INVESTIGATING GENE EXPRESSION CHANGES IN DAPHNIA: THE ROLE OF EMS-INDUCED MUTATIONS AND REPRODUCTIVE MODES.

By

MARELIZE SNYMAN

DISSERTATION

Submitted in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

at

The University of Texas at Arlington

August 2022

Arlington, Texas

Supervising Committee:

Sen Xu, Supervising Professor

Ester Betran

Todd Castoe

Matthew Walsh

Matthew Fujita

COPYRIGHT

ACKNOWLEDGEMENT

I would like to thank the following people, without whom I would not have been able to complete my doctoral degree. First, I would like to thank my supervisor, Dr. Sen Xu, for his guidance, expertise, support, and patience throughout this program. I would also like to thank each member of my dissertation committee, Dr. Ester Betran, Dr. Todd Castoe, Dr. Matthew Walsh, and Dr. Matthew Fujita for your support and guidance over the last 5 years. I would like to thank my colleagues, Thinh Pham, Dr. Swatantra Neupane, Trung Huynh, and Dr. Honjun Wang for your support and all the helpful discussions. I will always treasure our "tea-times", Panera coffee walks, and Gilligans nights. Further, I would like to thank all the undergraduate students in the lab, especially Matt Smith and Anish Gamadia who helped me with animal care and contributed to the experiments, your help is deeply appreciated. I also appreciate all the support that I received from my friends, especially Oswald Jenewein, thank you for all the phone calls, coffee breaks, walks, and writing sessions. They helped keep me sane during this busy and stressful time. The biggest thank you go to my family for your support and encouragement during the last 5 years. Without you, this thesis would not have been possible.

DEDICATION

This work is dedicated to my mother, Isabel Snyman, who shares my love for biology, and my grandmother, Dina van Vuuren. Thank you for your love, encouragement, and support during the last 5 years.

ABSTRACT

Gene expression variation is a common source of phenotypic variation between species. Understanding how changes in gene expression are associated with a particular phenotype will aid in understanding the molecular basis and evolution of complex traits. For the first chapter, I investigated the role that changes in gene expression play in the origin of obligate parthenogenesis using the microcrustacean *Daphnia pulex*. I conducted a genome-wide differential expression and splicing analysis comparing early subitaneous and early resting egg production in OP *D. pulex* isolates to investigate the genes and mechanisms underlying these parthenogenetic modes. Results from the KEGG pathway and GO term enrichment analysis revealed that early subitaneous egg production is associated with an upregulation of meiosis and cell-cycle genes as well as genes mapped to sugar and lipid metabolic processes. Downregulated genes were enriched in various metabolic processes, biosynthesis, and signaling pathways. For the second chapter, I developed a forward genetic screening approach utilizing EMS mutagenesis in *Daphnia* to study gene function. First, I showed that 10mM and 25mM EMS concentrations significantly elevated the base substitution rate of multiple *Daphnia* species to 1.17×10^{-6} and 1.75×10^{-6} per site per generation respectively, and that the base substitutions were dominated by G:C to A:T transitions. Furthermore, we showed that EMS-induced mutations were present in the first three consecutive broods of an exposed female, and additionally recommend screening 4-5 F_2s generated from sibling crosses per F_1 mutant line to detect EMS-induced mutations in the homozygous state with a 70-80% probability. For the last chapter, I studied the effects of EMS-induced mutations on gene expression. First, EMS-induced mutations were identified along with differentially expressed (DE) and spliced (DS) genes. Next, we showed that among the DE and

DS genes, a median of 51 and 12 genes were impacted by EMS-induced mutations per mutant line respectively. Of the DE genes, most of these variants were modifier variants (83%), followed by moderate impact variants (10%), low impact variants (6%), and high impact variants (1%). The DS genes were mainly impacted by modifier variants (64%) and moderate impact variants (22%).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction

Gene expression patterns within genomes are usually highly regulated since changes can significantly impact tissue differentiation, organogenesis, and development (Emerson 2002). Changes in gene expression are also a common source of phenotypic variation between species and understanding how changes in gene expression are associated with a particular phenotype will aid in understanding the molecular basis and evolution of complex traits. First, the following dissertation aims to understand the role that changes in gene expression play in the origin of obligate parthenogenesis, and second, investigate the effect that ethyl methanesulfonate (EMS) induced mutations have on gene expression in the microcrustacean *Daphnia*.

Origin of Obligate Parthenogenesis

Numerous studies have been conducted on the evolution of sexual reproduction and the emergence of asexuality (Smith 1971, 1978, Vrijenhoek 1979, Kondrashov 1988, Crow 1994, Peck 1994, Doncaster *et al.* 2000, Peck and Waxman 2000, Pound *et al.* 2002, 2004, Paland and Lynch 2006). Currently, sexual reproduction is the most predominant mode of reproduction in eukaryotes; however, the ability to reproduce asexually has developed in all major lineages. Most animal taxa, including cladocerans, monogonont rotifers, aphids, protists, cnidarians, bryozoans, and plants, contain cyclical parthenogens (CP) where they switch between sexual and asexual reproduction (De Meester *et al.* 2004). On the branch tips of the evolutionary tree resides some species that reproduce strictly via obligate parthenogenesis (OP). These species have emerged in most multicellular taxa yet, make up less than 1% of the animal kingdom due to being prone to extinction (Bell, 1982, Howard and Lively 1994, Lynch *et al.* 1993, Park and Krug 2013, Smith 1978). Various well-studied cytological modifications of meiosis could lead to

obligate parthenogenesis, such as automictic parthenogenesis, apomictic parthenogenesis, gynogenesis, and hybridogenesis (Stenberg and Saura 2009, Vrijenhoek 1998). However, the genes and mechanisms underlying these modifications remain unknown.

Further, obligate parthenogenesis can arise through four main origins nl. spontaneous mutations, hybridization, infection by microorganisms such as *Wolbachia*, or the spread of asexuality-conferring elements (Huigens *et al*. 2000, Lynch 1984, Neiman *et al*. 2014, Simon *et al*. 2003). Of the mentioned routes, hybridization is the most prevalent as numerous vertebrates, including fish, amphibians, and reptiles (Avise 2008, Avise 2015, Dawley and Bogart 1989, Neaves and Baumann 2011), and invertebrates including snails, crustaceans, and insects are obligate parthenogens with a hybrid genomic background (Johnson and Bragg 1999, Innes and Hebert 1988, Schwander *et al*. 2011, Stenberg and Lundmark 2004, White *et al.* 1977).

With most obligate parthenogens being hybrids, hybridization may play a vital role in the origin of obligate parthenogenetic species. In the second chapter of this dissertation, I investigated the genomic signatures of obligate parthenogenesis in the cladoceran microcrustacean, *Daphnia pulex*. Within the *Daphnia pulex* species complex, OP *D. pulex* isolates emerged via ancestral hybridization and introgression events between two cyclical parthenogenetic (CP) sister species, CP *Daphnia pulex* and *Daphnia pulicaria* (**Figure 1.1,** Decaestecker *et al.* 2009, Xu *et al.* 2013, 2015). These two sister species diverged about 800,000 – 2,000,000 years ago (Colbourne and Hebert 1996, Cristescu *et al.* 2012, Omilian and Lynch 2009), are morphologically similar (Brandlova *et al.* 1972), but inhabit very different environments. *Daphnia pulicaria* is commonly found in permanent stratified lakes, while CP *D. pulex* isolates inhabit ephemeral ponds and permanent fishless habitats (Crease *et al.* 1997, Pfrender *et al.* 2000).

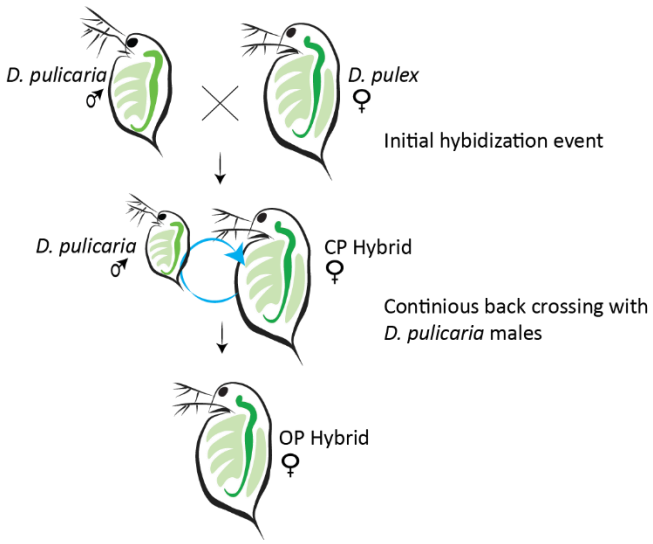


Figure 1.1. Hybridization and introgression events leading to obligate parthenogenetic *D. pulex* isolates.

Daphnia usually reproduces via cyclical parthenogenesis (switch between sexual and asexual reproduction); however obligate parthenogenetic isolates have been identified **(Figure 1.2 A and B)**. During favorable environmental conditions (low population density, high food abundance, higher temperatures, longer day length), both OP and CP *Daphnia* isolates produce subitaneous eggs through a modified meiosis **(Figure 1.3 A)**. During meiosis I, cell division arrests before the onset of anaphase I; thus, no segregation of homologous chromosomes or cytokinesis occurs. Meiosis II results in diploid eggs (Hiruta *et al.* 2010, Zaffagnini and Sabelli 1972, Ojima 1955). These eggs are deposited into the female brood chamber, where they continue developing for about three days before the offspring, genetically identical to their mother, are released. Once the environment deteriorates (high population density, colder temperatures, low food availability, shorter day length), both OP and CP isolates produce resting eggs, however, CP isolates follow conventional meiosis **(Figure 1.3 B)** and produce haploid eggs requiring fertilization before being deposited into a protective case i.e., ephippium (Innes and

Hebert 1988, Lynch 1984). In OP *Daphnia* isolates, resting eggs, similar to subitaneous eggs, are diploid and produced via a modified meiosis.

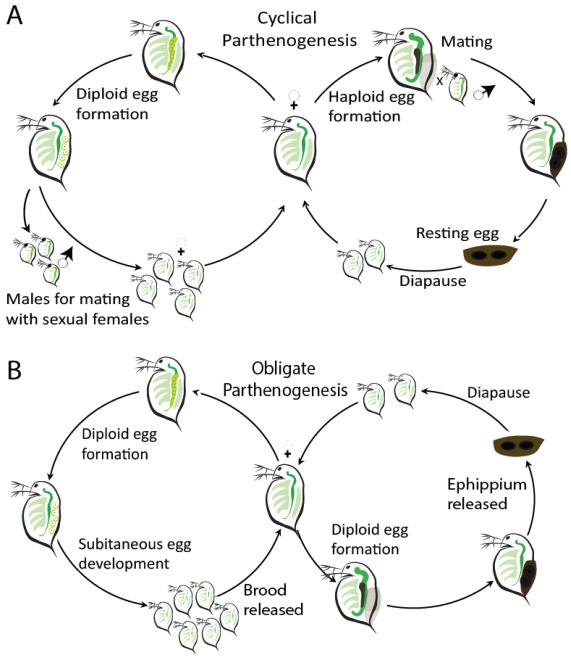


Figure 1.2. (A) Cyclic parthenogenesis and (B) obligate parthenogenesis.

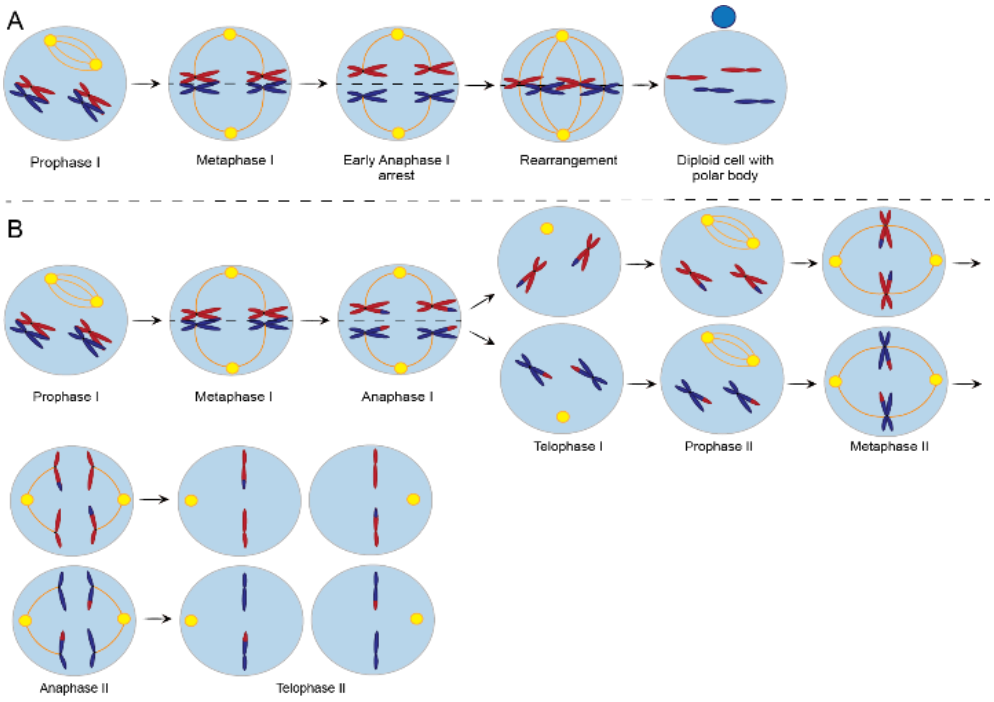


Figure 1.3. (A) Modified meiosis (adapted from Hiruta *et al.* 2010) and (B) conventional meiosis.

In the second chapter of this dissertation, I investigated the two parthenogenetic modes OP *D. pulex* isolates utilize to produce subitaneous and resting eggs. I conducted a genome-wide differential expression, differential splicing, and functional analysis between early subitaneous egg and early resting egg production to identify pathways, mechanisms, and candidate genes that may play a fundamental role in the evolution of obligate parthenogenesis.

EMS Mutagenesis

For the third chapter, I developed a forward genetic screening approach using the chemical mutagen, ethyl methanesulfonate (EMS), to study gene function in *Daphnia*. The mutagenic properties of EMS were first demonstrated by Brookes and Lawley in 1961. They showed that EMS primarily causes the alkylation of guanine, leading to O^6 ethylguanine mispairing with thymine instead of cytosine in subsequent replications. This makes EMS mutagenesis characteristically biased towards G:C to A:T transitions (Coulondre and Miller 1977). To a lesser extent, EMS can additionally induce indels (insertions and/or deletion) and chromosomal breaks (Sega 1984, Greene *et al.* 2003).

Forward genetic screening approaches utilizing EMS have successfully been used in a variety of other model organisms, including *Caenorhabditis elegans* (Brenner 1974), *Drosophila melanogaster* (Lewis and Bacher 1968), and *Arabidopsis thaliana* (McCallum *et al.* 2000, Greene *et al.* 2003, Martín *et al.* 2009) as a gene function discovery tool. With about a third of *Daphnia* genes lacking annotations due to being lineage-specific and lacking orthologs in other eukaryotic genomes (Colbourne *et al.* 2011, Ye *et al.* 2017), utilizing this approach will help us gain insight into important topics such as evolution, adaptation, as well as the genetic basis for novel phenotypes such as obligate parthenogenesis.

In the third chapter of this dissertation, I designed an EMS mutagenesis protocol for *Daphnia* utilizing 10mM and 25mM EMS concentrations and showed that offspring from the first three consecutive broods of a mutagenized female all contain unique mutations. Further, I described how this protocol could be utilized as part of a forward genetic screening approach to generate and screen F_1 or F_2 mutants for a phenotype of interest. In a future study we will employ this approach to generate and identify *Daphnia* mutants with reproductive abnormalities to further study the genes and mechanisms underlying obligate parthenogenesis. In the fourth and final chapter, I expanded on this study and examined the impact of these EMS-induced mutations on gene expression.

References

Avise J (2008) *Clonality: The Genetics, Ecology, and Evolution of Sexual Abstinence in Vertebrate Animals*. Oxford University Press, USA.

Avise JC (2015) Evolutionary perspectives on clonal reproduction in vertebrate animals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), 8867–8873.

Bell G (1982) *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. University of California Press.

Brandlova J, Brandl Z, Fernando CH (1972) The Cladocera of Ontario with remarks on some species and distribution. *Canadian Journal of Zoology*, *50*(11), 1373–1403.

Brenner S (1974) The Genetics of *Caenorhabditis Elegans*. *Genetics*, *77*(1), 71–94.

Brookes P, Lawley PD (1961) The reaction of mono- and di-functional alkylating agents with nucleic acids. *Biochemical Journal*, *80*(3), 496–503.

Colbourne JK, Hebert PDN (1996) The systematics of North American *Daphnia* (Crustacea: Anomopoda): a molecular phylogenetic approach. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1337), 349–360.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, *et al*. (2011) The ecoresponsive genome of *Daphnia pulex*. *Science (New York, N.Y.)*, *331*(6017), 555–561.

Coulondre C, Miller JH (1977) Genetic studies of the lac repressor. III. Additional correlation of mutational sites with specific amino acid residues. *Journal of Molecular Biology*, *117*(3), 525–567.

Crease TJ, Lee SK, Yu SL, Spitze K, Lehman N, Lynch M (1997) Allozyme and mtDNA variation in populations of the *Daphnia pulex* complex from both sides of the Rocky Mountains. *Heredity*, *79*(3), 242–251.

Cristescu ME, Constantin A, Bock DG, Cáceres CE, Crease TJ (2012) Speciation with gene flow and the genetics of habitat transitions. *Molecular Ecology*, *21*(6), 1411–1422.

Crow JF (1994) Advantages of sexual reproduction. *Developmental Genetics*, *15*(3), 205–213.

Dawley RM, Bogart JP (1989) *Evolution and Ecology of Unisexual Vertebrates*. University of the State of New York, State Education Department, New York State Museum.

De Meester L, Gomez A, Simon JC (2004) Evolutionary and ecological genetics of cyclical parthenogens. In *Evolution: From molecules to ecosystems.* (pp. 122–134). Oxford University Press; Oxford.

Decaestecker E, De Meester L, Mergeay J (2009) Cyclical Parthenogenesis in *Daphnia*: Sexual Versus Asexual Reproduction. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost Sex: The Evolutionary Biology of Parthenogenesis* (pp. 295–316). Springer Netherlands.

Doncaster CP, Pound GE, Cox SJ (2000) The ecological cost of sex. *Nature*, *404*(6775), 281–285.

Emerson BM (2002) Specificity of Gene Regulation. *Cell*, *109*(3), 267–270.

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*. *Genetics*, *164*(2), 731–740.

Hiruta C, Nishida C, Tochinai S (2010) Abortive meiosis in the oogenesis of parthenogenetic *Daphnia pulex*. *Chromosome Research*, *18*(7), 833–840.

Howard RS, Lively CM (1994) Parasitism, mutation accumulation and the maintenance of sex. *Nature*, *367*(6463), 554–557.

Huigens ME, Luck RF, Klaassen RHG, Maas MFPM, Timmermans MJTN, Stouthamer R (2000) Infectious parthenogenesis. *Nature*, *405*(6783), 178–179.

Innes DJ, Hebert PDN (1988) The Origin and Genetic Basis of Obligate Parthenogenesis in *Daphnia Pulex*. *Evolution*, *42*(5), 1024–1035.

Kondrashov AS (1988) Deleterious mutations and the evolution of sexual reproduction. *Nature*, *336*(6198), 435–440.

Lewis EB, Bacher F (1968) Method of feeding ethylmethane sulfonate (EMS) to *Drosophila* males.

Lynch M (1984) Destabilizing Hybridization, General-Purpose Genotypes and Geographic Parthenogenesis. *The Quarterly Review of Biology*, *59*(3), 257–290.

Lynch M, Bürger R, Butcher D, Gabriel W (1993) The mutational meltdown in asexual populations. *The Journal of Heredity*, *84*(5), 339–344.

Martín B, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C (2009) A high-density collection of EMS-induced mutations for TILLING in Landsberg erecta genetic background of *Arabidopsis*. *BMC Plant Biology*, *9*, 147.

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. *Nature Biotechnology*, *18*(4), 455–457.

Neaves WB, Baumann P (2011) Unisexual reproduction among vertebrates. *Trends in Genetics*, *27*(3), 81–88.

Neiman M, Sharbel TF, Schwander T (2014) Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. *Journal of Evolutionary Biology*, *27*(7), 1346–1359.

Omilian AR, Lynch M (2009) Patterns of Intraspecific DNA Variation in the *Daphnia* Nuclear Genome. *Genetics*, *182*(1), 325–336.

Paland S, Lynch M (2006) Transitions to asexuality result in excess amino acid substitutions. *Science (New York, N.Y.)*, *311*(5763), 990–992.

Park SC, Krug J (2013) Rate of adaptation in sexuals and asexuals: A solvable model of the Fisher-Muller effect. *Genetics*, *195*(3), 941–955.

Peck JR (1994) A ruby in the rubbish: Beneficial mutations, deleterious mutations and the evolution of sex. *Genetics*, *137*(2), 597–606.

Peck JR, Waxman D (2000) Mutation and sex in a competitive world. *Nature*, *406*(6794), 399–404.

Pfrender ME, Spitze K, Lehman N (2000) Multi-locus genetic evidence for rapid ecologically based speciation in *Daphnia*. *Molecular Ecology*, *9*(11), 1717–1735.

Pound GE, Cox SJ, Doncaster CP (2004) The accumulation of deleterious mutations within the frozen niche variation hypothesis. *Journal of Evolutionary Biology*, *17*(3), 651–662.

Pound GE, Doncaster CP, Cox SJ (2002) A lotka-volterra model of coexistence between a sexual population and multiple asexual clones. *Journal of Theoretical Biology*, *217*(4), 535–545.

Simon JC, Delmotte F, Rispe C, Crease T (2003) Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society*, *79*(1), 151–163.

Smith JM (1971) The Origin and Maintenance of Sex. In *Group Selection*. Routledge.

Smith JM (1978) Optimization Theory in Evolution. *Annual Review of Ecology and Systematics*, *9*(1), 31–56.

Stenberg P, Saura A (2009) Cytology of Asexual Animals. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost Sex: The Evolutionary Biology of Parthenogenesis* (pp. 63–74). Springer Netherlands.

Vrijenhoek R (1998) Animal Clones and Diversity. *Bioscience*, *48*, 617–628.

Vrijenhoek RC (1979) Factors Affecting Clonal Diversity and Coexistence. *American Zoologist*, *19*(3), 787–797.

Xu S, Innes DJ, Lynch M, Cristescu ME (2013) The role of hybridization in the origin and spread of asexuality in *Daphnia*. *Molecular Ecology*, *22*(17), 4549–4561.

Xu S, Spitze K, Ackerman MS, Ye Z, Bright L, Keith N, Jackson CE, Shaw JR, Lynch M (2015) Hybridization and the Origin of Contagious Asexuality in *Daphnia pulex*. *Molecular Biology and Evolution*, *32*(12), 3215–3225.

Ye Z, Xu S, Spitze K, Asselman J, Jiang X, Ackerman MS, Lopez J, Harker B, Raborn RT, Thomas WK, Ramsdell J, Pfrender ME, Lynch M (2017) A New Reference Genome Assembly for the Microcrustacean *Daphnia pulex*. *G3: Genes|Genomes|Genetics*, *7*(5), 1405–1416.

Zaffagnini F, Sabelli B (1972) Karyologic observations on the maturation of the summer and winter eggs of *Daphnia pulex* and *Daphnia middendorffiana*. *Chromosoma*, *36*(2), 193–203.

Ojima (1955) A cytological study on the development and maturation of the parthenogenetic and sexual eggs of *Daphnia pulex* (Crustacea-Cladocera). *Undefined*.

CHAPTER 2: On the Origin of Obligate Parthenogenesis in *Daphnia pulex*

Authors: Marelize Snyman and Sen Xu*

Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019, USA.

*Corresponding author. 501 S. Nedderman Dr., Arlington, Texas, 76019. Email: sen.xu@uta.edu

Keywords: *Daphnia*, obligate parthenogenesis, meiosis, cell-cycle

Running head: Origin of Obligate Parthenogenesis in *Daphnia pulex*.

Abstract

Despite the presence of obligate parthenogenetic (OP) lineages derived from sexual ancestors in diverse phylogenetic groups, the genetic mechanisms giving rise to obligate asexual hybrids remain poorly understood. The freshwater microcrustacean *Daphnia pulex* typically reproduces via cyclic parthenogenesis. However, some populations of OP *D. pulex* have emerged due to ancestral hybridization and introgression events between two cyclic parthenogenetic (CP) sister species *D. pulex* and *D. pulicaria*. These OP hybrids produce both subitaneous and resting eggs parthenogenetically, deviating from CP isolates where resting eggs are produced via conventional meiosis. This study examines the genome-wide expression and differential splicing patterns of early subitaneous and early resting egg production in OP *D. pulex* to gain insight into the genes and mechanisms underlying this transition to obligate parthenogenesis. Our differential expression, KEGG pathway, and GO term enrichment analysis revealed an upregulation of meiosis and cell-cycle genes and genes mapped to sugar and lipid metabolic processes in early subitaneous egg production. Downregulated genes were mainly enriched in various metabolic, biosynthesis, and signaling pathways. Enriched pathways of particular interest were the arginine metabolic process and sphingolipid metabolism, both pathways shown to play a role in controlling reproductive mode determination. Functional annotation of differentially spliced transcripts revealed further enrichment for the arginine metabolic process, indicating its relevance to reproductive mode initiation. Lastly, we compiled a list of differentially expressed meiosis and cell-cycle candidate genes for further investigation into their role in differentiating these two parthenogenetic modes.

Introduction

The existence of obligate parthenogenetic eukaryotic lineages that have entirely abandoned sexual reproduction has long fascinated evolutionary biologists. The transition from sexual reproduction to obligate parthenogenesis is phylogenetically widespread, having occurred independently in most multicellular taxa (Bell 2019, Liegeois *et al.* 2021, Neiman *et al.* 2014, van der Kooi *et al.* 2017). Various cytogenetic manifestations of obligate parthenogenesis have been well described in the literature, including automictic parthenogenesis, apomictic parthenogenesis, gynogenesis, and hybridogenesis (Stenberg and Saura 2009, Vrijenhoek 1998). However, the molecular mechanisms and genes underlying these cytogenetic modifications are poorly understood (Ferree *et al.* 2006, King and Hurst 2010, Riparbelli *et al.* 2005, Suomalainen *et al.* 1987).

Obligate parthenogenesis can originate from multiple evolutionary routes. Spontaneous mutations in meiosis or other reproductive genes could lead to the loss of sexual reproduction as in monogonont rotifers (Serra and Snell 2009). Contagious parthenogenesis could result due to the spread of asexuality conferring elements as in the pea aphid (Jaquiéry *et al.* 2014), and parasite-induced parthenogenesis could occur in haploid organisms such as wasps infected by *Wolbachia* (Simon *et al.* 2003). The most common route, interspecific hybridization, could disrupt meiosis due to genetic incompatibilities between parental species resulting in the loss of sex (Vrijenhoek 1998, White 1978). Among vertebrates, there are currently around 100 known asexual lineages of amphibians, reptiles, and fish due to interspecific hybridization (Avise 2008, Avise 2015, Dawley and Bogart 1989, Neaves and Baumann 2011). For invertebrates, the occurrence of hybrid asexuals have been demonstrated in snails (Johnson and Bragg 1999),

crustaceans (Innes and Hebert 1988), as well as many insects (Schwander *et al*. 2011, Stenberg and Lundmark 2004, White *et al.* 1977).

This widespread occurrence of obligate asexuals with a hybrid ancestry suggests that meiotic modifications due to hybridization may provide insight into the origin of obligately parthenogenetic (OP) species. In this study, we investigate the origin of obligate parthenogenesis in the cladoceran microcrustacean, *Daphnia pulex*, commonly found in freshwater habitats in North America. Previous studies have shown that obligate parthenogenetic *Daphnia pulex* originated due to ancestral hybridization and introgression events between two cyclic parthenogenetic (CP) sister species, *D. pulex* and *D. pulicaria* (Decaestecker *et al.* 2009, Xu *et al.* 2013, 2015). This hypothesis is supported by genome-wide association studies showing microsatellite and SNPs alleles on chromosomes 8 and 9 of OP *D. pulex*, usually only present in *D. pulicaria*, to be associated with obligate parthenogenesis (Lynch *et al.* 2008, Xu *et al.* 2015). No reproductive mode tests performed on *D. pulicaria* isolates have revealed OP lineages (Heier and Dudycha 2009). Having diverged about 800,000 – 2,000,000 years ago (Colbourne and Hebert 1996, Cristescu *et al.* 2012, Omilian and Lynch 2009), these two members of the *D. pulex* species complex are morphologically similar (Brandlova *et al.* 1972) yet can be distinguished due to inhabiting very different environments and utilizing microsatellite markers or allozyme loci such as the lactate dehydrogenase (Ldh) locus (Cristescu *et al.* 2014). CP *D. pulex* isolates, homozygous for the slow allele of Ldh (*SS*), are commonly found in ephemeral ponds and occasionally in permanent fishless habitats. In contrast, *D. pulicaria* isolates, homozygous for the fast allele of Ldh (*FF*), inhabit permanent stratified lakes (Crease *et al.* 1997, Cristescu *et al.* 2014, Pfrender *et al.* 2000). OP *D. pulex* isolates, heterozygous *SF* at the Ldh locus, have mainly been identified in areas that have undergone deforestation or where water bodies have been

contaminated with heavy metals (Hebert and Crease 1983, Hebert and Finston 2001, Shaw *et al.* 2007).

The main difference between the cyclically parthenogenetic sister species and obligately parthenogenetic hybrids is how they reproduce. Under favorable environmental conditions such as high food abundance, low population density, longer day length, and warmer temperatures, both cyclic parthenogenetic (CP) and obligate parthenogenetic (OP) *Daphnia* females (**Figure 2.1A, Figure 2.1B**) produce diploid subitaneous eggs through a modified meiosis. During meiosis I, bivalents align at the metaphase plate; however, cell division is arrested before the onset of anaphase I, after which each half-bivalent moves back to the metaphase plate and sister chromatids rearrange (Hiruta *et al.* 2010). Thus, during meiosis I, there is no segregation of homologous chromosomes and no cytokinesis resulting in daughter cells. Next, meiosis II proceeds and results in diploid embryos (Ojima 1958, Zaffagnini and Sabelli 1972, Hiruta *et al.* 2010). In the absence of rare events such as mitotic recombination, conversion, and mutation (Omilian *et al.* 2006, Xu *et al.* 2009), all offspring will be genetically identical to the mother.

Once the environment deteriorates (e.g., low food abundance, high population density, short day length, colder temperatures), some subitaneous embryos can develop into males through environmental sex determination (Gorr *et al.* 2006, Tatarazako *et al.* 2003), and CP *Daphnia* females start to produce haploid eggs (usually two) via meiosis, which upon fertilization by sperm become resting embryos (Innes and Hebert 1988, Lynch 1984). In contrast, OP *D. pulex* females produce chromosomally unreduced resting embryos without mating through a modified meiosis. Previous cytological observations have shown that parthenogenesis leading to subitaneous eggs in OP and CP isolates is highly similar to parthenogenesis leading to resting eggs in OP isolates (Zaffagnini and Sabelli 1972). The resulting resting embryos from both OP

and CP isolates are deposited into a protective case (i.e., ephippium) and remain dormant until environmental conditions turn favorable for hatching. Thus, environmental conditions are the main factor determining the reproductive mode in *Daphnia*, and that transcriptomic changes due to environmental variation may play a fundamental role and provide insight into the origin of obligate parthenogenesis.

In a previous study, we examined genome-wide expression differentiation in the early resting egg production stage between various OP *D. pulex* and parental CP *D. pulex* and *D. pulicaria* isolates **(Figure 2.2A).** Our results showed that early resting egg production in OP *D. pulex* is associated with a downregulation of meiosis and cell-cycle genes and an upregulation of metabolic and biosynthesis genes compared to early resting egg production in CP isolates (Xu *et al.* 2022), suggesting these gene expression changes are critical for the transition from sexual to asexual reproduction. Interestingly, when comparing the gene expression patterns between early subitaneous egg and early resting egg production in CP *D. pulex* and *D. pulicaria* isolates **(Figure 2.2B)**, meiosis and cell-cycle genes were downregulated while metabolic, and biosynthesis genes were upregulated in early subitaneous egg development (Huynh *et al.* 2021). Together, the results from these two studies strongly suggest that modes of parthenogenesis in both OP and CP isolates are associated with a downregulation of meiosis and cell-cycle genes and an upregulation of metabolic and biosynthesis genes and that this expression pattern may play an essential role in triggering parthenogenesis.

With our previous work examining expression differences between parthenogenesis and meiosis within CP isolates, as well as resting egg production between OP and CP isolates (Huynh *et al.* 2021, Xu *et al.* 2022), this work focuses on understanding the gene expression variation differentiating the two modes of parthenogenesis within hybrid OP *D. pulex* isolates

(Figure 2.2C). Previous cytological observations have shown that both subitaneous eggs and resting eggs undergo a single equational division where diploidy is maintained (Zaffagnini and Sabelli 1972). Yet, little is understood regarding the modifications made to the underlying genes and mechanisms allowing for meiosis I to be aborted and a single division to take place. With environmental conditions dictating the mode of reproduction, analyzing gene expression changes during the initiation of these two parthenogenetic modes will allow insight into the genes and mechanisms at play.

In this study, we investigate the differences in gene expression profiles between early subitaneous egg and early resting egg production via a pooled transcriptomic analysis as well as within three OP *D. pulex* isolates. Furthermore, we set out to identify patterns of differential splicing, as well as compile a list of candidate genes for further investigation that may give insight into the genes and mechanisms differentiating between subitaneous and resting egg production. With both subitaneous and resting eggs produced via parthenogenesis in hybrid OP *D. pulex* isolates, we hope to gain additional insight into the differences and similarities of these two mechanisms and the modifications required to transition from cyclical to obligate parthenogenesis.

Materials and Methods

Sampling of Isolates

A total of three obligate parthenogenetic (OP) *Daphnia pulex* isolates (DB4-4, K09, and Main 348-1) were used in this study. These isolates were previously collected from Texas, Ontario, and Maine, respectively **(Supplementary Table S2.1)** and have been maintained in the lab as clonal cultures. Initiated from a single female, each isolate has been kept as an asexually

reproducing line in artificial lake water (Kilham *et al.* 1998) under a 16:8 light/dark cycle at 18°C and fed with the green algae, *Scenedesmus obliquues*, twice a week.

Animal tissue collection

For each OP *D. pulex* isolate, experimental animals were maintained in the same environmental conditions for two generations to minimize maternal effects, which could significantly impact gene expression. Then, one-day-old neonates were continuously collected from each isolate and grown until sexual maturity in the same environmental conditions described above. Sexually mature animals were examined daily under a light microscope to collect females engaging in the early stage of producing subitaneous eggs and in the early stage of producing resting eggs. Early subitaneous embryo production is characterized by a thin clear, or greenish ovary extending along the gut with oil droplets visible. In contrast, early resting egg production is characterized by a small milky brown ovary starting to develop along the end of the gut **(Figure 2.1C and D)**. For each isolate, three replicates of each stage were collected (15-20 individuals) for RNA extraction.

