# What is the human germline mutation rate? methodological innovations, challenges, and evolutionary implications

Cecilia Padilla-Iglesias[1,2*], Paco Majic[3]

[1]HUMAN ORIGINS - Cluster of Excellence for Integrative Human Origins Studies (EXC 3101), University of Tübingen, Tübingen, Germany
[2]Evolutionary Ecology Group, Department of Zoology, University of Cambridge, Cambridge, United Kingdom
[3]European Molecular Biology Laboratory, Heidelberg, Germany

*Correspondence should be addressed to: cp654@cam.ac.uk

## Abstract

Germline mutations are the ultimate source of heritable genetic variation, driving evolution, enabling adaptation, and underlying disease. Despite their fundamental importance, key questions remain unanswered: How frequently do germline mutations arise? Do mutation rates vary systematically across individuals, populations, and local genomic context? And what determines whether a mutation arising in a germ cell is ultimately transmitted to offspring? Historically, mutation rates were assumed to be relatively uniform within species, but recent advances in sequencing and computational methods now allow direct measurement of *de novo* mutations in families and gametes, revealing unexpected variation. We review evidence for variation in human germline mutation rates and spectra, considering sources ranging from molecular mechanisms and life-history traits to developmental processes and evolutionary forces. A central theme is the distinction between factors shaping mutation occurrence in germ cells and those influencing mutation detection in sampled individuals and populations. We highlight major sources of uncertainty, including methodological heterogeneity across studies, confounding between molecular and developmental effects, and limited sampling diversity. We argue that resolving these questions requires technical standardization, broader sampling across diverse ancestries and environments, and explicit theoretical frameworks that account for the multiple scales—cellular, organismal, and populational—at which mutation rates vary. Understanding the origins and extent of germline mutation rate variation is essential for reconstructing human evolutionary history, predicting disease risk, and interpreting the genetic architecture of complex traits.

## Main text

Ninety years ago, J.B.S. Haldane suggested that "taking the generation as the unit of time, man is a rather more mutable species than Drosophila", after attempting, for the first time, to measure the rates of spontaneous germline mutations in members of our species[1] (but see[2] for an earlier discussion of the same approach). Given that this was still two decades before understanding the structure of

DNA[3,4], Haldane lacked the technical and conceptual foundations to estimate mutations based on changes on DNA sequences as we do today. Thus, he needed to rely on an indirect method based on the principle of mutation-selection-balance, or the idea that deleterious alleles exist in populations because their purging via purifying selection is counterbalanced by a continual influx of new mutations[5]. Today, although our understanding of genetics has dramatically improved, both through the increasing availability of DNA sequencing technologies and important theoretical advances in the field, answers to the question of *how frequently new mutations arise* in the human germline are still elusive. Despite major breakthroughs and advances in this direction, there are still vast uncertainties in relation to just how variable germline mutation rates are across and within human populations, what drives differences in such variation, and how have those differences evolved throughout our evolutionary history.

Because genetic mutations are the primary substrate of heritable variation in populations, providing answers to these questions is central to understanding species evolution in general, including our own, as well as health and disease and the heritability of traits. This requires insight into the forces driving both how frequently new mutations arise (their rate), and the type of mutations that occur (their spectrum), particularly at the level of the germline. Germline mutations arise in cell lineages that are capable of differentiating into gametes and are therefore the source of genetic variation within the gene pools of populations. Although research over the past few decades suggests that germline mutation rates and spectra might vary within and between populations due to an interplay of genetic, environmental, and life-history factors, many analyses in genomics, population genetics, and phylogenetics still rely on simplifying assumptions about mutation rates. These assumptions can significantly influence the conclusions we draw (e.g.[6,7]).

In this perspective, we examine recent findings on variation in both the rate and spectrum of human germline mutations—within and across populations—and consider their possible causes. We also highlight major sources of uncertainty and key research gaps (For recent reviews covering somatic mutations or broader taxonomic comparisons, see[8,9]). A central theme is the importance of distinguishing between two questions: (1) what determines the probability that particular mutations occur in an individual's germline, and (2) what determines the probability that such mutations are present in a population at the time we measure them. In other words, we argue that disentangling the causes of mutation occurrence from the forces that shape their probability of persisting within a successful gamete, as well as through development and evolutionary time is key to reconciling disparate findings—and to guiding future research. For this it is important to distinguish between the molecular mutation rate, and the generational mutation rate, at which new alleles are added in the gene pool of a population[10–13]. We also argue that, as for most aspects of genetics research, diversifying the samples we use for human is necessary in order to disentangle potential drivers and variability in mutation rates within our species (see[14–16] for a more general discussion).

## *1. Quantifying human germline mutation rates*

Some of the earliest attempts to estimate mutation rates go back to Haldane's work on haemophilia, a recessive X-linked disorder. Haldane reasoned that if haemophiliacs typically did not reproduce, that is, if the strength of selection *s* was close to 1, the mutation rate could be approximated by dividing the frequency of haemophiliac males by 3, since males carry one-third of the X chromosomes in the population[1]. Using data from haemophilia cases in London, he estimated a mutation rate of about one mutation per 50,000 generations[1,17]. Following this seminal work, these genetic principles were widely applied to estimate spontaneous per-generation mutation rates by counting individuals affected by autosomal dominant or X-linked recessive disorders born to unaffected parents[18], yielding per-locus, per-generation estimates ranging from $10^{-6}$ to $10^{-4}$[19].

This approach, however, inherently biases results toward mutations with significant phenotypic impact, typically those causing severe and easily identifiable disease symptoms[20], and requires estimating both the mutational target size (the size of the genomic region where mutations yield the observable phenotype) and the strength of selection, both of which carry substantial uncertainty. Controlling for target size by using 20 known disease-causing loci, Kondrashov[17] estimated a human mutation rate of about $2 \times 10^{-8}$ per site per generation, while Lynch[21] combined data from a broader set of loci to arrive at a lower estimate of $1.28 \times 10^{-8}$ per site per generation.

Parallel to these disease-based approaches, other methods were developed during the molecular revolution that followed the deciphering of the DNA structure and genetic code. In 1968, Kimura introduced the neutral theory of molecular evolution, suggesting that most genetic substitutions persisting over evolutionary time are neutral and their fixation rate matches the mutation rate[22]. According to the neutral model, the number of substitutions ($K$) accumulating in a lineage over time ($T$) equals $(\mu/G)T$, where $\mu$ is the per-generation mutation rate and $G$ is generation time. This provided the theoretical foundation for the "phylogenetic method," where mutation rates are estimated by dividing observed sequence divergence in orthologous neutral genomic regions between two species by their divergence time (typically estimated from fossil records) and generation lengths[10,23,24]. Using this approach, Nachman and Crowell [24] estimated a human mutation rate of $2.5 \times 10^{-8}$ per site per generation, closely matching Kondrashov's[17] per-site estimate and compatible with Haldane's earlier per-locus estimate, assuming ~$10^3$ coding sites per gene[1].

In recent decades, sequencing-based pedigree designs (including trio-based approaches) have offered a more direct route to estimating germline mutation rates[25], and have been complemented by methods that sequence single gametes, allowing mutations to be observed in the very cells that transmit

DNA to the next generation[26–28]. Yet as we discuss below, these advances have not so much resolved the question of how fast the human germline mutates as revealed a new set of puzzles.

## *2. The puzzle of inconsistent estimates and their technical underpinnings*

Genome-wide sequencing in family pedigrees[29–34], has yielded estimates of germline mutation rates by relying on the assumption that any mutations present in the offspring whilst not in the parents must be the product of a *de novo* mutation (DNM) in the parental germline. These estimates typically fall around 1.1–1.3 × $10^{-8}$ mutations per site per generation[30,34–39]. They are also often lower than estimates derived from exome sequencing, disease-incidence and phylogenetic approaches, in some cases by roughly a factor of two[40–42]. Part of this discrepancy, especially with phylogenetic approaches, may reflect problems in phylogenetic calibration, since fossil dates constrain speciation times only imperfectly, whereas sequence divergence reflects the average coalescence time between lineages and may therefore substantially predate speciation, especially when ancestral effective population sizes are large. However, calibration issues are unlikely to explain the entire gap, which is particularly puzzling given that if anything, one would expect pedigree-based estimates to be slightly higher (as they include deleterious mutations that would eventually be weeded out by selection). To illustrate the issue, Wu et al.[43], using multigeneration pedigrees of humans and baboons, estimated mutation rates of 1.23 × $10^{-8}$ and 0.58 × $10^{-8}$ per base pair per generation, respectively. Under simple molecular-clock assumptions, these rates imply an ape–baboon average divergence time of as much as 63–67 million years, which is closer to the Cretaceous–Paleogene (KPG) extinction rather than the ~35 million years indicated by fossil evidence[44,45]. Although the precise degree of mismatch depends on assumptions about ancestral population sizes and generation intervals[46], the question is not how much lower pedigree-based estimates are relative to phylogenetic or disease-incidence methods, but what is driving remaining discrepancies?

