Wu et al. *Zool. Res.* 2024, 45(4): 767−780
https://doi.org/10.24272/j.issn.2095-8137.2024.058

ZR | Zoological Research

Review Open Access

A new era of mutation rate analyses: Concepts and methods

Kun Wu[1], Danqi Qin[1], Yang Qian[1], Haoxuan Liu[1,*]

[1] *Center for Evolutionary & Organismal Biology and the Fourth Affiliated Hospital of Zhejiang University*, *Zhejiang University School of Medicine*, *Hangzhou*, *Zhejiang* 310058, *China*

ABSTRACT

The mutation rate is a pivotal biological characteristic, intricately governed by natural selection and historically garnering considerable attention. Recent advances in high-throughput sequencing and analytical methodologies have profoundly transformed our understanding in this domain, ushering in an unprecedented era of mutation rate research. This paper aims to provide a comprehensive overview of the key concepts and methodologies frequently employed in the study of mutation rates. It examines various types of mutations, explores the evolutionary dynamics and associated theories, and synthesizes both classical and contemporary hypotheses. Furthermore, this review comprehensively explores recent advances in understanding germline and somatic mutations in animals and offers an overview of experimental methodologies, mutational patterns, molecular mechanisms, and driving forces influencing variations in mutation rates across species and tissues. Finally, it proposes several potential research directions and pressing questions for future investigations.

Keywords: Mutation rate; Somatic mutations; Germline mutations; Animal

INTRODUCTION

In his seminal first chapter of “On the Origin of Species”, Darwin recognized the integral role of natural selection in evolution, deeply intertwined with trait variations among individuals within a population, although he was uncertain about the origins of these variations at the time (Darwin, 1859). Subsequent discoveries revealed that these trait variations predominantly arise from alterations in DNA sequences (Watson & Crick, 1953). Ronald Aylmer Fisher, a pivotal figure in population genetics, mathematically described evolution as shifts in allele frequencies within populations (Fisher, 1930) and identified five key mechanisms influencing these frequencies: mutation, recombination, natural selection, genetic drift, and gene flow.

Among these mechanisms, mutation is recognized as the primary and most fundamental source of genetic variation, providing new genetic material upon which other evolutionary forces can act. Consequently, the mutation rate and the fitness effects of emerging mutations are crucial in understanding genetic variation and evolution. These elements are central to theoretical frameworks in evolutionary biology and quantitative genetics (Barton & Keightley, 2002).

In recent decades, various scholars have provided insightful summaries and comprehensive overviews on mutation and mutation rates (Baer et al., 2007; Halligan & Keightley, 2009; Katju & Bergthorsson, 2019; Lynch, 2010; Lynch et al., 2016; Wang & Obbard, 2023). This review diverges from these previous works by adopting a more accessible, popular science approach, avoiding complex theoretical derivations. Our objective is to bridge the disciplines of biomedical science, structural biology, cell biology, and evolutionary biology and to present fundamental concepts of mutations, classic and modern methods for calculating mutation rates, and emerging methodologies based on whole-genome sequencing (WGS, also known as next-generation sequencing) and long-read third-generation sequencing. We also clarify definitions and types of mutations, theoretical underpinnings, methods, workflows, and considerations in mutation analysis. Finally, we explore the evolutionary patterns of mutation rates in animals and consider future directions in mutation research.

Definitions of genetic mutation, somatic mutation, and germline mutation

In evolutionary biology, a genetic mutation is defined as any change in the genetic material of an organism. For most cellular life and DNA viruses, this involves alterations in DNA sequences. While DNA typically replicates with high fidelity, occasional errors may arise. If these errors are not corrected by cellular repair mechanisms, they result in mutations. In RNA viruses and other organisms that use RNA as their genetic material, mutations constitute changes in the RNA sequence.

In multicellular organisms, the distinction between somatic and germline cells adds complexity to the concept of mutations. Germline mutations, which can be transmitted to subsequent generations, play a pivotal role in evolution (Figure 1A), underpinning various theories and models in evolutionary biology, including those related to inbreeding

Received: 06 April 2024; Accepted: 04 June 2024; Online: 11 June 2024
Foundation items: This work was supported by Zhejiang University
*Corresponding author, E-mail: haoxuan@zju.edu.cn

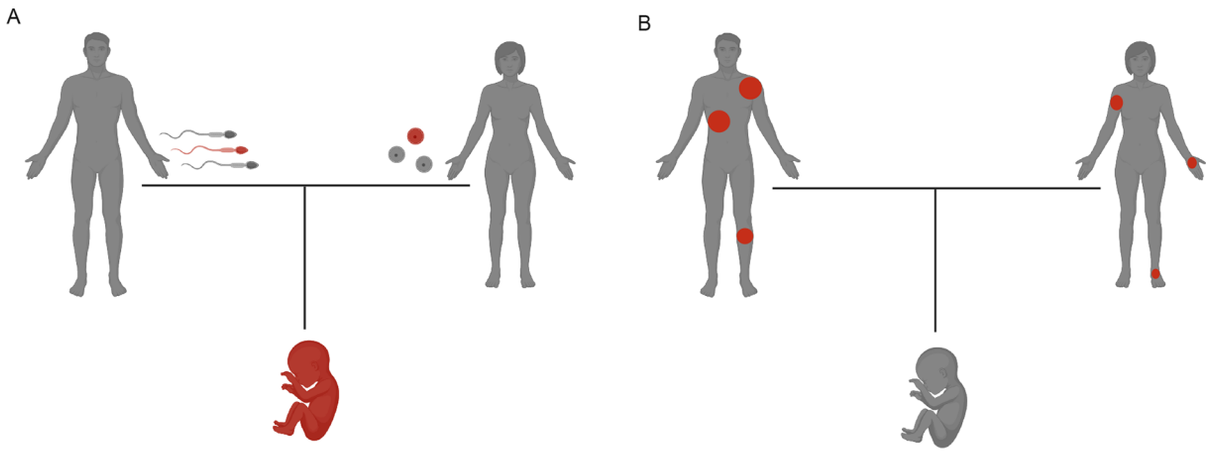

Figure 1 Schematic of germline mutations (A) and somatic mutations (B)

A: Germline mutations. Red sperm and egg cell represent germ cells carrying *de novo* mutations. Through a serendipitous encounter, these cells undergo fertilization and develop into an individual, transmitting newly acquired mutations to the offspring. B: Somatic mutations. Red circles represent somatic mutations occurring in different tissues of the body. Unlike germline mutations, somatic mutations do not get passed on to offspring, but they may contribute to the development of genetic diseases.

depression (Charlesworth et al., 1990), the maintenance of genetic variation (Caballero & Keightley, 1994), and the evolutionary fate of duplicate genes (Lynch, 2002; Lynch & Conery, 2000).

Conversely, somatic mutations, which do not transfer to offspring, can significantly impact the individual (Figure 1B). These mutations are of particular interest in the medical field as they are frequently associated with genetic disorders. Notably, the accumulation of somatic mutations over an individual's lifespan can contribute to cancer and aging (Stratton et al., 2009; Vijg & Dong, 2020). Traditionally, mutations implicated in the onset of cancer were thought to primarily originate from intrinsic processes and environmental factors. However, recent research has suggested that approximately two-thirds of cancer-causing mutations emerge from random errors during DNA replication (Tomasetti et al., 2017). These mutations, which occur independently of environmental influences, are considered unavoidable. As such, understanding these mutations is essential for advancing diagnosis, prediction, and treatment of various mutation-related conditions (Li et al., 2020).

CLASSIFICATIONS OF MUTATIONS

Mutations are diverse and can be classified in several ways. Using the framework established by Hermann Muller, one approach categorizes mutations according to their functional outcomes, known as "Muller's morphs". These include amorphs (loss-of-function mutations leading to product inactivation), hypomorphs (mutations causing partial functional defects), hypermorphs (mutations resulting in more active products), and neomorphs (mutations leading to novel functions). Alternatively, mutations can be categorized based on changes in nucleotides and number of nucleotides involved. Another classification differentiates between spontaneous mutations, arising during a cell's regular life cycle, and induced mutations, induced by external mutagens. This paper will focus on the latter two categorizations, which are more commonly used in contemporary scientific discourse.

Point mutations, indels, and inversions

Point mutations, indels, and inversions constitute the basis of mutation classification (Figure 2). Point mutations involve the replacement of a matching pair of nucleotides in the DNA double helix by another pair, also referred to as single nucleotide substitutions/variations (Figure 2A, B). These mutations can be further categorized as transitions, substitutions between adenine (A) and guanine (G) or cytosine (C) and thymine (T), and transversions, substitutions between pyrimidines and purines (Figure 2A).

Synonymous and nonsynonymous mutations

The distinction between synonymous and nonsynonymous mutations is crucial in evolutionary research. Synonymous mutations, which do not alter the amino acid sequence of a protein, have traditionally been viewed as neutral (Figure 2B). As such, the ratio of nonsynonymous/synonymous mutation rate has served as a key parameter for inferring selection, with ratios >1 indicating positive selection and those <1 indicating negative selection (Goldman & Yang, 1994; Kimura, 1980). However, recent studies have reported that synonymous mutations, while not altering the amino acid sequence encoded by a codon, may still influence protein function by creating new splicing sites, disrupting exonic splicing enhancers (Cartegni et al., 2002), affecting pre-mRNA splicing, mRNA folding, and mRNA degradation, altering transfer RNA (tRNA) translation efficiency (Grantham et al., 1980; Sharp et al., 1988), disrupting binding sites for transcription factor, and altering translational initiation, efficiency, accuracy, and co-translational protein folding, thereby potentially altering protein sequence or abundance (Buschauer et al., 2020; Deng et al., 2022; Kristofich et al., 2018; Shen et al., 2022). For example, while synonymous mutations have traditionally been viewed as nonfunctional in the context of cancer, Supek et al. proposed that these seemingly silent mutations may possess oncogenic potential by influencing transcript splicing and consequently impacting protein function (Supek et al., 2014; Zheng et al., 2014). Notably, recent advancements have revealed that synonymous mutations can markedly impact various biological processes, challenging earlier assumptions of their neutrality (Shen et al., 2022).