RNA extraction and sequencing

RNA of all samples was extracted using the Promega (Madison, WI, USA) SV Total RNA Isolation kit, following the manufacturer's instructions. RNA concentration was measured using a Qubit 4.0 Fluorometer (Thermo Scientific, Waltham, MA, USA). RNA integrity was checked with a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA sequencing libraries were prepared by Novogene Corporation Inc. (Sacramento, CA, USA) following standard Illumina sequencing library protocol. Each library was sequenced on an Illumina NovaSeq6000

platform with at least 20 million 150-bp paired-end reads. This project's raw RNA sequence data were deposited at NCBI SRA under PRJNA847604.

Sequencing quality control and mapping

Software package FastQC (Andrews 2010) was used to examine the quality of the raw reads. Because of no observed adapter contamination, reads were mapped directly to the *D. pulex* reference genome PA42 3.0 (Ye *et al.* 2017) using STAR aligner (Dobin *et al.* 2013) with default parameters. SAMtools (Li *et al.* 2009) was used to remove reads that mapped to multiple locations in the genome, and the program featureCounts (Liao *et al.* 2014) was used to obtain the raw counts for expressed genes in each sample.

Differential gene expression analysis

We performed differential expression (DE) analysis using DESeq2 v.1.34.0 (Love *et al.* 2014) in R (R Core Team 2020). Differentially expressed genes (DEG) were determined for early subitaneous *vs* early resting egg development for each isolate individually to investigate for intraspecific differences and by pooling all the samples to establish commonalities. The Wald negative binomial test using the design formula ~ Stage for individual isolates and ~ Clone + Stage for pooled isolates were utilized in DEseq2. P-values were adjusted for multiple testing using the Benjamini-Hochberg method, and genes were considered significantly differentially expressed if they had a p-value < 0.05 and a fold-change > 1.5 and < -1.5 for upregulation and downregulation, respectively.

KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway and GO term enrichment analysis

To investigate the biological relevance of the differentially expressed genes, we performed a functional enrichment analysis using the R package topGO (Alexa and Rahnenfuhrer 2016). The

default algorithm, weight01, was used along with the Fisher exact test, and GO terms were considered significantly enriched if the weightFisher statistic was < 0.05. Our script is available at https://github.com/Marelize007/RNAseq_obligate_parthenogenesis.

We examined whether any KEGG pathways were enriched for differentially expressed genes. We queried 18,440 gene sequences from the *D. pulex* transcriptome (Ye *et al*. 2017) in the KEGG Automatic Annotation Server (KAAS) using the GHOSTX program (Moriya *et al.* 2007). A set of 10 135 genes were assigned a KO (KEGG ortholog) number, and from these, 6 282 were assigned to a KEGG pathway map. Hypergeometric tests with Holm-Bonferroni corrected p-values (p-value < 0.05) were used to identify enriched pathways (script available at: https://github.com/Marelize007/RNAseq_obligate_parthenogenesis). The up-and down-regulated genes were analyzed separately to increase the power to detect biologically relevant pathways (Hong *et al.* 2014). The Gene Annotation Easy Viewer (GAEV) (Huynh and Xu 2019) was used to visualize the functional pathways that each gene is mapped to.

Differential splicing analysis

Differentially spliced events (DS) were identified using the software rMATS v4.1.1 (Shen *et al.* 2014). Reads mapped to both exons and splice junctions were used to detect the following alternatively spliced events: SE (skipped exon), A5SS (alternative 5' splice site), A3SS (alternative 3' splice site), RI (retained intron), and MXE (mutually exclusive exons). An SE event occurs when an entire exon including its flanking introns is spliced out. A5SS and A3SS events may result in the inclusion or exclusion of different parts of exons, while entire introns are retained during RI events. During MXE events only one out of two exons are spliced into the resulting mRNA (Pohl *et al.* 2013, Wang *et al.* 2015). Genes were considered to be differentially spliced if at least four uniquely mapped reads supported the events, reads had a minimum anchor

length of 10 nt, the Benjamini-Hochberg adjusted p-value < 0.05, and the difference in exon inclusion level ($\Delta|\psi|$) > 5% (Shen *et al.* 2014, Suresh *et al.* 2020). Similar to the differential expression analysis, this analysis was completed for each isolate to observe for intraspecific differences, as well as by pooling the isolates together to obtain a comprehensive overview of the splicing differences between the early subitaneous and early resting egg production stages. Chi-squared tests were performed using R in Rstudio to test for significant ($p < 0.05$) over-and-under representation of splicing events within each isolate and the pooled sample.

Results

Data quality

Three biological replicates were collected for all isolates during early subitaneous egg and early resting egg production leading to a total of 18 RNA-seq samples. An average of 26.4 million raw reads were sequenced per sample. Our quality check using FastQC revealed no issues related to read quality or adapter contamination. On average 93% (SD = 2%) of the reads uniquely mapped to the *D. pulex* reference genome were retained for differential gene expression and differential splicing analysis **(Supplementary Table S2.2)**.

Differential expression analysis

Mapped read counts were normalized using the regularized log (rlog) transformation function in DEseq2, and a principal component analysis (PCA) was performed to visualize the grouping of samples. The first two principal components accounted for 47% and 17% of the variance. The first principal component (PC1) is likely due to differences between the early subitaneous and early resting egg production stages, whereas the second principal component (PC2) could be due to clonal variance **(Figure 2.3A)**. The same separation pattern for PC1 and PC2 was seen among

individual isolates **(Supplementary Figure S2.1)**. These results strongly suggest that our data captured the differences between the early subitaneous and early resting egg production stages and some inter-clonal differences.

To increase statistical power and obtain a comprehensive overview of the main differences between early subitaneous and early resting egg production contributing to PC1, we performed a pooled analysis comparing the developmental stages across all three isolates. A total of 3263 genes were significantly differentially expressed (p-value < 0.05), with 1771 genes upregulated and 1492 genes downregulated in early subitaneous egg production compared to early resting egg production **(Supplementary Table S2.3)**.

To reveal lineage-specific differences and commonalities, the transcriptomes between the two reproductive stages were compared within and between each isolate. For DB4-4, K09, and M348, a total of 1115, 2591, and 3942 genes were differentially expressed between the two stages, respectively **(Supplementary Table S2.3)**. Of these genes, 475 were shared among these isolates and the pooled analysis, while 99, 1483, 712, and 428 genes were uniquely differentially expressed in DB4-4, M348, K09, and the pooled analysis sample respectively **(Figure 2.3B)**.

KEGG pathway enrichment

A KEGG pathway enrichment analysis comparing the two reproductive stages within each isolate and the pooled analysis revealed that the onset of early subitaneous egg production is associated with the upregulation of meiosis and cell-cycle genes, as well as genes mapped to sugar and lipid metabolic pathways. Downregulated pathways were mainly associated with various metabolic, biosynthesis, and signaling pathways **(Figure 2.4A)**. Specifically, of the upregulated genes obtained from the pooled analysis, 576 were mapped to KEGG pathways and

enriched in 28 pathways for early subitaneous egg production (p.adjust < 0.05) **(Supplementary Table S2.4)**. The most notable of the pathways with upregulated genes included the Hedgehog signaling pathway, which plays a vital role in embryonic development by coordinating cell proliferation, coordination, and migration (Carballo *et al*. 2018). Additionally, the cell-cycle, oocyte meiosis, and progesterone-mediated oocyte maturation pathways were all significantly enriched for upregulated genes in early subitaneous egg production. Significantly enriched sugar and lipid metabolic pathways included starch and sucrose metabolism, galactose metabolism, and sphingolipid metabolism **(Figure 2.4A, C)**. KEGG pathways maps for the cell-cycle, oocyte meiosis, and progesterone-mediated oocyte maturation pathways were produced using the software GAEV (Huynh and Xu 2019) and can be viewed in **Supplementary Figures 2.2, 2.3, and 2.4**. This pattern of upregulation for genes mapped to the Hedgehog signaling pathway, cell-cycle, oocyte meiosis, progesterone-mediated oocyte maturation and various sugar and lipid metabolic pathways observed in the pooled analysis was also replicated within the individual isolates **(Figure 2.4B)**.

Of the downregulated genes obtained from the pooled analysis 449 were mapped to KEGG pathways and enriched in 12 pathways for early subitaneous egg production (p.adjust < 0.05) **(Supplementary Table S2.5)**. Pathways significantly downregulated include metabolic and biosynthesis pathways such as various N-glycan, O-glycan, and glycosphingolipid biosynthesis, as well as the GnRH, Hippo, and calcium signaling pathways **(Figure 2.4A, C)**. The GnRH signaling pathway is a regulator of the reproductive system (Kraus *et al.* 2001), while the Hippo signaling pathway controls organ size and development (Boopathy and Hong 2019).

Collectively, these results suggest that the upregulation of meiosis and cell-cycle genes and genes mapped to various sugar and lipid metabolic pathways, are associated with early

subitaneous egg production. Additionally, early subitaneous egg production is associated with a downregulation of genes mapped to various metabolic and biosynthesis pathways including N-glycan, O-glycan, and glycosphingolipid biosynthesis, as well as the GnRH, Hippo, and calcium signaling pathways.

GO term enrichment analysis

Our GO term enrichment analysis further corroborated the idea that early subitaneous egg production is associated with an upregulation of meiosis and cell-cycle genes and genes mapped to various sugar and lipid metabolic processes. Upregulated genes in early subitaneous egg production revealed enrichment for GO terms associated with carbohydrate metabolic process (weightFisher = 0.00014), oogenesis (weightFisher = 0.00158), lipid transport (weightFisher = 0.00215), trehalose metabolic process (weightFisher = 0.00572), regulation of mitotic cell-cycle phase transition (weightFisher = 0.0212) and cellular carbohydrate biosynthetic process (weightFisher = 0.04016), **Supplementary Table S2.6**). In line with the KEGG pathway analysis, our GO term enrichment analysis revealed a downregulation of various metabolic processes, including arginine metabolic process (weightFisher = 0.00646) and proline metabolic process (weightFisher = 0.00957), as well as cell differentiation (weightFisher = 0.02096), protein glycosylation (weightFisher = 9e-07) and other biosynthetic and signaling processes **(Supplementary Table S2.7)**. This pattern of regulation observed during early subitaneous egg production was additionally observed within each isolate **(Supplementary Table S2.8-S2.13)**.

Differentially Spliced genes

Differentially spliced transcripts were identified between the two stages for each isolate and the pooled splicing analysis to investigate how differential splicing distinguishes between early

subitaneous egg and early resting egg production. Across all 18 samples, 293 differentially spliced transcripts (FDR corrected p-value < 0.05) were identified between early subitaneous egg and early resting egg production. For the individual isolates, 397, 466, and 515 differentially spliced transcripts were identified for DB4-4, M348, and K09, respectively **(Supplementary Table S2.14)**. A GO term enrichment analysis of the differentially spliced transcripts between early subitaneous and early resting egg production shared by at least two samples (Total transcripts = 213, **Figure 2.5A**) showed enrichment for various metabolic and biosynthetic processes **(Figure 2.5B)**.

Further, analysis of the differentially spliced events across all isolates and the pooled splicing analysis revealed that skipped exon (SE) events were most abundant, totaling 729 (35%) events across all comparisons, followed by 417 (20%) RI events, 350 A5SS events (17%), 345 A3SS events (17%) and 235 MXE events (11%). **(Figure 2.5C, Supplementary Table S2.14)**. Lastly, we tested for over-and under-representation of splicing type among transcripts differentially spliced between early subitaneous and early resting egg production for each isolate and the pooled splicing analysis. The DB4-4 isolate showed an over-representation of A3SS and RI events, while K09 showed an over-representation of RI and SE events (chi-squared test $p < 0.05$). For the pooled splicing analysis, there was a significant under-representation of A3SS and SE events (chi-squared test $p < 0.05$).

Genes of interest

With early subitaneous egg production showing significant upregulation of meiosis and cell-cycle genes, we compiled a list of upregulated consensus genes mapped to the following significant pathways: Hedgehog signaling pathway, cell-cycle, oocyte meiosis, and progesterone-mediated oocyte maturation pathway. Consensus genes had to be shared by at least 3 of the

individual isolates or a combination of individual isolates and the pooled analysis **(Figure 2.6)**. These consensus genes were mainly composed of cell-cycle regulators such as AURKA, which regulates spindle formation and controls chromosome segregation (Blengini *et al.* 2021), various cyclins, and CDC20, CDC4, REC8L, and SMC1. CDC20, cell division cycle 20, which activates the anaphase-promotion complex/cyclosome (APC/C) to initiate sister chromatid separation (Jin *et al*. 2010), and REC8L, a gene coding the meiosis-specific component of the cohesion complex, which regulates the separation of sister chromatids and homologous chromosomes (Ward *et al.* 2016), showed upregulation across all three isolates and the pooled analysis. Additionally, CDC4, or cell division control protein 4, essential for the transition from G1 to S phase, the onset of anaphase, and the transition from G2 to M phase (Goh and Surana, 1999), also showed upregulation in all four samples with three samples having an average log2 fold-change > 2. Furthermore, no consensus genes belonging to the four mentioned cell-cycle or meiosis pathways were identified for the downregulated genes.

Discussion

In this study, we investigated the transcriptomic signatures between two modes of parthenogenesis in three obligate parthenogenetic *Daphnia pulex* isolates to gain insight into the genetic mechanisms and genes central to the emergence of obligate parthenogenesis through interspecific hybridization. First, we performed a pooled transcriptomic analysis by comparing early subitaneous egg and early resting egg production across all three OP *D. pulex* isolates. This allowed for a comprehensive overview of the main differences and to compensate for environment-genotype interactions. Second, we compared the transcriptomes of these two stages within each isolate to reveal lineage-specific differences. Through comparing the early stages of egg development, we aimed to identify differentiating transcriptomic signatures involved in the

initiation of these two parthenogenetic modes. Additionally, we investigated the role of differential splicing and identified consensus candidate genes mapped to the Hedgehog signaling pathway, cell-cycle, oocyte meiosis, and progesterone-mediated oocyte maturation pathway. Consensus candidate genes were obtained from these pathways since they showed significant enrichment in our KEGG pathway enrichment analysis, and fulfill vital functions in vertebrate embryonic development, the cell-cycle, meiosis, and meiosis reinitiation. The next goal is to utilize the identified candidate genes in gene knock-out studies to further clarify their role in these two parthenogenetic modes and how it affects meiosis and cell division.

In a previous publication, we investigated the transcriptomic signatures of early resting egg production between OP *D. pulex* hybrid isolates and their CP parentals (**Figure 2.2 A**, Xu *et al.* 2022). This study revealed that the development of resting embryos in OP isolates is associated with a downregulation of meiosis and cell-cycle genes and an upregulation of metabolism and biosynthesis genes (Xu *et al.* 2022). Additionally, this pattern was observed when contrasting early subitaneous egg production, observed in both OP and CP isolates, with early resting egg production in CP isolates (**Figure 2.2 B**, Huynh *et al.* 2021). Together, these results suggest that the initiation of parthenogenesis is associated with an upregulation of metabolic and biosynthesis genes and a downregulation of meiosis and cell-cycle genes.

Building upon this previous work, our KEGG pathway enrichment and GO term enrichment analysis of differentially expressed genes revealed a pattern of upregulation of meiosis and cell-cycle genes, as well as genes mapped to various carbohydrate and lipid metabolic pathways in early subitaneous egg production **(Figure 2.4 A, B and C)**. One factor that may have contributed to this upregulation of meiosis and cell-cycle genes could be due to a single female *Daphnia* producing multiple subitaneous eggs during one clutch. In contrast, a

maximum of two resting eggs are produced per ephippium. However, since whole-animal tissue was utilized in this study, no definitive conclusions can be drawn regarding the meiosis and cell-cycle genes. Of the lipid metabolic pathways, sphingolipid metabolism was significantly upregulated **(Figure 2.3 A, B, and C)**. Sphingolipids have been shown to play a vital role in growth factor signaling and morphogenesis in arthropods (Varki *et al.* 2015), and changes in sphingolipid abundance could impact cell proliferation, apoptosis, senescence, and differentiation (Hannun and Obeid 2002). Like *Daphnia*, the brine shrimp, *Artemia franciscana*, has two reproductive modes, resulting in diapausing eggs or nauplii larvae depending on light/dark cycles and temperature (Nambu *et al.* 2004). With sphingolipids involved in signaling and signal transduction pathways, they are partly responsible for which reproductive mode ensue in *A. franciscana* (Kojima *et al.* 2013) and may therefore be of major significance in embryogenesis of *Daphnia* as well. Further, glycosphingolipid biosynthetic enzymes were shown to interact with the NOTCH signaling receptors (Kraut 2011). In another arthropod system, adult worker honeybees, the NOTCH signaling pathway regulates reproduction, particularly during the early stages of oogenesis (Duncan *et al.* 2016).

Downregulated genes were enriched in various metabolic pathways, including arginine metabolism, biosynthesis pathways, including glycosphingolipid biosynthesis, as well as signaling pathways, including the calcium signaling pathway **(Figure 2.4A, C)**. Previous work on the inducibility of resting embryos have revealed that supplementing specific dietary amino acids such as arginine could suppress resting embryo production, therefore playing a vital role in the switch between reproductive modes (Fink *et al.* 2011, Koch *et al.* 2011). With our results showing a downregulation of genes mapped to the arginine metabolic pathway during early subitaneous egg production, indicates that arginine may additionally play a vital role in switching

between parthenogenetic modes within OP *D. pulex* isolates **(Supplementary Table S2.7)**. The calcium signaling pathway was also significantly downregulated in early subitaneous egg production. Calcium in *Daphnia* is vital for growth, molting, and ephippia formation (Giardini *et al.* 2015). In contrast to subitaneous egg production, during resting egg formation, *Daphnia* undergo two molting cycles and forms an ephippium which contains a thick layer of calcium phosphate enclosing the resting eggs (Gerrish and Cáceres 2003). This suggests that the calcium requirement for subitaneous egg production may be lower than during resting egg production. Similarly, various metabolic pathways were downregulated in early subitaneous embryo production, indicating a lower energy requirement. Studies conducted in *Daphnia magna* have shown that under stressful environmental conditions such as short day lengths, there was an upregulation of metabolic pathways (Gust *et al.* 2019).

These findings suggest that differentially expressed genes mapped to various metabolic, biosynthesis, and signaling pathways control the initiation of different parthenogenetic modes in *Daphnia*. Future studies could focus on pinpointing the master regulators of these pathways and additionally, expand on this study by utilizing single-cell RNA sequencing to compare early subitaneous oocyte production with early resting oocyte production to gain a clearer picture of the differentially expressed meiosis and cell-cycle genes between these two modes of parthenogenesis.

The differential splicing analysis between early subitaneous and early resting egg production revealed 60 differentially spliced transcripts shared by all four samples. Further annotation of the shared differentially spliced transcripts revealed UBE2I (UBC9), Ubiquitin Conjugating Enzyme E2 I, as a gene of interest. Work done in mouse oocytes revealed that meiotic maturation was disrupted and defects in spindle organization were identified with an

inhibition of UBE2I in fully grown oocytes, while over-expression caused a stimulation of transcription in meiotically incompetent oocytes (Ihara *et al*. 2008, Yuan *et al.* 2014). These results suggest that UBE2I, and thus sumoylation, may play an essential role in regulating gene expression during oocyte growth and maturation (Ihara *et al.* 2008). None of the consensus transcripts mapped to either the meiosis or cell-cycle pathways, indicating that since obligate parthenogenesis in *Daphnia* has a polyphyletic origin, different isolates may utilize different mechanisms of regulating gene expression in the transition to obligate parthenogenesis (Crease *et al.* 1989). Additionally, a GO term enrichment analysis of differentially spliced transcripts shared by at least two samples showed enrichment for various metabolic and biosynthetic processes, including the arginine metabolic process (**Figure 2.5C**). This finding, along with the results from our differential expression analysis, further support the idea that genes mapped to the arginine metabolic process may play a vital role in determining which mode of parthenogenesis to initiate.

Lastly, we compiled a list of upregulated consensus genes mapped to the cell-cycle, oocyte meiosis, progesterone-mediated oocyte maturation, and Hedgehog signaling pathway for further investigation **(Figure 2.6)**. A gene of interest from this list previously identified as playing an essential role in parthenogenesis is CDC20 (Xu *et al.* 2022). During oocyte meiosis, CDC20, a subunit responsible for activating the anaphase-promoting complex/cyclosome (APC/C), promotes progression from metaphase to anaphase via the destruction of cyclin B1 and securin (Jin *et al*. 2010). Studies in mice have shown that a reduction in CDC20 increased the average time from metaphase entry to the onset of anaphase (Jin *et al.* 2010), while a lack of CDC20 caused metaphase arrest (Li *et al*. 2007). This metaphase arrest due to a deficiency in CDC20 was also observed in bovine oocytes (Yang *et al.* 2014) and budding yeast (Lim *et al.*

1998). Further studies such as single-cell RNA sequencing on *Daphnia* oocytes will aid in deciphering the role of CDC20 and other meiosis genes in differentiating these two parthenogenetic modes.

In conclusion, we identified a pattern of upregulation of meiosis and cell-cycle genes and genes mapped to various carbohydrate and lipid metabolic pathways during early subitaneous egg production. Genes downregulated during early subitaneous egg production were mainly enriched in multiple metabolic, biosynthesis, and signaling pathways. Pathways of particular interest were the arginine metabolic process, sphingolipid metabolism, and the calcium signaling pathway. Additionally, we compiled a list of consensus meiosis and cell-cycle genes, including CDC20, that were upregulated in early subitaneous egg production for further investigation to determine their role in differentiating these two parthenogenetic modes. Lastly, our differential splicing analysis further supported the idea of arginine metabolism as playing an essential role in switching between early subitaneous egg and early resting egg production, however, more studies need to be conducted in OP isolates.

References

Avise, J. (2008). *Clonality: The Genetics, Ecology, and Evolution of Sexual Abstinence in Vertebrate Animals*. Oxford University Press, USA.

Avise JC (2015) Evolutionary perspectives on clonal reproduction in vertebrate animals. *Proceedings of the National Academy of Sciences of the United States of America, 112*(29), 8867–8873.

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/

Bell G (1982) *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. Berkeley, CA: University of California Press.

Blengini CS, Ibrahimian P, Vaskovicova M, Drutovic D, Solc P, *et al.* (2021) Aurora kinase A is essential for meiosis in mouse oocytes. PLOS Genetics 17: e1009327.

Boopathy GTK, Hong W (2019) Role of Hippo Pathway-YAP/TAZ Signaling in Angiogenesis. *Frontiers in Cell and Developmental Biology, 7*.

Brandlova J, Brandl Z, Fernando CH (1972) The Cladocera of Ontario with remarks on some species and distribution. *Canadian Journal of Zoology, 50*(11), 1373–1403.

Carballo GB, Honorato JR, de Lopes GPF, de Sampaio e Spohr TCL (2018) A highlight on Sonic hedgehog pathway. *Cell Communication and Signaling, 16*(1), 11.

Colbourne JK, Hebert PDN (1996) The systematics of North American Daphnia (Crustacea: Anomopoda): a molecular phylogenetic approach. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1337), 349–360.

Crease TJ, Lee SK, Yu SL, Spitze K, Lehman N, Lynch M (1997) Allozyme and mtDNA variation in populations of the *Daphnia pulex* complex from both sides of the Rocky Mountains. *Heredity*, *79*(3), 242–251.

Crease TJ, Stanton DJ, Hebert PDN (1989) Polyphyletic origins of asexuality in *Daphnia*. II. Mitochondrial-DNA variation. *Evolution; International Journal of Organic Evolution*, *43*(5), 1016–1026.

Cristescu ME, Constantin A, Bock DG, Cáceres CE, Crease TJ (2012) Speciation with gene flow and the genetics of habitat transitions. *Molecular Ecology*, *21*(6), 1411–1422.

Cristescu ME, Demiri B, Altshuler I, Crease TJ (2014) Gene Expression Variation in Duplicate Lactate dehydrogenase Genes: Do Ecological Species Show Distinct Responses? *PLoS ONE*, *9*(7), e103964.

Dawley RM, Bogart JP (1989) *Evolution and Ecology of Unisexual Vertebrates*. University of the State of New York, State Education Department, New York State Museum.

Decaestecker E, De Meester L, Mergeay, J (2009) Cyclical Parthenogenesis in *Daphnia*: Sexual Versus Asexual Reproduction. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost Sex: The Evolutionary Biology of Parthenogenesis* (pp. 295–316). Springer Netherlands.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.

Duncan EJ, Hyink O, Dearden PK (2016) Notch signalling mediates reproductive constraint in the adult worker honeybee. *Nature Communications*, *7*, 12427.

Ferree PM, McDonald K, Fasulo B, Sullivan W (2006) The origin of centrosomes in parthenogenetic hymenopteran insects. *Current Biology: CB*, *16*(8), 801–807.

Fink P, Pflitsch C, Marin K (2011) Dietary Essential Amino Acids Affect the Reproduction of the Keystone Herbivore *Daphnia pulex*. *PLOS ONE*, *6*(12), e28498.

Gerrish GA, Cáceres CE (2003) Genetic versus environmental influence on pigment variation in the ephippia of *Daphnia pulicaria*. *Freshwater Biology*, *48*(11), 1971–1982.

Giardini JL, Yan ND, Heyland A (2015) Consequences of calcium decline on the embryogenesis and life history of *Daphnia magna*. *The Journal of Experimental Biology*, *218*(Pt 13), 2005–2014.

Gorr TA, Rider CV, Wang HY, Olmstead AW, LeBlanc GA (2006) A candidate juvenoid hormone receptor cis-element in the *Daphnia magna* hb2 hemoglobin gene promoter. *Molecular and Cellular Endocrinology*, *247*(1–2), 91–102.

Gust KA, Kennedy AJ, Laird JG, Wilbanks MS, Barker ND, Guan X, Melby NL, Burgoon LD, Kjelland ME, Swannack TM (2019) Different as night and day: Behavioural and life history responses to varied photoperiods in Daphnia magna. *Molecular Ecology*, *28*(19), 4422–4438.

Hannun YA, Obeid LM (2002) The Ceramide-centric universe of lipid-mediated cell regulation: Stress encounters of the lipid kind. *The Journal of Biological Chemistry*, *277*(29), 25847–25850.

Harris KDM, Bartlett NJ, Lloyd VK (2012) *Daphnia* as an Emerging Epigenetic Model Organism. *Genetics Research International*, *2012*, 1–8.

Hebert PDN, Crease T (1983) Clonal diversity in populations of *Daphnia pulex* reproducing by obligate parthenogenesis. *Heredity*, *51*(1), 353–369.

Hebert PDN, Finston TL (2001) Macrogeographic patterns of breeding system diversity in the *Daphnia pulex* group from the United States and Mexico. *Heredity*, *87*(2), 153–161.

Heier CR, Dudycha JL (2009) Ecological speciation in a cyclic parthenogen: Sexual capability of experimental hybrids between *Daphnia pulex* and *Daphnia pulicaria*. *Limnology and Oceanography*, *54*(2), 492–502.

Hiruta C, Nishida C, Tochinai S (2010) Abortive meiosis in the oogenesis of parthenogenetic *Daphnia pulex*. *Chromosome Research*, *18*(7), 833–840.

Hong G, Zhang W, Li H, Shen X, Guo Z (2014) Separate enrichment analysis of pathways for up- and downregulated genes. *Journal of the Royal Society Interface*, *11*(92), 20130950.

Huynh TV, Hall AS, Xu S (2021) *The transcriptomic signature of cyclical parthenogenesis* (p. 2021.09.27.461985). bioRxiv.

Huynh T, Xu S (2019) Gene Annotation Easy Viewer (GAEV): Integrating KEGG's Gene Function Annotations and Associated Molecular Pathways. *F1000Research*, *7*, 416.

Ihara M, Stein P, Schultz RM (2008) UBE2I (UBC9), a SUMO-conjugating enzyme, localizes to nuclear speckles and stimulates transcription in mouse oocytes. *Biology of Reproduction*, *79*(5), 906–913.

Innes DJ, Hebert PDN (1988). The Origin and Genetic Basis of Obligate Parthenogenesis in *Daphnia Pulex*. *Evolution*, *42*(5), 1024–1035.

Jaquiéry J, Stoeckel S, Larose C, Nouhaud P, Rispe C, Mieuzet L *et al.* (2014). Genetic Control of Contagious Asexuality in the Pea Aphid. *PLOS Genetics*, *10*(12), e1004838.

Jin F, Hamada M, Malureanu L, Jeganathan KB, Zhou W, Morbeck DE, van Deursen JM (2010) Cdc20 Is Critical for Meiosis I and Fertility of Female Mice. *PLOS Genetics*, *6*(9), e1001147.

Johnson SG, Bragg E (1999) Age and Polyphyletic Origins of Hybrid and Spontaneous Parthenogenetic Campeloma (Gastropoda: Viviparidae) from the Southeastern United States. *Evolution*, *53*(6), 1769–1781.

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.

King KC, Hurst GD (2010) Losing the desire: Selection can promote obligate asexuality. *BMC Biology*, *8*(1), 101.

Koch U, Martin-Creuzburg D, Grossart HP, Straile D (2011) Single dietary amino acids control resting egg production and affect population growth of a key freshwater herbivore. *Oecologia*, *167*(4), 981–989.

Kojima H, Tohsato Y, Kabayama K, Itonori S, Ito M (2013) Biochemical studies on sphingolipids of *Artemia franciscana*: Complex neutral glycosphingolipids. *Glycoconjugate Journal*, *30*(3), 257–268.

Kraus S, Naor Z, Seger R (2001) Intracellular signaling pathways mediated by the gonadotropin-releasing hormone (GnRH) receptor. *Archives of Medical Research*, *32*(6), 499–509.

Kraut R (2011) Roles of sphingolipids in *Drosophila* development and disease. *Journal of Neurochemistry*, *116*(5), 764–778.

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

Li M, York JP, Zhang P (2007) Loss of Cdc20 Causes a Securin-Dependent Metaphase Arrest in Two-Cell Mouse Embryos. *Molecular and Cellular Biology*, *27*(9), 3481–3488.

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.

Liegeois M, Sartori M, Schwander T (2021) Extremely Widespread Parthenogenesis and a Trade-Off Between Alternative Forms of Reproduction in Mayflies (Ephemeroptera). *Journal of Heredity*, *112*(1), 45–57.

Lim HH, Goh PY, Surana U (1998) Cdc20 is essential for the cyclosome-mediated proteolysis of both Pds1 and Clb2 during M phase in budding yeast. *Current Biology*, *8*(4), 231–237.

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.

Lynch M (1984) Destabilizing Hybridization, General-Purpose Genotypes and Geographic Parthenogenesis. *The Quarterly Review of Biology*, *59*(3), 257–290.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, *35*(Web Server issue), W182–W185.

Nambu Z, Tanaka S, Nambu F (2004) Influence of photoperiod and temperature on reproductive mode in the Brine shrimp, *Artemia franciscana*. *Journal of Experimental Zoology. Part A, Comparative Experimental Biology*, *301*(6), 542–546.

Neaves WB, Baumann P (2011) Unisexual reproduction among vertebrates. *Trends in Genetics*, *27*(3), 81–88.

Neiman M, Sharbel TF, Schwander T (2014) Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. *Journal of Evolutionary Biology*, *27*(7), 1346–1359.

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.

Omilian AR, Lynch M (2009) Patterns of Intraspecific DNA Variation in the Daphnia Nuclear Genome. *Genetics*, *182*(1), 325–336.

Pfrender ME, Spitze K, Lehman N (2000) Multi-locus genetic evidence for rapid ecologically based speciation in *Daphnia*. *Molecular Ecology*, *9*(11), 1717–1735.

Pohl M, Bortfeldt RH, Grützmann K, Schuster S (2013) Alternative splicing of mutually exclusive exons—A review. *Biosystems*, *114*(1), 31–38.

Riparbelli MG, Tagu D, Bonhomme J, Callaini G (2005) Aster self-organization at meiosis: A conserved mechanism in insect parthenogenesis? *Developmental Biology*, *278*(1), 220–230.

RStudio Team (2020) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Schwander T, Henry L, Crespi BJ (2011) Molecular Evidence for Ancient Asexuality in Timema Stick Insects. *Current Biology*, *21*(13), 1129–1134.

Serra M, Snell TW (2009) Sex loss in monogonont rotifers. In: I Schön, K Martens, P van Dijk, editors. Lost Sex. Dordrecht: Springer. p. 281-294.

Shaw JR, Colbourne JK, Davey JC, Glaholt SP, Hampton TH, Chen CY, Folt CL, Hamilton JW (2007) Gene response profiles for *Daphnia pulex* exposed to the environmental stressor cadmium reveals novel crustacean metallothioneins. *BMC Genomics*, *8*, 477.

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.

Simon JC, Delmotte F, Rispe C, Crease T (2003) Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society*, *79*(1), 151–163.

Stenberg P, Lundmark M (2004) Distribution, mechanisms and evolutionary significance of clonality and polyploidy in weevils. *Agricultural and Forest Entomology*, *6*(4), 259–266.

Stenberg P, Saura A (2009) Cytology of Asexual Animals. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost Sex: The Evolutionary Biology of Parthenogenesis* (pp. 63–74). Springer Netherlands.

Suomalainen E, Saura A, Lokki J (1987) *Cytology and evolution in parthenogenesis*. CRC Press.

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.

Tatarazako N, Oda S, Watanabe H, Morita M, Iguchi T (2003) Juvenile hormone agonists affect the occurrence of male *Daphnia*. *Chemosphere*, *53*(8), 827–833.

Alexa A, Rahnenfuhrer J (2021) topGO: Enrichment Analysis for Gene Ontology. R package version 2.46.0.

van der Kooi CJ, Matthey-Doret C, Schwander T (2017) Evolution and comparative ecology of parthenogenesis in haplodiploid arthropods. *Evolution Letters*, *1*(6), 304–316.

Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH (2015). *Essentials of Glycobiology* (3rd ed.). Cold Spring Harbor Laboratory Press.

Vrijenhoek R (1998) Animal Clones and Diversity. *Bioscience*, *48*, 617–628.

Wang Y, Liu J, Huang B, Xu YM, Li J, Huang LF, Lin J, Zhang J, Min QH, Yang WM, Wang XZ (2015) Mechanism of alternative splicing and its regulation (Review). *Biomedical Reports*, *3*(2), 152–158.

White MJ, Contreras N, Chency J, Webb GC (1977) Cytogenetics of the parthenogenetic grasshopper Warramaba (formerly Moraba) virgo and its bisexual relatives. II. Hybridization studies. *Chromosoma*, *61*(2), 127–148.

Wright S (1978) Modes of Speciation. Michael J. D. White W. H. Freeman and Co., San Francisco. 1978. *Paleobiology*, *4*(3), 373–379.

Xu S, Huynh TV, Snyman M (2022) The transcriptomic signature of obligate parthenogenesis. *Heredity*, *128*(2), 132–138.

Xu S, Innes DJ, Lynch M, Cristescu ME (2013) The role of hybridization in the origin and spread of asexuality in *Daphnia*. *Molecular Ecology*, *22*(17), 4549–4561.

Xu S, Spitze K, Ackerman MS, Ye Z, Bright L, Keith N, Jackson CE, Shaw JR, Lynch M (2015) Hybridization and the Origin of Contagious Asexuality in *Daphnia pulex*. *Molecular Biology and Evolution*, *32*(12), 3215–3225.

