Article Type: Letter

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

Short Running Title: mutation rates in eukaryotes

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.

Abstract Word Count: 280

Total Word Count: ~5200

Abstract

Mutation is the ultimate source of all genetic variation, and over the last ten years the ready availability of whole-genome sequencing has permitted direct estimation of mutation rate for many non-model species across the tree of life. In this meta-analysis we make a comprehensive search of the literature for mutation rate estimates in eukaryotes, identifying 140 mutation accumulation (MA) and parent-offspring (PO) sequencing studies covering 134 species. Based on these data, we revisit differences in single nucleotide mutation (SNM) rate between different phylogenetic lineages and update the known relationships between mutation rate and generation time, genome size, and 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 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 nucleotide mutation 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 amphibia and scarce for reptiles and fish.

Lay Summary

Over the past decade, the sequencing revolution has led to an ever-increasing number of mutation-rate estimates from mutation accumulation or parent-offspring sequencing studies in eukaryotes. However, studies rarely quantify to what extent the mutation rate varies among these species. Also, despite strong predictions as to how mutation rate should vary with (e.g.) generation time, there have been few recent or wide-ranging analyses of such predictors while accounting for the inherent similarity between closely-related species. Of particular note, there has been surprisingly little effort to robustly test the 'drift barrier' hypothesis that mutation rates should decrease with increasing effective population size. In this study, we used a comprehensive literature search to identify all the available experimental estimates of mutation rate in eukaryotes and subject them to phylogenetic mixed-model analyses. We find that per-nucleotide per-generation mutation rates differ by orders of magnitude among species: plants and mammals tend to have higher mutation rates than arthropods, and unicellular organisms have the lowest mutation rates. Our analysis also shows that mutation rates increase significantly with increasing generation time and genome size, and nucleotide diversity increases with mutation rate with a gradient less than one—as predicted by the drift-barrier hypothesis.

72 **Introduction**

73 The per-nucleotide per-generation *de novo* mutation rate, μ, is a key parameter in
74 population and evolutionary genetics, appearing either alone or combined with effective
75 population size (N_e) in the compound parameter θ. However, it seems implicit that μ is of no
76 interest to many evolutionary studies; researchers often use μ from one species when
77 analysing another—as if μ were exchangeable in a way that N_e is not (e.g. Wilding, 2017).
78 This is partly because the historically high cost of sequencing made estimates prohibitively
79 expensive, but may also be because variation in μ among species is widely thought to be
80 negligible compared to variation in N_e. But, if the mutation rate varied as widely among
81 multicellular eukaryotes as many other traits, there would be much less need to invoke
82 variation in N_e when explaining variation in (e.g.) genetic diversity. One reason why μ may
83 vary relatively little is if it is minimised to a lower limit imposed by drift: the 'drift barrier'
84 hypothesis (Lynch, 2008, 2010). However, it might be that in reality mutation rate is
85 optimised to well above the drift-barrier limit by selection in favour of adaptability or the
86 cost of replication fidelity (Kimura, 1967; Liu et al., 2021; Peck et al., 2018). If this were the
87 case, then we might find that mutation rates varied dramatically, even among closely related
88 taxa.

89 Early estimates of the mutation rate leveraged visible phenotypes, such as the occurrence of
90 genetic diseases in humans (Nute et al., 1984; Trimble et al., 1974). For example, using
91 haemoglobin M disease Stamatoyannopoulos et al. (1982) estimated the mutation rate in
92 the human haemoglobin beta gene at 7.4×10^{-9} per bp per generation. However, such
93 estimates are prone to bias, for example if the mutation is required to be autosomal

94 dominant and non-lethal (Nute et al., 1984), and such approaches are not easily applicable in
95 organisms for which disease is hard to observe. More commonly, using the equivalence
96 between mutation and neutral substitution rates (Kimura, 1968), many studies have used
97 time-calibrated sequence divergence to estimate μ. For example, comparison between
98 humans and chimpanzees gives rise to estimates on the order of 10^{-8} per bp per generation
99 (Drake et al., 1998; Kondrashov et al., 1993; Nachman et al., 2000). However, such indirect
100 estimates are also limited (Kondrashov, 2003). First, because generation time and calibration
101 dates are prone to substantial uncertainty (e.g. Obbard et al., 2012). Second, because it is
102 hard to distinguish between unconstrained and weakly constrained sites (Harmon et al.,
103 2021). Third, because mutation saturation in hotspots may lead to an underestimation,
104 especially for sequences with a long divergence time (Sigurðardóttir et al., 2000). In addition,
105 phylogenetic-calibration approaches reflect the long-term mutation rate, averaging across
106 biologically important factors such as generation time, sex differences, life-stage differences,
107 and inter-individual variation.

108 In contrast to phenotypic and phylogenetic approaches, mutation-accumulation (MA) and
109 parent-offspring (PO) sequencing can provide direct estimates of the per-generation
110 mutation rate. Mutation-accumulation utilises the accumulation of spontaneous mutations
111 in a single inbred or asexual genome over multiple generations, by minimising the
112 effectiveness of natural selection (Halligan et al., 2009). Given the low mutation rates in
113 eukaryotes, the multiple generations in an MA experiment allow sufficient mutations for
114 them to be detected at the end. Although the first use of MA to estimate μ phenotypically
115 dates back to a study by Mukai et al. (1977), the availability of whole genome sequencing
116 and its decreasing price have made sequencing approaches more widely applicable.

117 Mutation accumulation best suits easily-maintained species with short generation times, for
118 example nematodes (Denver et al., 2004; Konrad et al., 2019; Meier et al., 2014; Saxena et
119 al., 2019), fruit flies (Haag-Liautard et al., 2007; Huang et al., 2016; Keightley et al., 2009;
120 Schrider et al., 2013), green algae (Lopez-Cortegano et al., 2021; Ness et al., 2012), and
121 water fleas (Ho et al., 2020; Keith et al., 2021). However, MA may underestimate mutation
122 rates (Lynch et al., 1998) because natural selection will still come into play for mutations that
123 are very highly deleterious, such as those that cause complete sterility or are lethal (Baer et
124 al., 2007).

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

139 Relaxed-clock phylogenetic studies generally indicate there is substantial among-lineage
140 variation in the evolutionary rate, suggestive of variation in mutation rates among taxa (Ho
141 et al., 2015). But such studies cannot separate the impact of generation time and mutation
142 rate, and may be biased by variation in the action of (weak) selection (Harmon et al., 2021).
143 As more direct experimental estimates have become available, it has become possible to
144 draw comparisons within particular clades such as primates (Chintalapati et al., 2020), or
145 vertebrates (Bergeron et al., 2022; Bergeron et al., 2023; Yoder et al., 2021). However, there
146 has not been a wider analysis, nor one that includes both PO and MA studies.

147 In addition, many previous studies have not taken an explicitly phylogenetic analytical
148 approach (e.g. Bergeron et al., 2022; Katju et al., 2019; Yoder et al., 2021). This could in
149 principle lead to pseudoreplication in the analyses that is caused by the non-independence
150 among taxa that arises from their shared ancestry (reviewed in Freckleton, 2009). However,
151 this problem can be mitigated by explicitly modelling covariance among species that is
152 contributed by their patterns of relationship, and generalised linear mixed modelling
153 approaches are available to do this (Hadfield et al., 2010; Halliwell et al., 2022; Lynch, 1991).

154 Here we perform a comprehensive phylogenetic meta-analysis of published eukaryotic single
155 nucleotide mutation (SNM) and insertion-deletion (indel) mutation-rate estimates that use
156 direct sequencing of multiple generations, either through MA or PO comparison. We focus
157 on the variation in mutation rate among species, and the relationship between mutation
158 rate estimates and the experimental method, generation time, genome size, and genetic
159 diversity.

160

161 **Results and Discussion**

162 ***Studies included***

163 We searched for eukaryotic mutation-rate studies in Clarivate Web of Science published
164 prior to 21st September 2022, and then we manually examined these studies and identified
165 those containing mutation rate estimates (see Methods). For the majority of our analyses we
166 also included data from one further recent study of 68 vertebrate species by Bergeron et al.
167 (2023), published after our initial submission. In total, we included 140 studies using either
168 MA or PO approaches, covering 134 species (see Availability of data and materials). Studies
169 on primate mutation rate accounted for over a third of this literature, and 82% (45/55) of
170 the primate studies were on humans. Arthropods accounted for 17% of the identified
171 literature, followed by fungi (11%), plants (8%), non-primate mammals (7%), and nematodes
172 (5%) (**Fig. 1A)**. We only identified four studies on fish, two studies on birds and one study on
173 reptiles. Across all groups, most studies focused on model species: yeast in fungi, *Drosophila*
174 in arthropods, mice in mammals, and *Arabidopsis* in plants. These estimates tended to
175 derive from MA analyses rather than PO studies (61 vs. 43 among non-human studies).
176 However, as sequencing has become cheaper, there has been a rapid increase in the number
177 of published experimental estimates of the mutation rate, with an average of 11 studies
178 each year in eukaryotic species between 2010 and 2022 (**Fig. 1B**).

179

180 ***Single nucleotide mutation rates across eukaryotes***

181 To summarise the single nucleotide mutation rates in different lineages, we obtained a time-
182 scaled phylogenetic tree of all represented species from http://www.timetree.org/ (Kumar
183 et al., 2022), and used this to fit the covariance among species in a Phylogenetic Generalised
184 Linear Mixed Model (PGLMM) using the Bayesian mixed model package MCMCglmm
185 (Hadfield et al., 2010). Treating species as a random effect, this provides a posterior estimate
186 of mutation rate for each species, given the observations and relationships to other species
187 (see Methods). In these data, the phylogenetic effect accounts for 96.4% of the random-
188 effects variance, with a 95% HPD interval of 94.9% to 97.9%.

