and will likely get fixed in such population). This process resembles genetic hitchhiking, a mechanism of selection and fixation of alleles that are in a close proximity from other alleles that are under a high selection pressure.

In contrast, another experiment performed by Chelo and Teotonio using 100 generations of *Caenorhabditis elegans* failed to demonstrate the fixation of mutations, although it showed the substantially increased allelic variations [40]. The differences between this study and the study of Chandler [39] could be explained by the fact that Chelo and Teotonio [40] used a substantially larger number of genetic isolates as the initial founding population and at a much higher frequency of genetic outcrossings (over 20% versus under 5%).

In another example of fixation, Schwartz has performed a simple experiment allowing him to select wild-type maize plants over mutant plants impaired in the alcohol dehydrogenase activity encoded by the Adh1 gene [41]. Since the germination of maize seeds in flooding conditions resulted in the inability of the *adh1* mutant to germinate, a repetitive exposure of the progeny to flooding resulted in the elimination of a mutant allele from the experimental population. Similar experiments can be easily performed with plant mutants that are impaired in the germination under normal or induced conditions, also with mutants that are impaired in self-fertilization or outcrossing, and with partially or completely sterile mutants, and so on. The propagation of such plants for multiple generations may quickly eliminate homozygous mutants and severely reduce the presence of heterozygous ones, likely leading to a complete fixation.

All these examples support the notion that whereas selection has a great influence on fixation of traits, the effective population size (Ne) affects the rate of fixation to a large degree.

5. EVOLUTION OF MUTATION RATES

As stated earlier, the rate of evolution and fixation of traits largely depends on the rate of the introduction of new mutations. Mutations are an absolute prerequisite for the evolution of species and speciation (with a rare exception of epimutations that also can result in speciation). Were the mutation rate and the evolution rate constant across 3.5 billion years of life on our planet? Two major theories were proposed to describe the rate of evolution. *First*, *uniformitarianism* coined by William Whewell and originally applied to geology suggests that all laws of the universe operate at a steady constant rate. This theory contrasted *catastrophism*, a theory that suggests that the geological features of our planet have been influenced by rapid violent catastrophic events. Catastrophism as a theory can also be applied to the understanding of evolution rates.

It is still not very clear whether the rate of evolution remained constant. If it did, then the rate of mutations and the appearance of new species would have a constant rate throughout the history of life. However, several examples of rapid speciation events are known, one of them representing the Cambrian period. This period is characterized by the explosion of life diversity on our planet—the appearance of most modern animal phyla. Lee et al. employed Bayesian and maximum likelihood phylogenetic clock methods on an extensive anatomical and genomic data set for arthropods and found that the phenotypic evolution was about 4 times faster and the molecular evolution about 5.5 times faster during the Cambrian period compared to all subsequent periods [42]. The authors state that although the evolution of arthropods was much faster than the normal one, it is still within the range that can be considered acceptable and consistent with the evolution by natural selection.

The question remains whether mutation rates were constant throughout all periods and in all domains of life. An extensive research in the area of the patterning of mutation rates across various organisms was done by John Drake. In his work in 1991, John Drake analyzed the mutation rate per nucleotide per generation (u) in seven species, four DNA bacteriophage species, a bacterium, an yeast, and a filamentous fungus [43]. He found that while the average mutation rates per base pair varied by about 16,000-fold, mutation rates per genome (G) varied only by about 2.5-fold. Drake concluded that the average mutation rates per nucleotide inversely correlate with the genome size, and that the mutation rate per genome is nearly constant across all microbes [43].

Further analysis performed by Lynch shows a strong support for the initial analysis provided by John Drake [44]. A more detailed analysis that included RNA viruses, DNA viruses, archaea, and eubacteria showed a negative correlation

between u and G (Fig. 1.1A). The same analysis performed only for species that have a defined cell structure, including bacteria, archaea, and various eukaryotes, showed a completely opposite picture—an increase in the genome size resulted in an increase in mutations per nucleotide (nt) per genome per generation; a positive correlation between u and G was found (Fig. 1.1B). It is interesting to note that the mutation rate of dsDNA viruses per replication was comparable to the average mutation rate per generation in mammals.

How would one explain higher mutation rates in high eukaryotes compared to single cell eukaryotes or prokaryotes? One of the theories proposed by Drake suggests that organisms under strong selection for rapid replication cannot maximize the fidelity of DNA replacement without limiting the rate of DNA synthesis necessary for daughter cell production [43]. If this cost-of-repair hypothesis is correct, then the replication in multicellular species occurs at higher mutation costs as compared to prokaryotes [44]. This is likely true, however, if one considers that mutations in somatic cells of multicellular organisms do not have the same impact as mutations in unicellular organisms, there should be an alternative explanation for higher mutation rates in multicellular eukaryotes.

An intriguing hypothesis was proposed by Lynch [44]. He suggested that there is a certain lower limit to the mutation rate per generation in any given species, and this rate is defined by molecular or biochemical properties of a cell or by the physiology of organisms. Lynch rather suggests that the lower bound on mutation rates is set by the intrinsic inability of natural selection to push the rates any lower [44]. In neutral evolution, a spread of mutations in the population occurs through genetic drift or a random chance (see earlier). Genetic drift restricts the influence of natural selection on any mutation, and when the mutation rate is reduced to the level when any further incremental improvement conveys a fitness advantage that is smaller than the power of drift, selection will not reduce the mutation rate any lower.

Lynch further used the data on the equilibrium level of heterozygosity at silent nucleotide substitutions from major phylogenetic groupings and the data on the average mutation rates (Fig. 1.1) to calculate the average effective population size (Ne) [44]. Plotting the average base substitution rate per generation (u) versus Ne revealed a significant negative correlation between u and Ne (Fig. 1.2A). A similar negative correlation was found by Lynch [44] for mammalian mitochondrial genomes using the data from Piganeau and Eyre-Walker [45] (Fig. 1.2B).

