

## ARTICLE



# The genome-wide rate and spectrum of EMS-induced heritable mutations in the microcrustacean *Daphnia*: on the prospect of forward genetics

Marelize Snyman <sup>1</sup>, Trung V. Huynh<sup>1</sup>, Matthew T. Smith<sup>1</sup> and Sen Xu <sup>1</sup>✉

© The Author(s), under exclusive licence to The Genetics Society 2021

Forward genetic screening using the alkylating mutagen ethyl methanesulfonate (EMS) is an effective method for identifying phenotypic mutants of interest, which can be further genetically dissected to pinpoint the causal genetic mutations. An accurate estimate of the rate of EMS-induced heritable mutations is fundamental for determining the mutant sample size of a screening experiment that aims to saturate all the genes in a genome with mutations. This study examines the genome-wide EMS-induced heritable base-substitutions in three species of the freshwater microcrustacean *Daphnia* to help guide screening experiments. Our results show that the 10 mM EMS treatment induces base substitutions at an average rate of  $1.17 \times 10^{-6}$ /site/generation across the three species, whereas a significantly higher average mutation rate of  $1.75 \times 10^{-6}$  occurs at 25 mM. The mutation spectrum of EMS-induced base substitutions at both concentration is dominated by G:C to A:T transitions. Furthermore, we find that female *Daphnia* exposed to EMS ( $F_0$  individuals) can asexually produce unique mutant offspring ( $F_1$ ) for at least 3 consecutive broods, suggestive of multiple broods as  $F_1$  mutants. Lastly, we estimate that about 750  $F_1$ s are needed for all genes in the *Daphnia* genome to be mutated at least once with a 95% probability. We also recommend 4–5  $F_2$ s should be collected from each  $F_1$  mutant through sibling crossing so that all induced mutations could appear in the homozygous state in the  $F_2$  population at 70–80% probability.

*Heredity* (2021) 127:535–545; <https://doi.org/10.1038/s41437-021-00478-x>

## INTRODUCTION

Forward genetic screening is one of the most common and effective methods for identifying phenotypic mutants, which can be further genetically dissected to pinpoint the causal genetic mutations. Unfortunately, genetic screening using spontaneous mutations to study genotype-phenotype relationships in eukaryotes is infeasible due to the low mutation rates, which is usually on the order of  $10^{-8}$  to  $10^{-10}$  per nucleotide site (Baer et al. 2007; Lynch et al. 2008; Krasovec et al. 2017). This is because approximately 1/(mutation rate per gene) individuals need to be screened to obtain a mutation in a particular gene (Kutscher and Shaham 2014). Although this number can be reached with bacteria due to their fast reproduction rates and ease of maintenance, it is impractical to get to this large number for multi-cellular eukaryotic organisms with much lower reproduction rates, longer generation times, and purifying selection pressure.

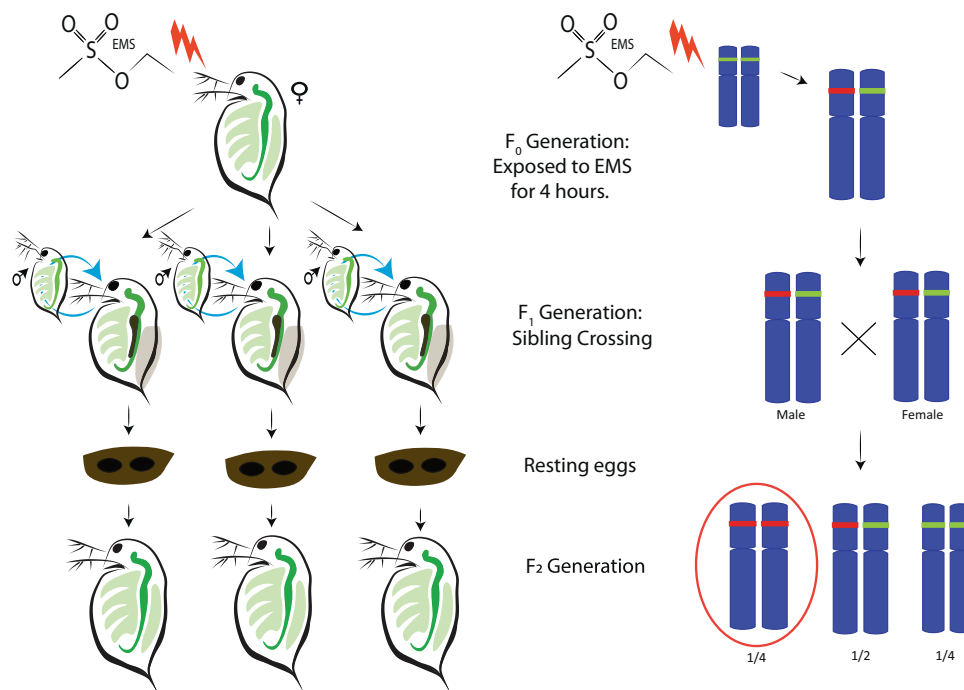
Since the discovery of X-ray induced mutations by H.J. Muller (Muller 1927), mutagens have been used to establish mutagenized screening populations with a manageable number of individuals, while keeping lethality and sterility to a minimum (Kutscher and Shaham 2014). To date a variety of mutagens, e.g., N-ethyl-N-nitrosourea (ENU), trimethylpsoralen with ultraviolet light (UV/TMP), are available to mutagenize both prokaryotes and eukaryotes, allowing researchers to efficiently implement forward genetic screenings.

Among these mutagens, ethyl methanesulfonate (EMS) has been commonly used for genetic screens in many different biological systems (Sega 1984). As an alkylating agent, EMS induces chemical modifications of nucleotides. It was first demonstrated by Brookes and Lawley (1961) that EMS primarily causes guanine alkylation, leading to the formation of O<sup>6</sup> ethylguanine, and results in mutations through mispairings with thymine in DNA replication and repair. Therefore, EMS mutagenesis is heavily biased towards G:C to A:T transitions (Coulondre and Miller 1977). In addition to single-base mutations, EMS has also been shown to cause, to a much lesser extent, indels (insertions and/or deletions) and chromosomal breaks (Sega 1984; Greene et al. 2003). Capable of inducing mutations randomly across the genome (Greene et al. 2003), EMS can therefore be used to generate loss- or gain-of-function mutants as well as weak nonlethal alleles (Lee et al. 2003).

EMS mutagenesis experiments were first done in the T2 viral system by Loveless and Haddow (1959), and was later expanded to *Drosophila melanogaster* (Lewis and Bacher 1968) and *Caenorhabditis elegans* (Brenner 1974). Although EMS mutagenesis has been applied to an increased number of organisms including *Arabidopsis thaliana* (McCallum et al. 2000; Greene et al. 2003; Martin et al. 2009) and *Saccharomyces cerevisiae* (Prakash and Higgins 1982; Mobini-Dehkordi et al. 2008), the list of species with an EMS mutagenesis screening protocol remains limited.

<sup>1</sup>Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA. Associate editor: Darren Obbard. ✉email: sen.xu@uta.edu

Received: 6 April 2021 Revised: 4 October 2021 Accepted: 5 October 2021  
Published online: 19 October 2021



**Fig. 1 Forward genetic approach for obtaining mutant lines in *Daphnia*.** This study used F<sub>1</sub> mutants to determine the mutation rate and spectrum of EMS-induced heritable mutations.

In this work, we aim to develop an effective genetic screening strategy based on EMS mutagenesis for the freshwater microcrustacean *Daphnia*. With world-wide distribution in nearly all kinds of freshwater habitats, *Daphnia* has been studied for more than 200 years (Ebert 2005) and has been a model system in ecology, toxicology, and evolution (Altshuler et al. 2011). As the first crustacean to have its whole genome sequenced (Colbourne et al. 2011), and with the development of new genomic tools, *Daphnia* finds itself with tremendous new momentum in empowering researchers to address many consequential biological questions with genomic insights.