189 Among all the studied species, a ciliate (*Tetrahymena thermophila*) had the lowest mutation
190 rate, estimated at 0.01×10^{-9} per generation per base pair, while a fungus (*Marasmius*
191 *oreades*) had the highest rate, estimated at 55.58×10^{-9}—a difference of over five thousand-
192 fold. The estimates for unicellular species were generally more than an order of magnitude
193 lower than rates in multicellular species (below; **Fig. 2**), even compared to species in the
194 same clade, e.g., fungi. This may either be attributable to there being a single cell division
195 per generation, or possibly to a limit set by the drift barrier in species with larger N_e (Sung et
196 al., 2012). We also estimated relatively low SNM rates for nematodes at 2.44×10^{-9},
197 arthropods at 2.82×10^{-9}, and fish at 5.55×10^{-9}. For plants, the clade root was predicted as
198 3.66×10^{-9}, but a large difference was observed between duckweed and other plant species
199 (0.22×10^{-9} vs. 19.8×10^{-9}). The high mutation rates in most plant species might be
200 attributable to the lack of a segregated germline (Burian, 2021; Cruzan et al., 2022; Lanfear,
201 2018; Wang et al., 2019). Mammals had moderate mutation rates with a root estimate at
202 8.36×10^{-9} and species (excluding primates) ranging from 5.64×10^{-9} in pigs to 11.68×10^{-9} in
203 killer whales. Primates had higher mutation rates, with a root estimate at 9.37×10^{-9}. Of

204 note, as the most studied species, human mutation rate was estimated as 13.41×10^{-9} with
205 95% CI 11.61 to 15.28×10^{-9}, which is also the highest among mammals. Similar to mammals,
206 birds and reptiles had high mutation rate, with root estimates at 10.43×10^{-9} and $12.41 \times 10^{-}$
207 ^{9}, respectively.

208 To assess a possible impact of the experimental method on rate estimates, we included the
209 experimental approach (MA or PO) as a fixed effect in the PGLMM. Overall, we did not find a
210 significant difference between the two methods (MCMC $P = 0.155$). We then examined
211 comparisons between MA and PO in fruit flies and mice more closely, as these two species
212 have been investigated using both methods. The rate estimated using MA in fruit flies was
213 5.28×10^{-9}, but 3.20×10^{-9} using PO (**Fig. S1**), and in mice the MA estimate was 5.40×10^{-9}
214 while the PO estimate was 4.05×10^{-9} (**Fig. S2**), and these differences were similarly not
215 significant. Note, however, that the mutation rate may be heterogeneous among different
216 populations: for *D. melanogaste*r, the estimated mutation rate from African populations was
217 lower than that from European population and North American population (Chan et al., 2012;
218 Wang et al., 2023). To make a more robust analysis would require samples of the same
219 population but using different methods (MA or PO), which are not yet available.

220

221 ***Biological predictors of the SNM rate***

222 Many biological factors have been proposed to be associated with mutation rate, including
223 generation time (Bailey et al., 1991; Keightley et al., 2000; Li et al., 1987; Wu et al., 1985)
224 and genome size (Drake, 1991; Lynch, 2010). To test for a dependence of mutation rate on
225 these factors, we ran three further models: the first two including these factors each

226 separately as an additional fixed effect in a PGLMM, and the last one including both of them
227 simultaneously.

228 We found a significant positive relationship with generation time (Bayesian PGLMM MCMC *P*
229 < 0.0001; **Fig. 3A**), such that a longer generation time predicted a higher per-site mutation
230 rate per generation. If mutations resulted only from replication errors, and most species had
231 a similar number of cell divisions per generation, then the per-generation mutation rate
232 would not scale with generation time (Thomas et al., 2014; Wu et al., 1985). However,
233 although few estimates are available, in reality the number of germline cell divisions per
234 generation varies within and between species. In humans, there are around 401 cell
235 divisions per 30-year generation in males and 31 in females (Drost et al., 1995; Ohno, 2019),
236 but in mice it is 62 per 9-month generation in males and 25 in females. In *Drosophila* the
237 sexes are much more similar, with an estimated 35.5 cell divisions in a 18-day generation for
238 males but 36.5 in an 25-day generation in females in *D. melanogaster* (Drost et al., 1995).
239 Within a species, a long generation time is likely to permit more cell divisions and thus a
240 higher mutation rate per generation, and many studies in humans and primates have shown
241 that mutations are heavily male-biased and correlated with paternal ages (Kaplanis et al.,
242 2022; Kong et al., 2012; Thomas et al., 2018; Wang et al., 2020).

243 Alternatively, mutations may not result predominantly from replication errors. Recent
244 studies suggest that not all germline mutations track cell divisions, and DNA damage may
245 contribute to the mutation rate (Wu et al., 2022; Wu et al., 2020). Independent of cell
246 divisions, this mutation rate would be correlated with calendar time, and thus species with a
247 longer generation time would have a higher rate. The relationship we observe between

248 generation time and per-generation mutation rate could therefore be a consequence of
249 either a greater number of cell divisions or of accumulating damage over time, and it will be
250 hard to separate these potential causes given the limited data available on the number of
251 germline cell divisions for different species (Gao et al., 2016; Wu et al., 2022).

252 We also found evidence that larger genome size predicts a higher per-generation mutation
253 rate (MCMC $P = 0.0020$; **Fig. 3B**). Mutation rate has previously been found to correlate
254 negatively with genome size in prokaryotes, a phenomenon termed "Drake's rule"
255 (Bourguignon et al., 2020; Drake, 1991; Marais et al., 2020), but positively in eukaryotes
256 (Lynch, 2010).

257 In each of the two models above, we fitted generation time and genome size individually as
258 predictors of mutation rate in a regression model. However, these factors might be highly
259 correlated: large organisms tend to have longer generation times and often larger genomes
260 (Buffalo, 2021; Martin et al., 1993), and is therefore hard to distinguish between casual
261 relationships and confounding correlations. Including generation time and genome size
262 together as fixed terms in a single model resulted in only generation time being significant
263 with MCMC $P < 0.001$ (versus MCMC $P = 0.301$ for genome size).

264 ***The drift-barrier hypothesis***

265 The most direct test of the drift-barrier hypothesis would be a regression of SNM rate on N_e.
266 However, long-term N_e can only be estimated from genetic diversity (i.e. π) by making use of
267 the relationship $\theta_\pi = 4N_e\mu$, and any uncertainty in μ could induce a spurious correlation in a
268 regression of μ on $\pi/4\mu$. Instead, error in both π and μ can be accounted for by fitting a
269 bivariate linear mixed model, parameterised in terms of a regression of π on μ (see

270 Methods). Then, on a logarithmic scale, the gradient of this relationship would be expected
271 to be 1 if μ and N_e are independent of each other, and lower than 1 if species with large N_e
272 tended to have lower μ. Nevertheless, such a relationship could be driven by other factors.
273 For example, if species with small N_e tend to have a longer generation time, and a longer
274 generation time causes higher mutation rates (i.e. **Fig. 3A**), then a higher μ in species with
275 low N_e could be driven by a mechanistic generation-time effect, rather than the drift barrier.
276 We attempted to account for this by including generation time in our analysis as a predictor
277 of both μ and π, thereby accounting for the direct impact of generation time on the
278 relationship between them. In this bivariate phylogenetic model, μ increased with
279 generation time (MCMC $P < 0.001$) whereas π did not (MCMC $P = 0.501$), and the gradient of
280 the regression of π on μ was significantly less than 1 (gradient = 0.535, 95% CI [0.259, 0.832];
281 MCMC $P = 0.004$). This suggests that populations with larger N_e tend to have a lower
282 mutation rate even after accounting for their shorter generation times—as predicted by the
283 drift-barrier hypothesis (Lynch, 2010). Note that excluding generation time as a predictor
284 resulted in an even stronger effect (gradient = 0.300 [0.098, 0.486]; MCMC $P < 0.001$; **Fig.**
285 **3C**), whereas removing unicellular species from the analysis (because of their low mutation
286 rates) resulted in a gradient that was still much less than 1 (gradient = 0.439 [0.155, 0.746];
287 MCMC $P < 0.001$).

288

289 ***The mutation rates of short indels across eukaryotes***

290 Although the rate of indel mutation is much less commonly estimated than the SNM rate,
291 our literature search identified estimates for 29 species. We found that the indel mutation

292 rate across eukaryotes displayed a similar pattern to that of the SNM rate (**Fig. 4**), with a
293 correlation coefficient between the two of 0.75 (95% HPD CI: 0.44, 0.98; MCMC $P < 0.001$)
294 using PGLMM, with the SNM rate being generally higher than the indel mutation rate,
295 except in two lineages (nematodes and amoebae) for which the indel mutation rate was
296 apparently higher (**Fig. 5A**). However, the mutations in the nematode study (Denver et al.,
297 2004) were specifically collected from mononucleotide repeats, which may artificially inflate
298 the indel mutation rate estimate in that study (Konrad et al., 2019; Rajaei et al., 2021). The
299 unicellular species had the lowest indel mutation rate, from 0.02×10^{-9} in yeast to 0.26×10^{-9}
300 in algae, followed by arthropods with 0.55 (0.08, 1.24) $\times 10^{-9}$. Mammals had a relatively
301 higher indel mutation rate with 1.13 (0.43, 2.05) $\times 10^{-9}$, with humans at the higher end of
302 this range at 1.74 (1.16, 2.50) $\times 10^{-9}$. Of all the studied species, plants had the highest indel
303 mutation rate, which was as high as 2.38 (0.51, 4.96) $\times 10^{-9}$.

304 Short deletions generally outnumbered short insertions across eukaryotes (**Fig. 5B, Fig.S3**).
305 The ratio of deletions over insertions was as high as 15.0 in the water flea, 6.0 in mouse, 3.4
306 in *D. melanogaster*, 2.6 in humans, 1.8 in *C. elegans*, and 1.6 in *Arabidopsis*. Honeybees,
307 diatoms and two plants (rice and peach) exhibited more insertions than deletions, but the
308 difference wasn't statistically significant ($P > 0.05$, binomial test). Fission yeast showed
309 significantly more insertions overall.

