The effects of mutations on gene expression and alternative splicing

Authors: Marelize Snyman[1,2] and Sen Xu[1,3,*]

[1]Department of Biology, University of Texas at Arlington, Arlington, Texas USA 76019.

[2]Present address: Department of Dermatology, University of Texas Southwestern Medical Center, Dallas, Texas, 75235, USA

[3]Present address: Division of Biological Sciences, University of Missouri, Columbia, Missouri, 65211, USA

*Correspondence author. Division of Biological Sciences, University of Missouri, Columbia, Missouri, 65211, USA. Email: sxwf7@umsystem.edu

Abstract

Understanding the relationship between mutations and their genomic and phenotypic consequences has been a longstanding goal of evolutionary biology. However, few studies have investigated the impact of mutations on gene expression and alternative splicing on the genome-wide scale. In this study, we aim to bridge this knowledge gap by utilizing whole-genome sequencing data and RNA sequencing data from 16 obligately parthenogenetic *Daphnia* mutant lines to investigate the effects of EMS-induced mutations on gene expression and alternative splicing. Using rigorous analyses of mutations, expression changes, and alternative splicing, we show that trans-effects are the major contributor to the variance in gene expression and alternative splicing between the wildtype and mutant lines, whereas cis mutations only affected a limited number of genes and do not always alter gene expression. Moreover, we show that there is a significant association between differentially expressed genes and exonic mutations, indicating that exonic mutations are an important driver of altered gene expression.

Introduction

One of the fundamental goals of evolutionary biology is to understand the genomic and phenotypic consequences of mutations. As mutations are rare, long-term mutation accumulation (MA) experiments in a variety of model organisms, e.g. *Drosophila* (Keightley *et al.* 2009, Krasovec 2021), *Caenorhabditis* (Denver *et al.* 2009), *Arabidopsis* (Keightley and Lynch 2003, Ossowski *et al.* 2010)*, Saccharomyces cerevisiae* (Lynch *et al.* 2008, Zhu *et al.* 2014), and *Daphnia* (Bull *et al.* 2019, Flynn *et al.* 2017, Xu *et al.* 2012) have been instrumental in our understanding of mutational consequences. These MA experiments minimized the power of natural selection relative to genetic drift, thus allowing for the persistence of non-lethal mutations in a nearly unbiased manner and offering insight into the genome-wide rates, spectrum, and fitness effects of spontaneous mutations. The findings of these studies have yielded important implications for our understanding of genotype-phenotype relationships, the evolution of mutation rate, and the impacts of mutations and selection on molecular evolution.

To date, many MA studies have moved beyond simply measuring the mutation rate, spectrum, and fitness effects on the organismal level. For example, MA studies have been utilized to study transposable element mutation rates (Díaz-González and Domínguez 2020, Ho *et al.* 2021), enzymatic activity (Aquadro *et al.* 1990, Harada 1995), methylation frequency (Denkena *et al.* 2021, Jiang *et al.* 2014), phenotypic plasticity (Latta *et al.* 2015), and patterns of genotype-environment interactions (Chu and Zhang 2021, Scheffer *et al.* 2022, Xu 2004). MA studies have also been supplemented with transcriptome data to study the evolution of copy number changes (Konrad *et al.* 2018), the rate at which mutations produce new variation in gene expression (Rifkin *et al.* 2005), effects of mutations on gene expression (Landry *et al.* 2007) and allele-specific expression (White *et al.* 2022). Moreover, how targeted mutagenesis (Iofrida *et al.*

2012, West *et al.* 2016) and somatic mutations in cancer genomes affect transcription has been investigated (Gerstung *et al.* 2015, Jia and Zhao 2017). Despite these previous studies, the impact of chemically induced germline mutations on gene expression on the genome-wide scale remains understudied (but see Gruber *et al.* 2012 and Vande Zande *et al.* 2022). Addressing this understudied issue would yield insight into how different types of environmentally-induced mutations (e.g., cis- and trans-mutations) can contribute to the evolution of gene expression in both short- and long-term Furthermore, the impact of chemically induced mutations on alternative splicing at the whole-genome level remains elusive. Gaining a better understanding of the impact of mutations affecting the splicing machinery and the role of cis- and trans-mutations in regulating alternative splicing will have important implications for understanding phenotypic changes such as disease (reviewed in Ward and Cooper 2010).

Gene expression is a major link between an organism's genotype and phenotype. Numerous studies have demonstrated the critical role of altered gene expression in phenotypic variation such as the beak morphology in Darwinian finches (Abzhanov *et al.* 2006), cold adaptation in lizards (Josephs 2021), melanin production in *Drosophila* (Rebeiz *et al.* 2009), body mass due to latitudinal adaptation and color pattern in mice (Mack *et al.* 2018, Manceau *et al.* 2011), and adaptation to predation in Trinidadian guppies (Ghalambor *et al.* 2015).

Variation in gene expression can arise through cis and trans-mutations. Cis-mutations are often located within a gene, or up- and downstream of a gene, affecting the interaction between the gene and its regulators (e.g., promotors and enhancers). On the other hand, trans-mutations can be located anywhere within the genome (other than within or nearby the gene of interest) and mediate the expression of genes through the use of e.g., transcription factors and long noncoding RNAs (Albert *et al.* 2018, Vande Zande *et al.* 2022). Trans-mutations are thought to be

primarily responsible for the variation in gene expression within species since trans-factors represent a much larger mutational target and are thus more likely to arise than cis-mutations (Chen *et al.* 2015, Coolon *et al.* 2014, Metzger *et al.* 2016, Rhoné *et al.* 2017, Wittkopp *et al.* 2004). However, presumably due to the pleiotropic effects of trans-mutations on expression (see Vande Zande *et al.* 2022), trans-mutations are often selected against as populations adapt and diverge. With cis-mutations mostly affecting the expression of specific genes, they survive selection more often, and therefore are mostly responsible for gene variation between populations and between species (Coolon *et al.* 2015, Schaefke *et al.* 2013).

These theories and empirical observations in natural populations lead us to expect that *de novo* mutations (spontaneous or environmentally induced) upon arising in the genome should exhibit stronger trans-effects than cis-effects. However, direct empirical support for this hypothesis remains limited. This is mainly because many studies inferred the impact of mutations on expression using F1 hybrids of two diverging species, because mutation accumulation studies are restricted to a small number of model organisms, and because many previous studies mainly utilized targeted mutagenesis to examine the effects of cis-mutations in a variety of different organisms (Gruber *et al.* 2012, Hornung *et al.* 2012, Kwasnieski *et al.* 2012, Maricque *et al.* 2017, Metzger *et al.* 2015).

Furthermore, a very much ignored aspect of mutations' impact resides in alternative splicing. Alternative splicing is critically involved in the origin of phenotypes, e.g., the flowering time of *Arabidopsis* (Macknight *et al.* 2002), seed development in sunflowers (Smith *et al.* 2018), pigmentation of cichlid fishes (Terai *et al.* 2003), and thermogenesis in mice (Vernia *et al.* 2016). With the involvement of trans-acting splicing factors and cis-acting regulatory motifs, alternative splicing is often highly regulated and vulnerable to mutations. One can imagine that

point mutations can occur in both introns and exons, leading to the alteration of existing splice sites or the generation of new ones, and affecting splicing enhancers and silencers (Anna and Monika, 2018). However, it remains unclear to what extent alternative splicing is affected by trans-mutations, which could directly or indirectly affect the various aspects of the transcription and translation of splicing factors.

