Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis

Yiguan Wang[1*], Darren J. Obbard[1]

1. Institute of Ecology and Evolution, University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK

- Corresponding author: Yiguan.Wang@ed.ac.uk

Abstract

Mutation is the ultimate source of all genetic variation, and over the last ten years the easy availability of whole-genome sequencing has permitted direct estimation of the mutation rate for many non-model species from across the tree of life. In this meta-analysis we make a comprehensive search of the literature for mutation rate estimates in eukaryotes, identifying 139 mutation accumulation (MA) and parent-offspring (PO) sequencing studies covering 71 species. Based on these recently accumulated data, we revisit differences in mutation rate between different phylogenetic lineages and update the known relationships between mutation rate and generation time, genome size, census population size, and neutral nucleotide diversity—while accounting for phylogenetic non-independence. We do not find a significant difference between MA and PO in estimated mutation rates, but we confirm that mammal and plant lineages have higher mutation rates than arthropods, and that unicellular eukaryotes have the lowest mutation rates. We find that mutation rates are higher in species with longer generation times and larger genome sizes, even when accounting for phylogenetic relationships. Moreover, although neutral nucleotide diversity is positively correlated with mutation rate, the gradient of the relationship is significantly less than one (on a logarithmic scale), consistent with higher mutation rates in populations with smaller effective size. For the 29 species for which data are available, we find that indel mutation rates are positively correlated with SNM rates, and that short deletions are generally more common than short insertions. Nevertheless, despite recent progress, no estimates of either SNM or indel mutation rates are available for the majority of deeply-branching eukaryotic lineages—or even for most animal phyla. Even among charismatic megafauna, experimental mutation rate estimates remain unknown for reptiles and amphibia and scarce for birds and fish.

Keywords: *de novo* mutation rate, eukaryotes, indels, phylogeny

Abbreviations: MA, mutation accumulation experiment; PO, parent-offspring experiment; SNMs, single nucleotide mutations; indels, insertions and deletion mutations; PGLMM, phylogenetic generalised linear mixed model; Mb, Mega base pair(s); HPD, highest posterior density.

Introduction

The *de novo* mutation rate, μ, is a key parameter in population and evolutionary genetics, appearing in almost every equation, either alone or combined with effective population size (N_e) in the compound parameter θ. However, it seems implicit that μ is of no interest to many evolutionary studies; researchers often use μ from one species when analysing another, simply for convenience—as if μ were exchangeable in a way that N_e is not (e.g. Wilding 2017). This is presumably because variation in μ among species is taken to be negligible compared to variation in N_e. But, if the mutation rate varied among multicellular eukaryotes as much as the recombination rate varies—around 1000-fold (Stapley, et al. 2017)—there would be much less need to invoke variation in N_e when explaining variation in (e.g.) genetic diversity.

One reason that we might expect mutation rate to vary little compared to recombination rate, is if recombination is optimised differently by different selective regimes (Stapley, et al. 2017), whereas mutation is minimised to a lower limit imposed by drift (the drift barrier hypothesis; Lynch, et al. 2016). However, it might be that in reality mutation rate is optimised to well above the drift-barrier limit by selection in favour of adaptability (Peck and Lauring 2018; Liu and Zhang 2021). If this were the case, then (as with recombination rates) we might find that mutation rates varied dramatically, even among closely related taxa.

Early estimates of the mutation rate leveraged visible phenotypes, such as the occurrence of genetic diseases in humans (Trimble and Doughty 1974; Nute and Stamatoyannopoulos 1984). For example, using haemoglobin M disease Stamatoyannopoulos and Nute (1982) estimated the mutation rate in the human haemoglobin beta gene at 7.4×10^{-9} per bp per generation. However, such estimates are prone to bias, for example if the mutation is required to be autosomal dominant and non-lethal (Nute and Stamatoyannopoulos 1984), and such approaches are not easily applicable in organisms for which disease is hard to observe. More commonly, using the equivalence between neutral mutation and substitution rates (Kimura 1968), many studies have used time-calibrated sequence divergence to estimate mutation rates. For example, comparison between humans and chimpanzees gives rise to estimates on the order of 10^{-8} per bp per generation (Kondrashov and Crow 1993; Drake, et al. 1998; Nachman and Crowell 2000). However, such indirect estimates are also limited (Kondrashov 2003). First, because generation time and calibration dates are prone to substantial uncertainty (e.g. Obbard, et al. 2012). Second, because it is hard to distinguish between unconstrained and weakly constrained sites (Harmon, et al. 2021). Third, because mutation saturation in hotspots may lead to an underestimation, especially for sequences with a long

divergence time (Sigurðardóttir, et al. 2000). In addition, phylogenetic-calibration approaches reflect the long-term mutation rate, averaging across biologically important factors such as generation time, sex differences, life-stage differences, and inter-individual variation.

In contrast to phenotypic and phylogenetic approaches, mutation-accumulation (MA) and parent-offspring (PO) sequencing can provide direct estimates of the per-generation mutation rate. Mutation accumulation utilises the accumulation of spontaneous mutations in a single inbred genome over multiple generations, by minimising the effectiveness of natural selection (Keightley, et al. 2009). Given the low mutation rates in eukaryotes, the multiple generations in an MA experiment allow sufficient mutations for them to be easily detected at the end. Although the first use of MA to estimate mutation rates phenotypically dates back to the 1970s (Mukai and Cockerham 1977), the availability of whole genome sequencing and its decreasing price have recently made it more widely applicable. Mutation accumulation best suits easily maintained species with short generation times, for example nematodes (Denver, et al. 2004; Meier, et al. 2014; Konrad, et al. 2019; Saxena, et al. 2019), fruit flies (Haag-Liautard, et al. 2007; Keightley, et al. 2009; Schrider, et al. 2013; Huang, et al. 2016), green algae (Ness, et al. 2012; Lopez-Cortegano, et al. 2021), and water fleas (Ho, et al. 2020; Keith, et al. 2021). However, MA may underestimate mutation rates (Lynch and Walsh 1998) because natural selection will still come into play for highly deleterious or lethal mutations (Baer, et al. 2007).

By sequencing parents and offspring, and counting the mutations arising over a single generation, the parent-offspring approach avoids losing lethal mutations (unless they are dominant) and provides the most direct estimate of mutation rate. It is not limited to model species (Keightley, et al. 2014; Yang, et al. 2015; Wang, et al. 2022b), and is suitable for large animals outside of the laboratory (Harland, et al. 2017; Wang, et al. 2022a). Parent-offspring sequencing has been particularly widely used in humans, where MA is not applicable. However, the challenge of robustly identifying new mutations in natural heterozygous genomes is formidable, as the number of mutations arising within one generation is small when compared to the impact of sequencing and mapping errors. For example, Bergeron, et al. (2022) compared the estimates from five different research groups for the same family of rhesus macaques and found that the highest estimates could be twice the lowest ones. Minimising the rates of false positive and false negative mutation-calls is therefore critical for the PO approach (Yoder and Tiley 2021).

Relaxed-clock phylogenetic studies generally identify substantial among-lineage variation in the evolutionary rate, suggestive of variation in mutation rates among taxa (Ho, et al. 2015). However, such studies cannot separate the impact of generation time and mutation rate, and may be biased

124 by variation in the action of (weak) selection (Harmon, et al. 2021). As more direct experimental
125 estimates have become available, it has become possible to draw comparisons within particular
126 clades such as primates (Chintalapati and Moorjani 2020), or vertebrates (Yoder and Tiley 2021;
127 Bergeron, et al. 2022). However, there has not been a wider analysis, nor one that includes both PO
128 and MA studies. Here we perform a comprehensive meta-analysis of published eukaryotic SNM and
129 indel mutation-rate estimates that use direct sequencing of multiple generations, either through MA
130 or PO comparison. We focus on the variation in mutation rate among species, and the role of the
131 host phylogeny, experimental approach, generation time and population size in determining
132 variation in estimates rates.