310 A bias toward deletion events has been known for at least 40 years, since de Jong et al.
311 (1981) found a four-fold excess of deletions over insertions in protein sequences. Petrov
312 (2002) argued that the deletion bias was likely due to a thermodynamic disparity between
313 short deletions and insertions that makes deletion events easier. However, studies in plants

314 have suggested that unequal homologous recombination and illegitimate recombination
315 may be another important cause of small deletions (Bennetzen et al., 2005). Whatever the
316 cause, the dominance of deletions over insertions seems convincing, and many early studies
317 argued that these short deletions resulted in genome contraction (Gregory, 2003; Petrov,
318 2002; Thomas et al., 2003). However, the role of deletions in shaping genome size remains
319 controversial (Gregory, 2003, 2004). A recent study in birds and mammals indicated that
320 short deletions alone cannot explain the DNA content loss (Kapusta et al., 2017), and we fail
321 to find any evidence of a relationship between deletion rate and genome size here (**Fig. S4;**
322 but note that power is likely to be low). It seems probable that other structural mutations,
323 such as large deletions, gene duplication and transposable elements, may have greater
324 impacts on genome size. For example, in *Drosophila* Wang et al. (2023) found transposable
325 element insertions to be more frequent than SNMs, and an order of magnitude more
326 frequent than short indels. Together, these studies suggest that the impact of short
327 deletions on genome size is likely to be small, despite their higher frequency relative to short
328 insertions.

329

330 **Concluding remarks**

331 All genetic variation arises from *de novo* mutations. To estimate the rate at which these arise,
332 studies have moved from investigating a few genes that result in phenotypic change, to
333 whole genome sequencing of MA lines or families. Ten years ago, Campbell et al. (2013)
334 called for more efforts to sequence genomes from non-human primate families to
335 understand 'how the rate has changed in different lineages'. And, in the present data-rich

336 and tool-rich era we have now seen the sequencing of not just multiple non-human primates,
337 but many taxa representing major clades across the tree of life. Nevertheless, the 'tree of
338 mutations' is far from complete, at either the large or small scale. Even within animals, the
339 vast majority of phyla have never been examined, and with the exception of primates, no
340 large groups of very close relatives have been examined. Even for vertebrates—and despite
341 of the recent large study by Bergeron et al. (2023) who examined 68 species–data are still
342 scarce for reptiles and entirely lacking for amphibians.

343 However, as progress is being made toward a tree of mutation rates, attention still needs to
344 focus on methodological heterogeneity. Different data filtering strategies may result in non-
345 negligible differences in mutation rate estimates (Bergeron et al., 2022). Filters that are too
346 conservative may discard some true mutations, while relaxed filters may inflate the false
347 positive rate. Although many strategies have been proposed (Bergeron et al., 2022), no
348 standardised pipeline has yet been agreed upon or widely adopted. All studies must
349 therefore carefully address the issues of false positives by manual curation and/or PCR
350 follow-up, and false negatives by simulation. Here, we did not examine the impact of the
351 false positive rate or false negative rate, but we remind the readers of the potential impact
352 from methodological difference, and caution that attention should be paid to the methods
353 or parameters used in each study when comparisons are made.

354 Another potential source of bias may also come from the specific choice of samples used in
355 each study. Beside the biological factors that affect mutation rates (e.g. age, sex),
356 environmental factors can also influence mutation rates, for example temperature (Belfield
357 et al., 2021; Waldvogel et al., 2021), osmotic stress (Hasan et al., 2022), or ultraviolet light

358 (Xu et al., 2019). Samples collected from the wild rather than reared in laboratories may
359 better reflect the 'real' mutation rate, but the variation that this introduces may also make it
360 harder to draw broad conclusions. Despite these challenges, the data to come are exciting.
361 In particular, as sequencing technology advances, we expect improved accuracy in long-read
362 sequencing facilitating the phasing of mutations and permitting the easy detection of
363 complex mutations and larger-scale structural changes.

364

365 **Methods**

366 We made an exhaustive literature search for publications on the experimental estimation of
367 mutation rate. We first searched Clarivate Web of Science up to the 21st September 2022 for
368 'Title/Keywords/Abstract' fields containing ("mutation rate" | "mutation rates" |
369 "mutational rate" | "mutational rates"). As much of the resulting literature was related to
370 somatic mutations (e.g., cancers) or virus mutations, we excluded '("tumor" | "cancer" |
371 "clinical" | "virus")' in the search field of 'Title/Keywords/Abstract'. This search strategy
372 resulted in 9462 studies with the earliest one dated 1928. We also searched for references
373 to mutation rate estimates in other papers not captured by our search. As bacteria are
374 outside of our study, we manually removed studies related to bacterial mutations. These
375 filters resulted in 174 studies. Of these studies, we further excluded 20 studies that used
376 phylogenetic approaches to infer mutation rates, 14 studies that used phenotypic markers,
377 and two other studies that used recombination landscape or the site frequency spectrum to
378 infer mutation rates statistically. In addition, we included one more recent study published
379 after our initial submission (Bergeron et al., 2023), which estimated the mutation rates for

380 68 vertebrates using PO sequencing. This resulted in a total of 140 studies in our final
381 dataset.

382 From each of the potential studies we extracted the year of the publication, the species, the
383 population, reproductive mechanism (sexual, asexual), mutation type (SNM, short insertion,
384 short deletion), the units of mutation rate (per bp per year, per bp per generation), the
385 mean and 95% CI of mutation rate, the number of identified mutation events, the number of
386 callable sites, the methods used to infer mutation rate (PO, MA or others), the sample size,
387 the number of generations (if MA), the sequencing technology, the sequencing depth and
388 other pertinent observations. Note that for multi-generation studies, such as MA, the
389 number of 'callable sites' incorporates the possibility that a mutation may have arisen in any
390 of several generations. If a 95% confidence or credible interval was not provided in the
391 original study, we calculated a 95% confidence interval using 'binconf()' from the Hmisc R
392 package (version 4.7-0), based on the number of mutation events and callable sites and
393 assuming a binomial distribution (Girard et al., 2011). If a study did not provide the mutation
394 count, we inferred an effective number of mutation events based on the reported mutation
395 rate and its confidence interval. For studies in which the number of callable sites was not
396 reported, an estimate was made from the number of mutation events and the reported
397 mutation rate. If a study *only* reported the mutation rate, we assumed one mutation event
398 and calculated the corresponding callable sites based on mutation rates. This latter approach
399 will necessarily underestimate the precision with which rates are estimated (minimising their
400 weight in the analysis), while keeping the estimate in the meta-analysis with unchanged
401 mean values.

402 To compare the SNM rates across different species, we conducted a phylogenetic meta-
403 analysis (i.e., a combined analysis of published data) using the mutation data from MA or PO
404 studies. We obtained the phylogenetic relationships and relative divergence dates from
405 http://www.timetree.org/ (Kumar et al., 2022) and visualised using ggtree (Yu, 2020). We
406 performed the analyses using a phylogenetic generalized linear mixed model with a unit-
407 scaled phylogeny as random effect, assuming a Poisson distribution for the occurrence of *de*
408 *novo* mutations, off-set by the number of callable sites (i.e. enforcing a directly proportional
409 relationship between the expected number of mutations and the number of sites observed).
410 The analysis was conducted using the Bayesian mixed-model R-package MCMCglmm
411 (version 2.32) (Hadfield, 2010) (see **Supporting Materials**). To investigate the effects of
412 methods (MA or PO) used to estimate mutation rate, we additionally included 'Method' as
413 an additional fixed effect in the above model.

414 We also investigated the relationships between mutation rate and generation time and
415 genome size using univariate phylogenetic mixed models that included these factors
416 separately as fixed effects, and a phylogenetic species term as a random effect. A log-
417 transformation was made on these predictors before fitting the model (see **Supporting**
418 **Materials**), and model predictions were made conditioning on each of the different species
419 as random effect levels. The estimated generation time for each species was gathered from
420 the literature (see Availability of data and materials) and we obtained estimates for 116
421 species. Genome sizes were taken to be the genome assembly size and gathered from the
422 NCBI database, which resulted in 108 species with genome size available.

423 To test for a relationship between nucleotide diversity (π) and the per generation per-site
424 SNM rate (μ) while accounting for uncertainty in both, we used MCMCglmm to fit a bivariate
425 PGLMM with mutation number and π as responses, again including the number of callable
426 sites as an offset on the number of mutations and generation time as a fixed effect (see
427 **Supporting Materials**). We collected nucleotide diversity data for 102 species from Buffalo
428 (2021), Bergeron et al. (2023) or literature (see Availability of data and materials).

429

430 **AVAILABILITY OF DATA AND MATERIALS**

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

433 **ACKNOWLEDGEMENTS**

434 We thank the authors of the many studies we analyse for making their data available, and
435 particularly Lucie Bergeron and co-authors who generously shared recently published
436 nucleotide diversity data with us. We thank Jarrod Hadfield for suggesting the overall
437 statistical approach, helping with MCMCglmm syntax and interpretation, and drafting much
438 of the supporting material. This work was funded by the UK Biotechnology and Biological
439 Sciences Research Council through grant BB/T007516/1. For the purpose of open access, the
440 authors have applied a Creative Commons Attribution (CC BY) licence to any Author
441 Accepted Manuscript version arising from this submission.

442 **CONFLICT OF INTEREST**

443 The authors declare that they have no competing interests.

444 **AUTHOR CONTRIBUTIONS**

445 YW participated in data collection, performed analyses, and wrote the first draft. DJO wrote
446 a part of the first draft and made extensive contributions to the development, drafting, and
447 editing of the manuscript and supporting material, as well as model explanation. All authors
448 gave final approval for publication.

449

450

451 **Figure legends**

452 **Figure 1**: A summary of *de novo* mutation rate studies published up to 21st September 2022.
453 (A) The proportion of studies in different taxonomic groups. (B) The number of studies by
454 year. MA: mutation accumulation studies, PO: parent-offspring studies.

455 **Figure 2**. A phylogenetic meta-analysis of SNM rates for different species identified from the
456 literature. Mutation rates and 95% CIs were estimated using a PGLMM. For each annotated
457 clade, we estimated the rate for their most recent common ancestor (clades marked in
458 colour; estimates in dashed lines and below taxon icons to the right). The vertical black
459 dashed line indicates the mutation rate of common ancestor of the tree.

460 **Figure 3**. Phylogenetic linear regressions of SNM rate on (A) generation time for 116 species,
461 (B) on genome size for 108 species. The solid black lines represent simple linear regressions;
462 the coloured lines represent different PGLMM regressions conditional on each of the species
463 random effects. (C) a regression of pairwise nucleotide diversity (π) on SNM rate (μ), fitted
464 as a bivariate model in MCMCglmm for 102 species. The solid grey line indicates the model

465 regression line; the dashed grey line indicates the expected slope for $\theta_\pi = 4N_e\mu$ on the log

466 scale. All axes are plotted on a log scale.