Whilst at first glance it might seem that pedigree studies provide the most direct and accurate means of measuring *de novo* mutations, recently, Beichman et al.[8] stated that "germline mutation calling is still more an art than a science*"*. Differences in the methods used to report such mutations, from the sequencing technology, sequencing depth, mapping to a reference, variant calling, filtering, and appropriately accounting for false-positive and false-negative rates can introduce significant variation. Bergeron et al.[47] illustrated this clearly in a "Mutationathon" competition, where five laboratories independently analysed the same rhesus macaque pedigree sequencing data, yet reported mutation rate estimates differing by more than twofold.

Whilst analysis pipelines could be standardised to ensure comparability between individuals, populations and species, another source of biases in mutation rate estimation from trio sequencing

derives from the fact that these studies typically compare somatic (i.e. blood or saliva) rather than germ cells. Early post-zygotic mutations, appearing in parental somatic tissues and offspring, may incorrectly be classified as inherited rather than *de novo*, thus underestimating the true mutation rate[48,49]. Conversely, early somatic mutations in offspring might erroneously be recorded as *de novo* germline mutations. Incorporating additional generations and more recently long-read family sequencing[32,43], can help mitigate this issue. Available data suggest that this source of error may downwardly bias conventional short-read trio estimates by roughly 5–10%, and in some datasets perhaps somewhat more, although the larger gains seen with long reads also reflect improved access to repetitive regions of the genome[32,50,51].

Unlike trio-based approaches that infer DNMs by comparing parents and offspring, single-gamete sequencing avoids confounding from early embryonic mosaicism and somatic mutations in blood or saliva. This makes it theoretically possible to disentangle true germline mutations from those arising post-zygotically and to track mutational processes at the level of individual gametes. Studies quantifying mutation rates directly in sperm have reported substantially higher estimates than those from blood in the same individuals[26], although sequential sampling over the lifespan in New York[28] revealed significant variability both between individuals and across ages. At present, however, these methods remain technically challenging, particularly for oocytes, which are rare and difficult to isolate, limiting their use compared to pedigree-based approaches. Whether and to what extent we can assume that gaining a better knowledge of the drivers of mutational processes in males is informative for understanding those in females remains an open question.

Identity-by-descent (IBD) analyses of whole-genome data offer an alternative to trio-based inference and should, in principle, reduce the same confounding from post-zygotic mosaicism[52,53]. Instead of detecting *de novo* mutations directly in parent–offspring trios, they identify chromosomal segments shared from a recent common ancestor in pedigrees[52,53] or unrelated individuals[54] and use the sequence differences that have accumulated on those shared haplotypes to infer mutation rates. Yet these studies also produce variable estimates ($1.66 \times 10^{-8}$, $1.24 \times 10^{-8}$, and $1.20 \times 10^{-8}$ per bp per generation, respectively).

Adding to the puzzle of explaining variation in human germline mutation rates, García-Salinas et al.[29] recently analysed over 10,000 trios from the 100,000 Genomes Project using a unified pipeline and reported that individuals of African ancestry had germline mutation rates 3–4% higher than those of European, South Asian, or American ancestry—even after controlling for maternal and paternal age (see also[55] for results in a similar direction). If this difference holds across generations, it will substantially influence our estimates of the timing of evolutionary events, as well as risk models in medical genomics.

Has the mutation rate evolved not just across lineages, but within populations? If not, what explains these differences?

The reasons why different methods yield different estimates may thus be biological, rather than technical, which begs the question of what the drivers of mutation rate and spectra variability are.

*3. Causes of germline mutations*

Going back to the discrepancies between the estimates for humans and baboons and the fossil record, Wu et al.[43] proposed a biological explanation. Specifically, they argued for a slowdown in germline mutation rates in both species over time. The broader disconnect between pedigree- and phylogeny-based estimates demands attention requires asking when have mutation rates changed across the phylogeny, but also, why.

Identifying potential sources of variation in mutation rate requires identifying the sources of mutations, both in germ cells (where all mutations emerge), as well as in individuals and or populations – which is what we ultimately sample when obtaining estimates[12]. Given that the different methods assess mutations at different timescales, and (most of the time) in adult organisms, in order to understand the drivers of observed differences, we need to understand, first and foremost, the emergence of germline mutations in single cells. Then, how do these mutations accumulate during development and therefore get to be sampled when taking blood/ saliva samples in individuals and last, how do those mutations accumulate over evolutionary timescales.

We will focus first on the causes of mutations at a molecular level, which can be classified under two main types of processes: The first one is copying errors that arise during DNA replication. The second are DNA changes that occur spontaneously (i.e., nonreplicatively) because of the instability of the structure of DNA[56,57] or are induced by endogenous or exogenous sources (i.e., UV radiation or other chemical or biological agents)[43,48]. Both types of errors lead to mutations if they are either not repaired or repaired wrongly before the next round of replication. Hence, mutation rate variation reflects a complex balance between lesion formation and DNA repair efficacy. Differences in either component could explain variation in mutation rates between sexes, individuals, and populations[41].

For decades, the predominant view was that replication errors primarily drove variation in human germline mutation rates[37,48]. Evidence supporting this notion included the significant excess of mutations of paternal origin compared to maternal origin, consistent with a higher number of replication cycles in males. In humans, by the onset of puberty, each primary oocyte has undergone approximately 31 cell divisions whilst in males, before spermatogenesis there are 34. However, through adulthood,

sperm is produced continously via the asymmetric division of spermatogonial stem cells every 16 days, with five additional cell divisions per cycle. Thus, by the age of 30, approximately 429 germ cell divisions occur in males, rising to about 659 by age 40[41,48]. Early research by Haldane[1] and subsequent studies[32,37,38,58] consistently documented a higher mutation contribution from fathers compared to mothers, called the paternal age effect.

Recent research has called into question the dominance of replication-driven mutations. Studies have shown a significant maternal age effect on mutation rates, although less pronounced than the paternal age effect, indicating that maternal age also contributes substantially to germline mutations [29,30,43]. Furthermore, young fathers contribute roughly three times more mutations than young mothers, a ratio that surprisingly remains relatively constant rather than increasing markedly with paternal age, as would be expected if replication cycles were the sole driving factor[48,59].

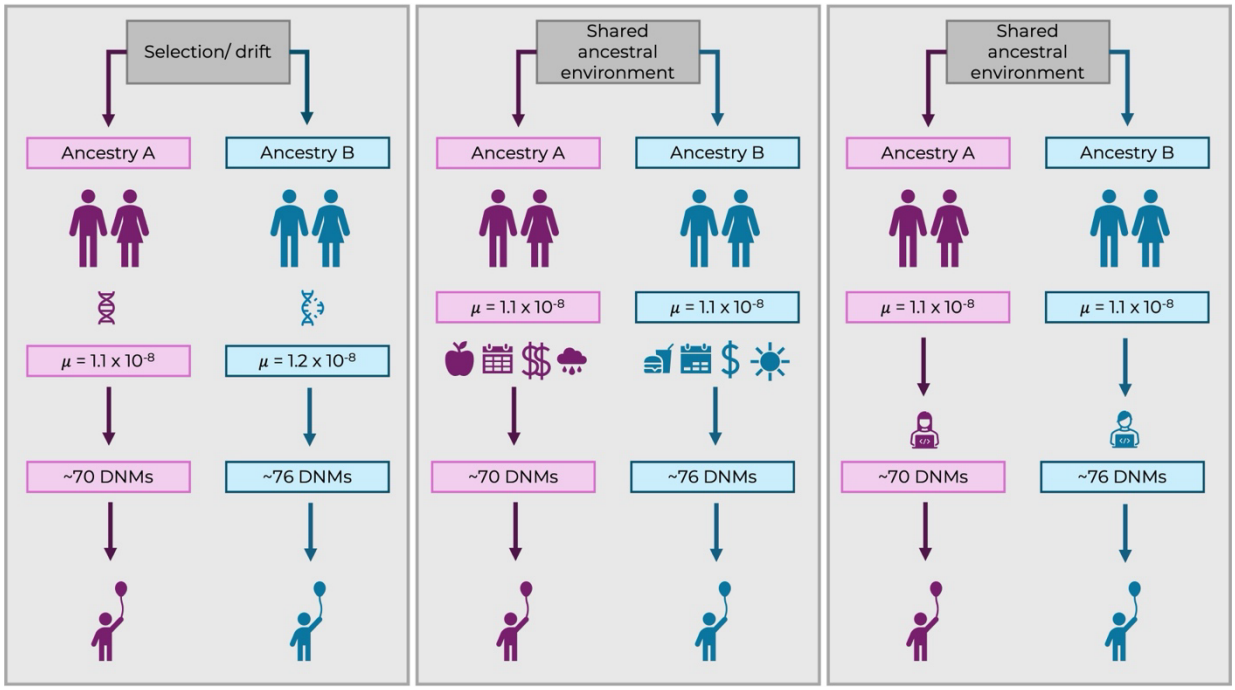


***Figure 1. Pathways generating observed differences in measured germline mutation rates between populations from pedigree (trio) studies.*** *Three distinct mechanisms that can produce between-population differences in measured de novo mutation (DNM) counts.* ***Left panel:*** *True differences in molecular mutation rates (μ) arise through lineage-specific evolution via selection or genetic drift, resulting in heritable differences in the per-generation mutation rate at the germline level.* ***Middle panel:*** *Shared molecular mutation rates can yield different measured DNM counts when populations experience distinct developmental, environmental, or life-history conditions (represented by icons: diet, life-history, socio-economic status, external environment). These factors alter the probability of sampling mutations in offspring without affecting the underlying per-cell mutation probability.* ***Right panel:*** *Identical molecular and developmental processes can produce apparent rate differences through variation in study design, laboratory protocols, or bioinformatic pipelines. Distinguishing among these scenarios is essential for interpreting reported mutation rate variation and requires careful experimental design, standardized methodologies, and representative sampling across diverse populations and environments.*

