



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Variation in the spectrum of new mutations among inbred strains of mice

Citation for published version:

López-Cortegano, E, Chebib, J, Jonas, A, Vock, A, Künzel, S, Tautz, D & Keightley, PD 2024, 'Variation in the spectrum of new mutations among inbred strains of mice', *Molecular Biology and Evolution*, vol. 41, no. 8, msae163. <https://doi.org/10.1093/molbev/msae163>

Digital Object Identifier (DOI):

[10.1093/molbev/msae163](https://doi.org/10.1093/molbev/msae163)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Molecular Biology and Evolution

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Variation in the spectrum of new mutations among inbred strains of mice

Eugenio López-Cortegano^{1,*}, Jobran Chebib^{1,*}, Anika Jonas², Anastasia Vock², Sven Künzel², Diethard Tautz², Peter D. Keightley¹

* These authors contributed equally to this work.

1. Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, EH9 3FL, UK
2. Department for Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, 24306, Plön, Germany

Abstract

The mouse serves as a mammalian model for understanding the nature of variation from new mutations, a question that has both evolutionary and medical significance. Previous studies suggest that the rate of single nucleotide mutations (SNMs) in mice is approximately 50% of that in humans. However, information largely comes from studies involving the C57BL/6 strain, and there is little information from other mouse strains. Here, we study the mutations that accumulated in 59 mouse lines derived from four inbred strains that are commonly used in genetics and clinical research (BALB/cAnNRj, C57BL/6JRj, C3H/HeNRj, and FVB/NRj), maintained for 8-9 generations by brother-sister mating. By analysing Illumina whole-genome sequencing data, we estimate that the average rate of new SNMs in mice is approximately $\mu = 6.7 \times 10^{-9}$. However, there is substantial variation in the spectrum of SNMs among strains, so the burden from new mutations also varies among strains. For example, the FVB strain has a spectrum that is markedly skewed towards C → A transversions, and is likely to experience a higher deleterious load than other strains, due to an increased frequency of nonsense mutations in glutamic

acid codons. Finally, we observe substantial variation in the rate of new SNMs among DNA sequence contexts, CpG sites and their adjacent nucleotides playing an important role.

Introduction

De novo germline mutations are the primary source of new adaptive genetic variation, and also give rise to many disorders that have a genetic basis. Knowledge of the mutation rate is important for modelling evolutionary processes, such as the molecular clock (Tiley et al. 2020), and for clinical applications in humans, such as in interpreting the incidence of genetic diseases (Veltman and Brunner 2012). However, mutation rates show substantial variation across taxa (Bergeron et al. 2023), caused by numerous genetic and non-genetic factors, including sequence repetitiveness, metabolic rate, parental age and the presence or absence of genetic and environmental stressors, among others (Baer et al. 2007; Sharp and Agrawal 2012; Rahbari et al. 2016; Liu and Zhang 2019). Moreover, the mutation rate varies considerably within species, both between families and between populations (Wang et al. 2023). Therefore, a thorough understanding of the underlying causes of variation in the mutation rate is essential for understanding evolutionary processes.

The mutation rate behaves as a quantitative trait that is subject to evolutionary forces such as selection and genetic drift (Lynch et al. 2016). It has been proposed that variation in the mutation rate across different taxonomic groups could be at least partly explained by the interplay between selection and genetic drift. According to the 'drift-barrier' hypothesis, most new mutations with effects on fitness are deleterious, so natural selection acts to reduce the mutation rate, but this process is more efficient in larger populations (Lynch 2011). Recently, Bergeron et al. (2023) found

evidence supporting this hypothesis, reporting a significant negative correlation between the mutation rate and the effective population size across 68 vertebrate species. If alleles with mutator effects negatively impact fitness, the evolutionary dynamics of the DNA repair machinery would therefore be expected to influence variation in the mutation rate on a phylogenetic scale.

Given its dependence on the DNA repair machinery, the spectrum of new mutations is fundamental for interpreting variation in the mutation rate. Variation in the DNA repair machinery can alter the spectrum of single nucleotide mutations (SNMs), i.e. the frequencies of every possible type of SNM (Lee et al. 2012; Sane et al. 2023). The SNM spectrum is important for understanding genome evolution, and also has applications in human genetics, as it can provide insights into the consequences of alterations of DNA repair pathways and can aid in the identification of mutational processes underlying tumour development in cancers (Alexandrov et al. 2020). For example, the single-base substitution (SBS) signature SBS6, typically found in microsatellite unstable tumours, has a pronounced enrichment of C → T transitions, and is associated with defective DNA mismatch repair in humans and mice (Alexandrov et al. 2020). Experimental studies also suggest that DNA mismatch repair genes have a pivotal role in determining the SNM spectrum, as evidenced in bacteria (Lee et al. 2012; Dillon et al. 2017) and *Caenorhabditis elegans* (Volkova et al. 2020). On an evolutionary time scale, variation in SNM spectra can be attributed to sequence divergence of DNA repair loci, and there is evidence of this in *Chlamydomonas* (López-Cortegano et al. 2021). Furthermore, alterations in the SNM spectrum may impose evolutionary constraints on changes in functionally important sequences, including coding sequences, and this possibility has received little attention.

Here, we present an in depth study on the rates and spectra of *de novo* mutations in common laboratory strains of the house mouse (*Mus musculus*). The mouse serves as a valuable model organism for genetic research in mammals, due to its relatively short generation time, large litters, and extensive genetic resources. Furthermore, the mouse has long been used as a model for understanding fundamental human biology and genetic disease (Schofield et al. 2012; Lilue et al. 2019). To study variation in the rate of new mutations, we adopted an experimental design involving the establishment of mutation accumulation (MA) lines derived from pairs of founder individuals of four different strains. The MA lines were maintained by full-sib mating, aiming to minimise the effective population size and the effectiveness of natural selection, thereby allowing new mutations to accumulate largely by genetic drift. We performed whole-genome sequencing of the founder individuals and individuals from the MA lines eight or nine generations later. Mutation accumulation combined with whole-genome sequencing has previously been employed in a wide range of taxa in order to estimate the mutation rate (e.g., Keightley et al. 2009; Schrider et al. 2013; Zhu et al. 2014; Ness et al. 2015; Belfield et al. 2018; Chu et al. 2018; Krasovec et al. 2019; Kucukyildirim et al. 2020; Katju et al. 2022). In mice, MA experiments have also been used to study the impact of mutator genotypes on the mutation rate and the incidence of abnormal phenotypes (Uchimura et al. 2015). However, the vast majority of MA studies do not include samples representing variation within species. In the case of mice, previous studies have been limited to the C57BL/6 strain or crosses derived from it (Adewoye et al. 2015; Uchimura et al. 2015; Milholland et al. 2017; Lindsay et al. 2019). Additionally, there is evidence of mutator alleles introducing mutational biases in mice (e.g., C → A transversions, Sasani et al. 2022, 2024), but the extent of variation of the SNM spectrum among mouse strains is

largely unknown. Comparative genomic analyses have provided information of the SNM spectrum for strains other than C57BL/6 (Dumont 2019). However, without an experiment designed to detect *de novo* mutations, it is difficult to assess the extent by which variation in the inferred spectra is influenced by variants that were segregating in the ancestors of the inbred strains.