Yang WL, Li J, An P, Lei AM (2014) CDC20 downregulation impairs spindle morphology and causes reduced first polar body emission during bovine oocyte maturation. *Theriogenology*, *81*(4), 535–544.

Ye Z, Xu S, Spitze K, Asselman J, Jiang X, Ackerman MS, Lopez J, Harker B, Raborn RT, Thomas WK, Ramsdell J, Pfrender ME, Lynch M (2017) A New Reference Genome Assembly for the Microcrustacean *Daphnia pulex*. *G3: Genes, Genomes, Genetics*, *7*(5), 1405–1416.

Yuan YF, Zhai R, Liu XM, Khan HA, Zhen YH, Huo LJ (2014) SUMO-1 plays crucial roles for spindle organization, chromosome congression, and chromosome segregation during mouse oocyte meiotic maturation. *Molecular Reproduction and Development*, *81*(8), 712–724.

Zaffagnini F, Sabelli B (1972) Karyologic observations on the maturation of the summer and winter eggs of *Daphnia pulex* and *Daphnia middendorffiana*. *Chromosoma*, *36*(2), 193–203.

Ojima Y (1958). A cytological study on the development and maturation of the parthenogenetic and sexual eggs of *Daphnia pulex*. Kwansei Gakuin Univ Ann Studies 6: 123 - 171.

Acknowledgements

This work is supported by NIH grant R35GM133730. We would like to thank the Xu lab members for their helpful discussions.

Data

The raw reads for this study are deposited at NCBI SRA PRJNA847604.

Author Contributions

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

Tables and Figures

Figure 2.1. (A) Cyclically parthenogenetic and (B) obligately parthenogenetic life cycles in *Daphnia*. (C) Early subitaneous egg and (D) early resting egg production as determined by the color and size of the ovaries (red circles).

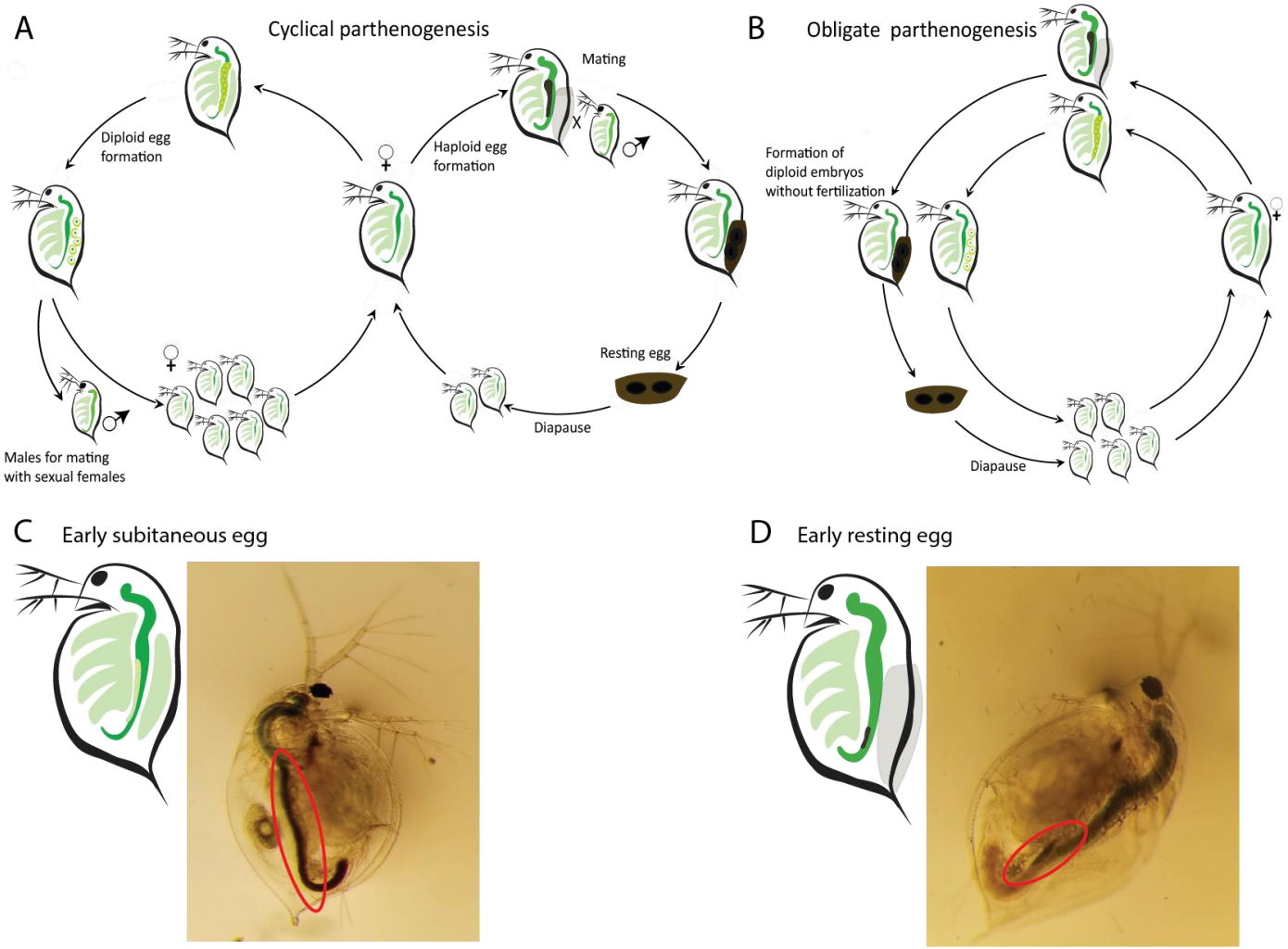


Figure 2.2. (A) Regulation of genes during early resting egg production between OP *D. pulex* and CP *D. pulex* and *D. pulicaria* isolates. (B) Regulation of genes during early resting egg production and early subitaneous egg production within CP isolates, (C) and within OP *D. pulex* isolates.

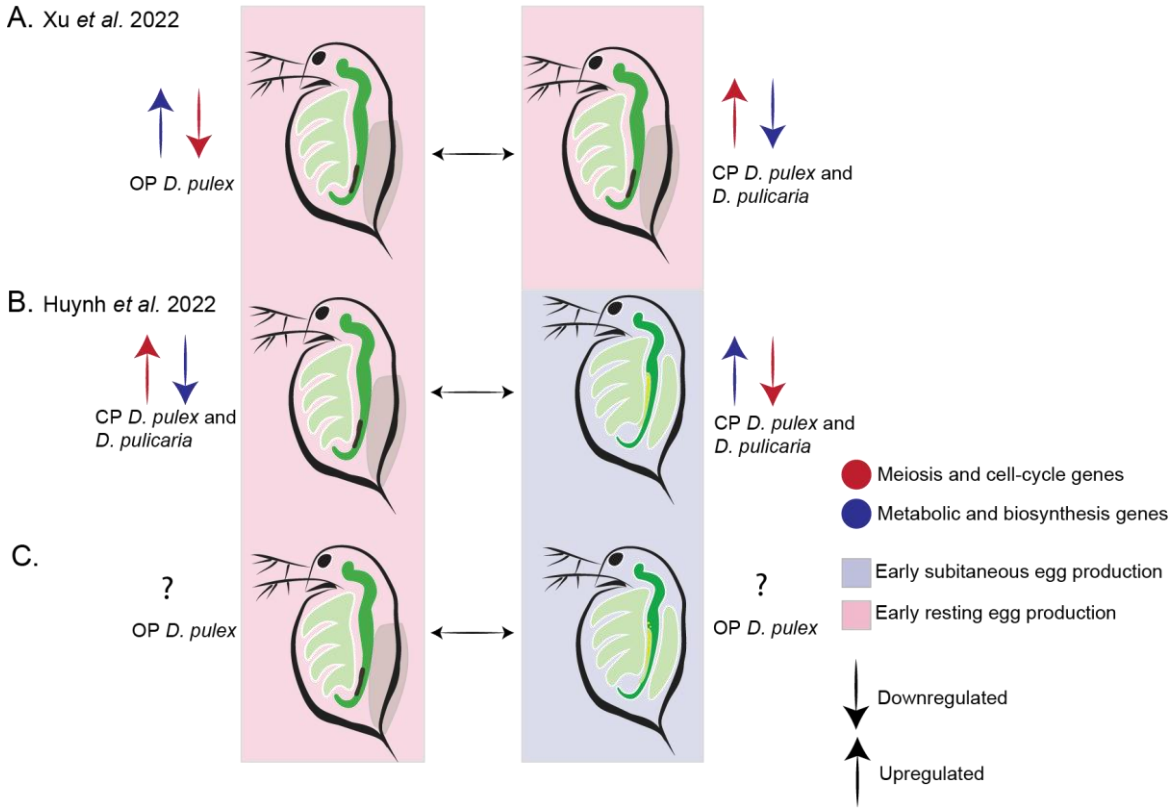


Figure 2.3. (A) PCA plot based on all samples. (B) Venn diagram showing the number of differentially expressed genes between early subitaneous and early resting egg production. OP represents the set of differentially expressed genes from the pooled analysis.

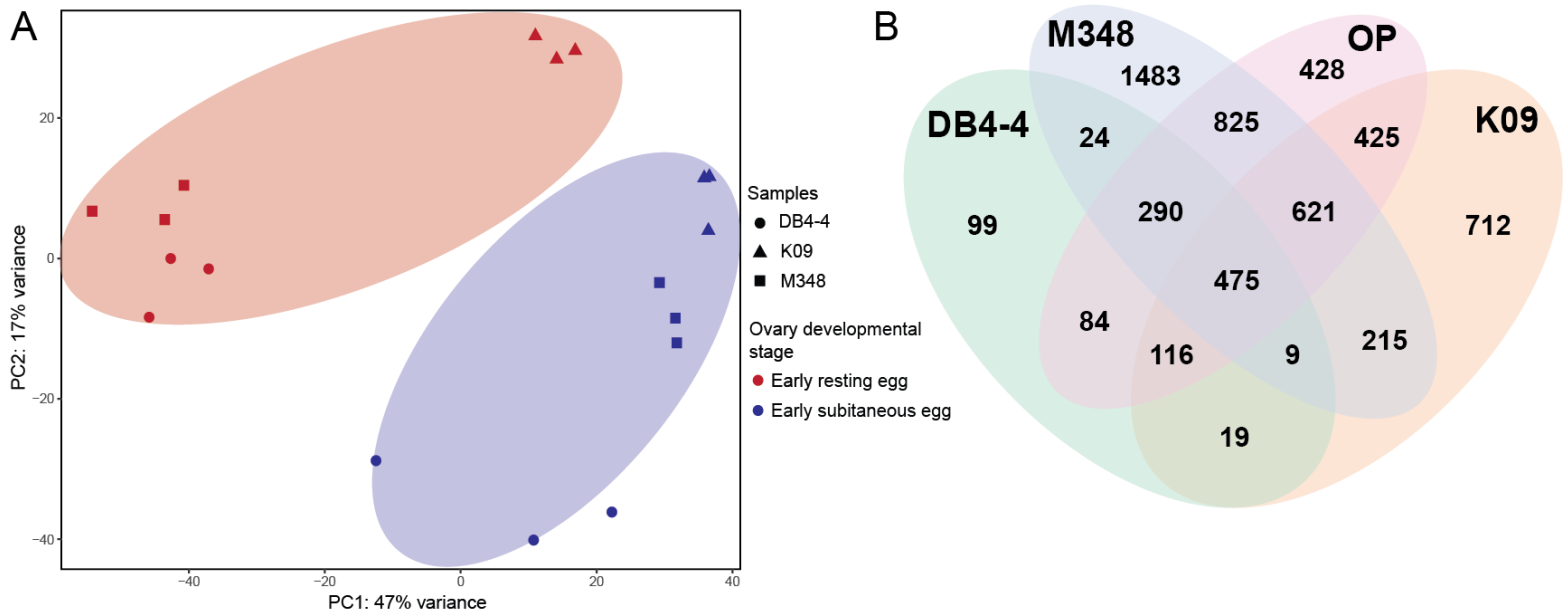


Figure 2.4. (A) Distribution of up-and down regulated DE genes during early subitaneous egg production for the pooled sample across KEGG pathways. (B) Heatmap illustrating a subset of up-and down regulated KEGG pathways across all replicates. (C) Distribution of up-and downregulated genes according to log2 fold-change and KEGG pathway.

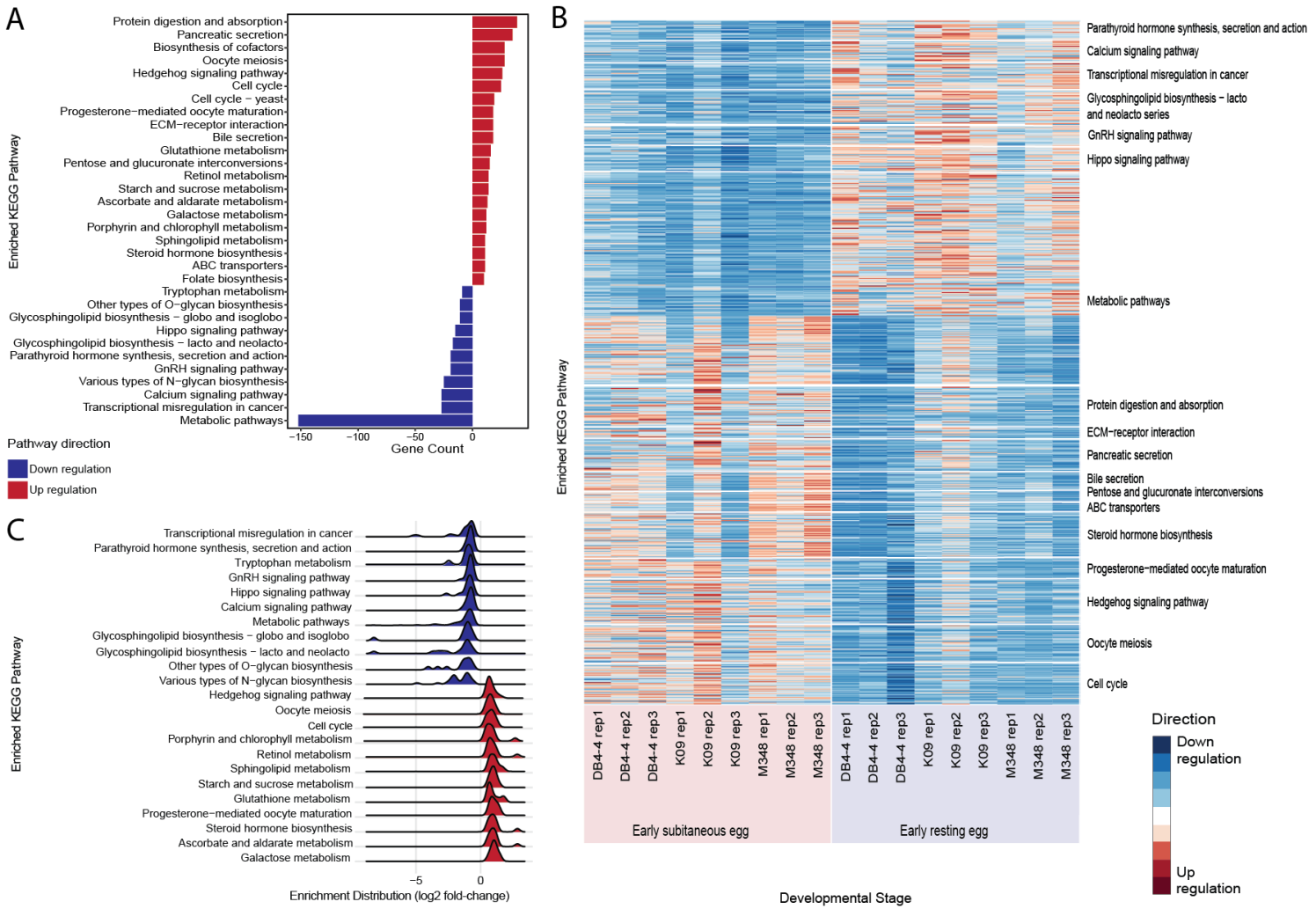


Figure 2.5. (A) Venn diagram showing the number of differentially spliced genes between early subitaneous egg and early resting egg production. OP represents the set of differentially spliced genes from the pooled analysis. (B) Significantly enriched GO terms of differentially spliced genes. (C) Composition of differentially spliced genes.

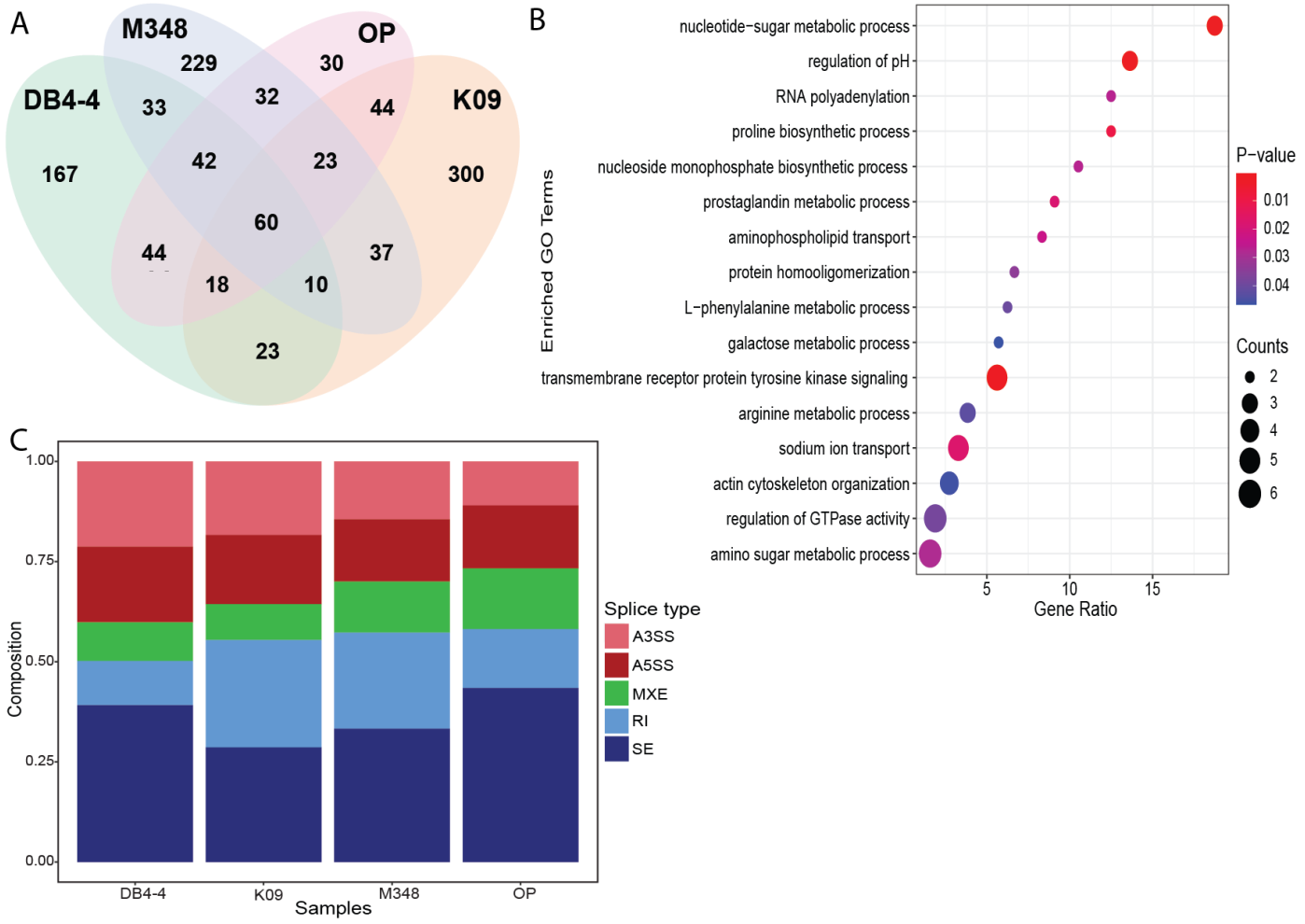
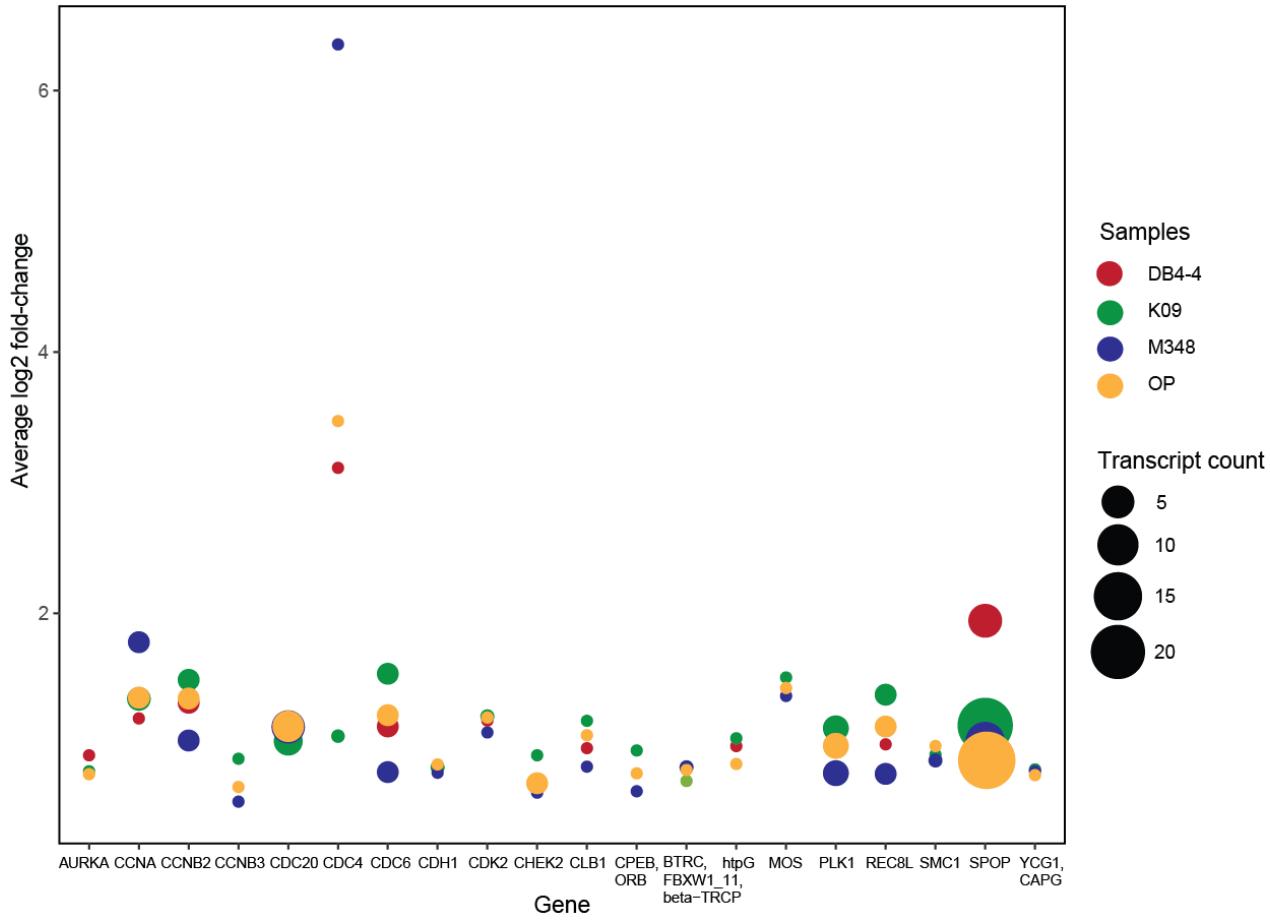


Figure 2.6. Average log2 fold-change for genes upregulated during early subitaneous embryo development which mapped to the cell-cycle, Oocyte meiosis, Progesterone-mediated oocyte maturation, and Hedgehog signaling pathway.



CHAPTER 3: The genome-wide rate and spectrum of EMS-induced heritable mutations in the microcrustacean *Daphnia*: on the prospect of forward genetics

Authors: Marelize Snyman, Trung Huynh, Matthew T. Smith, Sen Xu*

Department of Biology, University of Texas at Arlington, Arlington, Texas USA 76019.

*Correspondence author. 501 S Nedderman Dr., Arlington, Texas, USA 76019. Email: sen.xu@uta.edu

This article is published in Nature: Heredity 127:535-545

Abstract

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 10mM EMS treatment induces base substitutions at an average rate of 1.17×10^{-6}/site/generation across the three species, whereas a significantly higher average mutation rate of 1.75×10^{-6} occurs at 25mM. 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_1s 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_2s 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.

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^6 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; Martín *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.

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_0 individuals) can be exposed to EMS, during which mutations can be introduced into the genomes of oocytes (**Figure 3.1**). Females which have been exposed to EMS can then asexually produce mutant female offspring (F_1s), which would carry all EMS-induced germline mutations in the heterozygous state. Each of the F_1

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_2s) that carry 25% of the EMS-induced mutations in the homozygous state (**Figure 3.1**). Screening these F_2s 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 10mM vs 25mM 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_1s and F_2s) for genetic screening in *Daphnia* to guide such 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 S3.1**). They have been kept in the lab as clonally reproducing lines in artificial lake water (Kilham *et al.* 1998) under a 16:8 hour (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., 10mM, 25mM, 50mM and 100mM) 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 50mM EMS for 4 hours (Brenner 1974) to achieve a mutation rate of 2.5×10^{-3} per gene per generation (Gengyo-Ando and Mitani 2000), while *D. melanogaster* is fed 25mM EMS (Lewis and Bacher 1968) to achieve a mutation rate of 1×10^{-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 1mL of EMS solution at the concentrations of 10mM, 25mM, 50mM, and 100mM, respectively. The exposure lasted four hours. The treated animals were then transferred to artificial lake water, and the survival rate was recorded after 24 hours.

Although no animals survived the 4-hour exposure to 50mM and 100mM EMS, 100% and 60% of females survived the 10mM and 25mM treatments, respectively (see **Results** and **Supplementary Table S3.2**). 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 10mM and 25mM EMS treatments, sexually mature *Daphnia* females from each isolate were exposed to these two concentrations for 4 hours, 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_1s) from the F_0s were then collected and individually isolated because F_1s 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_1s clonally, with each F_1 propagating into a mass asexual culture (**Figure 3.2A**). These asexual progenies were whole-genome 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 10mM and 25mM EMS treatment (**Figure 3.2B**). 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. Additionally, 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) >= 20, sequencing depth (DP) >= 10, and a distance >= 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) 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 >=10 coverage and 50bp 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 genic 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 10mM and 25mM 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 25mM) 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_1s (Shaham 2007). The $(1 – r)^n$ term denotes the probability for none of the sampled F_1s 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 10mM, 25mM, 50mM and 100mM for four hours. All *Daphnia* exposed to 50mM and 100mM EMS died during or after treatment (within 24 hours). At lower EMS concentrations, 100% of the animals treated with 10mM EMS survived, whereas the average survival rate was 60.0% (SD=8.8%) for animals treated with 25mM EMS (**Supplementary Table S3.2**).

Whole-genome sequencing data

A total of 43 *Daphnia* mutant lines derived from 10mM or 25mM EMS treatment were whole-genome sequenced with 150bp Illumina paired-end reads (**Supplementary Table S3.3 and Table S3.4**). 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 << 0.05 (i.e., << 1 false positive out of 20 mutations, **Supplementary Table S3.11**).

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 10mM EMS and 14 mutant lines treated with 25mM EMS to detect germline mutations (**Supplementary Table S3.3**). 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 10mM treatment ranged from 9.40×10^{-7} to 1.32×10^{-6} (mean = 1.17×10^{-6}, SEM=1.84×10^{-7}, see **Table 3.1 and Figure 3.3**), a few hundred times higher than the spontaneous mutation rate which ranges from 2.30×10^{-9} to 7.17×10^{-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 10mM (ANOVA p value > 0.1). Across the threes *Daphnia* species, the average per gene mutation rate and per gene non-synonymous rate of the 10mM treatment lines are 2.65×10^{-3} (SEM=3.32×10^{-4}) and 1.19×10^{-3} (SEM=1.71×10^{-4}), respectively (**Table 3.1**).

Notably, mutant lines from the 25mM EMS treatment showed on average a higher base substitution mutation rate than those from the 10mM treatment, yielding strong support to our

first hypothesis. The base substitution rate for the 25mM ranged from 1.58×10^{-6} to 1.98×10^{-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 (1.75×10^{-6} per site per generation, SEM=6.82×10^{-7}, **Table 3.1, Figure 3.3**) across the three species was significantly higher (a 0.5-fold increase) than that at 10mM (mean = 1.17×10^{-6}, SEM=1.84×10^{-7}, t-test p = 0.0052). The average per gene mutation rate (4.09×10^{-3} per gene per generation, SEM=4.31×10^{-4}, **Table 3.1**) and average nonsynonymous mutation rate (1.91×10^{-3} per gene per generation, SEM=1.81×10^{-4}, **Table 3.1**) across the three species at 25mM also showed an increase of 0.5 and 0.6-fold compared to those at 10mM, respectively.

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_0s. To test this, we sequenced a total of 17 first-brood (BR1), second-brood (BR2), and third-brood (BR3) mutant lines treated with 10mM and 25mM EMS in three *Daphnia* isolates, Tex21 (CP *D. pulex*), AroMoose (CP *D. pulicaria*), and DB4-4 (OP *D. pulex*) (**Supplementary Table S3.4**).

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 10mM, the average base-substitution mutation rates for BR1, BR2, and BR3 progenies at 10mM are 6.58×10^{-7} (SEM=2.07×10^{-8}), 5.70×10^{-7} (SEM=9.20×10^{-8}) and 9.90×10^{-7} (SEM=2.90×10-7),

respectively (**Figure 3.3**). Similarly, the average base-substitution mutation rates for BR1, BR2, and BR3 lines at 25mM are 1.86×10^{-6} (SEM=6.95×10^{-7}), 3.75×10^{-6} (SEM=2.02×10^{-6}) and 4.49×10^{-6} (SEM=1.77×10^{-6}), respectively, significantly higher than those at 10mM (ANOVA p = 0.039) (**Figure 3.3**).

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 25mM when compared to lines treated with 10mM EMS (**Supplementary Table S3.4**). 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 10mM treatment lines are G:C to A:T transitions, resulting in an elevated transition-transversion ratio greater than 4.1 for all lines (**Figure 3.4A**, **Supplementary Table S3.3 and S3.5**). Mutant lines of 25mM 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 (**Figure 3.4A, Supplementary Table S3.3**). 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 10mM and 25mM EMS treatments, further substantiating the idea that EMS successfully induced heritable mutations in consecutive broods (**Figure 3.4B**, **Supplementary Table S3.4 and S3.6**).

Concordant with the notion that EMS induces mutation randomly across the genome, the distribution of EMS-induced mutations for the 10mM and 25mM 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 10mM, 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 25mM, 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 (**Figure 3.5A and Supplementary Table S3.7**).

Furthermore, regarding exonic mutations, for the 10mM treatment on average 23 (65.5%) were missense, 1 (3.2%) nonsense (stop-gained) and 11 (31.3%) silent. The 25mM treatment once again produced very similar results with 44 (67.8%) missense, 3 (5.2%) nonsense (stop-gained) and 18 (27%) silent (**Figure 3.5C and Supplementary Table S3.7**). 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 (**Figure 3.5B, D and Supplementary Table S3.8**). 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 (Grauer and Li 2000) .

Motif analysis of mutated sites

For the NAN and NTN trinucleotide motifs **(Figure 3.6)**, all of the trinucleotides were significantly under-represented (chi-squared test $p < 0.05$). Among the NCN trinucleotides **(Figure 3.6)**, 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 of the GGT, AGG, CGG, CGA, GGA, GGC, CGC, and GGG (chi-squared test $p < 0.05$, **Figure 3.6** and **Supplementary Table S3.9**).

Number of F_1s for reaching mutation saturation

Based on the average base-substitution rate per gene at 25mM EMS treatment across three *Daphnia* species (~4×10^{-3} per gene per generation), approximately 750 F_1s are needed to find at least one F_1 animal heterozygous for a mutation in a gene of interest with 95% probability. With ~750 F_1s, a total of 54, 000 genes would have been mutated, translating to roughly 3 mutations per gene given the ~18000 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 EMS concentrations. We demonstrate that exposure to 10mM or 25mM EMS solution for 4 hours can readily induce mutations in the oocytes that *Daphnia* females carry at a rate that is hundreds of times higher than the spontaneous mutation rate (Keith *et al.* 2016; Flynn *et al.* 2017; Bull *et al.* 2019), establishing a useful protocol that can be used for obtaining mutant lines for screening experiments. Since our ultimate goal is to establish a forward genetic method for *Daphnia*, we will compare our results to those of three model organisms (i.e., *C. elegans, D. melanogaster*, and *A. thaliana*) that have well demonstrated EMS mutagenesis protocols (Page and Grossniklaus 2002; St Johnston 2002; Jorgensen and Mango 2002).

As we hypothesized, the concentrations of EMS are indeed a major determinant of the induced mutation rate. The base substitution mutation rate is significantly higher for the mutant lines from 25mM treatment than from the 10mM treatment lines, showing a 0.5-fold increase, with both rates hundreds of times higher than the spontaneous base substitution rate (Keith *et al.* 2016; Flynn *et al.* 2017; Bull *et al.* 2019). Nonetheless, all lines from both treatments show the mutation spectra characteristic of EMS mutagenesis, with a strong bias towards G:C to A:T transitions, averaging 87% and 86% for 10mM and 25mM mutant lines, respectively. This is a substantial increase from the previously reported ~66% G/C to A/T ratio in *Daphnia* spontaneous mutation accumulation lines (Keith *et al.* 2016).

With an average of 78 (SD=13) genes affected by mutations per line treated with 25mM EMS, the EMS induced per gene per generation mutation rate was 4.1×10^{-3}, 3.2×10^{-3}, and 4.4×10^{-3} for OP *D. pulex*, CP *D. pulex* and *D. pulicaria*, respectively (**Table 3.1**). In comparison, this rate is higher than those in *C. elegans* (Gengyo-Ando and Mitani 2000), *D. melanogaster* (1.0×10^{-3}, Spradling 1997) and *A. thaliana* (Ossowski *et al.* 2010).

Our observed mutation spectrum of EMS-induced base substitutions is also consistent with earlier observations in other model organisms. In comparison, the proportion of G:C to A:T transitions in *Daphnia* is higher than that in *C. elegans* (66%, Sarin *et al.* 2010), similar to that in *D. melanogaster* (70-84%, Winkler *et al.* 2005; Cooper *et al.* 2008), and much lower than that in *A. thaliana* (>99%, Greene *et al.* 2003). Evidently, the spectrum of EMS-induced mutations is dominated by G:C to A:T transitions across eukaryotic species, although the ratio greatly varies. Presumably, the concentration of EMS and the means of exposure to EMS could contribute to this difference across species because each species has its own specific experimental procedures

(e.g., EMS exposure through feeding, soaking seeds). Nonetheless, whether species-specific DNA repair mechanisms are a contributing factor remains to be clarified by future investigation.