Damage-induced mutations play a significant role in shaping germline mutation patterns. One striking example is the prevalence of C>T transitions at methylated CpG sites. Despite CpGs comprising less than 2% of the genome, they account for nearly 20% of observed germline

mutations[48]. This disproportionate frequency is largely due to the spontaneous deamination of methylated cytosines. When unmethylated cytosines deaminate, they form uracil, which is efficiently recognized and excised by uracil-DNA glycosylase as part of the base excision repair pathway. In contrast, deamination of methylated cytosines produces thymine, resulting in T:G mismatches that are less effectively repaired[41,57,60]. As a result, methylated CpG sites are especially prone to C>T transitions. Transversions are also enriched at CpG sites, though their mechanisms remain unclear.

The frequency of damage and repair efficiency are further influenced by local sequence context. DNA in AT-rich regions, held together by fewer hydrogen bonds, is more easily unwound, facilitating repair by exonuclease enzymes[61,62]. In contrast, GC-rich regions, being more thermodynamically stable, may hinder repair accessibility—contributing to their elevated mutation rates[48,63]. This sequence-context effect may also help explain the observed increase in mutation frequency at GC-rich sites with advancing maternal age, likely reflecting accumulated DNA damage coupled with reduced repair efficiency in the oocytes of older mothers[48]. Additionally, the functional sequence context can also play a role in mutation rates. In recent years it has been proposed that, at least in some organisms, the binding of transcription factors may shield novel mutations from repair mechanisms[64,65], that the amount of transcriptional activity influences the frequency of mutations[66] and that certain epigenetic marks may differentially recruit repair mechanisms, thus locally decreasing mutation rates[67]. Different mechanisms shaping mutations of paternal and maternal origin has been argued based on analyses of the spatial clustering of paternal derived and maternally derived mutations in pedigree studies[30,68].

In population-level studies, a consistent enrichment of TCC>TTC mutations—up to 20% higher in individuals of European ancestry compared to those of African or Asian ancestry—has been documented in phylogenetic analyses and sometimes attributed to UV exposure[69–72]. However, Amos[73] has proposed a role of local sequence context in explaining this difference. Specifically, that flanking heterozygosity promotes extra rounds of repair/replication that slightly elevate mutation rates. Analysing 1000 Genomes data, he showed that as flanking heterozygosity increases, the relative probability of some triplet changes consistently rises while others fall, with broadly similar slopes across Africa, Europe, South/Central Asia, East Asia and the Americas. Because the "Out-of-Africa" bottleneck reduced heterozygosity outside Africa, this framework predicts a demographic, genome-wide shift in mutation spectra that could explain African vs. non-African spectrum differences (e.g., the well-known TCC→TTC contrast) without invoking population-specific modifiers for each triplet class. He extended this reasoning to propose that mutability might be overall higher in Africans[74]. And in fact, a recent study relying on the TOPMed research program – across 1,465 trios from diverse human populations living in the United States, that tested the relationship between heterozygosity and DNM rate across all individuals, we found a significant positive correlation even after accounting for parental age[75]. However, heterozygosity only explained ~0.3% of the variation in DNM count. Crucially,

however, this mechanism cannot straightforwardly account for the elevated mutation rates in African-ancestry individuals reported by García-Salinas et al.[29], whose African-descent sample actually showed *lower* genomic diversity and heterozygosity than their European-descent sample.

This leaves open the question of what drives the observed between-population and between-sex differences in mutation rates and spectra. Could environmentally induced damage be then responsible? The evidence suggests not. García-Salinas et al.[29] found no difference in mutation spectra between individuals of African and European ancestry in the same trio dataset — not even in the TCC>TTC class — despite a small but consistent difference in overall DNM rates, implying that the elevated rate in African-ancestry individuals is not driven by specific mutational signatures. Similarly, Jónsson et al.[30] found no significant sex differences in mutation spectrum, suggesting that oocytes and spermatogonia may experience similar mutagenic processes, even if mutation rates differ. So, can different evolutionary processes and histories account for the differences observed within and between human populations?

4. *The role of development on shaping generational mutation rates*

Development, while not altering the *molecular* mutation rate of germ cells, can strongly influence the *measured* mutation rate in individuals (i.e. the probability of sampling a mutant cell) and, ultimately, the number of heritable mutations present in populations—the substrate on which natural selection acts[12,76].

A clear example comes from a potential downward bias in germline mutation-rate estimates from family trio studies. Such studies generally compare somatic tissue from parents and offspring, rather than germ cells directly. An early post-zygotic mutation occurring *before* germline specification in either parent will be present in both the parent's soma (and thus in the sampled tissue) and in their germline, and so also in the offspring. Under standard short-read approaches, an early post-zygotic mutation arising before germline specification in a parent may be present in both the parent's soma and germline, and therefore also in the offspring, causing the site to be classified as shared rather than *de novo* and omitted from the estimate[29,49,53]. Recent long-read pedigree studies, however, have begun to resolve this problem by phasing mutations to parental haplotypes and distinguishing inherited variants from early embryonic/postzygotic mutations more directly, thereby increasing *de novo* mutation discovery and revealing additional mutations in repetitive regions inaccessible to short reads[32,50].

This bias could be substantial if mutation rates are especially high during the earliest embryonic cell divisions. Increased early mutation loads are biologically plausible: for example, two key components of base excision repair are missing in spermatozoa, meaning DNA lesions accumulated late

in spermatogenesis can only be repaired post-fertilization in the zygote[48,77]. At the same time, mammalian zygotes depend almost entirely on the maternal protein and mRNA complement until the four-cell stage[78–80]. If oocyte replication or repair capacity declines with maternal age[81,82], early embryonic mutation counts could increase, providing one mechanistic explanation for the maternal-age effect observed in recent trio studies[29,48].

This reasoning also predicts that species or populations with longer developmental time may show higher per-generation de novo mutation input, even if the underlying molecular mutation rate per cell division or per unit time is unchanged[12]. In this view, developmental duration can increase the probability that a gamete derives from a mutant cell lineage, thereby elevating the effective mutation rate observed at the population level.

Indeed, Zhu et al.[65] argue that this positive relationship between developmental time and measured mutation counts could mask an actual *negative* relationship between the true gametic mutation rate and generation time. In their framework, embryonic mutations reflect the action of genetic drift on mutation-rate modifiers (i.e. drift-barrier dynamics[21]), but for gamete-derived mutations, mutation-rate modifiers are constrained by both effective population size and generation time—making selection against mutator alleles more efficient in long-lived species, even if their $N_e$ is smaller. Regardless of the mechanism, this example is helpful to illustrate how distinguishing whether we measure mutations in cells, individuals or populations is key to investigate variation in rates and its evolutionary explanations.

Differences in the proportion of the body allocated to the germline could, in principle, also explain cross-species variation in mutation rates. For example, if chimpanzees have proportionally more germline stem cells than humans—due to their much larger testes—this could contribute to their higher measured mutation rate relative to humans[83,84].

More generally, the pronounced excess of paternally derived DNMs may partly reflect the far greater number of cell divisions in the male germline, which creates more opportunity for mutant spermatogonial lineages to expand and contribute disproportionately to the sperm pool. Such sex-specific biases may therefore arise not only from mutational input per se, but also from intraorganismal selection within the germline[12,85–87].