To gain insight into variation in the mutation rate and its spectrum, we maintained MA lines of four different inbred mouse strains that are commonly used in genetics research: BALB/cAnNRj, C57BL/6JRj, C3H/HeNRj, and FVB/NRj. The different MA lines were maintained under as similar environmental conditions as possible in order to minimise the impact of non-genetic factors on the mutation rate. In total, our study involved the maintenance of 59 MA lines for more than 4 years (Figure S1), resulting in the largest MA study conducted in mammals to date. Our primary objective was to provide an accurate description of the rates and spectra of new mutations in mice, while offering insights into the causes of variation across strains. We also explore the evolutionary consequences of variation in the SNM spectrum by estimating the predicted effect of mutations in coding sequences.

Results

The rate of new SNMs

We conducted whole-genome sequencing of a total of 59 mouse MA line samples of four different mouse strains (BALBC, BL6, C3H and FVB), in addition to the founder individuals. We sequenced a total of 120.9 Gb of 'callable' sites; these are sites chosen using the same quality criteria that were used to call *de novo* mutations. Approximately 80% of the genome was thus deemed callable. The C3H strain

contributed the majority of these sites, due to its larger sample size (47 MA lines), followed by FVB (5 lines), BL6 (4 lines), and BALBC (3 lines).

A total of 6,563 *de novo* single nucleotide mutations (SNMs) were found. There were 310 SNMs in BALBC, 482 in BL6, 5,041 SNMs in C3H, and 730 in FVB. To estimate the corresponding mutation rates, we employed a mutation-dropping approach into the pedigree to account for 1) the stochastic loss of new mutations segregating in the ancestors of the sequenced individuals and 2) mutations that occurred early in the MA experiment that were shared by several samples, and filtered out by our calling criteria. The mean estimated SNM rates were as follows: $\mu_{\text{BALBC}} = 6.0 \times 10^{-9}$, $\mu_{\text{BL6}} = 6.8 \times 10^{-9}$, $\mu_{\text{C3H}} = 6.3 \times 10^{-9}$ and $\mu_{\text{FVB}} = 7.7 \times 10^{-9}$ (Figure 1). Estimates for three strains are therefore highly consistent, but the estimate for FVB is somewhat higher. Our mutation rate estimation method incorporated the possibility of mutation clustering, i.e., whether or not mutations appearing in a parent's germline can be inherited by multiple offspring. Including clustering led to somewhat lower mutation rate estimates (Table S1). The gene-dropping simulations also provided an expectation for the proportion of heterozygous mutations, and this was in good agreement with the observed proportion (Figure S2). A list of all SNMs found in this study is provided as Supplementary Table S2.

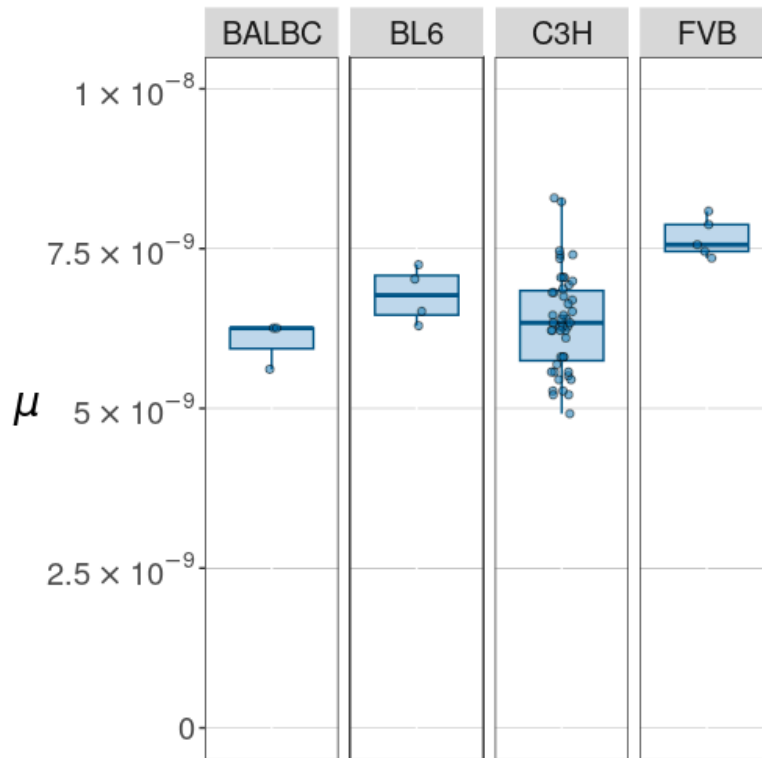


Figure 1. Rates of *de novo* SNM in MA samples of the mouse strains BALBC, BL6, C3H and FVB shown as points and boxplots. Points represent individual MA lines. Mutation rates were estimated assuming no mutation clustering.

Figure 1 suggests that there could be appreciable variation in the mutation rate within strains. To investigate this, we focused on C3H, since this strain had by far the largest number of samples sequenced. We computed the distribution of expected MA line variance of mutation number using computer simulations of the MA experiment. We found that the observed variance of the number of mutations among the C3H lines is close to the mode and within the 95% confidence interval of the distribution of expected variance from independent simulation runs (Figure S3A). This result suggests that mutations in our experiment distribute as expected if there is random accumulation in independent MA lines.

In contrast, overdispersion of mutation rates was previously observed in humans, presumably due to the effect of parental age (Kong et al. 2012). Although there is a two-fold range in the average age at conception of the ancestors of each sequenced C3H mouse (Figure S4), we did not see a significant correlation between the average age of male or female ancestors and the corresponding number of mutations per generation of each sequenced mouse (Figure S5; Pearson's product-moment correlation, $t_{45} \approx -0.69$, $r \approx -0.10$, $P > 0.45$). This lack of correlation is likely due to the narrow range of ages at which the mice were bred. This differs from a previous mouse experiment, which detected a positive correlation (Lindsay et al. 2019). The four strains in our experiment were bred at similar ages (Figure S6), so we only expect a minor effect of parental age on the mutation rate estimates within and between strains.

To assess the extent of variation in SNM rates between strains, we used a linear model with mouse strain included as a random effect. The amount of variation between strains was approximately 70% of the residual variance within strains (1.57 vs. 2.21). However, results from a bootstrapping analysis indicate that there is a large amount of variation for both the within and between-line variance estimates themselves (Figure S3B), suggesting that these two variances are similar in magnitude.

