Received: 18 January 2021 | Revised: 22 April 2021 | Accepted: 26 May 2021

DOI: 10.1111/mec.16007

SPECIAL ISSUE

MOLECULAR ECOLOGY WILEY

The challenge and promise of estimating the de novo mutation rate from whole-genome comparisons among closely related individuals

Anne D. Yoder | George P. Tiley

Department of Biology, Duke University, Durham, NC, USA

Correspondence
Anne D. Yoder, Department of Biology, Duke University, Durham, NC, USA.
Email: anne.yoder@duke.edu

Abstract

Germline mutations are the raw material for natural selection, driving species evolution and the generation of earth's biodiversity. Without this driver of genetic diversity, life on earth would stagnate. Yet, it is a double-edged sword. An excess of mutations can have devastating effects on fitness and population viability. It is therefore one of the great challenges of molecular ecology to determine the rate and mechanisms by which these mutations accrue across the tree of life. Advances in high-throughput sequencing technologies are providing new opportunities for characterizing the rates and mutational spectra within species and populations thus informing essential evolutionary parameters such as the timing of speciation events, the intricacies of historical demography, and the degree to which lineages are subject to the burdens of mutational load. Here, we will focus on both the challenge and promise of whole-genome comparisons among parents and their offspring from known pedigrees for the detection of germline mutations as they arise in a single generation. The potential of these studies is high, but the field is still in its infancy and much uncertainty remains. Namely, the technical challenges are daunting given that pedigree-based genome comparisons are essentially searching for needles in a haystack given the very low signal to noise ratio. Despite the challenges, we predict that rapidly developing methods for whole-genome comparisons hold great promise for integrating empirically derived estimates of de novo mutation rates and mutation spectra across many molecular ecological applications.

KEYWORDS
comparative genomics, life history evolution, molecular evolution, mutation rate, pedigree-based studies

1 | INTRODUCTION

> "The life of a new mutation is not an easy one."
> Phillips, 1997

Spontaneous de novo mutations (DNMs) in the germline fuel the engine of evolution. Without them, natural selection has no material on which to act. The distribution and frequency of DNMs across the genome contribute to virtually every aspect of an organism's function and fitness. Consequently, the mechanisms by which they are generated and transmitted from one generation to the next is of fundamental interest to the field of molecular ecology. Accurate mutation rate estimates within species and populations can inform essential evolutionary parameters such as the timing of speciation events and key

aspects of historical demography such as population growth, decline, and the timing of population bottlenecks. Indeed, it is difficult to conceive of a biological phenomenon more important to our understanding of speciation, population genetic theory, molecular adaptation, life history strategy – and ultimately, conservation biology – than the accurate measurement of the de novo mutation rate per generation (hereafter referred to as μ). Even so, the vast majority of mutations do not confer a selective advantage and can thus lead to genomes that are burdened with mutational loads that hamper or prevent opportunities for adaptation through negative selection at linked sites (Charlesworth et al., 1993). Indeed, as is becoming increasingly evident via μ studies in humans, both the magnitude of new mutations, as well as their distribution across the genome, can have profound consequences for individual health and evolutionary fitness (Coe et al., 2019; Gao et al., 2020; Goldmann et al., 2019; Mitra et al., 2021; Rahbari et al., 2016). Thus, the imperative for accurate determination of μ extends from questions of basic evolutionary genetic research to applications for human health.

In this review, we will focus on new insights into the causes and consequences of μ evolution that are being ushered in by the ever-expanding innovations in whole-genome sequencing (WGS). With the advent of massively-parallel sequencing technologies (Rogers & Venter, 2005), and more recently, with accelerating improvements to single-molecule long-read technologies (Miga et al., 2020; Nurk et al., 2020), we are steadily progressing from an era wherein whole-genome analysis was restricted to genetic model organisms to one wherein WGS applications can be applied to virtually any organism for which genomic DNA can be obtained. Here, we focus on whole-genome comparisons for the detection of germline mutations in a single generation among closely related individuals in what are typically described as "trio" analyses (Scally & Durbin, 2012). By sequencing and comparing genomes from parents and their offspring, as well as other relatives (i.e., extended pedigrees), investigators can ideally count and characterize the mutations that occur within the transmission of one generation to another. To date, these studies have nearly exclusively examined the appearance of single base-pair substitutions, typically referred to as de novo mutations (DNMs) and this review is focused accordingly. Given that structural variants are also a component of germline mutations (Besenbacher et al., 2015; Course et al., 2020; Harris & Pritchard, 2017; Tatsumoto et al., 2017), we also wish to emphasize that, although not covered here, there is a pressing need going forward to enlarge the focus of pedigree-based studies to capture insertion and deletion events, as well as other structural mutations.

Although estimating μ would ideally be accomplished by sequencing the germline itself, in practice this is rarely feasible outside of a model organism such as laboratory mice. Instead, investigators must take advantage of available tissues such as blood, skin, or in postmortem cases, organ tissue. For such studies, this has the unintended consequence that DNM studies based on these tissues may be capturing somatic mutations in addition to the heritable germline variation (Li, 2014). Although our understanding of somatic and germline mutation rates is still developing (Muyas et al., 2020), pedigree-based studies have already begun to characterize many features of the mutational spectrum such as the type of base pair changes, their frequency, biases in genomic regions, and patterns specific to parental origin (reviewed in Ségurel et al., 2014). Many of the insights derived from pedigree-based studies build upon earlier mutation-accumulation analysis in model organisms (Denver et al., 2000; Estes et al., 2004; Halligan & Keightley, 2009; Vassilieva & Lynch, 1999; Zhu et al., 2014) – studies that permit characterization of μ for organisms where pedigrees are not biologically feasible or sensible, such as asexual organisms (Krasovec et al., 2019, 2020; Long et al., 2016). Regardless of the precise study design for estimating μ, we predict that with the increasing power of WGS, this parameter will find increasing uses and conferred power for the field of molecular ecology.

2 | THE TECHNICAL CHALLENGE OF ESTIMATING THE DE NOVO MUTATION RATE

2.1 | Finding needles in a haystack

An understanding of the per generation de novo mutation rate (μ), both within and among phylogenetic lineages, can inform practical applications, such as divergence time estimation (Bergeron et al., 2020; Keightley et al., 2015; Martin & Hohna, 2018; Moorjani et al., 2016) and the underpinnings of genetic load in threatened species (Bataillon, 2000; Davenport et al., 2021; Glémin et al., 2003). It can also illuminate the genetic mechanisms that dictate the characteristics and distribution of genomic variation (Jónsson et al., 2017; Rodriguez-Galindo et al., 2020). The potential for deeper insight into speciation, historical demography, environmental contributions to mutation, the effects of aging, and disease risk is profound, yet the cruel reality is that we are looking for needles in a haystack. The technical challenges are enormous and largely relate to the difficulty of identifying true mutations against a background of sequencing errors. Even with the most accurate sequencing platforms and assembly methods available, yielding upwards of 99.999% sequencing accuracy (QV50) (Nurk et al., 2020), this will result in about 28,000 errors in a typical 2.8 Gb primate genome. Thus, de novo mutation rates are orders of magnitude lower than the sequencing error rate, even for the most accurate sequencing methods. Sequencing costs for high-fidelity platforms, such as PacBio HiFi, are such that they are also well beyond the reach of the average budget for molecular ecological studies. Therefore, as the field develops and becomes more inclusive, it is critical that measurements of μ from pedigree-based studies follow agreed-upon community standards that will – at a minimum – allow for meaningful comparisons across independent studies.

Further, given that confidence in calling mutations is dependent, at least in part, upon a high-quality reference genome, which were initially only available for humans and other genetic model organisms such as house mouse, pedigree-based DNM studies are at this

stage of the field's development correspondingly biased. Thanks in part to the enormous number of pedigrees made available through the 1000 Genomes Project (Genomes Project et al., 2015), the preponderance of our understanding comes from humans, although with more diverse nonhuman primates and other animal lineages increasingly targeted in recent years (summarized in Table 1). Although plants have also been investigated for the estimation of μ, those studies are not addressed here due to the biological complexities of germline segregation in plants versus animals (Wang et al., 2019; Xie et al., 2016; Yang et al., 2015).

2.2 | Estimating the de novo mutation rate using pedigree comparisons

From a naive perspective, the measurement of germline mutations should be straightforward: the investigator sequences parents and offspring for the species of interest and then simply counts the number of spontaneous mutations that have appeared in the offspring during genetic transmission; that is, those variants that appear in the offspring that are not present in either parent. This number (the numerator) is then divided by the total genome size (the denominator) to derive the rate. As the section above has foreshadowed, however, the signal to noise ratio of true DNMs versus sequencing error is daunting. Thus, there are detailed considerations relating to pedigree study design, choice of sequencing platform, depth of sequencing coverage, computational management and analysis of sequenced genomes, parental biases, as well as accumulating knowledge regarding the probability of certain types of mutations versus others (Figure 1). As is becoming the community standard, each genome in the pedigree is sequenced to approximately 30X coverage with few benefits found for sequencing at higher depths. These recommendations are largely based on early observations from short-read sequencing data from humans generated for detecting heterozygous variants with reasonable accuracy and has been a typical benchmark for variant discovery (e.g. Bentley

TABLE 1 Pedigree-based mutation rate estimates for animals

Species	Common name	Citation	μ[a]
Homo sapiens	Human	Roach et al. (2010)	1.1
		Kong et al. (2012)	1.2
		Besenbacher et al. (2015)	1.29
		Rahbari et al. (2016)	1.28
		Jónsson et al. (2017)	1.29
		Lindsay et al. (2019)	1.22
Pan troglodytes	Chimpanzee	Venn et al. (2014)	1.2
		Tatsumoto et al. (2017)	1.48
		Besenbacher et al. (2019)	1.26
Gorilla gorilla	Gorilla	Besenbacher et al. (2019)	1.13
Pongo abelii	Orangutan	Besenbacher et al. (2019)	1.66
Chlorocebus sabaeus	Green monkey	Pfieffer (2017)	0.94
Papio anubis	Baboon	Wu et al. (2020)	0.57
Macaca mulatta	Rhesus macaque	Wang et al. (2020)	0.58
		Bergeron et al. (2020)	0.77
Aotus nancymaae	Owl monkey	Thomas et al. (2018)	0.81
Microcebus murinus	Mouse lemur	Campbell et al. (2021)	1.52
Mus musculus	House mouse	Lindsay et al. (2019)	0.39
Canis lupus	Wolf	Koch et al. (2019)	0.45
Felic catus	Domestic cat	Wang et al. (2021)	0.86
Bos taurus	Cow	Harland et al. (2017)	1.17
Ornithorhynchus anatinus	Platypus	Martin et al. (2018)	0.7
Ficedula albicollis	Collared flycatcher	Smeds et al. (2016)	0.46
Clupea harengus	Atlantic herring	Feng et al. (2017)	0.2
Drosophila melanogaster	Fruit fly	Keightly et al. (2014)	0.28
Heliconius melopene	Butterfly	Keightly et al. (2015)	0.29
Apis mellefiera	Honeybee	Yang et al. (2015)	0.34
Bombus terrestris	Bumblebee	Liu et al. (2017)	0.36

[a]Rates are $\times 10^{-8}$.