Unlike synonymous mutations, nonsynonymous mutations typically exert more profound effects on individuals. Although many nonsynonymous mutations do not significantly alter

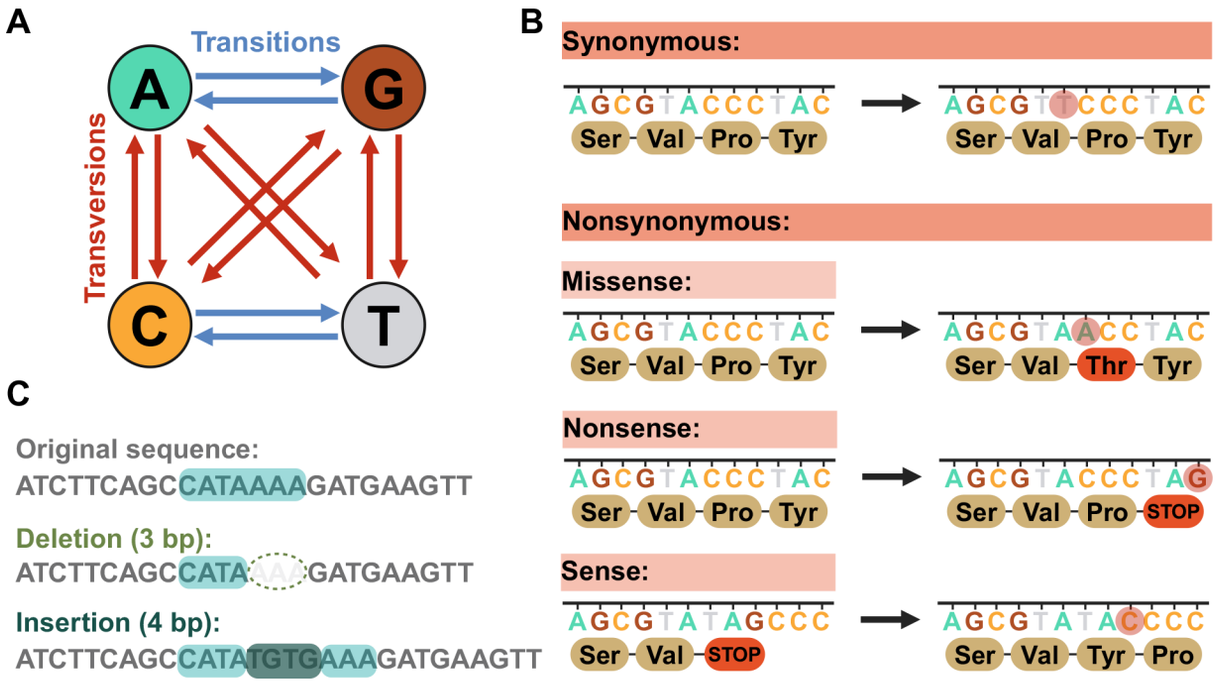

Figure 2 Classification of mutations

A: Transitions and transversions. Due to the complexity of distinguishing specific origins in many analysis scenarios, mutations are typically grouped into six categories for analytical simplicity: A:T→G:C, G:C→A:T, A:T→T:A, A:T→C:G, G:C→C:G, and G:C→T:A. B: Synonymous and nonsynonymous mutations. According to the impact of mutations, nonsynonymous mutations can be further categorized into missense, nonsense, and sense mutations. C: Deletions and insertions.

protein function or cause immediate lethality, they can lead to changes in the amino acid sequence of proteins. Such mutations can be subdivided into missense, nonsense, and sense mutations (Figure 2B) based on their specific impacts at the amino acid level.

Missense mutations alter encoded amino acids (Figure 2B), as exemplified in sickle cell anemia (Hoban et al., 2016; Rees et al., 2010) and amyotrophic lateral sclerosis (ALS), where altered superoxide dismutase (SOD) activity from mutations in the *SOD1* gene leads to misfolded proteins (Urushitani et al., 2006). A missense mutation at the start codon can be particularly damaging, potentially halting protein translation entirely. For instance, a nonsynonymous mutation at the initiation codon of the *HBB* gene can abolish the beta-globin chain, causing severe beta^0-thalassemia, characterized by growth delays, bone deformities, and organ enlargement (Cao & Galanello, 2010; Musallam et al., 2021).

Nonsense mutations convert a regular codon into a stop codon (Figure 2B), truncating the protein and often leading to dysfunctional products. These mutations account for about 20% of genetic diseases caused by mutations in coding regions (Mort et al., 2008), and are implicated in conditions like Hurler syndrome, neurofibromatosis type 1, and Duchenne muscular dystrophy (Li et al., 2019; Osum et al., 2023; Pichavant et al., 2011). While typically detrimental, nonsense mutations can occasionally become fixed in a population (Hahn & Lee, 2006); however, whether these mutations confer new beneficial traits or contribute to adaptive evolution and species-specific features remains a subject of controversy.

Sense mutations convert a stop codon into a regular codon, extending the protein by incorporating the 3' untranslated region (UTR) into the coding sequence (Graur & Li, 2000), resulting in longer proteins due to a shifted termination site, thus classified as elongating mutations (Figure 2B). For instance, an exon in the Metnase gene may potentially originate from the exonization of noncoding sequences induced by a sense mutation (Cordaux et al., 2006).

Indels

Insertions and deletions, collectively referred to as indels due to their shared underlying mechanisms (Figure 2C), arise from various processes, including asymmetric crossover, intra-strand exchange, slippage in repetitive regions, and DNA transposition (Figure 3). Asymmetric crossover may lead to DNA sequence loss on one chromosome and simultaneous insertion on another (Figure 3A). Tandem repeat regions with homologous sequences are particularly susceptible to asymmetric crossovers due to mismatches. Intra-strand exchange, a form of site-specific recombination, frequently leads to DNA deletions when similar segments on a chromosome align and exchange, especially in regions of simple or microsatellite repeats (Figure 3B) (Li et al., 2002).

Indels can vary greatly in size, from massive alterations involving millions of nucleotides to minor changes involving just one nucleotide. During DNA replication, slippage in repetitive sequences can lead to small indels (Figure 3C). In the human genome, areas with simple repeats (such as (CAG/CTG)n) are particularly prone to expansion or contraction (Polleys et al., 2023). Similarly, in single-cell organisms like bacteria or fungi, over 60% of short indel sequences occur in repetitive regions (Katju & Bergthorsson, 2019). Though rarer than single nucleotide mutations, these small indels can profoundly affect biological functions (Audano et al., 2019).

Previous studies have primarily focused on short indels (fragment lengths <50 bp), largely due to the low sensitivity and high false positive rates of short-read sequencing technologies in identifying structural variations (SVs), including

