



HHS Public Access

Author manuscript

Curr Opin Genet Dev. Author manuscript; available in PMC 2021 June 30.

Published in final edited form as:

Curr Opin Genet Dev. 2020 June ; 62: 50–57. doi:10.1016/j.gde.2020.05.024.

Inferring evolutionary dynamics of mutation rates through the lens of mutation spectrum variation

Jedidiah Carlson¹, William S. DeWitt^{1,2}, Kelley Harris^{1,2}

¹Department of Genome Sciences, Foege Hall, University of Washington, Seattle, WA 98105

²Computational Biology Program, Fred Hutchinson Cancer Research Center, 1100 Eastlake Ave E, Seattle, WA 98109

Abstract

There are many possible failure points in the transmission of genetic information that can produce heritable germline mutations. Once a mutation has been passed from parents to offspring for several generations, it can be difficult or impossible to identify its root cause; however, sometimes the nature of the ancestral and derived DNA sequences can provide mechanistic clues about a genetic change that happened hundreds or thousands of generations ago. Here, we review evidence that the sequence context “spectrum” of germline mutagenesis has been evolving surprisingly rapidly over the history of humans and other species. We go on to discuss possible causal factors that might underlie rapid mutation spectrum evolution.

Introduction

Like all other complex traits, the germline mutations present in an individual’s genome are ultimately governed by heritable genetic factors, environmental influences, and interactions thereof. Commonly studied properties of germline mutations include their rate (i.e., mutations per site per generation), spectrum (i.e., relative abundances of different mutation types), or spatial clustering throughout the genome. Because these phenotypic outcomes of mutation are embedded as variation in the genome, the evolutionary pressures acting on mutation phenotypes are intrinsically more complex than the forces that drive the evolution of other phenotypes. The theoretical complexities of this phenomenon have fascinated population geneticists for decades. In 1937, A.H. Sturtevant was the first to show that alleles that modify the mutation rate become linked over time to mutations that they beget in nearby genomic regions, leading to time-delayed selection on mutation rate modifiers that is ultimately dependent on the fitness effects of much younger genetic variants [1]. Thirty years after Sturtevant’s seminal paper, population geneticists remained daunted by the

carjed@uw.edu, Lab phone: 1-206-543-8918.

Author declaration

¹-Conflict of Interest

No conflict of interest exists.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

higher-order complexity of mutation rate evolution: Motoo Kimura wrote in 1967 that “[t]he whole question of the evolutionary modification of the spontaneous mutation rate is quite puzzling and more evidence is needed to clarify the problem” [2].

In the fifty-three years since Kimura’s landmark theoretical work, large amounts of genetic data have been produced, proffering new empirical evidence of past and ongoing mutation rate evolution. It is now feasible to directly measure variability in mutation rates by generating mutation accumulation lines and/or sequencing parent/offspring trios, then test hypotheses about the genetic and environmental causes of ongoing within-species germline mutation rate variability. Comparative genomics approaches have shown that, over long timescales, mutation rates evolve toward a level that is approximately proportional to effective population size divided by the size of the coding fraction of the genome [3].

However, between the extremes of inter-species mutation rate evolution and within-species mutation rate variation, there has been much less documentation of historical germline mutation rate variation within populations or between closely related populations, a fact that stands in the way of studying the evolution of this trait using standard techniques from quantitative genetics. To make matters even more complicated, environmental mutagens might also affect mutation rates, as do life history features such as the ages at which individuals tend to become parents. In a neutral model setting, disentangling historical mutation rate evolution from the effects of demographic history is a serious challenge for population genetic inference. Indeed, the rate of genetic drift influencing genomic variation is jointly determined by mutation rate and effective population size.

An indirect hint that mutational modifiers do segregate within populations is the existence of variation between populations in their mutation spectra. Mutations can be coarsely classified into 6 types according to the ancestral and derived alleles (A>C, A>G, A>T, C>A, C>G, and C>T are all distinct mutation types, but T>G, for example, is not distinct from this set because it is the strand complement of A>C). There is considerable variation between microorganism species (and even strains of the same species) in the relative abundances of these variant types [4], but to ascertain differences between closely related eukaryotes that have larger genomes and more stable mutation rates, it is useful to partition these mutation types further by trinucleotide context (yielding 96 unique types such as AAA>ACA, etc.) or even by extended sequence contexts as large as 7-mers [5,6]. Trinucleotide mutation spectrum variation has been extensively catalogued across cancer types, an effort that has led to the discovery of dozens of “mutational signatures” that appear to be associated with specific exogenous and endogenous DNA damage agents [7,8]. More recently, comparisons of mutation spectra between populations of humans, great apes, and laboratory mice has revealed that mutation spectrum phenotypes often vary so predictably between populations that they can be used to identify an individual’s population of origin [9–14] [see also: Taliun et al., *bioRxiv* doi: [10.1101/563866](https://doi.org/10.1101/563866); Goldberg and Harris, *bioRxiv* doi: [10.1101/805598](https://doi.org/10.1101/805598)].

The most striking mutation spectrum difference discovered thus far between closely related human populations is an excess of TCC>TTC mutations that exists in Europeans relative to East Asians and Africans. In the 1000 Genomes Phase I dataset, where this difference was first reported [9], private European variation is enriched for TCC>TTC mutations by about

50% compared to private African or East Asian variation. South Asian genomes contain an intermediate amount of TCC>TTC mutations, perhaps because this population was founded by admixture of a European-related Ancestral North Indian population into an East-Asian-related Ancestral South Indian population [15,16]. Based on the allele frequencies of the excess TCC>TTC mutations in Europe, it has been estimated that the rate of this mutation increased in Europe between 15,000 to 30,000 years ago [11,17] (or perhaps even earlier, depending on the details of ancient European demographic history). It is also clear from the mutation spectrum of rare variants that the TCC>TTC mutation rate decreased again around 1,000 to 2,000 years ago. In direct sequencing data from parent child trios, the rate of this mutation type is indistinguishable between Europeans and other populations, suggesting that the pulse of excess mutagenesis either subsided entirely or is confined to a small subpopulation that has not been included in any trio sequencing efforts.