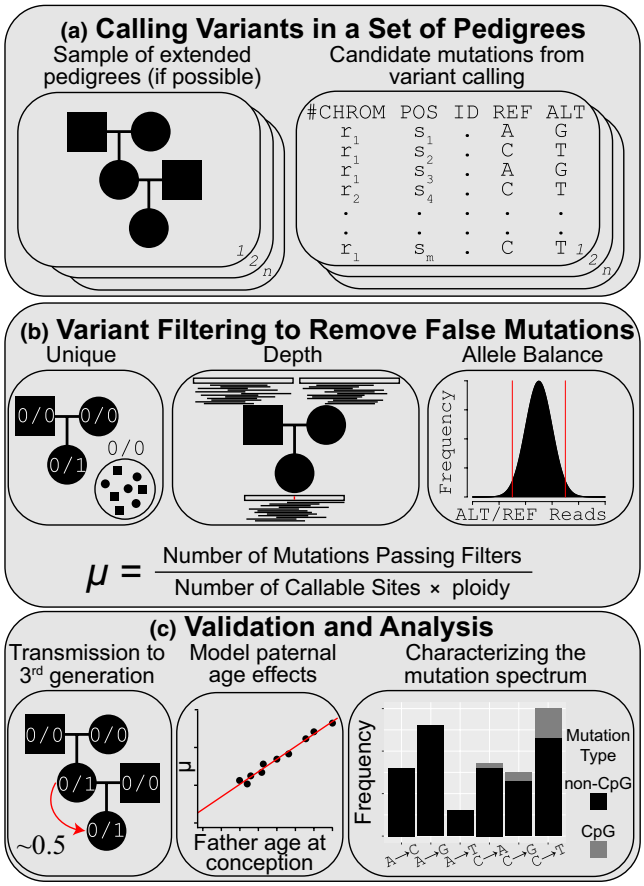

FIGURE 1 Estimating per-generation mutation rates from pedigrees. (a) Genomes are sequenced to at least 30x for *n* pedigrees. At a minimum, both parents and an offspring need to be sequenced, but pedigrees that include a third generation are the emerging recommendation. Variants are called jointly across pedigrees. Because genotyping is often sensitive as to not miss variants a number of filtering steps are required to identify putative mutations. (b) DNMs should be heterozygous in the offspring, but homozygous for the reference allele in both parents as well as all other individuals available across pedigrees or other population-level data. A mutation at a site needs to be callable at both parents and the offspring, which is typically determined by sufficient depth of unambiguously aligned reads. Variants are also filtered for allele balance, where between 30% and 70% of the reads should have the alternate allele. DNMs passing these filters are then divided by the number of callable sites multiplied by the ploidy level to estimate μ. (c) A number of analyses can be used to validate a mutation rate estimate. Where third generation pedigrees are available, approximately 50% of putative mutations should be observed in the third generation. An over-representation of mutations in the second generation would suggest a high number of false positives. Paternal age effects on mutation rate should also be observable. It is now well established that fathers contribute more mutations over time, and linear modelling can be used to estimate the mutation rate before puberty. It is also well-accepted that certain types of mutations are more common than others, such C-to-T transitions, and mutation spectra can be useful for evaluating pipelines used for estimating mutation rates [Colour figure can be viewed at wileyonlinelibrary.com]

et al., 2008; Gudbjartsson et al., 2015). Many nonhuman studies have also adopted this guidance, subsequently making 30× the standard read depth with the additional finding that increased depth of coverage has no beneficial effects for accounting for either false negative (Koch et al., 2019) or false positive mutations (Wu et al., 2020). Conversely, however, when sequencing depths are too low, there can be spurious errors that mischaracterize mutation spectra (Harris, 2015).

It has also become standard to include a third generation to track patterns of inheritance across two germline transmissions (Figure 1a). And with larger studies, joint calling across multiple pedigrees can identify a large number of variants, although these must be filtered to accurately identify the small fraction of variants that are putative DNMs (Figure 1b). According to biological expectations, DNMs should be heterozygous in the offspring, but homozygous in both parents, as well as in other individuals available for sequencing across pedigrees or other population-level data. Because sequencing coverage can be uneven, variants below a specific read depth are excluded to ensure that putative DNMs are not due to one or a few reads with sequencing or mapping errors, while at higher levels of coverage, variants with greater than two times the average sequencing depth, or some distribution-based upper threshold (Li, 2014), are removed given that these sites are likely to be mapping errors from repetitive or duplicated regions (Besenbacher et al., 2019; Keightley et al., 2014; Li, 2014; Thomas et al., 2018; Wang et al., 2020). Variants are also filtered for allele balance, where between 30% and 70% of the reads should have the alternate allele given that half of the reads are expected to carry the mutation in a diploid. Although more conservative allele balance intervals have been used (Thomas et al., 2018), studies that evaluate sensitivity over a range of filter settings reveal how exceptionally sensitive the estimation of μ can be to filtering options (e.g., Campbell et al., 2021).

Determining the denominator is an equally critical estimate given that not all sites in the genome will have sufficient information for identifying a DNM had one occurred. In other words, not all sites in the genome are "callable sites." Accordingly, there have been a number of approaches implemented for estimating the number of callable sites, with the most straightforward being that each site is determined as callable or not given the proposed filtering criteria. For example, a site that is homozygous in the child must also be homozygous in both parents, and with all three individuals falling between the desired depth of coverage cutoffs. A probabilistic approach has also been proposed that weights sites based on all of the sites at a given depth across pedigrees (Besenbacher et al., 2015). By estimating the number of putative DNMs and the number of callable sites, there is enough information to calculate μ (Figure 1b). But it is possible to also account for false positives (sites that are mistakenly identified as DNMs) as well as false negatives (sites that may contain true DNMs, but are mistakenly filtered out).

2.3 | False positives

Even though erroneous variant calls due to library preparation, sequencing, mapping, and genotyping can largely be mitigated through filtering (Figure 1b), there may also be somatic mutations that can be

mistaken for DNMs. This is especially problematic given that somatic mutations can sometimes generate signal in enough reads to pass filtering thresholds. Simulations based on human genomes suggest that at least 100× coverage is needed to differentiate DNMs from somatic mutations when they are present in 10% or more reads for a given site (Acuna-Hidalgo et al., 2015). Furthermore, somatic mutations may be more prevalent than previously appreciated. For example, a recent human transcriptome study across 49 tissue types estimated a somatic mutation rate of 1.32×10^{-8}, which is uncomfortably similar to germline mutation rate estimates (Muryas et al., 2020).

It is here that mutation spectra and extended three-generation pedigrees can be used for diagnosing false positives. Germline mutational *spectra* are remarkably consistent, at least within species, and not affected by parent age (Rahbari et al., 2016). Mutation rates, on the other hand, have been shown to have strong parental age effects (e.g., Monaghan & Metcalfe, 2019) with the number of DNMs observed to increase with paternal age in particular (Kong et al., 2012; Lindsay et al., 2019; Thomas et al., 2018; Wang et al., 2020). Although a maternal age effect has also been reported, it is less pronounced (Gao et al., 2019). Somatic mutations, on the other hand, can be caused by a number of mutational mechanisms for which profiles differ markedly from those of DNMs (Alexandrov et al., 2015; 2020). Thus, in addition to being a vital component of DNM analysis, the predictability of age effects and spectra (Figure 1c) can help evaluate pipelines used to call putative DNMs. Additional approaches for estimating false positive rates include estimating error rates from technical replicates (Campbell et al., 2021; Kessler et al., 2020) or sequencing monozygotic twins (Kessler et al., 2020). Even so, technical replication may not always be practical or advantageous compared to sampling three-generation pedigrees, and we are discovering that there can be substantial mutational differences between monozygotic twins due to prezygotic mutations (Jónsson et al., 2021).

The emerging community standard for evaluating the presence of false positives, therefore, is the "extended pedigree" approach that uses transmission rates from the second to third generation to measure the performance of DNM calling pipelines and filtering criteria (e.g. Besenbacher et al., 2019; Thomas et al., 2018; Wang et al., 2020; Wu et al., 2020) (Figure 1c). Validation of DNMs can be further confirmed with targeted Sanger sequencing when there are few candidate mutations or few pedigrees (Keightley et al., 2014, 2015; Koch et al., 2019). Although validating all detected DNMs with Sanger sequencing may not be operationally feasible, three-generation pedigree study design, combined with Sanger validation of a subset of DNMs (e.g. Bergeron et al., 2020; Jónsson et al., 2017; Wu et al., 2020), is presently the ultimate means for determining the biological validity of called DNMs.