To examine the effects of mutations on gene expression and alternative splicing on a genome-wide scale, we employed the chemical mutagen ethyl methanesulfonate (EMS) to induce heritable mutations in the microcrustacean *Daphnia*. EMS induces DNA damage through the alkylation of guanine, causing O^6 ethylguanine to mispair with thymine instead of cytosine in subsequent replications, resulting in EMS-induced mutations mainly consisting of G:C to A:T transitions (Greene *et al.* 2003). EMS-induced mutations are randomly distributed throughout the genome, enabling the generation of loss or gain of function mutants and weak nonlethal alleles (Greene *et al.* 2003, Lee *et al.* 2003).

Previously we established an EMS mutagenesis protocol for *Daphnia* (Snyman *et al.* 2021), where we showed that the base substitution rate for EMS exposed *Daphnia* reached 1.17×10^{-6} and 1.75×10^{-6} per base per generation for 10mM and 25mM EMS concentrations, respectively (Snyman *et al.* 2021). These EMS-induced mutation rates are greatly elevated compared to the spontaneous base substitution rate estimates in *Daphnia* of 2.30×10^{-9} and 7.17×10^{-9} per base per generation (Flynn *et al.* 2017, Keith *et al.* 2016). The EMS-induced mutations were also randomly distributed in different types of genomic regions (e.g., exon, intron, intergenic regions).

In this study, we selected an obligately parthenogenetic (OP) *Daphnia* isolate to examine the impact of EMS-induced heritable mutations on gene expression and alternative splicing. A

major motivation for using an OP isolate is that we can avoid the maternal effects imposed by EMS exposure in the ancestral mutant as much as possible by examining the asexual offspring of later generations that are genetically identical (albeit extremely rare spontaneous mutations) to the ancestors. *Daphnia* typically reproduces by cyclical parthenogenesis where they switch between asexual and sexual reproduction depending on environmental cues. However, these obligate parthenogenetic isolates **(Supplementary Figure S1)** reproduce solely via asexual reproduction (Xu *et al.* 2022).

In this study, we examined the gene expression and alternative splicing mechanisms functional in 16 *Daphnia* mutant lines derived from the same ancestor in comparison to the wildtype using whole-genome DNA and RNA sequencing data. Our main goals were three-fold. We first investigated the alterations of gene expression (DE) and alternative spicing (AS) in the mutant lines due to the EMS-induced mutations, second the contribution of cis- and trans-mutations to DE and AS, and lastly if whether any genomic regions carrying cis-mutations were significantly associated with DE and AS.

Materials and Methods

Sampling and maintenance of isolates

An OP *Daphnia* isolate collected from a pond in the US was used in this study. This isolate has been kept as a clonally reproducing line in artificial lake (Kilham *et al.* 1998) water under a 16:8 (light: dark) cycle at 18°C and fed with the green algae, *Scenedesmus obliquus* twice a week.

Generating Daphnia mutants using EMS

To generate *Daphnia* mutants with heritable mutations from the wildtype, we used EMS mutagenesis following the established methodology by Snyman *et al.* 2021. Specifically, eight

sexually mature *Daphnia* females were exposed to 25 mM ethyl methanesulfonate for 4 hours to introduce mutations into their oocytes. After exposure, these females were individually isolated **(Figure 1)**. We collected two progenies (G_1s) in the first asexual brood from each of the eight exposed females because each G_1 was derived from an oocyte that was individually mutagenized, thus all the G_1s are genetically distinct. In total, we established 16 mutant lines.

To avoid maternal effect (i.e., from exposure to EMS) on gene expression in the offspring, each of the mutant lines was propagated asexually for at least 3 generations to reach a high enough animal density to induce the parthenogenetic production of resting eggs. 5-10 ephippia (each containing 2 resting eggs) were collected from each mutant line, and the decapsulated resting eggs were hatched using the protocol established by Luu *et al.* 2020. Briefly, the collected resting embryos were kept in the dark at 18°C for two weeks before exposure to UV light to stimulate embryo development. If no sign of embryo development was observed after five days of UV exposure, the process was repeated until at least one resting embryo hatched per mutant line. The hatched mutant offspring was then asexually propagated until a high enough animal density was reached for DNA and RNA extraction.

DNA extraction and whole-genome DNA sequencing

A total of 30-40 clonal offspring were collected for each mutant line for DNA extraction using a CTAB (Cetyl Trimethyl Ammonium Bromide) method (Doyle and Doyle 1987). DNA quality and concentration were assessed by electrophoresis on a 2% agarose gel and a Qubit 4.0 Fluorometer (Thermo Scientific, Waltham, MA, USA). DNA sequencing libraries were prepared following standard MGI sequencing library protocol by the Beijing Genomics Institute (BGI, Cambridge, MA, USA). All 16 mutant lines were sequenced on an MGI DNBseq platform with

150-bp paired-end reads, with a targeted sequencing coverage of 30X per line. The raw DNA sequence data were deposited at NCBI SRA under PRJNA892919.

RNA extraction and transcriptomic sequencing

Experimental animals were maintained in the same environmental conditions i.e., 18℃ with a 16:8 light/dark cycle, and three replicates of 2 or 3-day old offspring were collected from the wild-type (i.e., control), and each mutant line. RNA was extracted using the Promega SV Total RNA Isolation kit (Madison, WI, USA) following the manufacturer's instructions. RNA quality was examined by electrophoresis on a 2% agarose gel, and RNA concentration was measured using a Qubit 4.0 Fluorometer (Thermo Scientific, Waltham, MA, USA). RNA sequencing libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina (Ipswich, MA, USA). Transcriptomic sequencing was done by Novogene Corporation Inc. (Sacramento, CA, USA) following standard Illumina sequencing protocol. Each library was sequenced on an Illumina NovaSeq6000 platform with at least 20 million 150-bp paired-end reads. The raw RNA sequence data were deposited at NCBI SRA underPRJNA892982.

Quality control and mapping

Quality of the raw reads was examined using FastQC (Andrews 2010). No adapter contamination was observed for the whole-genome sequencing data, thus further analysis was completed using the raw reads. Our RNAseq dataset showed adapter contamination, therefore Trimmomatic v.0.39 (Bolger *et al.* 2014) was used to perform adapter trimming and quality filtering. Lastly, reads were reassessed using FastQC to confirm the removal of low-quality reads and adapter sequences.

Identification of EMS-induced heritable mutations

We identified heterozygous EMS-induced heritable mutations following the EMS mutagenesis analysis protocol established by Snyman *et al.* (2021). Since these mutant lines are obligate parthenogens (i.e., no segregation and no sexual reproduction) all EMS-induced mutations should be in the heterozygous state, ignoring rare ameiotic gene conversions (Omilian et al 2006). We used the Burrows-Wheeler Alignment Tool BWA-MEM version 0.7.17 (Li and Durbin 2009) with default parameters to align the whole-genome DNA sequencing raw reads of each mutant line to the *Daphnia pulicaria* reference genome (Jackson *et al.* 2021). SAMtools (Li *et al.* 2009) was used to remove reads mapped to multiple locations, and the MarkDuplicates function of Picard tools (http://broadinstitute.github.io/picard/) was used to locate and tag PCR duplicates. We used BCFtools (Li 2011) mpileup and call functions with default parameters to generate genotype likelihoods and genotype calls in a VCF file. The following additional FORMAT and INFO tags were added 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). Tentative EMS-induced mutations for all treatment lines were further filtered with BCFtools' filter function to retain only single nucleotide polymorphisms (SNPs) with a sequencing depth (DP) >= 10 and <= 60, quality score (QUAL) >= 20, and a distance of at least 50-bp from an indel.