A comparison of the nuclear and mitochondrial genome mutation rates per generation shows a much larger u in mitochondrial genomes. A higher mutation rate in mitochondria is typically explained by a high metabolic rate and the presence of free radicals as well as a lower efficiency of DNA repair. In addition, mitochondria have a very low level of homologous

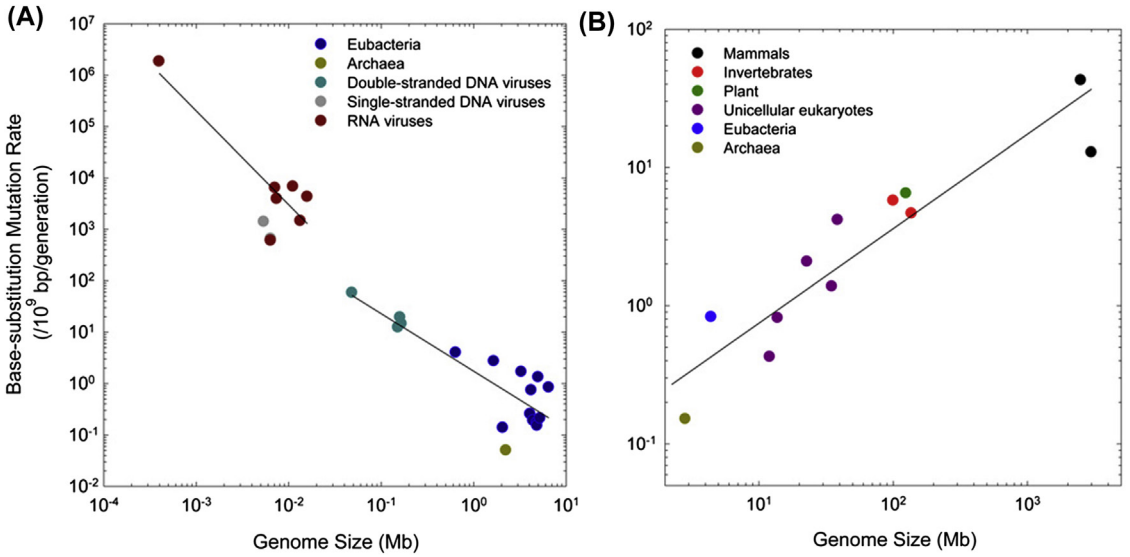

FIGURE 1.1 **The scaling of base substitution rate/nucleotide site/generation with genome size.** Each data point represents the average estimate for a separate taxon, although the results for most microbes are based on just one or a few reporter constructs (and hence, have a high sampling error), whereas those for most multicellular taxa are based on very large data sets (in several cases, whole genome sequences). (A) For noneukaryotes, two separate regressions are provided, one for RNA viruses alone, and the other for the pooled data from double-stranded DNA viruses, eubacteria, and archaea. The respective regressions of the $\log_{10}$ plotted mutation rates are $-0.17-1.83\log_{10}(G)$ and $0.24-1.12\log_{10}(G)$, with G denoting the genome size in megabases, and $r^2=0.78$ and 0.72, respectively. (B) The regression for cellular organisms is $-0.81+0.68\log_{10}(G)$, with $r^2=0.80$. Here, the results for various eubacteria (excluding *Buchnera* which has an unusually small genome) are averaged into a single point. The pattern is quite similar if prokaryotes are excluded (the slope = 0.59 and $r^2=0.83$). *Reproduced with permission from Lynch M. The rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci USA 2010;107:961–68.*

recombination and therefore do not have an efficient mechanism of elimination of deleterious mutations. Mitochondria are uniparentally inherited, and mtDNA is distributed to several primary oocytes where mtDNA is replicated during maturation of oocytes. This segregation of mtDNA variants represents a genetic bottleneck that allows a rapid propagation of at least some mutations [46]. At the same time, the mutation rate in plant mitochondria is significantly smaller than in animal mitochondria, likely due to the fact that plant mitochondria are larger in size than animal mitochondria. However, the mutation rate in plant mitochondria is much lower than in the nucleus, despite the fact that the nuclear genome is much larger in size. Hence, it is still unclear whether a low mutation rate in plant mitochondria is either due to a more efficient repair rate, or a higher level of recombination, or perhaps due to a different mechanism of inheritance of plant mitochondria by the progeny.

5.1 Evolution of Somatic Mutation Rates

Somatic mutations in multicellular organisms may not alter the germline mutation rate; however, the accumulation of somatic mutations may contribute to a reproductive success of an organism due to the development of cancer or other severe diseases. Since plants do not have a predetermined germline, but rather develop it from somatic cells, somatic mutations occurring early in plant development may actually be inherited. For other higher eukaryotes, it is not clear whether somatic mutations can substantially affect germline mutations or their heritability. It is possible that the somatic mutation rate influences the evolution of the germline mutation rate in animals. Also, it is possible that the germline mutation rate actually influences the somatic mutation rate [44].

In order to compare the somatic mutation rate with the germline mutation rate, Lynch has prorated the germline mutation rates to a single cell division [44]. This comparison was very revealing: in humans, the rate of nucleotide substitution per cell division in the germline is very low, although not as low as in microbes (Table 1.1). The somatic mutation rates in metazoans are much higher than the germline mutation rates. For example, in humans, the average somatic mutation rate calculated from four tissue types was 17-fold higher than in the germline, 1.02×10^{-9}/nt/cell division versus 0.6×10^{-10}. This somatic mutation rate is also higher (3.5-fold) than the average one in yeast and *Escherichia coli* [44]. The analysis of mutation rates in mice and rats using LacZ or LacI transgenes also showed that somatic cells have up to a sixfold higher mutation rate compared to the mutation rate in testes (Table 1.1). Moreover, it was clearly demonstrated that somatic mutations accumulated with age, whereas germline mutations remain relatively constant (Fig. 1.3).

Considering that mammalian genomes are large (often several Gbp) and adult organisms consist of billion cells, the genetic load of new mutations in somatic cells of adult organisms is enormous. With the diploid genome size of a human being about 6.5×10^{9} nt and the average cell number being about 10^{13}, an adult human might carry as many as 10^{16} point