2.4 | False negatives

Prior to validating a DNM, one has to observe it in the first place. Because the prevailing emphasis in pedigree-based DNM studies has been on filtering out false positives, it is possible that μ estimates may be overly conservative; that is, some fraction of true DNMs might be removed by filtering pipelines (Besenbacher et al., 2015). For example, the allele balance filter is critical for removing false positives, but can also remove true mutations due to sampling error among reads (Thomas et al., 2018). Thus, the accurate measurement of μ must deal with some probability of false negatives in addition to false positives (Scally, 2016; Ségurel et al., 2014). DNMs can be missed for largely technical reasons than can range from poor mapping to the use of an inappropriate reference genome (Garimella et al., 2020; Martin et al., 2018), applying filters with too much vigour (Ségurel et al., 2014), or simply poor sequencing quality at specific regions of the genome (Keightley et al., 2015; Pfeifer, 2017; Thomas, 2019).

The false negative rate can be determined empirically from the data by counting the number of otherwise good variants that are lost due to increasing stringency of the allele balance filter (Bergeron et al., 2020; Thomas et al., 2018). Alternatively, false negative rates can be approximated with simulation methods. One such approach has been to generate synthetic mutations in the sequencing reads and then run the genotyping and filtering pipeline to determine the proportion of synthetic mutations that are recovered (Campbell et al., 2021; Keightley et al., 2015; Koch et al., 2019; Pfeifer, 2017; Wu et al., 2020). This simulation procedure can be tedious, however, as it requires some reasonable understanding of where mutations are likely to occur and in what proportion of reads they will be recovered. Moreover, the need to re-run the analysis pipeline for simulation replicates requires both effort and time. As an additional concern, comparison of a simulation-based false negative rates and simpler filter-based estimates suggest that the simulation approach may provide an underestimate (Bergeron et al., 2020). It is debatable if μ calculations should attempt to correct for false negatives, since callability calculations already mitigate some of these effects, especially when using the probability that a site is callable rather than counting each site categorically as either callable or not callable (Besenbacher et al., 2019).

3 | THE PROMISE OF THE DE NOVO MUTATION RATE FOR MOLECULAR ECOLOGY

3.1 | Reconciling mutation and substitution rates across organismal scales

Biologists occasionally use the terms "mutation rate" and "substitution rate" interchangeably. They are, however, distinct phenomena both in the mode of change represented and in the specified unit of time. Mutations are the spontaneous result of errors in the DNA replication machinery (Ohno, 2019; Seoighe & Scally, 2017), appearing abruptly and with an unknown evolutionary fate. They can arise in any tissue and at any life stage, but it is only the DNMs that occur in an organism's germline that are inherited by the next generation.

Because DNMs are observed between parents and their offspring, mutation rates are typically measured as changes in the genome per-generation. Substitutions, on the other hand, are the genetic changes observed in a population after the forces of selection and drift have purged or fixed de novo mutations within a population (Haldane, 1927) and are generally estimated as per-site rates with time-reversible models of sequence evolution (Yang, 1995).

To compare the rates directly it is necessary to scale them appropriately. Phylogenetic branch lengths are usually measured in substitutions per-site, but with appropriate fossil or other methods of calibration (Duchene et al., 2014; Heath et al., 2014; Marshall, 2019), they can be scaled to absolute time that measure substitutions as per-site per-year (Zuckerkandl & Pauling, 1965). Similarly, μ can be converted to a per-year measure if the generation times are known for the targeted species (Angelis & dos Reis, 2015), such that generation time is the average parent age at the time of conception. Given that one of the fundamental assertions of the neutral theory is that substitution rates based on putatively neutral sites such as third codon positions should be equivalent to the neutral rate of molecular evolution (Kimura, 1968), and approximately μ, comparison of pedigree-based estimates of μ with substitution rates from phylogenies should theoretically agree. Surprisingly, however, disagreements between the two have been observed, as initially highlighted in humans (Scally & Durbin, 2012). Indeed, a recent survey of pedigree-based μ in primates compared with substitution rates from a phylogenetic relaxed clock analysis from the same species (dos Reis et al., 2018) has shown that although credible intervals overlapped between the two in many cases, there were also notable exceptions (Campbell et al., 2021). We suggest that the investigation of these discrepancies is critical to the field. By recognizing that μ and phylogenetic substitution rate estimates potentially disagree, we can begin to define the organismal level – from individual, to population, to species – wherein rates can be measured, and accordingly, for which level of evolutionary analysis they are best suited (Figure 2).

By explicitly considering the underlying processes, one that occurs nearly instantaneously in an individual, and the other that occurs over evolutionary time scales in populations and species, we can anticipate the observation that μ can be substantially higher than phylogenetically derived substitution rates (Denver et al., 2000) – which in certain cases has been observed to be as much as an order of magnitude higher (Ho et al., 2011; Howell et al., 2003). General expectations are that most DNMs will not be fixed in a population due to drift or will be purged by purifying selection, and thus the expected number of substitutions observed over a phylogeny should be lower than the number of DNMs when scaling both to the same per-unit time. It is thus noteworthy that μ is not always observed to be higher than the estimated substitution rate, such as in humans (Moorjani et al., 2016; Scally & Durbin, 2012; Ségurel et al., 2014). One explanation for the lower than expected μ in humans is that our de novo rate appears to have slowed down since humans diverged from chimps (Besenbacher et al., 2019; Moorjani, Gao, et al., 2016). The differential could also be attributed to an increase in generation time along the human branch (Amster & Sella, 2016; Langergraber et al., 2012) although recent comparisons

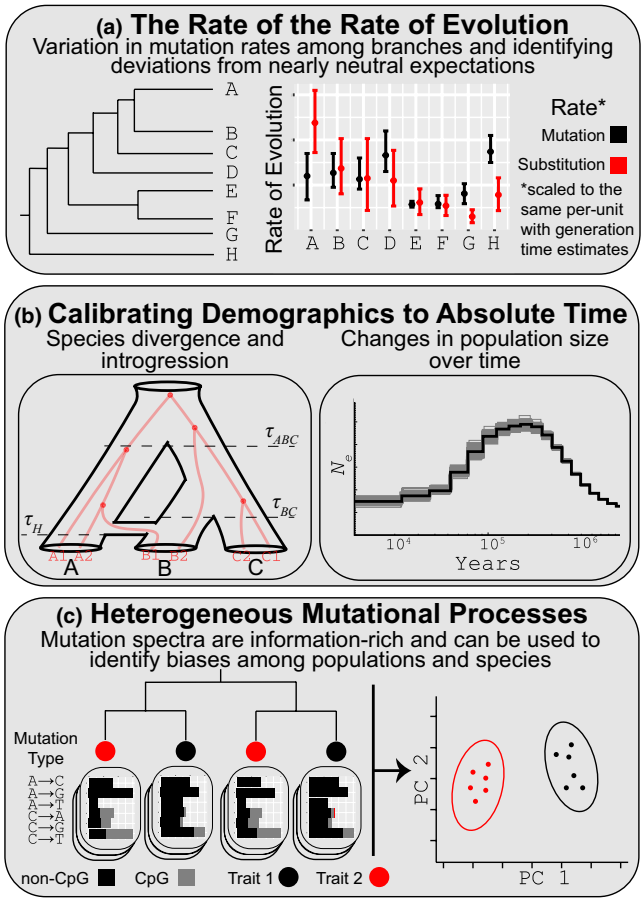

FIGURE 2 Applications of mutation rate estimates. (a) When rate estimates are available for multiple species, the evolution of the mutation rate itself can be studied. Comparisons between mutation rates and substitution rates can reveal life-history changes such as increased generation times, when mutation rates are much slower than expected from substitution rates. (b) Per-generation mutation rate estimates have been a powerful tool for understanding the demographic history of populations and species. The timing of species divergences and introgression events can be calibrated to absolute time from coalescent models that jointly account for incomplete lineage sorting and cross-species gene flow. Popular methods for studying change in population size over time such as PSMC require a mutation rate, which is often assumed to be 1×10^{-8} for many animal studies. (c) The mutation spectrum can also be utilized when studying structured populations to identify mutational biases between them. There are nine mutational categories (when considering mutations at CpG and non-CpG) that can be revealing about changes in methylation over time [Colour figure can be viewed at wileyonlinelibrary.com]

of baboon and human rates suggest that the slowdown may have occurred much earlier in the anthropoid lineage than has been previously appreciated (Wu et al., 2020). Although the human example currently appears exceptional for primates, this apparent deviation may become less remarkable as we continue to accumulate estimates of μ across the tree of life. An increased focus on large-scale comparisons between μ and phylogenetic substitution rates will help identify cases where de novo mutation rates have changed along a branch and thus appear to deviate from neutral expectations (Figure 2a).

Ultimately, we predict that pedigree-based estimates of μ will free investigators from the constraints of fossil calibrations for divergence time estimation (Tiley et al., 2020), with the caveat that generation times are known (Langergraber et al., 2012). In such a cases, we can simply divide μ by generation time (defined as the average parent age at the time of conception for the pedigree) to get a per-year rate that is not dependent on an external calibration (Amster & Sella, 2016). Although simplistic, this approach, may be more appropriate when estimating divergence times of recent speciation events where incomplete lineage sorting is a concern, compared to concatenated fossil-calibrated methods (Amster & Sella, 2016; Angelis & dos Reis, 2015; Martin & Hohna, 2018; Poelstra et al., 2021). In addition to estimating species divergence times, μ estimates can be used to calibrate various demographic analyses (Figure 2b). For example, μ and a generation time is needed to calibrate analyses of population size change over time with the popular PSMC program (Li & Durbin, 2011) or related methods. The age of introgression events can similarly be dated when using models that jointly account for incomplete lineage sorting and episodic gene flow (Flouri et al., 2020). The use of μ for calibrating divergence times and demographic change has been especially important in humans. For example, a pedigree-based μ applied to divergence of Yoruba African and non-African humans yielded older splits that are much more consistent with archaeological evidence from the Arabian Peninsula compared to more recent dates based on phylogenetic estimates (Scally & Durbin, 2012).