Majic et al.[12] demonstrated this principle experimentally in *Drosophila*: heterozygous flies carrying a mutation that increased cell proliferation produced a biased proportion of offspring inheriting that mutation, implying overrepresentation in the gametes. In humans, while such experiments are not feasible, convergent evidence points to the same process. Clinical studies have long shown that certain pathogenic mutations undergo clonal expansion in the male germline and rise in frequency with paternal

age[88,89]. More recently, large-scale analyses of human sperm have shown that positive selection on driver variants in spermatogonia creates apparent mutation hotspots, elevating the effective transmission rate at specific loci far above baseline expectations. These studies identified dozens of genes involved in such clonal expansions, and can produce a roughly 2–3-fold enrichment of likely disease-causing variants in sperm, with 3–5% of sperm from middle-aged to older individuals carrying a pathogenic mutation across the exome[13,90]. Together, these results indicate that intraorganismal selection within the testis can alter the spectrum and frequency of transmitted mutations, contributing to paternal-age effects and to an, albeit small, part of the paternal bias in germline mutagenesis.

The important point to take from this section is that even if the underlying germline mutation rate is identical, developmental differences could produce systematic variation in the *measured* rates among individuals and populations. This is particularly important when comparing populations in diaspora settings (as it is the case for most *biobank*-based studies): for example, African-descended individuals in the UK may develop under environmental, nutritional, or socio-economic conditions that differ from both their ancestral and local European-descended populations. If such conditions influence early development, germline turnover or other life history parameters, they could affect measured mutation rates in ways unrelated to genetic ancestry.

This possibility highlights a broader concern: observed inter-population differences in mutation rates may partly reflect *which* individuals are measured and the environments they experienced, rather than intrinsic, population-wide genetic differences. This further highlights the need of careful study designs when attempting to explore inter-population or inter-individual differences in mutation rates or spectra.

### *5. Evolutionary explanations*

All other things being equal, given the key role of mutations in introducing heritable sources of variation – i.e. the raw material for evolution - mutation rates (at a molecular level) are expected to be an evolving trait. Explanations for differences in molecular germline mutation rates across species, populations, and individuals have generally been framed in two ways: selectionist theories, which posit that mutation rates are shaped by natural selection, and neutralist theories, which emphasise the role of genetic drift and other stochastic forces. Disentangling their relative contributions matters for predicting our species' mutation rate and for assessing whether a single "universal" rate—often assumed in population-genetic models—adequately describes humans, or whether rates vary systematically among populations and how. Human per-generation mutation rates appear comparatively high among mammals[8,21,91]. Therefore, an evolutionary account is required to explain why.

In selectionist theories, mutation rates are treated as the product of natural selection acting on traits such as long-term survival, reproductive fitness, or evolvability[67,92–96]. One such theories, named the cost of fidelity model, proposes that the germline mutation rate reflects a trade-off between the benefits of avoiding deleterious mutations and the energetic and/or time costs of achieving accurate DNA replication. While this model is well supported in viruses and bacteria, its explanatory power in multicellular organisms remains uncertain[91].

A related idea treats mutation rates as an evolvability trait[95,97], as low mutation rates can impose a second-order fitness cost by slowing adaptation to changing environments. In bacteria, certain mutagenic pathways are upregulated under stress conditions[98]. If a similar mechanism operated in our lineage, it could plausibly have been advantageous during periods of environmental instability. Paleoclimate records indicate that after ~270 ka a relatively stable phase in Africa gave way to marked climatic cyclicity, which coincides with the ecological diversification and wider dispersal of early *Homo sapiens* populations[99,100]. Whether selection favoured slightly higher baseline mutation rates as part of a broader adaptive strategy for flexibility is therefore an untested, but intriguing possibility[101,102], (but see[103] for experimental evidence in bacteria).

Another selectionist hypothesis links germline and somatic mutation rates through shared DNA repair pathways. If long-lived, large-bodied species experience strong selection to reduce somatic mutation rates, for instance, to limit cancer risk, then improved DNA repair could incidentally lower germline mutation rates as well[8,104,105] success or increases parental investment, such selection could be especially strong. However, available data complicates this link. For instance, baleen whales (with life spans up to ~200 years) show per-generation germline mutation rates similar to humans, which are already high for mammals[106], suggesting that selection against somatic mutations might not always reduce germline mutation rates.

Comparative tests within primates could also explore the effects of contrasts in parental investment. Cooperative breeders such as marmosets (*Callithrix jacchus*)—which rely heavily on biparental and alloparental care and show a low per-generation mutation rate ($0.43\times10^{-8}$[107])—can be contrasted with taxa in which parental investment is more limited, including owl monkeys ($0.81\times10^{-8}$[108]; a closely related species), many Old World monkeys (e.g., green monkeys or macaques[69,109], and most great apes[83]. If selection for longevity and somatic maintenance is reinforced by high parental investment, all other things being equal, cooperative breeders should lie toward the lower end of the mutation-rate continuum. However, this prediction does not map cleanly onto apes: humans are arguably the most cooperative breeders among great apes[110,111], yet human per-generation mutation rates are amongst the highest (e.g.[8,83]) - implying that parental investment alone cannot explain interspecific differences,

though within-human variation in parental investment remains an interesting candidate for population-level differences to test.

On the other hand, one of the most influential neutralist theories of mutation rate evolution is the drift-barrier hypothesis[21,91,112,113]. This framework builds on Ohta's nearly neutral theory[114], and it proposes that since most mutations are neutral or deleterious, natural selection should act to reduce the mutation rate. However, this reduction is ultimately limited either by the cost of extremely high-fidelity replication or by the power of genetic drift. In species with small effective population sizes ($N_e$), selection against weakly deleterious "mutator" alleles (i.e. alleles that increase mutation rates) is therefore less effective, allowing higher mutation rates to persist. Yet this logic also presumes a supply of heritable modifier variants affecting mutation rate. If such variants arise rarely, or if mechanistic changes to replication and repair are strongly constrained, mutation rates may evolve more slowly, or over a narrower range, than drift-barrier arguments alone would suggest[12].

A recent study providing support for the drift-barrier hypothesis was that from Bergeron et al.[25], comparing germline mutation rates from family trios across 68 vertebrate species. The authors found a significant negative correlation between per-generation mutation rates and $N_e$, independent of phylogenetic autocorrelation. Traits that are typical of species with low $N_e$ (i.e. slow life history strategies), such as low fecundity and long generation times were also independently associated with higher mutation rates. However, Majic et al.[12] demonstrated how the association between *Ne* and generational mutation rates in this dataset could also result from the evolution of life-history traits. More recently, reinforcing this interpretation, Weinstein and Roy[115] reanalysed the same dataset used by Bergeron et al.[25] using a phylogenetic, causal inferential framework explicitly controlling for six life-history traits, and showed that the apparent effect of $N_e$ on per generation mutation rates was completely mediated by the generation time. Future work in our own species should therefore explore whether human populations with historically small $N_e$ (e.g., long-term isolated groups) may have slightly higher baseline mutation rates than those with larger $N_e$.

If mutation rates track the number of germline cell divisions, then life-history traits affect the number of cell divisions may indirectly influence mutation rates. The generation time hypothesis[116,117] predicts that species—or populations—with shorter generation times have higher per-year mutation rates because they undergo more germline replications per unit time. Longer reproductive ages should also increase the *proportion* of replication-driven, male-biased mutations[118].

Generation time changes have also been invoked to explain the discrepancy between trio-based and phylogenetic mutation rate estimates, as generation times have lengthened over the course of primate evolution—from Old World monkeys, to apes, to hominins, and finally to modern

humans[106,119,120]. However, if longer generation times are accompanied by later puberty onset (which seems to be the case across human evolution, see[121,122]), the total number of germline divisions may not scale linearly with generation time. In recent centuries, nonetheless age at puberty has fallen sharply in many Western populations[123], even as generation time has increased, raising the possibility of population-specific shifts in mutation rate components.

Within our own species, archaeological, ethnographic, and historical data indicate that both contemporary and prehistoric human populations vary widely in life-history traits[106,124–127]. Although such variation could, in principle, produce inter-population differences in mutation rates, its explanatory power in humans is debated[128].

Mating systems may also influence mutation rates indirectly through sperm production. In species with high sperm competition, males produce more sperm, increasing the number of germline cell divisions, and speed of cell division and thus the potential for mutations[129–131]. Humans are largely (historically) monogamous with intermediate relative testis size, while chimpanzees are promiscuous, with larger testes and faster sperm turnover (every 14 days vs every 16 days in humans[58,124,132]). This may contribute to a pronounced paternal age effect in chimpanzees, although current pedigree-based estimates remain imprecise because they are based on small sample sizes [133]. In fact, in Weinstein and Roy's[115] analyses of mutation rates across the vertebrate phylogeny, the authors found that after controlling for generation time, polygynous vs. monogamous mating systems was the strongest predictor of per-generation mutation rates, possibly due to differences in sperm competition. Within humans, Auger et al.[134] reported variation in the rate in sperm production exists across and even within populations across relatively short geographic distances, but whether these differences are large enough to influence mutation rates is unknown.