Furthermore, a custom python script (https://github.com/Marelize007/Functional_impact_of_EMS-induced_mutations) was used to filter the tentative mutations using a consensus method (Snyman *et al.* 2021). Briefly, all genotype data from the 16 EMS mutant lines were added to one VCF file, and a consensus genotype was established by majority rule. Out of the 16 mutant lines, at least 12 mutant lines had to agree to generate a consensus genotype call per site. This allowed for the inclusion of mutation sites where up to 3 out of the 16 mutant lines had no

genotype call due to inadequate depth. A tentative mutation was identified if the mutant line showed a different genotype than the consensus. All final mutations had to be supported by at least two forward and two reverse reads to limit false positives due to allele drop and inadequate sequence coverage.

Differential expression analysis

The trimmed RNAseq reads were mapped to the *D. pulicaria* (Jackson *et al.* 2021) reference genome utilizing the STAR aligner (Dobin *et al.* 2013) with default parameters. Reads mapped to multiple locations in the genome were removed using SAMtools (Li *et al.* 2009), and raw transcript counts were obtained for each sample using featureCounts (Liao *et al.* 2014). Differential expression analysis was performed using DESeq2 v.1.34.0 (Love *et al.* 2014). First, the regularized log (rlog) transformation function was used to normalize the mapped read counts, and DE genes were determined using the Wald negative binomial test with the design formula ~ genotype, where genotype represents either the mutant or wildtype genotype. The Benjamini-Hochberg method (FDR < 0.05) was used to adjust p-values for multiple testing, and genes with a p-value <0.05 were considered significantly differentially expressed. DE genes were additionally filtered according to fold-change. Genes with a fold-change > 1.5 were considered upregulated, and < -1.5 were considered downregulated in the mutant lines compared to the wildtype (i.e., control).

Mutation rate calculation

The formula, $\mu = m/n * l$ was used to calculate the per site per generation mutation rate for all 16 mutant lines, where m is the total number of mutations identified in each line, n is the total number of genomic sites with a sequencing depth >=10, and <=60, QUAL >= 20, and where

each site is at least 50-bp from the nearest indel in each mutant line. Furthermore, *l* represents the number of generations. To calculate the per gene per generation mutation rate, we used the formula $\mu_g = m_g/n_g * 1$, where m_g represents the total number of mutations detected in genic regions (including UTRs, exons, and introns), n_g represents the total number of genes analyzed in each mutant line, and *l* represents one generation. The non-synonymous mutation rate was calculated utilizing the same formula, except m_g represented the number of non-synonymous mutations per mutant line.

Annotation of EMS-induced mutations

Functional annotation based on genomic locations and effect prediction of EMS-induced heritable mutations were done using the cancer mode (-cancer) of SnpEff version 4.0 with default parameters (Cingolani *et al.* 2012). This mode was utilized since it allowed direct comparison between the EMS mutant genotypes and the wild type.

Alternative splicing analysis

We used the tool rMATS v4.1.1 (Shen *et al.* 2014) to detect alternatively spliced (AS) events using reads mapped to both exons and splice junctions. The following alternatively spliced events were detected: skipped exon (SE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), retained intron (RI), and mutually exclusive exons (MXE). SE events take place when an exon along with its flanking introns are spliced out, and A3SS and A5SS result when different parts of exons are either included or excluded from the resulting transcript. During RI events, introns are retained; during MXE events, only one of two exons are retained in the resulting mRNA (Pohl *et al.* 2013, Wang *et al.* 2015). Alternatively spliced events had to be supported by at least four uniquely mapped reads and have a minimum anchor length of 10nt.

Additionally, the Benjamini-Hochberg adjusted (FDR < 0.05) p-value had to be less than 0.05, and the difference in exon inclusion level ($\Delta|\psi|$) greater than 5% (Shen *et al.* 2014, Suresh *et al.* 2020).

Results

Whole-genome DNA and RNA sequencing data

We obtained ~6GB of raw DNA sequence data per mutant line, containing an average of ~34 million reads (SD = 3.7 million) per sample **(Supplementary Table S1)**. After removing PCR duplicates and reads mapped to multiple locations, each mutant line had on average ~29 million (SD = 2.5 million) mapped reads with an average coverage of 25 reads (SD = 2) per site. Furthermore, ~20 million (SD = 4.9 million) RNA reads were obtained per line, totaling 51 samples (3 replicates per mutant line and one wildtype (control)). After adapter trimming and quality control, reads were mapped to the *D. pulicaria* reference genome (Jackson *et al.* 2021) with an average mapping rate of 95% **(**SD = 2.37%, **Supplementary Table S2)**.

EMS-induced heritable mutation rate and spectrum

To identify EMS-induced mutations we used a rigorous bioinformatic EMS mutation identification pipeline previously tested with PCR mutation verification (false positive rate < 0.05, Snyman *et al.* 2021). The number of EMS-induced heritable heterozygous mutations in each mutant line ranged from 125 to 576 (median =152), translating to a base substitution rate ranging from 1.51 to 6.98×10^{-6} (median= 1.84×10^{-6}) per site per generation. The number of mutations found within genic regions (including UTRs, exons, and introns) ranged from 49 to 320 (median = 74) in each line, resulting in a mutation rate of 2.0×10^{-3} to 1.3×10^{-2} (median $=3.1\times10^{-3}$) per gene per generation **(Figures 2A and 2B, Supplementary Table S3)**.

Across all 16 EMS-induced mutant lines the number of non-synonymous mutations ranged from 21 to 140 (median = 29), resulting in a median non-synonymous mutation rate of 1.14×10^{-3} per gene per generation **(Figure 2B)**. The number of synonymous mutations ranged from 6 to 72 (median = 17, **Supplementary Table S3**) across mutant lines. Assuming a random distribution of EMS-induced mutations, we expect to see a non-synonymous *vs* synonymous mutation ratio of 3:1. However, in 6 of the mutant lines we observed a significant deviation (one proportion Z test p-value < 0.05) from the expected 3:1 ratio (**Figure 2 B**, Graur and Li 2000).

Lastly, in accordance with previous work on *Daphnia* (Snyman *et al.* 2021) and other model organisms such as *C. elegans* (Flibotte *et al.* 2010) and *D. melanogaster* (Pastink *et al.* 1991), the majority of EMS-induced mutations were G:C to A:T transitions (mean = 90%, SD = 4%), yielding a transition-transversion ratio greater than 5.06 for all mutant lines **(Figure 2C, Supplementary Table S4)**.

Differentially expressed genes in mutant lines

To assess how EMS-induced mutations altered the transcript abundance across the genome, we examined the transcriptomic dataset derived from 16 mutant lines in comparison to the wildtype. Our principal component analysis based on the normalized read counts (rlog transformation) showed the tight clustering of replicates derived from the same mutant, suggesting strong overall similarity among replicates of the same mutant background (**Figure 3A)**. The top two principal components (PC) were responsible for >50% of the variance, with PC1 accounting for 32% and PC2 for 21%. It was unclear what biological factors these two principal components potentially represent. Most likely, they jointly captured the variability between mutants, between sibling mutant lines, and between replicates.