3.2 | Connecting the dots between male bias, generation time, and rate variation

The section above is meant to emphasize the distinction between μ and substitution rates, although in this section we cover how life-history traits can correlate with both. Because phylogenetic rates measured by substitutions-per-site or calibrated to substitutions-per-site-per-year have been studied for decades, we can build off of these earlier observations and hypotheses. In a seminal paper that examined differential rates of mtDNA evolution across both endotherms and ectotherms, Martin and Palumbi (1993) noted that substitutions-per-site-per-year correlated with body size, with large mammals (such as whales) having slow rates, medium-sized mammals (such as primates) having intermediate rates, and small mammals (such as rodents) showing the fastest rates. They also noted a potentially confounding pattern wherein ectotherms, with their relatively slow metabolic rates, tend to have slower rates than similarly sized endotherms with their higher metabolic rates. They acknowledged, however, that body size, generation time, and metabolic rate are not entirely independent traits given that large-bodied organisms tend to have longer generation times, and often, slower metabolisms.

These life history traits have remained a subject of interest for explaining substitution rate variation (reviewed in Bromham, 2011) and although some hypotheses such as the metabolic rate hypothesis have been refuted as a general explanation (Lanfear et al., 2007; Qiu et al., 2014), generation time has remained a topic of interest both for animal studies – especially primates where it has long been observed that great apes have lower substitution rates than Old-World and New-World monkeys (Goodman, 1985; Moorjani, Amorim, et al., 2016) – and plant studies where slower-growing woody species have slower substitution rates than faster-growing ones (Lanfear et al., 2013; Smith & Donoghue, 2008). Given that the generation-time hypothesis was originally formulated with respect to per-year substitution rates, the negative correlation between generation time and rates of evolution could disappear when measuring rates of evolution with μ. There are several life-history traits that covary with generation time, however, that also correlate with variation in substitution rates and in μ, both within and among species. For example, comparison of substitution rates in sex chromosomes and autosomes have shown that males have higher substitution rates than females (Axelsson et al., 2004; Bartosch-Harlid et al., 2003), and that there can be a generation-time effect to this male bias (Amster & Sella, 2016; Ellegren, 2007; Goetting-Minesky & Makova, 2006). One explanation for the observed male bias is that there will be more cell divisions during spermatogenesis than in oogenesis (Ellegren & Fridolfsson, 1997), and in turn, this will be more pronounced in longer-lived species with longer generation times (Amster et al., 2019; Goetting-Minesky & Makova, 2006). Although most interest has focused on variation in μ, it is also possible to investigate trait associations with mutational spectra at the population level (Figure 2c). For example, within human populations where many spectra are available for multiple populations, an excess of specific mutation types are observable in ancestral Europeans following a known bottleneck (Harris & Pritchard, 2017).

3.3 | Cell division bias and the paternal age effect

The assumption that the number of cell divisions correlates with rates of evolution is perhaps most associated with the phenomenon of male mutation bias. Well before the advent of NGS technologies, investigators noted the propensity for mutation accumulation to be higher in males than in females (Crow, 2000; Ellegren & Fridolfsson, 1997; Hurst & Ellegren, 1998; Shimmin et al., 1993). The evidence for this has come both from comparisons of differential mutation accumulation on sex chromosomes (Axelsson et al., 2004; Bartosch-Harlid et al., 2003; Ellegren & Fridolfsson, 1997; Haldane, 1947; Shimmin et al., 1993) as well as from direct measurements of the paternal contribution of novel mutations in descendent offspring (see Crow, 2000 for a detailed review relevant to human disease phenotypes). It was also noted early on that rates accelerate as males age, with older fathers contributing more mutations than younger fathers, a phenomenon with the potential to contribute to disease risks in humans (Crow, 1997; Kong et al., 2012).

The explanation for both phenomena, male bias and the paternal age effect, appeared to be immediately self-evident: because of the many more cell divisions in spermatogenesis than in oogenesis – and thus more opportunity for errors in the replication process –the male germline will accumulate more DNA replication errors

throughout life (Ellegren, 2007; Goetting-Minesky & Makova, 2006). Despite the fact that spermatogonial stem cells are characterized by highly efficient DNA repair and one of the lowest spontaneous mutation rates in the human body (Aitken et al., 2020), this idea has dominated the mutation rate literature for years. Pedigree-based studies in humans and nonhuman primates have confirmed the paternal origin of a majority of DNMs, consistent with the hypothesis that the number of mutations increase with the father's age (Bergeron et al., 2020; Jónsson et al., 2017; Lindsay et al., 2019; Rahbari et al., 2016; Venn et al., 2014; Wang et al., 2020). The paternal mutation bias may be a more general feature of mammals, as it is also prevalent in domestic cats (Wang et al., 2021) and mice (Lindsay et al., 2019). Modelling of μ given paternal age has further clarified that it is specifically during the years following the onset of puberty that contribute new mutations with age (Gao et al., 2019; Thomas et al., 2018; Wu et al., 2020). This may explain why paternal biases appear to be less prevalent in quickly reproducing organisms like mice (Lindsay et al., 2019), and that some mutation types such as CpG to TpG mutations do not increase with the father's age given that they are not associated with DNA repair during replication (Gao et al., 2019; Jónsson et al., 2017; Thomas et al., 2018; Wu et al., 2020). Although mutation rates can also increase with maternal age, this may not be observable except in organisms with long generation times. In mammals, oocytes are differentiated early in development before birth and do not undergo additional rounds of replication. Nonetheless, these oocytes may sustain damage over time requiring double-strand break repair that can generate C-to-G mutations (Gao et al., 2019; Goldmann et al., 2016; Jónsson et al., 2017; Wong et al., 2016). It should be noted, however, that these observations and underlying hypotheses largely derive from studies limited to mammals and thus require further testing across the tree of life for validation.

As one potential counterexample, a study of mouse lemurs – a strepsirrhine primate that diverged from humans before the Cretaceous-Paleogene boundary (dos Reis et al., 2018) – found a very weak male bias (Campbell et al., 2021). Unfortunately, this study is limited to a single pedigree and lacks a third generation for verifying DNM heritability. Also, given that it is the first and only pedigree-based analysis of a strepsirrhine primate, it cannot yet be known if the results are representative of the strepsirrhine clade, or are more likely to be an artefact of study design. A reduced male mutational bias is plausible in this case; however, given that mouse lemurs reproduce early in life, coincident with puberty. These results are similar to those from a pedigree-based analysis of the collard flycatcher (Smeds et al., 2016), even though earlier phylogenetic analyses in birds supported a male-bias (Axelsson et al., 2004; Ellegren & Fridolfsson, 1997). Uncertainties have also been reported for monotremes, with comparative genomic (although not pedigree-based) studies failing to reveal a strong male mutation bias (Cortez et al., 2014). Subsequent analysis by the same group of investigators has reported nuances that could result from the action of purifying selection on the Y-chromosome in monotremes and as well as other mammals (Link et al., 2017).

Taken together, new research is adding nuance to the simple hypothesis that cell division from spermatogenesis explains parental variation in μ. Hurst and Ellegren (1998) asserted that the germ-line cell division model "is unlikely to be the whole truth" with regard to male mutation bias (p. 451). They argued that an array of other mutagenic mechanisms might apply, including patterns of methylation, exposure to oxygen free radicals, temperature effects, and possible metabolites that might act differentially on sperm. New analyses appear to be bearing them out, with DNA damage due to the hardships of aging taking on strength as a contributor to the age-related effects on μ (Ohno, 2019). Cellular aging is being examined from variety of empirical perspectives including cellular function (Monaghan & Metcalfe, 2019), mutagenesis (Aitken et al., 2020), maternal aging (Gao et al., 2019), and the timing of puberty (Ségurel et al., 2014; Thomas et al., 2018; Wang et al., 2020, 2021; Wu et al., 2020). The full resolution of these conflicts and complexities will benefit enormously from expanded pedigree-based studies across a more comprehensive phylogenetic sample with attention given not only to the parental source of DNMs, but also to their genomic and developmental context (Goldmann et al., 2016; Jónsson et al., 2017; Link et al., 2017; Narang & Wilson Sayres, 2016; Wong et al., 2016).

3.4 | Effective population size, selection, and mutator loci

It has been noted largely from mutation accumulation studies of single-celled organisms as well as with model organisms that μ varies widely across phylogenetic scales and is negatively correlated with effective population size (N_e): when N_e is small, μ tends to be high (Sung et al., 2012); when N_e is large, μ tends to be low (Long et al., 2016). Further, there appears to be a positive correlation between genome size and μ, such that organisms with large genomes have higher rates than those with smaller genomes (Smeds et al., 2016). This relationship was first noted in microbes wherein rates per base pair were observed to vary by approximately 16,000-fold whereas rates per genome varied only by 2.5-fold (Drake, 1991). Given the "largely mysterious" patterns observed on a site-by-site basis, Drake (1991) supposed that any underlying rules were likely to be observed in comparisons of the mutation rate per genome per round of DNA replication, with the further expectation that overall rates must have evolved under general evolutionary forces.

Here lie the underpinnings of the drift-barrier hypothesis (Lynch, 2010). Although there are various complexities relating to genome size, and to the differential phylogenetic characteristics of eukaryotes and prokaryotes (Sung et al., 2012), the drift-barrier hypothesis states that N_e can explain variation in μ across species due to selection as it acts more efficiently on DNA replication fidelity in larger populations. Rather than being a balance between the usually deleterious effects of mutation and selection to reduce those costs, the lower limits of the genome-wide de novo mutation rate must ultimately be set by the barriers imposed by genetic drift, which is determined by N_e. In such a model, the mutation rate will scale

negatively with N_e up to the point where further reductions in rate cannot overcome the selective disadvantage of even the weakest "mutator allele."