We were also interested in whether induced germline mutations occur in an independent manner in consecutive broods produced by the same female *Daphnia* exposed to EMS, i.e., whether these progenies are all genetically distinct due to the induced mutations. Our results show that EMS mutagenesis can successfully induce germline mutations in the first three consecutive broods, while the mutation spectrum remains highly similar between broods and between different EMS concentrations. Because all the identified mutations are unique across mutant lines, this supports that the progenitor cells of oocytes were independently affected by EMS. We suggest that the progenies of at least the first three consecutive broods can be used to establish F_1 mutant lines in screening experiments.

One important reason that EMS mutagenesis is used for screening experiments is because EMS is expected to induce mutations at random locations throughout the genome. Our results showed that the distribution of induced mutations in mutant lines from 10mM and 25mM treatments are highly similar, and that no genomic regions (e.g., exons, introns,) are significantly enriched with mutations (**Figure 3.5A**).

The trinucleotide motif analysis showed a significant under-representation of NAN and NTN trinucleotides, consistent with the strong preference EMS has to mutate G and C nucleotides. The analysis further shows that most trinucleotides enriched with EMS-induced mutations are characterized by at least two adjacent G/C nucleotides, a novel feature of EMS mutagenesis that has previously not been identified. Studies in *A. thaliana* reported an 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.

Lastly, 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 3.2**).

A simple screening can be easily performed on F_1s in *Daphnia*. As we calculated, 750 F_1s 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_1s, 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_1s 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_2s that are homozygous for the induced mutations so that recessive mutations can show their phenotypic effect (**Figure 3.1**). Technically speaking, if we start with 750 F_1s, 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 (**Figure 3.1**).

Although F_2s 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_2s collected from each F_1 line. This probability is written as $1 - (¾)^n$, where n is the number of F_2s and the term $(¾)^n$ denotes the probability of seeing non-homozygous mutants in n F_2 individuals. We can directly see that collecting 4 and 5 F_2s 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_2s. 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.

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.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Data Archiving

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*. *Genetics* **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.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, *et al.* (2011) The ecoresponsive 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 U S A* **106**: 16310–16314.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*.

Ebert D (2005) Introduction to *Daphnia* Biology. National Center for Biotechnology Information (US).

EMBOSS: Compseq. Retrieved from http://emboss.open-bio.org/rel/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.

Spradling, AC (1997) Fly pushing: The theory and practice of *Drosophila* genetics. Trends Genet, 13(10), 418.

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.

Graur D, Li W (2000) Fundamentals of Molecular Evolution (2nd Edition). Sinauer Associates.

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.

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*. *Genetics* **164**: 731–740.

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 LC 4th, 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.

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 Online Rev C Elegans Biol*: 1–26.

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.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical 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, Gutenkunst R, Ackerman M, Spitze K, Ye Z, Maruki T, *et al.* (2017) Population Genomics of *Daphnia pulex*. *Genetics* **206**: 315–332.

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.

Martín B, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C (2009) A high-density collection of EMS-induced mutations for TILLING in Landsberg erecta genetic background of *Arabidopsis*. *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**.

Page DR, Grossniklaus U (2002) The art and design of genetic screens: *Arabidopsis thaliana*. *Nat Rev Genet* **3**: 124–136.

Pastink A, Heemskerk 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 MGG* **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 on 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, pp 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 δ. *Int J Evol Biol* **2012**.

St Johnston D (2002) The art and design of genetic screens: *Drosophila melanogaster*. *Nat Rev Genet* **3**: 176–188.

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.

Figure 3.1. Forward genetic approach for obtaining mutant lines in *Daphnia*. This study used F_1 mutants to determine the mutation rate and spectrum of EMS-induced heritable mutations.

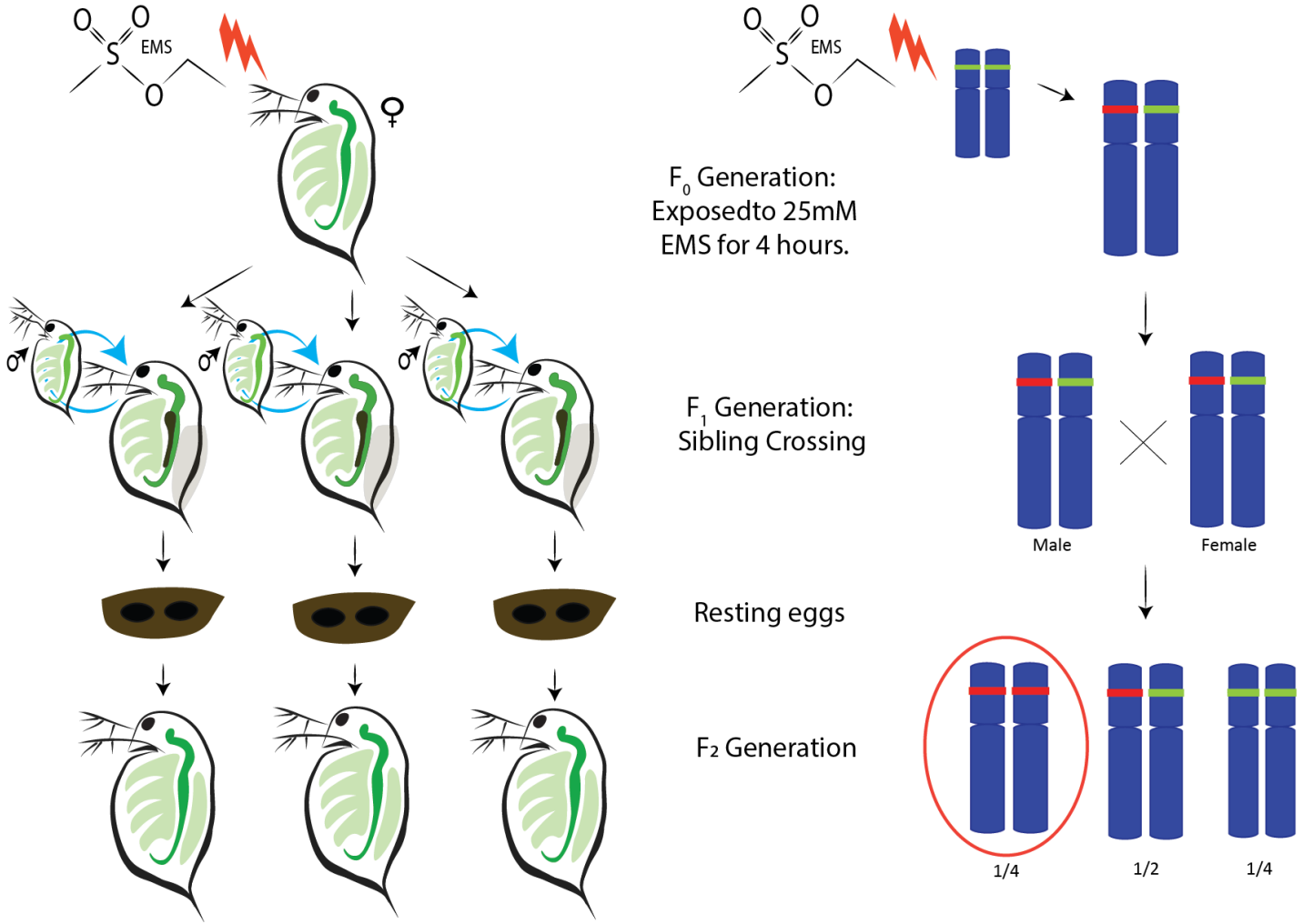


Figure 3.2. Experimental procedure for establishing EMS mutant lines. (A) Establishing replicate mutant lines of a *Daphnia* isolate. (B) Establishing brood-specific mutant lines of a *Daphnia* isolate.

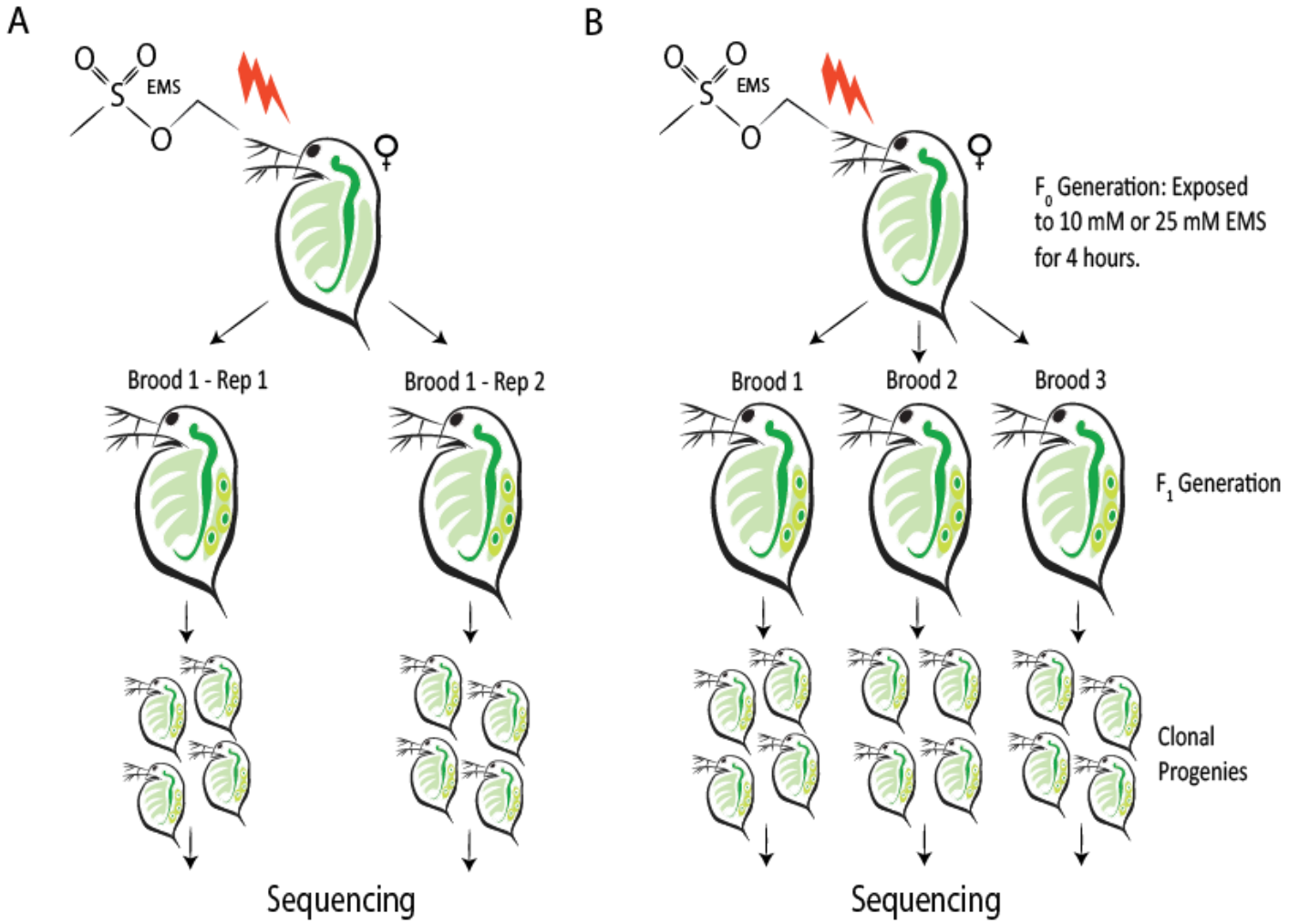


Figure 3.3. Base-substitution mutation rates of three *Daphnia* species at 10mM and 25mM EMS treatment. The bar plot summarizes the species-specific rates based on multiple isolates of each species, whereas the scatter plot represents brood-specific mutation rates in each species.

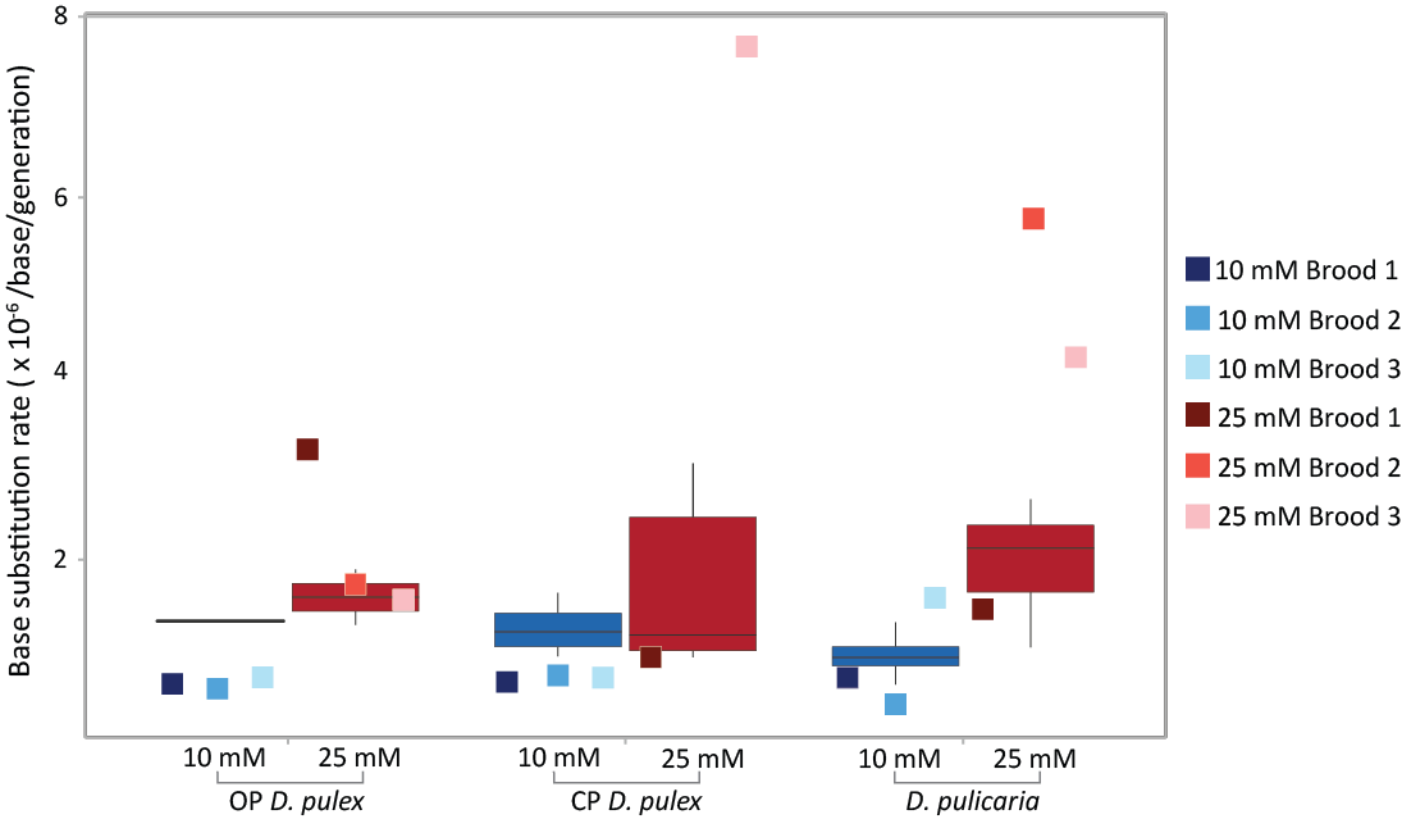


Figure 3.4. Average proportions of different types of base substitutions in each species at 10mM and 25mM EMS treatment. (A) Composition of base substitutions in each species. (B) Composition of base substitutions in different broods.

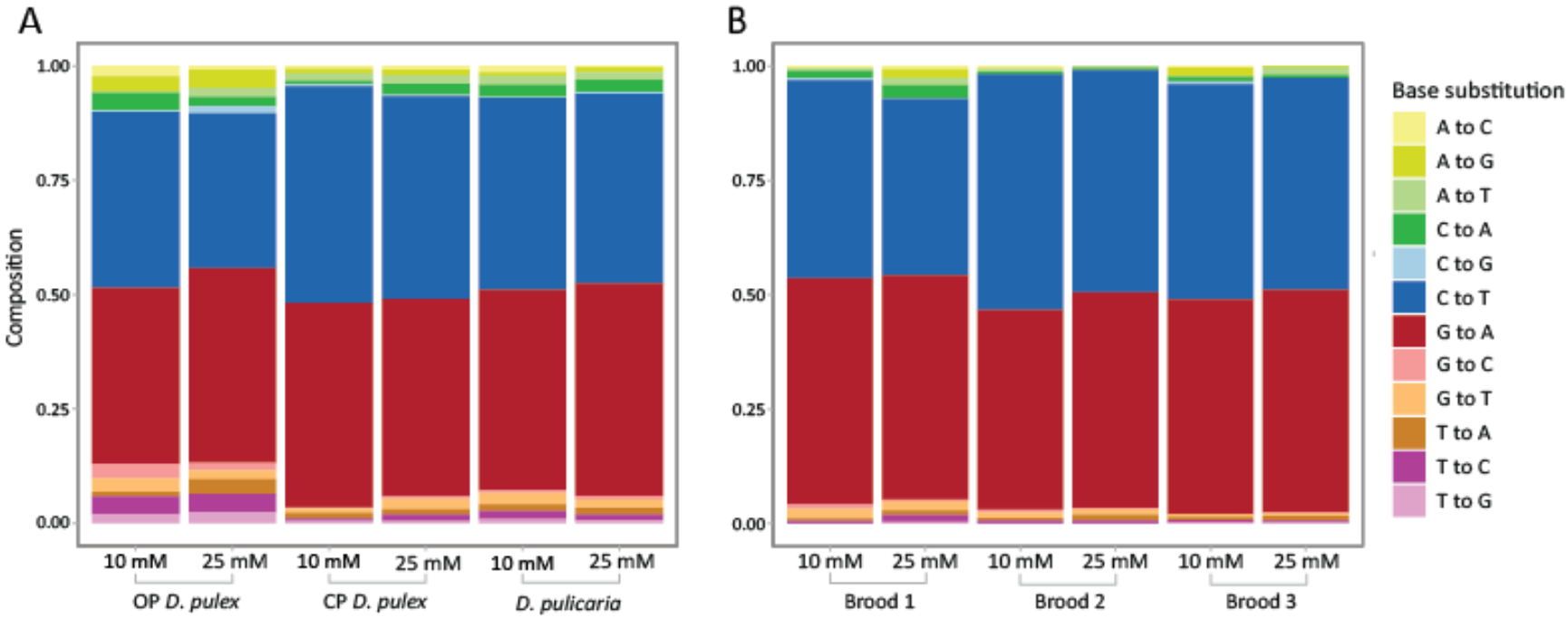


Figure 3.5. Average proportions of base substitutions in different genomic regions (A and B) and amino acid changing effects at 10mM and 25mM EMS concentration (C and D).

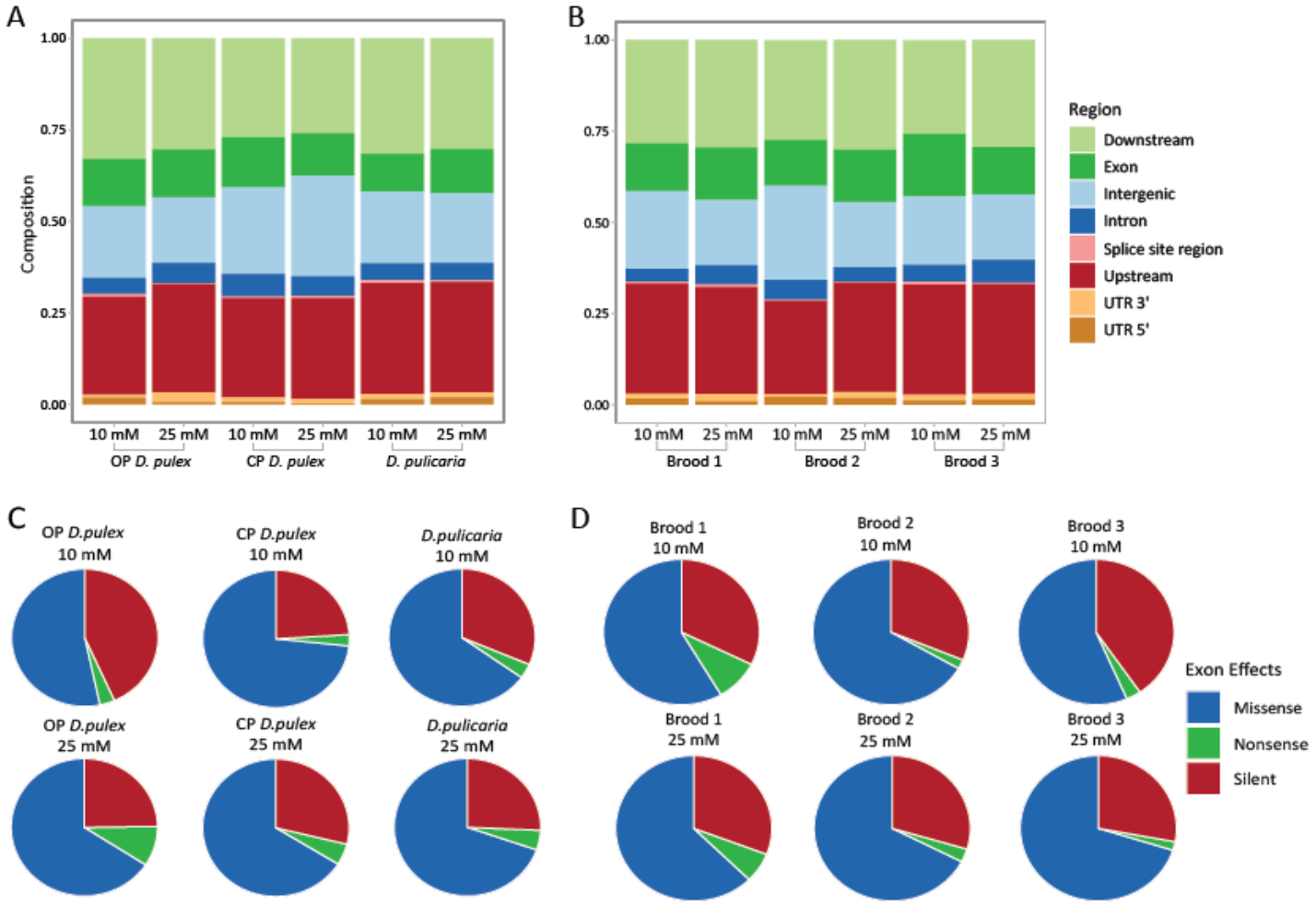


Figure 3.6. Bars represent the proportion of EMS-induced mutations at each trinucleotide motif centered at mutated sites (5′-3′ orientation), whereas the lines represent the observed proportion of trinucleotide frequencies observed in the *Daphnia* reference assemblies. NAN and NTN trinucleotides are significantly underrepresented, whereas many of the NGN and NCN motifs are overrepresented (indicated by asterisks).

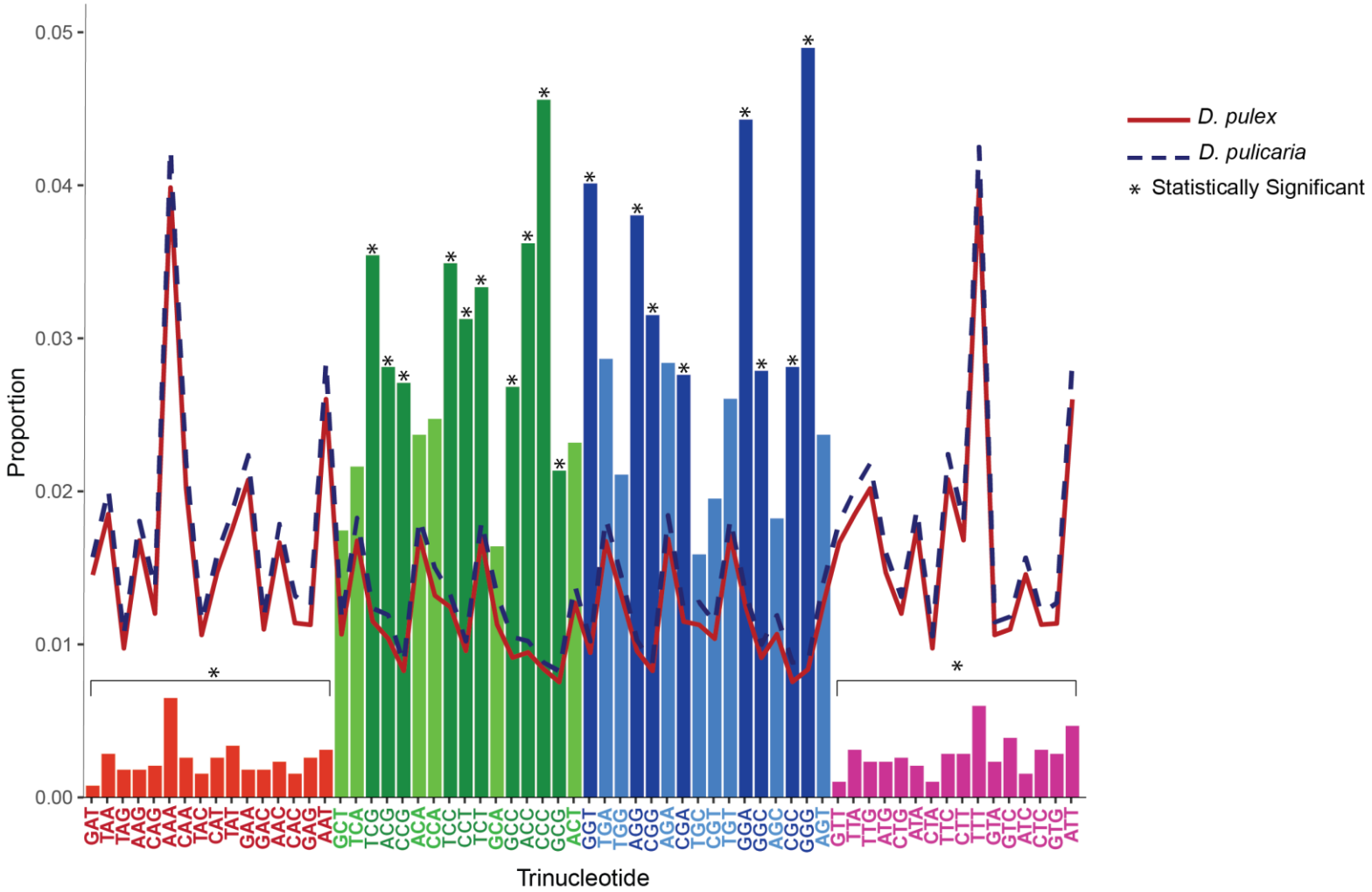


Table 3.1. Summary of mutations induced by 10mM and 25mM EMS.

EMS concentration	*Daphnia* 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)
10mM	*D. pulex* (OP)	126 (0.71)	1.24×10^{-6} (3.09×10^{-7})	48 (2)	3.9×10^{-3} (1.2×10^{-3})	21 (0)	1.9×10^{-3} (4.6×10^{-4})
	D. pulex (CP)	124 (32)	1.32×10^{-6} (8.47×10^{-9})	58 (18)	2.4×10^{-3} (1.3×10^{-4})	28 (5)	1.0×10^{-3} (3.4×10^{-6})
	D. pulicaria	94 (22)	9.40×10^{-7} (2.35×10^{-7})	40 (60)	1.9×10^{-3} (2.7×10^{-4})	16 (5)	7.8×10^{-4} (2.3×10^{-4})
25mM	*D. pulex* (OP)	154 (43)	1.69×10^{-6} (1.0×10^{-6})	65 (16)	4.1×10^{-3} (2.1×10^{-3})	36 (8)	1.8×10^{-3} (8.7×10^{-4})
	D. pulex (CP)	171 (104)	1.58×10^{-6} (4.34×10^{-7})	78 (40)	3.2x10-3 (7.5×10^{-4})	30 (14)	1.7×10^{-3} (3.7×10^{-4})
	D. pulicaria	201 (67)	1.98×10^{-6} (6.12×10^{-7})	91 (27)	4.4×10^{-3} (1.3×10^{-3})	43 (13)	2.1×10^{-3} (6.1×10^{-4})

Table 3.2. Number of genes, spontaneous base substitution rate, EMS treated mutation rate and estimated number of function affecting variants for different model organisms. *Daphnia* estimates are based on the results from 25mM EMS treatment.

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
C. elegans	[a]19 404	[e] 2.5×10^{-9}	[l]2.5×10^{-3}	49
D. melanogaster	[b]13 601	[f]3.5×10^{-9}; [g]5.49×10^{-9}	[m]1.0×10^{-3}	14
A. thaliana	[c]27 655	[h]6.95×10^{-9}, [i]5.90×10^{-9}	[i]3.0×10^{-3}	83
D. pulex (OP)	[d]18 440	[j]2.30×10^{-9}, [k]7.17×10^{-9}	4.1×10^{-3}	76
D. pulex (CP)	[d]18 440	[k]4.53×10^{-9}	3.2×10^{-3}	59
D. pulicaria	27 846	-	4.4×10^{-3}	123

Chen *et al.* 2005[a], Adams *et al.* 2000[b], Cheng *et al.* 2017[c], Ye *et al.* 2017[d], Denver *et al.* 2009[e], Keightley *et al.* 2009[f], Schrider *et al.* 2013[g], Weng *et al.* 2019[h], Ossowski *et al.* 2010[i], Flynn *et al.* 2017[j], Keith *et al.* 2016[k], Gengyo-Ando and Mitani, 2000[l], Spradling 1997[m]

CHAPTER 4: Effect of EMS-induced mutations on gene expression in *Daphnia*

Authors: Marelize Snyman and Sen Xu*

Department of Biology, University of Texas at Arlington, Arlington, Texas USA 76019.

*Correspondence author. 501 S Nedderman Dr., Arlington, Texas, USA 76019. Email: sen.xu@uta.edu

Keywords: *Daphnia*, mutagenesis, ethyl methanesulfonate, gene expression

Abstract

Forward genetic screens have aided in identifying mutations for a phenotype of interest in various model organisms. The alkylating agent, ethyl methanesulfonate, is a popular mutagen used in forward genetic screens to introduce point mutations throughout the genome to generate a phenotype of interest. In previous work, we demonstrated how EMS mutagenesis successfully induced mutations into the genome of the microcrustacean *Daphnia*. In this study, we expand on this earlier work by utilizing whole-genome sequencing data and RNA sequencing data for 16 OP *Daphnia* mutant lines to study how EMS-induced mutations affect gene expression. First, we showed that EMS mutagenesis successfully elevated the base substitution rate and per gene mutation rate, that most mutations were G:C to A:T transitions, and the effects were randomly distributed throughout the genome. Additionally, we showed that of the differentially expressed genes, a median of 51 genes per mutant line were directly impacted by an EMS-induced mutation. Most of these variants were modifier variants (83%), followed by moderate impact variants (10%), low impact variants (6%), and high impact variants (1%). Further, we analyzed differentially spliced genes and found that a median of 12 DS genes were impacted by an EMS-induced mutation. Most of these variants were modifier variants (64%) and moderate impact variants (22%).

Introduction

Forward genetic screens are a popular method used as a gene function discovery tool through the identification of mutations responsible for a phenotype of interest. With naturally occurring mutations being rare, scientists often utilize DNA damaging agents such as x-rays, ultraviolet radiation, or chemical mutagens to induce artificial mutations throughout the genome. The resulting mutants can then be screened for a mutant phenotype to study the biological process of interest. A popular chemical mutagen used to induce DNA damage for forward genetic screening is the alkylating agent ethyl methanesulfonate (EMS). 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). Additionally, 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). Forward genetic screens utilizing EMS mutagenesis have been developed and successfully used in a variety of different model organisms, including *C. elegans* (Brenner 1974, Flibotte *et al.* 2010, Kutscher and Shaham 2014), *D. melanogaster* (Blumenstiel *et al.* 2009, Bökel 2008), *Saccharomyces cerevisiae* (Mobini-Dehkordi *et al.* 2008, Prakash and Higgins 1982) and *Arabidopsis thaliana* (Greene *et al.* 2003, Martín *et al.* 2009, McCallum *et al.* 2000).

With the widespread success of utilizing EMS-induced mutations to study gene function, we established in a previous publication an EMS mutagenesis protocol for the microcrustacean *Daphnia*, a recognized model organism by the National Institute of Health (NIH) due to its ecoresponsive genome (Colbourne *et al.* 2011, Edison *et al.* 2016, Snyman *et al.* 2021). Despite having 27,276 protein-coding genes annotated in the *Daphnia* genome, around a third of these

genes are lineage-specific and lack detectable homologs in other eukaryotic genomes (Colbourne *et al.* 2011, Jackson *et al.* 2021). Thus, establishing an EMS mutagenesis protocol was the crucial first step in developing a forward genetic screening approach to help gain insight into these lineage-specific genes to answer fundamental questions related to invertebrate evolution, adaptation, and the evolution of novel phenotypes such as obligate asexuality. In our previous publication, we showed that exposing *Daphnia* to 10mM and 25mM EMS for 4 hours can successfully elevate the base substitution rate from around 2.30×10^{-9} to 7.17×10^{-9} per base per generation (Flynn *et al.* 2017, Keith *et al.* 2016), to 1.17×10^{-6} and 1.75×10^{-6} per base per generation for 10mM and 25mM EMS concentrations respectively. Further, we showed that around 86% - 87% of EMS-induced mutations were G/C to A/T transitions, and that these mutations were randomly distributed throughout the genome, both established characteristics of EMS-induced mutations (Greene *et al.* 2003). Exonic mutations were additionally annotated, and ~65-67% were missense, ~27-31% silent, and ~3-5% nonsense mutations (Snyman *et al.* 2021).

Building upon this previous work, we utilized an obligate parthenogenetic (OP) *Daphnia* isolate to investigate the effect that these EMS-induced mutations have on gene expression. Depending on where these mutations are located throughout the genome determines their effects. If mutations arise in transcriptional regulatory elements, mRNA expression could be affected. If mutations occur in genes, mRNA splicing, stability and translation could be affected. Non-synonymous (missense) mutations leading to amino acid changes can affect a protein's activity and potentially impact the expression and functions of other genes participating in the same pathway (Ding *et al.* 2015). Synonymous (silent) mutations may not alter the protein sequence but can impact mRNA half-life or translation rates, while the functional consequences of mutations in the 5' and 3' untranslated regions (UTRs) often go uncharacterized or unrecognized

(Robert and Pelletier 2018). Truncation mutations could affect mRNA transcript concentrations by inducing a premature stop codon, which could result in the removal of these transcripts by nonsense-mediated mRNA decay (Mendell and Dietz 2001, Nicholson *et al.* 2010; Noensie and Dietz 2001).

In this study, we chose an obligate parthenogenetic *Daphnia* isolate with the long-term goal of utilizing EMS mutagenesis in a forward genetic screening approach to study the genes and mechanisms underlying obligate asexuality. During favorable environmental conditions such as high food abundance, low population density, long photoperiod, and warmer temperatures, OP *Daphnia* females produce diploid subitaneous eggs giving rise to offspring genetically identical to the mother **(Figure 4.1)**. Under stressful environmental conditions (e.g., high population density, lack of food availability, shorter photoperiod, and colder temperatures), these subitaneous eggs can develop into genetically identical males, as sex is environmentally determined (Gorr *et al.* 2006, Tatarazako *et al.* 2003). Similar to subitaneous eggs, resting eggs produced during stressful environmental conditions also maintain diploidy, and no fertilization is required. These resting embryos are deposited into a protective case (i.e., ephippium) and can resume development once the environment becomes favorable (Frisch *et al.* 2014). This contrasts with cyclical parthenogenetic *Daphnia* isolates where resting eggs are produced by conventional meiosis, and fertilization is required before being deposited into an ephippium.