467 **Figure 4**. An analysis of short indel mutation rates for different species. The estimated

468 mutation rates are presented on a log scale for clarity.

469 **Figure 5**. The correlation between indel mutation rates and SNM rates plotted on a log scale

470 with a 1:1 ratio shown as a dashed line (A). The aggregated log ratios of deletion/insertion in

471 eukaryotes (B).

472

473 **References**

474 Baer, C. F., Miyamoto, M. M., & Denver, D. R. (2007). Mutation rate variation in multicellular
475 eukaryotes: causes and consequences. *Nature Reviews Genetics, 8*(8), 619-631.
476 doi:10.1038/nrg2158
477 Bailey, W. J., Fitch, D., Tagle, D. A., Czelusniak, J., Slightom, J. L., & Goodman, M. (1991). Molecular
478 evolution of the psi eta-globin gene locus: gibbon phylogeny and the hominoid slowdown.
479 *Molecular Biology and Evolution, 8*(2), 155-184.
480 Belfield, E. J., Brown, C., Ding, Z. J., Chapman, L., Luo, M., Hinde, E., . . . Zheng, S. J. (2021). Thermal
481 stress accelerates Arabidopsis thaliana mutation rate. *Genome Research, 31*(1), 40-50.
482 Bennetzen, J. L., Ma, J., & Devos, K. M. (2005). Mechanisms of recent genome size variation in
483 flowering plants. *Annals of Botany, 95*(1), 127-132.
484 Bergeron, L. A., Besenbacher, S., Turner, T., Versoza, C. J., Wang, R. J., Price, A. L., . . . Schierup, M. H.
485 (2022). The Mutationathon highlights the importance of reaching standardization in
486 estimates of pedigree-based germline mutation rates. *Elife, 11*. doi:10.7554/eLife.73577
487 Bergeron, L. A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M. F., Quintard, B., . . . Shao, C. (2023).
488 Evolution of the germline mutation rate across vertebrates. *Nature*, 1-7.
489 Bourguignon, T., Kinjo, Y., Villa-Martin, P., Coleman, N. V., Tang, Q., Arab, D. A., . . . Ohkuma, M.
490 (2020). Increased mutation rate is linked to genome reduction in prokaryotes. *Current*
491 *Biology, 30*(19), 3848-3855. e3844.
492 Buffalo, V. (2021). Quantifying the relationship between genetic diversity and population size
493 suggests natural selection cannot explain Lewontin's Paradox. *Elife, 10*, e67509.
494 Burian, A. (2021). Does shoot apical meristem function as the germline in safeguarding against excess
495 of mutations? *Frontiers in Plant Science, 12*, 707740.
496 Campbell, C. D., & Eichler, E. E. (2013). Properties and rates of germline mutations in humans. *Trends*
497 *in Genetics, 29*(10), 575-584.
498 Chan, A. H., Jenkins, P. A., & Song, Y. S. (2012). Genome-wide fine-scale recombination rate variation
499 in Drosophila melanogaster. *Plos Genetics, 8*(12), e1003090.
500 doi:10.1371/journal.pgen.1003090

Chintalapati, M., & Moorjani, P. (2020). Evolution of the mutation rate across primates. *Current Opinion in Genetics & Development, 62*, 58-64.

Cruzan, M. B., Streisfeld, M. A., & Schwoch, J. A. (2022). Fitness effects of somatic mutations accumulating during vegetative growth. *Evolutionary Ecology, 36*(5), 767-785.

de Jong, W. W., & Rydén, L. (1981). Causes of more frequent deletions than insertions in mutations and protein evolution. *Nature, 290*(5802), 157-159.

Denver, D. R., Morris, K., Lynch, M., & Thomas, W. K. (2004). High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. *Nature, 430*(7000), 679-682. doi:10.1038/nature02697

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences, 88*(16), 7160-7164.

Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics, 148*(4), 1667-1686.

Drost, J. B., & Lee, W. R. (1995). Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environmental and Molecular Mutagenesis, 25*(S2), 48-64.

Freckleton, R. (2009). The seven deadly sins of comparative analysis. *Journal of Evolutionary Biology, 22*(7), 1367-1375.

Gao, Z., Wyman, M. J., Sella, G., & Przeworski, M. (2016). Interpreting the dependence of mutation rates on age and time. *Plos Biology, 14*(1), e1002355.

Girard, S. L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., . . . Rouleau, G. A. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet, 43*(9), 860-863. doi:10.1038/ng.886

Gregory, T. R. (2003). Is small indel bias a determinant of genome size? *Trends in Genetics, 19*(9), 485-488.

Gregory, T. R. (2004). Insertion–deletion biases and the evolution of genome size. *Gene, 324*, 15-34.

Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D. L., Houle, D., . . . Keightley, P. D. (2007). Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. *Nature, 445*(7123), 82-85. doi:10.1038/nature05388

Hadfield, J. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software, 33*(2), 1-22. doi:DOI 10.18637/jss.v033.i02

Hadfield, J. D., & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology, 23*(3), 494-508. doi:10.1111/j.1420-9101.2009.01915.x

Halligan, D. L., & Keightley, P. D. (2009). Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics, 40*, 151-172.

Halliwell, B., Yates, L. A., & Holland, B. R. (2022). Multi-Response Phylogenetic Mixed Models: Concepts and Application. *bioRxiv*, 2022.2012. 2013.520338.

Harland, C., Charlier, C., Karim, L., Cambisano, N., Deckers, M., Mni, M., . . . Georges, M. (2017). Frequency of mosaicism points towards mutation-prone early cleavage cell divisions in cattle. *bioRxiv*, 079863.

Harmon, L. J., Pennell, M. W., Henao-Diaz, L. F., Rolland, J., Sipley, B. N., & Uyeda, J. C. (2021). Causes and consequences of apparent timescaling across all estimated evolutionary rates. *Annual Review of Ecology, Evolution, and Systematics, 52*, 587-609.

Hasan, A. R., Lachapelle, J., El‐Shawa, S. A., Potjewyd, R., Ford, S. A., & Ness, R. W. (2022). Salt stress alters the spectrum of de novo mutation available to selection during experimental adaptation of Chlamydomonas reinhardtii. *Evolution*.

Ho, E. K. H., Macrae, F., Latta, L. C., McIlroy, P., Ebert, D., Fields, P. D., . . . Schaack, S. (2020). High and Highly Variable Spontaneous Mutation Rates in Daphnia. *Molecular Biology and Evolution, 37*(11), 3258-3266. doi:10.1093/molbev/msaa142

Ho, S. Y., Duchêne, S., & Duchêne, D. (2015). Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. *Molecular Ecology Resources, 15*(4), 688-696.

Huang, W., Lyman, R. F., Lyman, R. A., Carbone, M. A., Harbison, S. T., Magwire, M. M., & Mackay, T. F. C. (2016). Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *Elife, 5*. doi:10.7554/eLife.14625

Kaplanis, J., Ide, B., Sanghvi, R., Neville, M., Danecek, P., Coorens, T., . . . McRae, J. (2022). Genetic and chemotherapeutic influences on germline hypermutation. *Nature, 605*(7910), 503-508.

Kapusta, A., Suh, A., & Feschotte, C. (2017). Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences, 114*(8), E1460-E1469.

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

Keightley, P. D., & Eyre-Walker, A. (2000). Deleterious mutations and the evolution of sex. *Science, 290*(5490), 331-333.

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-+. doi:10.1534/genetics.113.158758

Keightley, P. D., Trivedi, U., Thomson, M., Oliver, F., Kumar, S., & Blaxter, M. L. (2009). Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. *Genome Research, 19*(7), 1195-1201. doi:10.1101/gr.091231.109

Keith, N., Jackson, C. E., Glaholt, S. P., Young, K., Lynch, M., & Shaw, J. R. (2021). Genome-Wide Analysis of Cadmium-Induced, Germline Mutations in a Long-Term Daphnia pulex Mutation-Accumulation Experiment. *Environmental Health Perspectives, 129*(10). doi:10.1289/Ehp8932

Kimura, M. (1967). On the evolutionary adjustment of spontaneous mutation rates. *Genetics Research, 9*(1), 23-34.

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

Kondrashov, A. S. (2003). Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Human Mutation, 21*(1), 12-27. doi:10.1002/humu.10147

Kondrashov, A. S., & Crow, J. F. (1993). A molecular approach to estimating the human deleterious mutation rate. *Human Mutation, 2*(3), 229-234. doi:10.1002/humu.1380020312

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., . . . Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature, 488*(7412), 471-475. doi:10.1038/nature11396

Konrad, A., Brady, M. J., Bergthorsson, U., & Katju, V. (2019). Mutational Landscape of Spontaneous Base Substitutions and Small Indels in Experimental Caenorhabditis elegans Populations of Differing Size. *Genetics, 212*(3), 837-854. doi:10.1534/genetics.119.302054

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., . . . Hedges, S. B. (2022). TimeTree 5: An expanded resource for species divergence times. *Molecular Biology and Evolution, 39*(8), msac174.

Lanfear, R. (2018). Do plants have a segregated germline? *Plos Biology, 16*(5), e2005439.

Li, W.-H., Tanimura, M., & Sharp, P. M. (1987). An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution, 25*(4), 330-342.

Liu, H., & Zhang, J. (2021). The rate and molecular spectrum of mutation are selectively maintained in yeast. *Nature Communications, 12*(1), 1-11.