*Daphnia* represents an important pan-crustacean lineage in metazoan evolution. However, about a third of the *Daphnia* genes remain poorly understood for their functions because they are lineage-specific and lack orthologues in other eukaryotic genomes (Colbourne et al. 2011; Ye et al. 2017). Understanding the function of these lineage-specific genes is critical for gaining insights into invertebrate evolution, the genomic adaptation to a freshwater lifestyle, and the genetic basis of novel phenotypes in *Daphnia*. We therefore envision that a forward genetic screening approach would be a valuable tool to aid such efforts.

*Daphnia* typically reproduces by cyclical parthenogenesis, in which they switch between clonal (asexual) and sexual reproduction depending on environmental conditions. Under favorable conditions, females reproduce asexually, producing chromosomally unreduced, diploid embryos that directly develop into genetically identical daughters. These directly developing embryos can develop into males under stressful conditions (e.g., crowding, lack of food). Environmental stress can also induce females to switch and produce haploid eggs through meiosis, which upon fertilization by sperm become diapausing embryos. These diapausing embryos, deposited in a protective case (i.e., ephippium), can hatch under suitable environmental conditions and remain viable for many, often up to hundreds of years (Frisch et al. 2014). Interestingly, some *Daphnia* lineages have transitioned to obligate parthenogenesis (Lynch et al. 2008; Xu et al. 2015). These lineages forgo sex and use parthenogenesis to

produce diapausing embryos under stress, while still asexually producing directly developing embryos in favorable conditions.

Cyclically parthenogenetic reproduction and a few other life history characteristics make *Daphnia* well amenable to large-scale forward genetic screening using EMS mutagenesis. Multiple clonal females of the same genotype (F<sub>0</sub> individuals) can be exposed to EMS, during which mutations can be introduced into the genomes of oocytes (Fig. 1). Females which have been exposed to EMS can then asexually produce mutant female offspring (F<sub>1</sub>s), which would carry all EMS-induced germline mutations in the heterozygous state. Each of the F<sub>1</sub> individuals can be used to propagate genetically identical female and male progenies, and siblings can be crossed (equivalent to selfing) to produce sexual progeny (F<sub>2</sub>s) that carry 25% of the EMS-induced mutations in the homozygous state (Fig. 1). Screening these F<sub>2</sub>s can be performed to identify phenotypes of interest, followed by further genetic analyses to pinpoint the underlying genotypes. Furthermore, the short generation time (7–10 days), large number of broods per female, and easy animal maintenance in lab conditions together make it manageable to screen thousands of *Daphnia* mutant lines.

Implementing this EMS screening strategy requires an understanding of the genome-wide EMS-induced heritable mutation rate in *Daphnia*. Accurate estimates of this rate allow us to gauge the number of mutagenized individuals that are needed to reach a saturation point where nearly every gene in the genome has been mutated a few times. However, no studies have examined EMS-induced mutations in *Daphnia*. We therefore set out to perform a series of experiments to investigate the genome-wide EMS-induced heritable mutation rate and spectrum in *Daphnia*.

Our experiments mainly test three hypotheses that can have major impact on the screening design. First, we hypothesize that a higher, non-lethal concentration of EMS causes a higher germline base-substitution rate than a lower concentration, while the mutation spectrum between different concentrations remains similar due to the mutagenic properties of EMS. Understanding the impact of EMS concentrations on base-substitution rates can help determine how we can most efficiently introduce the desired

number of mutations into the screening population. In this study, we test the mutational effect of 10 vs 25 mM EMS solution.

Second, we hypothesize that the different broods of the same female are affected by EMS-induced mutations in an independent manner (e.g., location of mutations, mutation rate). In *Daphnia* females, all the primary oocyte nuclei are deposited in the germarium at the posterior end of ovary (Kato et al. 2012). Once exposed to an EMS solution, all the primary oocyte nuclei could be independently mutagenized by EMS. If this hypothesis holds true, it would mean that the different broods of females exposed to EMS can all be used to establish mutant lines. To this end, we specifically examine the heritable mutations in the first, second, and third broods of EMS-treated *Daphnia* females.

Third, we hypothesize that the EMS-induced heritable base-substitution rate should be highly similar between different *Daphnia* species. Although the spontaneous mutation rate in different *Daphnia* species/population may evolve to different levels largely due to their varying population genetic environments (Ho et al. 2020), the EMS-induced mutation rate is most likely highly similar between species because the EMS concentration and means of exposure are most likely the greatest determinants of the induced mutation rate. To test this hypothesis, we examined EMS-induced mutations in three species (cyclically parthenogenetic *D. pulex*, *D. pulicaria*, and obligately parthenogenetic *D. pulex*) and multiple genotypes from different populations in each species.

Lastly, based on our results of the EMS-induced mutation rate in *Daphnia*, we performed a power analysis of experimental design (e.g., number of required  $F_1$ s and  $F_2$ s) for genetic screening in *Daphnia* to guide screening efforts in the future.

## MATERIALS AND METHODS

### Experimental animals

A total of three cyclically parthenogenetic (CP) *Daphnia pulex* isolates (Tex21, SW4, and Povi4), three obligately parthenogenetic (OP) *Daphnia pulex* isolates (DB4-1, DB4-2, and DB4-4), and three CP *Daphnia pulicaria* isolates (AroMoose, RLSD26, and Warner5) were used in this study.

These isolates were previously collected from various pond and lake populations across the US and Canada (Supplementary Table S1). They have been kept in the lab as clonally reproducing lines in artificial lake water (Kilham et al. 1998) under a 16:8 h (light:dark) cycle at 18 °C. We fed them with the green algae *Scenedesmus obliquus* twice a week.

### Determining tolerable EMS concentrations

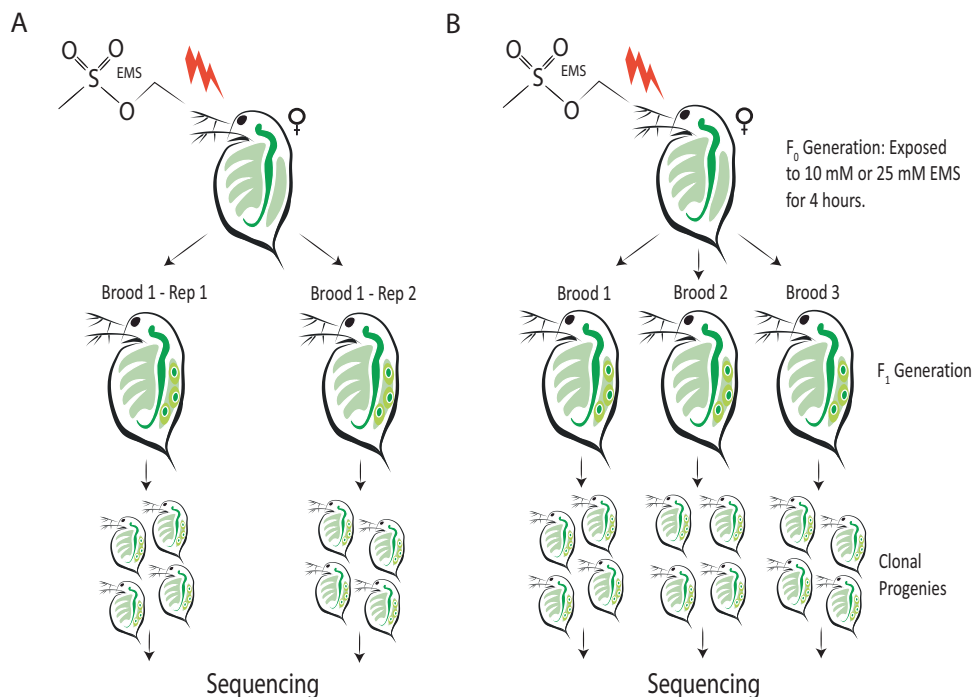
Survival experiments were performed at four different EMS concentrations (i.e., 10, 25, 50 and 100 mM) to determine a tolerable range for *Daphnia* females. Since no prior data were available regarding EMS tolerance in *Daphnia*, these four concentrations were established by referencing standard EMS mutagenesis protocols in other model organisms such as *C. elegans* and *D. melanogaster*. For *C. elegans*, the standard mutagenesis protocol entails exposure to 50 mM EMS for 4 h (Brenner 1974) to achieve a mutation rate of  $2.5 \times 10^{-3}$  per gene per generation (Gengyo-Ando and Mitani 2000), while *D. melanogaster* is fed 25 mM EMS (Lewis and Bacher 1968) to achieve a mutation rate of  $1 \times 10^{-3}$  per gene per generation (Greenspan 1997).