A few similar mutation types, including TCT>TTT, ACC>ATC, and CCC>CTC, are enriched in Europeans with the same marginal allele frequency distribution as TCC>TTC, suggesting that these are minor components of the same mutational signature [11,12]. However, other mutation types show differentiation between populations that becomes strictly stronger with decreasing allele frequency. The first principal component of human mutation spectrum variation is not defined by TCC>TTC mutation abundance at all, but instead separates Africans from Eurasians. African mutations are slightly more likely than Eurasian mutations to have G/C ancestral alleles compared to A/T ancestral alleles, a trend that is not very dependent on sequence context and does not dramatically affect the frequency distribution of any specific trinucleotide mutation type. In addition, certain complex mutations that occur at neighboring sites show evidence of even more population differentiation than simple SNPs do [18].

An apparent separate pulse of NAC>NCC mutations in the history of the Japanese population was also reported by Harris and Pritchard; however, this mutation type appears to be strongly influenced by a cell line artifact or other technical issue that caused context specific errors in 1000 Genomes variant calls [19]. An overall excess of NAC>NCC mutations in East Asians compared to Europeans has been confirmed in two independent datasets [11], but it is not clear how this relates to the presence of similar high-frequency artifactual mutations in Japanese genomes that were sequenced using an older technological platform.

At present, evidence of the causal mechanisms underlying germline mutation rate evolution in human populations is virtually nonexistent. In this review, we aim to summarize the recent proliferation of evidence for mutation spectrum variation at the levels of individuals, populations, and species, and explore the range of hypotheses about what may have caused these curious patterns of genetic variation.

The mutator allele hypothesis

DNA replication and repair are among the most essential of biological processes and are necessary for maintaining the fidelity of each species' genome from generation to generation. Sturtevant and Kimura posited that the mechanisms responsible for replication

and repair, if mutated themselves, would alter the efficiency of these essential processes. We now know that there are dozens, if not hundreds, of genes involved in DNA replication and repair [20], most of which are highly conserved across eukaryotic organisms [21,22]. Even so, the processes of replication and repair have evolved in each species' lineage, and the sequence and function of genes involved in highly conserved processes can diverge dramatically between species (an example is base excision repair [BER], a repair process that is operative in both yeast and mammals, but yeast BER pathways are far less efficient at repairing oxidative damage [23]). Inter-species variation in such genes is thus thought to be directly linked to inter-species variation in the mutation rate [3].

The inter-species variation in DNA replication and repair machinery and mutation rates leads us to question whether within-species variation of these essential genes can also explain within-species variation in mutation rates, both between individuals and looking back in time. Harris and Pritchard, who were the first to show evidence for a temporal pulse of elevated mutation rate at a specific triplet context in one population (TCC>TTC in Europeans), hypothesized that these dynamics are driven by the appearance and subsequent drift of mutator alleles that modify the germline mutation rate [11].

This hypothesis is bolstered by evidence that high mutation rates sometimes emerge spontaneously in experimentally evolving populations of microorganisms [24–27]. A recent landmark experimental evolution study found that mutator alleles arose spontaneously in six of twelve experimental *Escherichia coli* populations that were maintained in stable, identical environments for over 60,000 generations [26]. In some cases, mutator alleles in these populations even reached fixation, and in others, the arrival of a mutator allele was followed by the emergence of an “antimutator” allele that caused the accelerated rate of mutation induced by the initial mutator to slow [26] [see also Tracy et al., *bioRxiv* doi.org/10.1101/718163]. Importantly, mutator alleles are not exclusive to rapidly evolving bacteria or yeast. Mutator alleles are one of the hallmarks of human cancers [28], and viable strains of mice possessing germline mutator alleles have been bred in the lab (albeit with smaller litter sizes and lower rates of reproduction, pregnancy, and survival) [29].

The existence of mutator alleles in microorganisms, cancer cell populations, and lab-engineered mice has fueled tantalizing speculation that mutator alleles have been (and continue to be) a ubiquitous feature of every species' genome and evolutionary history. If mutator alleles are segregating nearly neutrally in human populations today, some will be predicted to fix over time. Some evidence for past mutator allele fixation can be found in great ape genomic diversity: closely related ape species like humans and chimps have more similar mutation spectra than distantly related species like humans and orangutans, despite the fact that chimps and orangutans inhabit much more similar environments today [Goldberg and Harris, *bioRxiv* doi: 10.1101/805598]. Proving that mutator alleles have arisen in and left a mark on the human genome, however, is an exceptionally challenging task. Since humans are a sexually-reproducing species, any mutator alleles arising in the genome will quickly become decoupled from any beneficial mutations they induce because of recombination, so, depending on the distribution of mutational fitness effects, most mutator alleles are expected to segregate neutrally or rapidly drop in frequency due to selection acting against the deleterious mutations that they create on an extended linked

haplotype [2]. Consequently, the only hope of identifying mutator alleles from population sequencing data in sexually-reproducing organisms is to search for haplotypes carrying an excess of derived alleles and identify candidate mutator alleles on the same haplotype that have remained in linkage disequilibrium with the mutations they have induced [30].