So far however, we have focused on both the potential causes and evolutionary explanations of differences in *molecular* mutation rates across individuals, populations and species. However, differences in the mutation rates we measure in adult individuals need not to be reflective of any differences in molecular machinery. In the section that follows, we argue that too much work has been devoted to investigating potential drivers of molecular mutations as explanations for differences we observe across methods, populations, and individuals – across species and within our own.

## 6. *So, is there population specific evolution of human mutation rates?*

Many reported population-level differences in mutation-rate estimates come from studies of groups with small effective population sizes, high levels of isolation, or cultural practices that increase consanguinity. These contexts are expected to alter the genomic background of individuals, and amplify

the role of drift acting on modifiers of the *molecular* mutation rate itself, but they can also confound interpretation by intersecting with developmental and environmental factors that influence the *number of mutations we measure in adult individuals*. In other words, it is important to recognise that biases in sampling can affect *both* the processes that genuinely alter the probability of new mutations per cell division, and processes that alter how many of those mutations are ultimately observed in offspring (such as developmental patterns, number of gametes produced, etc…).

Examples of populations in which these complexities arise, yet have been used to claim inter-population differences in mutation rates include immigrant groups in Europe and North America[29], isolated religious communities[75], societies with high levels of consanguineous marriage[55], and countries with exceptionally small effective populations[30,135]. Thus, observed inter-population differences may stabilise population-specific (adaptive or neutral) evolution of germline mutation rates, short-term effects of environmental damage, but also environmental and developmental effects on measured counts in sampled individuals.

Any attempt to disentangle the above-mentioned possibilities must first recognize that grouping individuals into broad continental ancestry categories often complicates the task. For instance, Africa harbours more genetic diversity than all populations outside Africa combined[136]. Treating African-descent individuals as a single category is therefore unjustified when trying to understand the evolutionary trajectories of mutation rates and spectra[16,137]. Hunter-gatherer and former hunter-gatherer groups in Central and Southern Africa descend from lineages that are estimated to have diverged from all other human populations as far as 250–300 kya[138–141]—four to five times older than the deepest divergence between any African and non-African populations[136]. This raises the question of what we really mean by "population-specific evolution of the human mutation rate" (cf.[70]) in studies relying on broad continental groupings, when some lineages diverged close to the origins of our species. It may also help explain the large within-group variance in mutation-rate estimates observed in datasets such as García-Salinas et al.[29].

Another key issue is whether the samples used to calculate human mutation rates are truly representative—not only of more or less "valid" continental groups but even of the specific populations they are meant to represent. For example, although African-descent women in the UK typically give birth at younger ages than women of European descent [142], the African-descent parents in García-Salinas et al.'s[29] dataset were significantly older, implying longer generation times. Moreover, while African populations generally harbour the highest levels of global genetic diversity[143], the African-descent participants in that study showed only about one-quarter of the genetic variation observed among the European-descent participants, indicating unusually low $N_e$ and high levels of homozygosity. This suggests that observed differences between "ancestry groups" may instead reflect drift-barrier effects

or even gene × environment interactions affecting mutation rate probabilities at a molecular level, but also developmental or environmental effects affecting the probability of sampling mutant cells from individuals.

If mutation rates had undergone stable, population-specific evolution, one expectation would be that mutation rate is highly heritable. Empirical evidence to date does not support this. The two largest trio-based studies—García-Salinas et al.[29], which estimated SNP heritability of parent-of-origin DNM burdens after adjusting for parental age and covariates, and Kessler et al.[75], which estimated narrow-sense heritability using genome-wide relatedness—both found near-zero heritability. This indicates minimal additive genetic contribution (particularly from common variants) to variation in DNM counts.

A further complication is that mutation-rate estimates depend on the timescale over which they are measured. Trio studies quantify mutations transmitted in a single generation, often after controlling for parental age and related covariates, whereas IBD-based approaches integrate mutation accumulation across several generations and thus reflect the realized number of mutations added along recent lineages. Unsurprisingly, the two approaches may yield different patterns. For example, Huang et al.[144], using IBD-based methods, reported significant differences in per-generation mutation-rate estimates across human populations, including higher estimates in European than African populations, seemingly contradicting the study by Garcia-Salinas et al.[29]. Such discrepancies may indicate that historical differences in parental age structure, environmental mutagen exposure, or other transient factors can leave measurable signatures in lineage-specific mutation accumulation even when present-day molecular mutation rates are similar. Yet for evolutionary and demographic inference, those accumulated differences may be precisely the relevant quantity, since they determine how many mutations were actually added along ancestral lineages.

When focusing on between-population differences in mutation spectra (rather than absolute rates)—for example, the well-known enrichment of TCC→TTC changes that is strongest in Europeans and weaker in South Asians[70,71]—genealogy-based methods have localized these shifts in time and population history rather than to particular genomic regions. For instance, Speidel et al.[72], relying on a dataset of 278 present-day and 430 ancient DNA human samples quantified the TCC→TTC "pulse" via an integrated mutation-intensity metric. The authors showed it is absent >34 kya, widespread by the Late Glacial Period, strongest in Neolithic farmers, and highly correlated with coalescence to a ~10 kya Anatolian genome, consistent with spread by gene flow rather than a locus-specific driver, and shedding light on the evolutionary history of a mutational pattern (albeit not its source; see also[69]).

Although still rare, new studies relying on multi-generational pedigrees may allow us to get one step closer to the question of heritability of mutation rates. Porubsky et al.[32] found that DNMs tended

to appear repeatedly within the same 32 genomic regions. Yet because this was a single family, this raises the question: would the same regions re-occur in others? In fact, even something like the extent of a paternal age effect has been shown to differ across human families[33,51]. A larger number of families will be needed to pinpoint why mutation rates differ between lineages and to what extent such variation is heritable or transient and environmentally modulated.

### *7. A way forward?*

So where do we go from here? Broadly speaking - despite ever-better sequencing—we still measure human germline mutations at a rate not too different from Haldane's first attempts almost a century ago. Yet variability in that rate is often treated as a given. We have suggestive evidence that paternal and maternal contributions to offspring DNMs differ, that paternal age (and to a lesser degree maternal age) shapes mutational load, and that ancestry groups may show distinct mutational profiles. But most of those inferences rest on Global-North datasets, modest sample sizes, and highly heterogeneous bioinformatic pipelines. And in many pedigree studies, germline and post-zygotic variants are lumped together—even though theory gives us good reasons to expect different evolutionary forces acting on the number of mutations that arise on germ cells versus those that occur during pre-puberty development[65]. As a result, it remains unclear in each case whether reported variation reflects molecular processes (neutral or adaptive), environmental/developmental factors that affect the likelihood of sampling mutant cells, or methodological discrepancies between studies—or, most likely, some complex interaction of all three.

The first thing we need to be able to disentangle between the above possibilities is to diversify our samples. Not just in terms of the ancestry composition of individuals, but in terms of the environments they inhabit – social, ecological and economic. For example, *All of Us* Research Program, a US-based biobank (one of the largest in the world), now includes >414k whole-genome sequences, yet the number of trios in the dataset is extremely small and parental ages at conception are not recorded[145]. Therefore, the potential to investigate questions concerning germline mutation rates is limited[146]. Similarly, the UK's Newborn Genomes "Generation Study" will sequence ~100,000 newborns and their mothers, but not their fathers. Whilst this design might be useful for early diagnostics, it makes it impossible to trace where mutations come from, or how often they arise, and thus, the heritability of DNMs.

Second, we need to *explicitly* incorporate uncertainty around mutation rate estimations though downstream analyses in studies of population history and inferences of historical relationships between populations and species, rather than treat point estimates as fixed constants.

Third, we must combine human data and theoretical studies with experiments in model systems. Mutation accumulation experiments remain one of the most powerful tools for identifying sources of mutations, and for disentangling mutational inputs from selective filters[147–154]. Therefore, manipulating exposure to mutagens, genomic background, life-history timing, and environments directly could allow us to directly test hypotheses with regards to their role in shaping mutagenesis.

Fourth, we need to create unified wet-lab and bioinformatic protocols to make sure "mutation rate" means the same across studies, species and contexts (c.f.[47]). For instance, whilst promising recent studies claim an ability to distinguish postzygotic mutations arising post-fertilisation, prior to primordial germ cell specification and *de novo* mutations in the Genomics England 100,000 genomes project[155], it remains to be determined whether such pipeline is applicable to other datasets.

Fifth, we should take advantage of technologies that let us measure germline mutations more directly and confidently. High-accuracy long-read sequencing (PacBio HiFi; ONT duplex[156]) and error-corrected Duplex Sequencing[13,26] now permit ultra-rare variant detection at single-gamete resolution[28] with far fewer false positives than short-read pipelines. These methods apply both genome-wide and at specific loci, allowing targeted tests of evolutionary hypotheses. Importantly, they do not just yield "better numbers" to plug into population history models—they enable asking genuinely new questions about when and where mutations arise, and how ancestry, environment, and life history modulate both their occurrence and fate.