We tested these concentrations on mature females from the three OP *D. pulex* isolates (DB4-1, DB4-2, and DB4-4). Three replicates of ten females from each isolate were simultaneously placed in 1 mL of EMS solution at the concentrations of 10, 25, 50, and 100 mM, respectively. The exposure lasted four hours. The treated animals were then transferred to artificial lake water, and the survival rate was recorded after 24 h.

Although no animals survived the 4-hour exposure to 50 and 100 mM EMS, 100% and 60% of females survived the 10 and 25 mM treatments, respectively (see Results and Supplementary Table S2). Therefore, we used these two concentrations in our subsequent mutagenesis experiments.

### Establishing EMS mutant lines

To examine the rate and spectrum of heritable mutations induced by 10 and 25 mM EMS treatments, sexually mature *Daphnia* females from each isolate were exposed to these two concentrations for 4 h, respectively. For females of each isolate, the exposed animals ( $F_0$  individuals) were isolated and kept individually in benign laboratory conditions. The first brood of asexually produced progenies ( $F_1$ s) from the  $F_0$ s were then collected and individually isolated because  $F_1$ s are derived from oocytes whose DNA may be differentially mutagenized by the EMS. For each natural *Daphnia* isolate at each concentration, we established two replicate mutant lines by growing two different  $F_1$ s clonally, with each  $F_1$  propagating into a mass asexual culture (Fig. 2A). These asexual progenies were whole-genome



**Fig. 2** Experimental procedure for establishing EMS mutant lines. **A** Establishing replicate mutant lines of a *Daphnia* isolate. **B** Establishing brood-specific mutant lines of a *Daphnia* isolate.

sequenced to detect EMS-induced heritable mutations that occur in the germline of  $F_0$  individuals.

Furthermore, to understand whether the EMS-induced mutation rate and spectrum differed between consecutive broods of the same  $F_0$  females, EMS mutant lines were established with the same procedure as above using one  $F_1$  from the first-brood (BR1), second-brood (BR2) and third-brood (BR3) at both 10 and 25 mM EMS treatment (Fig. 2B). We examined the brood effects in three isolates, AroMoose (*D. pulicaria*), Tex21 (CP *D. pulex*), and DB4-4 (OP *D. pulex*).

### Whole-genome sequencing of EMS mutant lines

We collected a total of 40–50 clonal offspring of each EMS mutant line for DNA extraction using a CTAB (Cetyl Trimethyl Ammonium Bromide) method (Doyle and Doyle 1987). The concentrations of the DNA samples were measured using a Qubit 4.0 Fluorometer (Thermo Fisher), and DNA quality was checked by electrophoresis on a 2% agarose gel. DNA sequencing libraries were prepared by the Novogene Company following standard Illumina sequencing library protocol. Each library was sequenced on an Illumina Novaseq 6000 platform with 150-bp paired-end reads, with a targeted sequencing coverage of 30x per mutant line.

### Computational pipeline for identifying mutations

Our computational pipeline for identifying mutations was constructed by incorporating the strengths of mutation calling procedures from previous *Daphnia* mutation accumulation studies (Keith et al. 2016; Flynn et al. 2017; Bull et al. 2019). We used the Burrows-Wheeler Alignment Tool BWA-MEM version 0.7.17 (Li and Durbin 2010) with default parameters to align the raw reads of each mutant line to either the *Daphnia pulex* (Ye et al. 2017) or *D. pulicaria* (Jackson et al. 2021) reference genome. SAMtools (Li et al. 2009) was used to remove reads that mapped to multiple locations in the genome and retain only uniquely mapped reads for downstream analyses, which helps to reduce false positive calls of mutations. The MarkDuplicates function of Picard tools (<http://broadinstitute.github.io/picard/>) was used to locate and tag PCR duplicates. We used the mpileup and call functions of BCFtools (Li 2011) to generate genotype likelihoods and genotype calls in a VCF file containing all EMS mutant lines derived from each natural *Daphnia* isolate. Default parameters were used for BCFtools mpileup and call functions. In addition, we added the following FORMAT and INFO tags to the VCF file: AD (allelic depth), DP (number of high-quality bases), ADF (allelic depth on forward strand) and ADR (allelic depth on reverse strand). We also used the filter function of BCFtools to retain only biallelic single nucleotide polymorphisms sites (SNPs) with a quality score (QUAL)  $\geq 20$ , sequencing depth (DP)  $\geq 10$ , and a distance  $\geq 50$  bp from an indel in each mutant line. We did not examine indels because previous work have shown a very low rate of EMS-induced indels, with  $< 2.8$  deletions and  $< 0.6$  insertion per mutant line (Flibotte et al. 2010; Shiwa et al. 2012; Henry et al. 2014).

A custom python script (all scripts in this study are available at [https://github.com/Marelize007/EMS\\_mutagenesis\\_daphnia](https://github.com/Marelize007/EMS_mutagenesis_daphnia)) was used to identify mutations using a consensus method. We generated one VCF file consisting of the genotype data of all EMS mutant lines derived from the same natural *Daphnia* isolate. For each SNP site, we established the consensus genotype call (i.e., genotype of natural isolate) using a majority rule. With  $N$  samples in a VCF file, the consensus genotype of a site needs to be supported by at least  $N-1$  samples. If an EMS line shows a genotype different from the consensus genotype, a tentative mutation is identified.

This approach allowed us to detect only mutations that were unique to one EMS line and were not shared between multiple lines derived from the same isolate. The rationale of this approach is that because EMS induces mutations at random locations in the genome, with a sample size of no greater than 10 mutant lines per natural isolate and a 200-Mb *Daphnia* genome size, it is highly unlikely that two lines would have mutations at the same site.

We further examined these tentative mutations to establish the final pool of mutations using two criteria. First, a mutant allele must be supported by at least two forward and two reverse reads to avoid false positives due to sequencing error. Second, a mutant genotype is recognized only when it is a heterozygous genotype derived from a homozygous consensus (i.e., wildtype) genotype. This criterion is to avoid false positives caused by allele drop due to insufficient sequence coverage or artifacts in library construction at heterozygous sites. We note that this criterion excludes less than 2% of genomic sites from our analyses as heterozygosity in natural *Daphnia* isolates is about 1–2% (Lynch et al. 2017).

### Mutation validation with Sanger sequencing

To evaluate the robustness of our mutation calling pipeline, Sanger sequencing was used to verify 20 randomly selected mutations from our final pool of mutations. Primers were designed using Primer 3 (Rozen and Skaletsky 1999) in order to amplify a 300–400 bp region of DNA centered at a mutation. We performed PCR on the genomic DNA of the mutant line from which the mutations were identified. BigDye Terminator v3.1 (ThermoFisher) was used for the sequencing reactions on the PCR amplicons, and the sequencing reaction products were sequenced on a 3130xL Genetic Analyzer (Applied Biosystems) at the Life Science Core Facility, University of Texas at Arlington. We examined the electropherograms in SnapGene® Viewer (GSL Biotech) to determine whether the Sanger genotype of the mutation site matches the genotype from our whole-genome sequencing data.

### Mutation rate calculation

The per site per generation mutation rate was calculated for each mutant line using the formula  $\mu = m/n * 1$ , where  $m$  represents the total number of mutations detected in each line,  $n$  is the total number of genomic sites with  $\geq 10$  coverage and 50 bp distance away from an indel in each line, and 1 represents one generation. This equation likely results in an underestimated mutation rate because the number of total sites is not subject to as much filtering as the mutations. The per gene per generation mutation rate was calculated using the following formula  $\mu_g = m_g/n_g * 1$ , where  $m_g$  is the total number of mutations located within gene regions (including UTRs, introns, and exons) in each line,  $n_g$  is the total number of genes analyzed in each line, and 1 represents one generation. To calculate the non-synonymous mutation rate per gene per generation, the same formula was used, with  $m_g$  representing the number of non-synonymous mutations.