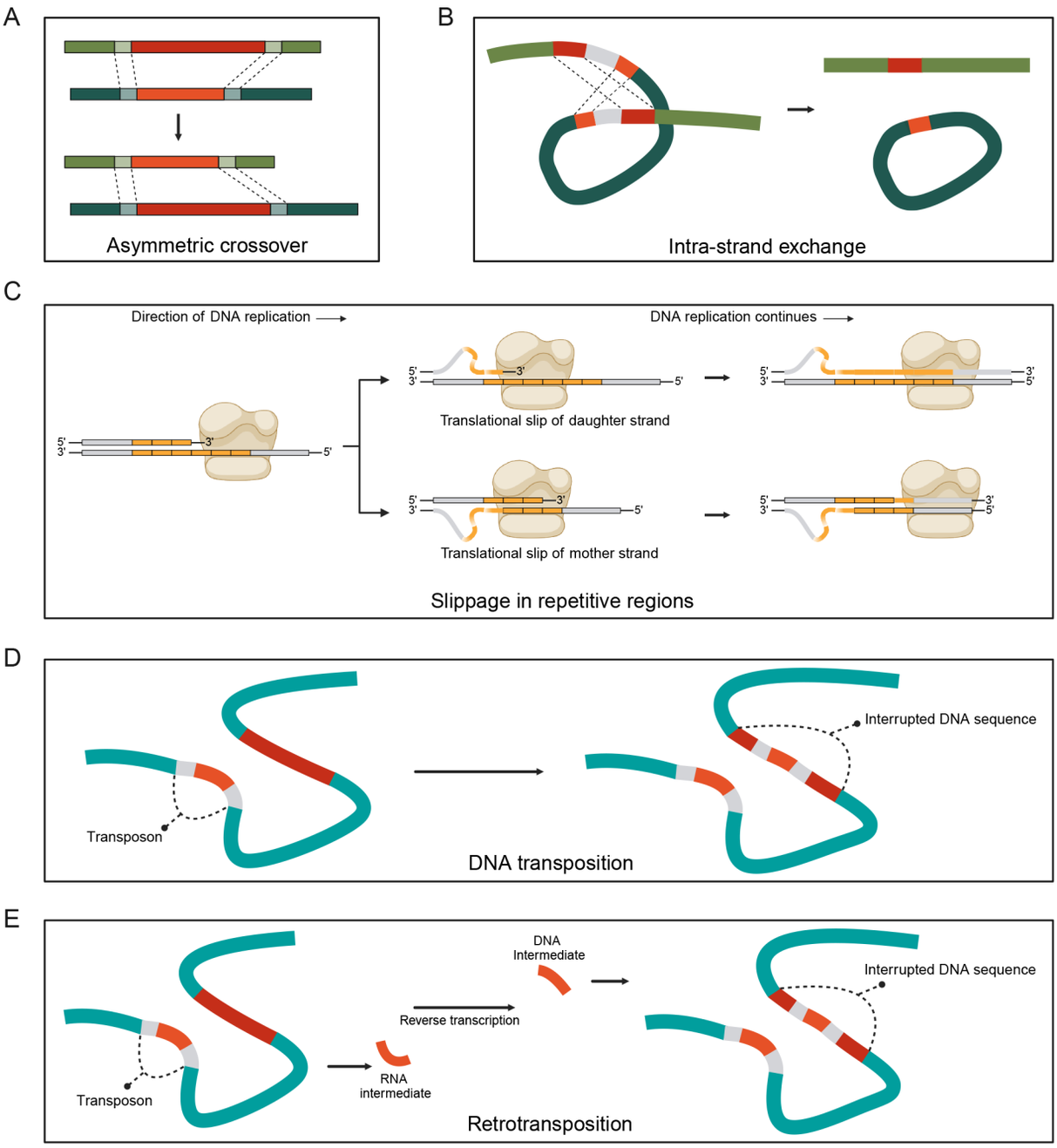

Figure 3 Mechanisms for insertions and deletions

A: Asymmetric crossover. Regions connected by dashed lines represent homologous sequences. Asymmetric crossover events can result in the insertion of a sequence in one strand and deletion in the other strand. B: Intra-strand exchange. Intra-strand crossover events lead to deletion of a sequence fragment. C: Indels caused by slippage in repetitive sequences. During DNA replication, slippage at the 3' end of the DNA strand may occur, and if not repaired, it can lead to the occurrence of insertions or deletions. D: DNA transposition. Transposons utilize transposase to randomly introduce insertions or deletions in the genome. E: RNA-mediated transposition. Transposon undergoes an initial transcription process, with conversion into RNA, followed by reverse transcription where it is transformed back into DNA. Subsequently, this DNA is inserted into a new genomic locus.

large indels, translocations, and DNA- and RNA-mediated transpositions (Figure 3D, E) (Huddleston et al., 2017; Jeffares et al., 2017). Large-scale SVs (>50 bp) and genome rearrangements, such as chromosomal fusion, fission, and polyploidy changes, play critical roles in species diversification or genetic polymorphism and are often associated with phenotypic alterations, including the development of cancer and other genetic disorders (Audano et al., 2019). The development and enhancement of long-read sequencing technologies, such as Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), have opened new avenues for studying SVs. With the ability to sequence millions of base pairs, long-read sequencing enables comprehensive coverage of SVs across the entire genome, thus enhancing the ability to detect large-scale segment alterations, including duplications, deletions, insertions, and translocations.

Inversions

Inversions, a type of DNA rearrangement, occur through

mechanisms such as ectopic recombination, chromosomal breakage, and repair (Feschotte & Pritham, 2007; Gray, 2000). These often involve reciprocal exchange of homologous sequences in opposite directions on a chromosome. Especially prevalent in bacterial and archaeal genomes (Eisen et al., 2000; Suyama & Bork, 2001), inversions can span long DNA segments. Although balanced inversions without genetic information loss do not typically cause phenotypic abnormalities, they can suppress recombination, leading to issues like infertility, cancer, and genetic disorders (Antonacci et al., 2009; Feuk, 2010; Guttenbach et al., 1997). Inversions play a crucial role in both evolution and disease. However, identifying their breakpoints is challenging as they are often embedded within similar segmental duplications (Chaisson et al., 2015). Advanced sequencing technologies, including single-cell DNA template strand sequencing and Hi-C, are improving our ability to understand inversions across pedigrees and lineages (Himmelbach et al., 2018; Porubsky et al., 2020).

Spontaneous and induced mutations

Spontaneous mutations arise from the cumulative errors that occur in DNA during an organism's normal lifecycle (Glickman et al., 1986). Most spontaneous mutations are single base substitutions, frequently resulting from mismatches during the DNA replication process. Typically, adenine (A) pairs with thymine (T) and guanine (G) with cytosine (C), but errors can lead to abnormal pairings like A:C and A:A (Harris et al., 2003; Topal & Fresco, 1976). The primary cause of these mismatches is the isomerization of bases, where DNA bases switch between keto and enol forms (for G and T) and amino and imino forms (for A and C). For example, G may pair with T when in the enol form, or A with C when in the imino form (Topal & Fresco, 1976), leading to spontaneous mutations if these isomerizations occur during replication. Furthermore, adenine can undergo deamination to form hypoxanthine, which pairs with cytosine, leading to an A to G transition (Figure 4A) (Jung et al., 2020).

Spontaneous mutations are crucial in studies related to spontaneous carcinogenesis, aging, and evolution (Kirkwood, 1989; Long et al., 2016; Totter, 1980). Initial research on spontaneous mutations was limited, with most insights derived from control groups in experiments focusing on induced mutations (Sargentini & Smith, 1985). In 1971, Hartman explored and classified spontaneous mutations, identifying single base substitutions, small-scale indels, and large-scale SVs (Hartman et al., 1971). However, the techniques available at the time, which primarily involved marker gene construction and phenotype assays, significantly underestimated the diversity and frequency of mutations. With the development of DNA sequencing technologies such as Sanger sequencing, more comprehensive mutation frequency data have been documented across various species, including yeast, *Escherichia coli*, and *Salmonella* (Albertini et al., 1982; Giroux et al., 1988; O'Hara & Marnett, 1991). However, earlier studies were limited by their focus on marker genes and the typically deleterious nature of spontaneous mutations, which often disappeared during culture passages, resulting in less accurate mutation data. In the 21st century, significant advancements and cost reductions in WGS have revolutionized the field, enabling researchers to merge mutation accumulation experiments with high-throughput sequencing to explore spontaneous mutations across a wide range of organisms, from viruses and single-cell microorganisms like bacteria, fungi, and protists, to more complex multicellular entities. This comprehensive and integrated approach has greatly enhanced the precision and reliability of spontaneous mutation evaluations, facilitating a thorough understanding of mutation rates and patterns across diverse life forms.

In addition to spontaneous mutations that occur as part of their natural life cycle, organisms are subjected to induced mutations from environmental and artificial factors, such as temperature, ultraviolet (UV) radiation, pH (Brash & Haseltine, 1982; Strauss et al., 2017; Witkin, 1953), mutagens, antibiotics, oxidizing agents, and other toxic substances (Bjelland, 2003; Long et al., 2016; Wu et al., 2023). These factors can directly or indirectly alter DNA due to their persistent effects, with the oxidation of guanine to form 8-oxoG, leading to a G to T mutation, noted as a common type of mutation (Figure 4B) (Tchou et al., 1991). Among physical factors, the deamination of nucleotides due to increased temperature is also considered common. For example, heating can cause cytosine to deaminate, transforming into uracil, a base typically found in RNA. If this uracil pairs with adenine and is not repaired before DNA replication, it can result in a C:G to U:A mutation (Graur & Li, 2000). Similarly, the deamination of methylated cytosine can directly produce thymine (Figure 4C). While cytosine to uracil changes are often recognized and corrected by DNA repair mechanisms, the conversion of methylated cytosine to thymine is more challenging to detect, making it a prevalent source of mutations in CG-rich regions of the genome (Graur & Li, 2000). Additionally, mutagenic UV radiation can induce covalent bonds between neighboring pyrimidines in DNA, leading to structural distortions and potential breakage, further contributing to the mutational burden (Brash & Haseltine, 1982).

METHODS FOR EVALUATING MUTATION RATES

Mutations are the fundamental source of genetic variations and are therefore critical for the study of genetics and evolutionary biology. However, quantifying mutation rates is a formidable task. In living organisms, mutation rates are exceedingly low —for example, the single-nucleotide substitution rate in *Escherichia coli* is approximately 2×10^{-10} per site per cell division. Moreover, most mutations are deleterious and tend to be eliminated by natural selection, rendering their observation at the population level particularly challenging. Historically, four principal approaches have been employed to assess mutation rates: theoretical estimation (neutral theory), reporter-based (Luria-Delbrück experiment), and sequencing-based methods (parent-progeny sequencing and mutation accumulation).