If the TCC>TTC pulse was the result of a mutator allele, that mutator has likely been lost from the population, so evidence of its genomic location and effect size may be lost to time. (Another possibility is that a mutator allele may remain in the population after attenuation by the arrival of an antimutator allele, which could explain why the TCC>TTC rate in present populations does not appear to have fallen back to its pre-pulse rate). However, it is possible that a TCC>TTC mutator still exists at extremely low frequency, or that other mutator alleles contribute to background levels of germline mutational signature differentiation. A mutator with a strong effect size could be identified by sequencing a carrier parent-child trio, and a mutator with a weaker effect size but high frequency could be identified by looking for a signal of elevated mutagenesis along an extended shared haplotype. However, attempting to detect this signal of spatially-clustered low-frequency variants may be extremely difficult because similar signals of genomic variation can be caused by sequencing artifacts [19], spatial variation in mutation rates across the genome [6] [Taliun et al., *bioRxiv* doi: [10.1101/563866](https://doi.org/10.1101/563866)], natural selection [31], and demographic history [32,33].

The generation time hypothesis

A second line of reasoning proposes that temporal shifts in the mutation spectrum evident from population sequencing data might reflect changes in life history traits among archaic human populations. There is abundant archaeological [34], anthropological [35], and genetic [36] evidence that life history traits have changed over the course of human history, and fundamental principles of biology dictate that life history traits—specifically generation time—are directly related to changes in the germline mutation rate [37]. In species that are genetically predisposed to have very short generation times and lifespans, this generation time hardwiring may relax selection against mutators due to reduced opportunity for deleterious mutations to occur over a shorter reproductive lifespan. This effect has been observed in the mitochondrial DNA of short-lived African killifish, which has the shortest known lifespan of any vertebrate [38].

The relationship between generation time and mutation rate in humans has been proven conclusively in sequencing studies of parent-offspring trios and larger pedigrees: the number of *de novo* point mutations increases roughly linearly by approximately 1 additional mutation per year of the father's age and 1 additional mutation per 3–4 years of the mother's age [39–45]. Recent evidence also suggests that the maternal mutation rate may actually increase exponentially with age rather than linearly [44]. Importantly, these effects of parental age are not universally true for all families: analyses of sequencing data from families with multiple children have found that paternal age effects can vary by an order of magnitude between families, indicating the mechanisms underlying this effect may themselves be under genetic and/or environmental control [40,45].

Humans have longer reproductive life cycles than their closest great ape relatives, and one recent study calculated that this delayed onset of puberty and parenthood was sufficient to explain the difference in mutation rate between human and owl monkey pedigrees [46]. However, a study of great ape pedigrees found that chimpanzees, orangutans, and gorillas accumulate about 50% more mutations per year of paternal age than humans do, [47] suggesting that the dependence of mutation rate on life history has been evolving in concert with life history itself.

Not only have studies found that the mutation *rate* increases with parental age, but also that the mutation *spectrum* changes. As fathers age, the rate of CpG>TpG mutations increases more rapidly than other types, and as mothers age, C>G mutations increase at a faster rate [42,44] and accumulate preferentially in specific hotspots [39,42,48], but to a lesser extent mark the entirety of the genome and inject a generation time signature into the mutation spectrum [Seplyarskiy et al., *bioRxiv* doi:10.1101/2020.01.10.893024]. Hypothetically, if a population sustains older generation times over a period of many generations, the spectrum of variation for mutations that arose during that period could look quite different from that of mutations that arose in previous generations [49]. In a recent study of Neanderthal introgression in the Icelandic population, Skov et al. [50] carried out a similar analysis and found that introgressed loci carried more C>G and fewer T>C and CpG>TpG variants compared to the non-introgressed regions of the genome. Given that these mutation types track maternal and paternal age, the authors concluded that, relative to humans, Neanderthals that interbred with humans may have had older mothers and younger fathers on average. Intriguingly, TCC>TTC is a mutation type that is significantly associated with maternal origin over paternal origin, though it is not clear how this association is related to the enrichment of TCC>TTC mutations in Europe and South Asia [51].

Although generation time variation can leave a predictable imprint on the mutation spectrum, this is unlikely to be a major driving cause of the historic mutation spectrum differences that exist between populations. Assuming that generation time has the same effect on mutagenesis in every genetic background, it can only explain a single dimension of mutation spectrum drift and will tend to explain variation within populations better than variation between populations. A caveat is the existence of hints that generation time has different effects on mutation rate in different genetic backgrounds, which might imply that it is capable of explaining more than one dimension of mutation spectrum variation [45]. In any case, exploring this dimension of variation could prove highly rewarding because of its multiplicity of links to cultural and environmental changes as well as genetic determinants of the ages of puberty and menopause. Even if generation time alone cannot perfectly explain mutation spectrum drift over time, principled estimation of the generation time is a major source of uncertainty for demographic inference [37], so using mutation spectra to estimate generation time distributions could increase the accuracy of demographic inference, provided the mutation spectra are not hopelessly confounded by biased gene conversion [52].

The environmental mutagen hypothesis

A third possibility is that the temporal variation in human mutation rate and spectra might be explained by some form of environmental exposure. We have long known that certain environmental mutagens can leave distinctive mutation signatures on the genome (e.g., mutagens in tobacco smoke are associated with G>T transversion mutations [53]), and in a recent study of the mutagenic effects of 79 environmental agents applied to somatic cell populations, over half were found to confer elevated mutation rates, often with distinctive spectra [54]. The mutation rate in healthy somatic tissues is highly variable, with the highest rates observed in tissues with greater exposure to environmental mutagens, including the skin, lung, blood and esophagus [55,56]. Fewer studies have investigated the effects of environmental mutagens on mutation rates specifically in the germline, but those that have demonstrated that many of the environmental mutagens known to affect somatic cells have similar effects in the germline. Mouse models have successfully determined that parental exposure to benzo(a)pyrene [57] and ionizing radiation [58,59] lead to elevated mutation rates in offspring. In human trio sequencing studies, it has been shown that offspring of fathers exposed to ionizing radiation also carry an increase in multi-site *de novo* mutations [60] and offspring of fathers exposed to dioxin are enriched for A>T mutations [61]. Another study that compared *de novo* mutation patterns ascertained in trios from diverse populations even found a significant reduction in A>G mutations in the offspring of Amish parents and suggested this might be due to the unique environmental conditions of this population [62].