SNM spectra

The spectrum of SNMs was characterised by a strong bias toward C → T transitions, a well-known pattern that is prevalent in both eukaryotes and prokaryotes (Lynch et al. 2023). When considering all the mouse strains together, C → T mutations were approximately three times more frequent than the expectation based on equally

frequent SNM types and a genomic GC content of 41.7%. Figure 2 shows the mutation rate for each SNM type after adjusting the number of callable sites by its genomic GC content. Figure 2 also shows that the average rate for C → T mutations is approximately 7.5×10^{-9} across strains, although variation is observed in the SNM spectra among the four mouse strains investigated (Figure 2). Specifically, FVB samples showed a SNM spectrum that has substantially higher rate of C → A transversions than the other strains (5.9×10^{-9} vs. 2.1×10^{-9} ; Wilcoxon rank-sum test, W test, $W = 270$, $P = 2.44 \times 10^{-4}$). Analysis of variance (ANOVA) followed by the Games-Howell post-hoc test confirmed the presence of differences in SNM spectra between several pairs of strains (Figures S7 and S8), most significantly for differences in the frequency of C → A transversions between FVB and all the other strains ($P < 10^{-3}$).

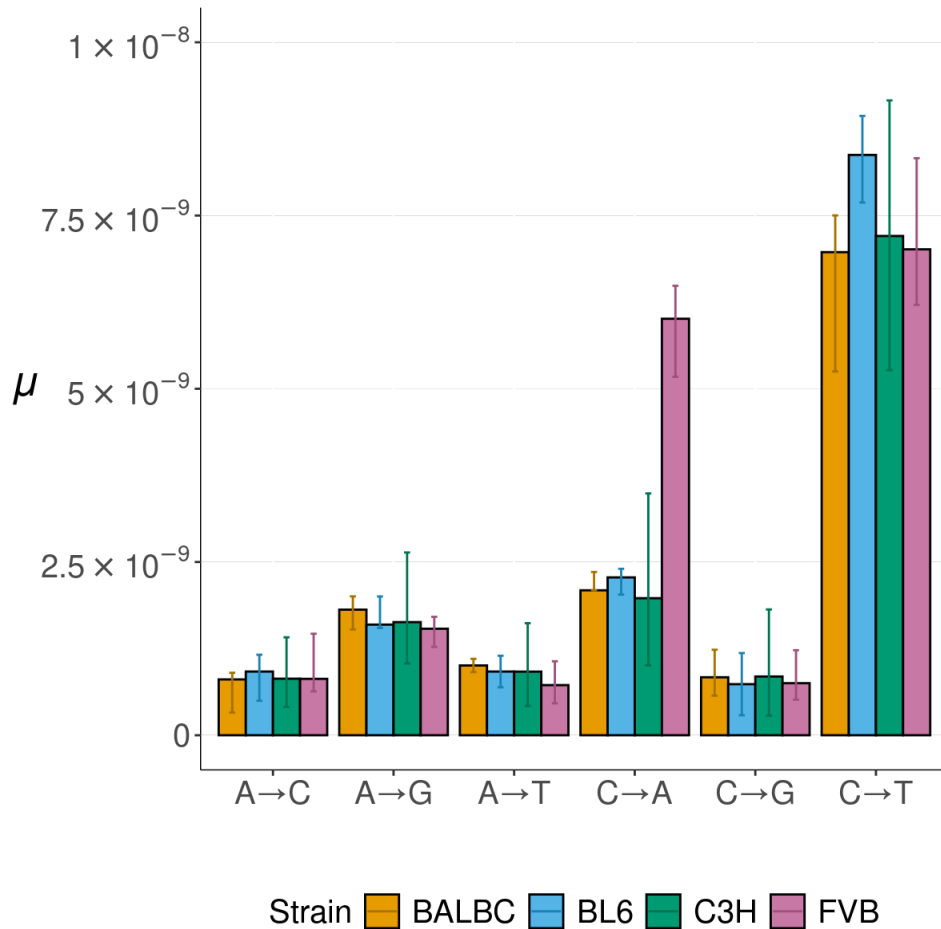


Figure 2. Spectrum of SNMs in four different mouse strains. The heights of the bars represent the median mutation rate (μ) for each type of mutation and strain (in colours). Mutation rates for each SNM type were calculated while adjusting the number of callable sites by their GC content. Error bars represent 95% confidence intervals.

The differences observed among the SNM spectra were strongly influenced by the sequence context. Across MA samples, the mutation rate at G/C sites was more than twice the rate of A/T sites and it was also more variable among strains (Figure S9). However, the FVB strain was an outlier, since G/C sites were more than four times more mutable than A/T sites (W test, $W = 25$, $P = 3.97 \times 10^{-3}$), and this is caused by the C → A transversion bias noted previously. CpG sites had a mutation rate more than 10 times higher than other dinucleotides (W test, $W = 3,481$, $P < 2.2$

$\times 10^{-16}$, Figures S10-S13). CpGs are known for their high mutability to TpG, and accordingly, we observed that more than 85% of cytosine mutations in CpG sites were C \rightarrow T transitions. Furthermore, our results demonstrate that the increased mutation rate at CpG sites extends to the adjacent nucleotides. Nucleotides downstream of a CpG site experienced a 10-fold higher mutation rate than nucleotides adjacent to randomly sampled dinucleotides (W test, $W \geq 601$, $P \leq 1.53 \times 10^{-6}$, Figure S11), and a similar effect was observed in other nucleotide contexts (Figures S12 and S13). In particular, the unusual C \rightarrow A mutation bias in the FVB strain is strongly associated with CpG sites upstream of the mutating C site (W test, $W = 25$, $P = 5.96 \times 10^{-3}$, Figure S12). Such C \rightarrow A mutations experienced a mutation rate approximately 60 times higher than the average for other upstream dinucleotide sequence contexts (Figure S12). This rate is 14 times higher than the next highest mutation rate in the same sequence context, and 185 times higher than the lowest one. Focusing on the C3H strain, since it has the highest number of MA samples sequenced, there was up to 150-fold variation in the mutation rate across dinucleotide sequence contexts and approximately 70-fold variation across trinucleotide contexts.

Predicted effects of mutations

Following the observed variation in the SNM spectra, we explored whether mutations in coding sequences (CDSs) differ in their fitness effects across strains due to the presence of distinct SNM spectra. Simulations were required for this exercise in order to estimate the expected effects of mutations, because only a few SNMs were detected in CDSs: 4 in BALBC, 10 in BL6, 100 in C3H, and 20 in FVB. We first

simulated 10,000 mutations in CDSs while accounting for the strain-specific SNM spectrum in order to assign weights to the mutation rates corresponding to the six different types of SNMs. Subsequently, we annotated the functional class of these mutations (e.g., silent, missense) using SnpEff (Cingolani 2022). This software primarily predicted “low” and “moderate” effects for silent and missense mutations, respectively, whereas nonsense mutations were predicted to have “high” effects. Based on 1,000 simulation replicates, we found that the expected proportion of mutations with a high effect size was substantially higher in FVB than the other strains (i.e., a ~16% increase, Figure 3A). This is mainly a consequence of a more than 2-fold increase in the proportion of nonsense mutations in codons encoding glutamic acid (Glu), associated with C → A transversions (Figure 3B, Figure S14). A higher proportion of high effect size mutations in FVB is also predicted using information from the trinucleotide sequence context (Figure S15). Additionally, the FVB strain experienced a reduction in the expected proportion of mutations of low effect size and a slight increase in the proportion of mutations with moderate effects (i.e., ~1-2% change, Figure 3A). These findings indicate that there is a higher expected burden of deleterious mutations resulting from new mutations in FVB than in the other strains.