133 **Results and Discussion**

134 ***Studies included***

135 We searched for eukaryotic mutation-rate studies in Clarivate Web of Science, and then we
136 manually examined these studies and identified those containing mutation rate estimates (see
137 Methods). In total, we identified 139 studies using either MA or PO approaches for inclusion,
138 covering 71 species (see Availability of data and materials). Studies on human mutation rate
139 accounted for a third of the identified literature, followed by unicellular organisms (20%),
140 arthropods (17%), primates (8%), and plants (8%) (**Figure 1A)**. However, across all groups, most
141 studies focused on model species: yeast in the unicellular group, *Drosophila* in arthropods, mice in
142 mammals, and *Arabidopsis* in plants. These estimates tended to derive from MA analyses rather
143 than PO studies (61 vs. 38 among non-human studies). Nevertheless, as sequencing has become
144 cheaper, there has been a rapid increase in the number of published experimental estimates of the
145 mutation rate, with an average of 11 studies each year in eukaryotic species since 2010 (**Figure 1B**).

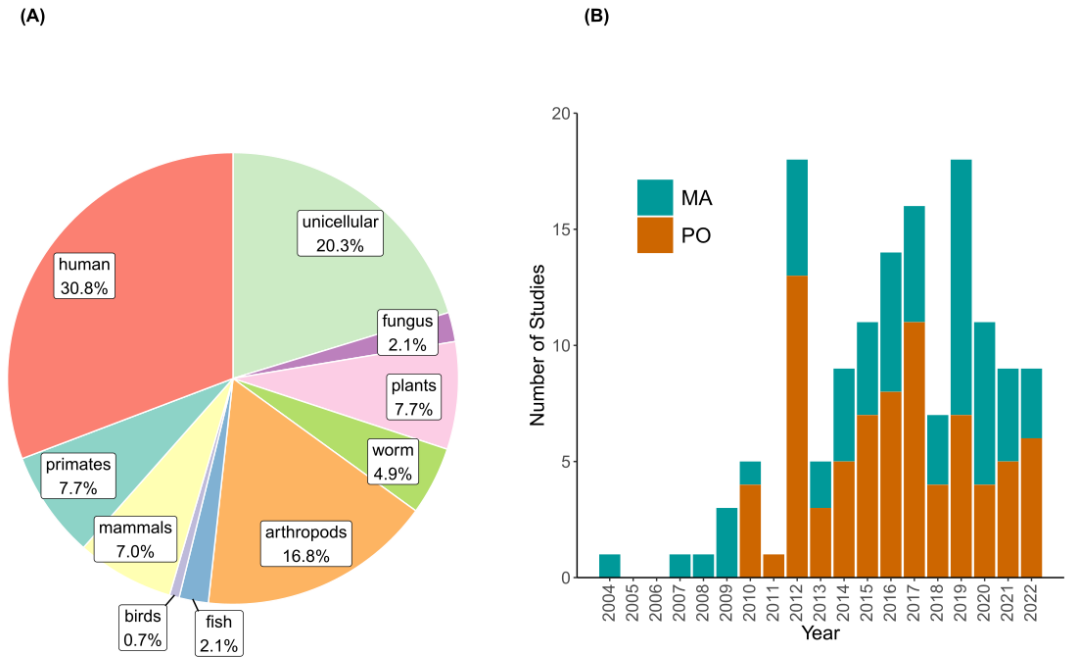


Figure 1: A summary of *de novo* mutation rate estimates included here. (A) The proportion of studies in different taxonomic groups. (B) The number of studies by year. MA: mutation accumulation studies, PO: parent-offspring studies.

Single nucleotide mutation rates across eukaryotes

To summarise the single nucleotide mutation (SNM) rates in different lineages, we obtained a time-scaled phylogenetic tree of all represented species from http://www.timetree.org/ (Kumar, et al. 2022), and used this to fit the covariance among species in a Phylogenetic Generalised Linear Mixed Model (PGLMM) using the Bayesian mixed model package MCMCglmm (e.g. Hadfield and Nakagawa 2010). Treating species as a random effect, this provides a posterior estimate of mutation rate for each species (and each internal node) under a Brownian motion model of trait evolution (see Methods), given the observations and relationships to other species. In these data, the phylogenetic effect accounts for 94.3% of the random-effects variance, with 95% HPD interval (85.4%, 98.6%).

Among the chosen species groupings, the unicellular species had the lowest mutation rate ranging from 0.02×10^{-9} per generation per base pair in ciliates to 0.74×10^{-9} in haptophyta. This estimate is more than an order of magnitude lower than rates in multicellular species (below; **Figure 2**), and may either be attributable to there being a single cell division per generation, or possibly to a lower

limit set by the drift barrier in species with larger N_e (Sung, et al. 2012a). We also observed relatively low SNM rates for a nematode at 2.27×10^{-9}, arthropods at 3.56×10^{-9}, fish at 7.37×10^{-9}, and a bird at 6.15×10^{-9} (random-effects estimates from our model, given the phylogeny). In particular, for arthropods (which were represented by multiple species) the lowest mutation rate was identified as 0.46×10^{-9} in pea aphids (per parthenogenic generation) and the highest one was only 4.97×10^{-9} in fruit flies. In contrast to these species, mammals and plants had relatively high mutation rates, including non-primate mammals at 8.76×10^{-9} and primates at 9.60×10^{-9}. The plants (excluding duckweed) and a multicellular fungus had the highest rates of all, ranging from 8.32 to 44.34×10^{-9} (**Figure 2**), and it is tempting to speculate that this might be attributable to the lack of a segregated germline (but see Wang, et al. (2019)).

To assess the impact of the experimental methods on rate estimates, we included the experimental approach (MA or PO) as a fixed effect in the PGLMM. Overall, we found that while the mean value from PO studies was higher than the value from MA studies, when accounting for variation among species and the covariance between them induced by the phylogeny, this effect was not significant (MCMC *p*-value = 0.659). We then examined comparisons between MA and PO in fruit flies and mice more closely, as these two species have been investigated using both methods. The rate estimated using MA in fruit flies was 5.28×10^{-9}, but 3.20×10^{-9} using PO (**supplementary Figure S1**), and in mice the MA estimate was 5.40×10^{-9} while the PO estimate was 4.03×10^{-9} (**supplementary Figure S2**), and these differences were likewise not significant. Note, however, that the mutation rate may be heterogeneous among different populations: for *D. melanogaster*, the estimated mutation rate from African populations was lower than that from European population and North American population (Chan, et al. 2012; Wang, et al. 2022c). To make a more robust analysis would require samples of multiple populations for each species, which are not yet available.