This study examined the functional footprints of EMS-induced germline mutations in 16 OP *Daphnia* mutant lines utilizing whole-genome and RNA sequencing data. First, we identified EMS-induced heterozygous mutations and differentially expressed (DE) genes throughout the genome in all 16 mutant lines. Next, we overlapped these two data sets to identify genes differentially expressed and impacted by a mutation. These genes were then further categorized

according to the impact type of the variant nl. high, moderate, low, or modifier, as well as their specific effects. Further, we examined the influence of high-impact variants on gene expression and the impact of EMS-induced mutations on differential splicing in each mutant line.

Materials and Methods

Sampling and maintenance of isolates

An Obligate parthenogenetic *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.

EMS-induced mutants

Sexually mature *Daphnia* females were exposed to 25mM ethyl methanesulfonate for 4 hours to introduce mutations into the genomes of oocytes before being individually isolated in artificial lake water **(Figure 4.2)**. Two first brood progenies (G_1s) of the exposed females were collected and individually isolated, establishing 16 mutant lines. Further, each of the isolated G_1 individuals were propagated until a high enough density was researched to allow for ephippia production. 5-10 ephippia were then collected for each mutant line and hatched according to the protocol established by Luu *et al.* 2020. Briefly, after collection, the ephippia were dissected to release the encased resting eggs. Afterwards, the resting eggs were kept in the dark at 18°C for two weeks before being placed under UV light to stimulate development. If no development was observed after five days, the resting eggs were placed back in the dark and the cycle repeated until at least one resting egg hatched per mutant line. The hatched mutant offspring (G_2) were then grown for DNA and RNA extraction.

DNA extraction and whole-genome sequencing

A total of 30-40 clonal offspring were collected per mutant line for DNA extraction using a CTAB (Cetyl Trimethyl Ammonium Bromide) method (Doyle and Doyle 1987). DNA quality and quantity 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 libraries were sequenced on an MGI DNBseq platform with 150-bp paired-end reads and a targeted sequencing coverage of 30X per mutant line.

RNA extraction and sequencing

Experimental animals were maintained in the same environmental conditions (18℃ with a 16:8 light/dark cycle), and G_2 offspring was used to account for maternal effects due to EMS exposure as this could significantly impact gene expression. Afterward, three replicates of 2-3 day old offspring were collected from all 16 mutant lines including the wild-type to act as a control line. RNA was extracted using the Promega SV Total RNA Isolation kit (Madison, WI, USA) by following the manufacturer's instructions. RNA quality was confirmed by electrophoresis on a 2% agarose gel, and 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) by following the manufacturer's instructions. Sequencing was done by the Beijing Genomics Institute (BGI, Cambridge, MA, USA). Each library was sequenced on an MGI DNBseq platform with 150-bp paired-end reads.

Quality control and mapping

Quality of the raw reads were 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, therefor Trimmomatic v.0.39 (Bolger *et al.* 2014) was used to perform adapter trimming and quality filtering using the following parameters: ILLUMINACLIP adapter.fasta:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 MINLEN:36. The following Illumina adapter sequences, found in adapter.fasta, were trimmed, read 1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA, read 2: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT. Lastly, reads were reassessed using FastQC to confirm the removal of low-quality reads and adapter sequences.

Identification of EMS-induced mutations

EMS-induced mutations were identified according to the EMS mutagenesis protocol established by Snyman *et al.* in 2021. Briefly, default parameters of the Burrows-Wheeler Alignment Tool BWA-MEM version 0.7.17 (Li and Durbin 2009) was used to align the whole-genome 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. Default parameters were used for BCFtools (Li 2011) mpileup and call functions 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 polymorphism sites (SNPs)

with a sequencing depth (DP) >= 10 and <= 60, quality score (QUAL) >= 20, and is located at least 50-bp from an indel.

Further, a custom python script (https://github.com/Marelize007/Functional_impact_of_EMS-induced_mutations) was used to identify the mutations. As described in Snyman *et al.* 2021, mutations were identified using a consensus method. 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 lines had to agree to generate a consensus genotype call. A tentative mutation was identified if the mutant line showed a different genotype than the consensus. Further, 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 STAR aligner (Dobin *et al.* 2013) and 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 completed in R (R Core Team 2017) using DESeq2 v.1.34.0 (Love *et al.* 2014). Differentially expressed genes were determined for each mutant line using the Wald negative binomial test with the design formula ~ genotype, where genotype represents either the mutant or wild-type genotype. The Benjamini-Hochberg method was used to adjust the p-values for multiple testing, and transcripts with a p-value <0.05 were considered significantly differentially expressed. Transcripts were additionally filtered according to fold-change. Transcripts with a fold-change > 1.5 were considered upregulated, and < -1.5 were considered downregulated.

Mutation rate calculation

The formula, $\mu = m/n * 1$, 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. Further, *l* represents one generation. To calculate the per gene per generation mutation rate, we used the formula $\mu_g = m_g/n_g * 1$, where *mg* represents the total number of mutations detected in genic regions, including UTRs, exons, and introns, *ng* 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 *mg* represented the number of non-synonymous mutations per mutant line.

Annotation of EM-induced mutations

Functional annotation based on genomic locations and effect prediction of EMS-induced 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.

Differential splicing analysis

The tool rMATS v4.1.1 (Shen *et al.* 2014) was utilized to detect differentially spliced (DS) 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). Differentially 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 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 and RNA sequencing

A total of 16 mutant lines were whole genome sequenced with 150 bp Illumina paired-end reads **(Supplementary Table S4.1)**. We obtained ~6GB of raw data for each mutant line with an average of ~34 million reads per sample. Read trimming was not performed since a FastQC analysis did not reveal any issues relating to adapter contamination. After removing PCR duplicates and reads mapped to multiple locations, each mutant line had on average ~29 million mapped reads with an average coverage of 25 reads (SD = 2) per site.

Further, RNA sequencing data was obtained for three replicates per mutant line and one wild-type line (control), totaling 51 samples. Reads were sequenced with 150-bp Illumina paired-end reads and ~ 20 million reads were obtained per sample. After adapter trimming and quality control, ~ 99% of the reads were kept and mapped to the *D. pulicaria* reference genome with an average mapping rate of 95% **(Supplementary Table S4.2)**.

EMS-induced mutation rate and spectrum

Multiple individuals of the same OP *Daphnia* isolate were treated with 25mM EMS for 4 hours. Offspring of the mutagenized females were collected; 16 mutant lines were established, and DNA was extracted and sequenced. The base substitution rate for these lines ranged from

1.56×10^{-6} to 7.43×10^{-6} with a median base substitution rate of 1.93×10^{-6} (SE = 4.67×10^{-6}) per site per generation. Further, we determined the per gene mutation rate and non-synonymous mutation rate. The per gene mutation rate ranged from 2.0×10^{-3} to 1.3×10^{-2} with a median of 3.1×10^{-3} (SE = 8.9×10^{-4}) per gene per generation. The non-synonymous mutation rate resulted in a median of 2.70×10^{-4} (SE = 1.7×10^{-4}) per gene per generation **(Figure 4.3 A and B, Supplementary Table S4.3)**.

In accordance with previous work (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 across all 16 mutant lines were G:C to A:T transitions (mean = 90%, SD = 4%) yielding a transition-transversion ratio greater than 5.06 for all treated lines **(Figure 4.3 C, Supplementary Table S4.4)**. The effects of the EMS-induced mutations were additionally randomly distributed throughout the genome (chi-squared test $p > 0.05$), with ~31% found in the downstream region, ~11% in exons, ~18% in intergenic, ~5% in introns, ~0.6% splice site region, ~32% in the upstream region, ~1% in 3' UTR and ~1% in 5' UTR **(Figure 4.3 D, Supplementary Table S4.5)**.

Differentially expressed genes

The regularized log (rlog) transformation function in DEseq2 was used to normalize the mapped read counts. A principal component analysis was utilized to visualize the grouping of the mutant samples **(Figure 4.4 A)**. The first principal component accounted for 32% of the variance, while the second principal component accounted for 21%.

Across all 16 mutant lines, the number of differentially expressed (DE) genes ranged from 1176 to 6606, with a median of 3545 differentially expressed genes per mutant line.

Further, of these DE genes, a median of 1716 DE genes were upregulated, and 2056 DE genes were downregulated **(Figure 4.4 B)**. The number of DE genes affected by EMS-induced mutations ranged from 5 to 423, with a median of 50.5 **(Figure 4.4 C)**. Of the DE genes affected by mutations, a median of 32 genes were upregulated, and 25 were downregulated per mutant line.

Impact and effect of EMS-induced mutations

Differentially expressed genes impacted by mutations were further analyzed using SnpEff to determine their effect region and functional consequences. Across all 16 mutant lines, the effects were distributed as follow; ~2% were found in both 3' UTR and 5' UTR region, ~34% in the downstream region, 37% in the upstream region, ~16% in exons, ~7% in introns and ~1% in the splice site region **(Figure 4.5 A)**. Next, mutations were categorized as high impact, moderate impact, low impact, or modifier variants. High impact variants disrupt protein function and could potentially lead to a loss of function, protein truncation, or nonsense-mediated RNA decay. Moderate impact variants influence a protein's effectiveness; low impact variants do not affect protein behavior, while modifier variants affect non-coding regions (Cingolani *et al.* 2012). Across all 16 mutant lines, we identified 27 (1%) high impact, 184 (10%) moderate impact, 113 (6%) low impact and 1553 (83%) modifier variants **(Figure 4.5 B, Supplementary Table S4.6)**.

Each impact category was further broken down into different variant types. High impact variants consisted of variants that mutated a start codon into a non-start codon (start-lost, 0.1%), variants introducing a stop codon (stop-gained, 1.2%), and variants impacting a splice acceptor site (splice acceptor variant, 0.1%). Moderate impact variants identified consisted only of missense variants (8.0%). Low impact variants identified were variants affecting the splice site region (splice region variant, 0.7%), synonymous variants (5.5%), and variants producing a

premature start codon in the 5' UTR region (5' UTR premature start codon gain variant, 0.2%). Lastly, modifier variants identified were variants 5K bases downstream of a gene (downstream variant, 35.0%), variants 5K bases upstream of a gene (upstream variant, 37.6%), variants within the 3' UTR region (3' UTR variant, 2.3%), variants within the 5' UTR region (5' UTR variant, 1.6%), and variants within introns (intron variant, 7.5%, **Figure 4.5 C, Supplementary Table S4.7**).

Lastly, we investigated gene regulation patterns of the identified high-impact variants across all 16 EMS-induced mutant lines. The two identified splice acceptor variants were both downregulated, while the two start-lost variants were both upregulated. Except for three of the 27 stop-gained variants identified, most were upregulated across mutant lines **(Figure 4.6, Supplementary Table S4.7)**.

Differential splicing

First, we identified differentially spliced (DS) genes by comparing all 16 mutant lines to the wild type. The number of differentially spliced genes ranged from 212 to 627, with a median of 393 differentially spliced genes per mutant line **(Figure 4.7 A)**. Of these DS genes, 11% were A3SS events, 11% A5SS events, 17% MXE events, 32% RI events, and 30% SE events **(Figure 4.7 B)**. The number of DS genes affected by an EMS-induced mutation ranged from 2 to 57, with a median of 12 genes per mutant line **(Figure 4.7 A)**. Of these genes, 8% were A3SS events, 9% A5SS events, 19% MXE events, 22% RI events, and 42% SE events. A median of 2 DS genes affected by a mutation were additionally differentially expressed per mutant line **(Figure 4.7 C, Supplementary Table S4.9)**. Mutations affecting differentially spliced genes were further categorized according to impact, with 3% being high impact, 9% low impact, 22% moderate impact, and 64% modifier impact variants **(Supplementary Table S4.10)**. High impact variants

consisted of 0.5% splice acceptor variant, 1% splice donor variant, and 2% stop-gained variants. Moderate impact variants consisted of 10% missense variants. Low impact variants consisted of 0.5% splice region variant and 7% synonymous variant. Modifier impact variants consisted of 2% 3' UTR variant, 20% downstream variant, 11% intron variant, and 49% upstream variant **(Supplementary Table S4.11)**.

Discussion

In this study, we examined the functional consequences of EMS-induced mutations across 16 OP *Daphnia* mutant lines by whole-genome and RNA sequencing. First, we identified EMS-induced mutations via a consensus method previously established in (Snyman *et al.* 2021). We calculated the genome-wide base substitution rate, per gene mutation rate, non-synonymous mutation rate as well as examined the types of base substitutions and the distribution of mutation effect across all 16 EMS mutant lines **(Figure 3 A, B, C, and D)**. Further, we identified differentially expressed and differentially spliced genes, as well as DE and DS genes impacted by EMS-induced mutations. Lastly, we examined the functional impacts and effects of DE and DS genes affected by EMS-induced mutations.

The EMS-induced base substitution rate, per gene mutations rate, and non-synonymous mutation rate across all 16 mutant lines were elevated to a median of 1.93×10^{-6} (SE = 4.67×10^{-6}) per site per generation, 3.1×10^{-3} (SE = 8.9×10^{-4}) per gene per generation and 2.70×10^{-4} (SE = 1.7×10^{-4}) respectively **(Figure 3 A, B)**. These results are in line with previous work, which showed that the base substitution rate for *Daphnia* exposed to 25mM EMS for 4 hours ranged from 1.58×10^{-6} to 1.98×10^{-6} per site per generation, and the average per gene mutation rate and non-synonymous mutation rate was 4.09×10^{-3} and 1.91×10^{-3} per gene per generation respectively (Snyman *et al.* 2021). Additionally, our EMS-induced mutation rate was higher than the

spontaneous mutation rate calculated from various mutation accumulation experiments, which ranged from 2.30×10^{-9} to 7.17×10^{-9} per site per generation in *Daphnia* (Bull *et al.* 2019, Flynn *et al.* 2017, Keith *et al.* 2016). Further, our results showed that 90% of base-substitutions were G:C to A:T transitions, and the mutational effects were randomly distributed throughout the genome **(Figure 3C, D)**, both known characteristics of EMS-induced mutations (Greene *et al.* 2003).

A median of 3545 differentially expressed genes were identified per mutant line; of these genes, a median of 51 genes were affected by EMS-induced mutations **(Figure 4B)**. With the number of differentially expressed genes significantly higher than the number of mutated genes could be indicative of trans-effects, where mutations affect the expression of genes located at other sites within the genome. This is in contrast to cis-effects, where mutations only impact the expression of the gene they are located in (Curtis *et al.* 2012).

Of the mutations identified in differentially expressed genes, 83% were modifier variants, 10% moderate impact variants, 6% low impact variants, and 1% high impact variants **(Figure 5A, Supplementary Table S6)**. High impact variants consisted of start-lost, stop-gained, and splice acceptor variants. For the stop-gained (nonsense) variants, we observed a decrease in mRNA expression for a couple of genes which could indicate nonsense-mediated RNA decay (NMD, **Figure 7**). 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). Most of the stop-gained variants, however, were upregulated compared to the wild-type, which 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). Future work could expand on this study by observing

expression at the protein level. If a decrease in protein expression is observed, it could indicate the formation of truncated proteins (Bordeira-Carriço *et al.* 2012).

Further, the identified star-lost variants were upregulated, while the splice acceptor variants were downregulated. The identified moderate impact variants consisted only of missense variants, and genes were roughly split between upregulated and downregulated genes. Since missense variants change the amino acid sequence, they are not generally expected to affect gene expression. This upregulation and downregulation of genes could indicate selection for and against advantageous and detrimental genes respectively (Jia and Zhao 2017). Missense variants have been shown to affect DNA-transcription factors resulting in the changed expression of the corresponding protein (Hanemann *et al.* 2000). These results indicate that mutational impacts differ across the genome, and to gain a clear understanding of the functional consequences mRNA data needs to be supplemented with protein expression data.

Changes in pre-mRNA splicing can be another source of phenotypic variation. Exon-intron boundaries consist of highly conserved splice site sequences that must be correctly identified by the spliceosome to perform a splicing reaction. Additionally, splicing factors interact with intronic and exonic sequences to determine how frequently an exon is included in the final transcript (Wang and Burge 2008). Premature stop codons and transcripts with an altered amino acid sequence can result if this process is not correctly regulated. In our dataset, only one of the mutant lines showed a differentially spliced gene affected by a stop-gained mutation. Most DS genes were affected by low impact variants (9%, splice region variant and synonymous variant) and modifier impact variants (64%, 3’ UTR variant, downstream variant, intron, and upstream variant).

In conclusion, we identified EMS-induced mutations in 16 OP *Daphnia* mutant lines and examined gene expression patterns throughout the genome. First, we determined the base substitution rate, per gene mutation rate, and non-synonymous mutation rate for all mutant lines showed that EMS-induced mutations were randomly distributed throughout the genome and enriched for G:C to A:T transitions. Lastly, we determined the functional consequences of DE and DS genes impacted by an EMS-induced mutation.

References

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/

Blumenstiel JP, Noll AC, Griffiths JA, Perera AG, Walton KN, Gilliland WD, Hawley RS, Staehling-Hampton K (2009) Identification of EMS-Induced Mutations in *Drosophila melanogaster* by Whole-Genome Sequencing. *Genetics*, *182*(1), 25–32.

Bökel C (2008) EMS screens: From mutagenesis to screening and mapping. *Methods in Molecular Biology (Clifton, N.J.)*, *420*, 119–138.

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

Bordeira-Carriço R, Pêgo AP, Santos M, Oliveira C (2012) Cancer syndromes and therapy by stop-codon readthrough. *Trends in Molecular Medicine*, *18*(11), 667–678.

Brenner S (1974) The genetics of *Caenorhabditis elegans*. *Genetics*, *77*(1), 71–94.

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.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, *et al.* (2011) The ecoresponsive genome of Daphnia pulex. *Science (New York, N.Y.)*, *331*(6017), 555–561.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ *et al.* (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, *486*(7403), 346–352.

Ding J, McConechy MK, Horlings HM, Ha G, Chun Chan F, Funnell T (2015) Systematic analysis of somatic mutations impacting gene expression in 12 tumour types. *Nature Communications*, *6*, 8554.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.

Doyle JJ, Doyle JL (1987) *A rapid DNA isolation procedure for small quantities of fresh leaf tissue* (RESEARCH). PHYTOCHEMICAL BULLETIN.

Edison AS, Hall RD, Junot C, Karp PD, Kurland IJ, Mistrik R, Reed LK, Saito K, Salek RM, Steinbeck C, Sumner LW, Viant, MR (2016) The Time Is Right to Focus on Model Organism Metabolomes. *Metabolites*, *6*(1), 8.

El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Günther S, Fukuda N *et al.* (2019) Genetic compensation triggered by mutant mRNA degradation. *Nature*, *568*(7751), 193–197.

Flibotte S, Edgley ML, Chaudhry I, Taylor J, Neil SE, Rogula A, Zapf R, Hirst M, Butterfield Y, Jones SJ, Marra MA, Barstead RJ, Moerman DG (2010) Whole-Genome Profiling of Mutagenesis in *Caenorhabditis elegans*. *Genetics*, *185*(2), 431–441.

Flynn JM, Caldas I, Cristescu ME, Clark AG (2017) Selection Constrains High Rates of Tandem Repetitive DNA Mutation in *Daphnia pulex*. *Genetics*, *207*(2), 697–710.

Frisch D, Morton PK, Chowdhury PR, Culver BW, Colbourne JK, Weider LJ, Jeyasingh PD (2014) A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecology Letters*, *17*(3), 360–368.

Gorr TA, Rider CV, Wang HY, Olmstead AW, LeBlanc GA (2006) A candidate juvenoid hormone receptor cis-element in the *Daphnia magna* hb2 hemoglobin gene promoter. *Molecular and Cellular Endocrinology*, *247*(1–2), 91–102.

Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, Henikoff S (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*, *164*(2), 731–740.

Hanemann CO, D'Urso D, Gabreëls-Festen AAWM, Müller HW (2000) Mutation-dependent alteration in cellular distribution of peripheral myelin protein 22 in nerve biopsies from Charcot–Marie–Tooth type 1A. *Brain*, *123*(5), 1001–1006.

Jackson CE, Xu S, Ye Z, Pfrender ME, Lynch M, Colbourne JK, Shaw JR (2021) *Chromosomal rearrangements preserve adaptive divergence in ecological speciation* (p. 2021.08.20.457158).

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

Keith N, Tucker AE, Jackson CE, Sung W, Lucas Lledó JI, Schrider DR, Schaack S, Dudycha JL, Ackerman M, Younge AJ, Shaw J R, Lynch M (2016) Generation of loss- or gain of function mutants and weak nonlethal alleles. *Genome Research*, *26*(1), 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*(1), 147–159.

Kutscher LM, Shaham S (2014) Forward and reverse mutagenesis in *C. elegans*. *WormBook : The Online Review of C. Elegans Biology*, 1–26.

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, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (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.

Martín B, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C (2009) A high-density collection of EMS-induced mutations for TILLING in Landsberg erecta genetic background of *Arabidopsis*. *BMC Plant Biology*, *9*, 147.

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. *Nature Biotechnology*, *18*(4), 455–457.

Mendell JT, Dietz HC (2001) When the Message Goes Awry: Disease-Producing Mutations that Influence mRNA Content and Performance. *Cell*, *107*(4), 411–414.

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. *Journal of Bioscience and Bioengineering*, *105*(4), 403–408.

Nicholson P, Yepiskoposyan H, Metze S, Zamudio Orozco R, Kleinschmidt N, Mühlemann O (2010) Nonsense-mediated mRNA decay in human cells: Mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cellular and Molecular Life Sciences*, *67*(5), 677–700.

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

Noensie EN, Dietz HC (2001) A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. *Nature Biotechnology*, *19*(5), 434–439.

Pastink A, Heemskerk E, Nivard MJ, van Vliet CJ, Vogel EW (1991) Mutational specificity of ethyl methanesulfonate in excision-repair-proficient and -deficient strains of *Drosophila melanogaster*. *Molecular & General Genetics: MGG*, *229*(2), 213–218.

Pohl M, Bortfeldt RH, Grützmann K, Schuster S (2013) Alternative splicing of mutually exclusive exons—A review. *Biosystems*, *114*(1), 31–38.

Prakash L, Higgins D (1982) Role of DNA repair in ethyl methanesulfonate-induced mutagenesis in *Saccharomyces cerevisiae*. *Carcinogenesis*, *3*(4), 439–444.

Robert F, Pelletier J (2018) Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. *Frontiers in Genetics*, *9*.

RStudio Team (2020) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

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.

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.

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.

Tatarazako N, Oda S, Watanabe H, Morita M, Iguchi T (2003) Juvenile hormone agonists affect the occurrence of male *Daphnia*. *Chemosphere*, *53*(8), 827–833.

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.

Wang Z, Burge CB (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*, *14*(5), 802–813.

Acknowledgements

We would like to thank Anish Gamadia and Matthew T Smith for their help with tissue collection, and Trung Huynh for his help with Python scripting. Additionally, we would like to thank all Xu lab members for their helpful discussions.

Author Contributions

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

Tables and Figures

Figure 4.1. The life cycle of obligate parthenogenetic *Daphnia* isolates.

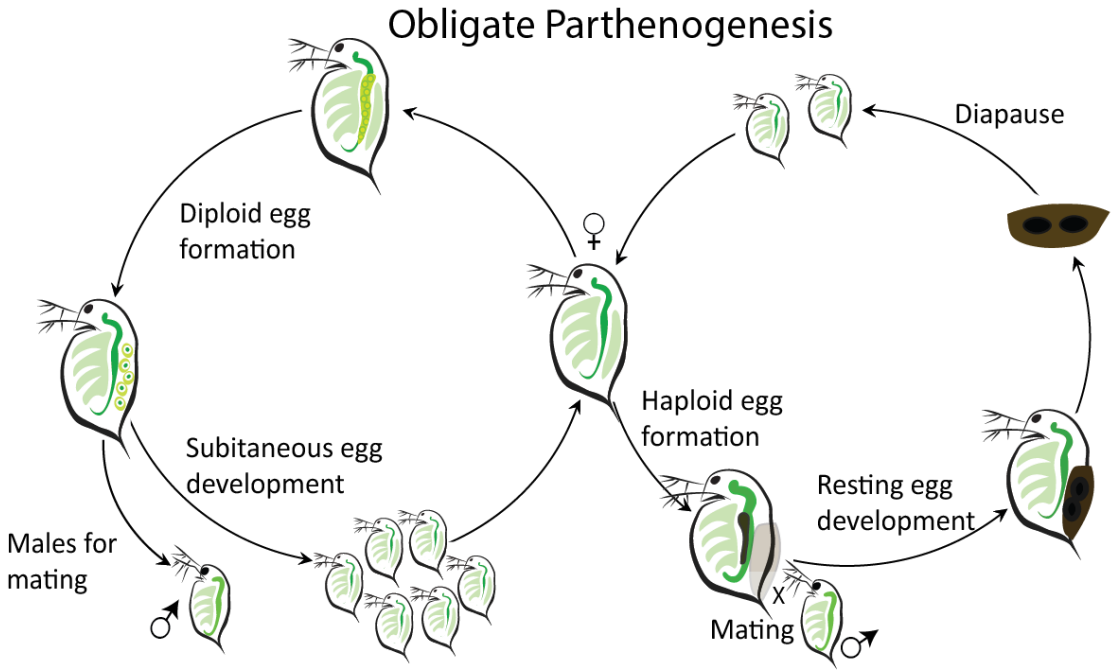


Figure 4.2. Illustration of exposure method to generate EMS mutant lines.

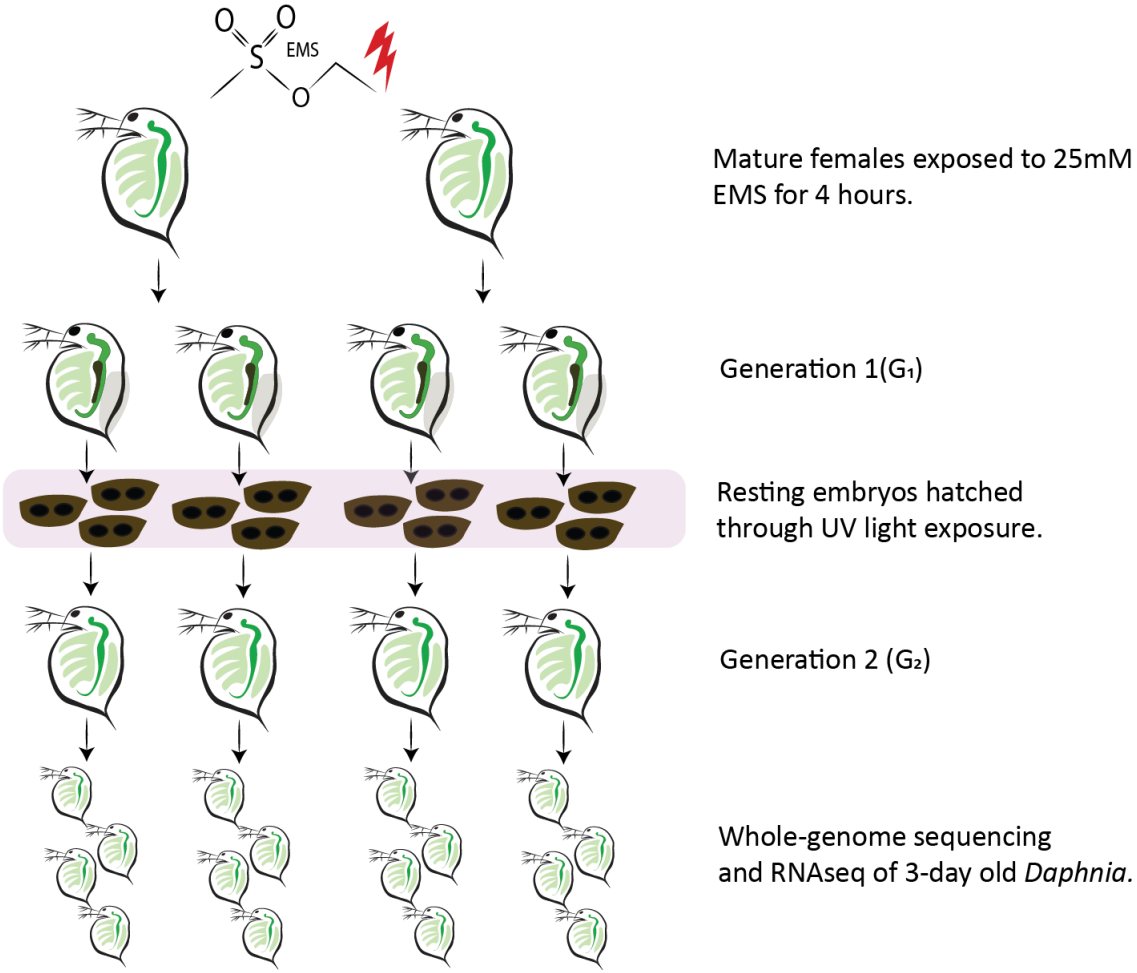


Figure 4.3. (A) The base substitution rate, (B) per gene and non-synonymous mutation rate across all 16 EMS mutant lines. (C) The proportion of base substitutions caused by EMS-induced mutations (D) and the regions affected.

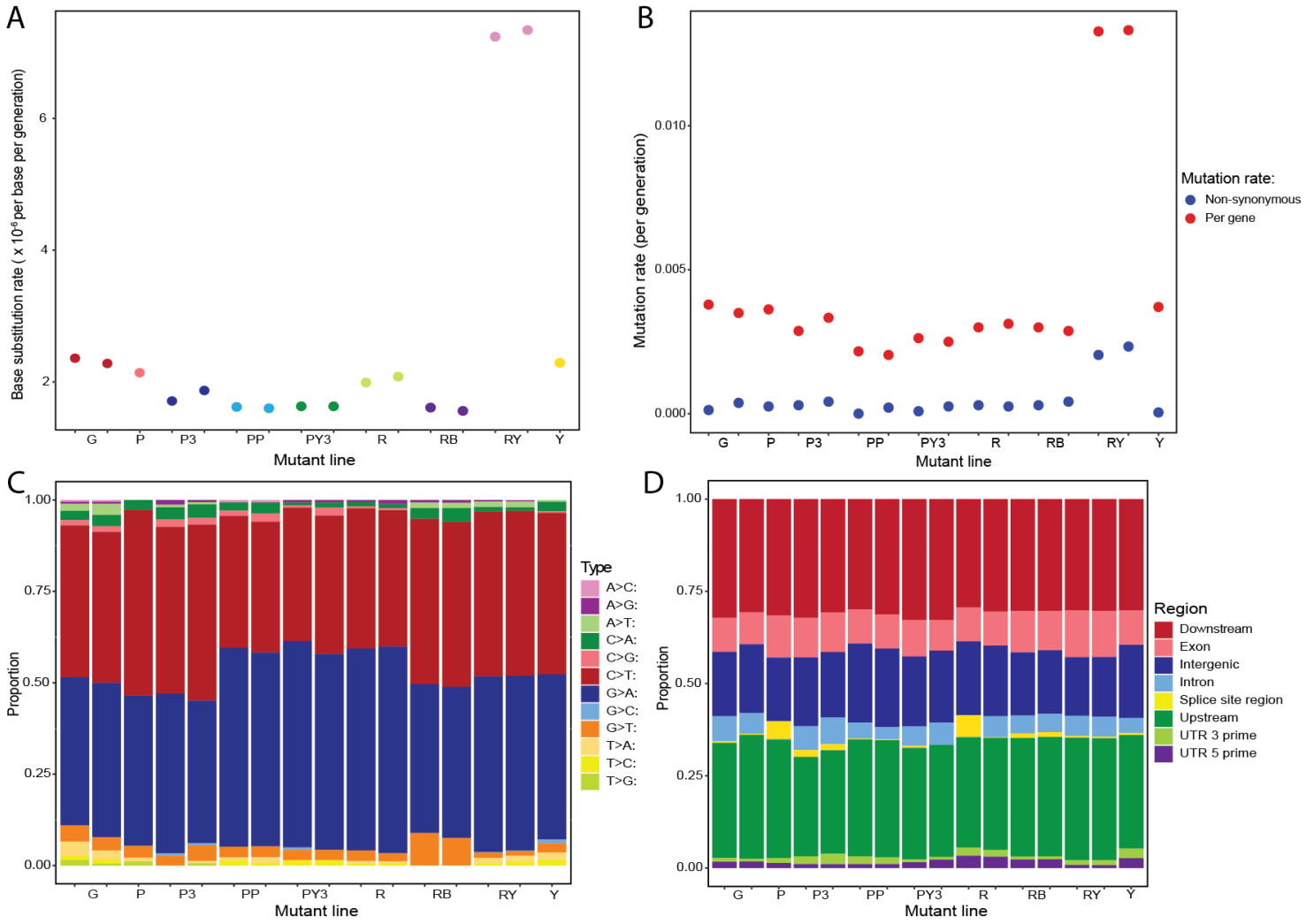


Figure 4.4. (A) PCA plot illustrating the variance between mutant lines due to EMS-induced mutations. (B) Direction of regulation for differentially expressed genes per mutant line. (C) Bar graph illustrating the number of DE genes, the number of genes affected by mutations, and the number of DE genes affected by mutations.