Regarding our focal example of the historic TCC>TTC pulse in European populations, we also know that the environment of the European continent was vastly different 15,000 years ago—the global climate was rapidly warming off the heels of the Last Glacial Maximum and during this period, massive flora and fauna extinction events took place, potentially due to direct and indirect activity of expanding hominin populations [63] and/or climate change [64]. There is also evidence of large-scale burning of biomass that occurred sometime during this period, possibly the result of a disintegrating comet that may have struck the Earth approximately 12,800 years ago [65,66]. These environmental disruptions could potentially have impacted the genomes of archaic human populations in multiple ways: particulate matter from burning of biomass has recently been linked to various forms of DNA damage both *in vivo* and *in vitro* [67,68], and extinction of primary food sources might have dramatically altered the diets of human populations, thus exposing them to new mutagens that humans had not yet evolved to avoid or metabolize [69].

During the late Pleistocene (between 11–19kya), human populations in Europe also acquired adaptive mutations in multiple genes that conferred lighter skin pigmentation, which rapidly swept to high frequency throughout most of Europe [70]. In the initial discovery of the TCC>TTC mutation pulse in Europeans, Harris hypothesized that this novel skin pigmentation phenotype might be indirectly responsible for elevated mutation rates, reasoning that light skin is much more sensitive to ultraviolet radiation, and UV radiation is known to degrade folate, which in turn is linked to endogenous mechanisms of DNA damage [9]. (Harris also rejected the possibility that direct exposure to UV radiation explains the TCC>TTC pulse, because UV radiation generates an excess of pyrimidine

dimers leading to CC>TT tandem mutations—no studies that have replicated the TCC>TTC pulse finding have detected that this was accompanied by an excess of CC>TT mutations in the germline [9,11,12]). However, excess TCC>TTC mutations cannot be an obligatory consequence of adaptation to higher latitudes, since it is not seen in light-skinned East Asian populations. Based on evidence that the mutational signature of the TCC>TTC pulse is similar to that of cancer genomes that were treated with chemotherapy drugs which included alkylating agents (which induce guanine methylation and subsequent G>A mutations) [12], Mathieson and Reich speculated that the ultimate environmental cause was not radiation-related but instead traceable to some unknown mutagenic alkylating agent that may have once been present in the human diet.

Another study found that a transmissible venereal tumor found in canines (CTVT) had a distinctive GTCCA>GTTCA pentamer mutation signature that was active roughly 1.9–8.5kya [71]. The fact that this pentamer is one of 16 possible NTCCN>NTTCN subtypes comprising TCC>TTC mutations, along with the closely-matched timing of the TCC>TTC pulse as estimated in humans [11], presents the rather bizarre possibility that a shared environmental mutagen might have played a role in both the human germline mutation process and the somatic mutation process in CTVT. Supplementary data from a manuscript by Aikens et al. shows that the GTCCA>GTTCA pentamer subtype that dominates the CTVT signature was indeed enriched in Europeans during the period of the TCC>TTC pulse in the human germline [72]. However, relative to other NTCCN>NTTCN subtypes, the *p*-value for enrichment of the GTCCA>GTTCA subtype in Europeans is only ranked 13 out of the 16 possible TCC-related pentamer mutations. This suggests that the dog tumor and the human TCC>TTC pulse may not be affected by the same mutational signature after all, despite their rapid appearance around the same time, unless the root cause was a shared mutagen whose sequence specificity exhibited subtle dependence on the genetic background of the cells being mutated.

Conclusion

Mutation spectrum drift appears to accompany many cases of population divergence, as seen so far in humans, apes, and mice. Despite this intriguing empirical pattern, disentangling the interactions of mutator alleles, changes in life history traits, and environmental mutagens remains almost as challenging today as it was over 50 years ago when Kimura first expressed his exasperation at the task. Understanding the causality of mutation spectrum drift, however, is vital to many of the central questions in population and evolutionary genetics, such as calibrating estimates of demographic and phylogenetic history and understanding historical changes in mutagenic exposures.

Most of this drift consists of small shifts in the relative frequencies of mutations in many trinucleotide contexts (and likely higher order contexts as well), which seems compatible with the random drift of weak, non-specific mutator alleles, subtly tweaking the efficiencies of replication and repair genes. The TCC>TTC pulses in European populations and CTVT cell populations in canines, however, seem fundamentally different from the typical drift pattern in that they are “fast” changes that affected very few sequence contexts and reversed themselves after short periods of time. It is entirely possible that the genetic, environmental,

and life history factors are all partially responsible for the evolution of mutation spectra. Further, these drivers of mutation rate evolution are not necessarily independent from one another; e.g., both environmental and genetic factors might influence life history traits or certain alleles might affect an organism's ability to metabolize mutagenic toxins in the environment.

The TCC>TTC pulse in European populations is the most striking example of mutation spectrum drift in humans known thus far, but it is possible that additional population-specific signatures may be found by studying more geographically diverse human populations, or that more pronounced signatures exist but are specific to certain genomic loci rather than genome-wide averages. Similarly, we speculate that exploring patterns of mutation spectrum variation within and between other species is likely to reveal novel signatures that reflect their unique evolutionary histories. As the field continues to move away from the outdated concept of mutation rate as a static constant and towards treating mutation rate like a dynamic, evolvable phenotype, there are many opportunities to incorporate new data sources and sophisticated simulation software [73], and develop new theoretical tools and inference methods to tackle these fundamental questions of genome evolution.