Lopez-Cortegano, E., Craig, R. J., Chebib, J., Samuels, T., Morgan, A. D., Kraemer, S. A., . . . Keightley, P. D. (2021). De Novo Mutation Rate Variation and Its Determinants in Chlamydomonas. *Molecular Biology and Evolution, 38*(9), 3709-3723. doi:10.1093/molbev/msab140

599 Lynch, M. (1991). Methods for the analysis of comparative data in evolutionary biology. *Evolution,*
600 *45*(5), 1065-1080.
601 Lynch, M. (2008). The cellular, developmental and population-genetic determinants of mutation-rate
602 evolution. *Genetics, 180*(2), 933-943.
603 Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics, 26*(8), 345-352.
604 Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1): Sinauer
605 Sunderland, MA.
606 Marais, G. A., Batut, B., & Daubin, V. (2020). Genome evolution: mutation is the main driver of
607 genome size in prokaryotes. *Current Biology, 30*(19), R1083-R1085.
608 Martin, A. P., & Palumbi, S. R. (1993). Body size, metabolic rate, generation time, and the molecular
609 clock. *Proceedings of the National Academy of Sciences, 90*(9), 4087-4091.
610 Meier, B., Cooke, S. L., Weiss, J., Bailly, A. P., Alexandrov, L. B., Marshall, J., . . . Campbell, P. J. (2014).
611 C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens
612 and DNA repair deficiency. *Genome Research, 24*(10), 1624-1636. doi:10.1101/gr.175547.114
613 Mukai, T., & Cockerham, C. C. (1977). Spontaneous mutation rates at enzyme loci in Drosophila
614 melanogaster. *Proceedings of the National Academy of Sciences, 74*(6), 2514-2517.
615 Nachman, M. W., & Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans.
616 *Genetics, 156*(1), 297-304.
617 Ness, R. W., Morgan, A. D., Colegrave, N., & Keightley, P. D. (2012). Estimate of the Spontaneous
618 Mutation Rate in Chlamydomonas reinhardtii. *Genetics, 192*(4), 1447-+.
619 doi:10.1534/genetics.112.145078
620 Nute, P. E., & Stamatoyannopoulos, G. (1984). Estimating Mutation-Rates Using Abnormal Human
621 Hemoglobins. *Yearbook of Physical Anthropology, 27*, 135-151.
622 Obbard, D. J., Maclennan, J., Kim, K. W., Rambaut, A., O'Grady, P. M., & Jiggins, F. M. (2012).
623 Estimating Divergence Dates and Substitution Rates in the Drosophila Phylogeny. *Molecular*
624 *Biology and Evolution, 29*(11), 3459-3473. doi:10.1093/molbev/mss150
625 Ohno, M. (2019). Spontaneous de novo germline mutations in humans and mice: rates, spectra,
626 causes and consequences. *Genes & Genetic Systems, 94*(1), 13-22.
627 Peck, K. M., & Lauring, A. S. (2018). Complexities of viral mutation rates. *Journal of Virology, 92*(14),
628 e01031-01017.
629 Petrov, D. A. (2002). Mutational equilibrium model of genome size evolution. *Theoretical Population*
630 *Biology, 61*(4), 531-544.
631 Rajaei, M., Saxena, A. S., Johnson, L. M., Snyder, M. C., Crombie, T. A., Tanny, R. E., . . . Baer, C. F.
632 (2021). Mutability of mononucleotide repeats, not oxidative stress, explains the discrepancy
633 between laboratory-accumulated mutations and the natural allele-frequency spectrum in C.
634 elegans. *Genome Research, 31*(9), 1602-1613.
635 Saxena, A. S., Salomon, M. P., Matsuba, C., Yeh, S. D., & Baer, C. F. (2019). Evolution of the
636 Mutational Process under Relaxed Selection in Caenorhabditis elegans. *Molecular Biology*
637 *and Evolution, 36*(2), 239-251. doi:10.1093/molbev/msy213
638 Schrider, D. R., Houle, D., Lynch, M., & Hahn, M. W. (2013). Rates and Genomic Consequences of
639 Spontaneous Mutational Events in Drosophila melanogaster. *Genetics, 194*(4), 937-+.
640 doi:10.1534/genetics.113.151670
641 Sigurðardóttir, S., Helgason, A., Gulcher, J. R., Stefansson, K., & Donnelly, P. (2000). The mutation
642 rate in the human mtDNA control region. *The American Journal of Human Genetics, 66*(5),
643 1599-1609.
644 Stamatoyannopoulos, G., & Nute, P. E. (1982). De novo mutations producing unstable Hbs or Hbs M.
645 II. Direct estimates of minimum nucleotide mutation rates in man. *Human Genetics, 60*(2),
646 181-188. doi:10.1007/BF00569709

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, 109*(45), 18488-18492.

Thomas, G. W., & Hahn, M. W. (2014). The human mutation rate is increasing, even as it slows. *Molecular Biology and Evolution, 31*(2), 253-257.

Thomas, G. W., Wang, R. J., Puri, A., Harris, R. A., Raveendran, M., Hughes, D. S., . . . Muzny, D. M. (2018). Reproductive longevity predicts mutation rates in primates. *Current Biology, 28*(19), 3193-3197. e3195.

Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H., . . . Green, E. D. (2003). Comparative analyses of multi-species sequences from targeted genomic regions. *Nature, 424*(6950), 788-793. doi:10.1038/nature01858

Trimble, B. K., & Doughty, J. H. (1974). The amount of hereditary disease in human populations. *Annals of Human Genetics, 38*(2), 199-223. doi:10.1111/j.1469-1809.1974.tb01951.x

Waldvogel, A.-M., & Pfenninger, M. (2021). Temperature dependence of spontaneous mutation rates. *Genome Research, 31*(9), 1582-1589.

Wang, L., Ji, Y., Hu, Y., Hu, H., Jia, X., Jiang, M., . . . Jia, Y. (2019). The architecture of intra-organism mutation rate variation in plants. *Plos Biology, 17*(4), e3000191.

Wang, R. J., Peña-Garcia, Y., Bibby, M. G., Raveendran, M., Harris, R. A., Jansen, H. T., . . . Hahn, M. W. (2022). Examining the effects of hibernation on germline mutation rates in grizzly bears. *Genome biology and evolution*.

Wang, R. J., Thomas, G. W., Raveendran, M., Harris, R. A., Doddapaneni, H., Muzny, D. M., . . . 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), 826-834.

Wang, Y., McNeil, P., Abdulazeez, R., Pascual, M., Johnston, S. E., Keightley, P. D., & Obbard, D. (2023). Variation in mutation, recombination, and transposition rates in Drosophila melanogaster and Drosophila simulans. *Genome Research*, gr. 277383.277122.

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*(6), 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.

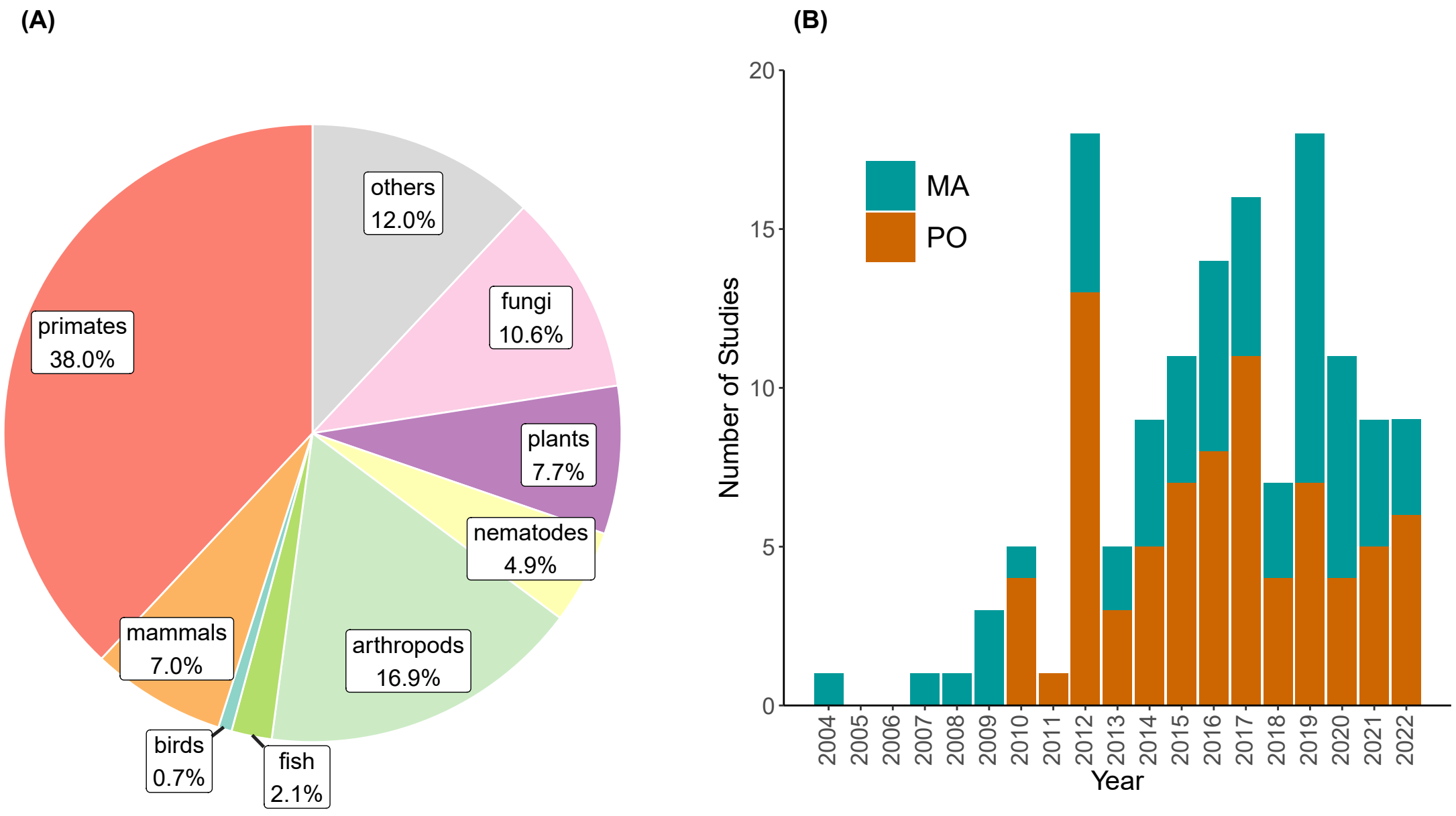
Wu, F. L., Strand, A. I., Cox, L. A., Ober, C., Wall, J. D., Moorjani, P., & Przeworski, M. (2020). A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *Plos Biology, 18*(8), e3000838.