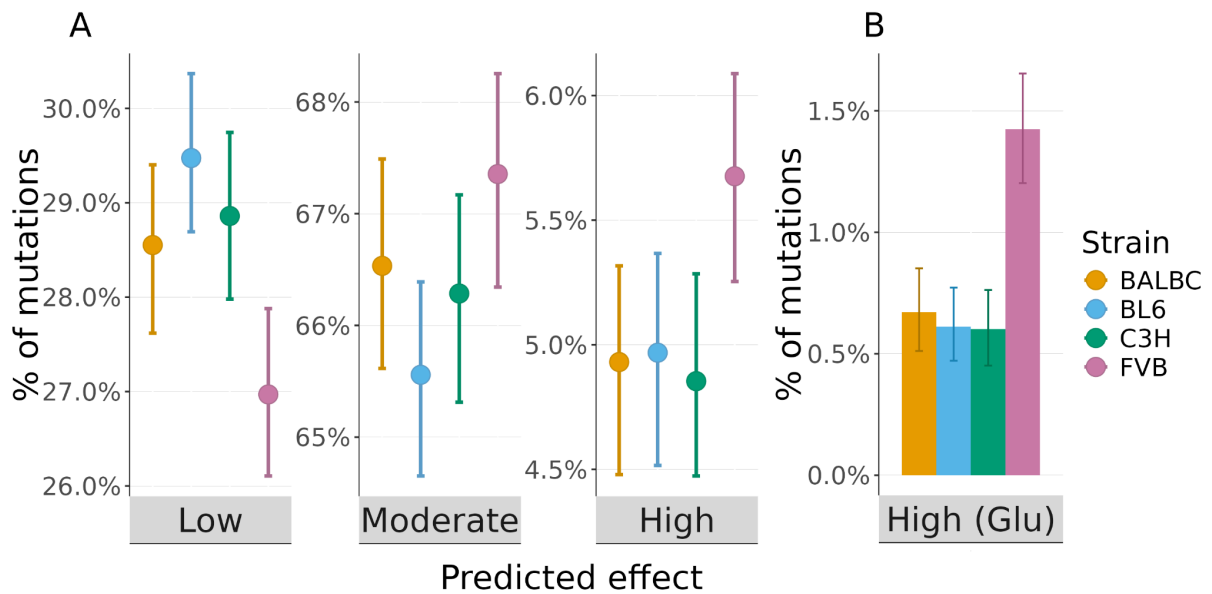


Figure 3. Proportion of mutations simulated in coding sites, grouped by their predicted effect using SnpEff. Median proportions are plotted with error bars showing 95% confidence intervals. A) Proportions for mutations with low, moderate and high effect sizes. B) Proportion of high effect size mutations in glutamic acid (Glu) codons.

Discussion

We conducted a MA experiment with four inbred mouse strains that are commonly used in genetics research, and performed whole-genome sequencing to identify new mutations. We limited our study to the analysis of SNMs in non-tandemly repeated sequences, due to the difficulties in validating variants called from short reads in microsatellites, and the prevalence of insertions and deletions (indels) in these sequences. Previous works based on polymorphism frequency data had suggested that microsatellites are highly prone to indel mutations (Montgomery et al. 2013), and our preliminary results suggested that at least 80% of new indels occur in microsatellite annotations. However, in the future, long-read sequencing technology should allow the accurate estimation of *de novo* indel rates and the mutation rate of tandem repeat sequences. Focusing on SNMs, our results with Illumina short-read

sequencing of 59 MA lines demonstrate that *de novo* mutation rates are similar among strains, but there is variation among SNM spectra. This is consistent with previous studies suggesting that the SNM spectrum evolves faster than the SNM rate (López-Cortegano et al. 2021). Additionally, we show that the spectrum of mutations may affect the distribution of fitness effects (DFE) of new mutations, since variation in the SNM spectra results in variation of the predicted deleterious load of new mutations.

We aligned all the samples to a common reference genome (GRCm38) and we used information from all the individuals sequenced in the mutation calling step. There were two reasons for this: first, it ensured that high-quality “callable” sites were consistently defined in all strains, and second, it prevented our filtering criteria from treating strains differently, e.g., due to variation in their sample sizes. The use of strain-specific reference genomes (Lilue et al. 2018) could have improved the read mapping for each strain, but would have made the comparison of SNM rates among strains more challenging. One limitation of our methods is that our mutation rates are only representative of the fraction of the reference genome that can be mapped with high quality using short-reads, which we estimate to be around 80%, and consists of the least repetitive regions.

For the four strains, the mean SNM rate is $\mu = 6.7 \times 10^{-9}$, a value that is in reasonable agreement with estimates previously reported for inbred mouse strains (reviewed in Chebib et al. 2021). These range from $\sim 3.8 \times 10^{-9}$ in pedigreed families derived from C57BL/6 crosses (Adewoye et al. 2015; Lindsay et al. 2019) to 7.9×10^{-9} estimated from the number of variants segregating in inbred C3H/HeNRj mice

(Chebib et al. 2021). Our results suggest that the SNM rate in mice is slightly greater than 50% of the rate in humans ($\mu = 1.2 \times 10^{-8}$, Kong et al. 2012). However, when considering the SNM rate per DNA replication in the male germline, mice are likely to have a higher rate than humans, since human germ cells undergo more than twice as many cell divisions as mice (Milholland et al. 2017; Ohno 2019).

A growing body of literature suggests that differences in the SNM spectrum are caused both by variation in the DNA repair machinery and by variation in environmental stress (Volkova et al. 2020). In mice, data from genome assemblies suggests that there is substantial variation in the SNM spectrum among 29 inbred mouse strains (Dumont 2019). It is likely, therefore, that alleles modifying the SNM spectrum, such as DNA repair genes, vary among inbred mouse strains (Dumont 2019). Our results on the spectrum of *de novo* SNMs including the presence of a biased C → A spectrum in the FVB strain validate Dumont's observations. It is likely that modifiers of the SNM spectrum segregate in natural and laboratory mouse populations, and became fixed during the establishment of the inbred lines. One hypothetical example could involve a mutator allele that fails to repair oxidatively modified bases, such as 8-oxoguanine, which is associated with an increased rate of C ↔ A transversions (Ohno 2019). Similar mutators appear to segregate in natural populations. For example, a recently described allele in mice appears to shift the SNM spectrum toward C → A mutations (Sasani et al. 2022), and is associated with the mutational signature SBS18 that results in a heritable cancer predisposition syndrome. Comparing our data against the version 3.4 of the Catalog of Somatic Mutations in Cancer (COSMIC, Alexandrov et al. 2020) supports the contribution of the SBS18 signature to the FVB SNM spectrum (Figure S16), and the contribution of

SBS6 to the spectra for the remaining strains. Although not validated by our data, it is plausible that alleles with similar mutator effects, as found by Sasani et al. (2022), became fixed in FVB, leading to a deviation of its SNM spectra and to a somewhat higher SNM rate compared to other strains (Figure 1). However, the contribution of an increased C → A mutation rate in the FVB strain appears to be specific to the sequence context (Figure S12), since SNM rates at A/T sites are broadly consistent among strains ($\sim 3.4 \times 10^{-9}$, Figure S9). Additionally, caution is advised in interpreting differences in mutation rates among strains, since there is substantial variation among the within-strain mutation rate estimates.