Acknowledgments

2. Funding

Funding was received for this work. All of the sources of funding for the work described in this publication are acknowledged below: We acknowledge support from the following funders, which played no role in study design, data analysis, or result interpretation: A Burroughs Wellcome Fund Career Award at the Scientific Interface (to K.H.) A Searle Scholarship (to K.H.) A Pew Biomedical Scholarship (to K.H.) An Alfred P. Sloan Fellowship (to K.H.) A Fred Hutchinson Cancer Research Center Translational Data Science Postdoctoral Fellowship (to J.C.) An NIH NIAID F-31 Predoctoral Fellowship (to W.S.D.)

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

References

1. Sturtevant AH: Essays on Evolution. I. On the Effects of Selection on Mutation Rate. *Q Rev Biol* 1937, 12:464–467.
2. Kimura M: On the evolutionary adjustment of spontaneous mutation rates*. *Genet Res* 1967, 9:23–34.
3. Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M: Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci U S A* 2012, 109:18488–18492. [PubMed: 23077252]
4. Lee H, Popodi E, Tang H, Foster PL: Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A* 2012, 109:E2774–83.
5. Aggarwala V, Voight BF: An expanded sequence context model broadly explains variability in polymorphism levels across the human genome. *Nat Genet* 2016, 48:349–355. [PubMed: 26878723]

6. Carlson J, Locke AE, Flickinger M, Zawistowski M, Levy S, Myers RM, Boehnke M, Kang HM, Scott LJ, Li JZ, et al.: Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. *Nat Commun* 2018, 9:3753. [PubMed: 30218074]
7. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR: Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 2013, 3:246–259. [PubMed: 23318258]
8. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, et al.: Signatures of mutational processes in human cancer. *Nature* 2013, 500:415–421. [PubMed: 23945592]
9. Harris K: Evidence for recent, population-specific evolution of the human mutation rate. *Proc Natl Acad Sci U S A* 2015, 112:3439–3444. [PubMed: 25733855]
10. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, et al.: The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 2016, 538:201–206. [PubMed: 27654912]
11. Harris K, Pritchard JK: Rapid evolution of the human mutation spectrum. *Elife* 2017, 6.
12. Mathieson I, Reich D: Differences in the rare variant spectrum among human populations. *PLoS Genet* 2017, 13:e1006581.
13. Narasimhan VM, Rahbari R, Scally A, Wuster A, Mason D, Xue Y, Wright J, Trembath RC, Maher ER, Heel DA van, et al.: Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. *Nat Commun* 2017, 8:303. [PubMed: 28827725]
14. Dumont BL: Significant Strain Variation in the Mutation Spectra of Inbred Laboratory Mice. *Mol Biol Evol* 2019, 36:865–874. [PubMed: 30753674] •• Using sequence data from the Collaborative Cross panel of inbred mouse strains, Dumont demonstrates the existence of population variation in the mutation spectrum that parallels what is seen in humans over a similar scale of genetic divergence.
15. Reich D, Thangaraj K, Patterson N, Price AL, Singh L: Reconstructing Indian population history. *Nature* 2009, 461:489–494. [PubMed: 19779445]
16. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M, et al.: The formation of human populations in South and Central Asia. *Science* 2019, 365.
17. Speidel L, Forest M, Shi S, Myers SR: A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet* 2019, 51:1321–1329. [PubMed: 31477933] • This paper presents a highly resolved ancestral recombination graph of the 1000 Genomes data, complete with a catalog of allele age estimates. They note that these allele age estimates support the existence of a rapid pulse of European TCC>TTC mutations that began at least 30,000 years ago.
18. Prendergast JGD, Pugh C, Harris SE, Hume DA, Deary IJ, Beveridge A: Linked Mutations at Adjacent Nucleotides Have Shaped Human Population Differentiation and Protein Evolution. *Genome Biol Evol* 2019, doi:10.1093/gbe/evz014. • The authors investigate the human population distribution of complex “sequential dinucleotide mutations” (SDMs) where two SNPs arise at adjacent sites in subsequent generations. Using principal component analysis, they find that SDMs exhibit more pronounced population differentiation than simple SNPs do.
19. Anderson-Troc me L, Farouni R, Bourgey M, Kamatani Y, Higasa K, Seo J-S, Kim C, Matsuda F, Gravel S: Legacy Data Confounds Genomics Studies. *Mol Biol Evol* 2019, doi:10.1093/molbev/msz201. •• The authors perform a GWAS on average genomic data quality within the 1000 Genomes Project and find that several Japanese-specific SNPs are associated with this data feature. They show that these data-quality-associated SNPs are the drivers of the Japanese-specific mutation pulse reported in Harris & Pritchard 2017, suggesting that genomes sequenced on older sequencing platforms should be phased out of use in population reference panels.
20. Chatterjee N, Walker GC: Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* 2017, 58:235–263. [PubMed: 28485537]
21. Taylor EM, Lehmann AR: Conservation of eukaryotic DNA repair mechanisms. *Int J Radiat Biol* 1998, 74:277–286. [PubMed: 9737531]
22. Lujan SA, Williams JS, Kunkel TA: DNA Polymerases Divide the Labor of Genome Replication. *Trends Cell Biol* 2016, 26:640–654. [PubMed: 27262731]