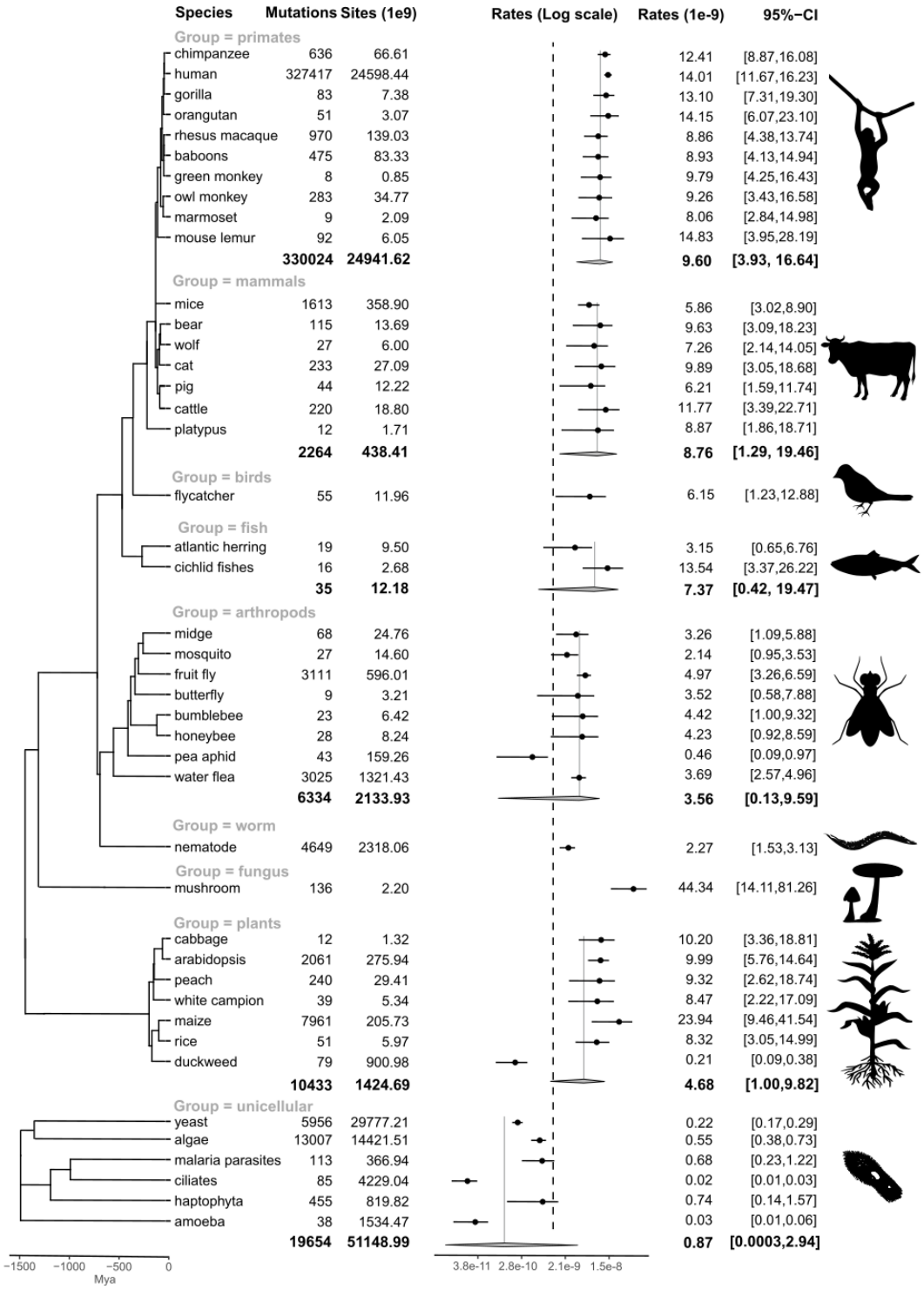


Figure 2. A phylogenetic meta-analysis of SNM rates for different species identified from the literature. Mutation rates and 95% CIs were estimated using a PGLMM. For each group, we estimated the rate for their most recent common ancestor (shown in bold under each group and

191 vertical solid grey lines). The vertical dashed line indicates the mutation rate of common ancestor of
192 the tree.

193 ***Biological predictors of the SNM rate***

194 Many biological factors have been proposed to be associated with mutation rate, including
195 generation time (Wu and Li 1985; Li, et al. 1987; Bailey, et al. 1991; Keightley and Eyre-Walker 2000),
196 genome size (Drake 1991; Lynch 2010), and the rate of genetic drift (Lynch, et al. 2016) — perhaps
197 through the impact of population size (Lanfear, et al. 2014; Krasovec, et al. 2020). To test for a
198 dependence of mutation rate on each of these factors, we ran three further models, including each
199 of them separately as an additional fixed effect in a PGLMM.

200 We found a significant positive relationship with generation time (Bayesian PGLMM *p*MCMC <
201 0.0002; **Figure 3A**), such that a longer generation time predicted a higher mutation rate per
202 generation. If mutations resulted only from replication errors, and most species had a similar
203 number of cell divisions per generation, then the per-generation mutation rate would not scale with
204 generation time (Wu and Li 1985; Thomas and Hahn 2014). However, in reality, the number of
205 germline cell divisions per generation varies within and between species. In humans, there are
206 around 401 cell divisions per 30-year generation in males and 31 in females (Drost and Lee 1995;
207 Ohno 2019), but in mice it is 62 per 9-month generation in males and 25 in females. In *Drosophila*
208 the sexes are much more similar, with an estimated at 35.5 cell divisions in a 18-day generation for
209 males but 36.5 in an 25-day generation in females in *D. melanogaster* (Drost and Lee 1995). Within a
210 species, a long generation time is likely to permit more cell divisions and thus a higher mutation rate
211 per generation, and many studies in humans and primates have shown that mutations are heavily
212 male-biased and correlated with paternal ages (Kong, et al. 2012; Thomas, et al. 2018; Wang, et al.
213 2020; Kaplanis, et al. 2022).

214 Alternatively, mutations may not result predominantly from replication errors. Recent studies
215 suggest that not all germline mutations track cell divisions, and DNA damage may contribute to the
216 mutation rate (Wu, et al. 2020; Wu and Przeworski 2022). Independent of cell divisions, this
217 mutation rate would be correlated with calendar time, and thus species with a longer generation
218 time would have a higher rate. The relationship we observe between generation time and per-
219 generation mutation rate could therefore be a consequence of either a greater number of cell
220 divisions or of accumulating damage over time, and it will be hard to separate these potential causes
221 (Gao, et al. 2016; Wu and Przeworski 2022).

222 We also found evidence that larger genome size predicts a higher per-generation mutation rate
223 (*p*MCMC = 0.0020; **Figure 3B**). Mutation rate has previously been found to correlate negatively with
224 genome size in prokaryotes (a phenomenon termed "Drake's rule" (Drake 1991; Bourguignon, et al.
225 2020; Marais, et al. 2020)), but positively in eukaryotes (Lynch 2010). Two 'outlier' unicellular
226 eukaryotes with large genomes but very low mutation rates have (*Paramecium tetraurelia* and
227 *Chlamydomonas reinhardtii*), have in the past been explained by appeal to their large effective
228 population size and the 'drift-barrier' hypothesis (Sung, et al. 2012a). With the increased availability
229 of mutation rates accrued over the past decade, we find that the mutation rate of *C. reinhardtii* is
230 not as low as previously thought, and this species now fits the positive trend in eukaryotes well
231 (**Figure 3B**). However, the mutation rates of *Tetrahymena thermophila* (Long, et al. 2016), *P.*
232 *tetraurelia*, and *Dictyostelium discoideum* (Kucukyildirim, et al. 2020) are indeed very low. Under
233 'drift barrier' hypothesis, these outliers could be caused by the large *N* in these species, and *T.*
234 *thermophila* would be expected to have the highest *N*. In fact, the N_e in *T. thermophila* is smaller
235 than that in *P. tetraurelia* (1.1×10^8 *vs.* 1.2×10^8) (Sung, et al. 2012b; Long, et al. 2016).

236 If census population size were a good predictor of the rate of genetic drift (N_e) then a significant
237 negative relationship between them could be consistent with the 'drift barrier' hypothesis; that is
238 that mutation rates are minimised by natural selection until the point at which the impact of
239 selection for lower mutation rate is counterbalanced by allele frequencies changes due to drift
240 (Lynch 2010). A negative relationship was indeed superficially apparent in our data (**Figure 3C**), but
241 this relationship was not significant when accounting for the non-independence among species in a
242 phylogenetic mixed model (*p*MCMC = 0.3350). Nevertheless, given the weak link between N_e and
243 census population size (Buffalo 2021), and the challenge of estimating census size (especially for
244 smaller organisms), the lack of a relationship is perhaps unsurprising.

245 In each of the three models above, we fitted generation time, population size, and genome size
246 individually as predictors of mutation rate in a regression model. However, these factors tend to be
247 highly correlated: large organisms tend to have longer generation times, smaller population sizes,
248 and often larger genomes (Martin and Palumbi 1993; Buffalo 2021) and is therefore hard to
249 distinguish between casual relationships and confounding correlations. Indeed, including all three
250 factors (generation time, genome size, population size) together as fixed terms in a single model
251 resulted in none of them being significant, although the effect of generation time was marginal with
252 *p*MCMC = 0.0781 (versus *p*MCMC = 0.3855 for genome size and 0.6844, for population size).