### Annotating effect of EMS-induced mutations

We used the cancer mode (-cancer) with default parameters of SnpEff version 4.0 to functionally annotate mutations and predict their effects (Cingolani et al. 2012). This mode allowed us to directly compare the mutant genotypes against the wildtype genotypes and infer the genomic effects of the mutations.

### Sequence motifs of EMS-induced mutations

To examine whether any sequence motifs are over- or under-represented surrounding the mutated sites, we performed a sequence motif enrichment analysis. We extracted the 3-bp sequence centered at the mutated sites (5'-3' orientation) from the 10 and 25 mM datasets, divided them into four groups based on the mutated site (NAN, NTN, NCN, and NGN), and calculated the observed frequency of the 16 motifs in each group in the *D. pulex* and *D. pulicaria* reference genome using Compseq (<http://emboss.open-bio.org/rel/rel6/apps/compseq.html>). The expected number of EMS-induced mutations for each motif under a random distribution hypothesis was calculated as the product of the total number of EMS-induced mutations from all mutant lines (10 and 25 mM) and the observed trinucleotide frequency. We then performed a chi-square test on each motif to test whether its observed number of mutations deviates significantly from the expectation under a random distribution hypothesis with Bonferroni-corrected  $p$ -values.

### Mutagenesis power analysis

Using our estimated EMS-induced heritable per-gene mutation rate, we calculated the probability for finding at least 1  $F_1$  animal heterozygous for a mutation in a gene of interest, using the equation  $1 - (1 - r)^n$ , where  $r$  is the per gene mutation rate and  $n$  is the number of  $F_1$ s (Shaham 2007). The  $(1 - r)^n$  term denotes the probability for none of the sampled  $F_1$ s carrying a mutation at the gene of interest.

## RESULTS

### *Daphnia* survival rate after EMS treatment

One of the major effects that EMS exposure had on *Daphnia* was survival. We obtained the survival rate for three OP *Daphnia pulex* isolates (DB4-1, DB4-2, DB4-4) exposed to EMS concentrations of 10, 25, 50mM and 100 mM for four hours. All *Daphnia* exposed to 50 and 100 mM EMS died during or after treatment (within 24 h).

At lower EMS concentrations, 100% of the animals treated with 10 mM EMS survived, whereas the average survival rate was 60.0% (SD = 8.8%) for animals treated with 25 mM EMS (Supplementary Table S2).

### Whole-genome sequencing data

A total of 43 *Daphnia* mutant lines derived from 10 or 25 mM EMS treatment were whole-genome sequenced with 150 bp Illumina paired-end reads (Supplementary Table S3 and Table S4). A total of ~6GB of raw sequence data was obtained for each mutant line. Each line had on average ~35 million mapped reads after removing PCR duplicates and reads that mapped to multiple locations, yielding an average coverage of 26 (SD = 3) reads per site in each line.

### Mutation validation using Sanger sequencing

Among the EMS-induced germline base substitutions identified using our computational pipeline (see below), 20 were selected for Sanger sequencing verification. We confirmed that all the selected mutations had concordant genotype calls between the Sanger sequencing and Illumina whole-genome data. This suggests that our computational pipeline for identifying EMS-induced germline mutations was robust and that the false positive rate in our dataset was low, most likely  $\ll 0.05$  (i.e.,  $\ll 1$  false positive out of 20 mutations, Supplementary Table S11).

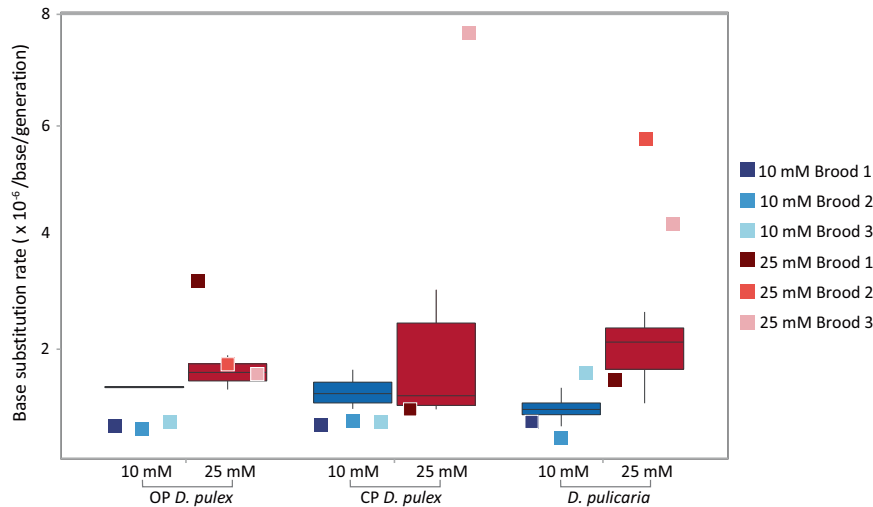
### EMS-induced heritable base-substitution rate

Across CP *D. pulex*, CP *D. pulicaria*, and OP *D. pulex*, we whole-genome sequenced 12 mutant lines treated with 10 mM EMS and 14 mutant lines treated with 25 mM EMS to detect germline mutations (Supplementary Table S3). Consistent with our expectation that EMS-induced mutations in the germline occur at an elevated rate relative to spontaneous mutations, the base substitution mutation rates for lines derived from 10 mM treatment ranged from  $9.40 \times 10^{-7}$  to  $1.32 \times 10^{-6}$  (mean =  $1.17 \times 10^{-6}$ , SEM =  $1.84 \times 10^{-7}$ , see Table 1 and Fig. 3), a few hundred times higher than the spontaneous mutation rate which ranges from  $2.30 \times 10^{-9}$  to  $7.17 \times 10^{-9}$  per site per generation (Keith et al. 2016; Flynn et al. 2017; Bull et al. 2019). Although there may be false positives in our dataset, the rate of such mis-identified mutations is most likely much smaller than 0.05 based on Sanger sequencing verification. We found no significant difference in the mean base substitution mutation rate or per gene mutation rate among the three *Daphnia* species at 10 mM (ANOVA  $p > 0.1$ ). Across the three *Daphnia* species, the average per gene mutation rate and per gene non-synonymous rate of the 10 mM treatment lines are  $2.65 \times 10^{-3}$  (SEM =  $3.32 \times 10^{-4}$ ) and  $1.19 \times 10^{-3}$  (SEM =  $1.71 \times 10^{-4}$ ), respectively (Table 1).

Notably, mutant lines from the 25 mM EMS treatment showed on average a higher base substitution mutation rate than those from the 10 mM treatment, yielding strong support to our first hypothesis. The base substitution rate for the 25 mM ranged from  $1.58 \times 10^{-6}$  to  $1.98 \times 10^{-6}$  across the three species. An ANOVA test also indicated no significant difference in EMS-induced base-substitution rates across the three species at this concentration ( $p > 0.1$ ). The average base substitution rate ( $1.75 \times 10^{-6}$  per site per generation, SEM =  $6.82 \times 10^{-7}$ , Table 1, Fig. 3) across the three species was significantly higher (a 0.5-fold increase) than that at 10 mM (mean =  $1.17 \times 10^{-6}$ , SEM =  $1.84 \times 10^{-7}$ ,  $t$  test  $p = 0.0052$ ). The average per gene mutation rate ( $4.09 \times 10^{-3}$  per gene per generation, SEM =  $4.31 \times 10^{-4}$ , Table 1) and average nonsynonymous mutation rate ( $1.91 \times 10^{-3}$  per gene per generation, SEM =  $1.81 \times 10^{-4}$ , Table 1) across the three species at 25 mM also showed an increase of 0.5 and 0.6 fold compared to those at 10 mM, respectively.