One example comes from recent sperm-sequencing studies. Melamed et al.[157] showed that the HbS mutation (20A→T in *HBB*), which is associated with malaria resistance, arises de novo at higher rates in donors of African descent. A follow-up srudy[27] found a similar pattern for the *APOL1 G1* variant (1024A→G), long known to protect against *Trypanosoma brucei gambiense*. The authors suggest that some conditionally adaptive mutations may originate more frequently in populations where they have been beneficial—seemingly at odds with the canonical view that mutations are random with respect to fitness[158]. Whether this claim holds up will depend on replication across labs, loci and methodologies, but the point is that this research illustrates how evolutionary principles and hypotheses can be tested directly.

The sequencing of single gametes could also enable scientists to track mutations across the lifespan on individual germ cells. This would help pinpoint not just exactly when and where mutations occur, but also how lifestyle, ancestry or the environment individuals are exposed to at different moments in life might shape both the occurrence and the fate of such mutations[13,90]. Meanwhile, if at least in part, lineage specific evolution of mutation rates is driven by the emergence of "mutator alleles" new

computational tools, combined with the rapidly increasing availability of ancient DNA, could help scientists trace when such "mutator" alleles emerged and spread in human history[72].

So, perhaps, "what is *the* human mutation rate" is the wrong question – and looking for a number is the wrong thing to look for. Mutations differ across genomic contexts[32], across developmental stages[28], and across individuals living in different environments. A more useful framework may be to ask: "what factors make certain mutations more likely to occur in particular germlines, at particular moments, and which of these factors are heritable? And, how have these forces shifted over the evolutionary history of our species?"

**Acknowledgements**: The authors wish to acknowledge Mikkel Heide Schierup for extensive and invaluable comments on earlier versions of this manuscript.

**References**

1. Haldane, J. B. The rate of spontaneous mutation of a human gene. *J. Genet.* **31**, 317–326 (1935).
2. Danforth, C. H. The frequency of mutation and the incidence of hereditary traits in man. in *Eugenics, genetics and the family, Scientific papers of the 2nd International Congress of Eugenics, NY, 1921* vol. 1 120–128 (Williams & Wilkins, Baltimore, 1923).
3. Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741 (1953).
4. Watson, J. D. & Crick, F. H. The structure of DNA. in *Cold Spring Harbor symposia on quantitative biology* vol. 18 123–131 (Cold Spring Harbor Laboratory Press, 1953).
5. Haldane, J. B. S. A mathematical theory of natural and artificial selection. in *Mathematical Proceedings of the Cambridge Philosophical Society* vol. 23 607–615 (Cambridge University Press, 1927).
6. Bromham, L. Mutation rate is central to understanding evolution. *Am. J. Bot.* **111**, (2024).
7. Ritchie, A. M., Hua, X. & Bromham, L. Investigating the reliability of molecular estimates of evolutionary time when substitution rates and speciation rates vary. *BMC Ecol. Evol.* **22**, 61 (2022).
8. Beichman, A. C., Zhu, L. & Harris, K. The Evolutionary Interplay of Somatic and Germline Mutation Rates. *Annu. Rev. Biomed. Data Sci.* **7**, 83–105 (2024).
9. Solís-Moruno, M., Batlle-Masó, L., Bonet, N., Aróstegui, J. I. & Casals, F. Somatic genetic variation in healthy tissue and non-cancer diseases. *Eur. J. Hum. Genet.* **31**, 48–54 (2023).
10. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
11. Kimura, M. Molecular evolutionary clock and the neutral theory. *J. Mol. Evol.* **26**, 24–33 (1987).
12. Majic, P., Srivastava, M. & Crocker, J. The evolution of mutation rates in the light of development and cell-lineage selection. *bioRxiv* 2025–01 (2025).
13. Neville, M. D. *et al.* Sperm sequencing reveals extensive positive selection in the male germline. *Nature* 1–8 (2025).
14. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat. Med.* **28**, 243–250 (2022).
15. Flegontov, P. *et al.* Modeling of African population history using f-statistics is biased when applying all previously proposed SNP ascertainment schemes. *PloS Genet.* **19**, e1010931 (2023).

16. Padilla-Iglesias, C. & Derkx, I. Hunter–gatherer genetics research: Importance and avenues. *Evol. Hum. Sci.* **6**, e15 (2024).
17. Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing mendelian diseases. *Hum. Mutat.* **21**, 12–27 (2003).
18. Cooper, D. N. & Krawczak, M. *Human Gene Mutation*. (Bios Scientific Publ, Oxford, 1995).
19. Vogel, F. & Motulsky, A. G. Mutation: Spontaneous Mutation in Germ Cells. in *Human Genetics* 385–430 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1997). doi:10.1007/978-3-662-03356-2_10.
20. Pfeifer, S. P. Spontaneous Mutation Rates. in *The Molecular Evolutionary Clock* (ed. Ho, S. Y. W.) 35–44 (Springer International Publishing, Cham, 2020). doi:10.1007/978-3-030-60181-2_3.
21. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci.* **107**, 961–968 (2010).
22. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
23. Kondrashov, A. S. & Crow, J. F. A molecular approach to estimating the human deleterious mutation rate. *Hum. Mutat.* **2**, 229–234 (1993).
24. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
25. Bergeron, L. A. *et al.* Evolution of the germline mutation rate across vertebrates. *Nature* **615**, 285–291 (2023).
26. Axelsson, J. *et al.* Frequency and spectrum of mutations in human sperm measured using duplex sequencing correlate with trio-based de novo mutation analyses. *Sci. Rep.* **14**, 23134 (2024).
27. Melamed, D. *et al.* De novo rates of a *Trypanosoma* -resistant mutation in two human populations. *Proc. Natl. Acad. Sci.* **122**, e2424538122 (2025).
28. Shoag, J. E. *et al.* Direct measurement of the male germline mutation rate in individuals using sequential sperm samples. *Nat. Commun.* **16**, 2546 (2025).
29. Garcia-Salinas, O. I. *et al.* The impact of ancestral, genetic, and environmental influences on germline de novo mutation rates and spectra. *Nat. Commun.* **16**, 4527 (2025).
30. Jónsson, H. *et al.* Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**, 519–522 (2017).
31. O'Roak, B. J. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* **43**, 585–589 (2011).
32. Porubsky, D. *et al.* Human de novo mutation rates from a four-generation pedigree reference. *Nature* 1–10 (2025).
33. Rahbari, R. *et al.* Timing, rates and spectra of human germline mutation. *Nat. Genet.* **48**, 126–133 (2016).
34. Roach, J. C. *et al.* Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. *Science* **328**, 636–639 (2010).
35. Awadalla, P. *et al.* Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am. J. Hum. Genet.* **87**, 316–324 (2010).
36. Besenbacher, S. *et al.* Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat. Commun.* **6**, 5969 (2015).
37. Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
38. Michaelson, J. J. *et al.* Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* **151**, 1431–1442 (2012).
39. Wong, W. S. *et al.* New observations on maternal age effect on germline de novo mutations. *Nat. Commun.* **7**, 10486 (2016).
40. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
41. Ségurel, L., Wyman, M. J. & Przeworski, M. Determinants of Mutation Rate Variation in the Human Germline. *Annu. Rev. Genomics Hum. Genet.* **15**, 47–70 (2014).
42. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223 (2013).
43. Wu, F. L. *et al.* A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *PLoS Biol.* **18**, e3000838 (2020).