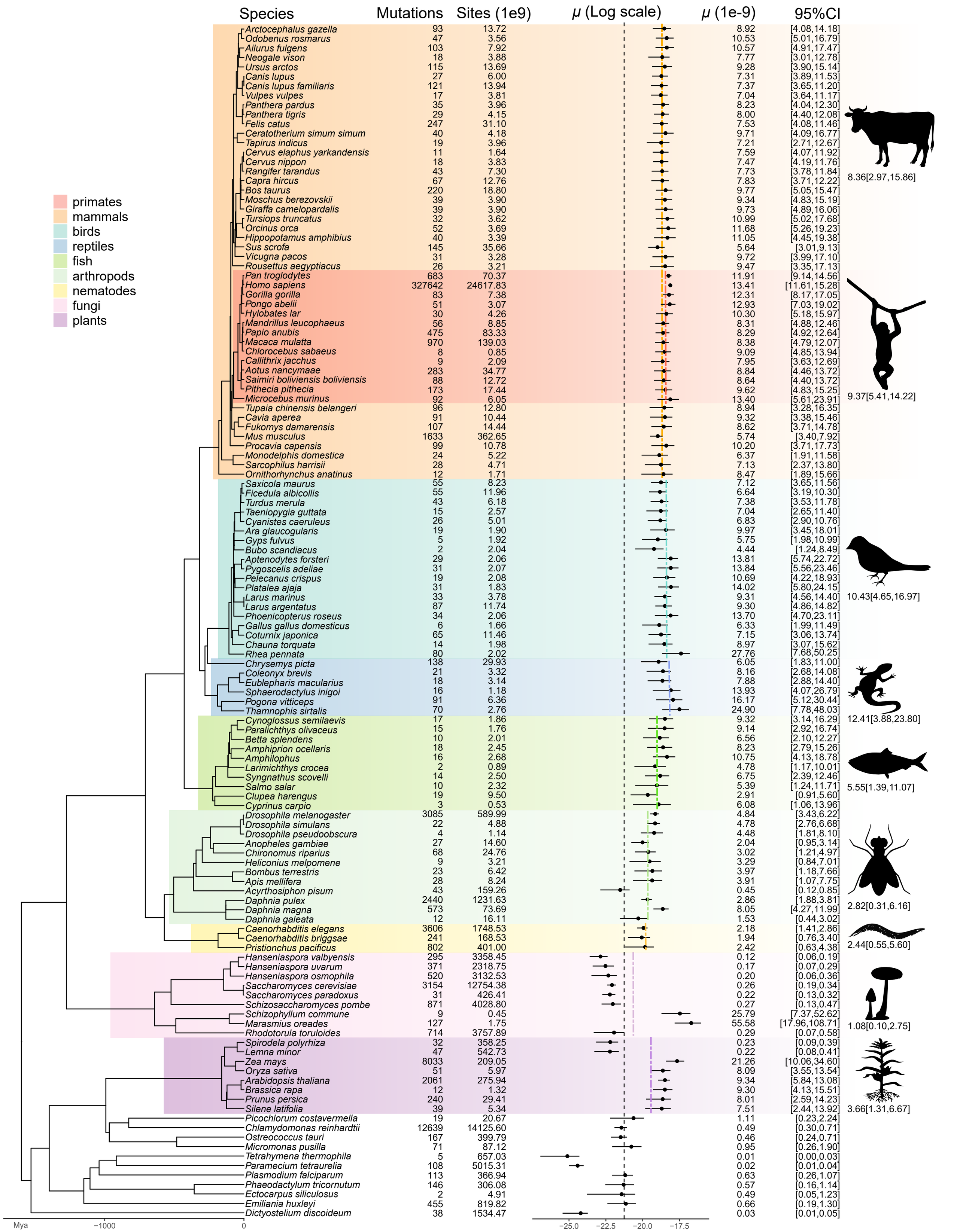
Xu, S., Stapley, J., Gablenz, S., Boyer, J., Appenroth, K. J., Sree, K. S., . . . Huber, M. (2019). Low genetic variation is associated with low mutation rate in the giant duckweed. *Nature Communications, 10*(1), 1-6.

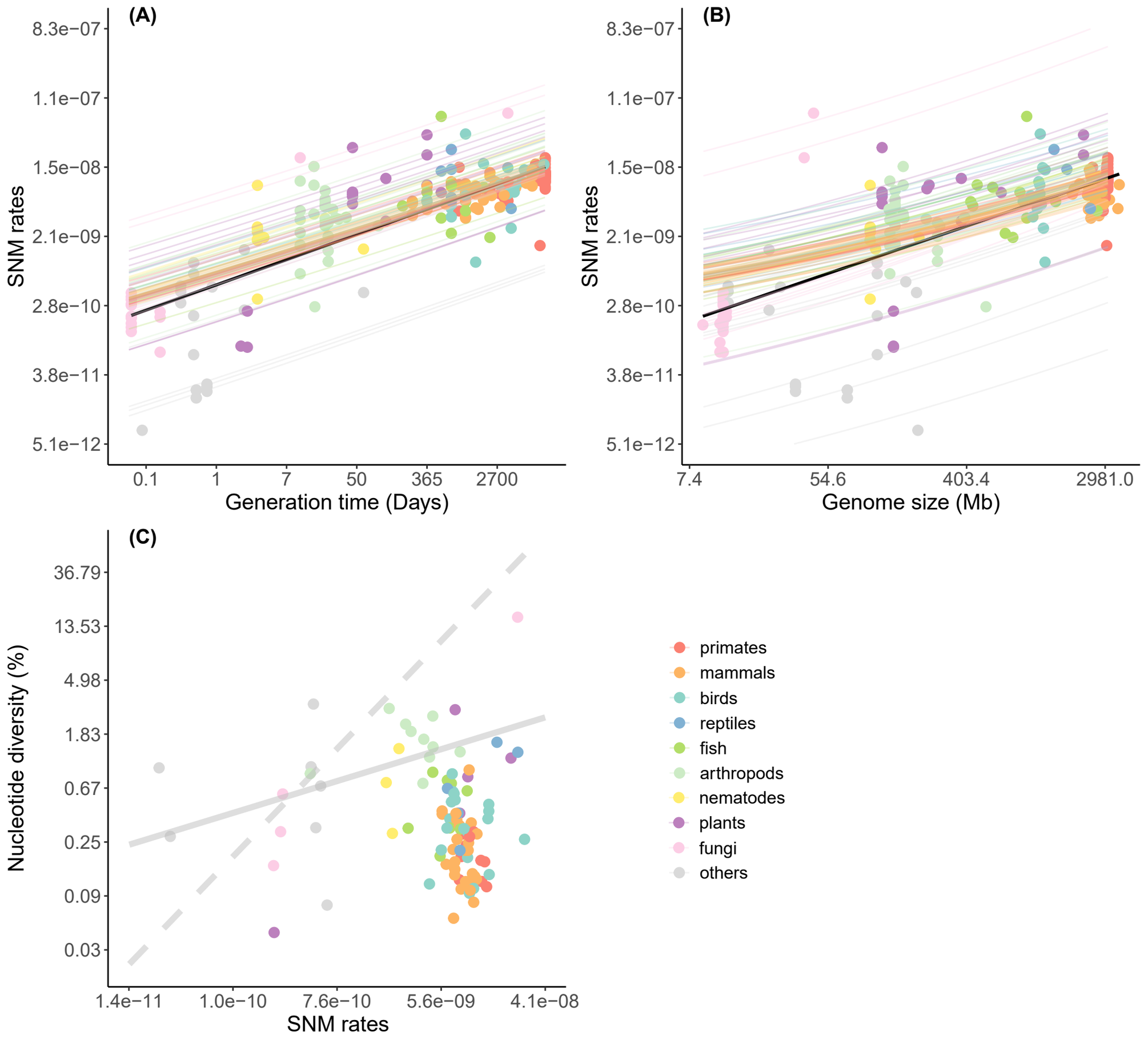
Yang, S. H., Wang, L., Huang, J., Zhang, X. H., Yuan, Y., Chen, J. Q., . . . Tian, D. C. (2015). Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature, 523*(7561), 463-U187. doi:10.1038/nature14649

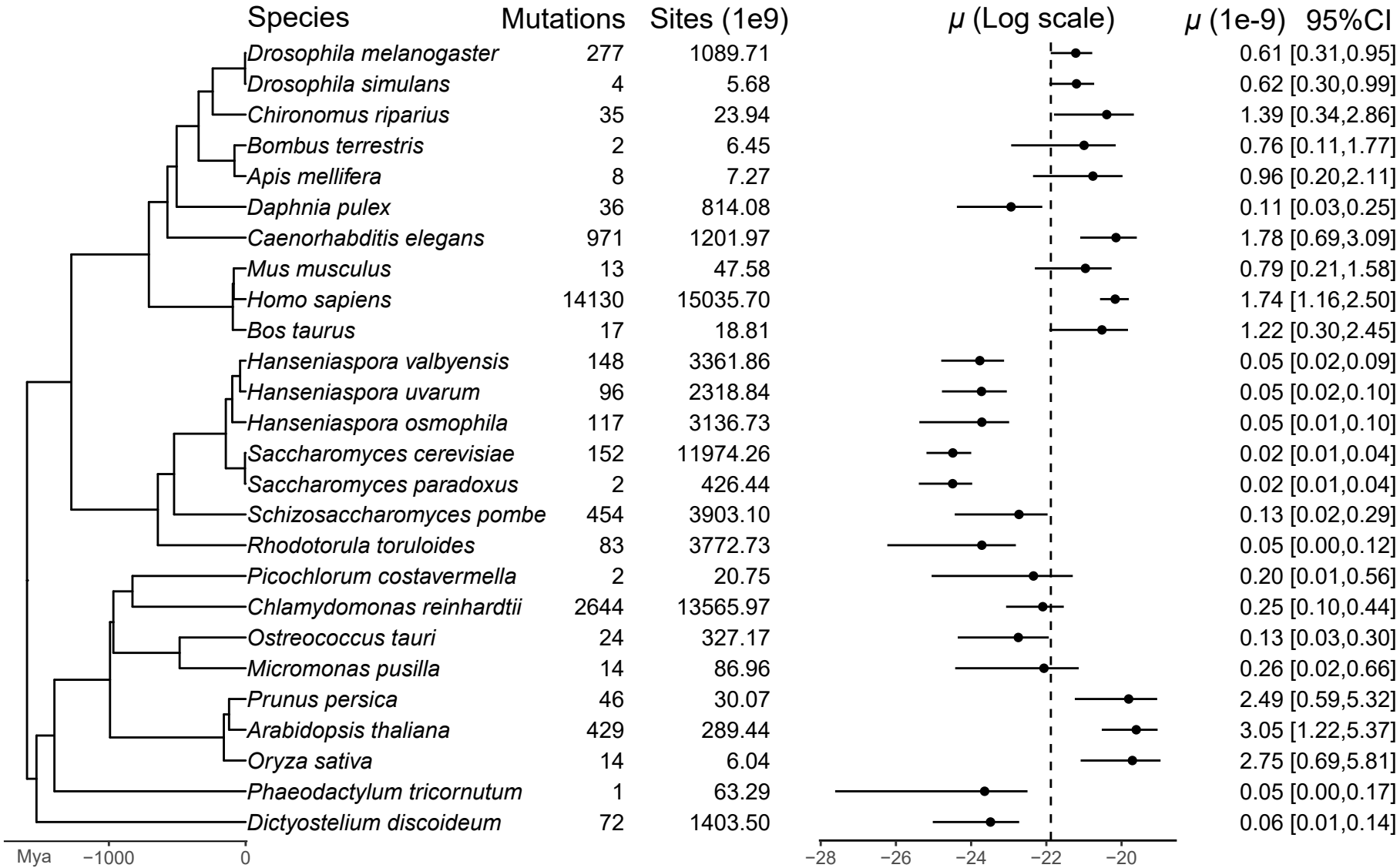
Yoder, A. D., & Tiley, G. P. (2021). The challenge and promise of estimating the de novo mutation rate from whole-genome comparisons among closely related individuals. *Molecular Ecology, 30*(23), 6087-6100. doi:10.1111/mec.16007