253 ***The drift-barrier hypothesis***

The most direct test of the drift-barrier hypothesis would be a regression of SNM rate on N_e. However, long-term N_e can only be estimated from putatively neutral genetic diversity (i.e. π_S) by making use of the relationship $\theta_\pi = 4N_e\mu$, and unfortunately any uncertainty in μ could induce a spurious correlation in a regression of μ on $\pi/4\mu$. Instead, error in both π_S and μ can be accounted for by fitting a bivariate linear mixed model, parameterised in terms of a regression of π_S on μ (see Methods). Then, on a logarithmic scale, the gradient of this relationship would be expected to be 1 if μ and N_e are independent of each other. Fitting only these terms, we found that π_S increased with μ, but that the gradient of this relationship was significantly less than 1 (gradient = 0.390 [0.078, 0.709]; *p*MCMC = 0.0002; **Figure 3D**). This suggests that π_S does not grow in direct proportion to μ, such that populations with larger N_e tend to have a lower mutation rate—as predicted by the drift barrier hypothesis (Lynch, et al. 2016).

Nevertheless, such a relationship could be driven by other factors. For example, if putatively silent sites are in fact weakly constrained, then the action of direct selection would lead to lower diversity in populations with larger N_e. Perhaps more seriously, if species with small N_e tend to have a longer generation time, and longer generation time causes higher mutation rates (i.e. **Figure 3A**), then a higher μ in species with low N_e could be driven by a mechanistic generation-time effect, rather than the drift barrier. We attempted to account for this latter hypothesis by fitting generation time as a predictor of both μ and π_S, thereby accounting for the direct impact of generation time on the relationship between them. In this final model, μ increased with generation time (*p*MCMC = 0.0004) and π did not (*p*MCMC = 0.9457), but the gradient of the regression of π_S on μ could no longer be distinguished from 1 (gradient = 0.688 [0.109, 1.230]; *p*MCMC = 0.1324).

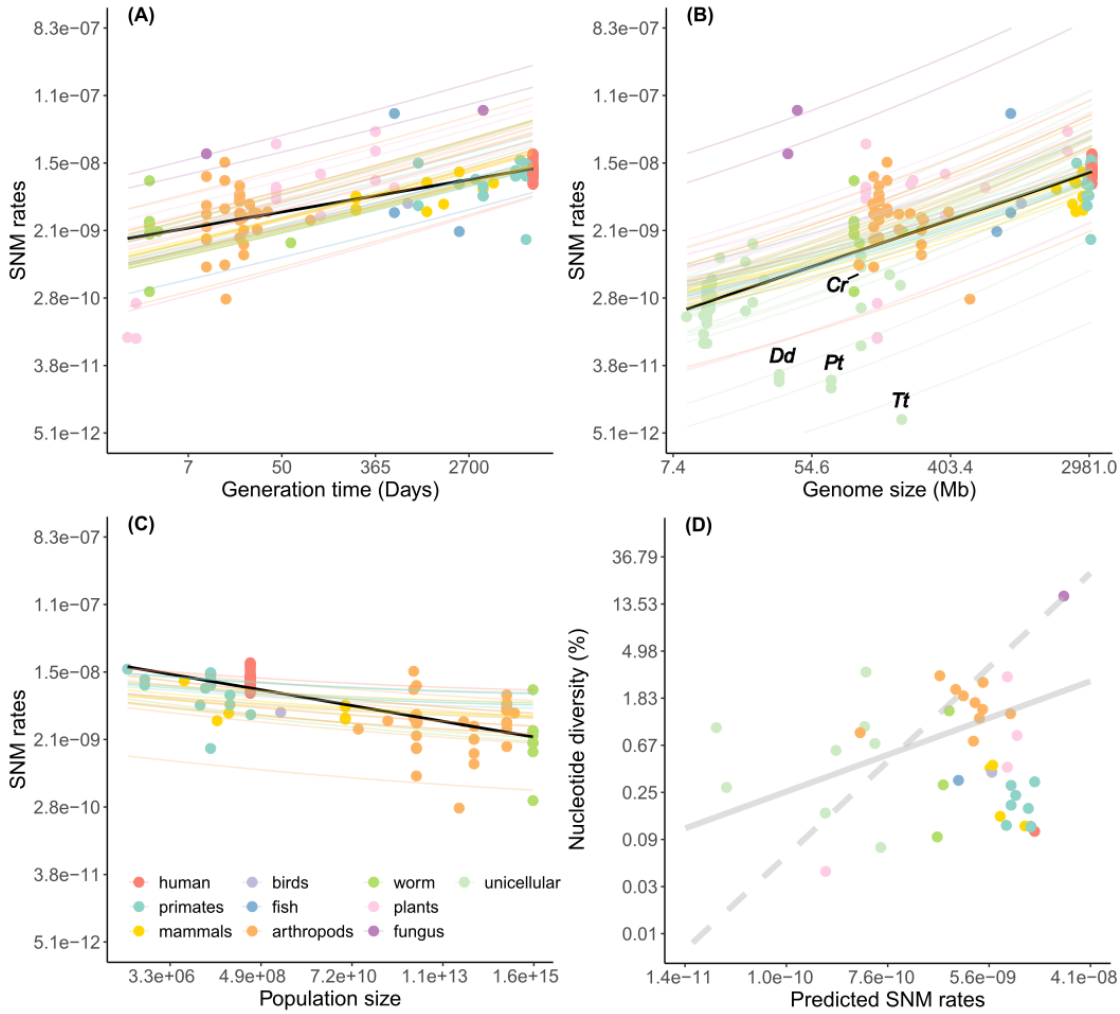
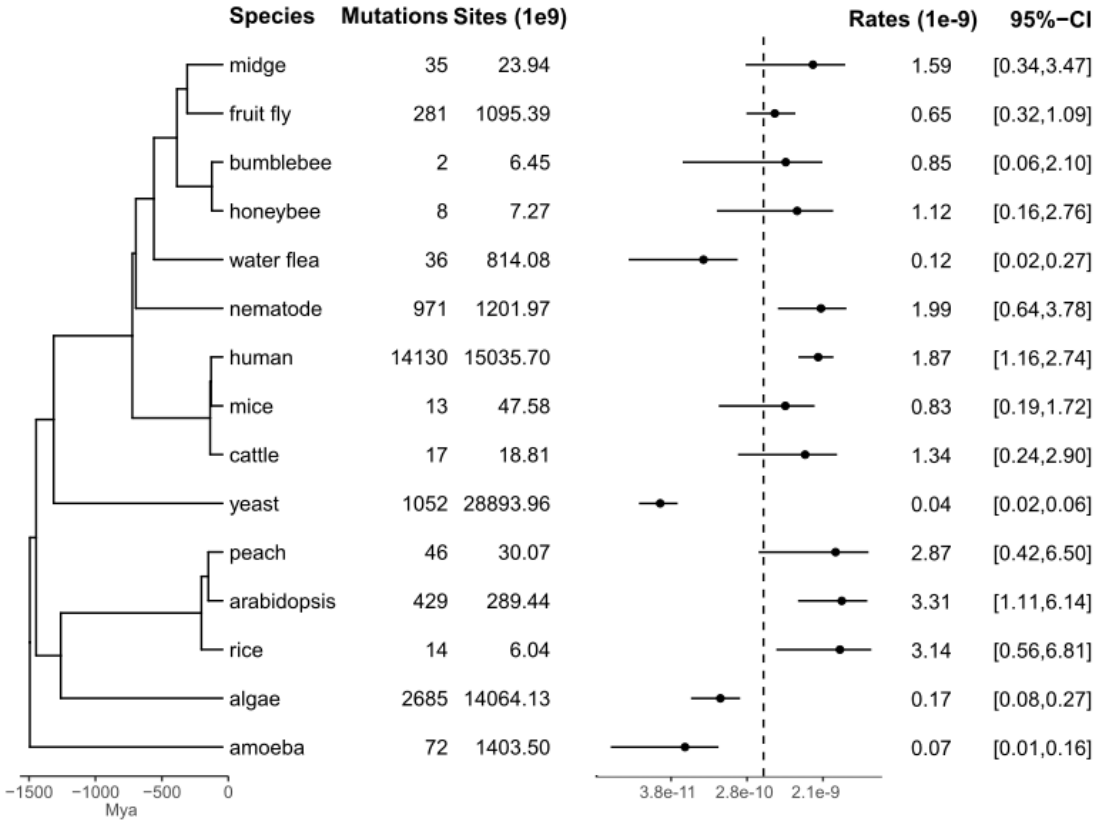