At a first glance, pedigree-based measurements of μ in animals appear to largely bear out these predictions (Figure 3a; Table S1), and can also be observed in traits that are autocorrelated with N_e such as body mass (Figure 3b). However, these relationships can explained by phylogenetic structure alone (Felsenstein, 1985; Whitney & Garland, 2010). When analysing phylogenetic independent contrasts instead of raw values (Supporting Information), the relationship between μ and N_e (Figure 3c) as well as body mass (Figure 3d) disappear. Some explanations have been offered as to why a simple negative correlation between μ and N_e would not be anticipated. The relationship between μ and N_e can change when mutations are strongly advantageous, and when slightly deleterious mutations are linked to selected sites. Further, these effects will have differential impacts across the expanse of a given genome (Lanfear et al., 2014; Martincorena & Luscombe, 2013). This latter nuance in particular has been noted to have relevance to endangered species and subsequent conservation strategies given that both N_e and μ can vary across the genome (Zeng et al., 2019).

Finally, some causes of variation in μ may be better observed at the population level as opposed to macrospecies phylogenetic comparisons. As sequencing technologies are becoming ever more precise, the relationship among lineages, their individual histories, and the idiosyncratic interactions of genotype and environment are coming into focus relative to the fixation of various hypermutator (Harris, 2015; Harris & Pritchard, 2017) and antimutator alleles (Maddamsetti & Grant, 2020; Sasani et al., 2021). Thus, an exciting frontier is within reach wherein investigators will have the analytical tools to explore the capacity for adaptive evolution as it relates to the interaction of mutation rate, genetic diversity, life history strategies, and environmental conditions (Rousselle et al., 2020).

4 | THE PATH FORWARD

Among the most pressing needs going forward for pedigree-based studies is for comparable methods wherein all studies confirm to a shared set of practices. These include, although are not limited to, agreed-upon standards for contiguity and phylogenetic proximity of reference genomes, methodologies for determining false positive and false negative rates, and a standard set of variant filters. Comprehensive analyses of multiple species and pedigrees with different estimators of μ may help develop best practices, and emphasize the degree to which variation in μ is due to methodological choices rather than biological differences. Even though estimates of μ may still be flawed, they will at least be similarly flawed, and thus comparable as we seek to discover the connections between genetic, environmental, and trait variation and the evolution of μ. Thankfully, efforts are being coordinated to this end (Bergeron et al., in prep), thus highlighting the future promise of pedigree-based mutation rate studies.

With each new rate estimate, we gain knowledge, particularly as we expand our inquiry across broader phylogenetic scales.

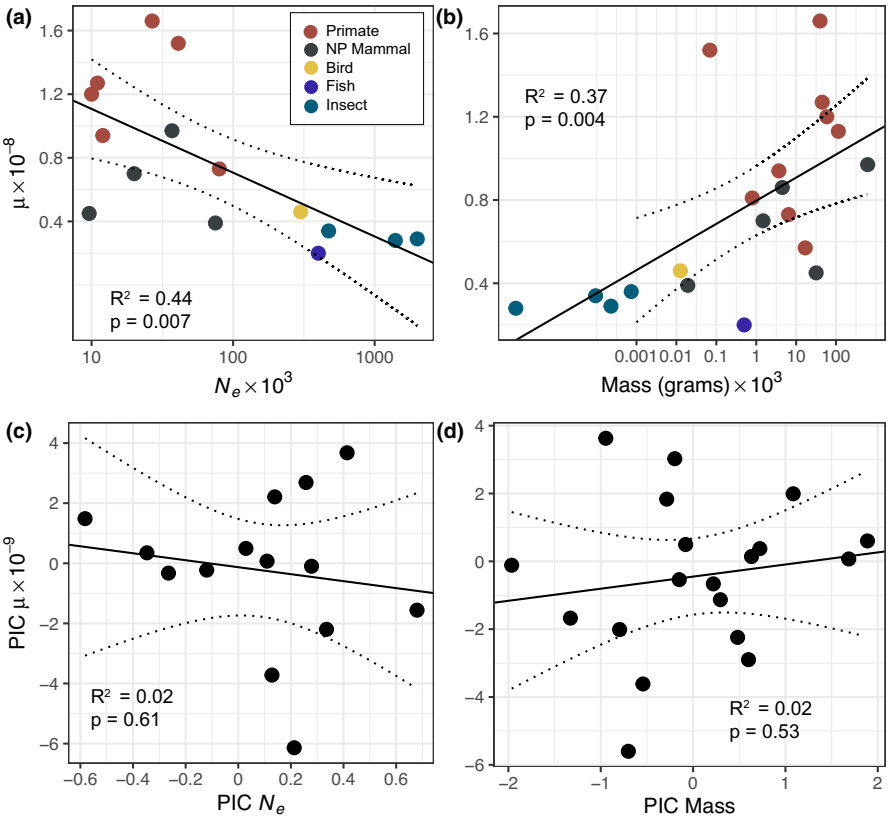

FIGURE 3 Relationship between life history and mutation rate in animals. (a) Effective population size estimates are plotted on a $\log_{10}$ scale. Mutation rates were estimated per-generation from pedigrees. A negative relationship seems evident when considering all animals as expected by the drift-barrier hypothesis, but this is not observed when looking within groups where multiple species are available for comparison such as primates. (b) Body mass estimates are plotted on a $\log_{10}$ scale. There are often autocorrelated traits such as body mass and effective population size that make teasing apart meaningful causal relationships between traits and μ tenuous. (c) Correcting for phylogenetic structure with phylogenetic independent contrasts (PICs) removes the significant negative relationship between μ and N_e as well as body mass (d)

Given the rate at which sequencing technologies are improving in accuracy and contiguity, pedigree-based studies should become increasingly more discerning. Also, as the associated costs continuing to drop, study designs of the breadth and depth previously available only for model organisms will become more common for non-model species. Here, it is worth noting the perhaps unappreciated value of zoos and other living-stock collections for enabling these leaps into new organismal systems. These collections offer precious opportunities for incorporating multigeneration pedigrees, as well as comprehensive databases of individual life history records including age, sex, number of offspring over lifespan, longitudinal health records, cause of death and other fundamental aspects of a given organism's biology (McCluskey et al., 2017). Thus, with a combination of technical, computational, and existing biological resources, we can be confident that classic hypotheses of molecular evolution will be increasingly refined with biological and functional sophistication.

ACKNOWLEDGEMENTS

We are grateful to our editor, Andrew Foote, and to three anonymous reviewers for their constructive comments and suggestions for improvement. This work was massively enhanced thanks to their collective recommendations and insights. We also thank Evan Eichler for sharing his expertise on state-of-the-art sequencing technologies. Finally, we express our gratitude to Lucie A. Bergeron, Søren Besenbacher, Mikkel H. Schierup, and Guojie Zhang for organizing the July 2020 virtual workshop wherein many of the technical challenges described here were discussed

ORCID

Anne D. Yoder https://orcid.org/0000-0002-1781-9552
George P. Tiley https://orcid.org/0000-0003-0053-0207

REFERENCES

Acuna-Hidalgo, R., Bo, T., Kwint, M. P., van de Vorst, M., Pinelli, M., Veltman, J. A., Hoischen, A., Vissers, L. E. L. M., & Gilissen, C. (2015). Post-zygotic point mutations are an underrecognized source of *de novo* genomic variation. *American Journal of Human Genetics*, *97*(1), 67–74. https://doi.org/10.1016/j.ajhg.2015.05.008

Aitken, R. J., De Iuliis, G. N., & Nixon, B. (2020). The sins of our forefathers: Paternal impacts on de novo mutation rate and development. *Annual Review of Genetics*, *54*, 1–24. https://doi.org/10.1146/annurev-genet-112618-043617

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nature Genetics*, *47*(12), 1402–1407. https://doi.org/10.1038/ng.3441

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Getz, G., … PCAWG Consortium (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94–101. https://doi.org/10.1038/s41586-020-1943-3

Amster, G., Murphy, D. A., Milligan, W. M., & Sella, G. (2019). Changes in life history and population size can explain relative neutral diversity levels on X and autosomes in extant human populations. bioRxiv. https://doi.org/10.1101/763524

Amster, G., & Sella, G. (2016). Life history effects on the molecular clock of autosomes and sex chromosomes. *Proceedings of the National Academy of Sciences, USA*, *113*(6), 1588–1593. https://doi.org/10.1073/pnas.1515798113

Angelis, K., & dos Reis, M. (2015). The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. *Current Zoology*, *61*(5), 874–885. https://doi.org/10.1093/czoolo/61.5.874

Axelsson, E., Smith, N. G., Sundstrom, H., Berlin, S., & Ellegren, H. (2004). Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and Turkey. *Molecular Biology and Evolution*, *21*(8), 1538–1547. https://doi.org/10.1093/molbev/msh157.

Bartosch-Harlid, A., Berlin, S., Smith, N. G., Moller, A. P., & Ellegren, H. (2003). Life history and the male mutation bias. *Evolution*, *57*(10), 2398–2406. https://doi.org/10.1554/03-036

Bataillon, T. (2000). Estimation of spontaneous genome-wide mutation rate parameters: Whither beneficial mutations? *Heredity*, *84*, 497–501.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53–59. https://doi.org/10.1038/nature07517

Bergeron, L. A., Besenbacher, S., Bakker, J., Zheng, J., Li, P., Pacheco, G., Sinding, M. H., Kamilari, M., Gilbert, M. T., Schierup, M. H., & Zhang, G. (2020). The germline mutational process in rhesus macaque 2 and its implications for phylogenetic dating. bioRxiv. https://doi.org/10.1101/2020.06.22.164178

Bergeron, L. A., Besenbacher, S., Hahn, M. W., Turner, T., Versoza, C., Price, A. L., Armstrong, E., Belles, M., Carlson, J., Chen, H.-Y., Harris, K., López-Nandam, E. H., Moorjani, P., Pfeifer, S. P., Tiley, G. P., Yoder, A. D., Wang, R., Zhang, G., & Schierup, M. H. (in prep). Mutationathon: Towards standardization to estimate pedigree-based germline mutation rates.