Our differential expression analysis contrasting each mutant line against the wildtype revealed that the total number of differentially expressed (DE) genes ranged from 1176 to 6606 across all lines ($p < 0.05$ with FDR < 0.05), with a median of 3545. In nearly all mutant lines the ratio of down- *vs.* upregulated genes was greater than 1 **(**1.05-3.63**, Figure 3B, Supplementary Table S5)**, demonstrating that downregulated genes were more abundant.

Next, we assessed the number of DE genes carrying an EMS-induced mutation and DE genes free of a mutation. We defined that a gene carries an EMS-induced mutation if a mutation was located 5kb upstream or downstream of the gene, in the 3' UTR, 5' UTR, introns, or exons. The number of mutation-carrying DE genes in the mutant lines ranged from 5 (0.43%) to 423 (6.40%), with a median of 50.5 **(Supplementary Table S5)**. Furthermore, the ratio of down- *vs*. upregulated mutation-carrying DE genes ranged from 0.5 to 3, with mutant lines equally split between showing a majority of downregulated *vs* a majority of upregulated genes **(Figure 3B, Supplementary Table S5)**.

Except for two lines where ~6% of DE genes carried a mutation, the remaining 14 lines all showed less than 2% of DE genes carrying a mutation **(Supplementary Table S5)**. On the other hand, mutation-free DE genes in each mutant line ranged from 1171 (94.6%) to 6383 (99.6%) with a median of 3502 (). Therefore, with most DE genes being mutation-free, it strongly suggests that trans-effects drive the transcriptional changes for most DE genes and tend to result in more downregulation than upregulation, whereas cis-effects only influence a small proportion of genes, although trans-effects cannot be excluded from imposing their impacts. Lastly, cis-mutations may not alter gene expression because we identified non-DE genes carrying a mutation. Out of the total number of genes carrying a mutation, the number of genes

that were non-DE ranged from 215 (51.6%) to 1021 (77.3%) across the lines, with a median of 303.

Impact of EMS-induced mutations

Across all 16 mutant lines, the EMS-induced mutations were randomly distributed throughout the genome (chi-squared test, all p values > 0.05), i.e., the proportion of mutations in a specific type of genomic region does not exceed expectation based on random distribution of mutations. The distribution of all mutations, mutation in DE genes, and mutations in non-DE genes across the genome can be viewed in **Table 1** and **Supplementary Tables S6, S7 and S8**. The distribution of mutations among these three categories was highly similar.

The occurrence of mutation-carrying DE and non-DE genes presents an opportunity to see whether differential expression is associated with EMS-induced mutations in a specific genomic region. Using a multiple logistic regression model, we find that there was a significant association (p-value = 0.028) between differential expression and mutations in exons. Our results showed that for every mutation occurring in the exonic region of a gene, there is an increase of 0.73 in the log odds of that gene being differentially expressed, suggesting that exonic mutations are an important driver of altered gene expression.

Moreover, we categorized the mutations in DE and non-DE genes as high impact, moderate impact, low impact, or modifier variants. The mutations identified in DE genes consisted of 27 (1%) high impact variants (i.e., loss of a start codon, gain of a stop codon, splice acceptor variant), 184 (10%) moderate impact variants (i.e., missense variants), 113 (6%) low impact variants (i.e., splice region variant, synonymous variant, and variants producing a premature start codon in the 5' UTR region), and 1553 (83%) modifier variants (i.e., variants 5kb

upstream or downstream of a gene, variants within the 3' UTR and 5' UTR region, and variants within introns **(Supplementary Table S9)**.

Interestingly, for the high-impact variants, two genes carrying splice acceptor variants were both downregulated, while two other genes carrying start-lost variants were both upregulated. Out of 23 genes carrying stop-gained variants, only four were downregulated whereas the remaining 19 were upregulated **(Supplementary Figure S2)**. From the genes carrying moderate impact (i.e., missense) variants, low impact variants and modifier variants; 129, 83 and 765 were upregulated, and 55, 36 and 793 were downregulated, respectively, suggesting moderate and low impact variants tend to be more associated with upregulated expression than downregulated expression **(Supplementary Table S10)**. On the other hand, mutations identified in non-DE genes have a similar distribution of the impact categories **(Supplementary Table S11 and S12)**. It is interesting to note that even some high-impact variants such as loss of a start codon and gain of a stop codon did not result in altered expression for the mutation carrying genes.

Alternative splicing

Alternatively spliced (AS) genes were identified by comparing all 16 EMS-induced mutant lines to the wildtype. The number of AS genes ranged from 212 to 627 across all mutant lines, with a median of 393 **(Supplementary Table S13)**. The number of mutation-carrying AS genes ranged from 2 (0.5%) to 57 (9.6%) with a median of 12, while that of mutation-free AS genes ranged from 202 (~90%) to 611 (~99%) across mutant lines, with a median of 373 **(Supplementary Table S14 and 15)**. The dominance of mutation-free AS genes strongly suggests that the trans-effects caused by mutations are a major driver of alternative splicing.

A summary of the AS events detected for all AS genes, mutation-carrying AS genes and mutation-free AS genes can be viewed in **Table 2**, with the full results in **Supplementary Table S13, S14 and S15**. No significant difference (chi-square test, p-value = 0.25) was observed in the distribution of AS events between mutation-carrying and mutation-free AS genes. Lastly, across all 16 EMS-induced mutant lines, a total of 44 genes were both AS and DE (median=2**, Supplementary Table S14)**.

Impact and effect of genic *EMS-induced mutations on alternative splicing*

Mutations located in the alternatively spliced genes were further categorized according to their impact, with 6 (~3%) being high impact, 14 (~7%) low impact, 20 (~10%) moderate impact, and 159 (~80%) modifier impact variants across all 16 EMS-induced mutant lines **(Supplementary Table S16)**. The high impact variants consisted of 1 (~0.5%) splice acceptor variant, 2 (~1%) splice donor variant, and 3 (~2%) stop-gained variants. Moderate impact variants consisted of 20 (~10%) missense variants, while low impact variants consisted of 1 (~0.5%) splice region variant and 13 (~7%) synonymous variants. Modifier impact variants consisted of 1 (~0.5%) 3' UTR variant, 40 (~20%) downstream variants, 22 (~11%) intron variants, and 97 (~49%) upstream variants across all mutant lines **(Supplementary Table S17)**. Lastly, results from a logistics regression showed that there was no significant association between AS event and variant type.

Discussion

How mutations alter genome-wide gene expression and alternative splicing has been understudied, with only a few attempting to address this issue using mutation-accumulation experiments in a small number of species including *C. elegans*, *D. melanogaster*, *S. cerevisiae,* and zebrafish (Denver *et al.* 2005, Huang *et al.* 2016, Konrad *et al.* 2018, Landry *et al.* 2007,

Rifkin *et al.* 2005, White *et al.* 2022, Zalts and Yanai 2017). In this study, we examined the alterations in genome-wide gene expression and alternative splicing patterns in EMS *Daphnia* mutants to bridge this knowledge gap. Furthermore, we dissected the contribution of cis- and trans-mutations to differential gene expression and alternative splicing.

Global gene expression changes due to EMS-induced mutations

In our mutant lines the EMS-induced base substitution rate, per gene mutations rate, and non-synonymous mutation rate were estimated to be (medians) 1.84×10^{-6} per site per generation, 3.1×10^{-3} per gene per generation and 1.14×10^{-3} per gene per generation respectively **(Figure 2A, B)**. These estimates are derived from a rigorous mutation filtering pipeline that has been calibrated with PCR verification of mutations (false discovery rate <0.05) in a previous study (Snyman *et al.* 2021). These rate estimates are also consistent with our previous work (Snyman *et al.* 2021), showing mutation rates significantly higher than the spontaneous mutation rates in *Daphnia* (Bull *et al.* 2019, Flynn *et al.* 2017, Keith *et al.* 2016). Additionally, the identified EMS-induced mutations showed the expected enrichment of G:C to A:T transitions **(Figure 2C)** and random distribution across the genome **(Table 1,** Greene *et al.* 2003).