Variation in the spectrum of mutations is expected to have evolutionary important consequences. In the reference genome of the inbred strain (C57BL/6J), the GC content is 41.7%, and this is similar to that observed in the genome assemblies of other inbred strains (BALBC, C3H and FVB; Lilue et al. 2018). However, the expected GC content, predicted on the basis of their SNM spectra at mutation equilibrium, is about 22% in BALBC, BL6, and C3H, and as low as 15% in FVB, due to its highly AT-biased SNM spectrum (Figure S17). These values underscore the importance of processes, such as selection and GC-biased gene conversion, that could vary among mouse strains, and possibly among natural populations.

We explored the consequences of alterations in the SNM spectrum on the expected genomic burden caused by new deleterious mutations (Figure 3, Figure S14). We demonstrated that changes in the SNM spectrum could thereby result in changes in the expected effect sizes of mutations. Consequently, we predict that the mutational load may be highly influenced not only by the presence of mutator alleles, but also

by the directional effects imposed by these mutators on the spectrum of new mutations. This observation is supported by previous results in bacteria, showing that alterations in the SNM spectrum create new opportunities for beneficial mutations to emerge (Sane et al. 2023). In other words, the SNM spectrum shapes the DFE for new mutations. For example, C → A mutators are likely to contribute to the deleterious load and to be strongly selected against, since they increase the frequency of nonsense mutations in the Glu-encoding codons GAA and GAG. These are among the five most prevalent codons in the mouse and human genomes (Alexaki et al. 2019). Even in the case of synonymous mutations, changes in the SNM spectrum could modify codon composition, leading to changes in selective pressures during RNA translation associated with codon usage bias (Hanson and Coller 2017) or selection on splicing signals (Parmley et al. 2006). Therefore, we emphasise the importance of a comprehensive description of the spectrum of *de novo* mutations when interpreting mutation rates. Arguably, not only does the mutation rate evolve as a quantitative trait, but also the spectrum of mutations also evolves, and its variation can impose differential selective pressures and introduce constraints on the evolution of protein sequences and genomes.

Methods

Biological material

One pair of mice of the C3H/HeNRj (C3H) inbred strain was obtained from the colony nucleus of Janvier Labs (<https://janvier-labs.com>), and pairs of mice from each of the three inbred strains BALB/cAnNRj (BALBC), C57BL/6JRj (BL6), and FVB/NRj (FVB) were obtained from Janvier Labs' production colonies (BALBC and

BL6 as live mice, and FVB as embryos). The colony nucleus is maintained by full-sib mating, but this is not necessarily the case for mice from production colonies, and therefore the mice from a given pair may be genetically more different from each other than expected from a line maintained by full-sib mating (Chebib et al. 2021). Full-sib matings were carried out using the offspring of each founder pair to establish MA lines for each strain. Four generations of full-sib mating were required to produce 47 independent lines of the C3H MA lines, and three generations were required to produce 3, 4, and 5 MA independent lines of BALBC, BL6, and FVB, respectively.

Each MA line was maintained by randomly selecting three sibling pairs of one female and one male from the offspring of a mating in one generation. These pairs were mated to create litters of the next generation. Occasionally, when a particular litter did not contain offspring of both sexes, a mating pair was chosen from another litter from the same MA line. In cases where it was not possible to choose full-sibs, first cousins of the opposite sex from different matings of the same line would be selected for mating. Pedigrees for each of the four mouse strains are shown in Figure S1. All mice were bred within the facilities of the Max Planck Institute for Evolutionary Biology, Plön, Germany. To limit the influence of environmental factors, mice from all strains were maintained under similar conditions. Mice were housed in Green Line GM500 IVC cages from Tecniplast (<https://www.tecniplast.it>) with water, food consisting of '1328 forti' pellets from Altromin (<https://altromin.de>), aspen bedding from Rettenmaier (Germany), nesting material and shelter. Water and food were given *ad libitum*. All material for maintenance (i.e. cages, bedding, water, food) were autoclaved before use and cages were changed weekly. The mice were kept at a temperature of $22^{\circ}\text{C} \pm 2^{\circ}\text{C}$, at 55-60% humidity, and room ventilation of 16 turnovers per hour, and were always handled under an air ventilated clean bench.

The mice facility is regularly germ tested by sentinels. Additionally, mice from all strains were bred at a young age and within a narrow range of ages to limit the effect of parental age on mutation rates (Figure S6). Maintenance and handling procedures adhered to both the German animal welfare law (Tierschutzgesetz) and the Federation of European Laboratory Animal Science Associations (FELASA) guidelines. The necessary permits for housing and caring for the mice were acquired from the local veterinary office “Veterinäramt Kreis Plön” (permit number: 1401-144153-5.2.3).

DNA extraction, sequencing and alignment

DNA was extracted from liver tissue of one randomly chosen male mouse from each MA line using a standard salt extraction method, which included an initial Proteinase K digestion step.

Whole genome sequencing of the genomic DNA from 67 mice was performed using the Illumina NovaSeq Platform at Edinburgh Genomics (Edinburgh, UK). The sequencing libraries were generated using a PCR-free approach. This yielded an average ~30× coverage of 150-bp paired-end sequences for each sample. Reads were aligned to the *Mus musculus* reference genome (GRCm38.p6) using BWA mem v0.7.13-r116 (Li and Durbin 2009). Subsequently, alignment data for each individual were processed through the following pipeline. The reads were sorted using Samtools v1.9 (Li et al. 2009), the read mate-pair information was synchronised using ‘fixMateInfo’ from the Picard Tools v2.2 suite (Broad Institute 2019), read groups were replaced using ‘setReadGroups’, and duplicate reads were marked using ‘markDuplicates’ from Picard Tools. The processed data were indexed

using Samtools. After the alignment processing, variant calling was performed for individual samples using HaplotypeCaller from GATK v4.1.2.0 (Van der Auwera and O'Connor 2020). The variant calling included options to enable calling at variant and invariant sites (using the option “--emit-ref-confidence BP_RESOLUTION”). HaplotypeCaller also used option “--pcr-indel-model NONE”, as recommended by the Broad Institute for calling insertions and deletions (indels) from PCR-free sequencing data. Variant calls from all MA samples and founder individuals across strains were then combined into one variant call format (VCF) file per strain using GATK’s CombineGVCFs. The final sets of variants were called from these VCFs, together with invariant sites, with GATK’s GenotypeGVCFs using the option “--include-non-variant-sites”.