Estimation based on substitution rate in selectively neutral regions

Originating from Kimura's neutral theory (Kimura, 1968), this approach posits that the mutation rate at neutral sites is equivalent to the substitution rate of neutral alleles (Figure 5A). Substitution involves one allele replacing another within a population. Kimura's diffusion approximations elucidate the fixation probability of an allele (P) as:

$$P = \frac{1 - e^{-4N_e sq}}{1 - e^{-4N_e s}} \quad (1)$$

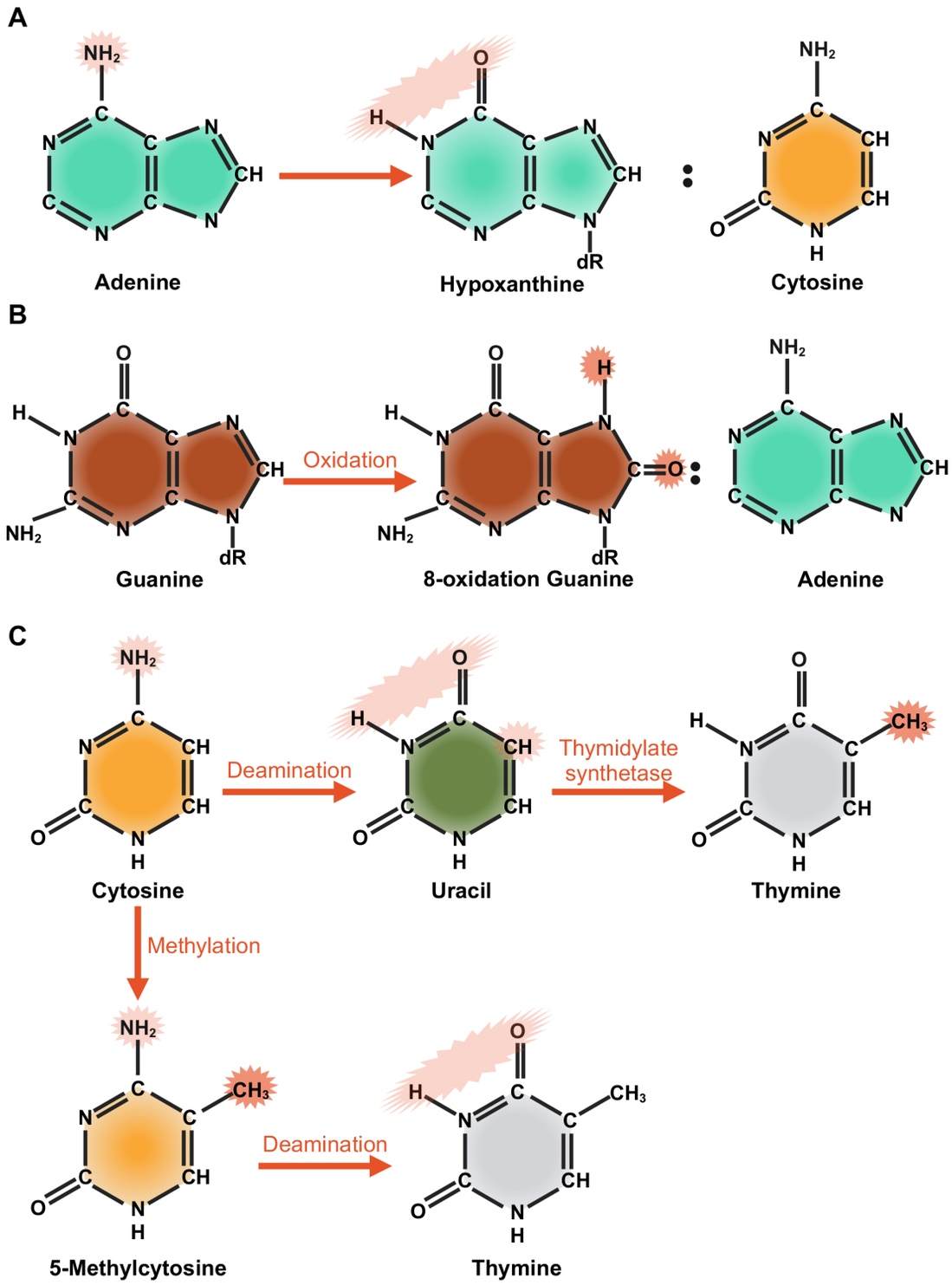

Figure 4 Molecular mechanisms of single base substitutions

A: Deamination of A leads to formation of hypoxanthine, which then pairs with C, causing an A to G mutation. B: Guanine oxidation leads to formation of 8-oxoG (8-oxo-7,8-dihydroguanine), which pairs with A, followed by replication, resulting in the replacement of G by T. C: Cytosine undergoes deamination to form uracil, while methylated cytosine directly undergoes deamination to form thymine.

where q is the initial frequency of an allele, N_e is the effective population size, and s is the selective advantage. In a diploid population of N individuals, the allele frequency of a newly arisen mutation is. If x is very small, $e^{-x} \approx 1-x$. Under neutral conditions (s=0) and replacing q with $\frac{1}{2N}$, the former equation can be simplified as:

$$P = \frac{1}{2N} \quad (2)$$

For neutral alleles, in a diploid population of N individuals, the number of mutations generated at a locus is $2N\mu$, where μ is the mutation rate per gene per generation. If the probability of neutral mutations fixed in a population equals $\frac{1}{2N}$, the rate of substitution of neutral alleles (K) equals the total mutation number multiplied by the fixation probability:

$$K = 2N\mu \times \frac{1}{2N} = \mu \quad (3)$$

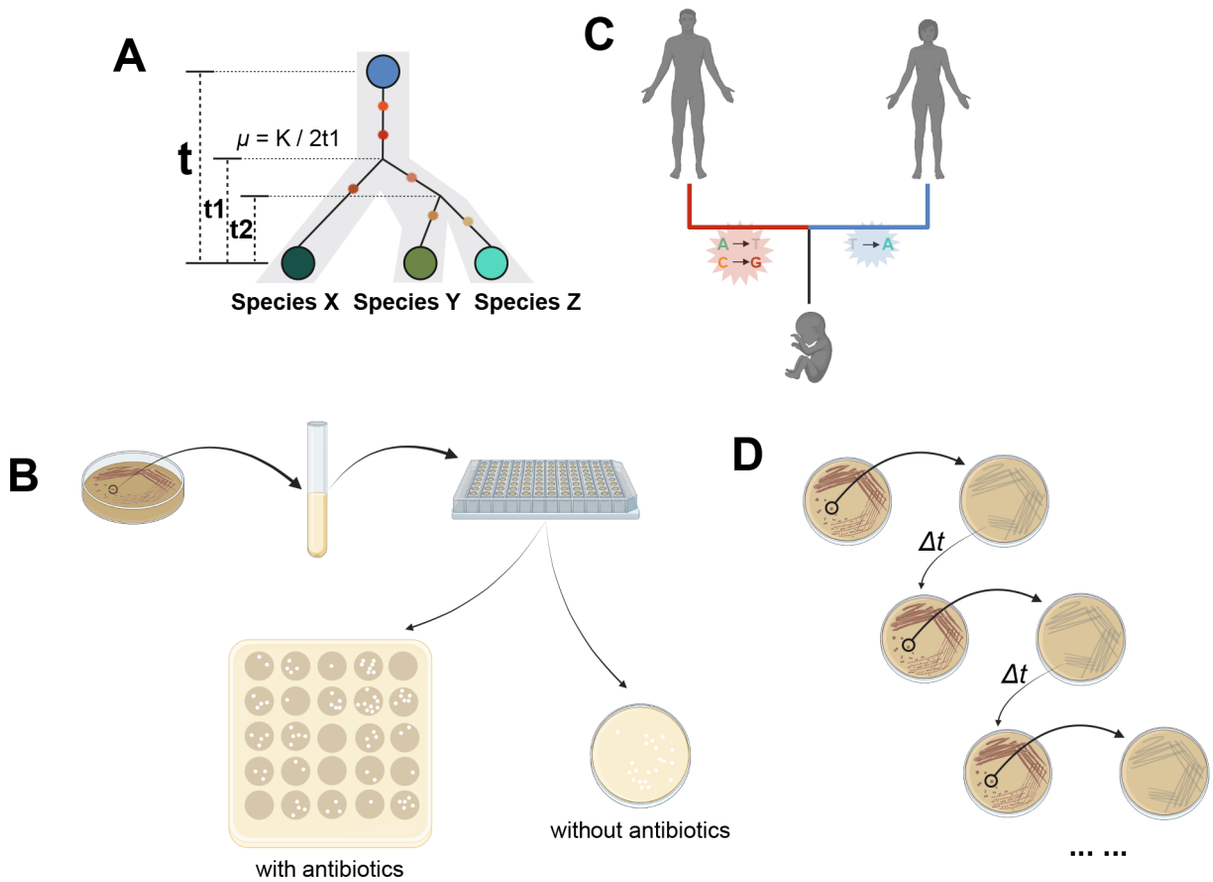

Figure 5 Methods for evaluating mutation rates

A: Estimation based on substitution rate in selectively neutral regions. Over time t, three species, X, Y, and Z, evolved. Colored dots along branches represent mutations that occurred during this process. Neutral theory suggests that polymorphisms within a population and substitutions between species are largely neutral, thus, their rate and divergent time can be used to infer the raw mutation rate. In μ=K/2t1, μ represents the mutation rate, K is the number of neutral substitutions between species X and Z, and t1 is the divergence time between two species. B: Luria-Delbrück experiment. C: Parent-offspring sequencing. Base substitutions in the shaded regions represent mutations. D: Mutation accumulation experiment.

This method has been used to estimate mutation rates in humans. Notably, comparing the sequences of pseudogenes between humans and closely related species revealed a mutation rate of 2.5×10^{-8} mutations per nucleotide site or 175 mutations per diploid genome per generation (Nachman & Crowell, 2000), within the same order as the estimation from direct parent-offspring sequencing (1.2×10^{-8}) (Milholland et al., 2017).

Luria-Delbrück experiment

In the 1940s, before the identification of DNA as the source of genetic material, the nature of mutations was a burning question: Did they pre-exist before selection, or were they adaptively induced during or after selection? Through pioneering experiments, Luria and Delbrück formulated a canonical experiment and a mathematical model that demonstrated genetic variations arise spontaneously rather than adaptively (Luria & Delbrück, 1943), a concept that earned them the 1969 Nobel Prize in Physiology or Medicine. The Luria-Delbrück model has since become a pivotal tool for estimating mutation rates in single-celled organisms and viruses.

In the typical fluctuation test, a small batch of cells from the same ancestor is cultured in parallel until reaching a specified density, during which mutations may occur. These cells are then plated on selective media to allow mutants with resistance mutations to form colonies (Figure 5B). By counting these colonies and evaluating the initial and total cell numbers, researchers can derive an accurate mutation rate estimate. Several methods exist for this, including the p_0 method, Luria-Delbrück's method of the mean, and Lea-Coulson method of the median, each based on specific assumptions (Drake, 1991; Lea & Coulson, 1949; Luria & Delbrück, 1943). To maximize experimental accuracy, certain conditions must be met, including consistent growth rates for mutants and non-mutants and negligible reverse mutations. With advancements in sequencing technologies, the fluctuation test remains a foundational method for investigating spontaneous mutations, offering precise mutation rates and spectra (Jiang et al., 2021).