Considering that EMS may cause acute gene expression changes in the exposed females and the lingering maternal effects may alter gene expression in immediate offspring, we measured gene expression in the clonal mutant lines several generations after exposure. This method ensures that the observed gene expression changes are most likely caused by the EMS-induced mutations, rather than by maternal effects.

Our differential expression analyses reveal that downregulated genes are more abundant than upregulated genes in nearly all mutant lines, with the ratio of down- *vs.* upregulated genes

ranging between 1.05 and 3.63 **(Figure 3B, Supplementary Table S5)**. It is not clear whether mutations induced by other mutagens would cause a similar pattern as gene expression changes in mutagen-induced mutant lines remains understudied. However, as we discuss below the dominance of trans-effects in gene expression changes, the higher abundance of downregulation is likely caused by trans-acting mutations mainly acting as silencers to reduce transcription (Johnson *et al.* 2015).

Examining whether DE genes carry EMS-induced mutations across our mutant lines unveiled a significant contribution of trans-mutations to expression changes. DE genes carrying EMS-induced mutations only accounted for a small portion of the total (ranging from 0.43% to 6.40% across our mutant lines), while mutation-free DE genes were more common (ranging from 94.6% to 99.6% across mutant lines). This pattern strongly suggests that the pleiotropic effects of trans-mutations are a major contributor to the variance in gene expression between mutant and wildtype lines **(Supplementary Table S5),** while cis-mutations only play a relatively small role. The pleiotropic effects of trans-mutations were previously reported in *D. melanogaster* and *C. elegans* mutation accumulation lines, where a relatively small number of trans-acting mutations with multiple downstream effects induced the majority of the transcriptional differences (Denver *et al.* 2005, Huang *et al.* 2016). The dominant pleiotropic effects of trans-mutations are probably due to trans-factors representing a larger mutational target and are more likely to arise than cis-mutations (Chen *et al.* 2015, Coolon *et al.* 2014, Gruber *et al.* 2012, Metzger *et al.* 2016, Rhoné *et al.* 2017, Wittkopp *et al.* 2004). As it is often assumed trans-mutations have negative pleiotropic effects on fitness, unfortunately we did not collect fitness data on these lines. It would be interesting for future studies to address this issue.

The presence of cis-mutations in DE genes indicates that cis-mutations may directly cause gene expression changes. However, it cannot be ruled out that trans-effects also contribute to the variance in gene expression of these genes. It is noteworthy that the presence of cis-mutations does not necessarily alter expression because 51.6% - 77.3% of the observed genes with cis-mutations were non-DE. However, our logistic regression analysis provides a clear signal that exonic mutations are strongly associated with altered gene expression. Interestingly, it was observed in EMS mutagenized *S. cerevisiae* mutants that trans regulatory mutations were also mostly located within the coding sequences (Duveau *et al.* 2021).

It is possible that some cis-mutations do not directly impact gene expression. However, the unchanged expression of genes with cis-mutations may also be a result of complex interactions between cis- and trans-effects, for which we do not have the capacity to further test in this study. Furthermore, because the mutated sites are heterozygous for the mutation and wildtype SNP allele, the invariance of total gene expression may have concealed the signature of allele-specific expression changes caused by the EMS-induced mutations. It would be of great interest for future studies to address whether and how mutations cause the changes in allele-specific transcript abundance.

Impact of different classes of mutations on gene expression

Our results also showed that moderate-impact (missense) and low-impact variants tend to be associated with an upregulation of gene expression **(**129 upregulated genes vs 55 downregulated genes for moderate impact**,** 83 upregulated vs 36 downregulated genes for low impact **)**. Missense variants are known to alter the amino acid sequence, and have been shown to impact the interacting DNA-transcription factors, resulting in changed expression of other genes (Ding *et al.* 2015). Low-impact variants including synonymous variants have also been shown to

influence gene expression by disrupting transcription and splicing, as well as mRNA stability (Pagani *et al.* 2005, Presnyak *et al.* 2015, Stergachis *et al.* 2013, Wang *et al.* 2021).

High impact variants consisted of two downregulated splice acceptor variants, two upregulated start-lost variants, and the upregulation of most stop-gained variants **(Supplementary Figure S2)**. The downregulation of some stop-gained (nonsense) variants could indicate nonsense-mediated RNA decay (NMD) is at play. NMD is a surveillance pathway that helps maintain RNA quality and cellular homeostasis by detecting and eliminating transcripts containing premature stop codons (Nickless *et al.* 2017). The upregulation of most stop-gained variants could indicate that these genes escape NMD, that a stop-codon read-through occurs, or that transcriptional adaptation is triggered causing an upregulation of the affected as well as related genes (El-Brolosy *et al.* 2019).

On the other hand, mutation-carrying non-DE genes showed a similar distribution to DE genes **(Supplementary Table S7 and S8)**. Interestingly, some high-impact mutations in these genes did not result in an altered expression. It is plausible that these mutations simply do not impact gene expression or some of these mutations are false positives that slipped through our mutation identification pipeline, although the rate of false positives should be < 5%. Another possibility is that our RNA-seq experimental design did not provide enough experimental power to detect the expression changes of these genes. It should also be noted that only one allele of the affected genes was mutated, which means that a functional wildtype allele still exists for the affected genes. Thus, it is possible that while the mutated allele may have had reduced transcription, the wildtype allele likely experiences enhanced transcription, resulting in unchanged expression of the gene in question.

Alternative splicing

Changes in pre-mRNA splicing are another important source of phenotypic variation. If the splicing process is not correctly regulated, premature stop codons and transcripts with an altered amino acid sequence can arise. Alternative splicing (AS) can result from both cis- and trans-regulatory mutations. Cis-mutations can interact with exonic or intronic splicing enhancers or silencers, whereas trans-mutations affect splicing factors potentially impacting genes throughout the genome (Juan *et al.* 2014, Kornblihtt *et al.* 2013). To date, the impact of cis- versus trans-mutations on alternative splicing remains understudied. Most studies have suggested that cis-regulatory (Ast 2004, Keren *et al.* 2010, Thatcher *et al.* 2014) mutations may be the main contributor to alternative splicing. However, a recent study raised the opposite view, suggesting that the pleiotropic effects of trans-mutations may be the main culprit (Smith *et al.* 2018).