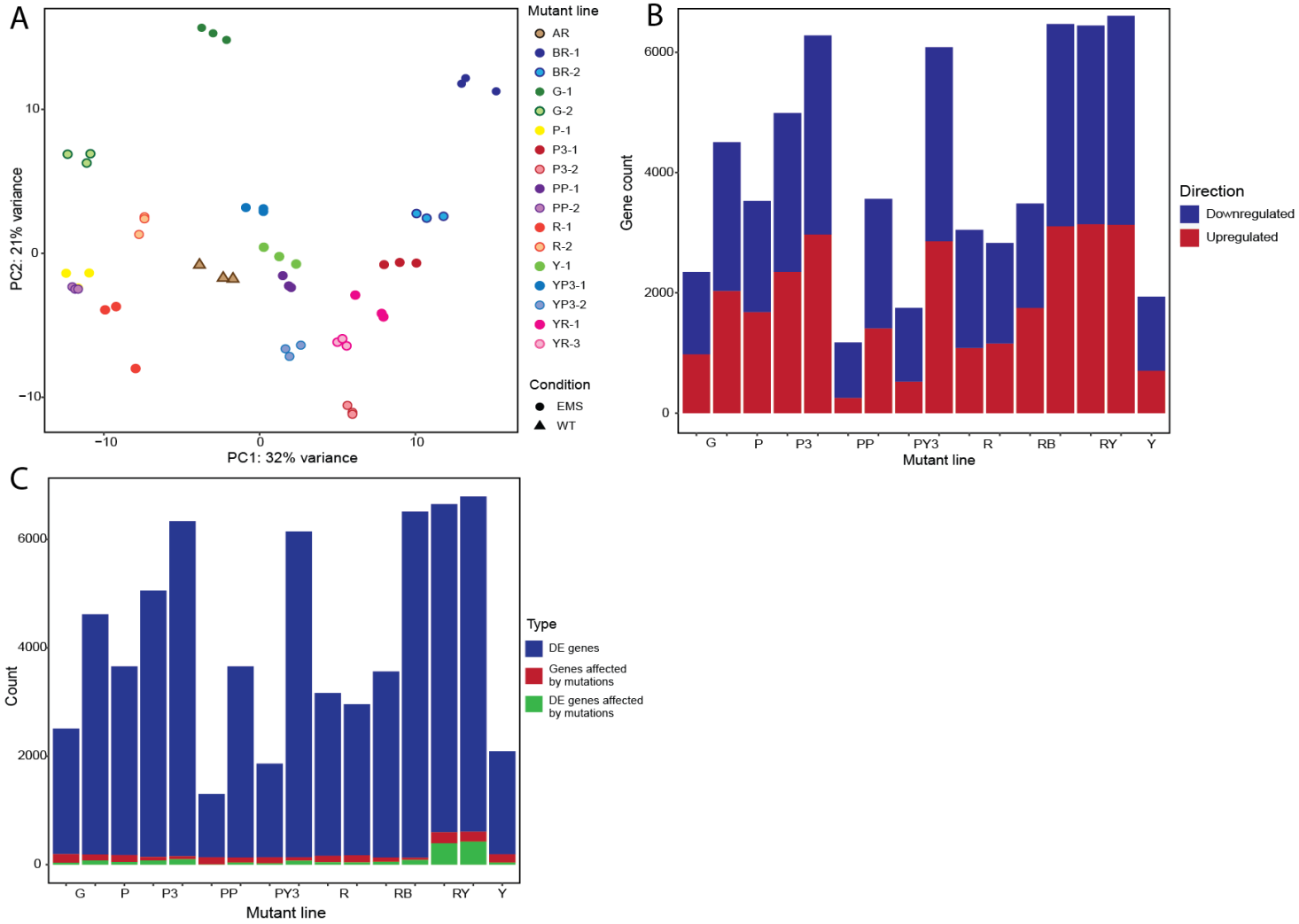


Figure 4.5. (A) Regions affected by genes that are DE and impacted by an EMS-induced mutation. (B)Variant impacts and (C) variant effects of differentially expressed genes affected by EMS-induced mutations.

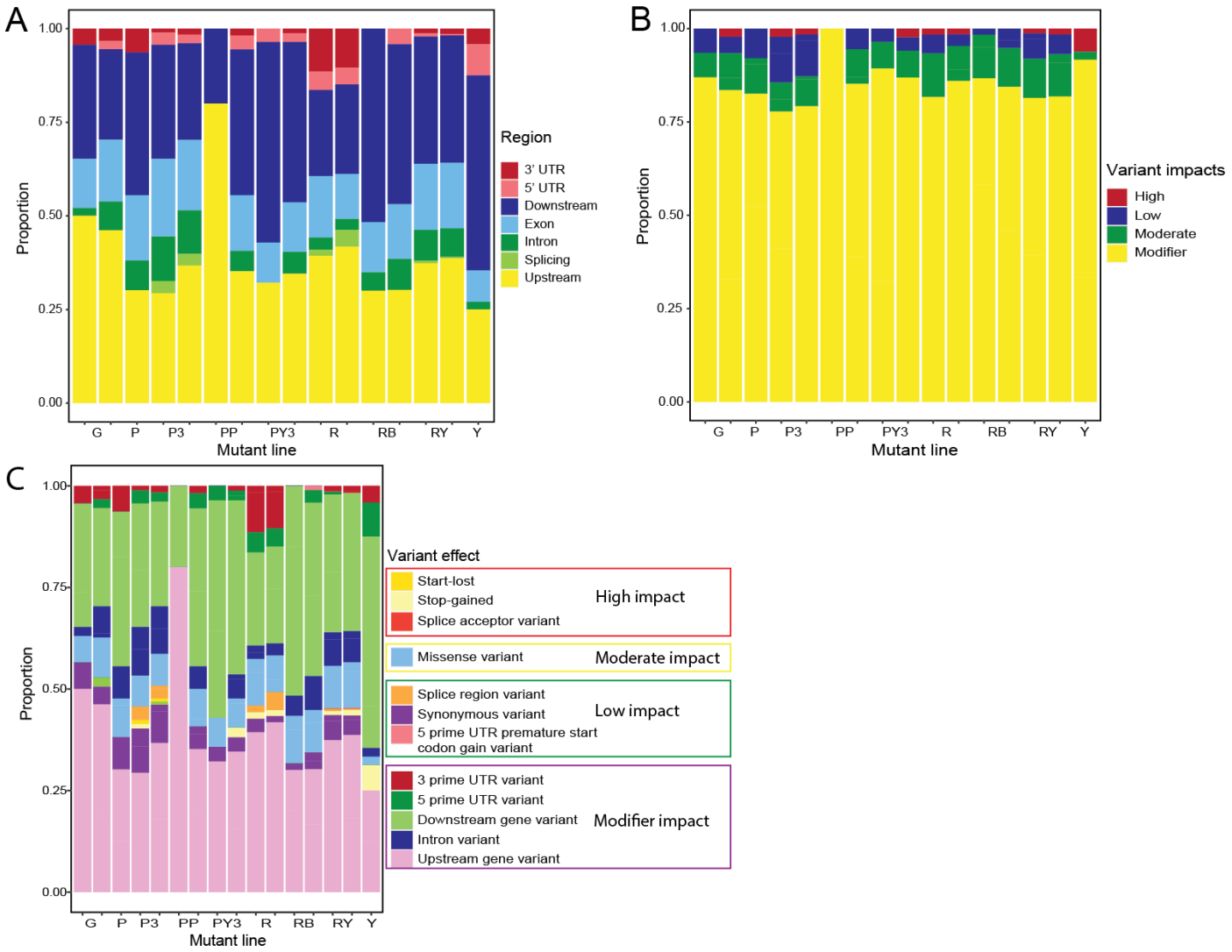


Figure 4.6. Log2 fold-change of high-impact and moderate-impact variants identified across all 16 EMS-induced mutant lines.

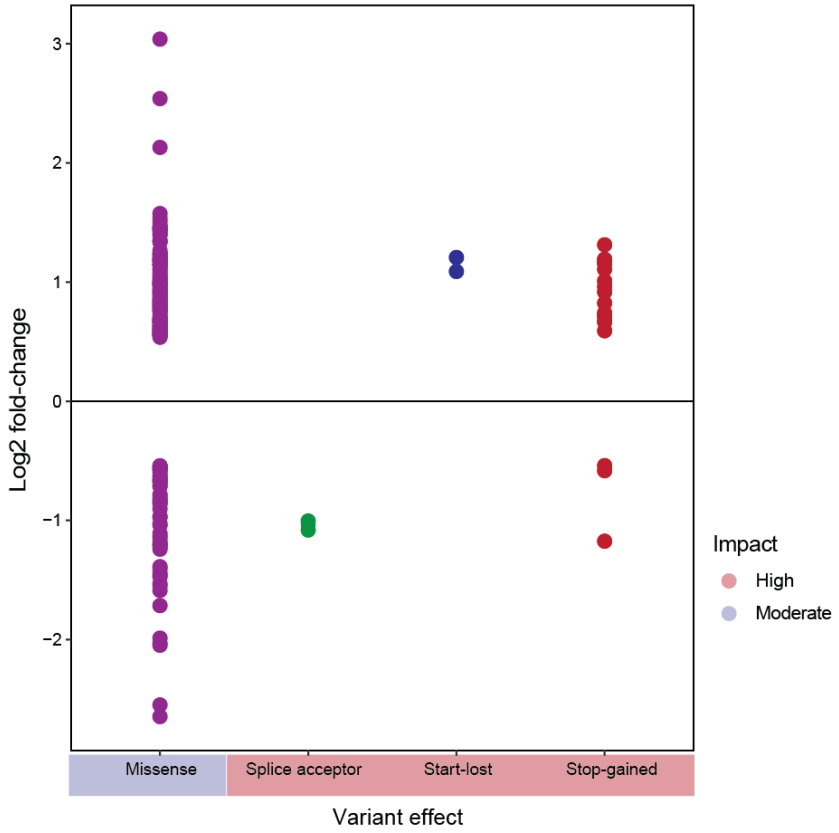
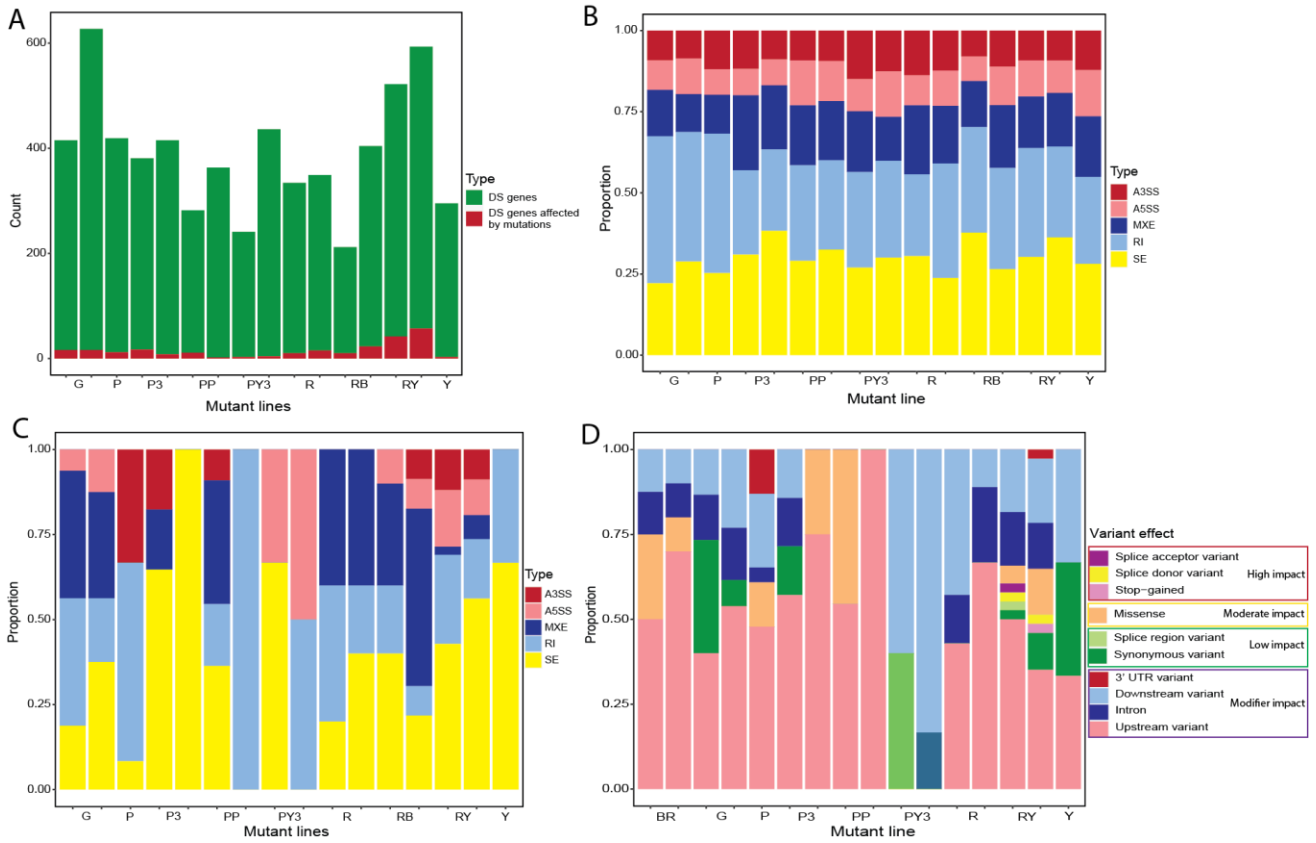


Figure 4.7. (A) Number of differentially spliced (DS) genes and DS genes affected by an EMS-induced mutation. (B) Composition of differentially spliced genes and (C) composition of differentially spliced genes impacted by an EMS-induced mutation. (D) Variant effects of differentially spliced genes affected by EMS-induced mutations.



CHAPTER 5: Conclusion

In this dissertation, I analyzed the gene expression changes associated with different reproductive modes and EMS-induced mutation in the microcrustacean *Daphnia*. The results from this dissertation will help give insight into the molecular basis of complex phenotypic traits that can arise through environmental changes or chemical mutagenesis.

The second chapter of this dissertation focuses on analyzing the gene expression patterns underlying two modes of parthenogenesis in obligate parthenogenetic (OP) *Daphnia pulex* isolates. During favorable environmental conditions, these isolates produce subitaneous eggs, and once the environment turns unfavorable, they switch to resting egg production. Both subitaneous and resting eggs are produced parthenogenetically. In the second chapter of this dissertation, I analyzed and compared the genome-wide expression, and differential splicing patterns between OP *D. pulex* isolates participating in early subitaneous and early resting egg production to elucidate the underlying genes and mechanisms of these two parthenogenetic modes. The analysis for this study was done in two different ways; through a pooled transcriptomic analysis and by comparing these two stages within each isolate. This allowed for a comprehensive overview of the main differences between early subitaneous egg and early resting egg production while controlling for environment-genotype interactions. The second comparison was made to reveal lineage-specific differences. Our differential expression, KEGG pathway, and GO term enrichment analysis revealed that early subitaneous egg production is associated with an upregulation of meiosis and cell-cycle genes and genes mapped to sugar and lipid metabolic processes. Downregulated genes were mainly enriched in various metabolic, biosynthesis, and signaling pathways. The arginine metabolic process and sphingolipid metabolism were two enriched pathways of particular interest. Both have previously been shown

to control reproductive mode determination in arthropods (Varki *et al.* 2015). In the brine shrimp, *Artemia franciscana*, sphingolipids play a role in signaling and signal transduction pathways which may be responsible for determining which reproductive mode will ensue (Kojima *et al.* 2013). Additionally, glycosphingolipid biosynthetic enzymes interact with NOTCH signaling receptors, which regulate reproduction in worker honeybees (Duncan et al., 2016; Kraut, 2011). In *Daphnia*, the amino acid arginine has been shown to play a role in reproductive mode switching by suppressing resting egg production (Fink *et al.* 2011, Koch *et al.* 2011). Functional annotation of differentially spliced transcripts further revealed enrichment for the arginine metabolic process, supporting its relevance to reproductive mode initiation.

These findings suggest that the initiation of different parthenogenetic modes in *Daphnia* may be controlled by genes mapped to various metabolic, biosynthesis, and signaling pathways. In future studies, sing-cell RNA sequencing could be utilized to identify the master regulators of these pathways, as well as obtain a clearer picture of the differential expression and splicing patterns of meiosis and cell-cycle genes underlying these two parthenogenetic modes. Additionally, our list of consensus genes can be further investigated utilizing CRISPR/Cas9 in "knock-out" studies to determine their role in the mechanisms of each reproductive mode.

The third chapter of this dissertation focused on developing an ethyl methanesulfonate (EMS) mutagenesis protocol that can be used in forward genetic screens to study gene function. In this study, we showed that exposure to 10mM and 25mM EMS for 4 hours resulted in a base substitution rate of 1.17×10^{-6} and 1.75×10^{-6} per base per generation, respectively (Snyman *et al.* 2021). Additionally, we showed that the base substitution type was dominated by G:C to A:T transitions, the mutations were randomly distributed throughout the genome, and that the first three consecutive broods of the same treated female had an elevated base substitution rate. This

indicates that the primary oocyte nuclei in the ovary of female *Daphnia* are independently mutated. Lastly, we estimated that about 750 F_1s are needed to mutate each gene in the *Daphnia* genome at least once with a 95% probability. To obtain mutations in the homozygous state with 70-80% probability, 4-5 F_2 mutants generated from our forward genetic screening approach must be collected and screened per F_1 mutant line.

For the fourth chapter, I expanded on the EMS project by analyzing the effects of EMS-induced mutations on gene expression and splicing. Our differential expression analysis showed that a median of 3545 genes were differentially expressed, while a median of 51 genes were affected by an EMS-induced mutation per mutant line. This could indicate trans-effects where the mutations affect genes located at other sites within the genome (Curtis *et al.* 2012). The impact of these mutations in differentially expressed genes were as follows; 83% were modifier variants, 10% were moderate impact variants, 6% were low impact variants, and 1% were high impact variants. The high-impact variants consisted of splice acceptor, start-lost, and stop-gained variants. A couple of genes affected by stop-gained variants were downregulated, which could indicate nonsense-mediated mRNA decay (NMD). However, most were upregulated, indicating an escape of NMD, stop-codon read-through, or transcriptional adaptation. Moderate impact variants consisted of missense mutations. Genes affected by missense mutations were up- and downregulated, indicating selection for and against advantages and detrimental genes, respectively (Jia and Zhao 2017). A median of 393 differentially spliced genes were identified per mutant line, with a median of 12 DS genes affected by mutations. Most of these mutations affecting differentially spliced genes were modifier variants (64%) and moderate impact variants (22%). These results indicate that the functional consequences of mutations depend on their

location within the genome, and to gain a clearer understanding, mRNA data should be supplemented with protein expression data.

References

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, *486*(7403), 346–352.

Duncan EJ, Hyink O, Dearden PK (2016) Notch signalling mediates reproductive constraint in the adult worker honeybee. *Nature Communications*, *7*, 12427.

Fink P, Pflitsch C, Marin K (2011) Dietary Essential Amino Acids Affect the Reproduction of the Keystone Herbivore *Daphnia pulex*. *PLOS ONE*, *6*(12), e28498.

Ihara M, Stein P, Schultz RM (2008) UBE2I (UBC9), a SUMO-conjugating enzyme, localizes to nuclear speckles and stimulates transcription in mouse oocytes. *Biology of Reproduction*, *79*(5), 906–913.

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

Koch U, Martin-Creuzburg D, Grossart HP, Straile D (2011) Single dietary amino acids control resting egg production and affect population growth of a key freshwater herbivore. *Oecologia*, *167*(4), 981–989.

Kojima H, Tohsato Y, Kabayama K, Itonori S, Ito M (2013) Biochemical studies on sphingolipids of *Artemia franciscana*: Complex neutral glycosphingolipids. *Glycoconjugate Journal*, *30*(3), 257–268.

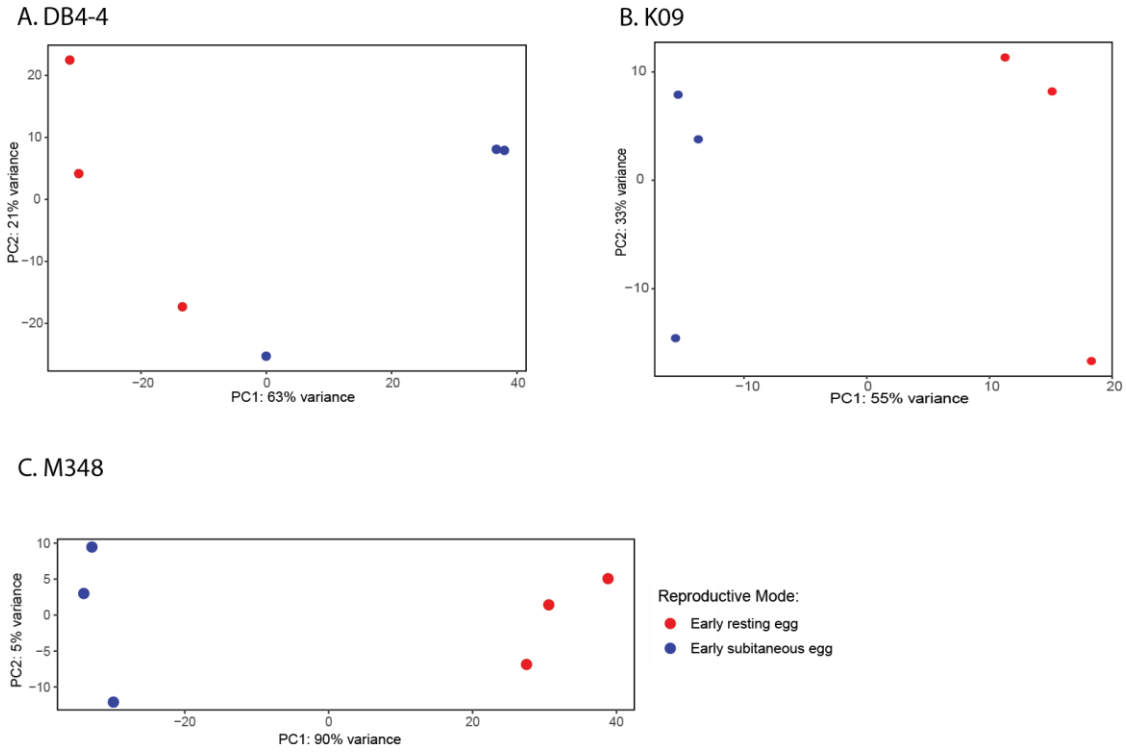
Kraut R (2011) Roles of sphingolipids in *Drosophila* development and disease. *Journal of Neurochemistry*, *116*(5), 764–778.

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.

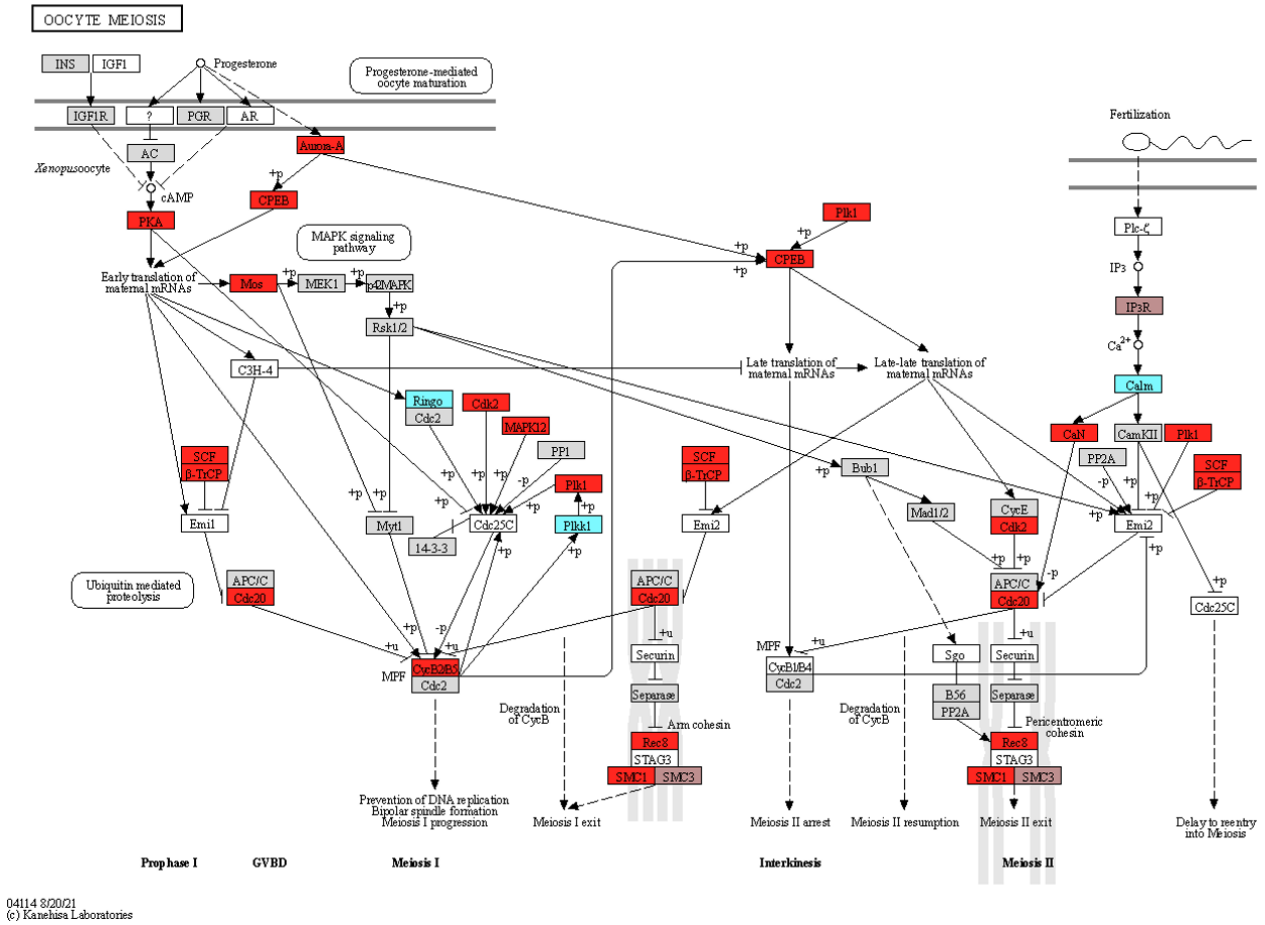
Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M *et al.* (2015) *Essentials of Glycobiology* (3rd ed.). Cold Spring Harbor Laboratory Press.

SUPPLEMENTARY MATERIAL: CHAPTER 2

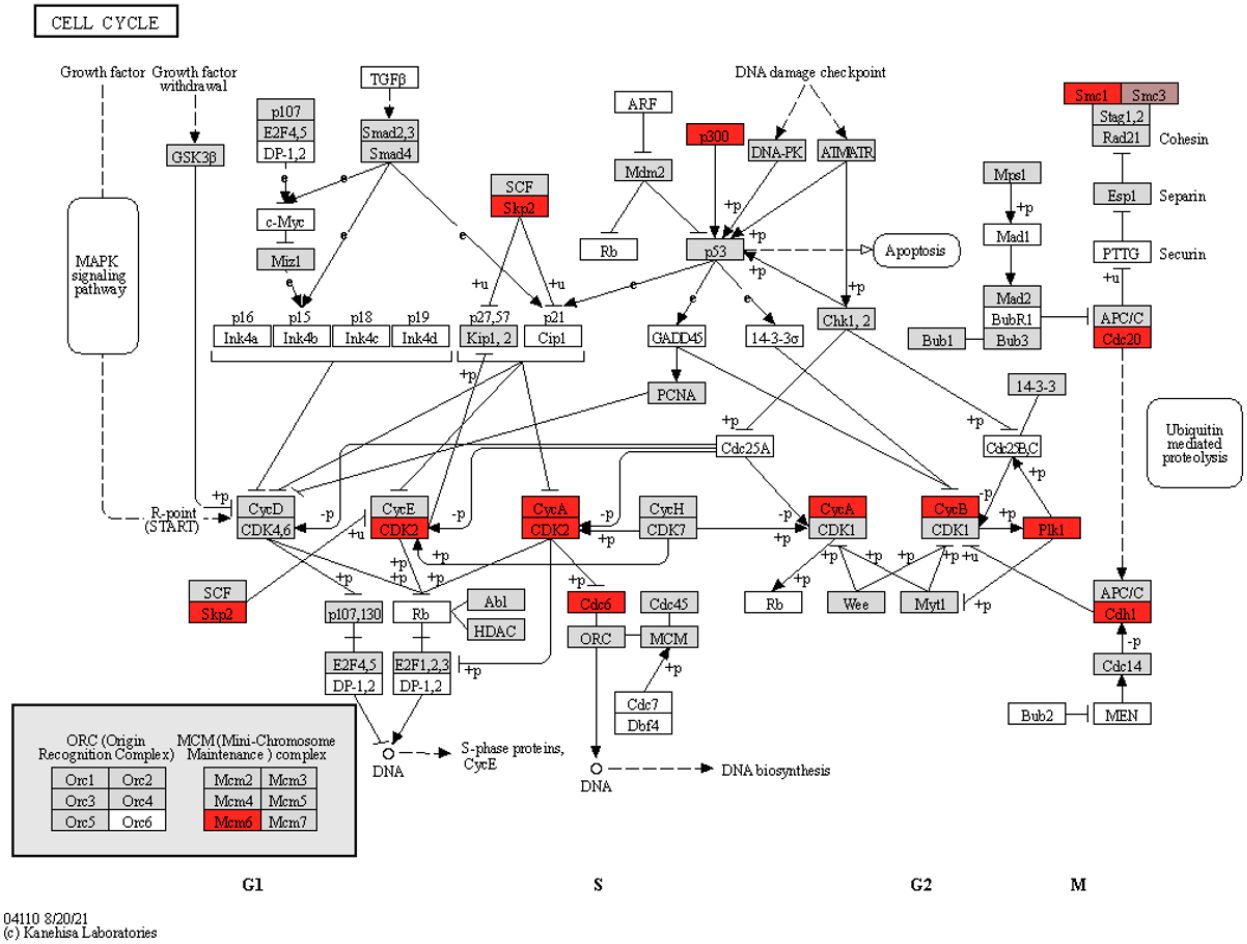
Supplementary Figure S2.1. PCA plot depicting the variance between early subitaneous egg and early resting egg production for all three OP *Daphnia pulex* isolates.



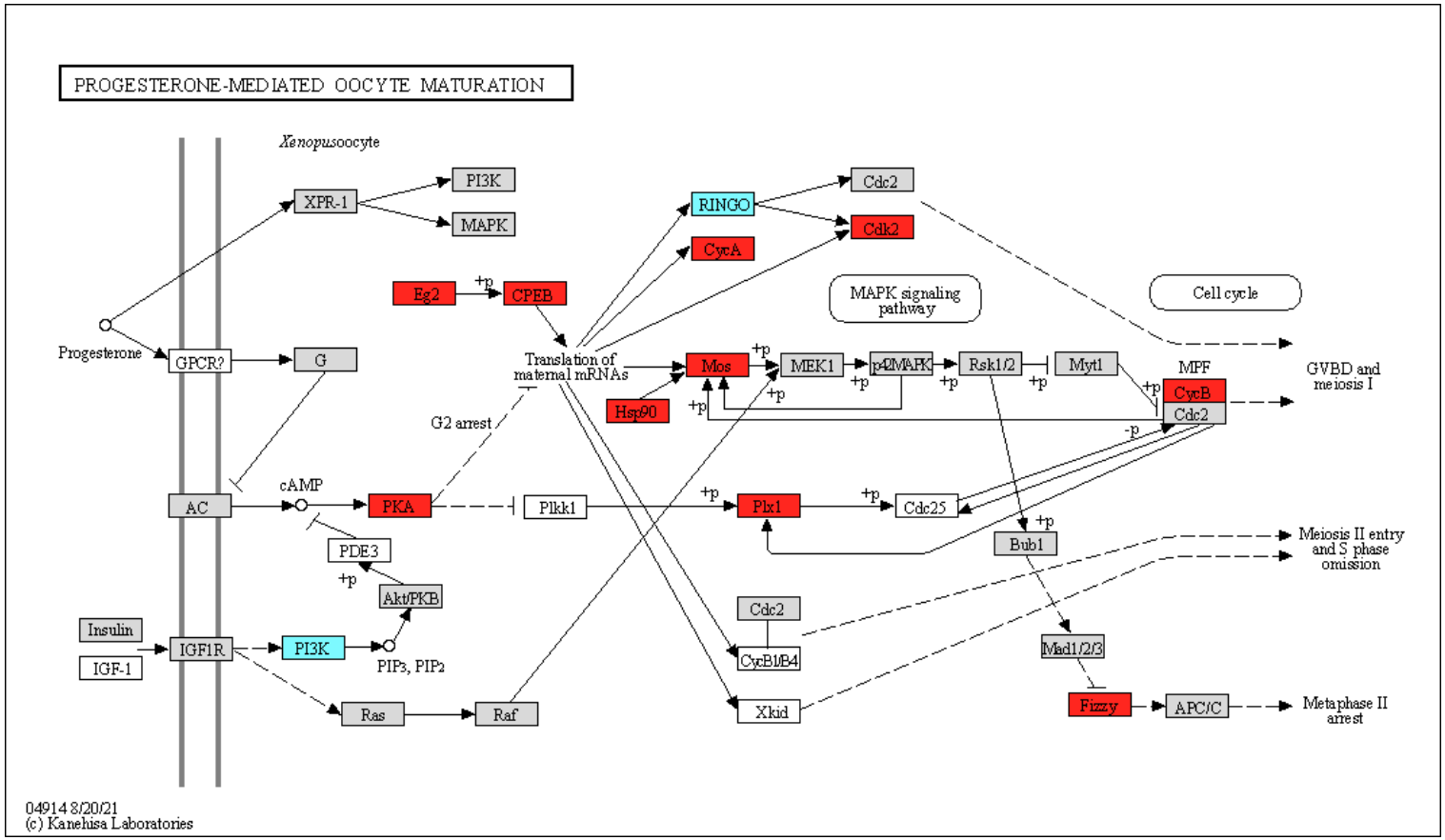
Supplementary Figure S2.2. KEGG pathway enrichment results for the oocyte meiosis pathway using the pooled OP *D. pulex* analysis as input. Genes upregulated in early subitaneous egg production are shown in red, while downregulated genes are shown in blue.



Supplementary Figure S2.3. KEGG pathway enrichment results for the cell-cycle pathway using the pooled OP *D. pulex* analysis as input. Genes upregulated in early subitaneous egg production are shown in red.



Supplementary Figure S2.4. KEGG pathway enrichment results for the progesterone-mediated oocyte maturation pathway using the pooled OP *D. pulex* analysis as input. Genes upregulated in early subitaneous egg production are shown in red, while downregulated genes are shown in blue.



Supplementary Table S2.1. *Daphnia* isolates with collection area.

Species	Isolates	Location
D.pulex (OP)	K09	46°26, -81°3, Kelley Lake, Sudbury, Ontario
	DB4-4	32° 47’17.6” N, 97° 07’ 27.9” W Drying Bed, Arlington, Texas
	Main 348-1	42°99, -76°01, Appledore Island, Maine

Supplementary Table S2.2. Raw read information per replicate and reproductive stage. Samples collected during early subitaneous egg production are labeled as EA, while samples collected during early resting egg production are labeled EM.

Sample name	Total Sequences	Total Alignments	% Aligned Reads	Assigned Alignments	% Assigned Alignments
DB4_EA1	27845395	24815965	89.12053501	19609955	79.00%
DB4_EA2	29435764	27592327	93.73742431	22212242	80.50%
DB4_EA3	26123157	24944712	95.4888875	20056922	80.40%
DB4_EM1	25187162	23723930	94.19056422	18740622	79.00%
DB4_EM2	23528407	20984854	89.18943811	16423446	78.30%
DB4_EM3	20566563	19785077	96.20021099	15694768	79.30%
K09_EA1	25523596	23758018	93.08256564	18819466	79.20%
K09_EA2	36536128	32943316	90.16641282	25607368	77.70%
K09_EA3	22160525	20540182	92.68815608	16128616	78.50%
K09_EM1	34232440	32332770	94.4506731	24549922	75.90%
K09_EM2	37900528	35810972	94.48673644	27946325	78.00%
K09_EM3	28134572	26360685	93.69499205	20131388	76.40%
M348_EA1	25078184	23532340	93.83590136	18963565	80.60%
M348_EA2	21549151	20211120	93.79079482	16277425	80.50%
M348_EA3	23547136	21965139	93.28157361	17523546	79.80%
M348_EM1	21138421	19459647	92.05818637	14468724	74.40%
M348_EM2	23203998	21694488	93.49461244	16291616	75.10%
M348_EM3	24143875	22852688	94.65211363	17376076	76.00%

Supplementary Table S2.3. Differentially expressed (DE) genes in early subitaneous egg production with direction per isolate and pooled analysis.