**Table 1.** Summary of mutations induced by 10 and 25 mM EMS.

EMS concentration	<i>Daphnia</i> species	Average number of mutations (SD)	Base substitution mutation rate (SEM)	Average number of mutations in genes (SD)	Per gene mutation rate (SEM)	Average number of non-synonymous mutations (SD)	Non-synonymous mutation rate (SEM)
10 mM	<i>D. pulex</i> (OP)	126 (0.71)	$1.24 \times 10^{-6}$ ( $3.09 \times 10^{-7}$ )	48 (2)	$3.9 \times 10^{-3}$ ( $1.2 \times 10^{-3}$ )	21 (0)	$1.9 \times 10^{-3}$ ( $4.6 \times 10^{-4}$ )
	<i>D. pulex</i> (CP)	124 (32)	$1.32 \times 10^{-6}$ ( $8.47 \times 10^{-9}$ )	58 (18)	$2.4 \times 10^{-3}$ ( $1.3 \times 10^{-4}$ )	28 (5)	$1.0 \times 10^{-3}$ ( $3.4 \times 10^{-6}$ )
	<i>D. pulicaria</i>	94 (22)	$9.40 \times 10^{-7}$ ( $2.35 \times 10^{-7}$ )	40 (60)	$1.9 \times 10^{-3}$ ( $2.7 \times 10^{-4}$ )	16 (5)	$7.8 \times 10^{-4}$ ( $2.3 \times 10^{-4}$ )
25 mM	<i>D. pulex</i> (OP)	154 (43)	$1.69 \times 10^{-6}$ ( $1.0 \times 10^{-6}$ )	65 (16)	$4.1 \times 10^{-3}$ ( $2.1 \times 10^{-3}$ )	36 (8)	$1.8 \times 10^{-3}$ ( $8.7 \times 10^{-4}$ )
	<i>D. pulex</i> (CP)	171 (104)	$1.58 \times 10^{-6}$ ( $4.34 \times 10^{-7}$ )	78 (40)	$3.2 \times 10^{-3}$ ( $7.5 \times 10^{-4}$ )	30 (14)	$1.7 \times 10^{-3}$ ( $3.7 \times 10^{-4}$ )
	<i>D. pulicaria</i>	201 (67)	$1.98 \times 10^{-6}$ ( $6.12 \times 10^{-7}$ )	91 (27)	$4.4 \times 10^{-3}$ ( $1.3 \times 10^{-3}$ )	43 (13)	$2.1 \times 10^{-3}$ ( $6.1 \times 10^{-4}$ )



**Fig. 3** Base-substitution mutation rates of three *Daphnia* species at 10 and 25 mM EMS treatment. The bar plot summarizes the species-specific rates based on multiple isolates of each species, whereas the scatter plot represents brood-specific mutation rates in each species.

### EMS-induced heritable base-substitution rate in consecutive broods

We hypothesized that consecutive broods produced by  $F_0$  females carry independent EMS-induced germline mutations as progenitor cells of oocytes are differentially affected by EMS in  $F_0$ s. To test this, we sequenced a total of 17 first-brood (BR1), second-brood (BR2), and third-brood (BR3) mutant lines treated with 10 and 25 mM EMS in three *Daphnia* isolates, Tex21 (CP *D. pulex*), AroMoose (CP *D. pulicaria*), and DB4-4 (OP *D. pulex*) (Supplementary Table S4).

Our results of the brood-specific mutation rate and spectrum in these three *Daphnia* isolates clearly supported our hypothesis. As our ANOVA tests indicated no significant variation in the base-substitution mutation rate among species/isolates, we do not distinguish among the species/isolate in the description below. Consistent with the base-substitution mutation rate at 10 mM, the average base-substitution mutation rates for BR1, BR2, and BR3 progenies at 10 mM are  $6.58 \times 10^{-7}$  (SEM =  $2.07 \times 10^{-8}$ ),  $5.70 \times 10^{-7}$  (SEM =  $9.20 \times 10^{-8}$ ) and  $9.90 \times 10^{-7}$  (SEM =  $2.90 \times 10^{-7}$ ), respectively (Fig. 3). Similarly, the average base-substitution mutation rates for BR1, BR2, and BR3 lines at 25 mM are  $1.86 \times 10^{-6}$  (SEM =  $6.95 \times 10^{-7}$ ),  $3.75 \times 10^{-6}$  (SEM =  $2.02 \times 10^{-6}$ ) and  $4.49 \times 10^{-6}$  (SEM =  $1.77 \times 10^{-6}$ ), respectively, significantly higher than those at 10 mM (ANOVA  $p = 0.039$ ) (Fig. 3).

When comparing the base-substitution mutation rates between the BR1, BR2 and BR3 lines of the same concentration, no significant difference was found (ANOVA  $p = 0.34$ ), indicating that the EMS induced base-substitution rate remained similar for at least the first three broods of the exposed  $F_0$  mother. The mean per gene mutation rate and non-synonymous mutation rate for the first three broods were also higher at 25 mM when compared to lines treated with 10 mM EMS (Supplementary Table S4). It should also be emphasized that the identified base substitutions in the first three consecutive broods of the same *Daphnia* isolate all occurred at unique sites in the genome, supporting that EMS induced heritable mutations in these broods in an independent manner.

### Spectrum and genomic effects of EMS-induced germline base substitutions

As expected and previously seen in other model organisms such as *C. elegans* (Flibotte et al. 2010) and *D. melanogaster* (Pastink et al. 1991), EMS primarily produced G:C to A:T transitions in all of the sequenced *Daphnia* mutant lines. On average 87% (SD = 8%) of the base substitutions in the 10 mM treatment lines are G:C to

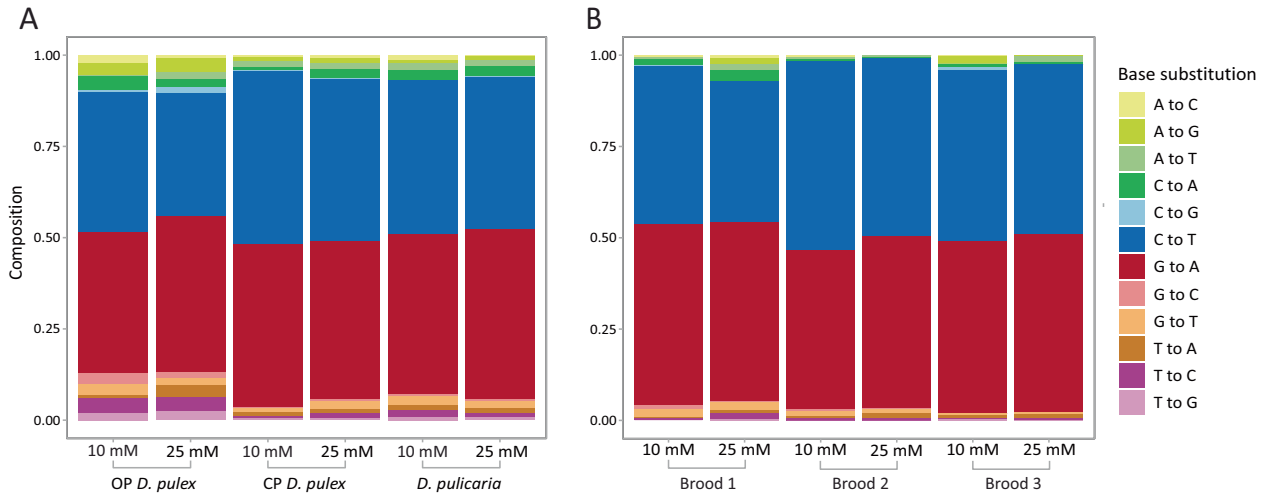
A:T transitions, resulting in an elevated transition-transversion ratio greater than 4.1 for all lines (Fig. 4A, Supplementary Table S3 and S5). Mutant lines of 25 mM EMS treatment were also highly biased towards G:C to A:T transitions (mean = 86%, SD = 7%), yielding a transition-transversion ratio greater than 4.8 for all lines (Fig. 4A, Supplementary Table S3). The observed transition-transversion ratio is much higher than those from spontaneous mutation accumulation experiments in *Daphnia* (e.g., Keith et al. 2016). Dominance of the mutation spectrum by G:C to A:T transitions was also seen in the BR1, BR2 and BR3 mutant lines for both 10 and 25 mM EMS treatments, further substantiating the idea that EMS successfully induced heritable mutations in consecutive broods (Fig. 4B, Supplementary Tables S4 and S6).