Our results are consistent with the latter view. Our analyses show that the number of mutation-carrying AS genes ranged from 0.5% to 9.6% across mutant lines **(Supplementary Table S14)**. Most AS genes are free of EMS-induced mutations (ranging from 90% to 99% of total AS genes), strongly suggesting that trans-effects are a major driver of AS **(Supplementary Table S15**). Furthermore, our results showed little overlap between DE and AS genes, a pattern further seen in *Drosophila* (Jakšić and Schlötterer 2016), aphids (Grantham and Brisson 2018), and salmonids (Jacobs and Elmer 2021), presumably because of non-overlapping regulation between DE and AS genes

Implications for understanding the biological impacts of anthropogenic events

As we are preparing this manuscript, on February 3rd 2023, a train derailment in East Palestine, Ohio, USA, resulted in the spill of > 100,000 gallons of vinyl chloride, a carcinogenic mutagen. Accidents like this (along with Chernobyl disaster in 1986 and Fukushima nuclear disaster in 2011) and many unreported, smaller-scale anthropogenic events occur at a sharply increased rate

since the Industrial Revolution. The epigenetic effects of some of these environmental chemicals have been investigated (reviewed in Baccarelli and Bollati 2009, Goyal *et al.* 2022, Hou *et al.* 2012). However, much work still needs to be done to determine the impact of chemically-induced mutations on gene expression and alternative splicing on the genome-wide scale. Based on our results, mutations exerting trans-effects can have a genome-wide impact by altering the expression and splicing of thousands of genes (up to 99% of the total DE and AS genes respectively in *Daphnia*), while high-impact and moderate-impact mutations could lead to amino acid changes, with potential functional implications. Understanding the consequences of these anthropogenic stressors are thus vital to protecting and preserving the health of our ecosystems and ourselves.

Conclusions

Our results show that trans-effects are the major contributor to the variance in gene expression and alternative splicing between the wildtype and EMS-induced mutant lines in *Daphnia*, while cis-mutations do not always alter gene expression and only affected a small portion of genes. Furthermore, our results showed a significant association between DE genes and exonic mutations, indicating that exonic mutations are an important driver of altered gene expression.

Acknowledgments

We thank A. Gamadia and M.T. Smith for their help with tissue collection and T. Huynh for help with script writing. We also thank two anonymous reviewers for their constructive suggestions on our manuscript. This work is supported by NIH grant R35GM133730 and NSF CAREER grant MCB-2042490 to SX.

Data Accessibility

DNA sequence data: NCBI SRA-PRJNA892919.

RNA sequence data: NCBI SRA-PRJNA892982.

Author Contributions

SX designed the study, and MS and SX wrote the manuscript. MS performed tissue collection, molecular work, and data analysis.

References

Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, *442*(7102), 563–567.

Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L (2018) Genetics of trans-regulatory variation in gene expression. *ELife*, *7*, e35471.

Anna A, Monika G (2018) Splicing mutations in human genetic disorders: Examples, detection, and confirmation. *Journal of Applied Genetics*, *59*(3), 253–268.

Aquadro CF, Tachida H, Langley CH, Harada K, Mukai T (1990) Increased variation in ADH enzyme activity in *Drosophila* mutation-accumulation experiment is not due to transposable elements at the Adh structural gene. *Genetics*, *126*(4), 915–919.

Ast G (2004) How did alternative splicing evolve? *Nature Reviews. Genetics*, *5*(10), 773–782.

Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Baccarelli A, Bollati V (2009) Epigenetics and environmental chemicals. *Current Opinion in Pediatrics*, *21*(2), 243–251.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120.

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, Genetics*, *9*(1), 61–71.

568 Chen J, Nolte V, Schlötterer C (2015) Temperature stress mediates decanalization and
569 dominance of gene expression in Drosophila melanogaster. *PLoS Genetics*, *11*(2),
570 e1004883.
571 Cherry S, Lynch KW (2020) Alternative splicing and cancer: Insights, opportunities, and
572 challenges from an expanding view of the transcriptome. *Genes & Development*, *34*(15–
573 16), 1005–1016.
574 Chu XL, Zhang QG (2021) Consequences of mutation accumulation for growth performance are
575 more likely to be resource-dependent at higher temperatures. *BMC Ecology and*
576 *Evolution*, *21*(1), 109.
577 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM
578 (2012) A program for annotating and predicting the effects of single nucleotide
579 polymorphisms, SnpEff. *Fly*, *6*(2), 80–92.
580 Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, *et al.*
581 (2011) The ecoresponsive genome of *Daphnia pulex*. *Science (New York, N.Y.)*,
582 *331*(6017), 555–561.
583 Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ (2014) Tempo and mode of
584 regulatory evolution in *Drosophila*. *Genome Research*, *24*(5), 797–808.
585 Coolon JD, Stevenson KR, McManus CJ, Yang B, Graveley BR, Wittkopp PJ (2015) Molecular
586 Mechanisms and Evolutionary Processes Contributing to Accelerated Divergence of
587 Gene Expression on the *Drosophila* X Chromosome. *Molecular Biology and Evolution*,
588 *32*(10), 2605–2615.
589 Denkena J, Johannes F, Colomé-Tatché M (2021) Region-level epimutation rates in Arabidopsis
590 thaliana. *Heredity*, *127*(2), Article 2.

591 Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, Lewis SC, Okamoto K,
592 Thomas WK, Lynch M, Baer CF (2009) A genome-wide view of *Caenorhabditis elegans*
593 base-substitution mutation processes. *Proceedings of the National Academy of Sciences*,
594 *106*(38), 16310–16314.
595 Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK (2005) The transcriptional
596 consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature*
597 *Genetics*, *37*(5), 544–548.
598 Díaz-González J, Domínguez A (2020) Different structural variants of roo retrotransposon are
599 active in *Drosophila melanogaster*. *Gene*, *741*, 144546.
600 Ding J, McConechy MK, Horlings HM, Ha G, Chun Chan F, Funnell T, Mullaly SC *et al.* (2015)
601 Systematic analysis of somatic mutations impacting gene expression in 12 tumour types.
602 *Nature Communications*, *6*, 8554.
603 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
604 TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.
605 Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf
606 tissue (RESEARCH). PHYTOCHEMICAL BULLETIN.
607 Duveau F, Vande Zande P, Metzger BP, Diaz CJ, Walker EA, Tryban S, Siddiq MA, Yang B,
608 Wittkopp PJ (2021) Mutational sources of trans-regulatory variation affecting gene
609 expression in *Saccharomyces cerevisiae*. *ELife*, *10*, e67806.
610 Edison AS, Hall RD, Junot C, Karp PD, Kurland IJ, Mistrik R, Reed LK *et al.* (2016) The Time
611 Is Right to Focus on Model Organism Metabolomes. *Metabolites*, *6*(1), 8.
612 El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Günther S, Fukuda N *et al.* (2019) Genetic
613 compensation triggered by mutant mRNA degradation. *Nature*, *568*(7751), 193–197.

614 Flibotte S, Edgley ML, Chaudhry I, Taylor J, Neil SE, Rogula A, Zapf R *et al.* (2010) Whole-
615 Genome Profiling of Mutagenesis in *Caenorhabditis elegans*. *Genetics*, *185*(2), 431–441.
616 Flynn JM, Caldas I, Cristescu ME, Clark AG (2017) Selection Constrains High Rates of Tandem
617 Repetitive DNA Mutation in *Daphnia pulex*. *Genetics*, *207*(2), 697–710.
618 Flynn JM, Chain FJJ, Schoen DJ, Cristescu ME (2017) Spontaneous Mutation Accumulation in
619 *Daphnia pulex* in Selection-Free vs. Competitive Environments. *Molecular Biology and*
620 *Evolution*, *34*(1), 160–173.
621 Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MGD, Jädersten M, Dolatshad H *et*
622 *al.* (2015) Combining gene mutation with gene expression data improves outcome
623 prediction in myelodysplastic syndromes. *Nature Communications*, *6*, 5901.
624 Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA (2015) Non-
625 adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature.
626 *Nature*, *525*(7569), 372–375.
627 Goyal K, Goel H, Baranwal P, Dixit A, Khan F, Jha NK, Kesari KK *et al.* (2022) Unravelling the
628 molecular mechanism of mutagenic factors impacting human health. *Environmental*
629 *Science and Pollution Research*, *29*(41), 61993–62013.
630 Grantham ME, Brisson JA (2018) Extensive Differential Splicing Underlies Phenotypically
631 Plastic Aphid Morphs. *Molecular Biology and Evolution*, *35*(8), 1934–1946.
632 Graur D, Li WH (2000) *Fundamentals of Molecular Evolution* (2nd Edition). Sinauer
633 Associates. Sunderland MA
634 Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC *et al.* (2003)
635 Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in
636 *Arabidopsis*. *Genetics*, *164*(2), 731–740.