44. Elango, N., Lee, J., Peng, Z., Loh, Y.-H. E. & Yi, S. V. Evolutionary rate variation in Old World monkeys. *Biol. Lett.* **5**, 405–408 (2009).
45. Seiffert, E. R. Early primate evolution in Afro-Arabia. *Evol. Anthropol. Issues News Rev.* **21**, 239–253 (2012).
46. Rivas-González, I. *et al.* Pervasive incomplete lineage sorting illuminates speciation and selection in primates. *Science* **380**, eabn4409 (2023).
47. Bergeron, L. A. *et al.* The Mutationathon highlights the importance of reaching standardization in estimates of pedigree-based germline mutation rates. *Elife* **11**, e73577 (2022).
48. Gao, Z. *et al.* Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc. Natl. Acad. Sci.* **116**, 9491–9500 (2019).
49. Scally, A. The mutation rate in human evolution and demographic inference. *Curr. Opin. Genet. Dev.* **41**, 36–43 (2016).
50. Noyes, M. D. *et al.* Long-read sequencing of families reveals increased germline and postzygotic mutation rates in repetitive DNA. *Nat. Commun.* https://www.nature.com/articles/s41467-026-70342-1 (2026).
51. Sasani, T. A. *et al.* Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *Elife* **8**, e46922 (2019).
52. Campbell, C. D. *et al.* Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* **44**, 1277–1281 (2012).
53. Palamara, P. F. *et al.* Leveraging distant relatedness to quantify human mutation and gene-conversion rates. *Am. J. Hum. Genet.* **97**, 775–789 (2015).
54. Tian, X., Browning, B. L. & Browning, S. R. Estimating the Genome-wide Mutation Rate with Three-Way Identity by Descent. *Am. J. Hum. Genet.* **105**, 883–893 (2019).
55. Kohailan, M. *et al.* Patterns and distribution of de novo mutations in multiplex Middle Eastern families. *J. Hum. Genet.* **67**, 579–588 (2022).
56. Enrico Bena, C. *et al.* Real-time monitoring of replication errors' fate reveals the origin and dynamics of spontaneous mutations. *Nat. Commun.* **15**, 2702 (2024).
57. Lindahl, T. Instability and decay of the primary structure of DNA. *nature* **362**, 709–715 (1993).
58. Heller, C. G. & Clermont, Y. Spermatogenesis in Man: An Estimate of Its Duration. *Science* **140**, 184–186 (1963).
59. Spisak, N., de Manuel, M., Milligan, W., Sella, G. & Przeworski, M. The clock-like accumulation of germline and somatic mutations can arise from the interplay of DNA damage and repair. *PLoS Biol.* **22**, e3002678 (2024).
60. Schmutte, C., Yang, A. S., Beart, R. W. & Jones, P. A. Base excision repair of U: G mismatches at a mutational hotspot in the p53 gene is more efficient than base excision repair of T: G mismatches in extracts of human colon tumors. *Cancer Res.* **55**, 3742–3746 (1995).
61. Bloom, L. B. *et al.* Pre-Steady-State Kinetic Analysis of Sequence-Dependent Nucleotide Excision by the 3'-Exonuclease Activity of Bacteriophage T4 DNA Polymerase. *Biochemistry* **33**, 7576–7586 (1994).
62. Petruska, J. & Goodman, M. F. Influence of neighboring bases on DNA polymerase insertion and proofreading fidelity. *J. Biol. Chem.* **260**, 7533–7539 (1985).
63. Harris, K. & Nielsen, R. Error-prone polymerase activity causes multinucleotide mutations in humans. *Genome Res.* **24**, 1445–1454 (2014).
64. Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).
65. Zhu, L., Beichman, A. & Harris, K. Population size interacts with reproductive longevity to shape the germline mutation rate. *Proc. Natl. Acad. Sci.* **122**, e2423311122 (2025).
66. Jinks-Robertson, S. & Bhagwat, A. S. Transcription-Associated Mutagenesis. *Annu. Rev. Genet.* **48**, 341–359 (2014).
67. Li, F. *et al.* The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell* **153**, 590–600 (2013).
68. Goldmann, J. M. *et al.* Parent-of-origin-specific signatures of de novo mutations. *Nat. Genet.* **48**, 935–939 (2016).

69. Gao, Z., Zhang, Y., Cramer, N., Przeworski, M. & Moorjani, P. Limited role of generation time changes in driving the evolution of the mutation spectrum in humans. *elife* **12**, e81188 (2023).
70. Harris, K. Evidence for recent, population-specific evolution of the human mutation rate. *Proc. Natl. Acad. Sci.* **112**, 3439–3444 (2015).
71. Harris, K. & Pritchard, J. K. Rapid evolution of the human mutation spectrum. *elife* **6**, e24284 (2017).
72. Speidel, L. *et al.* Inferring population histories for ancient genomes using genome-wide genealogies. *Mol. Biol. Evol.* **38**, 3497–3511 (2021).
73. Amos, W. Flanking heterozygosity influences the relative probability of different base substitutions in humans. *R. Soc. Open Sci.* **6**, 191018 (2019).
74. Amos, W. Signals interpreted as archaic introgression appear to be driven primarily by faster evolution in Africa. *R. Soc. Open Sci.* **7**, 191900 (2020).
75. Kessler, M. D. *et al.* De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population. *Proc. Natl. Acad. Sci.* **117**, 2560–2569 (2020).
76. Otto, S. P. & Orive, M. E. Evolutionary consequences of mutation and selection within an individual. *Genetics* **141**, 1173–1187 (1995).
77. Smith, M. J. *et al.* Loss-of-function mutations in SMARCE1 cause an inherited disorder of multiple spinal meningiomas. *Nat. Genet.* **45**, 295–298 (2013).
78. Braude, P., Bolton, V. & Moore, S. Human gene expression first occurs between the four-and eight-cell stages of preimplantation development. *Nature* **332**, 459–461 (1988).
79. Dobson, A. T. *et al.* The unique transcriptome through day 3 of human preimplantation development. *Hum. Mol. Genet.* **13**, 1461–1470 (2004).
80. Zhang, K. & Smith, G. W. Maternal control of early embryogenesis in mammals. *Reprod. Fertil. Dev.* **27**, 880–896 (2015).
81. Karabulut, S., Kutlu, P., Korkmaz, O. & Oria, L. Impact of Maternal Age on the Repairing Capacity of Oocytes on Paternal DNA Damage. *Reprod. Sci.* **32**, 2397–2403 (2025).
82. Sun, F. *et al.* Increased DNA damage in full-grown oocytes is correlated with diminished autophagy activation. *Nat. Commun.* **15**, 9463 (2024).
83. Besenbacher, S., Hvilsom, C., Marques-Bonet, T., Mailund, T. & Schierup, M. H. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat. Ecol. Evol.* **3**, 286–292 (2019).
84. Wong, A. Covariance between testes size and substitution rates in primates. *Mol. Biol. Evol.* **31**, 1432–1436 (2014).
85. Arnheim, N. & Calabrese, P. Germline Stem Cell Competition, Mutation Hot Spots, Genetic Disorders, and Older Fathers. *Annu. Rev. Genomics Hum. Genet.* **17**, 219–243 (2016).
86. Hodge, R. A. & Bach, E. A. Mechanisms of Germline Stem Cell Competition across Species. *Life* **14**, 1251 (2024).
87. Otto, S. P. & Hastings, I. M. Mutation and selection within the individual. *Genetica* **102–103**, 507–524 (1998).
88. Choi, S.-K., Yoon, S.-R., Calabrese, P. & Arnheim, N. A germ-line-selective advantage rather than an increased mutation rate can explain some unexpectedly common human disease mutations. *Proc. Natl. Acad. Sci.* **105**, 10143–10148 (2008).
89. Yoon Cho, S. *et al.* The proportion of uniparental disomy is increased in Prader–Willi syndrome due to an advanced maternal childbearing age in Korea. *J. Hum. Genet.* **58**, 150–154 (2013).
90. Seplyarskiy, V. *et al.* Hotspots of human mutation point to clonal expansions in spermatogonia. *Nature* 1–7 (2025).
91. Lynch, M. *et al.* Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **17**, 704–714 (2016).
92. Galhardo, R. S., Hastings, P. J. & Rosenberg, S. M. Mutation as a Stress Response and the Regulation of Evolvability. *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435 (2007).
93. Martincorena, I., Seshasayee, A. S. & Luscombe, N. M. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* **485**, 95–98 (2012).
94. Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E. & Merrikh, H. Accelerated gene evolution through replication–transcription conflicts. *Nature* **495**, 512–515 (2013).