Figure 3. Phylogenetic linear regressions of SNM rate on (A) generation time, for 45 species, (B) genome size, for 63 species and (C) population size, for 25 species. The solid black lines represent simple linear regressions; the coloured lines represent different PGLMM regressions conditional on each of the species random effects. (D) a regression of pairwise nucleotide diversity (π_S) on SNM rate (μ), fitted as a bivariate model in MCMCglmm for 41 species. Grey solid line indicates the model regression line; grey dashed line indicates slope = 1. All axes are plotted on a log scale; Cr: *C. reinhardtii*; Pt: *P. tetraurelia*; Tt: *T. thermophila*; Dd: *D. discoideum*.

The mutation rates of short indels across eukaryotes

Although the rate of indel mutation is estimated much less often than the SNM rate, our literature search identified estimates for 29 species. We found that the indel mutation rate across eukaryotes displayed a similar pattern to that of the SNM rate (**Figure 4**), with a correlation coefficient between

288 the two of 0.76 (95% HPD CI: 0.40, 0.99; *p*MCMC = 0.0030), with the SNM rate being generally higher
289 than the indel mutation rate, except in two lineages (nematodes and amoebae) for which the indel
290 mutation rate was higher (**Figure 5A**). The unicellular species had the lowest indel mutation rate,
291 from 0.04×10^{-9} in yeast to 0.17×10^{-9} in algae, followed by arthropods with 0.65 (0.05, 1.61) $\times 10^{-9}$.
292 Mammals had a relatively higher indel mutation rate with 1.23 (0.35, 2.33) $\times 10^{-9}$, with humans at
293 the higher end of this range at 1.87 (1.16, 2.74) $\times 10^{-9}$. Of all the studied species, plants had the
294 highest indel mutation rate, which was as high as 2.83 (0.40, 6.28) $\times 10^{-9}$.



295

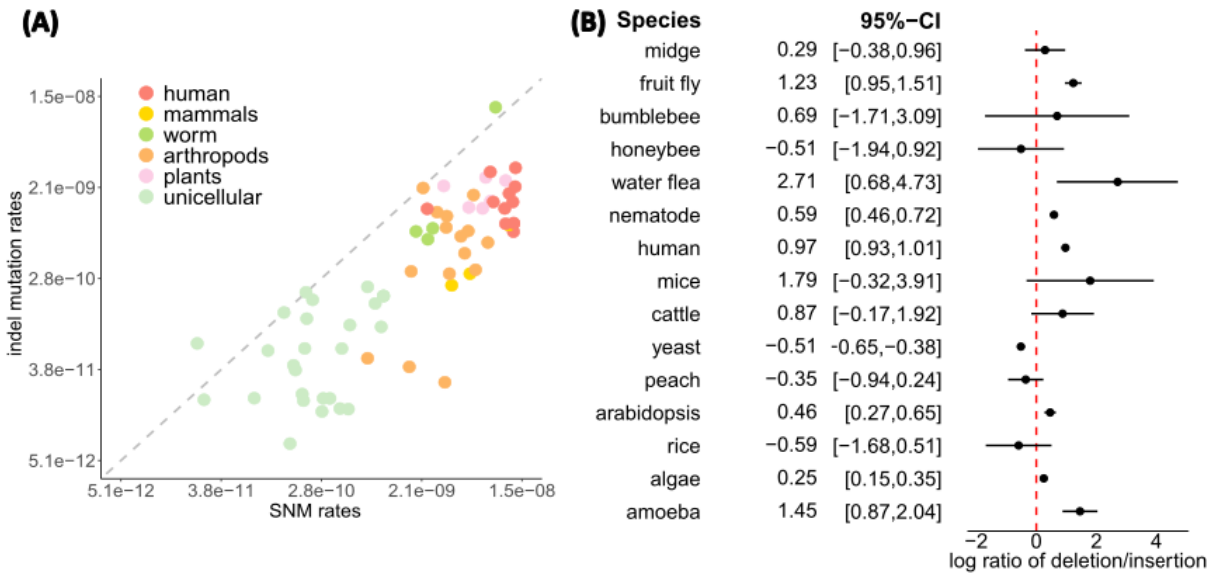
296 **Figure 4**. An analysis of short indel mutation rates for different species. The estimated mutation
297 rates are presented on a log scale for clarity.

298

299 Short deletions generally outnumbered short insertions across eukaryotes (**Figure 5B,**
300 **supplementary Figure S3**). The ratio of deletions over insertions was as high as 15.00 in the water
301 flea, 4.28 in amoeba, 3.43 in fruit flies, 2.64 in humans, 1.80 in nematode, 1.58 in Arabidopsis and
302 1.28 in algae. Honeybees and two plants (rice and peach) exhibited more insertions than deletions,

303 but the difference wasn't statistically significant (*p*-value > 0.05, binomial test). Although yeast
304 shows significantly more insertions overall, this estimate is dominated by two large studies and an
305 excess of deletions was still reported in yeast across other studies (**supplementary Figure S3**).

306 A bias toward deletion events has been known for at least 40 years since de Jong and Rydén (1981)
307 found a four-fold excess of deletions over insertions in protein sequences. Petrov (2002) argued that
308 the deletion bias was likely due to a thermodynamic disparity between short deletions and
309 insertions that makes deletion events easier. Studies in plants have suggested that unequal
310 homologous recombination and illegitimate recombination may be another important cause of small
311 deletions (Bennetzen, et al. 2005). Whatever the cause, the dominance of deletions over insertions
312 seems convincing, and many early studies argued that these short deletions resulted in genome
313 contraction (Petrov 2002; Gregory 2003; Thomas, et al. 2003). However, the role of deletions in
314 shaping genome size remains controversial (Gregory 2003, 2004). A recent study in birds and
315 mammals indicates that short deletions alone cannot explain the DNA content loss (Kapusta, et al.
316 2017), and we fail to find any evidence of a relationship between deletion rate and genome size here
317 (**supplementary Figure S4;** but note that power is likely to be low). It seems probable that other
318 structural mutations, such as large deletions, gene duplication and transposable elements, may have
319 greater impacts on genome size. For example, in *Drosophila* Wang, et al. (2022c) found transposable
320 element insertions to be more frequent than SNMs, and an order of magnitude more frequent than
321 short indels. Together, these studies suggest that the impact of short deletions on genome size is
322 likely to be small, despite their higher frequency relative to short insertions.



323

Figure 5. The correlation between indel mutation rates and SNM rates plotted on a log scale with a 1:1 ratio shown as a dashed line (A). The aggregated log ratios of deletion/insertion in eukaryotes, (B).

Concluding remarks

All genetic variation arises from *de novo* mutations. To estimate the rate at which these arise, studies have moved from investigating a few genes that result in phenotypic change, to whole genome sequencing of MA lines or pedigrees. Ten years ago, Campbell and Eichler (2013) called for more efforts to sequence genomes from non-human primate families to understand 'how the rate has changed in different lineages'. And, in the present data-rich and tool-rich era we have now seen the sequencing of not just multiple non-human primates, but many taxa representing major clades across the tree of life. Nevertheless, the 'tree of mutations' is far from complete, at either the large or small scale. Even within animals, the vast majority of phyla have never been examined, and with the exception of primates, no groups of close relatives have been examined.