23. Kelley MR, Kow YW, Wilson DM 3rd: Disparity between DNA base excision repair in yeast and mammals: translational implications. *Cancer Res* 2003, 63:549–554. [PubMed: 12566294]
24. Lenski RE, Rose MR, Simpson SC, Tadler SC: Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *Am Nat* 1991, 138:1315–1341.
25. Sniegowski PD, Gerrish PJ, Lenski RE: Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 1997, 387:703–705. [PubMed: 9192894]
26. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM: The dynamics of molecular evolution over 60,000 generations. *Nature* 2017, 551:45–50. [PubMed: 29045390]
27. Liu H, Zhang J: Yeast Spontaneous Mutation Rate and Spectrum Vary with Environment. *Curr Biol* 2019, 29:1584–1591.e3. [PubMed: 31056389]
28. Loeb LA: Human Cancers Express a Mutator Phenotype: Hypothesis, Origin, and Consequences. *Cancer Res* 2016, 76:2057–2059. [PubMed: 27197248]
29. Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, Miura I, Wakana S, Nishino J, Yagi T: Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res* 2015, 25:1125–1134. [PubMed: 26129709]
30. Seoighe C, Scally A: Inference of Candidate Germline Mutator Loci in Humans from Genome-Wide Haplotype Data. *PLoS Genet* 2017, 13:e1006549.
31. Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, Yengo L, Rocheleau G, Froguel P, McCarthy MI, et al.: Detection of human adaptation during the past 2000 years. *Science* 2016, 354:760–764. [PubMed: 27738015]
32. Plagnol V, Wall JD: Possible ancestral structure in human populations. *PLoS Genet* 2006, 2:e105. [PubMed: 16895447]
33. Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM: Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* 2018, 173:53–61.e9. [PubMed: 29551270]
34. Gilchrist R: Archaeological biographies: realizing human lifecycles, -courses and -histories. *World Archaeol* 2000, 31:325–328. [PubMed: 16475293]
35. Hill K, Kaplan H: Life history traits in humans: theory and empirical studies. *Annu Rev Anthropol* 1999, 28:397–430. [PubMed: 12295622]
36. Pettay JE, Kruuk LEB, Jokela J, Lummaa V: Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proc Natl Acad Sci U S A* 2005, 102:2838–2843. [PubMed: 15701704]
37. Moorjani P, Amorim CEG, Arndt PF, Przeworski M: Variation in the molecular clock of primates. *Proc Natl Acad Sci U S A* 2016, 113:10607–10612. [PubMed: 27601674]
38. Cui R, Medeiros T, Willemsen D, Iasi LNM, Collier GE, Graef M, Reichard M, Valenzano DR: Relaxed Selection Limits Lifespan by Increasing Mutation Load. *Cell* 2019, 178:385–399.e20. [PubMed: 31257025]
39. Wong WSW, Solomon BD, Bodian DL, Kothiyal P, Eley G, Huddleston KC, Baker R, Thach DC, Iyer RK, Vockley JG, et al.: New observations on maternal age effect on germline de novo mutations. *Nat Commun* 2016, 7:10486. [PubMed: 26781218]
40. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, Dominiczak A, Morris A, Porteous D, Smith B, et al.: Timing, rates and spectra of human germline mutation. *Nat Genet* 2016, 48:126–133. [PubMed: 26656846]
41. Goldmann JM, Wong WSW, Pinelli M, Farrah T, Bodian D, Stittrich AB, Glusman G, Vissers LELM, Hoischen A, Roach JC, et al.: Parent-of-origin-specific signatures of de novo mutations. *Nat Genet* 2016, doi:10.1038/ng.3597.
42. Jónsson H, Sulem P, Kehr B, Kristmundsdóttir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, et al.: Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* 2017, 1:16027.
43. Jónsson H, Sulem P, Arnadóttir GA, Pálsson G, Eggertsson HP, Kristmundsdóttir S, Zink F, Kehr B, Hjorleifsson KE, Jensson BÖ, et al.: Multiple transmissions of de novo mutations in families. *Nat Genet* 2018, 50:1674–1680. [PubMed: 30397338]

44. Gao Z, Moorjani P, Sasani TA, Pedersen BS, Quinlan AR, Jorde LB, Amster G, Przeworski M: Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc Natl Acad Sci U S A* 2019, 116:9491–9500. [PubMed: 31019089] • This study reports that the paternal mutations consistently accumulate three times faster than mutations of maternal origin, regardless of the ages of the parents. This contradicts the classical assumption that the paternal age effect is a result of replication-driven mutation accumulation during spermatogenesis, which would imply that older parents should have a larger paternal age effect than younger parents. The authors propose that their results imply a greater role for DNA damage in mutagenesis than previously believed, and a lesser role for DNA replication errors.
45. Sasani TA, Pedersen BS, Gao Z, Baird L, Przeworski M, Jorde LB, Quinlan AR: Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *Elife* 2019, 8. • This paper reports the existence of variation in parental age effect on mutation rates in a number of human pedigrees originally sampled by the Centre d'Etude Polymorphisme Humaine (CEPH). In each of these families, the same set of parents conceived on the order of ten children, and the collection of these pedigrees exhibits variation in the number of new mutations that accumulate with each additional year of parental age.
46. Thomas GWC, Wang RJ, Puri A, Harris RA, Raveendran M, Hughes DST, Murali SC, Williams LE, Doddapaneni H, Muzny DM, et al.: Reproductive Longevity Predicts Mutation Rates in Primates. *Curr Biol* 2018, 28:3193–3197.e5. [PubMed: 30270182]
47. Besenbacher S, Hvilsom C, Marques-Bonet T, Mailund T, Schierup MH: Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat Ecol Evol* 2019, 3:286–292. [PubMed: 30664699] •• This paper presents de novo mutation rate estimates from multiple great ape families with diverse parental ages and suggests that delayed puberty, delayed reproduction, and a slowdown in the extra mutations accumulated per year of parental age all contribute to the slowdown of the mutation rate during human evolution.
48. Goldmann JM, Seplyarskiy VB, Wong WSW, Vilboux T, Neerinx PB, Bodian DL, Solomon BD, Veltman JA, Deeken JF, Gilissen C, et al.: Germline de novo mutation clusters arise during oocyte aging in genomic regions with high double-strand-break incidence. *Nat Genet* 2018, doi:10.1038/s41588-018-0071-6. •• This paper investigates previously reported hotspots of maternal-age-associated mutagenesis in detail, showing that many of the contributing G>C mutations are actually multinucleotide mutation clusters.
49. Moorjani P, Gao Z, Przeworski M: Human Germline Mutation and the Erratic Evolutionary Clock. *PLoS Biol* 2016, 14:e2000744.
50. Skov L, Coll Macià M, Sveinbjörnsson G, Mafessoni F, Lucotte EA, Einarisdóttir MS, Jonsson H, Halldorsson B, Gudbjartsson DF, Helgason A, et al.: The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* 2020, doi:10.1038/s41586-020-2225-9. • This paper infers the landscape of Neanderthal and Denisovan introgression in a large dataset of Icelandic genomes and finds that a complex web of interbreeding among Neanderthals, Denisovans, and humans is needed to explain the data. Upon testing for mutation spectrum differences between archaic and modern human haplotypes, they find that the axis of differentiation lines up well with the C>G-driven axis of generation time differentiation.
51. Agarwal I, Przeworski M: Signatures of replication timing, recombination, and sex in the spectrum of rare variants on the human X chromosome and autosomes. *Proc Natl Acad Sci U S A* 2019, 116:17916–17924. [PubMed: 31427530] •• The authors report several mutation spectrum differences between compartments of the human genome, including the X chromosome and regions that replicate last during the cell cycle. They go on to show that unique mutation spectrum of the X chromosome is likely shaped by differences between the spectra of maternal and paternal mutations.
52. Pouyet F, Aeschbacher S, Thiéry A, Excoffier L: Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *Elife* 2018, 7.
53. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P: Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* 2002, 21:7435–7451. [PubMed: 12379884]
54. Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, Gomez C, Degasperi A, Harris R, Jackson SP, et al.: A Compendium of Mutational Signatures of Environmental Agents. *Cell* 2019, 177:821–836.e16. [PubMed: 30982602]