Identification of new variants in high-quality genomic sites

Candidate *de novo* mutations were identified from the set of sites where variation was unique to one MA sample and where variation was absent from any of the founder mice. Each candidate variant was further filtered according to the following criteria:

1. Phred-scaled quality score for the variant (QUAL) ≥ 30 .
2. The read depth of every sample ≥ 10 .
3. The read depth of every sample < 60 .
4. If the mutation was called heterozygous, the proportion of reads supporting it was in the range [0.25, 0.75].
5. The total number of variant reads in non-mutated MA samples did not exceed 25 (Figure S18).

These criteria were coded into Cython scripts (see Data Availability), which incorporated the Python wrapper cyvcf2 0.30.18 (Pedersen and Quinlan 2017). The variant sites that passed the above criteria were then subjected to a manual verification using the Integrative Genomics Viewer v2.16.0 (IGV, Robinson et al. 2011). Mutations were rejected if they lacked unambiguous support from the read alignments or were not unique to a single MA line. In addition to the above requirements, we filtered candidate mutations according to the following criteria.

6. Variants are not in phase with other variants.
7. Variants in sex chromosomes are homozygous.
8. The variant site has no more than 2 alleles.
9. Variants do not have more than 1 read whose read pair was aligned to another chromosome.

Haplotype phase distance between variants in criterion 6 was determined by GATK v4.1.2.0 HaplotypeCaller “active site” defining algorithms (Poplin et al. 2018). Criteria 3 and 6 through 9 were specifically intended to filter out false positives due to misaligned paralogous reads. Many regions containing misaligned paralogous reads were recognizable because they tended to contain groups of linked variants in phase.

To estimate mutation rates with single nucleotide precision, we defined a fraction of the genome as ‘callable’ for each strain. Here, the callable genome was determined following the filtering criteria 1 to 3 defined above, which can be applied to invariant as well as variant sites, so that the callable genome has an equivalent quality as required to detect mutations. The callable sites were further restricted to the autosomes and to the Chromosome X. The mouse Chromosome Y is dominated by

the presence of ampliconic genes (Morgan et al. 2017), and was excluded from our callability criteria, since mutations were effectively unmappable.

Insertion and deletion variants were initially included in our data set. However, validating indels in short-read sequencing data posed a significant challenge owing to their widespread occurrence in microsatellite annotation (see below). In addition, there were many instances of tandem repeats that failed to be annotated as such. Therefore, we focused on single nucleotide variants occurring outside of microsatellite annotations, following previous studies on *de novo* mutation in humans and mice (Kong et al. 2012; Lindsay et al. 2019).

Calculation of the mutation rate

The mutation rate (μ) was calculated for sample i of strain j as $\mu_{i(j)} = (N_{\mu(i)} \times \alpha) / (2N_c \times N_{a(j)})$, where $N_{\mu(i)}$ is the total number of homozygous and heterozygous mutations found in sample i , N_c is the length of the haploid callable genome, $N_{a(j)}$ is the number of ancestors in j 's pedigree, and α is a correction factor required to account for two possible sources of bias affecting $N_{\mu(i)}$. First, a mutation that occurred early on in the pedigree could be lost in unsequenced ancestors of the MA samples. Second, our filtering criteria only considered mutations unique to a single MA sample (see above), so genuine mutations present in more than one MA line that occurred early on in the experiment would be excluded. The correction factor was obtained by dropping mutations into each strain's pedigree and simulating their segregation following the rules of Mendelian inheritance. This was done using a custom script 'sim-ped.pl' v.1.0.3 (see Data Availability). For each simulation iteration, the program recorded whether a mutation was recovered in multiple or just one of the sequenced

individuals, and when it was present in only one individual, whether it was in the heterozygous or homozygous state. The correction factor was calculated simply as the reciprocal of the proportion of the total number of homozygous and heterozygous mutations uniquely recovered in one of the sequenced samples. To make the simulations more realistic, a clustering parameter was included specifying whether mutations occurring in a parent's germline could be inherited by multiple offspring (with clustering) or not (without clustering). Unless otherwise stated, our results assume no mutation clustering. However, it should be noted that a previous study observed that up to 18% of new SNMs in mice are clustered (Lindsay et al. 2019). Table 1 shows the expected proportions of heterozygous and homozygous mutations, and the applied correction factor. Point estimates of the correction factor were obtained for each pedigree (i.e. strain) by running a large number of gene-dropping simulations (10^6 replicates, Table1). Using a large number of replicates ensures that point estimates of the correction factor are accurate for each pedigree. By simulating 100 correction factors for each strain as described above, we calculated that the interquartile range was less than 10^{-2} in all cases and similar across strains (6.4×10^{-3} for BALBC, 5.0×10^{-3} for BL6, 5.2×10^{-3} for C3H, and 5.8×10^{-3} for FVB).

Table 1. Parameters used in the calculation of the mutation rate. For each strain, the total number of individuals in the MA pedigree (N_{ped}) is given. Additionally, the proportions of heterozygous [E(%Het)] and homozygous [E(%Hom)] mutations expected to be observed, based on gene-dropping simulations, along with the correction factor (α), are given for cases with and without mutation clustering.

Strain	N_{ped}	With clustering			Without clustering		
		E(%Het)	E(%Hom)	α	E(%Het)	E(%Hom)	α
BALBC	39	0.312	0.072	2.60	0.290	0.040	3.04

BL6	52	0.295	0.105	2.50	0.285	0.063	2.88
C3H	618	0.309	0.077	2.59	0.296	0.042	2.96
FVB	72	0.277	0.115	2.55	0.267	0.070	2.97

Variation in the mutation rate

To investigate the extent of variation in the mutation rate among MA lines, we used a simulation approach to compare the observed variation in number of mutations among MA lines with the variation expected based on the random accumulation and inheritance of mutations. We focused on the C3H MA lines since this strain has by far the largest amount of data. We used the software SLiM 4.2.2 (Haller and Messer, 2023) and input the C3H pedigree in order to replicate the MA experiment. The simulated genome had a length equal to that of the callable genome (2,049,374,058 bp). Mutations were randomly introduced into each offspring's genome following a Poisson distribution with rate parameter λ equal to the mutation rate estimated here ($\mu_{\text{C3H}} = 6.3 \times 10^{-9}$) multiplied by the two times the length of the callable genome. From the output returned by SLiM, mutations were filtered so that they were unique among the sequenced MA samples, as in the actual experiment. For each 10,000 replicates, we calculated the variance of the number of mutations among MA lines. The SLiM script is available on GitLab (see Data Availability).