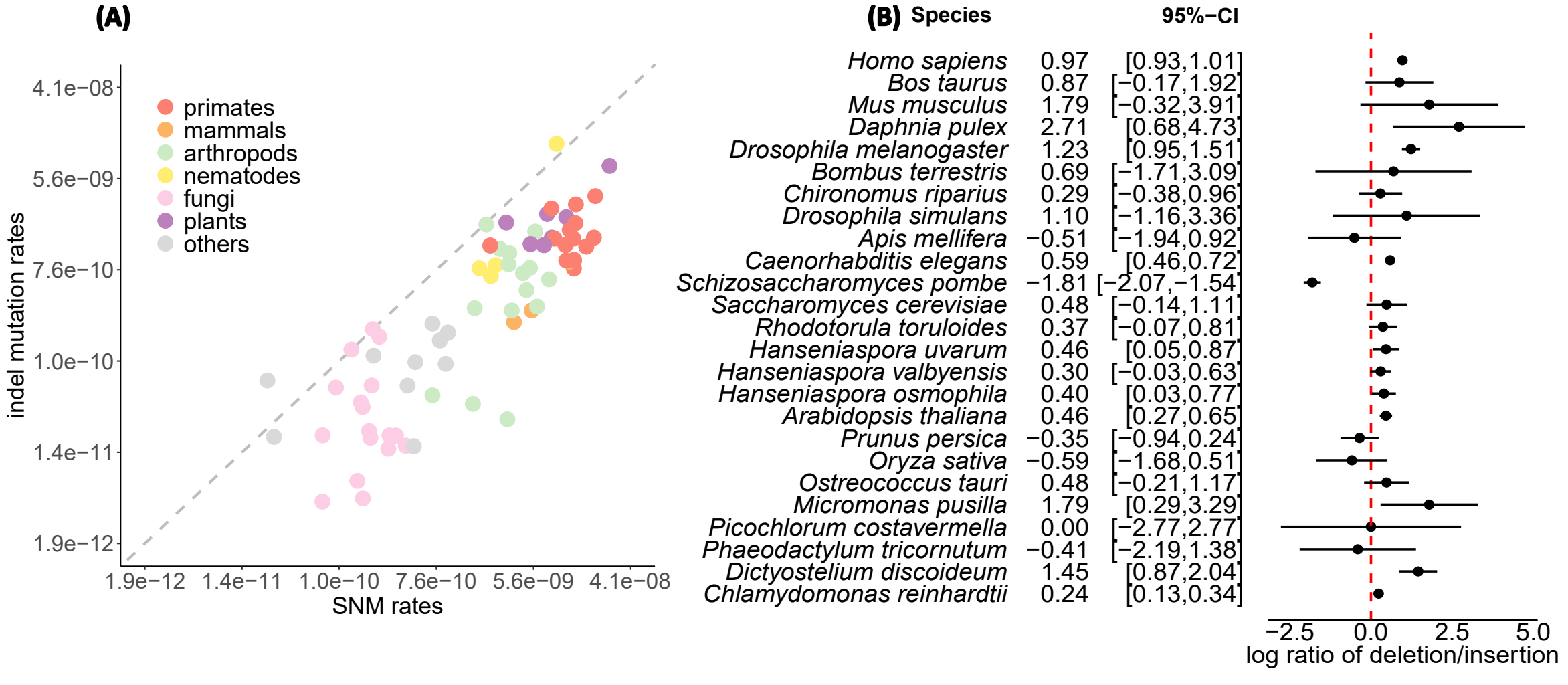
Yu, G. (2020). Using ggtree to visualize data on tree‐like structures. *Current protocols in bioinformatics, 69*(1), e96.


(A)
others
12.0%
fungi
10.6%
plants
7.7%
nematodes
4.9%
arthropods
16.9%
fish
2.1%
birds
0.7%
mammals
7.0%
primates
38.0%
(B)
MA
PO
Number of Studies
0
5
10
15
20
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Year