Concordant with the notion that EMS induces mutation randomly across the genome, the distribution of EMS-induced mutations for the 10 and 25 mM mutant lines were highly similar (ANOVA  $p = 1$ ) and did not show enrichment in specific genomic regions (chi-squared test  $p = 0.40$ ). In mutant lines treated with 10 mM, on average 34 (12%) of the induced mutations reside in exons, 14 (5%) in introns, 4 (1.3%) in 3' UTR, 3 (1.1%) in 5' UTR, and 59 (20.8%) in intergenic regions, whereas for lines treated with 25 mM, on average 65 (12.3%) mutations reside in exons, 27 (5.2%) in introns, 9 (1.7%) in 3' UTR, 6 (1.1%) in 5' UTR, and 114 (21.6%) in intergenic regions (Fig. 5A and Supplementary Table S7).

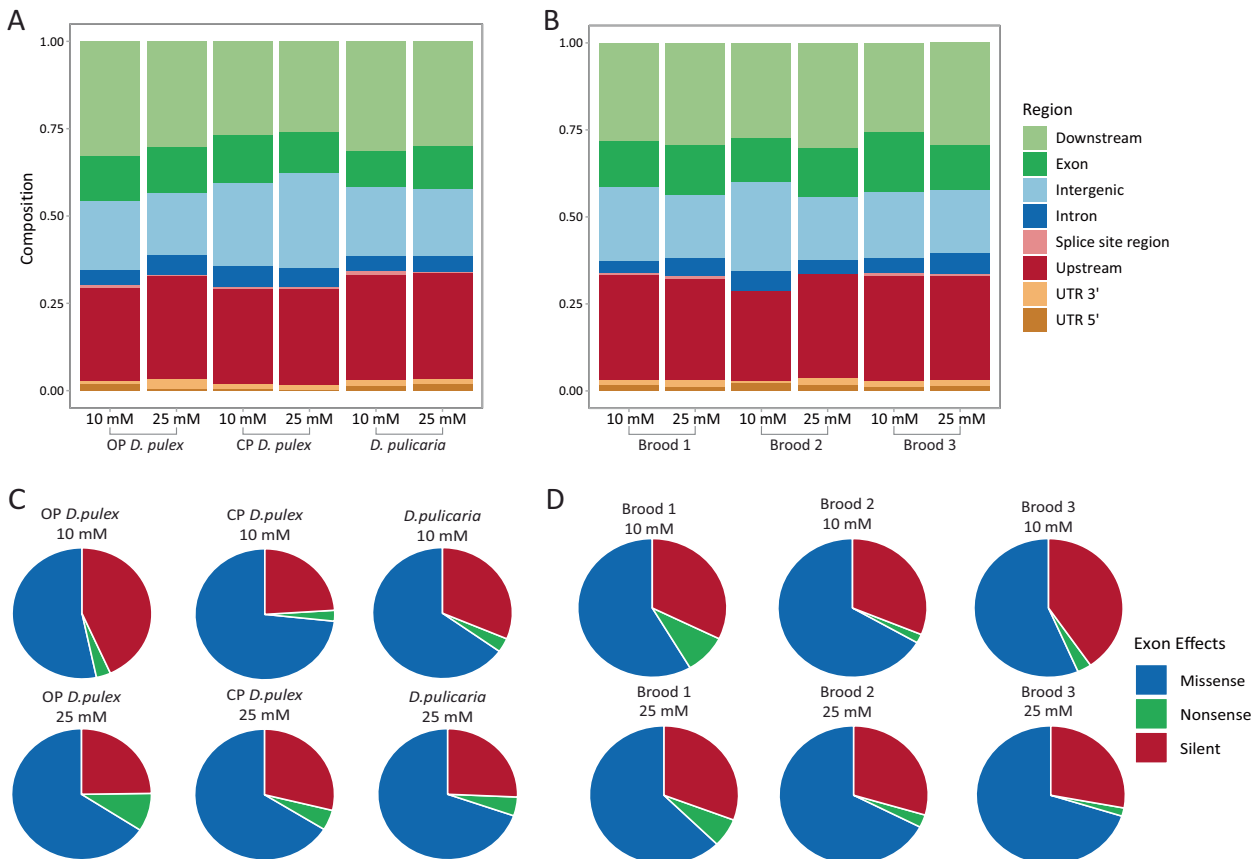
Furthermore, regarding exonic mutations, for the 10 mM treatment on average 23 (65.5%) were missense, 1 (3.2%) nonsense (stop-gained) and 11 (31.3%) silent. The 25 mM treatment once again produced very similar results with 44 (67.8%) missense, 3 (5.2%) nonsense (stop-gained) and 18 (27%) silent (Fig. 5C and Supplementary Table S7). The genomic distribution of mutations and exonic effects for the BR1, BR2 and BR3 lines also remained similar between the different broods and treatments, and mirrored the results summarized above (Fig. 5B, D and Supplementary Table S8). The observed ratios of non-synonymous vs synonymous changes do not significantly deviate from the 3:1 ratio based on considering all possible base substitutions at all codon sites (Graur and Li 2000).

### Motif analysis of mutated sites

For the NAN and NTN trinucleotide motifs (Fig. 6), all of the trinucleotides were significantly under-represented (chi-squared test  $p < 0.05$ ). Among the NCN trinucleotides (Fig. 6), the TCG, ACG, CCG, TCC, CCT, TCT, GCC, ACC, CCC and GCG (5'-3' orientation) were significantly over-represented (chi-squared test  $p < 0.05$ ). For the NGN trinucleotides, we found significant over-representation



**Fig. 4** Average proportions of different types of base substitutions in each species at 10 and 25 mM EMS treatment. **A** Composition of base substitutions in each species. **B** Composition of base substitutions in different broods.



**Fig. 5** Summary of distribution of base substitutions and amino acid changing effects. Average proportions of base substitutions in different genomic regions (**A** and **B**) and amino acid changing effects at 10 and 25 mM EMS concentration (**C** and **D**).

of the GGT, AGG, CGG, CGA, GGA, GGC, CGC, and GGG (chi-squared test  $p < 0.05$ , Fig. 6 and Supplementary Table S9).

#### Number of $F_1$ s for reaching mutation saturation

Based on the average base-substitution rate per gene at 25 mM EMS treatment across three *Daphnia* species ( $\sim 4 \times 10^{-3}$  per gene per generation),  $\sim 750 F_1$ s are needed to find at least one  $F_1$  animal heterozygous for a mutation in a gene of interest with 95%

probability. With  $\sim 750 F_1$ s, a total of 54,000 genes would have been mutated, translating to roughly 3 mutations per gene given the  $\sim 18,000$  number of genes in the *D. pulex* genome.

#### DISCUSSION

This study examines the genome-wide EMS-induced heritable mutations in three microcrustacean *Daphnia* species at different





**Table 2.** Number of genes, spontaneous base substitution rate, EMS treated mutation rate and estimated number of function affecting variants for different model organisms.

Specie	Number of genes	Spontaneous per base per generation mutation rate	EMS treated per gene per generation mutation rate	Estimated number of function affecting variants produced
<i>C. elegans</i>	<sup>a</sup> 19,404	<sup>e</sup> $2.5 \times 10^{-9}$	<sup>l</sup> $2.5 \times 10^{-3}$	49
<i>D. melanogaster</i>	<sup>b</sup> 13,601	<sup>f</sup> $3.5 \times 10^{-9}$ ; <sup>g</sup> $5.49 \times 10^{-9}$	<sup>m</sup> $1.0 \times 10^{-3}$	14
<i>A. thaliana</i>	<sup>c</sup> 27,655	<sup>h</sup> $6.95 \times 10^{-9}$ ; <sup>i</sup> $5.90 \times 10^{-9}$	<sup>j</sup> $3.0 \times 10^{-3}$	83
<i>D. pulex</i> (OP)	<sup>d</sup> 18,440	<sup>j</sup> $2.30 \times 10^{-9}$ ; <sup>k</sup> $7.17 \times 10^{-9}$	$4.1 \times 10^{-3}$	76
<i>D. pulex</i> (CP)	<sup>d</sup> 18,440	<sup>k</sup> $4.53 \times 10^{-9}$	$3.2 \times 10^{-3}$	59
<i>D. pulicaria</i>	27,846	–	$4.4 \times 10^{-3}$	123

*Daphnia* estimates are based on the results from 25 mM EMS treatment.