Correlations between number of new mutations and ages of parents

To test whether there was an effect of parental age on the number of *de novo* mutations, we estimated the age at conception of the ancestors of each of our

sequenced mice. We did this by subtracting 21 days from the date of birth of the offspring (21 days is the average length of pregnancy in mice). We then subtracted the date of birth of the parent from the conception date to obtain the age at conception in days. For each sequenced mouse (i.e. each line) we averaged the age at conception of its ancestors. We then performed a linear regression in R (R Core Team 2024) of the number of new mutations found in a line (divided by the number of generations separating the sequenced individual from its founders) against the average age at conception of that line. This analysis was limited to the C3H MA lines because we do not have sufficient lines to obtain useful estimates for the other strains (but see Figure S6 for the distributions in age at conception for the four strains).

Genome annotation

We annotated microsatellites in the GRCm38 genome using the software Tandem Repeats Finder (Benson 1999) with options “2 7 7 80 10 50 500 -f -d -m -h -ngs”. In addition, we ran the software RepeatMasker v4.1.2 (<https://www.repeatmasker.org>) using the default library for mice (option “-species ‘Mus musculus’”), to include additional repeat annotation. All annotation files were processed into BED format and manipulated using the programs provided in the BEDTools suite v2.30.0 (Quinlan and Hall 2010).

Mutational effects

To investigate how variation in the SNM spectrum might impact the effects of mutations on fitness, we used a simulation approach that samples mutations from each strain’s SNM spectrum, and annotates their predicted effects on CDSs with the software SnpEff v5.1 (Cingolani 2022). This approach consisted of three steps: 1)

random sampling of single nucleotide coordinates in CDSs, 2) introduction of synthetic mutations using information from the SNM spectrum, and 3) annotation of the predicted effect of mutations. These steps are described in more detail below. A pipeline to reproduce them is available on GitLab (see Data Availability)

For each strain, we first sampled a large number of random nucleotide sites in the GRCm38 mouse genome using 'bedtools random' with options '-n 20000 -l 1'. We then used 'bedtools shuffle' to resample these coordinates, ensuring they were constrained to CDS annotations. Coding sequences were extracted from the Ensemble annotation file of the GRCm38 reference with release number 102 ('Mus_musculus.GRCm38.102.chr.gff3.gz'), by extracting annotations of type "CDS". The BED file containing random sites within CDSs was then sorted, and reference nucleotides were added to the annotation using 'bedtools getfasta' with the option '-bedOut'.

We then used a custom R script to introduce synthetic mutations in the annotated BED file (see Data Availability), using a mutation table containing the count of each type of SNM. From this table, the script calculates the probabilities of the different SNM types. These probabilities were calculated separately for reference adenosine ('A') and cytosine ('C'). Based on the observed frequency of GC sites in the annotated BED file, genomic coordinates were downsampled to the desired number of mutations (10,000). It should be noted that this simulation approach is based on point estimates of the SNM spectrum (the raw count of SNM types), and does not account for variation in the SNM spectra estimates, e.g., due to different sample sizes. Analyses were performed using R v.4.3.3 (R Core Team 2024).

Finally, variant annotation for the predicted effect of mutations on coding sequences was performed using SnpEff with a database for the mouse GRCm38.102 genome. The option “-formatEff” was used to incorporate codon and amino acid change annotation, as well as the default predicted functional class of mutations (e.g., silent, nonsense, etc.) and their expected effect impact (e.g., low, high, etc.).

Acknowledgements

Analyses performed here made use of the high performance computing resources at the Ashworth Compute Co-operative Cluster (AC3) at the Institute of Ecology and Evolution of the University of Edinburgh. This project has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement no. 694212).

Data Availability

Raw FASTQ files will be deposited at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1017978. The python scripts for extracting unique mutations from a VCF file, ‘MutFinder_allbait.py, and determining the number of callable sites from a VCF file, ‘SiteFinder_nobait.py’, are available on github:

<https://github.com/jchebib/MutFinder.git>. The Perl script ‘sim-ped.pl’ used to estimate from gene-dropping simulations is available for download at:

<https://sourceforge.net/projects/sim-ped/>. A pipeline to simulate mutations in coding sequences using information from the SNM spectrum and to predict their effects is available on GitLab: <https://gitlab.com/elcortegano/expLoadSNM>. The SLiM script

used to measure variation in the count of mutations among MA lines is available on GitLab: https://gitlab.com/elcortegano/SLiM_recipes.

References

- Adewoye AB, Lindsay SJ, Dubrova YE, Hurles ME. 2015. The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. *Nat. Commun.* 6(1): 1-8.
- Alexaki A, Kames J, Holcomb DD, Athey J, Santana-Quintero LV, Lam PVN, Hamasaki-Katagiri N, Osipova E, Simonyan V, Bar H, et al. 2019. Codon and codon-pair usage tables (CoCoPUTs): Facilitating genetic variation analyses and recombinant gene design. *Journal of Molecular Biology* 431(13): 2434-2441.
- Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Ng AWT, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, et al. 2020. The repertoire of mutational signatures in human cancer. *Nature* 578: 94-101.
- Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics* 8: 619-631.
- Belfield EJ, Ding ZJ, Jamieson FJC, Visscher AM, Zheng SJ, Mithani A, Harberd NP. 2018. DNA mismatch repair preferentially protects genes from mutation. *Genome Res.* 28(1): 66-74.

- Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, Leger JS, Shao C, et al. 2023. Evolution of the germline mutation rates across vertebrates. *Nature* 615: 295-291.
- Broad Institute. 2019. Picard Toolkit. Available from:
<https://broadinstitute.github.io/picard/>.
- Chebib J, Jackson BC, López-Cortegano E, Tautz D, Keightley PD. 2021. Inbred lab mice are not isogenic: genetic variation within inbred strains used to infer the mutation rate per nucleotide site. *Heredity* 126: 107-116.
- Cingolani P. 2022. Variant annotation and functional prediction: SnpEff. *Methods Mol. Biol.* 2493: 289-314.
- Chu XL, Zhang BW, Zhang QG, Zhu BR, Lin K, Zhang DY. 2018. Temperature responses of mutation rate and mutational spectrum in an Escherichia coli strain and the correlation with metabolic rate. *BMC Evolutionary Biology* 18(1): 126.
- Dillon MM, Sung W, Sebra R, Lynch M, Cooper VS. 2017. Genome-wide biases in the rate and molecular spectrum of spontaneous mutations in *Vibrio cholerae* and *Vibrio fischeri*. *Mol. Biol. Evol.* 34(1): 93-109.
- Dumont BL. 2019. Significant strain variation in the mutation spectra of inbred laboratory mice. *Mol. Biol. Evol.* 36(5): 865-874.
- Haller BC, Messer PW. 2023. SLiM 4: Multispecies eco-evolutionary modeling. *The American Naturalist* 201(5): E127-E139.
- Hanson G, Collier J. 2017. Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology* 19: 20-30.
- Katju V, Konrad A, Deiss T, Bergthorsson U. 2022. Mutation rate and spectrum in obligately outcrossing *Caenorhabditis elegans* mutation accumulation lines

- subjected to RNAi-induced knockdown of the mismatch repair gene *msh-2*.
G3 12(1): jkab364.
- Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* 19(7): 1195-1201.
- Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488: 471-475.
- Krasovec M, Sanchez-Brosseau S, Piganeau G. 2019. First estimation of the spontaneous mutation rate in diatoms. *Genome Biol. Evol.* 11(7): 1829-1837.
- Kucukyildirim S, Behringer M, Williams EM, Doak TG, Lynch M. 2020. Estimation of the genome-wide mutation rate and spectrum in the archaeal species *Haloferax volcanii*. *Genetics* 215(4): 1107-1116.
- Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 109(41): E2774-E2783.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 27(21): 2987-2993.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25(16): 2078-2079.