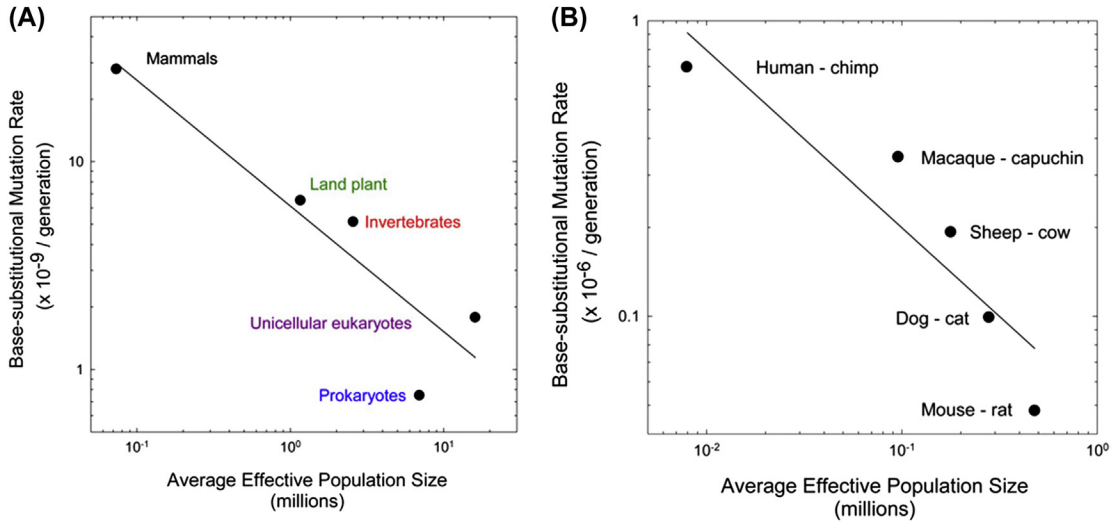

FIGURE 1.2 **The scaling of the base substitution mutation rate per generation (u) and the effective number of genes per locus (2Ne for diploids, and Ne for haploids).** (A) The slope of the log–log regression for the nuclear genome of major phylogenetic groupings is –0.60 (0.16), where the number in parentheses denotes the standard error, with $r^2=0.84$, although if the estimated Ne for prokaryotes is assumed to be 10 times too low [29], the slope is modified to –0.52 (0.02) with $r^2=0.99$. (B) The slope of the log–log regression for the mitochondrial genome of mammalian lineages is –0.60 (0.15), with $r^2=0.84$. The data are the average estimates from analyses assuming fixed and variable substitution rates in Piganeau and Eyre-Walker [45]. *Reproduced with permission from Lynch M. The rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci USA 2010;107:961–68.*

TABLE 1.1 Mutation Rates per Nucleotide Site ($\times 10^{-9}$) in a Variety of Tissues

Species	Tissue	Cell Divisions per Generation[a]	Mutation Rates[b]	
			Per Generation	Per Cell Division
Homo sapiens	Germline	216	12.85	0.06
	Retina	55	54.45	0.99
	Intestinal epithelium	600	162	0.27
	Fibroblast (culture)			1.34
	Lymphocytes (culture)			1.47
Mus musculus	Male germline	39	38	0.97
	Brain		76.94	
	Colon		83.35	
	Epidermis		90.38	
	Intestine		117.69	
	Liver		237.88	
	Lung		166.83	
	Spleen		130	
Rattus norvegicus	Colon		178.38	
	Kidney		167.45	
	Liver		179.92	
	Lung		223.22	
	Mammary gland		57.7	
	Prostate		448.9	
	Spleen		101.62	
Drosophila melanogaster	Germline	36	4.65	0.13
	Whole body		380.92	
Caenorhabditis elegans	Germline	9	5.6	0.62
Arabidopsis thaliana	Germline	40	6.5	0.16
Saccharomyces cerevisiae		1	0.33	0.33
Escherichia coli		1	0.26	0.26

[a]*References to data on the number of germline cell divisions: human [81];* D. melanogaster *and mouse [82];* C. elegans *[83]; and* A. thaliana *[84]. The number of cell divisions is unknown for the mouse and rat rates.*
[b]*Mammalian tissue-specific rates are given only for tissues in which at least two independent estimates have been acquired. All data on human mutation rates are taken from Lynch (2010) [85]. Data for somatic mutation rates in mice and rats are derived from references contained within the Supplementary Material, see Lynch (2010) [85]. References to data on germline mutation rates:* D. melanogaster *[86];* C. elegans *[87];* A. thaliana *[59];* S. cerevisiae *[53]; and* E. coli *[19].*
Reproduced with permission from Lynch M. The rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci USA 2010;107:961–68.

mutations [44]. Although only a fraction of these mutations occur in coding regions, and even smaller fractions may be deleterious, adult organisms easily accumulate over 10^{10} mutations.

This analysis also showed that tissues accumulate more mutations with age. Similar data are reported for medaka fish [47] and *Drosophila melanogaster* [48]. Lynch concluded that if not for the separation of generations via the germline, the heritable per generation mutation rates for animals would be much higher than reported now [44]. This

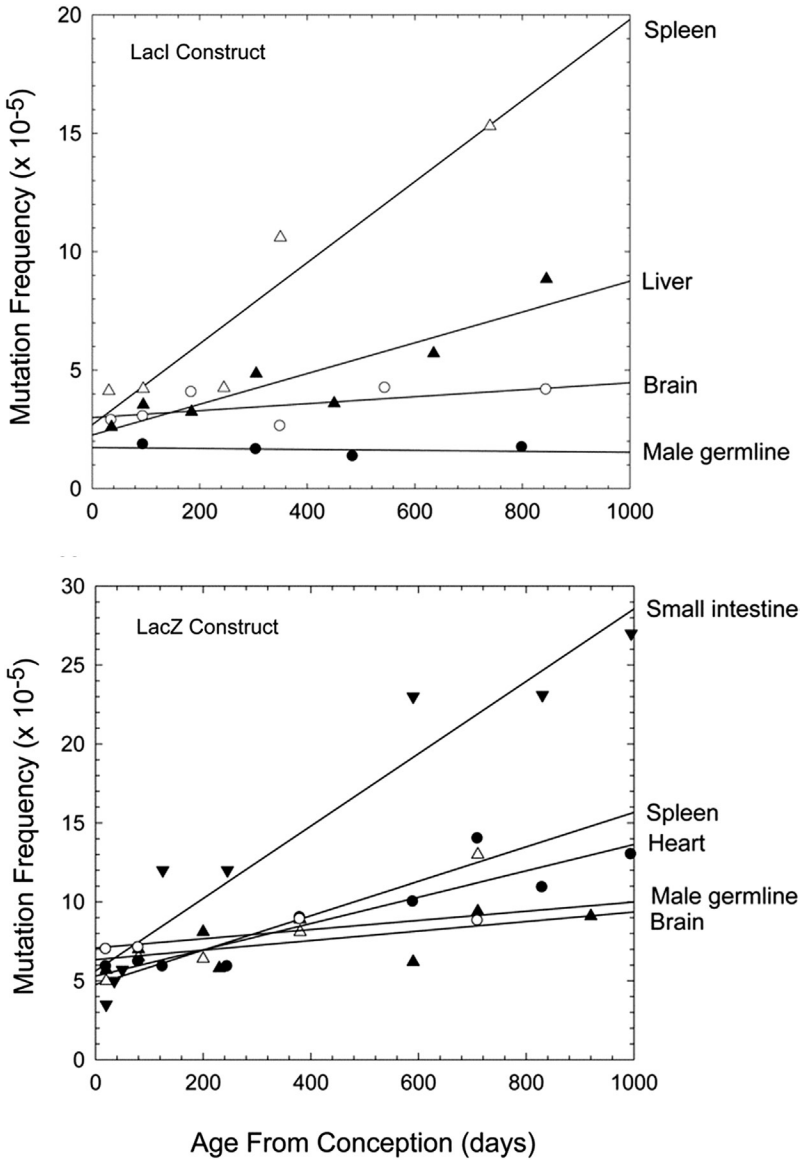

FIGURE 1.3 **Tissue-specific frequencies of mutations as a function of age in mouse lines carrying Lac reporter constructs.** Results are averaged over multiple studies. *Reproduced with permission from Lynch M. The rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci USA 2010;107:961–68.*

likely would have dramatic and likely negative effect on animals, due to substantial increase in heterogeneity and deleterious mutations in relatively stable environment. It would be interesting to run simulations to confirm or disprove this hypothesis.