<sup>a</sup>Chen et al. (2005).

<sup>b</sup>Adams et al. (2000).

<sup>c</sup>Cheng et al. (2017).

<sup>d</sup>Ye et al. (2017).

<sup>e</sup>Denver et al. (2009).

<sup>f</sup>Keightley et al. (2009).

<sup>g</sup>Schrider et al. (2013).

<sup>h</sup>Weng et al. (2019).

<sup>i</sup>Ossowski et al. (2010).

<sup>j</sup>Flynn et al. (2017).

<sup>k</sup>Keith et al. (2016).

<sup>l</sup>Gengyo-Ando and Mitani (2000).

<sup>m</sup>Spradling (1997).

excess of purines in the  $-1$  and  $+1$  positions, with adenine favored over guanine, a deficiency of guanine in the  $-2$  position, and an excess of guanine in the  $+2$  position (Greene et al. 2003). In *D. melanogaster* a strong purine bias, mostly of guanine, was reported in the positions flanking the mutation site (Bentley et al. 2000). These observations suggest that EMS might show preference for certain sequence motifs, but the target motifs can differ between species, likely due to nucleotide composition differences.

Last, we offer a few recommendations for performing genetic screening in *Daphnia*. We determined the number of function affecting mutations induced by EMS, with 76, 59 and 123 per generation for OP *D. pulex*, CP *D. pulex* and *D. pulicaria* respectively. In comparison, EMS mutagenesis in *C. elegans*, *D. melanogaster*, and *A. thaliana* produces around 49, 14, and 83 function affecting variants per generation respectively (Table 2).

A simple screening can be easily performed on  $F_1$ s in *Daphnia*. As we calculated, 750  $F_1$ s would be sufficient to contain 3 mutated copies of each gene, with at least 1 mutation residing in any single gene at 95% probability. Because all the induced mutations exist in the heterozygous state in the  $F_1$ s, only mutants with dominant mutations causing observable morphological alterations can be scored. However, we note that high-throughput molecular assays can be applied for screening  $F_1$ s to detect molecular phenotypic changes caused by recessive mutations if the costs for screening ~1000 individuals are manageable.

Furthermore, we can perform sibling crossing between progenies of  $F_1$  individuals to obtain  $F_2$ s that are homozygous for the induced mutations so that recessive mutations can show their phenotypic effect (Fig. 1). Technically speaking, if we start with 750  $F_1$ s, each  $F_1$  mutant line can be clonally expanded to a large clonal culture, which can then be crowded to induce clonal male production and sexual reproduction in females. Since clonally produced males and females of the same  $F_1$  mutant line are genetically identical, sibling crossing will produce  $F_2$  offspring that have 25% induced mutations in the homozygous state (Fig. 1). Although  $F_2$ s have to be hatched from resting embryos that only develop under a strict set of conditions, an optimal hatching procedure for the *Daphnia* species used in this study have already been developed (Luu et al. 2020).

The probability of obtaining mutants carrying mutations in homozygous state in a gene of interest depends on the number of  $F_2$ s collected from each  $F_1$  line. This probability is written as  $1 - (3/4)^n$ , where  $n$  is the number of  $F_2$ s and the term  $(3/4)^n$  denotes the probability of seeing non-homozygous mutants in  $n$   $F_2$  individuals. We can directly see that collecting 4 and 5  $F_2$ s from each  $F_1$  mutant line would have nearly 70 and 80% chance of getting a homozygous mutation, respectively. Therefore, a near saturated  $F_2$  screening in *Daphnia* would require 3000–4000  $F_2$ s. It is obvious that the  $F_1/F_2$  ratio determines the amount of resources that will be devoted to the screening experiments. Depending on the types of mutants of interest, a different  $F_1/F_2$  ratio can be adopted (Shaham 2007).

Efficiently scoring mutants in the large  $F_2$  population is another critical factor for a successful genetic screening experiment. Depending on the phenotypic traits of interest, high throughput phenotypic assay methods need to be developed for *Daphnia*, which seems to be underdeveloped at this moment. *Daphnia* has a nearly transparent carapace and many body parts (e.g., heart, appendages) are directly visible under a microscope, which are desirable characteristics for high throughput phenotypic screening. We hope that many novel phenotyping methods will emerge as forward screening in *Daphnia* or other small crustaceans gains more popularity.

In conclusion, we demonstrated that EMS mutagenesis can successfully induce heritable mutations in the genome of *Daphnia*. Our analyses of the mutation rate caused by different concentrations of EMS and mutation patterns in consecutive broods provide possible ways to increase the efficiency of a genetic screening experiments. Lastly, we provide some guidance on the sample sizes required for  $F_1$  and  $F_2$  screening experiments in the hope that genetic screening will become a powerful tool in the study of *Daphnia* genomics and evolution.

#### DATA AVAILABILITY

The raw DNA sequence data for this project can be found at NCBI SRA under PRJNA715913.