However, as progress is being made toward a tree of mutation rates, attention still needs to focus on methodological heterogeneity. Different data filtering strategies may result in non-negligible differences in mutation rate estimates (Bergeron, et al. 2022). Although many strategies have been proposed (Bergeron, et al. 2022), no standardised pipeline has yet been agreed upon or widely adopted. All studies must therefore carefully address the issues of false positives, by manual curation and/or PCR follow-up, and false negatives, by simulation. Biases may also come from the samples used in each study. Besides the biological factors that affect mutation rates (age, sex), environmental factors can also exert influences on mutation rates, such as temperature (Belfield, et al. 2021; Waldvogel and Pfenninger 2021), osmotic stress (Hasan, et al. 2022) or ultraviolet light (Xu, et al. 2019). Samples collected from the wild rather than reared in laboratories may reflect the *real* mutation rate for the studied species, but the variation that this introduces may also make it harder to draw broad conclusions. Despite these challenges, the data to come are exciting. In particular, as sequencing technology advances, we expect improved accuracy in long-read sequencing facilitating the phasing of mutations and permitting the easy detection of complex mutations and larger-scale structural changes.

Methods

355 We made an exhaustive literature search for publications on the experimental estimation of
356 mutation rate. We first searched Clarivate Web of Science up to the date of 21 September 2022 for
357 'Title/Keywords/Abstract' fields containing ("mutation rate" | "mutation rates" | "mutational rate" |
358 "mutational rates"). As much of the resulting literature was related to somatic mutations (e.g.,
359 cancers) or virus mutations, we excluded '("tumor" | "cancer" | "clinical" | "virus")' in the search
360 field of 'Title/Keywords/Abstract'. This search strategy resulted in 9462 studies with the earliest one
361 dated 1928. We also searched for references to mutation-rate estimates in other papers not
362 captured by our search. As bacteria are outside of our study, we manually removed studies related
363 to bacterial mutations. These filters resulted in 174 studies. Of these studies, we further excluded 20
364 studies that used phylogenetic approaches to infer mutation rates, 14 studies that used phenotypic
365 markers, and two other studies that used recombination landscape or the site frequency spectrum
366 to infer mutation rates statistically. This resulted in a total of 139 studies in our final dataset.

367 From each of the potential studies we extracted the year of the publication, the species, the
368 population, mutation type (SNM, short insertion, short deletion), the units of mutation rate (per bp
369 per year, per bp per generation), the mean and 95% CI of mutation rate, the number of identified
370 mutation events, the number of callable sites, the methods used to infer mutation rate (PO, MA or
371 others), the sample size, the number of generations (if MA), the sequencing technology, the
372 sequencing depth and other pertinent observations. If a 95% confidence or credible interval was not
373 provided in the original study, we calculated a 95% confidence interval using 'binconf()' from the
374 Hmisc R package (version 4.7-0), based on the number of mutation events and callable sites and
375 assuming a binomial distribution (Girard, et al. 2011). If a study did not provide the mutation counts,
376 we inferred an effective number of mutation events based on the reported mutation rate and its
377 confidence interval. For studies in which the number of callable sites was not reported, an estimate
378 was made from the number of mutation events and the reported mutation rate. If a study *only*
379 reports mutation rate, we assumed one mutation event and calculated the corresponding callable
380 sites based on mutation rates. This will necessarily underestimate the precision with which rates are
381 estimated (minimising their weight in the analysis), while keeping it in the meta-analysis with
382 unchanged mean values.

383 To compare the SNM rates across different species, we conducted a phylogenetic meta-analysis (i.e.
384 a combined analysis of published data) using the mutation data from MA or PO studies. We obtained
385 the phylogenetic tree and relative divergence dates for the species included in this study from
386 http://www.timetree.org/ (Kumar, et al. 2022). We performed the analyses using a phylogenetic
387 mixed model with a unit-scaled phylogeny as random effect, assuming a Poisson distribution for the
388 occurrence of *de novo* mutations, off-set by the number of callable sites (i.e. enforcing a directly

proportional relationship between the expected number of mutations and the number of sites observed). The analysis was conducted using the Bayesian mixed-model R-package MCMCglmm (version 2.32) (Hadfield 2010) with a weak prior for the variance of random effects and a fully-informative prior fixed at 1 for the fixed effect of 'callable sites', using the following code in R:

```
vv <- diag(2)
diag(vv) <- c(1e6, 1e-8)
prior <- list(B=list(mu=c(0,1), V=vv),
              G=list(G1=list(V=1, nu=1, alpha.mu=0, alpha.V=100)))
```

The syntax for the basic model in R was:

```
m1 <- MCMCglmm(Mutation ~ log(Callable),
          random= ~ Species,
          ginverse=list(Species=InverseTree),
          prior = prior, data=aa,
          nitt=5001000, thin=1000, burnin=10000,
          family = "poisson", pr=TRUE)
```

where 'Mutation' was the number of mutations and 'Callable' was the number of callable sites. 'Species' is the random effect and 'InverseTree' is the expected phylogenetic covariance among species based on relative branch lengths. To achieve stationarity, we ran 5010000 iterations with a thinning interval of 1000 and a burn-in of 10000 steps for the model. To investigate the effects of methods (MA or PO) used to estimate mutation rate, we additionally included 'Method' as a second fixed effect in the above model.

We also investigated the relationships between mutation rate and generation time, population size, and genome size using univariate phylogenetic mixed models that included these factors separately as fixed effects, and a phylogenetic species term as a random effect. A log-transformation was made on these variables before fitting the model, and model predictions were made conditioning on each of the different species as random effect levels. The estimated generation time for each species was gathered from the literature (see Availability of data and materials) and we obtained estimates for 45 species. Genome sizes were taken to be assembly size, and gathered from the NCBI database,

which resulted in 63 species with genome size available. Estimates of census population size data were taken from Buffalo (2021), who used body length and species range size to infer the population size. As the estimation of population size is extremely challenging and subject to many uncertainties, we chose to use only this source to avoid methodological heterogeneity, which covered 25 species.

To test for a relationship between putatively neutral nucleotide diversity (π_S) and the SNM rate (μ) while accounting for uncertainty in both, we used MCMCglmm to fit a bivariate PGLMM with mutation number and π_S as responses. We collected nucleotide diversity data for 41 species either from (Buffalo 2021) or literature (see Availability of data and materials). Because we were interested in the gradient of regression of π_S on μ (on the logarithmic scale), we modelled the covariance between them using an antedependence structure that parameterises the covariance matrix in terms of a regression (see supporting material to Thomson, et al. (2017)). As before, we modelled the number of mutations as a Poisson distribution offset by the number of callable sites and included a phylogenetic term to account for non-independence among species in either SNM rate or π_S. The basic syntax for this model was:

```
biv_model <- MCMCglmm(y ~ trait + at.level(trait, 'mu'):log_callable,

            random = ~ante1(trait):Species2_phylo,

            rcov = ~idh(trait):units,

            data = d12,

            ginverse = list(Species2_phylo=InverseTree),

            prior = prior,

            nitt=5001000, thin=1000, burnin=10000,

            family = NULL

            )
```

In addition, to test for the potential role of generation time in driving this negative relationship, we fitted a final model that additionally included log generation time as a fixed effect predicting both SNM rate and π_S.

Availability of data and materials

The datasets and the code to analyse the data are available on GitHub: https://github.com/Yiguan/mutation_literature

Acknowledgements

We thank the authors of the many studies we analyse for making their data available. We thank Jarrod Hadfield for suggesting the statistical approach and helping with MCMCglmm syntax and interpretation.

Funding

This work was funded by the UK Biotechnology and Biological Sciences Research Council through grant BB/T007516/1. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Competing interests

The authors declare that they have no competing interests.

Supplementary figure legends

Supplementary Figure S1: The SNM rates comparison in *D. melanogaster* across different studies using MA or PO method to estimate mutation rates. 1a: line33; 1b: line39; 2a: West African population; 2b: European population.