Isolate	Total DE genes	Direction	Total per Direction
DB4	1115	Upregulated	553
		Downregulated	582
K09	2591	Upregulated	1299
		Downregulated	1292
M348	3942	Upregulated	2125
		Downregulated	1817
Pooled	3263	Upregulated	1771
		Downregulated	1492

Supplementary Table S2.4. KEGG pathways significantly upregulated in early subitaneous egg production for the pooled OP *D. pulex* analysis.

pathway	wht.drawn	wht.in.urn	blk.in.urn	total.draw	p.value	p.adjust
4341 Hedgehog signaling pathway - fly	26	90	6192	576	6.27904028430267e-08	0
4340 Hedgehog signaling pathway	25	94	6188	576	6.4771192493499e-07	1e-04
0980 Metabolism of xenobiotics by cytochrome P450	17	54	6228	576	3.37925446287116e-06	4e-04
0982 Drug metabolism - cytochrome P450	16	51	6231	576	6.907930072752e-06	5e-04
0040 Pentose and glucuronate interconversions	15	46	6236	576	7.86780515332708e-06	5e-04
4974 Protein digestion and absorption	39	207	6075	576	8.46658026184089e-06	5e-04
0500 Starch and sucrose metabolism	14	46	6236	576	3.81206471605243e-05	0.0018
5164 Influenza A	29	146	6136	576	4.50932176972073e-05	0.0018

4976 Bile secretion	18	71	6211	576	4.86636237248086e-05	0.0018
0053 Ascorbate and aldarate metabolism	13	42	6240	576	5.9223539719106e-05	0.002
0983 Drug metabolism - other enzymes	18	73	6209	576	7.23257925641939e-05	0.0022
0830 Retinol metabolism	14	52	6230	576	0.00016748868216794	0.0039
4512 ECM-receptor interaction	18	78	6204	576	0.00018078553737925	0.0039
5204 Chemical carcinogenesis	16	65	6217	576	0.00018476339613695 3	0.0039
4114 Oocyte meiosis	28	150	6132	576	0.00018802729525193 9	0.0039
2010 ABC transporters	11	35	6247	576	0.00018855468572481 6	0.0039
4110 Cell cycle	25	128	6154	576	0.00019807746204694 7	0.0039
4972 Pancreatic secretion	35	207	6075	576	0.00024172394619655	0.0045
0860 Porphyrin and chlorophyll metabolism	12	42	6240	576	0.00026575332819255 8	0.0047
1240 Biosynthesis of cofactors	28	158	6124	576	0.00046012486743408 9	0.0077
0790 Folate biosynthesis	10	33	6249	576	0.00051314488133693 7	0.0082
5146 Amoebiasis	18	91	6191	576	0.00129779315335033	0.0197

0052 Galactose metabolism	12	50	6232	576	0.00147655191393492	0.0214
4914 Progesterone-mediated oocyte maturation	18	95	6187	576	0.0021587295681261	0.03
0480 Glutathione metabolism	16	83	6199	576	0.00311442598430835	0.0416
0140 Steroid hormone biosynthesis	11	48	6234	576	0.00340253101662544	0.0421
0600 Sphingolipid metabolism	11	48	6234	576	0.00340253101662544	0.0421
4111 Cell cycle - yeast	19	108	6174	576	0.0039582896692776	0.0472

Supplementary Table S2.5. KEGG pathways significantly downregulated in early subitaneous egg production for the pooled OP *D. pulex* analysis.

pathway	wht.drawn	wht.in.urn	blk.in.urn	total.draw	p.value	p.adjust
0513 Various types of N-glycan biosynthesis	25	100	6182	449	1.84525706291771e-08	0
1100 Metabolic pathways	152	1491	4791	449	3.01038079929269e-07	1e-04
0601 Glycosphingolipid biosynthesis - lacto and neolacto series	17	69	6213	449	4.52204472130551e-06	5e-04
5202 Transcriptional misregulation in cancer	27	169	6113	449	5.5284548925444e-05	0.0047
0603 Glycosphingolipid biosynthesis - globo and isoglobo series	11	42	6240	449	0.000123527200223239	0.0085
4391 Hippo signaling pathway - fly	15	73	6209	449	0.000157409163045342	0.009
4912 GnRH signaling pathway	19	109	6173	449	0.000222340323313731	0.0109
4020 Calcium signaling pathway	27	192	6090	449	0.000484524928547292	0.0208

4928 Parathyroid hormone synthesis, secretion and action	19	117	6165	449	0.000564510265219281	0.0215
0514 Other types of O-glycan biosynthesis	11	52	6230	449	0.000908713580803324	0.0312
4392 Hippo signaling pathway - multiple species	6	19	6263	449	0.00156764174079492	0.0477
0380 Tryptophan metabolism	9	40	6242	449	0.00166753562728767	0.0477

Supplementary Table S2.6. Significantly enriched GO terms upregulated during early subitaneous egg production for the pooled OP *D. pulex* sample.

GO.ID	Term	Annotated	Significant	Expected	weightFisher	p.adj
GO:0006508	proteolysis	669	80	58.17	5.2e-05	0.070564
GO:0005975	carbohydrate metabolic process	275	38	23.91	0.00014	0.18984
GO:0055085	transmembrane transport	501	61	43.57	0.00129	1
GO:0048477	oogenesis	16	6	1.39	0.00158	1
GO:0006979	response to oxidative stress	69	14	6	0.00209	1
GO:0006869	lipid transport	53	11	4.61	0.00215	1
GO:0005991	trehalose metabolic process	5	3	0.43	0.00572	1
GO:0051493	regulation of cytoskeleton organization	26	3	2.26	0.00758	1
GO:0071840	cellular component organization or bioge...	725	46	63.04	0.0081	1
GO:0006606	protein import into nucleus	16	5	1.39	0.0095	1
GO:0015858	nucleoside transport	6	3	0.52	0.01071	1
GO:0006541	glutamine metabolic process	7	3	0.61	0.01755	1

GO:0042554	superoxide anion generation	7	3	0.61	0.01755	1
GO:0016998	cell wall macromolecule catabolic proces...	32	7	2.78	0.01788	1
GO:1901990	regulation of mitotic cell cycle phase t...	14	5	1.22	0.0212	1
GO:0006030	chitin metabolic process	128	21	11.13	0.02214	1
GO:0006810	transport	1322	139	114.96	0.02326	1
GO:0003333	amino acid transmembrane transport	20	5	1.74	0.02524	1
GO:0006749	glutathione metabolic process	27	6	2.35	0.02569	1
GO:0051258	protein polymerization	36	5	3.13	0.02813	1
GO:0006812	cation transport	288	28	25.04	0.03532	1
GO:0006032	chitin catabolic process	29	6	2.52	0.03558	1
GO:0034637	cellular carbohydrate biosynthetic proce...	9	3	0.78	0.04016	1
GO:0007017	microtubule-based process	153	14	13.3	0.04478	1
GO:0023052	signaling	974	52	84.7	0.04552	1

Supplementary Table S2.7. Significantly enriched GO terms downregulated during early subitaneous egg production for the pooled OP *D. pulex* analysis.

GO.ID	Term	Annotated	Significant	Expected	weightFisher	p.adj
GO:0006486	protein glycosylation	138	30	10.98	9e-07	0.0012213
GO:0007156	homophilic cell adhesion via plasma memb...	30	12	2.39	1.3e-06	0.0017628
GO:0030198	extracellular matrix organization	24	10	1.91	6.6e-06	0.008943
GO:0007155	cell adhesion	114	29	9.07	8.2e-06	0.0111028
GO:0006569	tryptophan catabolic process	5	4	0.4	0.00019	0.25707
GO:0006508	proteolysis	669	70	53.23	3e-04	0.4056
GO:0018401	peptidyl-proline hydroxylation to 4-hydr...	11	5	0.88	0.00096	1
GO:0006032	chitin catabolic process	29	8	2.31	0.00146	1
GO:0006693	prostaglandin metabolic process	8	4	0.64	0.00214	1
GO:0006355	regulation of transcription, DNA-templat...	500	53	39.78	0.00255	1
GO:0016998	cell wall macromolecule	32	8	2.55	0.00289	1

	catabolic proces...					
GO:0007169	transmembrane receptor protein tyrosine ...	33	8	2.63	0.00356	1
GO:0019530	taurine metabolic process	9	4	0.72	0.00361	1
GO:0006691	leukotriene metabolic process	9	4	0.72	0.00361	1
GO:0006525	arginine metabolic process	29	7	2.31	0.00646	1
GO:0060429	epithelium development	11	4	0.88	0.00832	1
GO:0006560	proline metabolic process	30	7	2.39	0.00957	1
GO:0009435	NAD biosynthetic process	7	3	0.56	0.01375	1
GO:0007186	G protein-coupled receptor signaling pat...	263	34	20.92	0.01664	1
GO:0038032	termination of G protein-coupled recepto...	8	3	0.64	0.02072	1
GO:0030154	cell differentiation	51	6	4.06	0.02096	1
GO:0006030	chitin metabolic process	128	22	10.18	0.02227	1

GO:0006555	methionine metabolic process	15	4	1.19	0.02667	1
GO:0006979	response to oxidative stress	69	10	5.49	0.04501	1

Supplementary Table S2.8. Significantly enriched GO terms upregulated during early subitaneous egg production for the DB4-4 isolate.

GO.ID	Term	Annotated	Significant	Expected	weightFisher	p.adj
GO:0051493	regulation of cytoskeleton organization	26	4	0.68	0.00067	0.90048
GO:0006979	response to oxidative stress	69	7	1.81	0.00206	1
GO:0051258	protein polymerization	36	5	0.94	0.00508	1
GO:0005991	trehalose metabolic process	5	2	0.13	0.00648	1
GO:0006810	transport	1322	45	34.63	0.00689	1
GO:0048477	oogenesis	16	3	0.42	0.0077	1
GO:0055085	transmembrane transport	501	21	13.13	0.01242	1
GO:0003333	amino acid transmembrane transport	20	3	0.52	0.01452	1
GO:0051301	cell division	21	3	0.55	0.01662	1
GO:0031667	response to nutrient levels	8	2	0.21	0.01723	1
GO:0033993	response to lipid	39	4	1.02	0.0259	1
GO:0009725	response to hormone	41	4	1.07	0.02591	1
GO:0006955	immune response	23	3	0.6	0.02656	1
GO:0006508	proteolysis	669	25	17.53	0.03	1
GO:0006270	DNA replication initiation	11	2	0.29	0.03214	1
GO:0006749	glutathione metabolic process	27	3	0.71	0.03261	1
GO:0005975	carbohydrate metabolic process	275	14	7.2	0.03276	1
GO:0006801	superoxide metabolic process	19	3	0.5	0.03759	1
GO:0006030	chitin metabolic process	128	7	3.35	0.04472	1

Supplementary Table S2.9. Significantly enriched GO terms downregulated during early subitaneous egg production for the DB4-2 isolate.

GO.ID	Term	Annotated	Significant	Expected	weightFisher	p.adj
GO:0006486	protein glycosylation	138	19	4.37	5.9e-07	0.00079296
GO:0007156	homophilic cell adhesion via plasma memb...	30	6	0.95	3e-04	0.4029
GO:0006693	prostaglandin metabolic process	8	3	0.25	0.0016	1
GO:0019530	taurine metabolic process	9	3	0.29	0.0023	1
GO:0006691	leukotriene metabolic process	9	3	0.29	0.0023	1
GO:0009395	phospholipid catabolic process	33	5	1.05	0.0035	1
GO:0006979	response to oxidative stress	69	7	2.19	0.0059	1
GO:0006629	lipid metabolic process	260	16	8.24	0.0069	1
GO:0007275	multicellular organism development	106	6	3.36	0.0123	1
GO:0006493	protein O-linked glycosylation	7	2	0.22	0.0189	1
GO:0038032	termination of G protein-coupled recepto...	8	2	0.25	0.0247	1
GO:0006508	proteolysis	669	30	21.19	0.031	1
GO:0006030	chitin metabolic process	128	9	4.05	0.0362	1
GO:0046339	diacylglycerol metabolic process	10	2	0.32	0.038	1
GO:0030198	extracellular matrix organization	24	3	0.76	0.0389	1
GO:0006807	nitrogen compound metabolic process	3369	112	106.71	0.0392	1

GO:0018401	peptidyl-proline hydroxylation to 4-hydr...	11	2	0.35	0.0455	1
GO:0060429	epithelium development	11	2	0.35	0.0455	1

Supplementary Table S2.10. Significantly enriched GO terms upregulated during early subitaneous egg production for the K09 isolate.

GO.ID	Term	Annotated	Significant	Expected	weightFisher	p.adj
GO:0006270	DNA replication initiation	11	7	0.75	1.7e-06	0.0022848
GO:0006260	DNA replication	76	23	5.19	2.4e-06	0.0032232
GO:0006508	proteolysis	669	72	45.65	4.5e-06	0.006039
GO:0006869	lipid transport	53	11	3.62	0.00032	0.42912
GO:0006606	protein import into nucleus	16	6	1.09	0.00043	0.5762
GO:0006541	glutamine metabolic process	7	4	0.48	0.00063	0.84357
GO:0006338	chromatin remodeling	41	11	2.8	0.00169	1
GO:0006406	mRNA export from nucleus	5	3	0.34	0.00285	1
GO:0051493	regulation of cytoskeleton organization	26	3	1.77	0.00466	1
GO:0071840	cellular component organization or bioge...	670	44	45.72	0.00473	1
GO:0007093	mitotic cell cycle checkpoint signaling	7	3	0.48	0.00898	1
GO:1901991	negative regulation of mitotic cell cycl...	7	3	0.48	0.00898	1
GO:0006334	nucleosome assembly	13	4	0.89	0.00932	1
GO:0051258	protein polymerization	36	5	2.46	0.01241	1
GO:1901990	regulation of mitotic cell cycle phase t...	14	6	0.96	0.01313	1
GO:2000112	regulation of cellular macromolecule bio...	556	39	37.94	0.01335	1
GO:0030071	regulation of mitotic metaphase/anaphase...	8	3	0.55	0.01365	1
GO:0051252	regulation of RNA metabolic process	553	44	37.73	0.01454	1

GO:0048477	oogenesis	16	4	1.09	0.02017	1
GO:0007017	microtubule-based process	153	13	10.44	0.02021	1
GO:0034637	cellular carbohydrate biosynthetic proce...	9	3	0.61	0.02535	1
GO:0000209	protein polyubiquitination	10	3	0.68	0.0264	1
GO:0000079	regulation of cyclin-dependent protein s...	10	3	0.68	0.0264	1
GO:0023052	signaling	984	47	67.14	0.02816	1
GO:0072488	ammonium transmembrane transport	5	2	0.34	0.04047	1
GO:0006207	'de novo' pyrimidine nucleobase biosynth...	5	2	0.34	0.04047	1
GO:0005991	trehalose metabolic process	5	2	0.34	0.04047	1
GO:0006979	response to oxidative stress	69	9	4.71	0.04364	1

Supplementary Table S2.11. Significantly enriched GO terms downregulated during early subitaneous egg production for the K09 isolate.

GO.ID	Term	Annotated	Significant	Expected	weightFisher	p.adj
GO:0006508	proteolysis	669	85	48.22	5.5e-09	7.392e-06
GO:0007156	homophilic cell adhesion via plasma memb...	30	13	2.16	4.7e-08	6.3121e-05
GO:0005975	carbohydrate metabolic process	275	31	19.82	0.00052	0.69784
GO:0006032	chitin catabolic process	29	8	2.09	0.00076	1
GO:0008152	metabolic process	4286	325	308.95	0.00153	1
GO:0016998	cell wall macromolecule catabolic proces...	32	8	2.31	0.00154	1
GO:0009395	phospholipid catabolic process	33	8	2.38	0.00191	1
GO:0006569	tryptophan catabolic process	5	3	0.36	0.00334	1
GO:0018401	peptidyl-proline hydroxylation to 4-hydr...	11	4	0.79	0.00585	1
GO:0006486	protein glycosylation	138	19	9.95	0.01219	1
GO:0048731	system development	62	5	4.47	0.01489	1
GO:0048519	negative regulation of biological proces...	130	5	9.37	0.01519	1
GO:0006633	fatty acid biosynthetic process	23	5	1.66	0.02168	1
GO:0006979	response to oxidative stress	69	10	4.97	0.02506	1
GO:0030198	extracellular matrix organization	24	5	1.73	0.02581	1
GO:0098656	anion transmembrane transport	18	4	1.3	0.03042	1

GO:0007275	multicellular organism development	106	12	7.64	0.03605	1
GO:0072359	circulatory system development	5	2	0.36	0.04481	1
GO:0006689	ganglioside catabolic process	5	2	0.36	0.04481	1

Supplementary Table S2.12. Significantly enriched GO terms upregulated during early subitaneous egg production for the M348 isolate.

GO.ID	Term	Annotated	Significant	Expected	weightFisher	p.adj
GO:0005975	carbohydrate metabolic process	275	70	33.87	4e-11	5.376e-08
GO:0006810	transport	1322	194	162.8	5.6e-08	7.5208e-05
GO:0055085	transmembrane transport	501	86	61.7	3.2e-07	0.00042944
GO:0006508	proteolysis	669	113	82.39	6.9e-07	0.00092529
GO:0006979	response to oxidative stress	69	24	8.5	1e-06	0.00134
GO:0006635	fatty acid beta-oxidation	5	5	0.62	2.8e-05	0.037492
GO:0006869	lipid transport	53	17	6.53	5.7e-05	0.076266
GO:0006030	chitin metabolic process	128	33	15.76	1e-04	0.1337
GO:0006633	fatty acid biosynthetic process	23	9	2.83	0.001	1
GO:0006637	acyl-CoA metabolic process	13	5	1.6	0.001	1
GO:0030206	chondroitin sulfate biosynthetic process	5	4	0.62	0.001	1
GO:0008152	metabolic process	4286	537	527.82	0.0021	1
GO:0006040	amino sugar metabolic process	140	40	17.24	0.0037	1
GO:0006629	lipid metabolic process	260	56	32.02	0.0046	1
GO:0006027	glycosaminoglycan catabolic process	11	5	1.35	0.0068	1
GO:0001575	globoside metabolic process	8	4	0.99	0.0106	1
GO:0016998	cell wall macromolecule catabolic proces...	32	9	3.94	0.0126	1
GO:1901071	glucosamine-containing compound metaboli...	130	35	16.01	0.0145	1
GO:0044264	cellular polysaccharide metabolic proces...	20	5	2.46	0.0151	1

GO:0005991	trehalose metabolic process	5	3	0.62	0.0154	1
GO:0071840	cellular component organization or bioge...	670	39	82.51	0.017	1
GO:0007015	actin filament organization	34	5	4.19	0.0262	1
GO:0042886	amide transport	6	3	0.74	0.0279	1
GO:0015858	nucleoside transport	6	3	0.74	0.0279	1
GO:0018401	peptidyl-proline hydroxylation to 4-hydr...	11	4	1.35	0.037	1
GO:0048477	oogenesis	16	5	1.97	0.0382	1
GO:0006812	cation transport	289	35	35.59	0.0414	1
GO:0009395	phospholipid catabolic process	33	8	4.06	0.0428	1
GO:0006687	glycosphingolipid metabolic process	30	11	3.69	0.0479	1

Supplementary Table S2.13. Significantly enriched GO terms downregulated during early subitaneous egg production for the M348 isolate.

GO.ID	Term	Annotated	Significant	Expected	weightFisher	p.adj
GO:0042254	ribosome biogenesis	223	73	22.41	5.6e-24	7.5264e-21
GO:0006412	translation	235	69	23.62	9.6e-19	1.28928e-15
GO:0006555	methionine metabolic process	15	8	1.51	3.4e-05	0.045628
GO:0006730	one-carbon metabolic process	31	11	3.12	0.00013	0.17433
GO:0030198	extracellular matrix organization	24	9	2.41	0.00032	0.4288
GO:0007156	homophilic cell adhesion via plasma memb...	30	10	3.02	0.00046	0.61594
GO:0006569	tryptophan catabolic process	5	4	0.5	0.00047	0.62886
GO:0000097	sulfur amino acid biosynthetic process	5	4	0.5	0.00047	0.62886
GO:0006486	protein glycosylation	138	28	13.87	0.00098	1
GO:0006518	peptide metabolic process	272	78	27.34	0.00299	1
GO:0006807	nitrogen compound metabolic process	3369	371	338.6	0.00405	1
GO:0007601	visual perception	17	6	1.71	0.00473	1
GO:0007155	cell adhesion	114	24	11.46	0.00479	1
GO:0006123	mitochondrial electron transport, cytoch...	8	4	0.8	0.00509	1
GO:0042157	lipoprotein metabolic process	29	5	2.91	0.00867	1
GO:0006030	chitin metabolic process	128	21	12.86	0.00899	1
GO:0009084	glutamine family amino acid biosynthetic...	12	4	1.21	0.016	1

GO:0018401	peptidyl-proline hydroxylation to 4-hydr...	11	4	1.11	0.01876	1
GO:0042537	benzene-containing compound metabolic pr...	11	4	1.11	0.01876	1
GO:0006525	arginine metabolic process	29	7	2.91	0.0219	1
GO:0035235	ionotropic glutamate receptor signaling ...	17	5	1.71	0.02243	1
GO:0006493	protein O-linked glycosylation	7	3	0.7	0.02597	1
GO:0044093	positive regulation of molecular functio...	63	6	6.33	0.02624	1
GO:0006560	proline metabolic process	30	7	3.02	0.02796	1
GO:0006414	translational elongation	18	4	1.81	0.0282	1
GO:0006805	xenobiotic metabolic process	17	3	1.71	0.02827	1
GO:0007602	phototransduction	18	5	1.81	0.02856	1
GO:0035249	synaptic transmission, glutamatergic	18	5	1.81	0.02856	1
GO:0007218	neuropeptide signaling pathway	13	4	1.31	0.03457	1
GO:0015671	oxygen transport	8	3	0.8	0.0385	1
GO:0006693	prostaglandin metabolic process	8	3	0.8	0.0385	1
GO:0007186	G protein-coupled receptor signaling pat...	259	36	26.03	0.03968	1

Supplementary Table S2.14. The number of differentially spliced events between early subitaneous egg and early resting egg production per isolate and pooled sample.

Comparison	A3SS	A5SS	RI	SE	MXE	Total
DB4-4	104	92	54	191	47	488
M348	84	90	139	193	74	580
K09	115	108	168	179	56	626
Pooled	42	60	56	166	58	382
Total (Percentage)	345 (16.6)	350 (16.9)	417 (20.1)	729 (35.1)	235 (11.3)	2076

SUPPLEMENTARY MATERIAL: CHAPTER 3

Table S3.1. *Daphnia* isolates used in this study and their sampling locations.

Species	Isolates	Location
D. pulex (OP)	DB4-1	32° 47’17.6” N, 97° 07’ 27.9” W Drying Bed, Arlington, Texas
	DB4-2	32° 47’17.6” N, 97° 07’ 27.9” W Drying Bed, Arlington, Texas
	DB4-4	32° 47’17.6” N, 97° 07’ 27.9” W Drying Bed, Arlington, Texas
D. pulex (CP)	Tex21	42°12, -83°12, Textile Road, Michigan
	Povi4	42°45, -85°21, Battle Creek, Michigan
	SW4	Illinois
D. pulicaria (CP)	RLSD26	44°57, -96°49, Round Lake, South Dakota
	AroMoose	44°50, -69°16, Sebasticook Lake, Maine
	Warner5	42°8, -85°3, Warner Lake, Michigan

Table S3.2. Average survival rate of three *Daphnia pulex* clones after exposure to four EMS concentrations. Numbers in brackets represent standard deviation.

Isolate	10mM	25mM	50mM	100mM
DB4-1	100.0%	53.3% (±0.6)	0%	0%
DB4-2	100.0%	70.0% (±1.0)	0%	0%
DB4-4	100.0%	56.7% (±0.6)	0%	0%

Table S3.3. Summary statistics of germline base substitution mutations in 10mM and 25mM EMS mutant lines. All mutation rates are per generation rates.

EMS	Isolate	No. of mutations	TsTv ratio	Per base mutation rate	GC to AT transitions (%)	No. of genic mutations (non-synonymous mutations)	Mutation rate per gene	Non-syn mutation rate per gene
10mM	AroMoose-rep1	63	14.75	$6.19x10^{-7}$	93.65	31 (14)	$1.5x10^{-3}$	$6.8x10^{-4}$
	AroMoose-rep2	82	40	$8.16x10^{-7}$	97.56	39 (14)	$1.9x10^{-3}$	$6.8x10^{-4}$
	DB4-1-rep1	126	4.73	$1.32x10^{-6}$	74.6	50 (21)	$2.5x10^{-3}$	$1.0x10^{-3}$
	DB4-1-rep2	125	6.35	$1.33x10^{-6}$	78.4	46 (21)	$2.3x10^{-3}$	$1.0x10^{-3}$
	Povi4-rep1	164	12.67	$1.62x10^{-6}$	90.85	77 (31)	$4.7x10^{-3}$	$1.9x10^{-3}$
	Povi4-rep2	133	15.62	$1.34x10^{-6}$	91.73	65 (32)	$3.9x10^{-3}$	$1.9x10^{-3}$
	RLSD26-rep1	128	4.12	$1.31x10^{-6}$	75	47 (19)	$2.3x10^{-3}$	$9.2x10^{-4}$
	RLSD26-rep2	93	4.81	$9.85x10^{-7}$	80.65	41 (20)	$2.0x10^{-3}$	$9.8x10^{-4}$
	Tex21-rep1	91	12	$9.3x10^{-7}$	91.21	34 (21)	$2.1x10^{-3}$	$1.3x10^{-3}$
	Tex21-rep2	106	25.5	$1.07x10^{-6}$	95.28	56 (28)	$4.8x10^{-3}$	$2.4x10^{-3}$
	Warner5-rep1	88	7	$8.57x10^{-7}$	86.36	36 (9)	$1.7x10^{-3}$	$4.3x10^{-4}$
	Warner5-rep2	107	14.29	$1.05x10^{-6}$	89.72	43 (21)	$2.1x10^{-3}$	$1.0x10^{-3}$
25mM	AroMoose-rep1	242	16.29	$2.39x10^{-6}$	92.98	112 (58)	$5.4x10^{-3}$	$2.8x10^{-3}$
	AroMoose-rep2	105	51.5	$1.03x10^{-6}$	98.1	54 (26)	$2.6x10^{-3}$	$1.3x10^{-3}$
	DB4-2-rep1	184	5.34	$1.89x10^{-6}$	75.54	76 (41)	$3.7x10^{-3}$	$2.0x10^{-3}$

DB4-2-rep2	123	5.15	$1.28x10^{-6}$	78.86	54 (30)	$2.6x10^{-3}$	$1.5x10^{-3}$
Povi4-rep1	312	13.86	$3.07x10^{-6}$	92.31	116 (47)	$7.0x10^{-3}$	$2.8x10^{-3}$
Povi4-rep2	299	8.65	$2.88x10^{-6}$	85.28	109 (48)	$6.5x10^{-3}$	$2.9x10^{-3}$
RLSD26-rep1	175	7.33	$1.9x10^{-6}$	83.43	85 (35)	$4.2x10^{-3}$	$1.74x10^{-3}$
RLSD26-rep2	153	6.29	$1.56x10^{-6}$	83.00	66 (36)	$3.2x10^{-3}$	$1.75x10^{-3}$
Tex21-rep1	93	14.5	$9.55x10^{-7}$	93.55	46 (22)	$2.8x10^{-3}$	$1.3x10^{-3}$
Tex21-rep2	90	17	$9.22x10^{-7}$	94.44	41 (17)	$2.5x10^{-3}$	$1.0x10^{-3}$
Warner5-rep1	247	8.5	$2.37x10^{-6}$	87.85	110 (43)	$5.3x10^{-3}$	$2.1x10^{-3}$
Warner5-rep2	283	7.58	$2.67x10^{-6}$	85.87	121 (57)	$5.8x10^{-3}$	$2.7x10^{-3}$
SW4-rep1	121	5.37	$1.2x10^{-6}$	78.51	57 (32)	$3.5x10^{-3}$	$1.9x10^{-3}$
SW4-rep2	112	4.89	$1.1x10^{-6}$	78.57	37 (16)	$2.2x10^{-3}$	$9.6x10^{-4}$

Table S3.4. Summary statistics of germline base substitution mutations in first-, second-, and third-brood mutant lines. All mutation rates are per generation rates.

EMS	Isolate (Species)	Brood	No. of mutations	TsTv ratio	Per base mutation rate	GC to AT transitions (%)	No. of genic mutations (no. of non-syn mutations)	Mutation rate per gene	Non-syn mutation rate per gene
10mM	AroMoose (*D. pulicaria*)	1	71	13.2	7.0×10^{-7}	91.55	25(9)	1.2×10^{-3}	4.3×10^{-4}
		2	41	19.5	4.0×10^{-7}	95.12	17(6)	8.2×10^{-4}	2.9×10^{-4}
		3	163	39.75	1.58×10^{-6}	96.32	91(40)	4.4×10^{-3}	1.9×10^{-3}
	DB4-4	1	60	9	6.28×10^{-7}	90	31(13)	1.5×10^{-3}	6.4×10^{-4}
	(OP *D. pulex*)	2	54	9.8	5.74×10^{-7}	90.74	26(12)	1.3×10^{-3}	5.9×10^{-4}
		3	67	8.57	7.0×10^{-7}	83.58	29(14)	1.4×10^{-3}	6.9×10^{-4}
	Tex21	1	63	30.5	6.5×10^{-7}	96.83	32(20)	1.9×10^{-3}	1.2×10^{-3}
	(CP *D. pulex*)	2	70	100% Transition	7.2×10^{-7}	98.57	24(14)	1.5×10^{-3}	8.5×10^{-4}
		3	68	100% Transition	7.0×10^{-7}	98.53	30(14)	1.8×10^{-3}	8.5×10^{-4}
25mM	AroMoose (*D. pulicaria*)	1	145	15.11	1.45×10^{-6}	92.41	80(34)	3.9×10^{-3}	1.7×10^{-3}

	2	574	29.21	$5.77x10^{-6}$	96.52	277(140)	$1.3x10^{-2}$	$6.8x10^{-3}$
	3	411	20.63	$4.24x10^{-6}$	94.4	228(91)	$1.1x10^{-3}$	$4.5x10^{-3}$
DB4-4	1	319	10.81	$3.22x10^{-6}$	89.66	161(80)	$7.8x10^{-3}$	$3.9x10^{-3}$
(OP *D. pulex*)	2	175	20.88	$1.73x10^{-6}$	93.14	84(42)	$4.0x10^{-3}$	$2.0x10^{-3}$
	3	157	25.17	$1.55x10^{-6}$	95.54	69(36)	$3.3x10^{-3}$	$1.7x10^{-3}$
Tex21	1	74	4.69	$9.16x10^{-7}$	68.92	30(13)	$1.8x10^{-3}$	$8.0x10^{-4}$
(CP *D. pulex*)	2				Died			
	3	482	23.1	$7.67x10^{-6}$	95.44	255(108)	$1.5x10^{-2}$	$6.5x10^{-3}$

Table S3.5. Proportions of different types of EMS-induced base substitutions in each species treated with 10mM and 25mM EMS.

Base substitution	**OP *D. pulex***	**OP *D. pulex***	**CP *D. pulex***	**CP *D. pulex***	***D. pulicaria***	***D. pulicaria***
	10mM	**25mM**	**10mM**	**25mM**	**10mM**	**25mM**
A to C	0.008	0.023	0.005	0.010	0.012	0.002
A to G	0.040	0.033	0.010	0.010	0.006	0.011
A to T	0.020	0.003	0.017	0.020	0.018	0.015
C to A	0.020	0.039	0.008	0.027	0.025	0.028
C to G	0.016	0.003	0.005	0.002	0.002	0.003
C to T	0.339	0.384	0.473	0.454	0.420	0.412
G to A	0.426	0.384	0.450	0.418	0.451	0.470
G to C	0.016	0.033	0.002	0.007	0.005	0.008
G to T	0.020	0.029	0.009	0.025	0.024	0.019
T to A	0.032	0.010	0.012	0.011	0.012	0.014
T to C	0.040	0.039	0.005	0.013	0.016	0.012
T to G	0.024	0.020	0.003	0.003	0.009	0.007

Table S3.6. Average proportions of different types of EMS-induced base substitutions across all first-brood (BR1), second-brood (BR2), and third-brood (BR3) mutant lines at 10mM and 25mM concentration.

Base Substitution	10mM BR1	25mM BR1	10mM BR2	25mM BR2	10mM BR3	25mM BR3
A to C	0.005	0.007	0.006	0.000	0.003	0.000
A to G	0.000	0.019	0.000	0.000	0.020	0.003
A to T	0.005	0.017	0.006	0.005	0.000	0.016
C to A	0.015	0.028	0.006	0.004	0.010	0.007
C to G	0.005	0.002	0.000	0.000	0.007	0.000
C to T	0.433	0.385	0.515	0.486	0.470	0.464
G to A	0.495	0.491	0.436	0.471	0.470	0.487
G to C	0.010	0.004	0.006	0.003	0.000	0.002
G to T	0.021	0.020	0.012	0.012	0.007	0.006
T to A	0.005	0.009	0.006	0.012	0.007	0.010
T to C	0.005	0.015	0.006	0.007	0.003	0.004
T to G	0.000	0.004	0.000	0.000	0.003	0.002

Table S3.7. The average proportions of EMS-induced based substitutions affecting different genomic regions and the average proportions of mutations with different amino acid changing effects in each species at 10mM and 25mM EMS concentration**.**

Regions	**OP *D. pulex***	**OP *D. pulex***	**CP *D. pulex***	**CP *D. pulex***	***D. pulicaria***	***D. pulicaria***
	10mM	**25mM**	**10mM**	**25mM**	**10mM**	**25mM**
Downstream	0.329	0.303	0.266	0.263	0.312	0.293
Exon	0.129	0.131	0.139	0.124	0.106	0.120
Intergenic	0.196	0.180	0.237	0.262	0.194	0.219
Intron	0.043	0.055	0.059	0.050	0.046	0.047
Splice site region	0.009	0.003	0.005	0.007	0.010	0.006
Upstream	0.268	0.295	0.277	0.279	0.305	0.290
UTR 3'	0.009	0.027	0.012	0.013	0.015	0.013
UTR 5'	0.017	0.006	0.005	0.003	0.013	0.013
Amino acid changing effects						
Missense	0.533	0.658	0.737	0.665	0.650	0.694
Nonsense	0.033	0.094	0.025	0.050	0.036	0.046
Silent	0.433	0.248	0.238	0.285	0.314	0.260

Table S3.8. Average proportions of EMS-induced based substitutions affecting different genomic regions and the average proportions of mutations with different amino acid changing effects across all first-brood (BR1), second-brood (BR2), and third-brood (BR3) mutant lines at 10mM and 25mM EMS concentration.