6. GENOME INSTABILITY: IS IT RANDOM?

As we discussed earlier, mutations (whether they are genetic or epigenetic in nature) are an absolute prerequisite of evolution. Assuming that all mutations have a random nature, it is only the selection process that decides which mutation is beneficial or deleterious, and thus it is ultimately up to the environment to fix the mutation or make it disappear. If the environment has an impact on fixation of certain mutations, does this mean that evolution as a process is directed toward the fixation of beneficial traits that give advantage in a specific environment? Does it mean that the mutation process might be random, whereas the evolution process is directed—directed to the survival of the fittest? The answers to these questions are not necessarily obvious.

But are mutations actually truly random? If this was the case, then a mutation of a nucleotide in any genomic position would always have the same frequency. Nowadays, we know that this is not the case as certain genomic regions evolve faster than others. However, this is the result of mutations and forces of evolution acting upon them (genetic drift or/and natural selection). This is where the problem lies; it is difficult to separate the two events—mutations and an evolutionary force. For example, it is well known that cancers demonstrate higher mutation rates in genes and genomic regions, clustering at recurrent mutation hot spots that actually aid cancer cells in their immortality [49,50]. It is also possible, however, that nonrandomness of appearance of mutations in different genomic regions is unrelated to randomness of accumulation of mutations in the population under selective pressure.

6.1 A Bias in Mutations in Different Genomic Regions

Mutations may occur in certain regions more frequently due to the intrinsic properties of endogenous and exogenous DNA-damaging agents and chemical modifications of nucleotides (oxidation, spontaneous deamination, and so on), and the associated DNA repair processes [51]. Purines are damaged by alkylation more often than pyrimidines; as a result, depurination is a frequent process. A bias in the type of DNA damage results in a bias in the type of mutations that occurs in the genome. For example, for every two possible transversions (changes from purine to pyrimidine and vice versa), there is one possible transition (a change from purine to purine or from pyrimidine to pyrimidine), giving the rate of transitions (ti) to transversions (tv) as 0.5. Despite this fact, the ti/tv rate in the human genome is about 2.0, and even higher at the exons—about 3.0 [52]. For other organisms, it has been reported that the ti/tv ratio is 0.62 for yeast [53] and 1.5 for maize [54]. Curiously, mitochondrial DNA (mtDNA) has been reported to have a much stronger bias toward transitions over transversions compared with nuclear genes [55,56]. In the human genome, A:T→G:C (A→G) and G:C→A:T (G→A) changes are predominant among other types of point mutations, including T→C transitions [44]; this difference is even more dramatic in the transcribed regions, likely influenced by transcription-coupled repair [57].

In addition to a bias in the mutation rate at a single nucleotide site, there is a substantial bias in point mutations at dinucleotides. For example, in mammals, and specifically in primates, C→T transitions that arise at CpG dinucleotides are 15 times more frequent than at other dinucleotides [58]; this is likely explained by the common spontaneous oxidative deamination of methylated cytosines at CpGs. A high frequency of C→T mutations at CpGs may explain an extremely rare occurrence of CpGs in the human genome. Considering that the human genome is about 42% GC rich, the occurrence of G or C nucleotide should be 21%, and thus, the frequency of occurrence of CpG should be $0.21 \times 0.21 = 4.41\%$. Instead, there is less than 1% of CpGs in the human genome. Similar data were also reported for *Arabidopsis* plants; the mutation rate at methylated cytosines at CpGs is substantially higher than at nonmethylated ones. There is also a higher rate of G→A substitutions at CpG sites, which cannot be easily explained [59].

Also, the rate of nucleotide substitutions varies at the base pair scale. In primates, substitution rates at the G:C base pairs (excluding the CpG sites) are up to 85% higher than at the A:T base pairs [60,61], it is possibly because cytosine is intrinsically more mutable than other bases [62].

Finally, it is well known that some short sequence motifs (such as minisatellite and microsatellite repeats, 1–100 nt repeated multiple times) are highly mutable. The mutation rate at some microsatellite loci is 10^3–10^5 higher than in the coding regions [63]. Among the possible mechanisms of such high mutability are replication slippage, a gain or loss of one or more repeat units [64], an unequal crossing over, nucleotide substitutions, and duplication events [65].

The existence of cryptic mutation hot spots was demonstrated by Hodgkinson et al. [66]. The authors compared the pattern of single nucleotide polymorphism (SNP) in the human and chimpanzee genomes. They hypothesized that if some genomic regions are more mutable than others, there should be more sites that are similarly polymorphic in both species (the so-called coincident SNPs). Indeed, they have found that the number of coincident SNPs was three times higher than the number of SNPs that are randomly distributed in the two genomes.

One of the most common and perhaps the easiest ways to test the randomness of mutations was to subject cells to nonlethal selection and allow them to mutate in a right direction, so that they are able to grow and multiply. The most commonly known experiment demonstrated that the *E. coli* strain with a nonsense mutation in the *lacZ* gene rendering it unable to use the lactose mutated back to the wild-type allele only when the lactose was present in the medium; the absence of the lactose in the medium did not induce this mutation [67]. Another experiment demonstrated that a strain in which *lacZ* was deleted gained two mutations that allowed to utilize the lactose; the first mutation occurred in a cryptic gene *ebgA* in such a way that it was able to hydrolyze the lactose, and the second mutation inactivated the *ebgR* gene, a negative regulator of *ebgA*. Considering that either of these mutations requires the frequency of 10^{-8}, these two mutations can coincide in the same cell with the frequency of 10^{-16} [68]. Both of these examples demonstrate the fact that mutations can likely occur in a nonrandom manner, and that such mutations are triggered by the environmental factor.