Supplementary Figure S2: The SNM rates comparison in *M. musculus* across different studies using MA or PO method to estimate mutation rates.

Supplementary Figure S3: The ratio of deletions over insertions across all the identified studies.

Supplementary Figure S4: The relationship between genome size and the ratio of deletion/insertion without (A) or with (B) human and mammal groups. Log-transformed has been made on both variables.

481 **References**

482 Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes:
483 causes and consequences. Nature Reviews Genetics 8:619-631.
484 Bailey WJ, Fitch D, Tagle DA, Czelusniak J, Slightom JL, Goodman M. 1991. Molecular evolution of the
485 psi eta-globin gene locus: gibbon phylogeny and the hominoid slowdown. Molecular Biology
486 and Evolution 8:155-184.
487 Belfield EJ, Brown C, Ding ZJ, Chapman L, Luo M, Hinde E, Van Es SW, Johnson S, Ning Y, Zheng SJ.
488 2021. Thermal stress accelerates Arabidopsis thaliana mutation rate. Genome Research
489 31:40-50.
490 Bennetzen JL, Ma J, Devos KM. 2005. Mechanisms of recent genome size variation in flowering
491 plants. Annals of Botany 95:127-132.
492 Bergeron LA, Besenbacher S, Turner T, Versoza CJ, Wang RJ, Price AL, Armstrong E, Riera M, Carlson J,
493 Chen HY, et al. 2022. The Mutationathon highlights the importance of reaching
494 standardization in estimates of pedigree-based germline mutation rates. Elife 11.
495 Bourguignon T, Kinjo Y, Villa-Martin P, Coleman NV, Tang Q, Arab DA, Wang Z, Tokuda G, Hongoh Y,
496 Ohkuma M. 2020. Increased mutation rate is linked to genome reduction in prokaryotes.
497 Current Biology 30:3848-3855. e3844.
498 Buffalo V. 2021. Quantifying the relationship between genetic diversity and population size suggests
499 natural selection cannot explain Lewontin's Paradox. Elife 10:e67509.
500 Campbell CD, Eichler EE. 2013. Properties and rates of germline mutations in humans. Trends in
501 Genetics 29:575-584.
502 Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in
503 Drosophila melanogaster. Plos Genetics 8:e1003090.
504 Chintalapati M, Moorjani P. 2020. Evolution of the mutation rate across primates. Current Opinion in
505 Genetics & Development 62:58-64.
506 de Jong WW, Rydén L. 1981. Causes of more frequent deletions than insertions in mutations and
507 protein evolution. Nature 290:157-159.
508 Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of
509 insertions in the Caenorhabditis elegans nuclear genome. Nature 430:679-682.
510 Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proceedings of
511 the National Academy of Sciences 88:7160-7164.
512 Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics
513 148:1667-1686.
514 Drost JB, Lee WR. 1995. Biological basis of germline mutation: comparisons of spontaneous germline
515 mutation rates among drosophila, mouse, and human. Environmental and Molecular
516 Mutagenesis 25:48-64.
517 Gao Z, Wyman MJ, Sella G, Przeworski M. 2016. Interpreting the dependence of mutation rates on
518 age and time. Plos Biology 14:e1002355.
519 Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrion E,
520 Diallo O, et al. 2011. Increased exonic de novo mutation rate in individuals with
521 schizophrenia. Nat Genet 43:860-863.
522 Gregory TR. 2004. Insertion–deletion biases and the evolution of genome size. Gene 324:15-34.
523 Gregory TR. 2003. Is small indel bias a determinant of genome size? Trends in Genetics 19:485-488.
524 Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B, Keightley PD.
525 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in
526 Drosophila. Nature 445:82-85.
527 Hadfield J. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The
528 MCMCglmm R Package. Journal of Statistical Software 33:1-22.

529 Hadfield JD, Nakagawa S. 2010. General quantitative genetic methods for comparative biology:
530 phylogenies, taxonomies and multi-trait models for continuous and categorical characters.
531 Journal of Evolutionary Biology 23:494-508.
532 Harland C, Charlier C, Karim L, Cambisano N, Deckers M, Mni M, Mullaart E, Coppieters W, Georges
533 M. 2017. Frequency of mosaicism points towards mutation-prone early cleavage cell
534 divisions in cattle. bioRxiv:079863.
535 Harmon LJ, Pennell MW, Henao-Diaz LF, Rolland J, Sipley BN, Uyeda JC. 2021. Causes and
536 consequences of apparent timescaling across all estimated evolutionary rates. Annual
537 Review of Ecology, Evolution, and Systematics 52:587-609.
538 Hasan AR, Lachapelle J, El‐Shawa SA, Potjewyd R, Ford SA, Ness RW. 2022. Salt stress alters the
539 spectrum of de novo mutation available to selection during experimental adaptation of
540 Chlamydomonas reinhardtii. Evolution.
541 Ho EKH, Macrae F, Latta LC, McIlroy P, Ebert D, Fields PD, Benner MJ, Schaack S. 2020. High and
542 Highly Variable Spontaneous Mutation Rates in Daphnia. Molecular Biology and Evolution
543 37:3258-3266.
544 Ho SY, Duchêne S, Duchêne D. 2015. Simulating and detecting autocorrelation of molecular
545 evolutionary rates among lineages. Molecular Ecology Resources 15:688-696.
546 Huang W, Lyman RF, Lyman RA, Carbone MA, Harbison ST, Magwire MM, Mackay TFC. 2016.
547 Spontaneous mutations and the origin and maintenance of quantitative genetic variation.
548 Elife 5.
549 Kaplanis J, Ide B, Sanghvi R, Neville M, Danecek P, Coorens T, Prigmore E, Short P, Gallone G, McRae
550 J. 2022. Genetic and chemotherapeutic influences on germline hypermutation. Nature
551 605:503-508.
552 Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals.
553 Proceedings of the National Academy of Sciences 114:E1460-E1469.
554 Keightley PD, Eyre-Walker A. 2000. Deleterious mutations and the evolution of sex. Science 290:331-
555 333.
556 Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the Spontaneous Mutation Rate
557 per Nucleotide Site in a Drosophila melanogaster Full-Sib Family. Genetics 196:313-+.
558 Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the genome
559 sequences of three Drosophila melanogaster spontaneous mutation accumulation lines.
560 Genome Research 19:1195-1201.
561 Keith N, Jackson CE, Glaholt SP, Young K, Lynch M, Shaw JR. 2021. Genome-Wide Analysis of
562 Cadmium-Induced, Germline Mutations in a Long-Term Daphnia pulex Mutation-
563 Accumulation Experiment. Environmental Health Perspectives 129.
564 Kimura M. 1968. Evolutionary rate at the molecular level. Nature 217:624-626.
565 Kondrashov AS. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing
566 Mendelian diseases. Human Mutation 21:12-27.
567 Kondrashov AS, Crow JF. 1993. A molecular approach to estimating the human deleterious mutation
568 rate. Human Mutation 2:229-234.
569 Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A,
570 Jonasdottir A, Jonasdottir A, et al. 2012. Rate of de novo mutations and the importance of
571 father's age to disease risk. Nature 488:471-475.
572 Konrad A, Brady MJ, Bergthorsson U, Katju V. 2019. Mutational Landscape of Spontaneous Base
573 Substitutions and Small Indels in Experimental Caenorhabditis elegans Populations of
574 Differing Size. Genetics 212:837-854.
575 Krasovec M, Rickaby RE, Filatov DA. 2020. Evolution of mutation rate in astronomically large
576 phytoplankton populations. Genome biology and evolution.
577 Kucukyildirim S, Behringer M, Sung W, Brock DA, Doak TG, Mergen H, Queller DC, Strassmann JE,
578 Lynch M. 2020. Low base-substitution mutation rate but high rate of slippage mutations in