55. Werner B, Sottoriva A: Variation of mutational burden in healthy human tissues suggests non-random strand segregation and allows measuring somatic mutation rates. *PLoS Comput Biol* 2018, 14:e1006233.
56. García-Nieto PE, Morrison AJ, Fraser HB: The somatic mutation landscape of the human body. *Genome Biol* 2019, 20:298. [PubMed: 31874648]
57. Beal MA, Meier MJ, Williams A, Rowan-Carroll A, Gagné R, Lindsay SJ, Fitzgerald T, Hurles ME, Marchetti F, Yauk CL: Paternal exposure to benzo(a)pyrene induces genome-wide mutations in mouse offspring. *Commun Biol* 2019, 2:228. [PubMed: 31240266]
58. Adewoye AB, Lindsay SJ, Dubrova YE, Hurles ME: The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. *Nat Commun* 2015, 6:6684. [PubMed: 25809527]
59. Satoh Y, Asakawa J-I, Nishimura M, Kuo T, Shinkai N, Cullings HM, Minakuchi Y, Sese J, Toyoda A, Shimada Y, et al.: Characteristics of induced mutations in offspring derived from irradiated mouse spermatogonia and mature oocytes. *Sci Rep* 2020, 10:37. [PubMed: 31913321]
60. Holtgrewe M, Knaus A, Hildebrand G, Pantel J-T, Santos MR de L, Neveling K, Goldmann J, Schubach M, Jäger M, Coutelier M, et al.: Multisite de novo mutations in human offspring after paternal exposure to ionizing radiation. *Sci Rep* 2018, 8:14611. [PubMed: 30279461]
61. Ton ND, Nakagawa H, Ha NH, Duong NT, Nhung VP, Hien LTT, Hue HTT, Hoang NH, Wong JH, Nakano K, et al.: Whole genome sequencing and mutation rate analysis of trios with paternal dioxin exposure. *Hum Mutat* 2018, 39:1384–1392. [PubMed: 29969170]
62. Kessler MD, Loesch DP, Perry JA, Heard-Costa NL, Taliun D, Cade BE, Wang H, Daya M, Ziniti J, Datta S, et al.: De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population. *Proc Natl Acad Sci U S A* 2020, doi:10.1073/pnas.1902766117. •• The authors estimate mutation rates in the most diverse set of human trios sequenced to date, including multiple African-American and Afro-Caribbean populations. They report that the Amish have the lowest mutation rate of the populations in their study and hypothesize that the “clean” Amish environment may be responsible.
63. Smith FA, Elliott Smith RE, Lyons SK, Payne JL: Body size downgrading of mammals over the late Quaternary. *Science* 2018, 360:310–313. [PubMed: 29674591]
64. Broughton JM, Weitzel EM: Population reconstructions for humans and megafauna suggest mixed causes for North American Pleistocene extinctions. *Nat Commun* 2018, 9:5441. [PubMed: 30575758]
65. Pino M, Abarzúa AM, Astorga G, Martel-Cea A, Cossio-Montecinos N, Navarro RX, Lira MP, Labarca R, LeCompte MA, Adedeji V, et al.: Sedimentary record from Patagonia, southern Chile supports cosmic-impact triggering of biomass burning, climate change, and megafaunal extinctions at 12.8 ka. *Sci Rep* 2019, 9:4413. [PubMed: 30867437]
66. Firestone RB, West A, Kennett JP, Becker L, Bunch TE, Revay ZS, Schultz PH, Belgya T, Kennett DJ, Erlanson JM, et al.: Evidence for an extraterrestrial impact 12,900 years ago that contributed to the megafaunal extinctions and the Younger Dryas cooling. *Proc Natl Acad Sci U S A* 2007, 104:16016–16021. [PubMed: 17901202]
67. Park CG, Cho HK, Shin HJ, Park KH, Lim HB: Comparison of Mutagenic Activities of Various Ultra-Fine Particles. *Toxicol Res* 2018, 34:163–172. [PubMed: 29686778]
68. de Oliveira Alves N, Vessoni AT, Quinet A, Fortunato RS, Kajitani GS, Peixoto MS, Hacon S de S, Artaxo P, Saldiva P, Menck CFM, et al.: Biomass burning in the Amazon region causes DNA damage and cell death in human lung cells. *Sci Rep* 2017, 7:10937. [PubMed: 28883446]
69. Ye K, Gu Z: Recent advances in understanding the role of nutrition in human genome evolution. *Adv Nutr* 2011, 2:486–496. [PubMed: 22332091]
70. Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, Cameron E, Shriver MD, Parra EJ, Rocha J: The timing of pigmentation lightening in Europeans. *Mol Biol Evol* 2013, 30:24–35. [PubMed: 22923467]
71. Baez-Ortega A, Gori K, Strakova A, Allen JL, Allum KM, Bansse-Issa L, Bhutia TN, Bisson JL, Briceño C, Castillo Domracheva A, et al.: Somatic evolution and global expansion of an ancient transmissible cancer lineage. *Science* 2019, 365. •• The authors reconstruct the phylogeny and transmission history of the Canine Transmissible Venereal Tumor and show, among other things,