Besenbacher, S., Hvilsom, C., Marques-Bonet, T., Mailund, T., & Schierup, M. H. (2019). Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nature Ecology and Evolution*, *3*(2), 286–292. https://doi.org/10.1038/s41559-018-0778-x

Besenbacher, S., Liu, S., Izarzugaza, J. M. G., Grove, J., Belling, K., Bork-Jensen, J., Huang, S., Als, T. D., Li, S., Yadav, R., Rubio-García, A., Lescai, F., Demontis, D., Rao, J., Ye, W., Mailund, T., Friborg, R. M., Pedersen, C. N. S., Xu, R., … Rasmussen, S. (2015). Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nature Communications*, *6*, 5969. https://doi.org/10.1038/ncomms6969

Bromham, L. (2011). The genome as a life-history character: Why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1577), 2503–2513. https://doi.org/10.1098/rstb.2011.0014

Campbell, C. R., Tiley, G. P., Poelstra, J. W., Hunnicutt, K. E., Larsen, P. A., & Lee, H.-J., Thorne, J. L., dos Reis, M., & Yoder, A. D. (2021). Pedigree-based and phylogenetic methods support surprising patterns of mutation rate and spectrum in the gray mouse lemur. *Heredity*, in press.

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular evolution. *Genetics*, *134*, 1289–1303.

Coe, B. P., Stessman, H. A. F., Sulovari, A., Geisheker, M. R., Bakken, T. E., Lake, A. M., Dougherty, J. D., Lein, E. S., Hormozdiari, F., Bernier, R. A., & Eichler, E. E. (2019). Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity.

Nature Genetics, *51*(1), 106–116. https://doi.org/10.1038/s41588-018-0288-4

Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., Grützner, F., & Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature*, *508*(7497), 488–493. https://doi.org/10.1038/nature13151

Course, M. M., Gudsnuk, K., Smukowski, S. N., Winston, K., Desai, N., Ross, J. P., Sulovari, A., Bourassa, C. V., Spiegelman, D., Couthouis, J., Yu, C.-E., Tsuang, D. W., Jayadev, S., Kay, M. A., Gitler, A. D., Dupre, N., Eichler, E. E., Dion, P. A., Rouleau, G. A., & Valdmanis, P. N. (2020). Evolution of a human-specific tandem repeat associated with ALS. *American Journal of Human Genetics*, *107*(3), 445–460. https://doi.org/10.1016/j.ajhg.2020.07.004

Crow, J. F. (1997). The high spontaneous mutation rate: Is it a health risk? *Proceedings of the National Academy of Sciences USA*, *94*, 8380–8386. https://doi.org/10.1073/pnas.94.16.8380

Crow, J. F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics*, *1*, 40–47. https://doi.org/10.1038/35049558

Davenport, E. S., Agrelius, T. C., Harmon, K. B., & Dudycha, J. L. (2021). Fitness effects of spontaneous mutations in a warming world. *Evolution*. https://doi.org/10.1111/evo.14208

Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L., & Thomas, W. K. (2000). High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science*, *289*(5488), 2342–2344.

dos Reis, M., Gunnell, G. F., Barba-Montoya, J., Wilkins, A., Yang, Z., & Yoder, A. D. (2018). Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Systematic Biology*, *67*(4), 594–615. https://doi.org/10.1093/sysbio/syy001

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences USA*, *88*, 7160–7164. https://doi.org/10.1073/pnas.88.16.7160

Duchene, S., Lanfear, R., & Ho, S. Y. W. (2014). The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution*, *78*, 277–289. https://doi.org/10.1016/j.ympev.2014.05.032

Ellegren, H. (2007). Characteristics, causes and evolutionary consequences of male-biased mutation. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1606), 1–10. https://doi.org/10.1098/rspb.2006.3720

Ellegren, H., & Fridolfsson, A.-K. (1997). Male-driven evolution of DNA sequences in birds. *Nature Genetics*, *17*, 182–184. https://doi.org/10.1038/ng1097-182

Estes, S., Phillips, P. C., Denver, D. R., Thomas, W. K., & Lynch, M. (2004). Mutation accumulation in populations of varying size: the distribution of mutational effects for fitness correlates in *Caenorhabditis elegans*. *Genetics*, *166*, 1269–1279.

Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*, *125*(1), 1–15. https://doi.org/10.1086/284325

Feng, C., Pattersson, M., Lamichhaney, S., Rubin, C. J., Rafati, N., Casini, M., Folkvord, A., & Andersson, L. (2017). Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. *Elife*, *6*, e23907.

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2020). A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution*, *37*(4), 1211–1223. https://doi.org/10.1093/molbev/msz296

Gao, F., Pan, X., Dodd-Eaton, E. B., Recio, C. V., Montierth, M. D., Bojadzieva, J., Mai, P. L., Zelley, K., Johnson, V. E., Braun, D., Nichols, K. E., Garber, J. E., Savage, S. A., Strong, L. C., & Wang, W. (2020). A pedigree-based prediction model identifies carriers of deleterious de novo mutations in families with Li-Fraumeni syndrome. *Genome Research*, *30*(8), 1170–1180. https://doi.org/10.1101/gr.249599.119

Gao, Z., Moorjani, P., Sasani, T. A., Pedersen, B. S., Quinlan, A. R., Jorde, L. B., Amster, G., & Przeworski, M. (2019). Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proceedings of the National Academy of Sciences USA*, *116*(19), 9491–9500. https://doi.org/10.1073/pnas.1901259116

Garimella, K. V., Iqbal, Z., Krause, M. A., Campino, S., Kekre, M., Drury, E., Kwiatkowski, D., Sá, J. M., Wellems, T. E., & McVean, G. (2020). Detection of simple and complex de novo mutations with multiple reference sequences. *Genome Research*, *30*(8), 1154–1169. https://doi.org/10.1101/gr.255505.119

Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Glémin, S., Ronfort, J., & Bataillon, T. (2003). Patterns of inbreeding depression and architecture of the load in subdivided populations. *Genetics*, *165*, 2193–2212. https://doi.org/10.1093/genetics/165.4.2193

Goetting-Minesky, M. P., & Makova, K. D. (2006). Mammalian male mutation bias: Impacts of generation time and regional variation in substitution rates. *Journal of Molecular Evolution*, *63*(4), 537–544. https://doi.org/10.1007/s00239-005-0308-8

Goldmann, J. M., Veltman, J. A., & Gilissen, C. (2019). De novo mutations reflect development and aging of the human germline. *Trends in Genetics*, *35*(11), 828–839. https://doi.org/10.1016/j.tig.2019.08.005

Goldmann, J. M., Wong, W. S. W., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A. B., Glusman, G., Vissers, L. E. L. M., Hoischen, A., Roach, J. C., Vockley, J. G., Veltman, J. A., Solomon, B. D., Gilissen, C., & Niederhuber, J. E. (2016). Parent-of-origin-specific signatures of *de novo* mutations. *Nature Genetics*, *48*(8), 935–939. https://doi.org/10.1038/ng.3597

Goodman, M. (1985). Rates of molecular evolution: the hominoid slowdown. *BioEssays*, *3*(1), 9–14. https://doi.org/10.1002/bies.950030104

Gudbjartsson, D. F., Helgason, H., Gudjonsson, S. A., Zink, F., Oddson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B. V., Hjartarson, E., Sigurdsson, G. T., Stacey, S. N., Frigge, M. L., Holm, H., Saemundsdottir, J., Helgadottir, H. T., Johannsdottir, H., Sigfusson, G., Thorgeirsson, G., … Stefansson, K. (2015). Large-scale whole-genome sequencing of the Icelandic population. *Nature Genetics*, *47*(5), 435–444. https://doi.org/10.1038/ng.3247

Haldane, H. B. S. (1947). The mutation rate of the gene for haemophilia, and its segregation ratios in males and females. *Annals of Eugenics*, *13*, 262–271. https://doi.org/10.1111/j.1469-1809.1946.tb02367.x

Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection. V. Selection and mutation. Mathematical Proceedings of the Cambridge Philosophical Society, volume 23 (pp. 838–844). Cambridge University Press.

Halligan, D. L., & Keightley, P. D. (2009). Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 151–172. https://doi.org/10.1146/annurev.ecolsys.39.110707.173437

Harland, C., Charlier, C., Karim, L., Cambisano, N., Deckers, M., Mni, M., Mullaart, E., Coppieters, W., & Georges, M. (2017). Frequency of mosaicism points towards mutation-prone early cleavage cell divisions in cattle. bioRxiv. https://doi.org/10.1101/079863

Harris, K. (2015). Evidence for recent, population-specific evolution of the human mutation rate. *Proceedings of the National Academy of Sciences U S A*, *112*(11), 3439–3444. https://doi.org/10.1073/pnas.1418652112

Harris, K., & Pritchard, J. K. (2017). Rapid evolution of the human mutation spectrum. *Elife*, *6*, e24284.

Heath, T. A., Huelsenbeck, J. P., & Stadler, T. (2014). The fossilized birth-death process for coherent calibration of divergence-time

estimates. *Proceedings of the National Academy of Sciences U S A*, *111*(29), E2957–2966. https://doi.org/10.1073/pnas.1319091111

Ho, S. Y. W., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., & Cooper, A. (2011). Time-dependent rates of molecular evolution. *Molecular Ecology*, *20*(15), 3087–3101. https://doi.org/10.1111/J.1365-294x.2011.05178.X

Howell, N., Smejkal, C. B., Mackey, D. A., Chinnery, P. F., Turnbull, D. M., & Herrnstadt, C. (2003). The pedigree rate of sequence divergence in the human mitochondrial genome: There is a difference between phylogenetic and pedigree rates. *American Journal of Human Genetics*, *72*(3), 659–670. https://doi.org/10.1086/368264

Hurst, L. D., & Ellegren, H. (1998). Sex biases in the mutation rate. *Trends in Genetics*, *14*, 446–452. https://doi.org/10.1016/S0168-9525(98)01577-7

Jonsson, H., Magnusdottir, E., Eggertsson, H. P., Stefansson, O. A., Arnadottir, G. A., Eiriksson, O., Zink, F., Helgason, E. A., Jonsdottir, I., Gylfason, A., Jonasdottir, A., Jonasdottir, A., Beyter, D., Steingrimsdottir, T., Norddahl, G. L., Magnusson, O. T., Masson, G., Halldorsson, B. V., Thorsteinsdottir, U., … Stefansson, K. (2021). Differences between germline genomes of monozygotic twins. *Nature Genetics*, *53*(1), 27–34. https://doi.org/10.1038/s41588-020-00755-1.

Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., Hardarson, M. T., Hjorleifsson, K. E., Eggertsson, H. P., Gudjonsson, S. A., Ward, L. D., Arnadottir, G. A., Helgason, E. A., Helgason, H., Gylfason, A., Jonasdottir, A., Jonasdottir, A., Rafnar, T., Frigge, M., … Stefansson, K. (2017). Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*, *549*(7673), 519–522. https://doi.org/10.1038/nature24018

Keightley, P. D., Ness, R. W., Halligan, D. L., & Haddrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics*, *196*(1), 313–320. https://doi.org/10.1534/genetics.113.158758

Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., Davey, J. W., & Jiggins, C. D. (2015). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, *32*(1), 239–243. https://doi.org/10.1093/molbev/msu302

Kessler, M. D., Loesch, D. P., Perry, J. A., Heard-Costa, N. L., Taliun, D., Cade, B. E., Wang, H., Daya, M., Ziniti, J., Datta, S., Celedón, J. C., Soto-Quiros, M. E., Avila, L., Weiss, S. T., Barnes, K., Redline, S. S., Vasan, R. S., Johnson, A. D., Mathias, R. A., … O'Connor, T. D. (2020). De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population. *Proceedings of the National Academy of Sciences U S A*, *117*(5), 2560–2569. https://doi.org/10.1073/pnas.1902766117

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*, 624–626. https://doi.org/10.1038/217624a0

Koch, E., Schweizer, R. M., Schweizer, T. M., Stahler, D. R., Smith, D. W., Wayne, R. K., & Novembre, J. (2019). De novo mutation rate estimation in wolves of known pedigree. *Molecular Biology and Evolution*, *36*(11), 2536–2547. https://doi.org/10.1093/molbev/msz159

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S. W., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. T., … Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, *488*(7412), 471–475. https://doi.org/10.1038/nature11396.

Krasovec, M., Rickaby, R. E. M., & Filatov, D. A. (2020). Evolution of mutation rate in astronomically large phytoplankton populations. *Genome Biology and Evolution*, *12*(7), 1051–1059. https://doi.org/10.1093/gbe/evaa131

Krasovec, M., Sanchez-Brosseau, S., & Piganeau, G. (2019). First estimation of the spontaneous mutation rate in diatoms. *Genome Biology and Evolution*, *11*(7), 1829–1837. https://doi.org/10.1093/gbe/evz130

Lanfear, R., Ho, S. Y. W., Jonathan Davies, T., Moles, A. T., Aarssen, L., Swenson, N. G., Warman, L., Zanne, A. E., & Allen, A. P. (2013). Taller plants have lower rates of molecular evolution. *Nature Communications*, *4*, 1879. https://doi.org/10.1038/ncomms2836

Lanfear, R., Kokko, H., & Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology & Evolution*, *29*(1), 33–41. https://doi.org/10.1016/j.tree.2013.09.009

Lanfear, R., Thomas, J. A., Welch, J. J., Brey, T., & Bromham, L. (2007). Metabolic rate does not calibrate the molecular clock. *Proceedings of the National Academy of Sciences USA*, *104*(39), 15388–15393. https://doi.org/10.1073/Pnas.0703359104

Langergraber, K. E., Prufer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J. C., Muller, M. N., Robbins, M. M., Schubert, G., Stoinski, T. S., Viola, B., Watts, D., Wittig, R. M., Wrangham, R. W., Zuberbuhler, K., Paabo, S., & Vigilant, L. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences U S A*, *109*(39), 15716–15721. https://doi.org/10.1073/pnas.1211740109

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, *30*(20), 2843–2851. https://doi.org/10.1093/bioinformatics/btu356

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Lindsay, S. J., Rahbari, R., Kaplanis, J., Keane, T., & Hurles, M. E. (2019). Similarities and differences in patterns of germline mutation between mice and humans. *Nature Communications*, *10*(1), 4053. https://doi.org/10.1038/s41467-019-12023-w

Link, V., Aguilar-Gomez, D., Ramirez-Suastegui, C., Hurst, L. D., & Cortez, D. (2017). Male mutation bias is the main force shaping chromosomal substitution rates in monotreme mammals. *Genome Biology and Evolution*, *9*(9), 2198–2210. https://doi.org/10.1093/gbe/evx155

Liu, H., Jia, Y., Sun, X., Tian, D., Hurst, L. D., & Yang, S. (2017). Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Molecular Biology and Evolution*, *34*(1), 119–130. https://doi.org/10.1093/molbev/msw226

Long, H., Winter, D. J., Chang, A. Y., Sung, W., Wu, S. H., Balboa, M., & Zufall, R. A. (2016). Low base-substitution mutation rate in the germline genome of the ciliate Tetrahymena thermophil. *Genome Biology and Evolution*, *8*(12), 3629–3639. https://doi.org/10.1093/gbe/evw223

Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, *26*(8), 345–352. https://doi.org/10.1016/j.tig.2010.05.003

Maddamsetti, R., & Grant, N. A. (2020). Divergent evolution of mutation rates and biases in the long-term evolution experiment with *Escherichia coli*. *Genome Biology and Evolution*, *12*(9), 1591–1603. https://doi.org/10.1093/gbe/evaa178

Marshall, C. R. (2019). Using the Fossil Record to Evaluate Timetree Timescales. *Frontiers in Genetics*, *10*, 1049. https://doi.org/10.3389/fgene.2019.01049

Martin, A. P., & Palumbi, S. R. (1993). Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences USA*, *90*, 4087–4091. https://doi.org/10.1073/pnas.90.9.4087

Martin, C. H., & Hohna, S. (2018). New evidence for the recent divergence of Devil's Hole pupfish and the plausibility of elevated mutation rates in endangered taxa. *Molecular Ecology*, *27*(4), 831–838. https://doi.org/10.1111/mec.14404

Martin, H. C., Batty, E. M., Hussin, J., Westall, P., Daish, T., Kolomyjec, S., Piazza, P., Bowden, R., Hawkins, M., Grant, T., Moritz, C., Grutzner, F., Gongora, J., & Donnelly, P. (2018). Insights into platypus

population structure and history from whole-genome sequencing. *Molecular Biology and Evolution*, *35*(5), 1238–1252. https://doi.org/10.1093/molbev/msy041

Martincorena, I., & Luscombe, N. M. (2013). Non-random mutation: The evolution of targeted hypermutation and hypomutation. *BioEssays*, *35*(2), 123–130. https://doi.org/10.1002/bies.201200150

McCluskey, K., Boundy-Mills, K., Dye, G., Ehmke, E., Gunnell, G. F., Kiaris, H., Polihronakis Richmond, M., Yoder, A. D., Zeigler, D. R., Zehr, S., & Grotewold, E. (2017). The challenges faced by living stock collections in the USA. *eLife*, *6*, e24611.

Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G. A., Schneider, V. A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., … Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, *585*(7823), 79–84. https://doi.org/10.1038/s41586-020-2547-7

Mitra, I., Huang, B., Mousavi, N., Ma, N., Lamkin, M., Yanicky, R., Shleizer-Burko, S., Lohmueller, K. E., & Gymrek, M. (2021). Patterns of de novo tandem repeat mutations and their role in autism. *Nature*, *589*(7841), 246–250. https://doi.org/10.1038/s41586-020-03078-7

Monaghan, P., & Metcalfe, N. B. (2019). The deteriorating soma and the indispensable germline: Gamete senescence and offspring fitness. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1917), 20192187. https://doi.org/10.1098/rspb.2019.2187

Moorjani, P., Amorim, C. E., Arndt, P. F., & Przeworski, M. (2016). Variation in the molecular clock of primates. *Proceedings of the National Academy of Sciences U S A*, *113*(38), 10607–10612. https://doi.org/10.1073/pnas.1600374113

Moorjani, P., Gao, Z., & Przeworski, M. (2016). Human germline mutation and the erratic evolutionary clock. *PLoS Biology*, *14*(10), e2000744. https://doi.org/10.1371/journal.pbio.2000744

Muryas, F., Zapata, L., Guigo, R., & Ossowski, S. (2020). The rate and spectrum of mosaic mutations during embryogenesis revealed by RNA sequencing of 49 tissues. *Genome Medicine*, *12*(1), 49. https://doi.org/10.1186/s13073-020-00746-1

Muyas, F., Zapata, L., Guigo, R., & Ossowski, S. (2020). The rate and spectrum of mosaic mutations during embryogenesis revealed by RNA sequencing of 49 tissues. *Genome Medicine*, *12*(1), 49. https://doi.org/10.1186/s13073-020-00746-1

Narang, P., & Wilson Sayres, M. A. (2016). Variable autosomal and X divergence near and far from genes affects estimates of male mutation bias in great apes. *Genome Biology and Evolution*, *8*(11), 3393–3405. https://doi.org/10.1093/gbe/evw232

Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., Miga, K. H., Eichler, E. E., Phillippy, A. M., & Koren, S. (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Research*, *30*(9), 1291–1305. https://doi.org/10.1101/gr.263566.120