579 the sequence repeat-rich genome of Dictyostelium discoideum. G3: Genes, Genomes,
580 Genetics 10:3445-3452.
581 Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022.
582 TimeTree 5: An expanded resource for species divergence times. Molecular Biology and
583 Evolution 39:msac174.
584 Lanfear R, Kokko H, Eyre-Walker A. 2014. Population size and the rate of evolution. Trends in Ecology
585 & Evolution 29:33-41.
586 Li W-H, Tanimura M, Sharp PM. 1987. An evaluation of the molecular clock hypothesis using
587 mammalian DNA sequences. Journal of Molecular Evolution 25:330-342.
588 Liu H, Zhang J. 2021. The rate and molecular spectrum of mutation are selectively maintained in
589 yeast. Nature Communications 12:1-11.
590 Long H, Winter DJ, Chang AY-C, Sung W, Wu SH, Balboa M, Azevedo RB, Cartwright RA, Lynch M,
591 Zufall RA. 2016. Low base-substitution mutation rate in the germline genome of the ciliate
592 Tetrahymena thermophila. Genome biology and evolution 8:3629-3639.
593 Lopez-Cortegano E, Craig RJ, Chebib J, Samuels T, Morgan AD, Kraemer SA, Bondel KB, Ness RW,
594 Colegrave N, Keightley PD. 2021. De Novo Mutation Rate Variation and Its Determinants in
595 Chlamydomonas. Molecular Biology and Evolution 38:3709-3723.
596 Lynch M. 2010. Evolution of the mutation rate. Trends in Genetics 26:345-352.
597 Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift,
598 selection and the evolution of the mutation rate. Nature Reviews Genetics 17:704-714.
599 Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits: Sinauer Sunderland, MA.
600 Marais GA, Batut B, Daubin V. 2020. Genome evolution: mutation is the main driver of genome size
601 in prokaryotes. Current Biology 30:R1083-R1085.
602 Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock.
603 Proceedings of the National Academy of Sciences 90:4087-4091.
604 Meier B, Cooke SL, Weiss J, Bailly AP, Alexandrov LB, Marshall J, Raine K, Maddison M, Anderson E,
605 Stratton MR, et al. 2014. C. elegans whole-genome sequencing reveals mutational signatures
606 related to carcinogens and DNA repair deficiency. Genome Research 24:1624-1636.
607 Mukai T, Cockerham CC. 1977. Spontaneous mutation rates at enzyme loci in Drosophila
608 melanogaster. Proceedings of the National Academy of Sciences 74:2514-2517.
609 Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics
610 156:297-304.
611 Ness RW, Morgan AD, Colegrave N, Keightley PD. 2012. Estimate of the Spontaneous Mutation Rate
612 in Chlamydomonas reinhardtii. Genetics 192:1447-+.
613 Nute PE, Stamatoyannopoulos G. 1984. Estimating Mutation-Rates Using Abnormal Human
614 Hemoglobins. Yearbook of Physical Anthropology 27:135-151.
615 Obbard DJ, Maclennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating Divergence
616 Dates and Substitution Rates in the Drosophila Phylogeny. Molecular Biology and Evolution
617 29:3459-3473.
618 Ohno M. 2019. Spontaneous de novo germline mutations in humans and mice: rates, spectra, causes
619 and consequences. Genes & Genetic Systems 94:13-22.
620 Peck KM, Lauring AS. 2018. Complexities of viral mutation rates. Journal of Virology 92:e01031-
621 01017.
622 Petrov DA. 2002. Mutational equilibrium model of genome size evolution. Theoretical Population
623 Biology 61:531-544.
624 Saxena AS, Salomon MP, Matsuba C, Yeh SD, Baer CF. 2019. Evolution of the Mutational Process
625 under Relaxed Selection in Caenorhabditis elegans. Molecular Biology and Evolution 36:239-
626 251.
627 Schrider DR, Houle D, Lynch M, Hahn MW. 2013. Rates and Genomic Consequences of Spontaneous
628 Mutational Events in Drosophila melanogaster. Genetics 194:937-+.

Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P. 2000. The mutation rate in the human mtDNA control region. The American Journal of Human Genetics 66:1599-1609.
Stamatoyannopoulos G, Nute PE. 1982. De novo mutations producing unstable Hbs or Hbs M. II. Direct estimates of minimum nucleotide mutation rates in man. Human Genetics 60:181-188.
Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos Trans R Soc Lond B Biol Sci 372:20160455.
Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012a. Drift-barrier hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences 109:18488-18492.
Sung W, Tucker AE, Doak TG, Choi E, Thomas WK, Lynch M. 2012b. Extraordinary genome stability in the ciliate Paramecium tetraurelia. Proceedings of the National Academy of Sciences 109:19339-19344.
Thomas GW, Hahn MW. 2014. The human mutation rate is increasing, even as it slows. Molecular Biology and Evolution 31:253-257.
Thomas GW, Wang RJ, Puri A, Harris RA, Raveendran M, Hughes DS, Murali SC, Williams LE, Doddapaneni H, Muzny DM. 2018. Reproductive longevity predicts mutation rates in primates. Current Biology 28:3193-3197. e3195.
Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, et al. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424:788-793.
Thomson CE, Bayer F, Crouch N, Farrell S, Heap E, Mittell E, Zurita‐Cassinello M, Hadfield JD. 2017. Selection on parental performance opposes selection for larger body mass in a wild population of blue tits. Evolution 71:716-732.
Trimble BK, Doughty JH. 1974. The amount of hereditary disease in human populations. Annals of Human Genetics 38:199-223.
Waldvogel A-M, Pfenninger M. 2021. Temperature dependence of spontaneous mutation rates. Genome Research 31:1582-1589.
Wang L, Ji Y, Hu Y, Hu H, Jia X, Jiang M, Zhang X, Zhao L, Zhang Y, Jia Y. 2019. The architecture of intra-organism mutation rate variation in plants. Plos Biology 17:e3000191.
Wang RJ, Peña-Garcia Y, Bibby MG, Raveendran M, Harris RA, Jansen HT, Robbins CT, Rogers J, Kelley JL, Hahn MW. 2022a. Examining the effects of hibernation on germline mutation rates in grizzly bears. Genome biology and evolution.
Wang RJ, Thomas GW, Raveendran M, Harris RA, Doddapaneni H, Muzny DM, Capitanio JP, Radivojac P, Rogers J, Hahn MW. 2020. Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. Genome Research 30:826-834.
Wang Y, Allen SL, Reddiex AJ, Chenoweth S. 2022b. Direct and indirect impacts of positive selection on genomic variation in Drosophila serrata. bioRxiv.
Wang Y, McNeil P, Abdulazeez R, Pascual M, Johnston SE, Keightley PD, Obbard DJ. 2022c. Variation in mutation, recombination, and transposition rates in Drosophila melanogaster and Drosophila simulans. bioRxiv.
Wilding C. 2017. Genetic diversity of the African malaria vector Anopheles gambiae. Nature 552:96-100.
Wu C-I, Li W-H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proceedings of the National Academy of Sciences 82:1741-1745.
Wu F, Przeworski M. 2022. A paternal bias in germline mutation is widespread in amniotes and can arise independently of cell division numbers. Elife 11:e80008-e80008.
Wu FL, Strand AI, Cox LA, Ober C, Wall JD, Moorjani P, Przeworski M. 2020. A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. Plos Biology 18:e3000838.

679 Xu S, Stapley J, Gablenz S, Boyer J, Appenroth KJ, Sree KS, Gershenzon J, Widmer A, Huber M. 2019.
680 Low genetic variation is associated with low mutation rate in the giant duckweed. Nature
681 Communications 10:1-6.
682 Yang SH, Wang L, Huang J, Zhang XH, Yuan Y, Chen JQ, Hurst LD, Tian DC. 2015. Parent-progeny
683 sequencing indicates higher mutation rates in heterozygotes. Nature 523:463-U187.
684 Yoder AD, Tiley GP. 2021. The challenge and promise of estimating the de novo mutation rate from
685 whole-genome comparisons among closely related individuals. Molecular Ecology 30:6087-
686 6100.

687