A very interesting work (that still continues) was done using *E. coli* that has demonstrated how the environment drives the appearance of certain type of mutations. *E. coli* cells cannot grow on citrate. In the experiment that started in 1988, Blount et al. propagated 12 identical populations of *E. coli* in the abundance of citrate and a limited amount of glucose [69]. They have frozen bacterial cultures every 500 generations. This gave them the possibility to analyze the potential genetic changes. The propagation of all 12 populations for 30,000 generations did not result in the appearance of the capacity to metabolize citrate, although such number of generations should have been sufficient to mutate every possible nucleotide more than once. After 31,500 generations, one of the populations has developed the ability to use citrate. It took another 2000 generations for this ability in the population to become very common. Although Cit+ cells continued to use glucose, they did not drive Cit− cells to extinction because the Cit− cells were superior competitors for glucose (even though it was present in low amounts). The authors performed replay evolution starting from different generations (at 500 increments). If they had started regrowing bacteria from stocks frozen before generation 20,000, they would not be able to obtain Cit+ cells. Between 20,000 and 27,000 generations, Cit+ cells were obtained, but they were extremely rare, after the 27,000th generation, their appearance was just rare [69].

Therefore, the authors concluded that the appearance of mutations leading to Cit+ cells was contingent upon the appearance of one or several previous enabling mutations. A population in which Cit+ cells appeared was no hypermutable, suggesting that the potentiation of the appearance of Cit+ cells was the event that was specific to the Cit function. The authors proposed two possible mechanisms [69]. The first mechanism suggests the epistatic interaction whereby the locus where a mutation leading to the Cit+ phenotype has occurred likely interacts with another locus where an earlier mutation has occurred. The second mechanism suggests that the first mutation(s) create a new sequence that allows the Cit+ mutation to occur; this can be due to the insertion/deletion of a sequence or the rearrangement or the insertion/excision of a TE, and so on. The exact sequence of the enabling mutation and a mutation leading to Cit+ are not known; the authors are currently sequencing bacteria from before and after generation 20,000.

Our research also demonstrated some directionality in the appearance of mutations. We used two types of tobacco plants, one was resistant to infection with tobacco mosaic virus (TMV) due to the presence of the resistance gene *N*, and the other one that was sensitive to TMV because plants lacked the *N* gene [70]. We collected seeds from the infected plants and analyzed genomic rearrangements in genes that carried homology to the *N* gene, the gene that conferred resistance to TMV. In order to do this, we performed the Southern blot analysis using the genomic DNA of progeny plants with DNA of fourth exon (the conservative region encoding leucine-rich repeats required for the recognition of pathogens) of the *N* gene as a probe. The experiment showed that tobacco plants carried as many as 30 loci with a certain degree of sequence homology (about 50–70%) to the fourth exon. The analysis showed great variations in the pattern of bands in the progeny of infected plants that did not contain the *N* gene; in fact, an eightfold higher rearrangement frequency was found in the progeny of infected plants as compared to the progeny of control plants [70]. The same analysis using probes against actin loci did not show any difference between the progeny of infected plants and the progeny of control plants, suggesting that an increase in the rearrangement frequency was locus specific. This analysis allowed us to conclude that the environmental pressure can induce potentially beneficial mutations (as it was not confirmed whether rearrangements in these loci resulted in the appearance of a viable resistance gene). Unfortunately, at the time of experiment, no comparison was done between mutation rates in the progeny of plants that had the *N* gene versus plants that did not have the *N* gene. Therefore, we do not know whether or not the absence of the active resistance gene (*N*) that triggers an increase in rearrangements at the loci mentioned earlier in response to TMV infection. More details on transgenerational changes in response to the environmental pressure, including changes in genome stability can be found in Chapter 35 and Chapter 36.

7. GENOME EVOLUTION MAY START FROM CHANGES AT THE LEVEL OF DNA METHYLATION OR CHROMATIN MODIFICATION

So far, we have only mentioned the possibility that epigenetic modifications can be critical components (or perhaps even a driving force) of evolution. If genetic variations in a population can be fixed either through random genetic drift or through selection, the same can likely happen with epigenetic variations. Epigenetic modifications such as DNA methylation and histone modifications (see Chapter 36) regulate gene expression and thus contribute to the phenotype appearance. Epigenetic modifications are highly sensitive to environmental changes allowing organisms to respond to the environmental cues more efficiently. Epigenetic modifications are inherited from somatic cell to somatic cell and from generation to generation. However, epigenetic modifications are reversible; therefore, changes in phenotypes that occur due to epigenetic modifications may also be reversible. Also, epigenetic changes typically occur in a large part of a population (perhaps in an entire population if it was exposed to certain environmental changes). Hence, epigenetic modifications may bring many critical advantages as far as the mechanisms of evolution are concerned (Table 1.2).

TABLE 1.2 Comparison of Genetic and Epigenetic Mechanisms of Evolution

	Genetic	Epigenetic
The frequency of occurrence	Extremely low	Very high
Response to the environment	Slow and likely random	Immediate
The directionality of changes (in terms of immediate benefits)	Likely neutral	Directed
The specificity of a response	Unspecific	Highly specific
The reversibility of changes	Rarely reversible	Mostly reversible
The directionality of changes (in terms of immediate benefits)	Likely neutral	Directed
Requirements of a constant environmental pressure	Required for selection	Required for the maintenance of changes
Costs to an organism	Very low (unless mutation is deleterious)	Very high
Long-term solution	Typical	Problematic

Since epigenetic changes occur immediately in response to the environment, they may represent the first level of defense, they may also be the first step in the process of evolution [71]. However, since epigenetic changes are reversible and need to be maintained, they may not represent an ideal long-term solution for an organism or a population. Yet, because the presence of certain epigenetic modifications may increase the chance for certain genetic mutations to occur, epigenetic modifications may aid genetic changes at certain loci, directing changes toward them. For example, methylated cytosines are prone to deamination leading to C→T mutations; when such mutations occur, they eliminate cytosines; therefore, they may, for example, change the sequence of certain genetic elements in the promoter region or other regions with regulatory elements, thus changing the ability of regulatory proteins to bind and affect gene expression [72]. In addition, DNA methylation influences the chromatin structure—higher methylation levels lead to a more closed chromatin structure, and vice versa [73]. Similarly, histone modifications also alter the chromatin structure; some of them make it more open, whereas the others make it more closed (see Chapter 36). Since genomic rearrangements occur more frequently in open chromatin, epigenetic changes that alter the chromatin structure may alter the frequency of rearrangements [74].