Parent-offspring sequencing

Parent-offspring or trio sequencing offers the most direct approach to infer mutation rates by analyzing genetic differences between parents and offspring (Figure 5C). This approach calculates mutation rates by tallying the number of mutations over a specific time scale, from years to generations. With the advancement of WGS, this method has been applied to many species, especially those with large genome sizes, such as humans, mice (Milholland et al., 2017), and *Arabidopsis* (Xue et al., 2009; Yang et al., 2015). Parent-offspring sequencing involves sequencing both parents and the affected individual to study genetic disorders, complex diseases, and other conditions related to gene mutations. By

comparing the genome sequences of parents and progeny, scientists can identify genetic mutations that cause specific diseases and traits in offspring (Lelieveld et al., 2016).

Mutation accumulation experiments

The fate of mutations in a population is influenced by the mutation rate as well as evolutionary forces such as natural selection and genetic drift (Halligan & Keightley, 2009; Kimura, 1968; Long et al., 2016; Otto & Michalakis, 1998). The deleterious nature of most non-neutral mutations presents a challenge as they are quickly eliminated from natural populations, making it difficult to glean useful information (Drake, 2006; Keightley & Eyre-Walker, 1999). To overcome this, Muller conceptualized various mutation accumulation experiments in 1928, which were later refined by Mukai and Ohnishi (Mukai, 1964; Mukai et al., 1972; Ohnishi, 1977a, 1977b, 1977c). While these experiments are conceptually straightforward, the practical workload is substantial, requiring independent culturing of numerous sublines, all originating from identical inbred ancestors, with each subline undergoing periodic bottlenecks, which reduce selection effectiveness, thereby inducing genetic drift and the accumulation of non-lethal mutations. For example, in microbial systems, this bottleneck can be achieved by streaking individual colonies onto agar plates, with each colony beginning from a single cell (Figure 5D). By employing WGS and mutation calling, researchers can pinpoint specific mutation sites and facilitate analysis of mutation rates and spectra. Given their comprehensive and accurate approach, mutation accumulation experiments in conjunction with WGS are widely acknowledged as one of the most precise methodologies for determining mutation rates.

The four approaches mentioned above each possess unique specializations and data requirements. One of the most concise and elegant conclusions of Kimura's neutral theory is that "the rate of substitution of neutral alleles (*K*) equals the neutral mutation rate", establishing a crucial link between molecular evolution and mutation rates. This concept has been widely applied in various evolutionary fields, including population genomics and molecular evolutionary theory. Mutation rates derived through Kimura's neutral theory align closely with those obtained from modern methods, underscoring its remarkable efficacy. However, as it does not account for certain variables, such as environmental variations, it may not meet the precision requirements for tasks demanding high accuracy. While the fluctuation test is convenient, precise, and efficient at estimating mutation rates, its applicability is restricted to bacteria, single-celled fungi, and viruses, and confined to estimating mutation rates for single loci within the genome. Trio sequencing is among the most direct, efficient, and accurate methods for estimating mutation rates in multicellular organisms, although it is limited to species with large genome sizes, such as vertebrates. For organisms with smaller genomes and shorter generation times, such as bacteria, fungi, fruit flies, and nematodes, mutation accumulation experiments are invaluable. These experiments facilitate the accumulation of a relatively high number of mutations, making mutation rate estimation more cost-effective.

VARIATIONS IN GERMLINE AND SOMATIC MUTATION RATES ACROSS THE ANIMAL KINGDOM

In the animal kingdom, both germ and somatic cells are susceptible to mutations. Yet, due to their potential for inheritance, germ cell mutations have garnered more attention. Estimating mutation rates is crucial in evolutionary studies, and recent years have seen an increased focus on understanding germline mutation rates and their evolution through analyses within and across species (Bergeron et al., 2022, 2023). To date, WGS has directly measured the germline mutation rate in at least 10 invertebrates and 80 vertebrates (Liu et al., 2023). These rates are typically expressed as number of mutations per site per generation and number of mutations per site per year, with generation time, body mass, and genome size being the strongest predictors of these rates (Figure 6A–F; Supplementary Table S1).

Conversely, somatic mutations in animals have been less studied due to technical challenges. Somatic mutations accumulate as cells divide and the organism develops, resulting in a unique mutational landscape across different tissues and cells (Dou et al., 2018). The identification of these mutations often necessitates single-cell sequencing, which currently lacks the accuracy required for such precise tasks. Consequently, most research in this area has been limited to specific organs, such as intestinal crypts, which originate from a single cell. Recent research sequenced hundreds of intestinal crypts from 16 species, revealing significant differences in mutation numbers between germline and somatic mutations (Cagan et al., 2022). For example, a 40-year-old human may accumulate about 1 500 somatic mutations in one intestinal crypt, compared to only around 40 in a germ cell. There is a near-perfect linear correlation between the accumulation of somatic mutations and age, suggesting a time-dependent mechanism for mutagenic effects. Furthermore, across species, the rate of somatic mutations (per sample per year) shows a correlation with generation time, body mass, and genome size (Figure 6G–L; Supplementary Table S2).

Variations in somatic mutation rates across different tissues in humans

Somatic mutations, which accumulate within normal tissues, play pivotal roles in aging, cancer, and various diseases (Stratton et al., 2009; Vijg & Dong, 2020). Both aging and cancer are associated with mutation accumulation over time, precipitating a marked decline in cellular function. However, detecting somatic mutations poses considerable challenges, primarily due to genetic heterogeneity and low allele frequencies in affected cells (Huddleston et al., 2017; Vijg & Dong, 2020). Genetic heterogeneity, or somatic mosaicism, arises from variations in postzygotic mutations across different cells and tissues. These mutations, unique to each cell in every individual, are typically present at low allele frequency in bulk sequencing, necessitating sequencing with sufficient depth and accuracy for detection. Various methods have been developed in recent years to address these challenges, including single-cell *ex vivo* expansion followed by sequencing (Bae et al., 2018), high-depth sequencing of microdissected tissues or single-cell-derived cell clusters (Li et al., 2021), single-cell sequencing (Huang et al., 2022), and duplex sequencing (Abascal et al., 2021). Although differences in sample collection and sequencing methodologies may hinder direct comparison, several discernible trends have emerged with sufficient data collection (Figure 7).

The somatic mutation rates across different tissues in humans, collated from three studies sampling multiple somatic

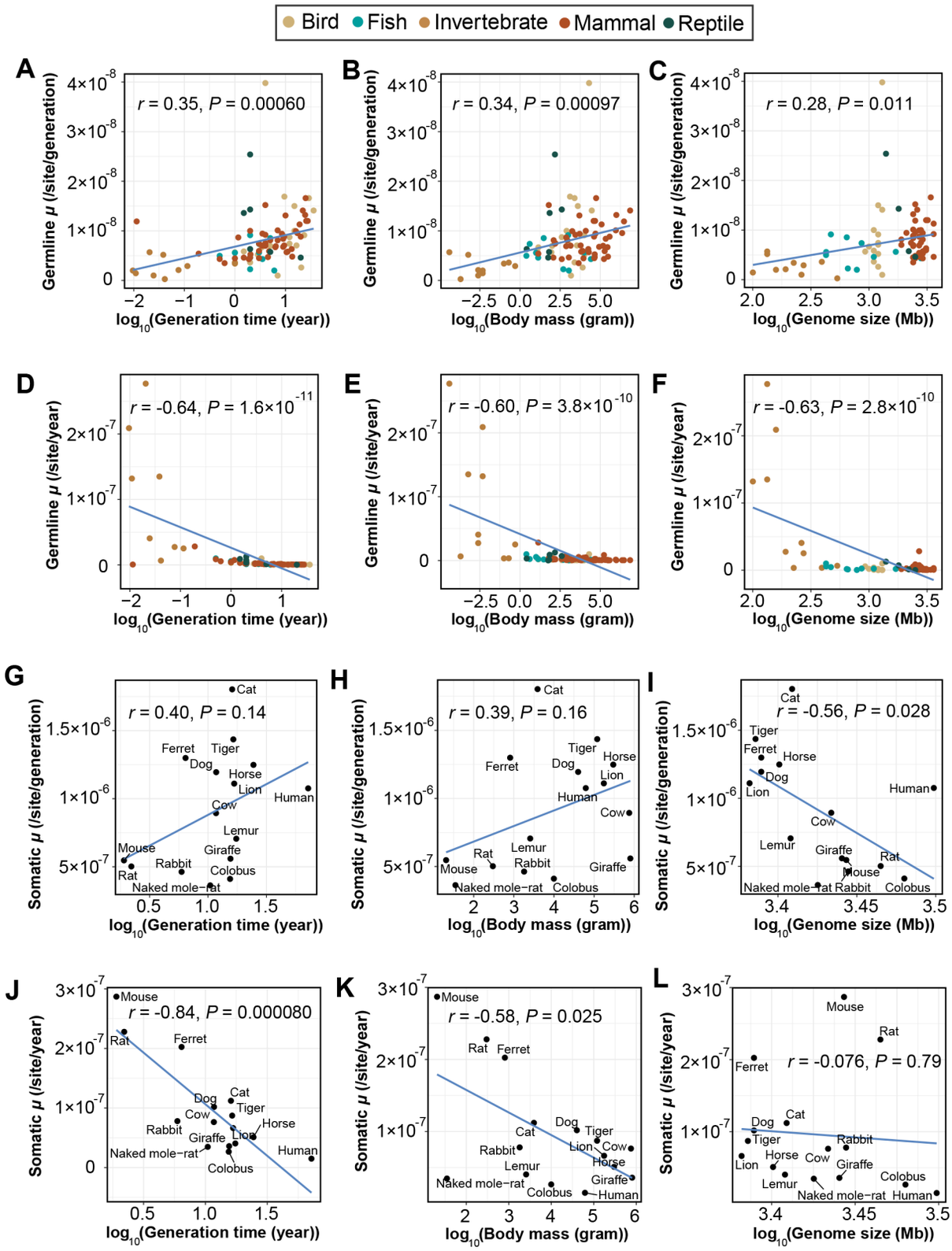

Figure 6 Correlations between germline and somatic mutation rates (μ) with generation time, body mass, and genome size