637 Gruber JD, Vogel K, Kalay G, Wittkopp PJ (2012) Contrasting properties of gene-specific
638 regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: Frequency,
639 effects, and dominance. *PLoS Genetics*, *8*(2), e1002497.
640 Harada K (1995) A quantitative analysis of modifier mutations which occur in mutation
641 accumulation lines in Drosophila melanogaster. *Heredity*, *75*(6), Article 6.
642 Ho EKH, Bellis ES, Calkins J, Adrion JR, Iv LCL, Schaack S (2021) Engines of change:
643 Transposable element mutation rates are high and variable within Daphnia magna. *PLOS*
644 *Genetics*, *17*(11), e1009827.
645 Hornung G, Oren M, Barkai N (2012) Nucleosome organization affects the sensitivity of gene
646 expression to promoter mutations. *Molecular Cell*, *46*(3), 362–368.
647 Hou L, Zhang X, Wang D, Baccarelli A (2012) Environmental chemical exposures and human
648 epigenetics. *International Journal of Epidemiology*, *41*(1), 79–105.
649 Huang W, Lyman RF, Lyman RA, Carbone MA, Harbison ST, Magwire MM, Mackay TF
650 (2016) Spontaneous mutations and the origin and maintenance of quantitative genetic
651 variation. *ELife*, *5*, e14625.
652 Iofrida C, Melissari E, Mariotti V, Guglielmi C, Guidugli L, Caligo MA, Pellegrini S (2012)
653 Effects on human transcriptome of mutated BRCA1 BRCT domain: A microarray study.
654 *BMC Cancer*, *12*(1), 207.
655 Jackson CE, Xu S, Ye Z, Pfrender ME, Lynch M, Colbourne JK, Shaw JR (2021) Chromosomal
656 rearrangements preserve adaptive divergence in ecological speciation (p.
657 2021.08.20.457158).

Jacobs A, Elmer KR (2021) Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Molecular Ecology*, *30*(20), 4955–4969.

Jakšić AM, Schlötterer C (2016) The Interplay of Temperature and Genotype on Patterns of Alternative Splicing in *Drosophila melanogaster*. *Genetics*, *204*(1), 315–325.

Jia P, Zhao Z (2017) Impacts of somatic mutations on gene expression: An association perspective. *Briefings in Bioinformatics*, *18*(3), 413–425.

Jiang C, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP (2014) Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Research*, *24*(11), 1821–1829.

Johnson WC, Ordway AJ, Watada M, Pruitt JN, Williams TM, Rebeiz M (2015) Genetic Changes to a Transcriptional Silencer Element Confers Phenotypic Diversity within and between *Drosophila* Species. *PLoS Genetics*, *11*(6), e1005279.

Josephs EB (2021) Gene expression links genotype and phenotype during rapid adaptation. *Molecular Ecology*, *30*(1), 30–32.

Juan WC, Roca X, Ong ST (2014) Identification of cis-Acting Elements and Splicing Factors Involved in the Regulation of BIM Pre-mRNA Splicing. *PLOS ONE*, *9*(4), e95210.

Keightley PD, Lynch M (2003) Toward a realistic model of mutations affecting fitness. *Evolution; International Journal of Organic Evolution*, *57*(3), 683–685; discussion 686-689.

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 Research*, *19*(7), 1195–1201.

Keith N, Tucker AE, Jackson CE, Sung W, Lucas Lledó JI, Schrider DR, Schaack S *et al.* (2016) Generation of loss- or gain of function mutants and weak nonlethal alleles. *Genome Research*, *26*(1), 60–69.

Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: Diversification, exon definition and function. *Nature Reviews. Genetics*, *11*(5), 345–355.

Kilham SS, Kreeger DA, Lynn SG, Goulden CE, Herrera L (1998) COMBO: A defined freshwater culture medium for algae and zooplankton. *Hydrobiologia*, *377*(1), 147–159.

Konrad A, Flibotte S, Taylor J, Waterston RH, Moerman DG, Bergthorsson U, Katju V (2018) Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), 7386–7391.

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ (2013) Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nature Reviews Molecular Cell Biology*, *14*(3), Article 3.

Krasovec M (2021) The spontaneous mutation rate of *Drosophila pseudoobscura*. *G3 Genes|Genomes|Genetics*, *11*(7), jkab151.

Kwasnieski JC, Mogno I, Myers CA, Corbo JC, Cohen BA (2012) Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19498–19503.

Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL (2007) Genetic properties influencing the evolvability of gene expression. *Science (New York, N.Y.)*, *317*(5834), 118–121.

Latta LC, Peacock M, Civitello DJ, Dudycha JL, Meik JM, Schaack S, Baer AECF, Kalisz ES (2015) The Phenotypic Effects of Spontaneous Mutations in Different Environments. *The American Naturalist*, *185*(2), 243–252.

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 Reports*, *22*(3), 218–223.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079.

Liao Y, Smyth GK, Shi W (2014) featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

Luu DHK, Vo HP, Xu S (2020) An efficient method for hatching diapausing embryos of *Daphnia pulex* species complex (Crustacea, Anomopoda). *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, *333*(2), 111–117.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ *et al.* (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, *105*(27), 9272–9277.

Mack KL, Ballinger MA, Phifer-Rixey M, Nachman MW (2018) Gene regulation underlies environmental adaptation in house mice. *Genome Research*, *28*(11), 1636–1645.

Macknight R, Duroux M, Laurie R, Dijkwel P, Simpson G, Dean C (2002) Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter FCA. *The Plant Cell*, *14*(4), 877–888.

Manceau M, Domingues VS, Mallarino R, Hoekstra HE (2011) The developmental role of Agouti in color pattern evolution. *Science (New York, N.Y.)*, *331*(6020), 1062–1065.

Maricque BB, Dougherty JD, Cohen BA (2017) A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of cis-regulatory activity in neural cells. *Nucleic Acids Research*, *45*(4), e16.

Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ (2016) Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affecting Gene Expression. *Molecular Biology and Evolution*, *33*(5), 1131–1146.

Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ (2015) Selection on noise constrains variation in a eukaryotic promoter. *Nature*, *521*(7552), 344–347.

Nickless A, Bailis JM, You Z (2017) Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell & Bioscience*, *7*(1), 26.

Omilian AR, Cristescu MEA, Dudycha JL, Lynch M (2006) Ameiotic recombination in asexual lineages of *Daphnia*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(49), 18638–18643.