## REFERENCES

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195
- Altshuler I, Demiri B, Xu S, Constantin A, Yan ND, Cristescu ME (2011) An integrated multi-disciplinary approach for studying multiple stressors in freshwater ecosystems: *Daphnia* as a model organism. *Integr Comp Biol* 51:623–633
- Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat Rev Genet* 8:619–631
- Bentley A, MacLennan B, Calvo J, Dearolf CR (2000) Targeted recovery of mutations in *Drosophila*. *Genet* 156:1169–1173
- Brenner S (1974) The genetics of *Caenorhabditis Elegans*. *Genetics* 77:71–94
- Brookes P, Lawley PD (1961) The reaction of mono- and di-functional alkylating agents with nucleic acids. *Biochem J* 80:496–503
- Bull JK, Flynn JM, Chain FJJ, Cristescu ME (2019) Fitness and genomic consequences of chronic exposure to low levels of Copper and Nickel in *Daphnia pulex* mutation accumulation lines. *G3 Genes Genomes Genet* 9:61–71
- Chen N, Harris TW, Antoshechkin I, Bastiani C, Bieri T, Blasiar D et al. (2005) WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Res* 33:D383–D389
- Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J* 89:789–804
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 6:80–92
- Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH et al. (2011) The recombining genome of *Daphnia pulex*. *Science* 331:555–561
- Cooper JL, Greene EA, Till BJ, Codomo CA, Wakimoto BT, Henikoff S (2008) Retention of induced mutations in a *Drosophila* reverse-genetic resource. *Genetics* 180:661–667
- Coulondre C, Miller JH (1977) Genetic studies of the lac repressor. III. Additional correlation of mutational sites with specific amino acid residues. *J Mol Biol* 117:525–567
- Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK et al. (2009) A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc Natl Acad Sci USA* 106:16310–16314
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15
- Ebert D (2005) Introduction to *Daphnia* Biology. National Center for Biotechnology Information (US)
- EMBOSS: Compseq. Retrieved from <http://emboss.open-bio.org/rel6/apps/compseq.html>
- Flibotte S, Edgley ML, Chaudhry I, Taylor J, Neil SE, Rogula A et al. (2010) Whole-Genome Profiling of Mutagenesis in *Caenorhabditis elegans*. *Genetics* 185:431–441
- Flynn JM, Chain FJJ, Schoen DJ, Cristescu ME (2017) Spontaneous mutation accumulation in *Daphnia pulex* in selection-free vs. competitive environments. *Mol Biol Evol* 34:160–173
- Frisch D, Morton PK, Chowdhury PR, Culver BW, Colbourne JK, Weider LJ et al. (2014) A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecol Lett* 17:360–368
- Gengyo-Ando K, Mitani S (2000) Characterization of mutations induced by ethyl methanesulfonate, UV, and trimethylpsoralen in the nematode *Caenorhabditis elegans*. *Biochem Biophys Res Commun* 269:64–69
- Graur D, Li W (2000) *Fundamentals of molecular evolution* (2nd Edition). Sinauer Associates. Sunderland MA
- Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH et al. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genet* 164:731–740
- Greenspan RJ (1997) Fly pushing: the theory and practice of *Drosophila* genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Henry IM, Nagalakshmi U, Lieberman MC, Ngo KJ, Krasileva KV, Vasquez-Gross H et al. (2014) Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. *Plant Cell* 26:1382–1397
- Ho EKH, Macrae F, Latta IV LC, McIlroy P, Ebert D, Fields PD et al. (2020) High and Highly Variable Spontaneous Mutation Rates in *Daphnia*. *Mol Biol Evol* 37:3258–3266
- Jackson CE, Xu S, Ye Z, Pfrender ME, Lynch M, Colbourne JK, et al. (2021) Chromosomal rearrangements preserve adaptive divergence in ecological speciation. *bioRxiv*. 457158; <https://doi.org/10.1101/2021.08.20.457158>
- St Johnston D (2002) The art and design of genetic screens: *Drosophila melanogaster*. *Nat Rev Genet* 3:176–188
- Jorgensen EM, Mango SE (2002) The art and design of genetic screens: *Caenorhabditis elegans*. *Nat Rev Genet* 3:356–369
- Kato Y, Matsuura T, Watanabe H (2012) Genomic integration and germline transmission of plasmid injected into Crustacean *Daphnia magna* Eggs. *PLOS ONE* 7:e45318
- 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* 19:1195–1201
- Keith N, Tucker AE, Jackson CE, Sung W, Lucas Lledó JI, Schrider DR et al. (2016) High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. *Genome Res* 26:60–69
- Kilham SS, Kreeger DA, Lynn SG, Goulden CE, Herrera L (1998) COMBO: a defined freshwater culture medium for algae and zooplankton. *Hydrobiologia* 377:147–159
- Krasovec M, Eyre-Walker A, Sanchez-Ferandin S, Piganeau G (2017) Spontaneous mutation rate in the smallest photosynthetic eukaryotes. *Mol Biol Evol* 34:1770–1779
- Kutscher LM, Shaham S (2014) Forward and reverse mutagenesis in *C. elegans*. *WormBook*. 1–26. <https://doi.org/10.1895/wormbook.1.167.1>
- Lee SY, Cheong JI, Kim TS (2003) Production of doubled haploids through anther culture of M1 rice plants derived from mutagenized fertilized egg cells. *Plant Cell Rep.* 22:218–223
- Lewis EB, Bacher F (1968) Method of feeding ethylmethane sulfonate (EMS) to *Drosophila* males. *Drosophila Information Service* 43:193
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26:589–595
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The sequence alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079
- Loveless A, Haddow A (1959) The influence of radiomimetic substances on deoxyribonucleic acid synthesis and function studied in *Escherichia coli*/ phage systems - III. Mutation of T2 bacteriophage as a consequence of alkylation in vitro: the uniqueness of ethylation. *Proc R Soc Lond Ser B - Biol Sci* 150:497–508
- Luu DH-K, Vo HP, Xu S (2020) An efficient method for hatching diapausing embryos of *Daphnia pulex* species complex (Crustacea, Anomopoda). *J Exp Zool Part Ecol Integr Physiol* 333:111–117
- Lynch M, Seyfert A, Eads B, Williams E (2008) Localization of the genetic determinants of meiosis suppression in *Daphnia pulex*. *Genetics* 180:317–327
- Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB et al. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci* 105:9272–9277
- Lynch M, Gutenkunst R, Ackerman M, Spitze K, Ye Z, Maruki T et al. (2017) Population genomics of *Daphnia pulex*. *Genetics* 206:315–332
- Martin B, Ramiro M, Martinez-Zapater JM, Alonso-Blanco C (2009) A high-density collection of EMS-induced mutations for TILLING in *Arabidopsis thaliana* Landsberg erecta genetic background. *BMC Plant Biol* 9:147
- McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. *Nat Biotechnol* 18:455–457
- Mobini-Dehkordi M, Nahvi I, Zarkesh-Esfahani H, Ghaedi K, Tavassoli M, Akada R (2008) Isolation of a novel mutant strain of *Saccharomyces cerevisiae* by an ethyl methane sulfonate-induced mutagenesis approach as a high producer of bioethanol. *J Biosci Bioeng* 105:403–408
- Muller HJ (1927) Artificial transmutation of the gene. *Science* 66:84–87
- Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327:92–94
- Page DR, Grossniklaus U (2002) The art and design of genetic screens: *Arabidopsis thaliana*. *Nat Rev Genet* 3:124–136
- Pastink A, Heemskerck E, Nivard MJ, van Vliet CJ, Vogel EW (1991) Mutational specificity of ethyl methanesulfonate in excision-repair-proficient and -deficient strains of *Drosophila melanogaster*. *Mol Gen Genet* 229:213–218
- Picard. Broad Institute, GitHub repository. GitHub. Retrieved from <http://broadinstitute.github.io/picard/>
- Prakash L, Higgins D (1982) Role of DNA repair in ethyl methanesulfonate-induced mutagenesis in *Saccharomyces cerevisiae*. *Carcinogenesis* 3:439–444
- Rozen S, Skaletsky H (1999) Primer3 for the WWW for General Users and for Biologist Programmers. In: Misener S, Krawetz SA (eds) *Bioinformatics Methods and Protocols, Methods in Molecular Biology*. Humana Press, Totowa, NJ, p 365–386
- Sarin S, Bertrand V, Bigelow H, Boyanov A, Doitsidou M, Poole RJ et al. (2010) Analysis of multiple ethyl methanesulfonate-mutagenized *Caenorhabditis elegans* strains by whole-genome sequencing. *Genetics* 185:417–430
- Schrider DR, Houle D, Lynch M, Hahn MW (2013) Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* 194:937–954
- Sega GA (1984) A review of the genetic effects of ethyl methanesulfonate. *Mutat Res Genet Toxicol* 134:113–142
- Shaham S (2007) Counting mutagenized genomes and optimizing genetic screens in *Caenorhabditis elegans*. *PLOS ONE* 2:e1117

- Shiwa Y, Fukushima-Tanaka S, Kasahara K, Horiuchi T, Yoshikawa H (2012) Whole-genome profiling of a novel mutagenesis technique using proofreading-deficient DNA polymerase  $\delta$ . *Int J Evol Biol* 2012:860797
- Spradling AC (1997) Fly pushing: the theory and practice of *Drosophila* genetics. *Trends Genet* 13(10):418
- Weng M-L, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG et al. (2019) Fine-grained analysis of spontaneous mutation spectrum and frequency in *Arabidopsis thaliana*. *Genetics* 211:703–714
- Winkler S, Schwabedissen A, Backasch D, Bökel C, Seidel C, Bönisch S et al. (2005) Target-selected mutant screen by TILLING in *Drosophila*. *Genome Res* 15:718–723
- Xu S, Spitze K, Ackerman MS, Ye Z, Bright L, Keith N et al. (2015) Hybridization and the origin of contagious asexuality in *Daphnia pulex*. *Mol Biol Evol* 32:3215–3225
- Ye Z, Xu S, Spitze K, Asselman J, Jiang X, Ackerman MS et al. (2017) A new reference genome assembly for the microcrustacean *Daphnia pulex*. *G3 GenesGenomesGenetics* 7:1405–1416

## ACKNOWLEDGEMENTS

We thank D. Luu for his assistance in the early stage of this study. We thank three anonymous reviewers for their constructive comments. This work is supported by NIH grant R35GM133730 to SX.

## AUTHOR CONTRIBUTIONS

MS and SX designed the experiment. MS and MTS performed the experiments. MS and TVH performed data analyses. MS and SX wrote the manuscript with input from other coauthors.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41437-021-00478-x>.

**Correspondence** and requests for materials should be addressed to Sen Xu.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.