Each dot represents a specific species, and blue lines represent linear regression. Pearson's *r* and *P* values between X-axis and Y-axis are annotated in each panel. Data sources are referenced in Supplementary Tables S1 and S2.

tissues, are summarized in Figure 7. Moore et al. (2021) conducted WGS on hundreds of microdissected samples from dozens of donors, revealing significant within-organ heterogeneity in mutational load. Notably, in a 47-year-old donor, the mutation rates in skin tissues ranged from 7.6×10^{-11} to 1.9×10^{-8} per site per year, whereas rates in stomach tissues varied by less than 3%. Furthermore, mutation rates varied substantially between tissues, with the lowest rate observed in heart tissue (averaging 8.0×10^{-12} per site per year) and the highest in appendix tissue (averaging 1.0×10^{-8} per site per year) (Figure 7). Abascal et al. (2021) applied a low-error-rate duplex sequencing protocol to analyze different cell types and

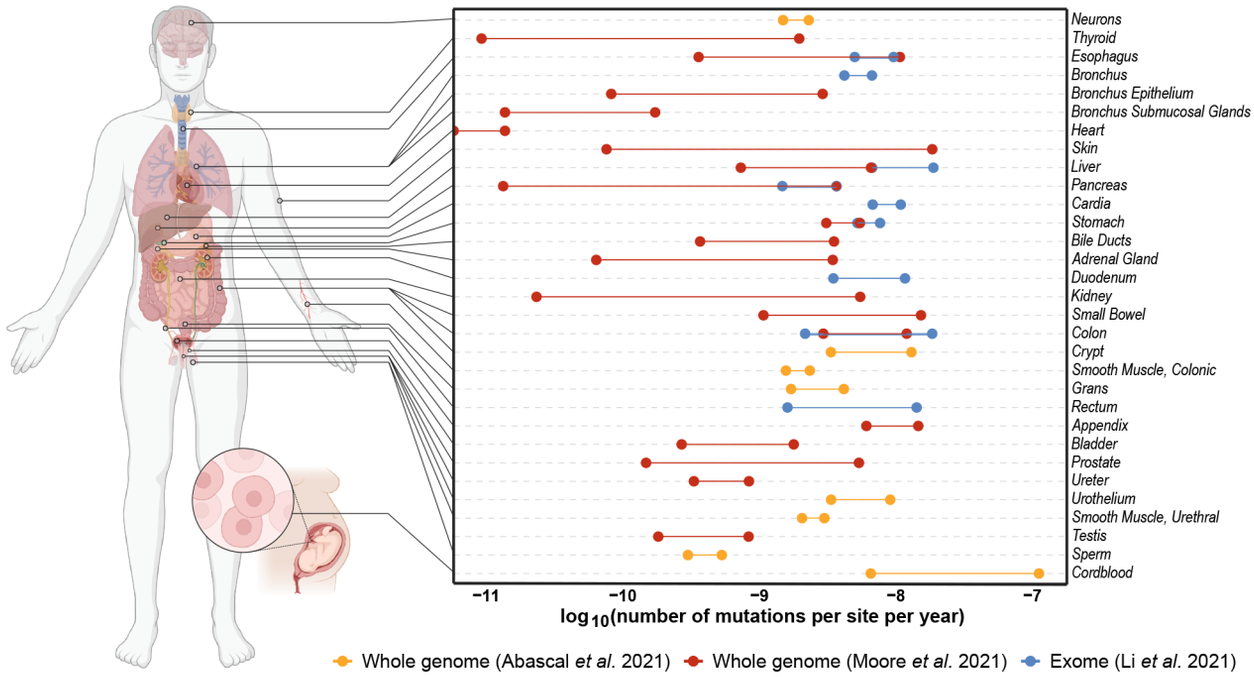

Figure 7 Somatic mutation rates across different tissues in humans

Estimated minimal and maximal mutation rates of 31 tissues are shown in the plot panel. Different colors represent different data sources. Age of cord blood was set to 1. For detailed data and references, see Supplementary Table S3.

identified the lowest mutation rate in sperm (Figure 7), consistent with low germline mutation rates identified in parent-offspring studies (Bergeron et al., 2023). Li et al. (2021) performed exome sequencing of 1 737 microdissected tissues from nine organs across five donors, with their findings largely aligning with the aforementioned WGS studies (Figure 7).

Underlying causes of differences in somatic mutation rates across tissues

Variations in mutational burden across tissues may stem from multiple intrinsic and environmental factors. Mutation directly impacts the genomes of somatic cells, involving factors such as DNA replication timing, chromatin structure, and gene expression levels (Haradhvala et al., 2016; Polak et al., 2015). Polak et al. (2015) observed asymmetrical distribution of cancer mutations between the two DNA strands, possibly linked to mutations associated with DNA replication and transcription. Moreover, different tissues possess distinct developmental, environmental, and functional characteristics, which can contribute to variation in mutation rates. Early research, mainly on blood cells, indicated that the mutational burden in these cells is primarily influenced by immune-driven somatic mutations, ultimately driving enhanced cell renewal and clonal expansion (Xie et al., 2014).

An alternative hypothesis suggests that the number of somatic cell divisions is correlated with mutation accumulation, as more cell divisions may lead to greater cell-division-dependent mutations. However, accurately estimating the number of cell divisions for each tissue is technologically challenging, and direct evidence is lacking. Contrary to expectations, Abascal et al. (2021) observed that the number of cell divisions may not be the sole determining factor affecting mutation accumulation in some tissues. Notably, despite mature blood cells undergoing significantly more divisions, granulocytes and colonic epithelial cells exhibited comparable mutational burdens and signatures. Additionally, post-mitotic neurons and polyclonal smooth muscle cells accumulated mutations throughout their lifespan, even in the absence of cell division, at rates comparable to actively dividing cells.

Lastly, exposure to environmental mutagens is a major determinant of somatic mutational load. For example, smoking tobacco is a well-recognized preventable risk factor for cancer. Alexandrov et al. (2016) evaluated differences in somatic mutational load between smokers and non-smokers using a dataset of 5 243 cancer samples covering 13 tissues and found significantly increased mutational load in smokers in four tissues—lung, larynx, liver, and kidney—characterized by a C>A mutation signature. Sunlight exposure is another significant mutagenic factor. Notably, Martincorena et al. (2015) observed increased mutational burdens and CC>TT dinucleotide substitutions, some leading to clonal expansion, in normal skin tissues upon sequencing. Moreover, UV radiation exposure is also associated with increases in mutations during aging (García-Nieto et al., 2019). *Aristolochia* and related plants, integral to traditional Chinese pharmacopeias, are known to contain aristolochic acids and similar compounds, which significantly increase A:T>T:A mutations in trinucleotide contexts and contribute to the development of multiple cancers, including bladder, upper tract urothelial (Debelle et al., 2008), kidney (Jelaković et al., 2015), intrahepatic cholangiocarcinoma (Zou et al., 2014), and liver cancers (Ng et al., 2017). Notably, previous research has estimated that 47% of liver cancers in mainland China and 78% in Taiwan are associated with aristolochic acid exposure (Ng et al., 2017).

In conclusion, the variability in mutation numbers among tissues can be attributed to multiple interacting factors, including the molecular mechanisms of mutations, cellular development, tissue-specific regulation, and environmental exposure. Importantly, mutations arising from different mechanisms exhibit distinct mutation spectra, which can be used to infer the mutational environment of a cell. For

instance, the mutation spectrum of a virus often closely mirrors that of its host; for example, the highly similar mutation spectra between bats and SARS-CoV-2 suggest bats as the host (Deng et al., 2022), while the SARS-CoV-2 Omicron variant is believed to have originated from mice, as evidenced by their similar mutation spectra (Wei et al., 2021).

FUTURE OUTLOOK

As genetic research continues to lead scientific exploration, the trajectory of mutation studies appears expansive and full of potential. Key areas are emerging that promise to shape and amplify the impact of this field significantly.

Technological advancements, particularly in sequencing technologies such as long-read sequencing and single-cell genomics, are set to revolutionize the accurate detection and characterization of mutations. This evolution is expected to deepen our comprehension of complex mutation patterns and their significance across various biological contexts. For instance, while SVs are prevalent in cancer, their detection using short-read sequencing has posed challenges. In a groundbreaking study, Xu et al. (2023) utilized Nanopore sequencing of colorectal cancer samples, achieving unprecedented accuracy and efficiency in detecting SVs. Such advancements indicate a burgeoning future for similar studies research endeavors.