747 Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D,
748 Lynch M (2010) The Rate and Molecular Spectrum of Spontaneous Mutations in
749 *Arabidopsis thaliana*. *Science (New York, N.Y.)*, *327*(5961).
750 Pagani F, Raponi M, Baralle FE (2005) Synonymous mutations in CFTR exon 12 affect splicing
751 and are not neutral in evolution. *Proceedings of the National Academy of Sciences*,
752 *102*(18), 6368–6372.
753 Pastink A, Heemskerk E, Nivard MJ, van Vliet CJ, Vogel EW (1991) Mutational specificity of
754 ethyl methanesulfonate in excision-repair-proficient and -deficient strains of *Drosophila*
755 *melanogaster*. *Molecular & General Genetics: MGG*, *229*(2), 213–218.
756 Pohl M, Bortfeldt RH, Grützmann K, Schuster S (2013) Alternative splicing of mutually
757 exclusive exons—A review. *Biosystems*, *114*(1), 31–38.
758 Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S *et al.* (2015) Codon
759 Optimality Is a Major Determinant of mRNA Stability. *Cell*, *160*(6), 1111–1124.
760 Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB (2009) Stepwise modification of a
761 modular enhancer underlies adaptation in a *Drosophila* population. *Science (New York,*
762 *N.Y.)*, *326*(5960), 1663–1667.
763 Rhoné B, Mariac C, Couderc M, Berthouly-Salazar C, Ousseini IS, Vigouroux Y (2017) No
764 Excess of Cis-Regulatory Variation Associated with Intraspecific Selection in Wild Pearl
765 Millet (*Cenchrus americanus*). *Genome Biology and Evolution*, *9*(2), 388–397.
766 Rifkin SA, Houle D, Kim J, White KP (2005) A mutation accumulation assay reveals a broad
767 capacity for rapid evolution of gene expression. *Nature*, *438*(7065), Article 7065.

Schaefke B, Emerson JJ, Wang TY, Lu MYJ, Hsieh LC, Li WH (2013) Inheritance of gene expression level and selective constraints on trans- and cis-regulatory changes in yeast. *Molecular Biology and Evolution*, *30*(9), 2121–2133.

Scheffer H, Coate JE, Ho EKH, Schaack S (2022) Thermal stress and mutation accumulation increase heat shock protein expression in *Daphnia*. *Evolutionary Ecology*, *36*(5), 829–844.

Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y (2014) rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), E5593-5601.

Smith CCR, Tittes S, Mendieta JP, Collier-zans E, Rowe HC, Rieseberg LH, Kane NC (2018) Genetics of alternative splicing evolution during sunflower domestication. *Proceedings of the National Academy of Sciences*, *115*(26), 6768–6773.

Snyman M, Huynh TV, Smith MT, Xu S (2021) The genome-wide rate and spectrum of EMS-induced heritable mutations in the microcrustacean *Daphnia*: On the prospect of forward genetics. *Heredity*, *127*(6), 535–545.

Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, Raubitschek A *et al.* (2013) Exonic Transcription Factor Binding Directs Codon Choice and Affects Protein Evolution. *Science*, *342*(6164), 1367–1372.

Suresh S, Crease TJ, Cristescu ME, Chain FJJ (2020) Alternative splicing is highly variable among *Daphnia pulex* lineages in response to acute copper exposure. *BMC Genomics*, *21*(1), 433.

Terai Y, Morikawa N, Kawakami K, Okada N (2003) The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(22), 12798–12803.

Thatcher SR, Zhou W, Leonard A, Wang BB, Beatty M, Zastrow-Hayes G, Zhao X, Baumgarten A, Li B (2014) Genome-wide analysis of alternative splicing in Zea mays: Landscape and genetic regulation. *The Plant Cell*, *26*(9), 3472–3487.

Vande Zande P, Hill MS, Wittkopp PJ (2022) Pleiotropic effects of trans-regulatory mutations on fitness and gene expression. *Science*, *377*(6601), 105–109.

Vernia S, Edwards YJ, Han MS, Cavanagh-Kyros J, Barrett T, Kim JK, Davis RJ (2016) An alternative splicing program promotes adipose tissue thermogenesis. *ELife*, *5*, e17672.

Wang SY, Cheng YY, Liu SC, Xu YX, Gao Y, Wang CL, Wang ZG *et al.* (2021) A synonymous mutation in IGF-1 impacts the transcription and translation process of gene expression. *Molecular Therapy - Nucleic Acids*, *26*, 1446–1465.

Wang Y, Liu J, Huang B, Xu YM, Li J, Huang LF *et al.* (2015) Mechanism of alternative splicing and its regulation (Review). *Biomedical Reports*, *3*(2), 152–158.

Ward AJ, Cooper TA (2010) The Pathobiology of Splicing. The Journal of Pathology, 220(2), 152–163.

West DB, Engelhard EK, Adkisson M, Nava AJ, Kirov JV, Cipollone A, Willis B *et al.* (2016) Transcriptome Analysis of Targeted Mouse Mutations Reveals the Topography of Local Changes in Gene Expression. *PLOS Genetics*, *12*(2), e1005691.

White RJ, Mackay E, Wilson SW, Busch-Nentwich EM (2022) Allele-specific gene expression can underlie altered transcript abundance in zebrafish mutants. *ELife*, *11*, e72825.

Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in cis and trans gene regulation. *Nature*, *430*(6995), 85–88.

Xu J (2004) Genotype-Environment Interactions of Spontaneous Mutations for Vegetative Fitness in the Human Pathogenic Fungus *Cryptococcus neoformans*. *Genetics*, *168*(3), 1177–1188.

Xu S, Schaack S, Seyfert A, Choi E, Lynch M, Cristescu ME (2012) High Mutation Rates in the Mitochondrial Genomes of *Daphnia pulex*. *Molecular Biology and Evolution*, *29*(2), 763–769.

Zalts H, Yanai I (2017) Developmental constraints shape the evolution of the nematode mid-developmental transition. *Nature Ecology & Evolution*, *1*(5), Article 5.

Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences*, *111*(22), E2310–E2318.

827 **Figure captions.**

828 **Figure 1.** Illustration of exposure method to generate EMS mutant lines.

829 **Figure 2.** (A) The base substitution rates. (B) Non-synonymous and synonymous mutation rate, and non-synonymous *vs* synonymous
830 mutation ratio in all mutant lines. (C) The proportion of base substitutions caused by EMS-induced mutations.

831 **Figure 3.** (A) PCA plot based on transcriptomic data of mutant lines. (B) Log2 fold-change of DE genes (red dots) and DE genes with
832 mutations (blue dots) with the ratio of down- *vs.* upregulated DE genes for each line.

833

834

835

836 **Table 1.** The number of mutations found in different genomic regions.

Category	5kb downstream region	5kb upstream region	Exons	Intergenic regions	Introns	Splice Sites	3'UTR	5'UTR
All mutations	3047 (~31%)	3142 (~32%)	1068 (~11%)	1762 (~18%)	538 (~5%)	58 (~0.6%)	144 (~1.5%)	147 (~1.5%)
Mutation in DE genes	650 (~34%)	698 (~37%)	311 (~17%)	N/A	140 (~7%)	16 (~1%)	43 (~2%)	33 (~2%)
Mutation-in non-DE genes	2397 (~38%)	2444 (~39%)	757 (~12%)	N/A	398 (~6%)	42 (~1%)	101 (~2%)	114 (~2%)

837

838

839 **Table 2.** Summary of alternative splicing events of genes.

Category	A3SS	A5SS	MXE	RI	SE
All AS genes	665 (~11%)	665 (~11%)	1051 (~17%)	2027 (~32%)	1880 (~30%)
Mutation-carrying AS genes	20 (~8%)	22 (~9%)	48 (~19%)	55 (~22%)	104 (~42%)
Mutation-free AS genes	645 (~11%)	643 (~11%)	1003 (~17%)	1972 (~33%)	1776 (~29%)

840