that it was affected by a burst of TCC>TTC mutations in a specific pentamer context around the same time period that the rate of TCC>TTC mutagenesis was elevated in Europe.

72. Aikens RC, Johnson KE, Voight BF: Signals of Variation in Human Mutation Rate at Multiple Levels of Sequence Context. *Mol Biol Evol* 2019, 36:955–965. [PubMed: 30753705] • The authors revisit mutation spectrum differentiation between human populations and catalog the extent to which this differentiation is enriched in specific higher-order sequence contexts such as pentamers and heptamers.
73. Haller BC, Messer PW: SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Mol Biol Evol* 2019, 36:632–637. [PubMed: 30517680]

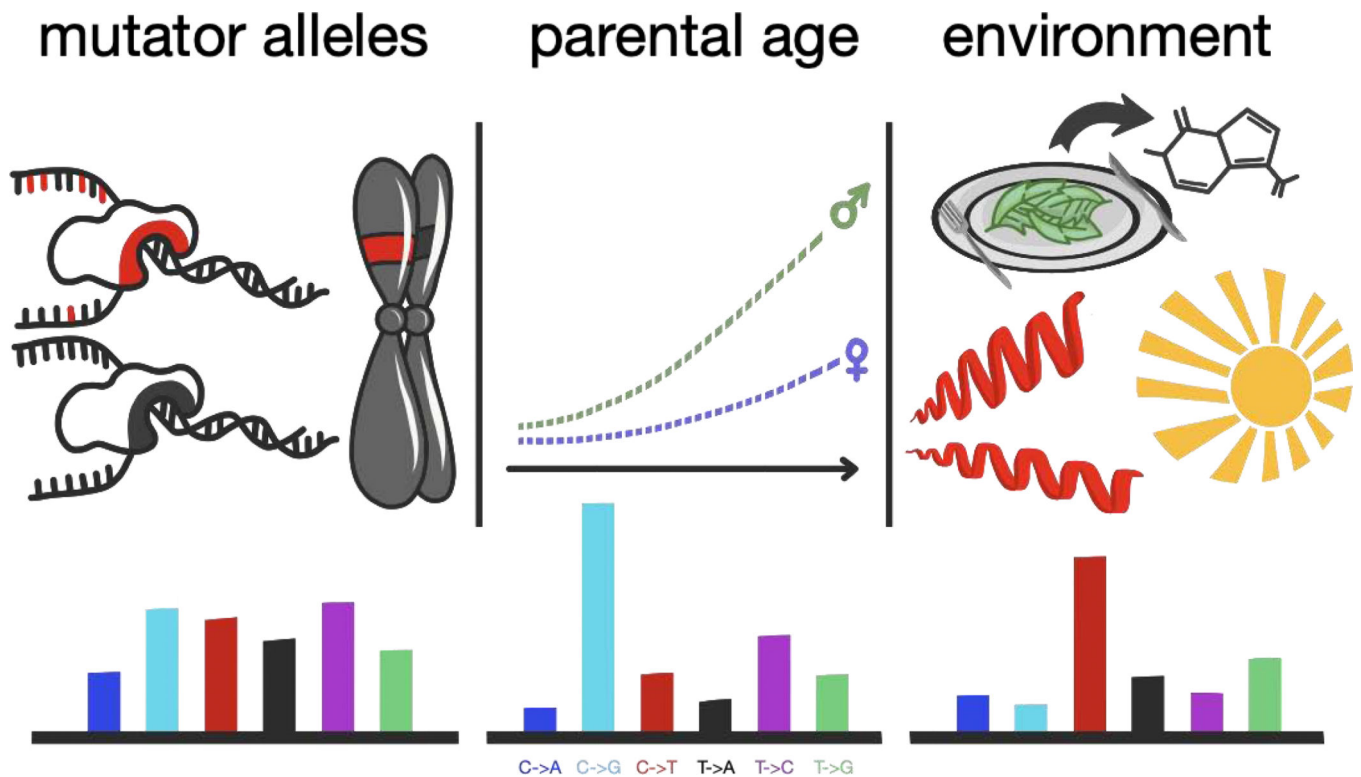


Figure 1:

Three hypotheses have been proposed to explain why mutation spectra appear to evolve rapidly within populations: mutator alleles, life history trait evolution, and environmental mutagenesis. Firstly, mutator alleles are genetic variants in DNA polymerases, repair enzymes, or other genes that can raise or lower the mutation rate by altering DNA replication or repair fidelity in a heritable way. Secondly, changes in lifespan and reproductive age have the potential to change mutation spectra because of age-related changes in the rate and spectrum of mutagenesis in the male and female germlines. Finally, exposure to radiation, dietary mutagens, and toxic air particulates are all possible hazards of past and present existence on earth. These have the potential to modify mutation rate in a way that is dependent on time rather than genetic background. (Image copyright Natalie Telis)