Expanding somatic mutation studies to encompass diverse tissues and species holds the potential to enrich our understanding of genetic variation, natural selection, and evolutionary dynamics. While germline mutation rates across numerous animal species have been measured directly through WGS, data on somatic mutation rates remains relatively scarce, with only a few cross-species estimates available. This scarcity is particularly evident in studies of between-tissue variations, where currently, comprehensive cross-tissue data exist solely for humans, underscoring the urgent need to expand research on somatic mutations to encompass more species and tissues. With the accumulation of data, opportunities arise to probe the underlying mechanisms driving between-tissue variation in mutation rates. Pertinent questions emerge, such as why the heart has the lowest mutational load among organs, potentially explaining its rarity in cancer occurrences, and what molecular mechanisms underlie the disparities in mutation rates between germline and somatic cells. Such investigations harbor immense biomedical potential, as exemplified by the naked mole rat, which exhibits an exceptionally low somatic mutation rate (Cagan et al., 2022), associated with its reduced cancer incidence and extended lifespan (Oka et al., 2023). Investigating the factors that govern the naked mole rat's unique mutation profile promises novel insights for biomedical research.

Ultimately, the mutation rate is shaped by evolutionary forces, including deleterious mutations, beneficial mutations, and the cost of fidelity (Sniegowski et al., 2000). While theoretical studies abound regarding the evolution of germline mutation rates, discussions on somatic mutation rates are notably lacking. Given that somatic mutations seldom confer benefits, their evolution likely hinges on deleterious mutations and the cost of fidelity. In addition to the rate at which mutations occur, the debate on whether the distribution of mutations is selected remains unresolved, with conflicting perspectives regarding the distribution of mutations between genic and intergenic regions (Liu & Zhang, 2020; Xia et al., 2020) and among genes of varying functional importance (Liu & Zhang, 2022; Monroe et al., 2022). As such, a definitive conclusion on these matters remains elusive, underscoring the enduring complexities inherent in mutation research.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

K.W., D.Q., Y.Q., and H.L. conceived the review, prepared the draft and designed the figures. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank members of the Liu laboratory for valuable comments.

REFERENCES

Abascal F, Harvey LMR, Mitchell E, et al. 2021. Somatic mutation landscapes at single-molecule resolution. *Nature*, **593**(7859): 405–410.

Albertini AM, Hofer M, Calos MP, et al. 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell*, **29**(2): 319–328.

Alexandrov LB, Ju YS, Haase K, et al. 2016. Mutational signatures associated with tobacco smoking in human cancer. *Science*, **354**(6312): 618–622.

Antonacci F, Kidd JM, Marques-Bonet T, et al. 2009. Characterization of six human disease-associated inversion polymorphisms. *Human Molecular Genetics*, **18**(14): 2555–2566.

Audano PA, Sulovari A, Graves-Lindsay TA, et al. 2019. Characterizing the major structural variant alleles of the human genome. *Cell*, **176**(3): 663–675. e19.

Bae T, Tomasini L, Mariani J, et al. 2018. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science*, **359**(6375): 550–555.

Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics*, **8**(8): 619–631.

Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nature Reviews Genetics*, **3**(1): 11–21.

Bergeron LA, Besenbacher S, Turner T, et al. 2022. The Mutationathon highlights the importance of reaching standardization in estimates of pedigree-based germline mutation rates. *eLife*, **11**: e73577.

Bergeron LA, Besenbacher S, Zheng J, et al. 2023. Evolution of the germline mutation rate across vertebrates. *Nature*, **615**(7951): 285–291.

Bjelland S. 2003. Mutagenicity, toxicity and repair of DNA base damage induced by oxidation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **531**(1-2): 37–80.

Brash DE, Haseltine WA. 1982. UV-induced mutation hotspots occur at DNA damage hotspots. *Nature*, **298**(5870): 189–192.

Buschauer R, Matsuo Y, Sugiyama T, et al. 2020. The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science*, **368**(6488): eaay6912.

Caballero A, Keightley PD. 1994. A pleiotropic nonadditive model of variation in quantitative traits. *Genetics*, **138**(3): 883–900.

Cagan A, Baez-Ortega A, Brzozowska N, et al. 2022. Somatic mutation rates scale with lifespan across mammals. *Nature*, **604**(7906): 517–524.

Cao A, Galanello R. 2010. Beta-thalassemia. *Genetics in Medicine*, **12**(2): 61–76.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and

understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics*, **3**(4): 285–298.

Chaisson MJP, Wilson RK, Eichler EE. 2015. Genetic variation and the *de novo* assembly of human genomes. *Nature Reviews Genetics*, **16**(11): 627–640.

Charlesworth B, Charlesworth D, Morgan MT. 1990. Genetic loads and estimates of mutation rates in highly inbred plant populations. *Nature*, **347**(6291): 380–382.

Cordaux R, Udit S, Batzer MA, et al. 2006. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(21): 8101–8106.

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

Debelle FD, Vanherweghem JL, Nortier JL. 2008. Aristolochic acid nephropathy: a worldwide problem. *Kidney International*, **74**(2): 158–169.

Deng SJ, Xing K, He XL. 2022. Mutation signatures inform the natural host of SARS-CoV-2. *National Science Review*, **9**(2): nwab220.

Dou YM, Gold HD, Luquette LJ, et al. 2018. Detecting somatic mutations in normal cells. *Trends in Genetics*, **34**(7): 545–557.

Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences of the United States of America*, **88**(16): 7160–7164.

Drake JW. 2006. Chaos and order in spontaneous mutation. *Genetics*, **173**(1): 1–8.

Eisen JA, Heidelberg JF, White O, et al. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biology*, **1**(6): research0011.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, **41**: 331–368.

Feuk L. 2010. Inversion variants in the human genome: role in disease and genome architecture. *Genome Medicine*, **2**(2): 11.

Fisher R. 1930. The Genetical Theory of Natural Selection. Oxford: Clarendon.

García-Nieto PE, Morrison AJ, Fraser HB. 2019. The somatic mutation landscape of the human body. *Genome Biology*, **20**(1): 298.

Giroux CN, Mis JRA, Pierce MK, et al. 1988. DNA sequence analysis of spontaneous mutations in the *SUP4*-o gene of *Saccharomyces cerevisiae. Molecular and Cellular Biology*, **8**(2): 978–981.

Glickman BW, Burns PA, Fix DF. 1986. Mechanisms of spontaneous mutagenesis: clues from altered mutational specificity in DNA repair-defective strains. *In*: Shankel DM, Hartman PE, Kada T, et al. Antimutagenesis and Anticarcinogenesis Mechanisms. New York: Springer, 259–281.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, **11**(5): 725–736.

Grantham R, Gautier C, Gouy M, et al. 1980. Codon catalog usage and the genome hypothesis. *Nucleic Acids Research*, **8**(1): r49–r62.

Graur D, Li WH. 2000. Fundamentals of Molecular Evolution. 2nd ed. Sunderland: Sinauer Associates.

Gray YHM. 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends in Genetics*, **16**(10): 461–468.

Guttenbach M, Engel W, Schmid M. 1997. Analysis of structural and numerical chromosome abnormalities in sperm of normal men and carriers of constitutional chromosome aberrations. A review. *Human Genetics*, **100**(1): 1–21.

Hahn Y, Lee B. 2006. Human-specific nonsense mutations identified by genome sequence comparisons. *Human Genetics*, **119**(1-2): 169–178.

Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics*, **40**: 151–172.

Haradhvala NJ, Polak P, Stojanov P, et al. 2016. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *Cell*, **164**(3): 538–549.

Harris VH, Smith CL, Jonathan Cummins W, et al. 2003. The effect of tautomeric constant on the specificity of nucleotide incorporation during DNA replication: support for the rare tautomer hypothesis of substitution mutagenesis. *Journal of Molecular Biology*, **326**(5): 1389–1401.

Hartman PE, Hartman Z, Stahl RC, et al. 1971. Classification and mapping of spontaneous and induced mutations in the histidine operon of *Salmonella. Advances in Genetics*, **16**: 1–34.

Himmelbach A, Ruban A, Walde I, et al. 2018. Discovery of multi-megabase polymorphic inversions by chromosome conformation capture sequencing in large-genome plant species. *The Plant Journal*, **96**(6): 1309–1316.

Hoban MD, Orkin SH, Bauer DE. 2016. Genetic treatment of a molecular disorder: gene therapy approaches to sickle cell disease. *Blood*, **127**(7): 839–848.

Huang ZQ, Sun SX, Lee M, et al. 2022. Single-cell analysis of somatic mutations in human bronchial epithelial cells in relation to aging and smoking. *Nature Genetics*, **54**(4): 492–498.

Huddleston J, Chaisson MJP, Steinberg KM, et al. 2017. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Research*, **27**(5): 677–685.

Jeffares DC, Jolly C, Hoti M, et al. 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, **8**: 14061.

Jelaković B, Castells X, Tomić K, et al. 2015. Renal cell carcinomas of chronic kidney disease patients harbor the mutational signature of carcinogenic aristolochic acid. *International Journal of Cancer*, **136**(12): 2967–2972.

Jiang PY, Ollodart AR, Sudhesh V, et al. 2021. A modified fluctuation assay reveals a natural mutator phenotype that drives mutation spectrum variation within *Saccharomyces cerevisiae. eLife*, **10**: e68285.

Jung H, Hawkins MA, Lee S. 2020. Structural insights into the bypass of the major deaminated purines by translesion synthesis DNA polymerase. *Biochemical Journal*, **477**(24): 4797–4810.

Katju V, Bergthorsson U. 2019. Old trade, new tricks: insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-throughput genomic approaches. *Genome Biology and Evolution*, **11**(1): 136–165.

Keightley PD, Eyre-Walker A. 1999. Terumi mukai and the riddle of deleterious mutation rates. *Genetics*, **153**(2): 515–523.

Kimura M. 1968. Evolutionary rate at the molecular level. *Nature*, **217**(5129): 624–626.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**(2): 111–120.