- Lilue J, Doran AG, Fiddes IT, Abrudan M, Armstrong J, Bennet R, Chow W, Collins J, Collins S, Czechanski A, et al. 2018. Sixteen diverse laboratory mouse reference genomes define strain specific haplotypes and novel functional loci. *Nat. Genet.* 50(11): 1574-1583.
- Lilue J, Shivalikanjli A, Adams DJ, Keane TM. 2019. Mouse protein coding diversity: What's left to discover? *PLoS Genet.* 15(11): e1008446.
- Lindsay SJ, Rahbari R, Kaplanis J, Keane T, Hurles ME. 2019. Similarities and differences in patterns of germline mutation between mice and humans. *Nat. Commun.* 10: 4053.
- Liu H, Zhang J. 2019. Yeast spontaneous mutation rate and spectrum vary with environment. *Current Biology* 29(10): 1584-1591.
- López-Cortegano E, Craig RJ, Chebib J, Samuels T, Morgan AD, Kraemer SA, Böndel KB, Ness RW, Colegrave N, Keightley PD. 2021. De novo mutation rate variation and its determinants in *Chlamydomonas*. *Mol. Biol. Evol.* 38(9): 3709-3723.
- Lynch M. 2011. The lower bound to the evolution of mutation rates. *Genome Biol. Evol.* 3: 1107-1118.
- Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* 17: 704-714.
- Lynch M, Ali F, Lin T, Wang Y, Ni J, Long H. 2023. The divergence of mutation rates and spectra across the tree of life. *EMBO Rep.* 24: e57561.

- Maksimov MO, Wu C, Ashbrook DG, Villani F, Colonna V, Mousavi N, Ma N, Lu L, Pritchard JK, Goren A, et al. 2023. A novel quantitative trait locus implicates Msh3 in the propensity for genome-wide short tandem repeat expansions in mice. *Genom. Res.* 33: 689-702.
- Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. 2017. Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* 8: 15183.
- Montgomery SB, Goode DL, Kvikstad E, Albers CA, Zhang ZD, Mu J, Ananda G, Howie B, Karczewski KJ, Smith KS, et al. 2013. The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. *Genome Res.* 23: 749-761.
- Morgan AP, Pardo-Manuel de Villena F. 2017. Sequence and structural diversity of mouse Y chromosomes. *Molecular Biology and Evolution* 34(12): 3186-3204.
- Ness RW, Morgan AD, Vasanthakrishnan RB, Colegrave N, Keightley PD. 2015. Extensive de novo mutation rate variation between individuals across the genome of *Chlamydomonas reinhardtii*. *Genome Res.* 25(11): 1739-1749.
- Ohno M. 2019. Spontaneous de novo germline mutations in humans and mice: rates, spectra, causes and consequences. *Genes Genet. Syst.* 94: 13-22.
- Parmley JL, Chamary JV, Hurst LD. 2005. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol. Biol. Evol.* 23(2): 301-309.

Pedersen, BS, Quinlan AR. 2017. cyvcf2: fast, flexible variant analysis with Python.

Bioinformatics. 33(12):1867-1869. doi: 10.1093/bioinformatics/btx057

Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D,

Dijamco J, Nguyen N, Afshar PT, et al. 2018. A universal SNP and small-indel variant caller using deep neural networks. *Nature biotechnology*, 36(10), pp.983-987.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842.

R Core Team. 2024. R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria.

<https://www.R-project.org/>.

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S,

Domiczak A, Morris A, Porteous D, Smith B, et al. 2016. Timing, rates and spectra of human germline mutation. *Nat. Genet.* 48(2): 126-133.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G,

Mesirov JP. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29: 24-26.

Sane M, Diwan GD, Bhat BA, Wahl LM, Agashe D. 2023. Shifts in mutation spectra

enhance access to beneficial mutations. *Proc. Natl. Acad. Sci. U.S.A.* 120(22): e2207355120.

Sasani TA, Ashbrook DG, Beichman AC, Lu L, Palmer AA, Williams RW, Pritchard

JK, Harris K. 2022. A natural mutator allele shapes mutation spectrum variation in mice. *Nature* 605: 497-502.

- Sasani TA, Quinlan AR, Harris K. 2024. Epistasis between mutator alleles contributes to germline mutation spectrum variability in laboratory mice. *Elife* 21:12:RP89096.
- Schofield PN, Hoehndorf R, Gkoutos GV. 2012. Mouse genetic and phenotypic resources for human genetics. *Human Mutation* 33: 826-836.
- Schrider DR, Houle D, Lynch M, Hahn MW. 2013. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* 194(4): 937-954.
- Sharp NP, Agrawal AF. 2012. Evidence for elevated mutation rates in low-quality genotypes. *PNAS* 109(16): 6142-6146.
- Sun W, Ackerman MS, Dillon MM, Platt TG, Fuqua C, Cooper VS, Lynch M. 2016. Evolution of the insertion-deletion mutation rate across the tree of life. *G3* 6(8): 2583-2591.
- Tiley GP, Poelstra JW, dos Reis M, Yang Z, Yoder AD. 2020. Molecular clocks without rocks: new solutions for old problems. *Trends in Genetics* 36(11): 845-856.
- Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, Miura I, Wakana S, Nishino J, Yagi T. 2015. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res.* 25(8): 1125-1134.
- Van der Auwera GA, O'Connor BD. 2020. Genomics in the cloud: using docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media.

Veltman JA, Brunner HG. 2012. De novo mutations in human genetic disease. *Nat. Rev. Genet.* 13: 565-575.

Volkova NV, Meier B, González-Huici V, Bertolini S, Gonzalez S, Vöhringer H, Abascal F, Martincorena I, Campbell PJ, Gartner A, et al. 2020. Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* 11: 2169.

Wang Y, McNeil P, Abdulazeez R, Pascual M, Johnston SE, Keightley PD, Obbard DJ. 2023. Variation in mutation, recombination, and transposition rates in *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res.* 33(4): 587-598.

Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 111(22): E2310-2318.