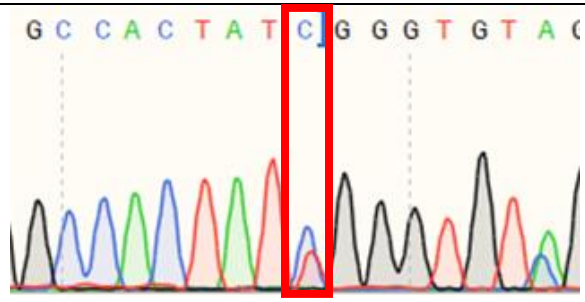
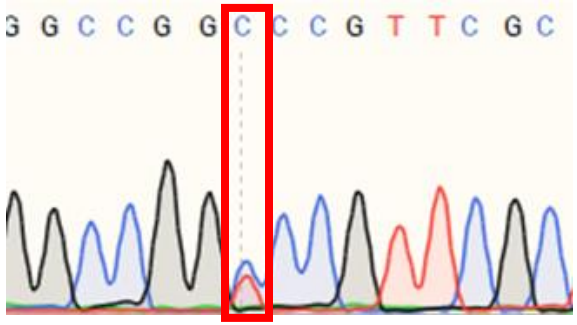
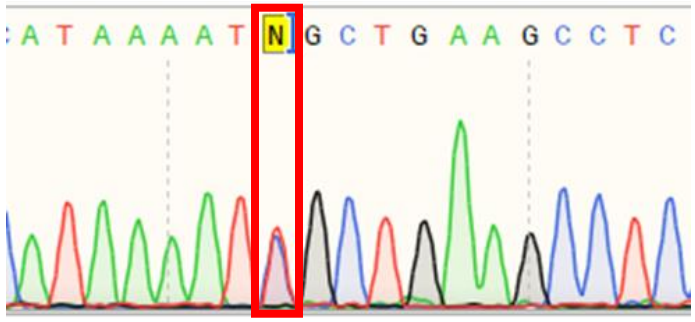
Regions	**10mM BR1**	**25mM BR1**	**10mM BR2**	**25mM BR2**	**10mM BR3**	**25mM BR3**
Downstream	0.284	0.295	0.274	0.301	0.258	0.293
Exon	0.130	0.144	0.125	0.143	0.171	0.131
Intergenic	0.214	0.180	0.258	0.179	0.188	0.178
Intron	0.034	0.052	0.055	0.040	0.046	0.063
Splice site region	0.006	0.007	0.003	0.002	0.008	0.004
Upstream	0.302	0.293	0.256	0.300	0.302	0.299
UTR 3'	0.014	0.021	0.008	0.018	0.017	0.018
UTR 5'	0.016	0.009	0.021	0.017	0.010	0.013
Amino acid changing effects						
Missense	0.585	0.626	0.667	0.674	0.565	0.701
Nonsense	0.092	0.066	0.021	0.030	0.031	0.019
Silent	0.323	0.308	0.313	0.296	0.405	0.279

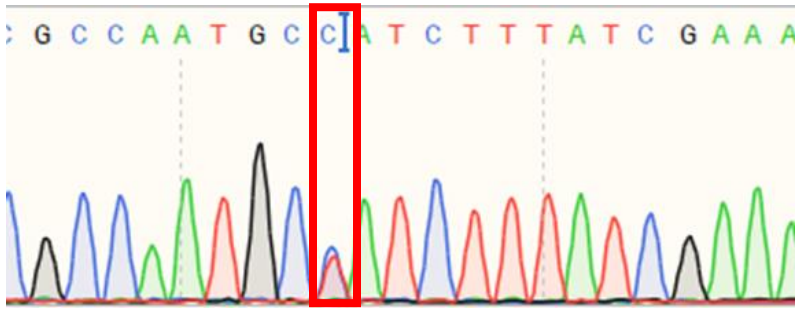
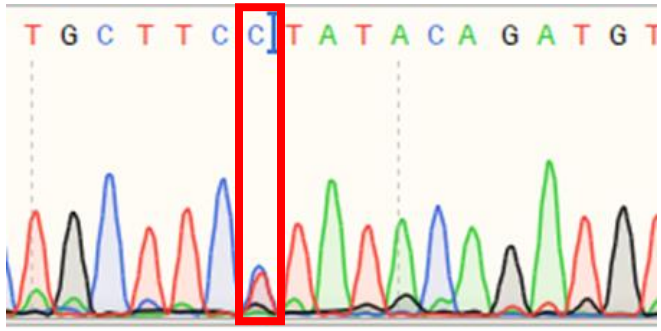
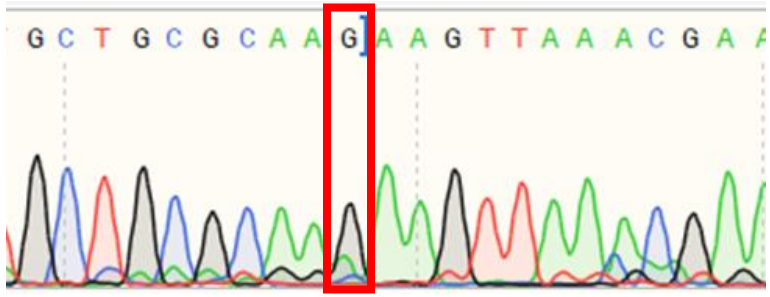
Table S3.9. Significantly (Bonferroni corrected chi squared p-value < 0.05) over-and underrepresented trinucleotides identified from 10mM and 25mM EMS treatment lines.

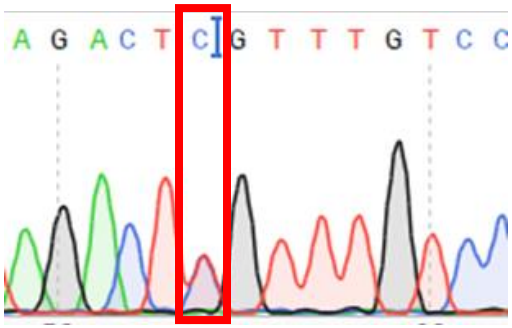
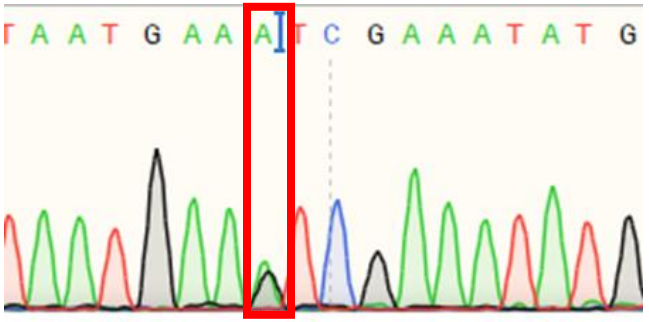
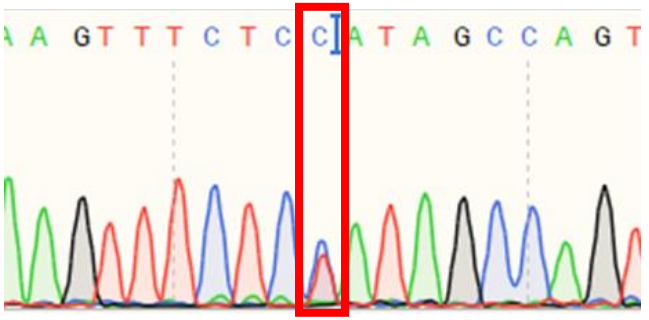
Trinucleotide	**Observed**	**Expected (*D. pulex*/ *D. pulicaria*)**	**Over-represented/ Under-represented**
NAN	-	-	Under-represented
NTN	-	-	Under-represented
TCG	136	44, 47	Over-represented
ACG	108	40, 46	Over-represented
CCG	104	32, 34	Over-represented
TCC	134	48, 51	Over-represented
CCT	120	37, 39	Over-represented
TCT	128	65, 69	Over-represented
GCC	103	35, 40	Over-represented
ACC	139	36, 39	Over-represented
CCC	175	32, 34	Over-represented
GCG	82	29, 32	Over-represented

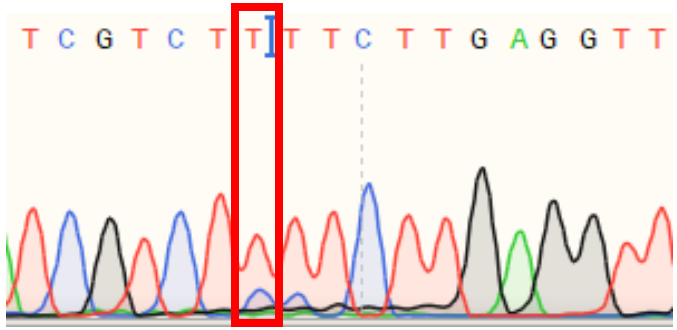
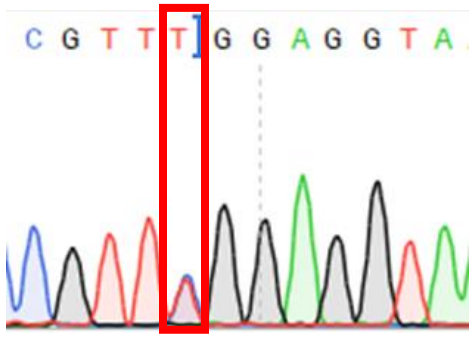
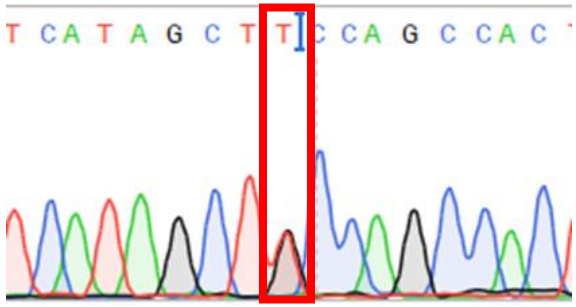
GGT	154	36, 39	Over-represented
AGG	146	37, 39	Over-represented
CGG	121	32, 34	Over-represented
CGA	106	44, 47	Over-represented
GGA	170	48, 51	Over-represented
GGC	107	35, 38	Over-represented
CGC	108	29, 34	Over-represented
GGG	188	32, 34	Over-represented

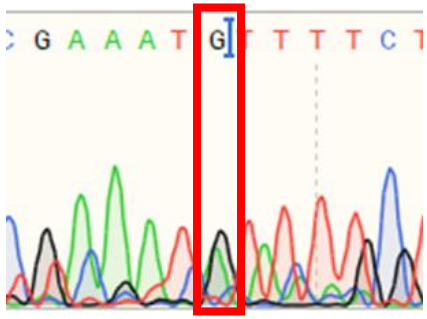
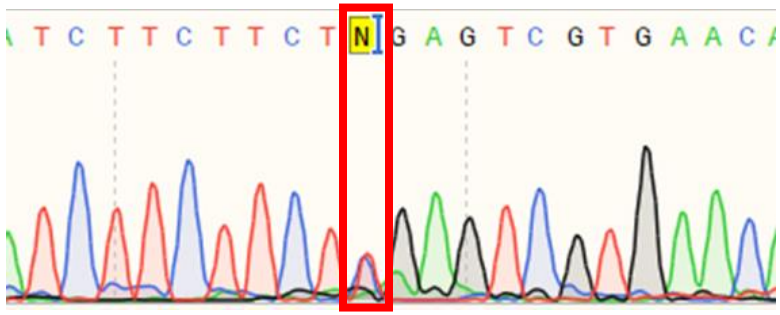
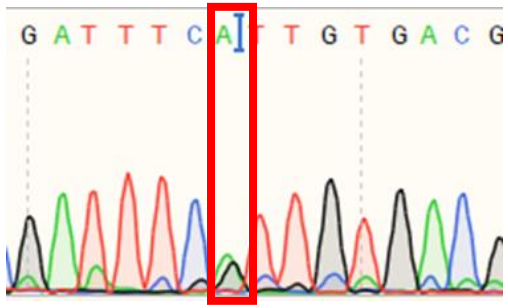
Table S3.10. Sanger sequencing validation for 20 mutations identified through whole-genome sequencing.

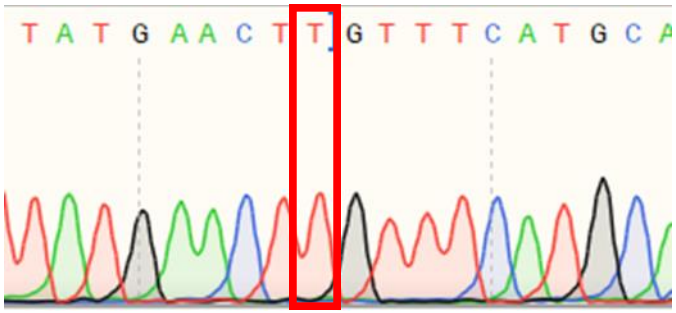
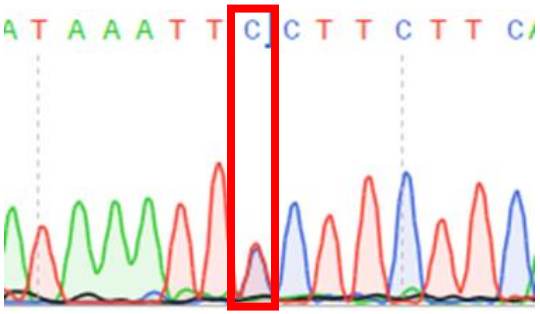
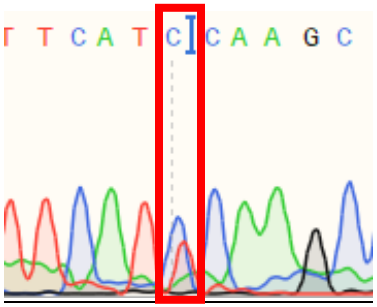
Sample	Position	Mutation type	Sanger results
DB_10_1	2:7110525	CC (depth of 53 reads) to CT (37 C and 17 T reads).	
	9:8600466	CC (depth of 35 reads) to CT (20 C and 17 T reads).	
	7:10050871	CC (depth of 41 reads) to CT (13 C and 9 T reads).	

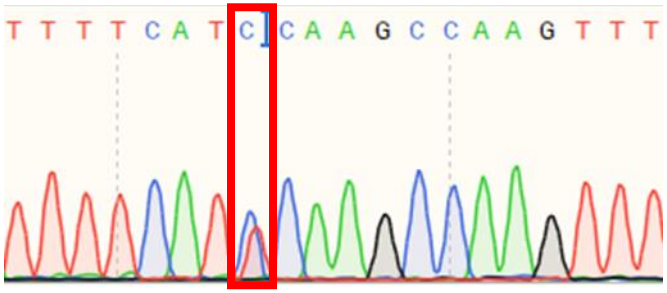
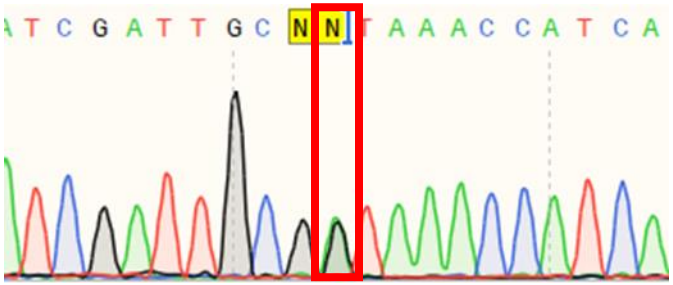
	10:9881179	CC (depth of 30 reads) to CT (13 C and 16T reads).	
	11:7864040	CC (depth of 29 reads) to CT (14 C and 11 T reads).	
Sample: DB_25_BR1	1:2914097	GG (depth of 45 reads) to GA (22A and 28G reads).	

9:3519253	CC (depth of 36 reads) to CT (19 C and 19 T reads).	
1:4871668	GG (depth of 20 reads) to GA (13 A and 10 G reads).	
3:6796285	CC (depth of 36 reads) to CT (16 C and 17 T reads).	

	5:3202467	CC (depth of 24 reads) to CT (20 C and 18 T reads).	
Sample: DB_25_BR2	4:248038	CC (depth of 16 reads) to CT (22 C and 25 T reads).	
	9:2561038	GG (depth of 38 reads) to GT (18 G and 20 T reads).	

1:8386876	GG (depth of 25 reads) to GA (depth of 23 G and 10 A reads).	
1:15272778	CC (depth of 25 reads) to CT (19 C and 18 T reads).	
7:4988557	GG (depth of 31 reads) to AG (17 A and 16 G reads).	

	3:12091607	CC (depth of 20 reads) to CT (depth of 6C and 10T reads). Sanger sequencing results appears to be homozygous TT.	
Sample: DB_25_BR3	6:789538	CC (depth of 50 reads) to CT (14 C and 20 T reads).	
	3:9954972	CC (depth of 55 reads) to CT (26 C and 15 T reads).	

6:1942968	CC (depth of 29 reads) to CT (14 C and 20 T reads).	
8:6072067	GG (depth of 20 reads) to GA (18 A and 19 G reads).	

SUPPLEMENTARY MATERIAL: CHAPTER 4

Supplementary Table S4.1. Whole-genome sequencing data obtained, mapping rate and depth per mutant line.

Mutant line	**Total Sequences**	**Mapped reads**	**Mapping rate**	**Average Depth**
P3-1	27876025	24108745	86.49	20.7
P3-2	39453960	28728608	72.82	24.4
PP-1	38439774	33372636	86.82	28.5
PP-2	38320557	33296967	86.89	28.3
RB-1	38401662	31664358	82.46	26.88
RB-2	33778759	28540895	84.49	24.34
RY-1	39329907	31304410	79.59	26.67
RY-2	38492505	32728183	85.02	28.07
G-1	31372511	27797569	88.6	23.4
G-2	32425011	27524999	84.89	22.8
P-1	33458597	28904143	86.39	24.4
PY3-1	31732028	27582657	86.92	23.2
PY3-2	32518996	27936904	85.91	23.51
R-1	31786328	28539697	89.79	24
R-2	31278659	28378273	90.73	24.12
Y-1	32457899	29329067	90.36	24.6
Average	**34445198.63**	**29358631.94**	**85.510625**	**24.868125**

Supplementary Table S4.2. RNA sequencing data obtained, trimming and alignment information per mutant line.

Mutant line	Total sequences	After trimming	% Retained	Total Alignments	% Alignments	Assigned Alignments	% Assigned Alignments
BR-1-1	14479195	14478436	99.994758	12951968	89.45695516	9325578	72
BR-1-2	16932167	16930930	99.99269438	16192746	95.64002686	11680561	72.1
BR-1-3	14189028	14188059	99.99317078	13275258	93.56641384	9447465	71.2
BR-2-1	16220129	16219008	99.99308883	15434890	95.16543798	11051189	71.6
BR-2-2	19217575	19216143	99.99254849	18366658	95.57931579	13209020	71.9
BR-2-3	17846654	17845552	99.99382517	16977986	95.13847484	12156042	71.6
G-1-1	24459488	24455295	99.98285737	23487807	96.04385063	16511894	70.3
G-1-2	30819608	30812005	99.97533064	29367458	95.31173969	20571186	70
G-1-3	30651687	30635229	99.94630638	28050037	91.5613753	19576672	69.8
G-2-1	36943663	36911220	99.9121825	33825831	91.64105386	23925693	70.7
G-2-2	24740613	24706683	99.86285708	21476386	86.92541204	14940137	69.6
G-2-3	24025269	23963266	99.74192589	22500249	93.89475124	15601966	69.3
P-1-1	21769002	21750579	99.91537049	20632572	94.85987476	14150118	68.6
P-1-2	20165358	20158104	99.96402742	19121744	94.85884188	13227292	69.2
P-1-3	20324587	20318905	99.97204371	19314091	95.05478273	13519362	70
P3-1-1	12237329	12236352	99.99201623	11791275	96.36266593	8434981	71.5
P3-1-2	15468266	15467375	99.99423982	14837156	95.92549479	10607410	71.5
P3-1-3	22413807	22411779	99.99095201	21086238	94.08551637	15053244	71.4
P3-2-1	18298659	18297673	99.99461163	17773108	97.13316005	12531327	70.5
P3-2-2	19651101	19649799	99.99337442	19064869	97.02322655	13572644	71.2
P3-2-3	21319219	21317714	99.99294064	20597702	96.62247087	14623764	71
PP-1-1	14945627	14944124	99.98994355	13438416	89.92441444	9443583	70.3
PP-1-2	15480770	15479392	99.99109863	14803742	95.63516448	10471623	70.7
PP-1-3	18436550	18434983	99.99150058	17471538	94.77382214	12380363	70.9
PP-2-1	22502559	22495319	99.96782588	20866653	92.7599782	14669481	70.3
PP-2-2	18596333	18594022	99.98757282	17746164	95.44015813	12557463	70.8

PP-2-3	26090199	26084679	99.97884263	23932493	91.74923333	16810075	70.2
R-1-1	20824328	20815102	99.95569605	20232898	97.20297311	13929949	68.8
R-1-2	22603348	22594946	99.96282852	22100482	97.81161681	15280094	69.1
R-1-3	24283518	24271625	99.95102439	24520442	101.0251353	16882371	68.9
R-2-1	23931170	23926138	99.97897303	23176052	96.86499342	16098785	69.5
R-2-2	19131280	19123768	99.96073446	18298671	95.68548939	12637576	69.1
R-2-3	22231762	22215201	99.92550748	21068811	94.83961455	14392456	68.3
Y-1-1	18655231	18649148	99.96739252	17866136	95.80135243	12696850	71.1
Y-1-2	15526481	15523321	99.97964767	14936343	96.21873438	10733888	71.9
Y-1-3	15725993	15720326	99.96396412	15134673	96.27454927	10781626	71.2
PY3-1-1	16353705	16350553	99.98072608	15576268	95.26447209	11094820	71.2
PY3-1-2	21144915	21141396	99.9833577	19886973	94.06650819	14296820	71.9
PY3-1-3	16455933	16452396	99.97850623	15633173	95.02064623	11165956	71.4
PY3-2-1	20896695	20891799	99.97657046	19905749	95.28020541	13991977	70.3
PY3-2-2	17672154	17668746	99.98071542	16841622	95.31871702	11828711	70.2
PY3-2-3	16284381	16281382	99.98158358	15526272	95.36212589	10853768	69.9
RY-1-1	17819218	17818332	99.99502784	16296216	91.45758425	11501911	70.6
RY-1-2	24091306	24090139	99.99515593	22043782	91.50541639	15560335	70.6
RY-1-3	14803027	14802171	99.9942174	13461606	90.94345688	9480941	70.4
RY-2-1	22069817	22068348	99.99334385	20957309	94.96546366	14947649	71.3
RY-2-2	26069363	26067705	99.99364004	24788003	95.09085284	17739412	71.6
RY-2-3	15774528	15773551	99.99380647	14943889	94.74016979	10679539	71.5
AR-1	25962592	25956036	99.97474828	24346962	93.80077143	17128549	70.4
AR-2	12452945	12449443	99.97187814	12001705	96.40354994	8521557	71
AR-3	19289167	19285232	99.97959995	18548829	96.18151858	13158109	70.9
Average	**20162299.98**	**20155675.08**	**99.97091278**	**19068194.14**	**94.69136332**	**13439878.08**	**70.5745098**

Supplementary Table S4.3. Summary statistics of the EMS-induced mutations per mutant line. All mutation rates are per one generation.

Mutant line	No. of EMS-induced mutations	Mutation rate	No. of genes with mutations	Per gene mutation rate	No. of genes affected by EMS-induced mutations	No. of DE genes	No. of DE genes affected by EMS-induced mutations	No. of non-synonymous mutations	Non-synonymous mutation rate
P3-1	141	1.71×10^{-6}	69	0.0029	269	4990	76	7	0.00029
P3-2	154	1.87×10^{-6}	80	0.0033	297	6280	100	10	0.00042
PP-1	134	1.62×10^{-6}	52	0.0022	272	1176	5	0	0
PP-2	132	1.60×10^{-6}	49	0.0020	276	3562	37	5	0.00021
RB-1	133	1.61×10^{-6}	72	0.0030	292	3484	53	7	0.000291
RB-3	129	1.56×10^{-6}	69	0.0029	284	6470	87	10	0.00042
RY-1	598	7.24×10^{-6}	319	0.013	1275	6444	393	49	0.00204
RY-3	606	7.34×10^{-6}	320	0.013	1289	6606	423	56	0.00234
G-1	195	2.36×10^{-6}	91	0.0038	412	2347	36	3	0.00012
G-2	188	2.28×10^{-6}	84	0.0035	398	4504	76	9	0.00037
P-1	177	2.14×10^{-6}	87	0.0036	375	3527	48	6	0.00025
PY3-1	135	1.63×10^{-6}	63	0.0026	287	1750	24	2	8.32×10^{-5}
PY3-2	135	1.63×10^{-6}	60	0.0025	284	6084	74	6	0.00025
R-1	164	1.99×10^{-6}	72	0.0030	309	3046	44	7	0.00029
R-2	172	2.08×10^{-6}	75	0.0031	334	2830	43	6	0.00025
Y-1	189	2.29×10^{-6}	89	0.0037	368	1937	37	1	4.16×10^{-5}

Supplementary Table S4.4. The number of different types of base substitutions per mutant line.

Base Substitution	G-1	G-2	P3-1	P3-2	PP-1	PP-2	P-1	PY3-1	PY3-2	RB-1	RB-2	R-1	R-2	RY-1	RY-2	Y-1
A>C:	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0
A>G:	1	1	2	1	0	0	0	1	1	1	1	1	2	3	2	0
A>T:	4	6	1	1	0	0	0	0	0	2	2	0	0	9	10	1
C>A:	5	6	5	6	3	4	5	1	2	4	5	2	2	8	6	5
C>G:	3	3	3	3	2	3	0	1	3	0	0	1	1	0	0	1
C>T:	83	80	68	78	49	48	94	51	53	61	60	65	66	282	287	87
G>A:	81	82	65	63	74	71	76	79	75	55	55	94	100	301	304	89
G>C:	0	0	1	1	0	0	0	1	0	0	0	0	0	1	1	2
G>T:	9	7	4	7	4	4	6	4	4	12	10	5	4	9	8	5
T>A:	8	5	0	1	1	2	2	0	0	0	0	1	1	9	11	4
T>C:	2	2	0	0	2	1	0	2	2	0	0	1	1	4	6	3
T>G:	3	1	0	1	0	0	2	0	0	0	0	0	0	0	0	0
Ts/Tv ratio:	**5.06**	**5.69**	**9.64**	**7.1**	**11.36**	**8.57**	**11.33**	**19**	**14.56**	**6.5**	**6.82**	**17.89**	**21.12**	**16.39**	**16.19**	**9.94**
Total mutations:	**190**	**182**	**133**	**146**	**132**	**130**	**169**	**130**	**130**	**131**	**125**	**158**	**167**	**570**	**576**	**181**

Supplementary Table S4.5. The number of effects found per region across all mutant lines.

Effect Region	PP-1	PP-2	P3-1	P3-2	RB-1	RB-2	RY-1	RY-2	G-1	G-2	P-1	PY3-1	PY3-2	R-1	R-2	Y-1
Downstream	116	122	123	128	119	116	536	543	192	176	168	130	131	134	151	162
Exon	36	36	41	45	44	41	224	225	55	50	61	39	33	42	46	50
Intergenic	83	83	71	74	67	66	281	289	104	107	92	75	78	91	95	106
Intron	17	13	25	30	19	19	98	97	40	32	26	21	24	27	28	22
Splice site region	1	1	7	7	5	5	8	8	3	2	1	2	0	1	1	3
Upstream	123	124	103	117	126	124	589	593	186	193	171	120	121	136	151	165
UTR 3 prime	8	7	8	12	3	3	23	24	6	4	7	3	3	10	9	14
UTR 5 prime	4	4	4	4	9	9	14	13	10	10	7	6	9	15	15	14

Supplementary Table S4.6. Impact of mutations affecting differentially expressed genes per mutant line.

Mutant line	Direction of regulation	High	Low	Moderate	Modifier
G-1	Upregulation	0	0	1	18
	Downregulation	0	3	2	22
G-2	Upregulation	0	3	3	30
	Downregulation	2	1	6	46
P3-1	Upregulation	2	7	3	37
	Downregulation	0	4	4	33
P3-2	Upregulation	2	12	9	53
	Downregulation	0	2	1	46
PP-1	Downregulation	0	0	0	5
PP-2	Upregulation	0	2	1	21
	Downregulation	0	1	4	25
P-1	Upregulation	0	5	5	33
	Downregulation	0	0	1	19
PY3-1	Upregulation	0	0	0	9
	Downregulation	0	1	2	16
PY3-2	Upregulation	2	3	5	35
	Downregulation	0	0	1	38
RB-1	Upregulation	0	1	5	35
	Downregulation	0	0	2	17
RB-2	Upregulation	0	2	7	44
	Downregulation	0	3	3	37
R-1	Upregulation	0	0	5	23
	Downregulation	1	3	2	26
R-2	Upregulation	0	0	2	29
	Downregulation	1	2	4	26
RY-1	Upregulation	5	25	37	184

	Downregulation	1	7	12	197
RY-2	Upregulation	7	21	45	197
	Downregulation	1	5	11	208
Y-1	Upregulation	3	0	1	16
	Downregulation	0	0	0	28
Total		**27**	**113**	**184**	**1553**

Supplementary Table S4.7. Effect of mutations impacting differentially expressed genes.

Mutant line	Direction of Regulation	3' UTR	5' UTR premature start codon gain	5' UTR	Down-stream	Intron	Mis-sense	Start-lost	Splice acceptor	Splice region	Stop gained	Synonymous	Up-stream
G-1	Up	0	0	0	4	0	1	0	0	0	0	0	14
	Down	2	0	0	10	1	2	0	0	0	0	3	9
G-2	Up	2	0	1	9	1	3	0	0	0	0	3	17
	Down	1	0	1	13	6	6	0	0	0	2	1	25
P3-1	Up	1	0	3	13	6	3	1	0	2	1	6	14
	Down	0	0	0	15	5	4	0	0	1	0	4	13
P3-2	Up	2	0	3	18	8	9	1	0	3	1	10	23
	Down	0	0	0	15	7	1	0	0	1	0	2	24
PP-1	Down	0	0	0	1	0	0	0	0	0	0	0	4
PP-2	Up	0	0	0	10	0	1	0	0	0	0	2	11
	Down	1	0	2	11	3	4	0	0	0	0	1	8
P-1	Up	4	0	0	17	4	5	0	0	0	0	5	8
	Down	0	0	0	7	1	1	0	0	0	0	0	11
PY3-1	Up	0	0	0	6	0	0	0	0	0	0	0	3
	Down	0	0	1	9	0	2	0	0	0	0	1	6
PY3-2	Up	1	0	1	16	4	5	0	0	0	2	3	13
	Down	0	0	1	20	1	1	0	0	0	0	0	16
RB-1	Up	0	0	0	22	1	5	0	0	0	0	1	12
	Down	0	0	0	9	2	2	0	0	0	0	0	6
RB-2	Up	0	0	2	19	7	7	0	0	0	0	2	16
	Down	0	1	1	22	1	3	0	0	0	0	2	13
R-1	Up	6	0	1	4	1	5	0	0	0	0	0	11
	Down	1	0	2	10	1	2	0	0	1	1	2	13
R-2	Up	6	0	0	9	0	2	0	0	0	0	0	14

	Down	1	0	3	7	2	4	0	0	3	1	1	14
RY-1	Up	4	1	3	70	31	37	0	0	2	5	22	78
	Down	2	0	0	90	8	12	0	1	0	0	7	98
RY-2	Up	5	1	1	73	29	45	0	0	1	7	19	90
	Down	2	0	0	96	9	11	0	1	0	0	5	102
Y-1	Up	2	0	2	5	1	1	0	0	0	3	0	6
	Down	0	0	2	20	0	0	0	0	0	0	0	6
Total		**43**	**3**	**30**	**650**	**140**	**184**	**2**	**2**	**14**	**23**	**102**	**698**

Supplementary Table S4.8. Differentially spliced events detected across all 16 EMS mutant lines.

Mutant line	No. of DS genes	A3SS	A5SS	MXE	RI	SE
P3-1	381	45	31	88	99	118
P3-2	415	37	33	82	104	159
PP-1	282	26	39	52	83	82
PP-2	363	34	45	66	100	118
RB-1	212	17	16	30	69	80
RB-2	404	45	48	78	126	107
RY-1	522	48	58	83	175	158
RY-2	593	55	59	98	166	215
G-1	415	38	38	59	188	92
G-2	627	54	69	73	250	181
P-1	419	50	33	50	180	106
PY3-1	241	36	24	45	71	65
PY3-2	436	55	61	59	130	131
R-1	334	46	31	71	84	102
R-2	349	43	38	62	123	83
Y-1	295	36	42	55	79	83
Total	**6288**	**665**	**665**	**1051**	**2027**	**1880**

Supplementary Table S4.9. Differentially spliced events affected by an EMS-induced mutation per mutant line.

Mutant line	No. of DS genes	A3SS	A5SS	MXE	RI	SE	DS genes also DE
P3-1	17	3	0	3	0	11	3
P3-2	8	0	0	0	0	8	1
PP-1	11	1	0	4	2	4	0
PP-2	2	0	0	0	2	0	2
RB-1	10	0	1	3	2	4	4
RB-2	23	2	2	12	2	5	4
RY-1	42	5	7	1	11	18	11
RY-2	57	5	6	4	10	32	12
G-1	16	0	1	6	6	3	0
G-2	16	0	2	5	3	6	2
P-1	12	4	0	0	7	1	2
PY3-1	3	0	1	0	0	2	1
PY3-2	4	0	2	0	2	0	1
R-1	10	0	0	4	4	2	0
R-2	15	0	0	6	3	6	1
Y-1	3	0	0	0	1	2	0
Total	**249**	**20**	**22**	**48**	**55**	**104**	**44**

Supplementary Table S4.10. Impact of mutations affecting differentially spliced genes per mutant line.

Mutant line	High Impact	Low impact	Moderate impact	Modifier
G_1	0	5	0	10
G_2	0	1	0	12
P3_1	0	1	0	6
P3_2	0	0	1	3
PP_1	0	5	6	0
PP_2	0	0	7	0
P_1	0	0	20	3
PY3_1	2	0	0	3
PY3_2	0	0	1	5
BR_1	0	0	2	6
BR_2	0	0	1	9
R_1	0	0	0	7
R_2	0	0	0	9
RY_1	2	2	2	29
RY_2	2	4	5	25
Y_1	0	1	0	2
Total	**6**	**19**	**45**	**129**

Supplementary Table S4.11. Effect of mutations impacting differentially spliced genes.

Mutant line	3' UTR	Downstream	Intron	Missense	Splice region	Splice acceptor	Splice donor	Stop-gained	Synonymous variant	Upstream
G-1	0	2	2	0	0	0	0	0	5	6
G-2	0	3	2	0	0	0	0	0	1	7
P3-1	0	1	1	0	0	0	0	0	1	4
P3-2	0	0	0	1	0	0	0	0	0	3
PP-1	0	0	0	5	0	0	0	0	0	6
PP-2	0	0	0	0	0	0	0	0	0	7
P-1	0	5	1	3	0	0	0	0	0	11
PY3-1	0	3	0	0	0	0	0	2	0	0
PY3-2	0	5	0	1	0	0	0	0	0	0
BR-1	0	1	1	2	0	0	0	0	0	4
BR-2	0	1	1	1	0	0	0	0	0	7
R-1	0	3	1	0	0	0	0	0	0	3
R-2	0	1	2	0	0	0	0	0	0	6
RY-1	0	7	6	2	1	1	1	0	1	19
RY-2	1	7	5	5	0	0	1	1	4	13
Y-1	0	1	0	0	0	0	0	0	1	1
Total	**1**	**40**	**22**	**20**	**1**	**1**	**2**	**3**	**13**	**97**