Spontaneous epigenetic variations are far more common in *Arabidopsis* than genetic variations. The propagation of 10 *Arabidopsis* lines for 30 generations resulted in a substantial genetic variation, but a far larger epigenetic variation was observed; specific epimutations were detected in all lines tested [59,75]. Johannes et al. used two parental *Arabidopsis* lines that had DNA sequence differences, but substantial DNA methylation differences demonstrate the stability of epigenetic variations [76]. They propagated these lines for eight generations by selfing, thus obtaining epigenetic Recombinant Inbred Lines (epiRILs). The epiRILs showed a variation and a high heritability of flowering time and plant height that lasted at least eight generations. This work demonstrated the fact that numerous epialleles can be stable over many generations in the absence of selection or an extensive DNA sequence variation [76].

Both of the abovementioned examples derive from highly inbred lines that were selfed and propagated in the laboratory environment. Nevertheless, even these lines grown in similar environments quickly gain a substantial epigenetic diversity. The effect of the environment on epigenetic variations should be far greater. The analysis of two nearby habitats of *Laguncularia racemosa* grown at a riverside or near a salt marsh revealed great dissimilarities, with individual plants showing little genetic but abundant DNA methylation differentiation. Moreover, plants grown near a salt marsh had the hypomethylated genome as compared to plants grown in a riverside [77]. This study showed that phenotypic variations in plants grown in the contrasting environments can be primarily triggered by epigenetic changes. It is possible that these two populations will eventually undergo genetic differentiation, and new species may arise as a consequence of both epigenetic and genetic changes, with epigenetic changes being a primary cause.

Differences in methylation patterns in plants grown in the contrasting environments can be propagated for a number of generations. In oil seed crop, *Brassica napus*, variations in methylation patterns analyzed by methylation-sensitive amplification polymorphism (me-AFLP) were demonstrated to be extremely stable. Only 5% of 627 me-AFLP markers identified in several independent parental lines were variable at different developmental stages following the growth in the contrasting

environments (11 different natural environments with winter and spring planting). Moreover, for two distinct parental lines, 97% of epialleles were transmitted through five meioses and segregated in a mapping population [78].

In the study by Yi et al., five populations of *Jatropha curcas* plants collected from China (CN), Indonesia (MD), Suriname (SU), Tanzania (AF), and India (TN) were planted in one farm under the same agronomic practices (and the environmental conditions) [79]. The analysis of polymorphism showed a very low level of genetic diversity (a polymorphic band <0.1%) within and among populations, despite the fact that plants were grown in the contrasting environments. In contrast, intermediate but significant epigenetic diversity was detected (25.3% of bands were polymorphic) within and among populations. The authors identified 39 different polymorphic epimarkers, with 30 of them being heritable and following Mendelian segregation [79].

8. CONCLUSION

DNA damage caused by external and internal stresses must be efficiently resolved through the activity of various DNA repair pathways. DNA damage repair is the primary mechanism that preserves the intactness of the plant genome and insures its stability. Noteworthy, many of DNA repair pathways lack high fidelity, and their widespread activity may actually destabilize the genome, compromise its integrity, and may even be lethal for a cell [80].

At the same time, a certain level of genome flexibility is absolutely required for the successful evolution of plant species. In fact, mistakes made during DNA repair may serve as the raw material of evolution. Similarly, the rearrangement and duplication of existing DNA sequences may lead to the generation of new traits conferring a selective advantage under new growth conditions. In view of this, a very delicate balance must exist between different DNA repair pathways to ensure the continuous generation of new DNA sequence variants without affecting genome functions and cell vitality. In fact, different groups of organisms seem to preferentially use specific DNA repair pathways, depending on their genome content. Some organisms with a small genome size and a low content of repetitive sequences preferentially use HR that ensures the minimum number of mistakes being made. An increase in genome size leads to an increase in the probability that a wrong template may be chosen for the HR-mediated repair of DNA damage. The presence of a high number of repetitive DNA elements further complicates this problem as an improper recombination event may result in large deletions and duplications. Hence, a shift occurs toward using a more error-prone pathway which is less dangerous to the genome. For example, plants usually prefer using nonhomologous end joining (NHEJ) to HR. At the same time, HR is mainly used for the reshaping of the plant genome. Indeed, in the genome of *Arabidopsis thaliana*, a substantial part of genetic material was created by the duplication of existing DNA sequences which was possibly mediated by the HR pathway.

The emerging evidences suggest that epigenetic marks and stress-induced sequence-specific signals such as small RNAs (sRNAs) may play an important role in the maintenance of genome stability and chromatin. Decreased DNA methylation and an open chromatin structure attract the activity of the HR pathway and lead to the activation of transposons, thus decreasing genome stability. At the same time, frequent rearrangements at unmethylated loci may help accelerate the evolution of targeted sequences. On the contrary, increased genome methylation permits to preserve genome stability at the time of stress and prevents undesirable rearrangements and transpositions.

The epigenetic landscape of eukaryotic genomes is rather complex, and it is subject to continuous modifications in response to various stimuli. The forgoing leads to an interesting speculation that epigenetic marks are not only used to control the transcriptional activity of the chromatin helping to adjust transcriptional profiles to new growth conditions, but they also contribute to maintain genome stability and direct genome evolution. The fact that stress can cause sequence-specific epigenetic changes permits further speculation about the involvement of epigenetic modifications in stress-directed genome evolution. Unfortunately, there is still no direct evidence that links together stress, epigenetic modifications, and mechanisms that control genome stability in a multifaceted system that regulates and directs genome evolution. It is the challenge for future studies to unravel these links and provide a better understanding of how stress-directed genome evolution and adaptation occur in nature.

GLOSSARY

Catastrophism A theory that suggests that geologic features of our planet have been influenced by rapid violent catastrophic events.

Coincident SNPs Polymorphisms found at similar genomic positions upon comparison between closely related species.

Commensalistic interaction Represents a symbiotic interaction between two organisms where one organism benefits from it and another one is not affected by it.

Endosymbiotic gene transfer A transfer of genes from the chloroplast to the nuclear genome; refers to the evolutionary process.

Fecundity Fertility of an organism.

Fixation Reaching 100% or near 100% occurrence of a certain mutation/allele in a population.

Genetic drift Refers to variations in the frequency of certain alleles in a population due to randomness of sampling. For haploid and diploid species, the variance in allele frequency resulting from drift is proportional to 1/Ne and 1/(2Ne), respectively [40].

Genetic hitchhiking A mechanism of selection and fixation of alleles that are in close proximity from alleles that are under a high selection pressure.

Macroevolution Typically reflects larger changes, across the boundaries of species; often referred to major changes, such as *the appearance of new substantially different species*, the appearance of flowering plants, placental animals, and so on.