Species
Mutations
Sites (1e9)
μ (Log scale)
μ (1e-9)
95%CI
primates
mammals
birds
reptiles
fish
arthropods
nematodes
fungi
plants
Arctocephalus gazella 93 13.72 8.92 [4.08,14.18]
Odobenus rosmarus 47 3.56 10.53 [5.01,16.79]
Ailurus fulgens 103 7.92 10.57 [4.91,17.47]
Neogale vison 18 3.88 7.77 [3.01,12.78]
Ursus arctos 115 13.69 9.28 [3.90,15.14]
Canis lupus 27 6.00 7.31 [3.89,11.53]
Canis lupus familiaris 121 13.94 7.37 [3.65,11.20]
Vulpes vulpes 17 3.81 7.04 [3.64,11.17]
Panthera pardus 35 3.96 8.23 [4.04,12.30]
Panthera tigris 29 4.15 8.00 [4.40,12.08]
Felis catus 247 31.10 7.53 [4.08,11.46]
Ceratotherium simum simum 40 4.18 9.71 [4.09,16.77]
Tapirus indicus 19 3.96 7.21 [2.71,12.67]
Cervus elaphus yarkandensis 11 1.64 7.59 [4.07,11.92]
Cervus nippon 18 3.83 7.47 [4.19,11.76]
Rangifer tarandus 43 7.30 7.73 [3.78,11.84]
Capra hircus 67 12.76 7.83 [3.71,12.22]
Bos taurus 220 18.80 9.77 [5.05,15.47]
Moschus berezovskii 39 3.90 9.34 [4.83,15.19]
Giraffa camelopardalis 39 3.90 9.73 [4.89,16.06]
Tursiops truncatus 32 3.62 10.99 [5.02,17.68]
Orcinus orca 52 3.69 11.68 [5.26,19.23]
Hippopotamus amphibius 40 3.39 11.05 [4.45,19.38]
Sus scrofa 145 35.66 5.64 [3.01,9.13]
Vicugna pacos 31 3.28 9.72 [3.99,17.10]
Rousettus aegyptiacus 26 3.21 9.47 [3.35,17.13]
8.36[2.97,15.86]
Pan troglodytes 683 70.37 11.91 [9.14,14.56]
Homo sapiens 327642 24617.83 13.41 [11.61,15.28]
Gorilla gorilla 83 7.38 12.31 [8.17,17.05]
Pongo abelii 51 3.07 12.93 [7.03,19.02]
Hylobates lar 30 4.26 10.30 [5.18,15.97]
Mandrillus leucophaeus 56 8.85 8.31 [4.88,12.46]
Papio anubis 475 83.33 8.29 [4.92,12.64]
Macaca mulatta 970 139.03 8.38 [4.79,12.07]
Chlorocebus sabaeus 8 0.85 9.09 [4.85,13.94]
Callithrix jacchus 9 2.09 7.95 [3.63,12.69]
Aotus nancymaae 283 34.77 8.84 [4.46,13.72]
Saimiri boliviensis boliviensis 88 12.72 8.64 [4.40,13.72]
Pithecia pithecia 173 17.44 9.62 [4.83,15.25]
Microcebus murinus 92 6.05 13.40 [5.61,23.91]
9.37[5.41,14.22]
Tupaia chinensis belangeri 96 12.80 8.94 [3.28,16.35]
Cavia aperea 91 10.44 9.32 [3.38,15.46]
Fukomys damarensis 107 14.44 8.62 [3.71,14.78]
Mus musculus 1633 362.65 5.74 [3.40,7.92]
Procavia capensis 99 10.78 10.20 [3.71,17.73]
Monodelphis domestica 24 5.22 6.37 [1.91,11.58]
Sarcophilus harrisii 28 4.71 7.13 [2.37,13.80]
Ornithorhynchus anatinus 12 1.71 8.47 [1.89,15.66]
Saxicola maurus 55 8.23 7.12 [3.65,11.56]
Ficedula albicollis 55 11.96 6.64 [3.19,10.30]
Turdus merula 43 6.18 7.38 [3.53,11.78]
Taeniopygia guttata 15 2.57 7.04 [2.65,11.40]
Cyanistes caeruleus 26 5.01 6.83 [2.90,10.76]
Ara glaucogularis 19 1.90 9.97 [3.45,18.01]
Gyps fulvus 5 1.92 5.75 [1.98,10.99]
Bubo scandiacus 2 2.04 4.44 [1.24,8.49]
Aptenodytes forsteri 29 2.06 13.81 [5.74,22.72]
Pygoscelis adeliae 31 2.07 13.84 [5.56,23.46]
Pelecanus crispus 19 2.08 10.69 [4.22,18.93]
Platalea ajaja 31 1.83 14.02 [5.80,24.15]
Larus marinus 33 3.78 9.31 [4.56,14.40]
Larus argentatus 87 11.74 9.30 [4.86,14.82]
Phoenicopterus roseus 34 2.06 13.70 [4.70,23.11]
Gallus gallus domesticus 6 1.66 6.33 [1.99,11.49]
Coturnix japonica 65 11.46 7.15 [3.06,13.74]
Chauna torquata 14 1.98 8.97 [3.07,15.62]
Rhea pennata 80 2.02 27.76 [7.68,50.25]
10.43[4.65,16.97]
Chrysemys picta 138 29.93 6.05 [1.83,11.00]
Coleonyx brevis 21 3.32 8.16 [2.68,14.08]
Eublepharis macularius 18 3.14 7.88 [2.88,14.40]
Sphaerodactylus inigoi 16 1.18 13.93 [4.07,26.79]
Pogona vitticeps 91 6.36 16.17 [5.12,30.44]
Thamnophis sirtalis 70 2.76 24.90 [7.78,48.03]
12.41[3.88,23.80]
Cynoglossus semilaevis 17 1.86 9.32 [3.14,16.29]
Paralichthys olivaceus 15 1.76 9.14 [2.92,16.74]
Betta splendens 10 2.01 6.56 [2.10,12.27]
Amphiprion ocellaris 18 2.45 8.23 [2.79,15.26]
Amphilophus 16 2.68 10.75 [4.13,18.78]
Larimichthys crocea 2 0.89 4.78 [1.17,10.01]
Syngnathus scovelli 14 2.50 6.75 [2.39,12.46]
Salmo salar 10 2.32 5.39 [1.24,11.71]
Clupea harengus 19 9.50 2.91 [0.91,5.60]
Cyprinus carpio 3 0.53 6.08 [1.06,13.96]
5.55[1.39,11.07]
Drosophila melanogaster 3085 589.99 4.84 [3.43,6.22]
Drosophila simulans 22 4.88 4.78 [2.76,6.68]
Drosophila pseudoobscura 4 1.14 4.48 [1.81,8.10]
Anopheles gambiae 27 14.60 2.04 [0.95,3.14]
Chironomus riparius 68 24.76 3.02 [1.21,4.97]
Heliconius melpomene 9 3.21 3.29 [0.84,7.01]
Bombus terrestris 23 6.42 3.97 [1.18,7.66]
Apis mellifera 28 8.24 3.91 [1.07,7.75]
Acyrthosiphon pisum 43 159.26 0.45 [0.12,0.85]
Daphnia pulex 2440 1231.63 2.86 [1.88,3.81]
Daphnia magna 573 73.69 8.05 [4.27,11.99]
Daphnia galeata 12 16.11 1.53 [0.44,3.02]
2.82[0.31,6.16]
Caenorhabditis elegans 3606 1748.53 2.18 [1.41,2.86]
Caenorhabditis briggsae 241 168.53 1.94 [0.76,3.40]
Pristionchus pacificus 802 401.00 2.42 [0.63,4.38]
2.44[0.55,5.60]
Hanseniaspora valbyensis 295 3358.45 0.12 [0.06,0.19]
Hanseniaspora uvarum 371 2318.75 0.17 [0.07,0.29]
Hanseniaspora osmophila 520 3132.53 0.20 [0.06,0.36]
Saccharomyces cerevisiae 3154 12754.38 0.26 [0.19,0.34]
Saccharomyces paradoxus 31 426.41 0.22 [0.13,0.32]
Schizosaccharomyces pombe 871 4028.80 0.27 [0.13,0.47]
Schizophyllum commune 9 0.45 25.79 [7.37,52.62]
Marasmius oreades 127 1.75 55.58 [17.96,108.71]
Rhodotorula toruloides 714 3757.89 0.29 [0.07,0.58]
1.08[0.10,2.75]
Spirodela polyrhiza 32 358.25 0.23 [0.09,0.39]
Lemna minor 47 542.73 0.22 [0.08,0.41]
Zea mays 8033 209.05 21.26 [10.06,34.60]
Oryza sativa 51 5.97 8.09 [3.55,13.54]
Arabidopsis thaliana 2061 275.94 9.34 [5.84,13.08]
Brassica rapa 12 1.32 9.30 [4.13,15.51]
Prunus persica 240 29.41 8.01 [2.59,14.23]
Silene latifolia 39 5.34 7.51 [2.44,13.92]
3.66[1.31,6.67]
Picochlorum costavermella 19 20.67 1.11 [0.23,2.24]
Chlamydomonas reinhardtii 12639 14125.60 0.49 [0.30,0.71]
Ostreococcus tauri 167 399.79 0.46 [0.24,0.71]
Micromonas pusilla 71 87.12 0.95 [0.26,1.90]
Tetrahymena thermophila 5 657.03 0.01 [0.00,0.03]
Paramecium tetraurelia 108 5015.31 0.02 [0.01,0.04]
Plasmodium falciparum 113 366.94 0.63 [0.26,1.07]
Phaeodactylum tricornutum 146 306.08 0.57 [0.16,1.14]
Ectocarpus siliculosus 2 4.91 0.49 [0.05,1.23]
Emiliania huxleyi 455 819.82 0.66 [0.19,1.30]
Dictyostelium discoideum 38 1534.47 0.03 [0.01,0.05]
Mya
−1000
0
−25.0
−22.5
−20.0
−17.5


(A)
SNM rates
8.3e−07
1.1e−07
1.5e−08
2.1e−09
2.8e−10
3.8e−11
5.1e−12
0.1
1
7
50
365
2700
Generation time (Days)
(B)
SNM rates
8.3e−07
1.1e−07
1.5e−08
2.1e−09
2.8e−10
3.8e−11
5.1e−12
7.4
54.6
403.4
2981.0
Genome size (Mb)
(C)
Nucleotide diversity (%)
36.79
13.53
4.98
1.83
0.67
0.25
0.09
0.03
1.4e−11
1.0e−10
7.6e−10
5.6e−09
4.1e−08
SNM rates
primates
mammals
birds
reptiles
fish
arthropods
nematodes
plants
fungi
others


Species
Mutations
Sites (1e9)
μ (Log scale)
μ (1e-9) 95%CI
Drosophila melanogaster 277 1089.71 0.61 [0.31,0.95]
Drosophila simulans 4 5.68 0.62 [0.30,0.99]
Chironomus riparius 35 23.94 1.39 [0.34,2.86]
Bombus terrestris 2 6.45 0.76 [0.11,1.77]
Apis mellifera 8 7.27 0.96 [0.20,2.11]
Daphnia pulex 36 814.08 0.11 [0.03,0.25]
Caenorhabditis elegans 971 1201.97 1.78 [0.69,3.09]
Mus musculus 13 47.58 0.79 [0.21,1.58]
Homo sapiens 14130 15035.70 1.74 [1.16,2.50]
Bos taurus 17 18.81 1.22 [0.30,2.45]
Hanseniaspora valbyensis 148 3361.86 0.05 [0.02,0.09]
Hanseniaspora uvarum 96 2318.84 0.05 [0.02,0.10]
Hanseniaspora osmophila 117 3136.73 0.05 [0.01,0.10]
Saccharomyces cerevisiae 152 11974.26 0.02 [0.01,0.04]
Saccharomyces paradoxus 2 426.44 0.02 [0.01,0.04]
Schizosaccharomyces pombe 454 3903.10 0.13 [0.02,0.29]
Rhodotorula toruloides 83 3772.73 0.05 [0.00,0.12]
Picochlorum costavermella 2 20.75 0.20 [0.01,0.56]
Chlamydomonas reinhardtii 2644 13565.97 0.25 [0.10,0.44]
Ostreococcus tauri 24 327.17 0.13 [0.03,0.30]
Micromonas pusilla 14 86.96 0.26 [0.02,0.66]
Prunus persica 46 30.07 2.49 [0.59,5.32]
Arabidopsis thaliana 429 289.44 3.05 [1.22,5.37]
Oryza sativa 14 6.04 2.75 [0.69,5.81]
Phaeodactylum tricornutum 1 63.29 0.05 [0.00,0.17]
Dictyostelium discoideum 72 1403.50 0.06 [0.01,0.14]
Mya −1000 0
−28 −26 −24 −22 −20


(A)
primates
mammals
arthropods
nematodes
fungi
plants
others
indel mutation rates
4.1e−08
5.6e−09
7.6e−10
1.0e−10
1.4e−11
1.9e−12
1.9e−12
1.4e−11
1.0e−10
7.6e−10
5.6e−09
4.1e−08
SNM rates
(B)
Species
95%−CI
Homo sapiens 0.97 [0.93,1.01]
Bos taurus 0.87 [−0.17,1.92]
Mus musculus 1.79 [−0.32,3.91]
Daphnia pulex 2.71 [0.68,4.73]
Drosophila melanogaster 1.23 [0.95,1.51]
Bombus terrestris 0.69 [−1.71,3.09]
Chironomus riparius 0.29 [−0.38,0.96]
Drosophila simulans 1.10 [−1.16,3.36]
Apis mellifera −0.51 [−1.94,0.92]
Caenorhabditis elegans 0.59 [0.46,0.72]
Schizosaccharomyces pombe −1.81 [−2.07,−1.54]
Saccharomyces cerevisiae 0.48 [−0.14,1.11]
Rhodotorula toruloides 0.37 [−0.07,0.81]
Hanseniaspora uvarum 0.46 [0.05,0.87]
Hanseniaspora valbyensis 0.30 [−0.03,0.63]
Hanseniaspora osmophila 0.40 [0.03,0.77]
Arabidopsis thaliana 0.46 [0.27,0.65]
Prunus persica −0.35 [−0.94,0.24]
Oryza sativa −0.59 [−1.68,0.51]
Ostreococcus tauri 0.48 [−0.21,1.17]
Micromonas pusilla 1.79 [0.29,3.29]
Picochlorum costavermella 0.00 [−2.77,2.77]
Phaeodactylum tricornutum −0.41 [−2.19,1.38]
Dictyostelium discoideum 1.45 [0.87,2.04]
Chlamydomonas reinhardtii 0.24 [0.13,0.34]
−2.5
0.0
2.5
5.0
log ratio of deletion/insertion