Ohno, M. (2019). Spontaneous de novo germline mutations in humans and mice: rates, spectra, causes and consequences. *Genes and Genetic Systems*, *94*(1), 13–22. https://doi.org/10.1266/ggs.18-00015

Pfeifer, S. P. (2017). Direct estimate of the spontaneous germ line mutation rate in African green monkeys. *Evolution*, *71*(12), 2858–2870. https://doi.org/10.1111/evo.13383

Phillips, P. C. (1997). The rise and fall of new mutations. *Trends in Ecology and Evolution*, *12*(12), 466–468. https://doi.org/10.1016/S0169-5347(97)01215-9

Poelstra, J. W., Salmona, J., Tiley, G. P., Schüßler, D., Blanco, M. B., Andriambeloson, J. B., Bouchez, O., Campbell, C. R., Etter, P. D., Hohenlohe, P. A., Hunnicutt, K. E., Iribar, A., Johnson, E. A., Kappeler, P. M., Larsen, P. A., Manzi, S., Ralison, J. É. M., Randrianambinina, B., Rasoloarison, R. M., … Yoder, A. D. (2021). Cryptic patterns of speciation in cryptic primates: Microendemic mouse lemurs and the multispecies coalescent. *Systematic Biology*, *70*(2), 203–218. https://doi.org/10.1093/sysbio/syaa053

Qiu, F., Kitchen, A., Burleigh, J. G., & Miyamoto, M. M. (2014). Scombroid fishes provide novel insights into the trait/rate associations of molecular evolution. *Journal of Molecular Evolution*, *78*(6), 338–348. https://doi.org/10.1007/s00239-014-9621-4

Rahbari, R., Wuster, A., Lindsay, S. J., Hardwick, R. J., Alexandrov, L. B., Al Turki, S., Dominiczak, A., Morris, A., Porteous, D., Smith, B., Stratton, M. R., & Hurles, M. E. (2016). Timing, rates and spectra of human germline mutation. *Nature Genetics*, *48*(2), 126–133. https://doi.org/10.1038/ng.3469

Roach, J. C., Glusman, G., Smit, A. F. A., Huff, C. D., Hubley, R., Shannon, P. T., Rowen, L., Pant, K. P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L. B., Hood, L., & Galas, D. J. (2010). Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. *Science*, *328*(5978), 636–639. https://doi.org/10.1126/science.1186802

Rodriguez-Galindo, M., Casillas, S., Weghorn, D., & Barbadilla, A. (2020). Germline de novo mutation rates on exons versus introns in humans. *Nature Communications*, *11*(1), 3304. https://doi.org/10.1038/s41467-020-17162-z

Rogers, Y.-H., & Venter, J. C. (2005). Genomics: Massively parallel sequencing. *Nature*, *437*, 326–327. https://doi.org/10.1126/science.1117389(2005)

Rousselle, M., Simion, P., Tilak, M. K., Figuet, E., Nabholz, B., & Galtier, N. (2020). Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *PLoS Genetics*, *16*(4), e1008668. https://doi.org/10.1371/journal.pgen.1008668

Sasani, T. A., Ashbrook, D. G., Lu, L., Palmer, A. A., Williams, R. W., Pritchard, J. K., & Harris, K. (2021). A wild-derived antimutator drives germline mutation spectrum differences in a genetically diverse murine family. *bioRxiv*. https://doi.org/10.1101/2021.03.12.435196

Scally, A. (2016). The mutation rate in human evolution and demographic inference. *Current Opinion in Genetics and Development*, *41*, 36–43. https://doi.org/10.1016/j.gde.2016.07.008

Scally, A., & Durbin, R. (2012). Revising the human mutation rate: Implications for understanding human evolution. *Nature Reviews Genetics*, *13*(10), 745–753. https://doi.org/10.1038/nrg3295

Ségurel, L., Wyman, M. J., & Przeworski, M. (2014). Determinants of mutation rate variation in the human germline. *Annual Review of Genomics and Human Genetics*, *15*, 47–70. https://doi.org/10.1146/annurev-genom-031714-125740

Seoighe, C., & Scally, A. (2017). Inference of candidate germline mutator loci in humans from genome-wide haplotype data. *PLoS Genetics*, *13*(1), e1006549. https://doi.org/10.1371/journal.pgen.1006549

Shimmin, L. C., Chang, B.-H.-J., & Li, W.-H. (1993). Male-driven evolution of DNA sequences. *Nature*, *362*, 745–747. https://doi.org/10.1038/362745a0

Smeds, L., Qvarnstrom, A., & Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research*, *26*(9), 1211–1218. https://doi.org/10.1101/gr.204669.116

Smith, S. A., & Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science*, *322*, 86–89. https://doi.org/10.1126/science.1163197

Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G., & Lynch, M. (2012). Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences USA*, *109*(45), 18488–18492. https://doi.org/10.1073/pnas.1216223109

Tatsumoto, S., Go, Y., Fukuta, K., Noguchi, H., Hayakawa, T., Tomonaga, M., Hirai, H., Matsuzawa, T., Agata, K., & Fujiyama, A. (2017). Direct estimation of de novo mutation rates in a chimpanzee parent-offspring trio by ultra-deep whole genome sequencing. *Scientific Reports*, *7*(1), 13561. https://doi.org/10.1038/s41598-017-13919-7

Thomas, G. W. C. (2019). The causes of mutation and substitution rate variation in primates. (Ph.D.), ed. Indiana University.

Thomas, G. W. C., Wang, R. J., Puri, A., Harris, R. A., Raveendran, M., Hughes, D. S. T., Murali, S. C., Williams, L. E., Doddapaneni, H., Muzny, D. M., Gibbs, R. A., Abee, C. R., Galinski, M. R., Worley, K. C., Rogers, J., Radivojac, P., & Hahn, M. W. (2018). Reproductive longevity predicts mutation rates in primates. *Current Biology*, *28*(19), 3193–3197 e3195. https://doi.org/10.1016/j.cub.2018.08.050

Tiley, G. P., Poelstra, J. W., dos Reis, M., Yang, Z., & Yoder, A. D. (2020). Molecular clocks without rocks: new solutions for old problems. *Trends in Genetics*, *36*(11), 845–856. https://doi.org/10.1016/j.tig.2020.06.002

Vassilieva, L. L., & Lynch, M. (1999). The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics*, *151*, 119–129. https://doi.org/10.1093/genetics/151.1.119

Venn, O., Turner, I., Mathieson, I., de Groot, N., Bontrop, R., & McVean, G. (2014). Strong male bias drives germline mutation in chimpanzees. *Science*, *344*(6189), 1272–1275.

Wang, L., Ji, Y., Hu, Y., Hu, H., Jia, X., Jiang, M., Zhang, X., Zhao, L., Zhang, Y., Jia, Y., Qin, C., Yu, L., Huang, J. U., Yang, S., Hurst, L. D., & Tian, D. (2019). The architecture of intra-organism mutation rate variation in plants. *PLoS Biology*, *17*(4), e3000191. https://doi.org/10.1371/journal.pbio.3000191

Wang, R. J., Raveendran, M., Harris, R. A., Murphy, W. J., Lyons, L. A., Rogers, J., & Hahn, M. W. (2021). De novo mutations in domestic cat are consistent with an effect of reproductive longevity on both the rate and spectrum of mutations. bioRxiv. https://doi.org/10.1101/2021.04.06.438608

Wang, R. J., Thomas, G. W. C., Raveendran, M., Harris, R. A., Doddapaneni, H., Muzny, D. M., Capitanio, J. P., Radivojac, P., Rogers, J., & Hahn, M. W. (2020). Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. *Genome Research*, *30*(6), 826–s834. https://doi.org/10.1101/gr.255174.119

Whitney, K. D., & Garland, T. Jr (2010). Did genetic drift drive increases in genome complexity? *PLoS Genetics*, *6*(8), https://doi.org/10.1371/journal.pgen.1001080

Wong, W. S. W., Solomon, B. D., Bodian, D. L., Kothiyal, P., Eley, G., Huddleston, K. C., Baker, R., Thach, D. C., Iyer, R. K., Vockley, J. G., & Niederhuber, J. E. (2016). New observations on maternal age effect on germline de novo mutations. *Nature Communications*, *7*, 10486. https://doi.org/10.1038/ncomms10486

Wu, F. L., Strand, A. I., Cox, L. A., Ober, C., Wall, J. D., Moorjani, P., & Przeworski, M. (2020). A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *PLoS Biology*, *18*(8), e3000838. https://doi.org/10.1371/journal.pbio.3000838

Xie, Z., Wang, L., Wang, L., Wang, Z., Lu, Z., Tian, D., Yang, S., & Hurst, L. D. (2016). Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. A low rate in woody perennials and a higher mutagenicity in hybrids. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1841), https://doi.org/10.1098/rspb.2016.1016

Yang, S., Wang, L., Huang, J. U., Zhang, X., Yuan, Y., Chen, J.-Q., Hurst, L. D., & Tian, D. (2015). Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature*, *523*(7561), 463–467. https://doi.org/10.1038/nature14649

Yang, Z. (1995). A space-time process model for the evolution of DNA sequences. *Genetics*, *139*, 993–1005. https://doi.org/10.1093/genetics/139.2.993

Zeng, K., Jackson, B. C., & Barton, H. J. (2019). Methods for estimating demography and detecting between-locus differences in the effective population size and mutation rate. *Molecular Biology and Evolution*, *36*(2), 423–433. https://doi.org/10.1093/molbev/msy212

Zhu, Y. O., Siegal, M. L., Hall, D. W., & Petrov, D. A. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences U S A*, *111*(22), E2310–2318. https://doi.org/10.1073/pnas.1323011111

Zuckerkandl, E., & Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, *8*(2), 357–366. https://doi.org/10.1016/0022-5193(65)90083-4

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Yoder, A. D., & Tiley G. P. The challenge and promise of estimating the de novo mutation rate from whole-genome comparisons among closely related individuals. *Mol Ecol.* 2021; 30, 6087–6100. https://doi.org/10.1111/mec.16007