Kirkwood TBL. 1989. DNA, mutations and aging. *Mutation Research/DNAging*, **219**(1): 1–7.

Kristofich J, Morgenthaler AB, Kinney WR, et al. 2018. Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLoS Genetics*, **14**(8): e1007615.

Lea DE, Coulson CA. 1949. The distribution of the numbers of mutants in bacterial populations. *Journal of Genetics*, **49**(3): 264–285.

Lelieveld SH, Reijnders MRF, Pfundt R, et al. 2016. Meta-analysis of 2, 104 trios provides support for 10 new genes for intellectual disability. *Nature Neuroscience*, **19**(9): 1194–1196.

Li HY, Yang Y, Hong WQ, et al. 2020. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms,

advances and prospects. *Signal Transduction and Targeted Therapy*, **5**(1): 1.

Li RY, Di L, Li J, et al. 2021. A body map of somatic mutagenesis in morphologically normal human tissues. *Nature*, **597**(7876): 398–403.

Li YC, Korol AB, Fahima T, et al. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, **11**(12): 2453–2465.

Li YJ, Tang XY, Meng Y, et al. 2019. A novel compound mutation in alpha-L-iduronidase gene causes mucopolysaccharidosis type I. *Journal of Genetics*, **98**(3): 65.

Liu HX, Sun MY, Zhang JZ. 2023. Genomic estimates of mutation and substitution rates contradict the evolutionary speed hypothesis of the latitudinal diversity gradient. *Proceedings of the Royal Society B: Biological Sciences*, **290**(2009): 20231787.

Liu HX, Zhang JZ. 2020. Higher germline mutagenesis of genes with stronger testis expressions refutes the transcriptional scanning hypothesis. *Molecular Biology and Evolution*, **37**(11): 3225–3231.

Liu HX, Zhang JZ. 2022. Is the mutation rate lower in genomic regions of stronger selective constraints?. *Molecular Biology and Evolution*, **39**(8): msac169.

Long H, Miller SF, Strauss C, et al. 2016. Antibiotic treatment enhances the genome-wide mutation rate of target cells. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(18): E2498–E2505.

Luria SE, Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, **28**(6): 491–511.

Lynch M. 2002. Gene duplication and evolution. *Science*, **297**(5583): 945–947.

Lynch M. 2010. Evolution of the mutation rate. *Trends in Genetics*, **26**(8): 345–352.

Lynch M, Ackerman MS, Gout JF, et al. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, **17**(11): 704–714.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*, **290**(5494): 1151–1155.

Martincorena I, Roshan A, Gerstung M, et al. 2015. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, **348**(6237): 880–886.

Milholland B, Dong X, Zhang L, et al. 2017. Differences between germline and somatic mutation rates in humans and mice. *Nature Communications*, **8**: 15183.

Monroe JG, Srikant T, Carbonell-Bejerano P, et al. 2022. Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature*, **602**(7895): 101–105.

Moore L, Cagan A, Coorens THH, et al. 2021. The mutational landscape of human somatic and germline cells. *Nature*, **597**(7876): 381–386.

Mort M, Ivanov D, Cooper DN, et al. 2008. A meta-analysis of nonsense mutations causing human genetic disease. *Human Mutation*, **29**(8): 1037–1047.

Mukai T. 1964. The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics*, **50**(1): 1–19.

Mukai T, Chigusa SI, Mettler LE, et al. 1972. Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster. Genetics*, **72**(2): 335–355.

Musallam KM, Bou-Fakhredin R, Cappellini MD, et al. 2021. 2021 update on clinical trials in β-thalassemia. *American Journal of Hematology*, **96**(11): 1518–1531.

Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics*, **156**(1): 297–304.

Ng AWT, Poon SL, Huang MN, et al. 2017. Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Science Translational Medicine*, **9**(412): eaan6446.

O'Hara SM, Marnett LJ. 1991. DNA sequence analysis of spontaneous and β-methoxy-acrolein-induced mutations in *Salmonella typhimurium hisD 3052. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **247**(1): 45–56.

Ohnishi O. 1977a. Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. I. Recessive lethal mutations. *Genetics*, **87**(3): 519–527.

Ohnishi O. 1977b. Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effect of polygenic mutations. *Genetics*, **87**(3): 529–545.

Ohnishi O. 1977c. Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. III. Heterozygous effect of polygenic mutations. *Genetics*, **87**(3): 547–556.

Oka K, Yamakawa M, Kawamura Y, et al. 2023. The naked mole-rat as a model for healthy aging. *Annual Review of Animal Biosciences*, **11**: 207–226.

Osum SH, Oribamise EI, Corbière SMAS, et al. 2023. Combining nonsense mutation suppression therapy with nonsense-mediated decay inhibition in neurofibromatosis type 1. *Molecular Therapy-Nucleic Acids*, **33**: 227–239.

Otto SP, Michalakis Y. 1998. The evolution of recombination in changing environments. *Trends in Ecology & Evolution*, **13**(4): 145–151.

Pichavant C, Aartsma-Rus A, Clemens PR, et al. 2011. Current status of pharmaceutical and genetic therapeutic approaches to treat DMD. *Molecular Therapy*, **19**(5): 830–840.

Polak P, Karlić R, Koren A, et al. 2015. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature*, **518**(7539): 360–364.

Polleys EJ, Del Priore I, Haber JE, et al. 2023. Structure-forming CAG/CTG repeats interfere with gap repair to cause repeat expansions and chromosome breaks. *Nature Communications*, **14**(1): 2469.

Porubsky D, Sanders AD, Höps W, et al. 2020. Recurrent inversion toggling and great ape genome evolution. *Nature Genetics*, **52**(8): 849–858.

Rees DC, Williams TN, Gladwin MT. 2010. Sickle-cell disease. *The Lancet*, **376**(9757): 2018–2031.

Sargentini NJ, Smith KC. 1985. Spontaneous mutagenesis: the roles of DNA repair, replication, and recombination. *Mutation Research/Reviews in Genetic Toxicology*, **154**(1): 1–27.

Sharp PM, Cowe E, Higgins DG, et al. 1988. Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*; A review of the considerable within-species diversity. *Nucleic Acids Research*, **16**(17): 8207–8211.

Shen XK, Song SL, Li C, et al. 2022. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature*, **606**(7915): 725–731.

Sniegowski PD, Gerrish PJ, Johnson T, et al. 2000. The evolution of mutation rates: separating causes from consequences. *Bioessays*, **22**(12): 1057–1066.

Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. *Nature*, **458**(7239): 719–724.

Strauss C, Long H, Patterson CE, et al. 2017. Genome-wide mutation rate response to pH change in the coral reef pathogen *Vibrio shilonii* AK1. *mBio*, **8**(4): e01021–17.

Supek F, Miñana B, Valcárcel J, et al. 2014. Synonymous mutations frequently act as driver mutations in human cancers. *Cell*, **156**(6): 1324–1335.

Suyama M, Bork P. 2001. Evolution of prokaryotic gene order: genome rearrangements in closely related species. *Trends in Genetics*, **17**(1): 10–13.

Tchou J, Kasai H, Shibutani S, et al. 1991. 8-oxoguanine (8-

hydroxyguanine) DNA glycosylase and its substrate specificity. *Proceedings of the National Academy of Sciences of the United States of America*, **88**(11): 4690–4694.

Tomasetti C, Li L, Vogelstein B. 2017. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*, **355**(6331): 1330–1334.

Topal MD, Fresco JR. 1976. Complementary base pairing and the origin of substitution mutations. *Nature*, **263**(5575): 285–289.

Totter JR. 1980. Spontaneous cancer and its possible relationship to oxygen metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, **77**(4): 1763–1767.

Urushitani M, Sik A, Sakurai T, et al. 2006. Chromogranin-mediated secretion of mutant superoxide dismutase proteins linked to amyotrophic lateral sclerosis. *Nature Neuroscience*, **9**(1): 108–118.

Vijg J, Dong X. 2020. Pathogenic mechanisms of somatic mutation and genome mosaicism in aging. *Cell*, **182**(1): 12–23.

Wang YG, Obbard DJ. 2023. Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis. *Evolution Letters*, **7**(4): 216–226.

Watson JD, Crick FHC. 1953. The structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, **18**: 123–131.

Wei CS, Shan KJ, Wang WG, et al. 2021. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *Journal of Genetics and Genomics*, **48**(12): 1111–1121.

Witkin EM. 1953. Effects of temperature on spontaneous and induced mutations in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, **39**(5): 427–433.

Wu K, Li HC, Wang YH, et al. 2023. Silver nanoparticles elevate mutagenesis of eukaryotic genomes. *G3 Genes*| *Genomes*| *Genetics*, **13**(3): jkad008.

Xia B, Yan Y, Baron M, et al. 2020. Widespread transcriptional scanning in the testis modulates gene evolution rates. *Cell*, **180**(2): 248–262. e21.

Xie MC, Lu C, Wang JY, et al. 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature Medicine*, **20**(12): 1472–1478.

Xu LM, Wang XY, Lu XH, et al. 2023. Long-read sequencing identifies novel structural variations in colorectal cancer. *PLoS Genetics*, **19**(2): e1010514.

Xue YL, Wang QJ, Long Q, et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current Biology*, **19**(17): 1453–1457.

Yang SH, Wang L, Huang J, et al. 2015. Parent–progeny sequencing indicates higher mutation rates in heterozygotes. *Nature*, **523**(7561): 463–467.

Zheng SY, Kim H, Verhaak RGW. 2014. Silent mutations make some noise. *Cell*, **156**(6): 1129–1131.

Zou SS, Li JR, Zhou HB, et al. 2014. Mutational landscape of intrahepatic cholangiocarcinoma. *Nature Communications*, **5**: 5696.