Microevolution Reflects changes in traits/phenotypes within a population or a species.

Modern evolutionary synthesis The modern theory of evolution that was synthetized on the basis of Darwin's theory of evolution and completed by additional knowledge of genes and genetics.

Mutualistic interaction Represents a symbiotic interaction between two organisms where both organisms benefit.

Ne An effective population size which reflects the size of a population with respect to a random distribution of alleles. Ne is typically smaller than the actual size of a population due to variations in sex-ratio bias, variations in family size, the nonrandomness of mating, to name a few factors. The smaller is the population, the closer is Ne to the actual population size [44].

Nonsynonymous A mutation that changes the amino acid composition of the encoded protein.

Silent mutation (silent site) A nucleotide substitution that does not result in changes in the protein sequence.

Single nucleotide polymorphism Variations at specific genomic positions existing in a population of certain species.

Somatic mutation A mutation in somatic (nongermline) cells of a multicellular organism. The somatic mutation rate may correlate and influence the germline mutation rate and may influence the fitness of an organism leading to changes in inheritance.

Synonymous mutation A mutation that does not change the amino acid composition of the encoded protein.

Transposable element A mobile element of a transposon, a genetic element capable of moving/duplicating itself (either through excision or through copying) from one genomic position to another.

u The mutation rate per nucleotide per generation.

Uniformitarianism A theory that suggests that all laws of the universe operate at a steady constant rate.

LIST OF ABBREVIATIONS

HT Horizontal transfer

MES Modern evolutionary synthesis

TE Transposable element

REFERENCES

[1] Darwin CR. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 1st ed. London: John Murray; 1859.

[2] Futuyma DJ. Evolutionary biology. Sinauer Associates; 1986. p. 12.

[3] Eldredge N, Gould SJ. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM, editor. Models in paleobiology. San Francisco: Freeman Cooper; 1972. p. 82–115. Reprinted in N. Eldredge Time frames. Princeton: Princeton Univ. Press, 1985, pp. 193–223.

[4] Rieseberg LH, Wood TE, Baack EJ. The nature of plant species. Nature 2006;440(7083):524–7.

[5] Martin WF, Garg S, Zimorski V. Endosymbiotic theories for eukaryote origin. Philos Trans R Soc Lond B Biol Sci 2015;370(1678):20140330.

[6] Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet February 2004;5(2):123–35.

[7] Martin W, Hoffmeister M, Rotte C, Henze K. An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. Biol Chem November 2001;382(11):1521–39.

[8] Holt AL, Vahidinia S, Gagnon YL, Morse DE, Sweeney AM. Photosymbiotic giant clams are transformers of solar flux. J R Soc Interface December 6, 2014;11(101):20140678.

[9] Sandelius AS. The chloroplast: interactions with the environment. Springer; 2009. ISBN: 978-3-540-68696-5. p. 18.

[10] Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA September 17, 2002;99(19):12246–51.

[11] Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 2008;9:605–18.

[12] Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J October 1991;10(10):3073–8.

[13] Mackiewicz P, Bodył A, Moszczyński K. The case of horizontal gene transfer from bacteria to the peculiar dinoflagellate plastid genome. Mob Genet Elem 2013;3(4):e25845.

[14] Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell 2011;23(7):2499–513.

[15] Selosse M, Albert B, Godelle B. Reducing the genome size of organelles favours gene transfer to the nucleus. Trends Ecol Evol 2001;16(3):135–41.

[16] Smith DR. Mutation rates in plastid genomes: they are lower than you might think. Genome Biol Evol 2015;7(5):1227–34.

[17] Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol 2007;7:135.

[18] Sloan DB, Alverson AJ, Chuckalovcak JP, et al. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol 2012;10:e1001241.
[19] Lynch M. The origins of eukaryotic gene structure. Mol Biol Evol 2006;23:450–68.
[20] Nabholz B, Glémin S, Galtier N. Strong variations of mitochondrial mutation rate across mammals–the longevity hypothesis. Mol Biol Evol January 2008;25(1):120–30.
[21] Gillooly JF, Allen AP, West GB, Brown JH. The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proc Natl Acad Sci USA 2005;102:140–5.
[22] James JE, Piganeau G, Eyre-Walker A. The rate of adaptive evolution in animal mitochondria. Mol Ecol November 17, 2015. http://dx.doi.org/10.1111/mec.13475.
[23] Gilbert C, Cordaux R. Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes. Genome Biol Evol 2013;5(5):822–32.
[24] Newton ILG, Bordenstein SR. Correlations between bacterial ecology and mobile DNA. Curr Microbiol 2011;62:198–208.
[25] Wagner A. Cooperation is fleeting in the world of transposable elements. PLoS Comput Biol 2006;2:1522–9.
[26] Filée J, Siguier P, Chandler M. Insertion sequence diversity in archaea. Microbiol Mol Biol Rev March 2007;71(1):121–57.
[27] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
[28] Schnable PS, Ware D, Fulton RS, et al. The B73 maize genome: complexity, diversity, and dynamics. Science 2009;326:1112–5.
[29] Lynch M. The origins of genome architecture. Sunderland (MA, USA): Sinauer Associates Inc.; 2007.
[30] Whitney KD, Garland Jr T. Did genetic drift drive increases in genome complexity? PLoS Genet 2010;6:e1001080.
[31] Startek M, Le Rouzic A, Capy P, Grzebelus D, Gambin A. Genomic parasites or symbionts? Modeling the effects of environmental pressure on transposition activity in asexual populations. Theor Popul Biol December 2013;90:145–51.
[32] Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. Genome Biol 2007;8(2):R18.
[33] Kimura M. Evolutionary rate at the molecular level. Nature February 17, 1968;217(5129):624–6.
[34] Kimura M, Ohta T. The average number of generations until fixation of a mutant gene in a finite population. Genetics 1968;61:763–71.
[35] Patwa Z, Wahl LM. The fixation probability of beneficial mutations. J R Soc Interface 2008;5(28):1279–89.
[36] Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, et al. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 2010;467:587–90.
[37] Tobler R, Franssen SU, Kofler R, Orozco-terWengel P, Nolte V, et al. Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. Mol Biol Evol 2014;31:364–75.
[38] Turner TL, Miller PM. Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. Genetics 2012;191:633–642.
[39] Chandler CH. Parallel genome-wide fixation of ancestral alleles in partially outcrossing experimental populations of *Caenorhabditis elegans*. G3 2014;4(9):1657–65.
[40] Chelo IM, Teotonio H. The opportunity for balancing selection in experimental populations of *Caenorhabditis elegans*. Evolution 2013;67:142–56.
[41] Schwartz D. An example of gene fixation resulting from selective advantage in suboptimal conditions. Am Nat 1969;103(933):479–81.
[42] Lee MS, Soubrier J, Edgecombe GD. Rates of phenotypic and genomic evolution during the Cambrian explosion. Curr Biol October 7, 2013;23(19):1889–95.
[43] Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA 1991;88:7160–4.
[44] Lynch M. Evolution of the mutation rate. Trends Genet August 2010;26(8):345–52. http://dx.doi.org/10.1016/j.tig.2010.05.003.
[45] Piganeau G, Eyre-Walker A. Evidence for variation in the effective population size of animal mitochondrial DNA. PLoS One 2009;4:e4396.
[46] Wai T, Teoli D, Shoubridge EA. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. Nat Genet 2008;40:1484–8.
[47] Winn RN, Norris MB, Brayer KJ, Torres C, Muller SL. Detection of mutations in transgenic fish carrying a bacteriophage lambda cII transgene target. Proc Natl Acad Sci USA 2000;97:12655–60.
[48] Edman U, Garcia AM, Busuttil RA, Sorensen D, Lundell M, Kapahi P, et al. Lifespan extension by dietary restriction is not linked to protection against somatic DNA damage in *Drosophila melanogaster*. Aging Cell 2009;8:331–8.
[49] Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nat Genet June 2014;46(6):573–82.
[50] Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W. Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet November 2014;46(11):1160–5.
[51] Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet September 2014;15(9):585–98.
[52] Guo Y, Ye F, Sheng Q, Clark T, Samuels DC. Three-stage quality control strategies for DNA re-sequencing data. Brief Bioinform November 2014;15(6):879–89.
[53] Lynch M, Sung W, Morris K, et al. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA 2008;105:9272–7.
[54] Morton BR, Bi IV, McMullen MD, et al. Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. Genetics 2006;172:569–77.
[55] Lanave C, Tommasi S, Preparata G, et al. Transition and transversion rate in the evolution of animal mitochondrial DNA. Biosystems 1986;19:273–83.
[56] Guo Y, Cai Q, Samuels DC, et al. The use of next generation sequencing technology to study the effect of radiation therapy on mitochondrial DNA mutation. Mutat Res 2012;744:154–60.