95. Payne, J. L. & Wagner, A. The causes of evolvability and their evolution. *Nat. Rev. Genet.* **20**, 24–38 (2019).
96. Rosenberg, S. M. Evolving responsively: adaptive mutation. *Nat. Rev. Genet.* **2**, 504–515 (2001).
97. Pal, C., Maciá, M. D., Oliver, A., Schachar, I. & Buckling, A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**, 1079–1081 (2007).
98. Foster, P. L. Stress responses and genetic variation in bacteria. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **569**, 3–11 (2005).
99. Foerster, V. *et al.* Pleistocene climate variability in eastern Africa influenced hominin evolution. *Nat. Geosci.* **15**, 805–811 (2022).
100. Scerri, E. M. L. *et al.* Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter? *Trends Ecol. Evol.* **33**, 582–594 (2018).
101. Hallett, E. Y. *et al.* Major expansion in the human niche preceded out of Africa dispersal. *Nature* 1–7 (2025).
102. Timbrell, L., Grove, M., Manica, A., Rucina, S. & Blinkhorn, J. A spatiotemporally explicit paleoenvironmental framework for the Middle Stone Age of eastern Africa. *Sci. Rep.* **12**, 3689 (2022).
103. Barnett, M., Meister, L. & Rainey, P. B. Experimental evolution of evolvability. *Science* **387**, eadr2756 (2025).
104. Cagan, A. *et al.* Somatic mutation rates scale with lifespan across mammals. *Nature* **604**, 517–524 (2022).
105. Tollis, M., Schiffman, J. D. & Boddy, A. M. Evolution of cancer suppression as revealed by mammalian comparative genomics. *Curr. Opin. Genet. Dev.* **42**, 40–47 (2017).
106. Wang, R. J., Al-Saffar, S. I., Rogers, J. & Hahn, M. W. Human generation times across the past 250,000 years. *Sci. Adv.* **9**, eabm7047 (2023).
107. Yang, C. *et al.* Evolutionary and biomedical insights from a marmoset diploid genome assembly. *Nature* **594**, 227–233 (2021).
108. Thomas, G. W. *et al.* Reproductive longevity predicts mutation rates in primates. *Curr. Biol.* **28**, 3193–3197 (2018).
109. Pfeifer, S. P. Direct estimate of the spontaneous germ line mutation rate in African green monkeys. *Evolution* **71**, 2858–2870 (2017).
110. Hrdy, S. B. Mothers and others. (2009).
111. Marlowe, F. W. Hunter-gatherers and human evolution. *Evol. Anthropol. Issues News Rev. Issues News Rev.* **14**, 54–67 (2005).
112. Lynch, M. The lower bound to the evolution of mutation rates. *Genome Biol. Evol.* **3**, 1107–1118 (2011).
113. Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. & Lynch, M. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci.* **109**, 18488–18492 (2012).
114. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
115. Weinstein, B. & Roy, S. W. Controlling for life-history traits in vertebrates reveals that effective population size does not affect mutation rate or genome size. *Proc. Natl. Acad. Sci.* **123**, e2519649123 (2026).
116. Li, W.-H., Ellsworth, D. L., Krushkal, J., Chang, B. H.-J. & Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation–time effect hypothesis. *Mol. Phylogenet. Evol.* **5**, 182–187 (1996).
117. Ohta, T. An examination of the generation-time effect on molecular evolution. *Proc. Natl. Acad. Sci.* **90**, 10676–10680 (1993).
118. Chang, B. H., Shimmin, L. C., Shyue, S. K., Hewett-Emmett, D. & Li, W. H. Weak male-driven molecular evolution in rodents. *Proc. Natl. Acad. Sci.* **91**, 827–831 (1994).
119. Elango, N., Thomas, J. W., NISC Comparative Sequencing Program§ & Yi, S. V. Variable molecular clocks in hominoids. *Proc. Natl. Acad. Sci.* **103**, 1370–1375 (2006).
120. Langergraber, K. E. *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci.* **109**, 15716–15721 (2012).

121. Dean, M. C. Measures of maturation in early fossil hominins: events at the first transition from australopiths to early *Homo*. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150234 (2016).
122. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
123. Bellis, M. A., Downing, J. & Ashton, J. R. Adults at 12? Trends in puberty and their public health consequences. *Journal of Epidemiology & Community Health* vol. 60 910–911 (2006).
124. Kelly, R. L. *The Foraging Spectrum: Diversity in Hunter-Gatherer Lifeways*. (Smithsonian Inst Press, 1995).
125. Mace, R. Evolutionary ecology of human life history. *Anim. Behav.* **59**, 1–10 (2000).
126. Robson, S. L. & Wood, B. Hominin life history: reconstruction and evolution. *J. Anat.* **212**, 394–425 (2008).
127. Schwartz, G. T. Growth, Development, and Life History throughout the Evolution of *Homo*. *Curr. Anthropol.* **53**, S395–S408 (2012).
128. Thomas, G. W. & Hahn, M. W. The human mutation rate is increasing, even as it slows. *Mol. Biol. Evol.* **31**, 253–257 (2014).
129. Parker, G. A. Sperm competition and its evolutionary effect on copula duration in the fly Scatophaga stercoraria. *J. Insect Physiol.* **16**, 1301–1328 (1970).
130. Ramm, S. A., Schärer, L., Ehmcke, J. & Wistuba, J. Sperm competition and the evolution of spermatogenesis. *Mol. Hum. Reprod.* **20**, 1169–1179 (2014).
131. Whitlock, M. C. & Agrawal, A. F. Purging the genome with sexual selection: reducing mutation load through selection on males. *Evolution* **63**, 569–582 (2009).
132. Smithwick, E. B., Young, L. G. & Gould, K. G. Duration of spermatogenesis and relative frequency of each stage in the seminiferous epithelial cycle of the chimpanzee. *Tissue Cell* **28**, 357–366 (1996).
133. Venn, O. *et al.* Strong male bias drives germline mutation in chimpanzees. *Science* **344**, 1272–1275 (2014).
134. Auger, J., Eustache, F., Chevrier, C. & Jégou, B. Spatiotemporal trends in human semen quality. *Nat. Rev. Urol.* **19**, 597–626 (2022).
135. Bataillon, T. *et al.* The effective size of the Icelandic population and the prospects for LD mapping: inference from unphased microsatellite markers. *Eur. J. Hum. Genet.* **14**, 1044–1053 (2006).
136. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, (2020).
137. Lewis, A. C. F. *et al.* Getting genetic ancestry right for science and society. *Science* **376**, 250–252 (2022).
138. Fan, S. *et al.* Whole-genome sequencing reveals a complex African population demographic history and signatures of local adaptation. *Cell* **186**, 923–939 (2023).
139. Lipson, M. *et al.* High continuity of forager ancestry in the Neolithic period of the eastern Maghreb. *Nature* 1–7 (2025).
140. Padilla-Iglesias, C. *et al.* Pan-African metapopulation model explains Homo sapiens genetic and morphological evolution. *bioRxiv* 2025–05 (2025).
141. Schlebusch, C. M. *et al.* Khoe-San Genomes Reveal Unique Variation and Confirm the Deepest Population Divergence in Homo sapiens. *Mol. Biol. Evol.* **37**, 2944–2954 (2020).
142. Rivers, F., Grollman, C., Khan, Z. & Daniele, M. A. S. The relationship between ethnicity and place of birth in England: a mixed-methods study. *BMC Pregnancy Childbirth* **24**, 798 (2024).
143. Gomez, F., Hirbo, J. & Tishkoff, S. A. Genetic Variation and Adaptation in Africa: Implications for Human Evolution and Disease. *Cold Spring Harb. Perspect. Biol.* **6**, a008524–a008524 (2014).
144. Huang, Z., Kelleher, J., Chan, Y. & Balding, D. Estimating evolutionary and demographic parameters via ARG-derived IBD. *PLOS Genet.* **21**, e1011537 (2025).
145. Biobank, M. B. A. L. 18 W. S. J. 18 W. V. A. 18 M. J. G. 18 C. M. S. 10 18 & A. 15, A. of U. R. D. P. T. C. S. H. 14 http://orcid org/0000-0002-0322-8970 W. X. 14 http://orcid org/0000-0001-6042-4487 R. E. Genomic data in the all of us research program. *Nature* **627**, 340–346 (2024).

146. Biobank, M. B. A. L. 18 W. S. J. 18 W. V. A. 18 M. J. G. 18 C. M. S. 10 18 & A. 15, A. of U. R. D. P. T. C. S. H. 14 http://orcid org/0000-0002-0322-8970 W. X. 14 http://orcid org/0000-0001-6042-4487 R. E. Genomic data in the all of us research program. *Nature* **627**, 340–346 (2024).
147. Keightley, P. D., Ness, R. W., Halligan, D. L. & Haddrill, P. R. Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics* **196**, 313–320 (2014).
148. Keightley, P. D. *et al.* Estimation of the spontaneous mutation rate in Heliconius melpomene. *Mol. Biol. Evol.* **32**, 239–243 (2015).
149. Monroe, J. G. *et al.* Mutation bias reflects natural selection in Arabidopsis thaliana. *Nature* **602**, 101–105 (2022).
150. Mukai, T. The genetic structure of natural populations of Drosophila melanogaster. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**, 1 (1964).
151. Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. Mutation rate and dominance of genes affecting viability in Drosophila melanogaster. *Genetics* **72**, 335–355 (1972).
152. Muller, H. J. The measurement of gene mutation rate in Drosophila, its high variability, and its dependence upon temperature. *Genetics* **13**, 279 (1928).
153. Ohnishi, O. Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in Drosophila melanogaster. II. Homozygous effect of polygenic mutations. *Genetics* **87**, 529–545 (1977).
154. Wu, K., Qin, D., Qian, Y. & Liu, H. A new era of mutation rate analyses: Concepts and methods. *Zool. Res.* **45**, 767 (2024).
155. Garcia-Salinas, O. I. *et al.* Landscape of parental postzygotic mutations in> 11,000 rare disease trios. *medRxiv* 2025–10 (2025).
156. Charmouh, A. P. *et al.* Estimating gene conversion tract length and rate from PacBio HiFi data. *Mol. Biol. Evol.* **42**, msaf019 (2025).
157. Melamed, D. *et al.* De novo mutation rates at the single-mutation resolution in a human HBB gene region associated with adaptation and genetic disease. *Genome Res.* **32**, 488–498 (2022).
158. Lenski, R. E. & Mittler, J. E. The Directed Mutation Controversy and Neo-Darwinism. *Science* **259**, 188–194 (1993).