[57] Green P, Ewing B, Miller W, Thomas PJ, Green ED. Transcription-associated mutational asymmetry in mammalian evolution. Nat Genet 2003;33:514–7.
[58] Kim SH, Elango N, Warden C, Vigoda E, Yi SV. Heterogeneous genomic molecular clocks in primates. PLoS Genet 2006;2:e163.
[59] Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science January 1, 2010;327(5961):92–4.
[60] Arndt PF, Hwa T, Petrov DA. Substantial regional variation in substitution rates in the human genome: importance of GC content, gene density, and telomere-specific effects. J Mol Evol 2005;60:748–63.
[61] Duret L, Arndt PF. The impact of recombination on nucleotide substitutions in the human genome. PLoS Genet 2008;4(5):e1000071.
[62] Birdsell JA. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Mol Biol Evol 2002;19:1181–97.
[63] Li YC, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 2002;11(12):2453–65.
[64] Schlötterer C, Tautz D. Slippage synthesis of simple sequence DNA. Nucleic Acids Res 1992;20(2):211–5.
[65] Hancock JM. Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: Goldstein D, Schlötterer C, editors. Microsatellites: evolution and applications. New York: Oxford University Press; 1999. p. 1–9.
[66] Hodgkinson A, Ladoukakis E, Eyre-Walker A. Cryptic variation in the human mutation rate. PLoS Biol February 3, 2009;7(2):e1000027.
[67] Cairns J, Overbaugh J, Miller S. The origin of mutants. Nature 1988;335:142–5.
[68] Hall BG, Hartl DL. Regulation of newly evolved enzymes 1. Selection of a novel lactase regulated by lactose in *Escherichia coli*. Genetics 1975;76:391–400.
[69] Blount ZD, Borland CZ, Lenski RE. Inaugural article: Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc Natl Acad Sci USA 2008;105(23):7899–906.
[70] Boyko A, Kathiria P, Zemp FJ, Yao Y, Pogribny I, Kovalchuk I. Transgenerational changes in the genome stability and methylation in pathogen-infected plants: (virus-induced plant genome instability). Nucleic Acids Res 2007;35(5):1714–25.
[71] Boyko A, Kovalchuk I. Epigenetic control of plant stress response. Environ Mol Mutagen January 2008;49(1):61–72.
[72] Boyko A, Kovalchuk I. Genome instability and epigenetic modification–heritable responses to environmental stress?. Curr Opin Plant Biol June 2011;14(3):260–6. http://dx.doi.org/10.1016/j.pbi.2011.03.003. Epub 2011 Mar 26.
[73] Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes Cancer June 2011;2(6):607–17.
[74] Déjardin J. How chromatin prevents genomic rearrangements: locus colocalization induced by transcription factor binding. Bioessays February 2012;34(2):90–3.
[75] Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, et al. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. Nature September 20, 2011;480(7376):245–9.
[76] Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, et al. Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genet 2009;5(6):e1000530.
[77] Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PC. Epigenetic variation in mangrove plants occurring in contrasting natural environment. PLoS One April 26, 2010;5(4):e10326.
[78] Long Y, Xia W, Li R, Wang J, Shao M, Feng J, et al. Epigenetic QTL mapping in *Brassica napus*. Genetics 2011;189:1093–102.
[79] Yi C, Zhang S, Liu X, Bui HT, Hong Y. Does epigenetic polymorphism contribute to phenotypic variances in *Jatropha curcas* L.? BMC Plant Biol November 23, 2010;10:259.
[80] Wu J, Starr S. Low-fidelity compensatory backup alternative DNA repair pathways may unify current carcinogenesis theories. Future Oncol 2014;10(7):1239–53.
[81] Crow JF. The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet October 2000;1(1):40–7.
[82] Drost JB, Lee WR. Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among *Drosophila*, mouse, and human. Environ Mol Mut 1995;25(Suppl. 26):48–64.
[83] Wilkins AS. Genetic analysis of animal development. 2nd ed. NewYork: Wiley-Liss; 1992.
[84] Hoffman PD, Leonard JM, Lindberg GE, Bollmann SR, Hays JB. Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective *Arabidopsis*. Genes Dev 2004;18:2676–85.
[85] Lynch M. The rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci USA 2010;107:961–8.
[86] Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res 2009;19:1195–201.
[87] Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, et al. A genome-wide view of *Caenorhabditis elegans* base substitution mutation processes. Proc Natl Acad Sci USA 2009;106:16310–4.