

Genome-Wide Analysis of Cadmium-Induced, Germline Mutations in a Long-Term *Daphnia pulex* Mutation-Accumulation Experiment

Nathan Keith,¹ Craig E. Jackson,¹ Stephen P. Glaholt,¹ Kimberly Young,² Michael Lynch,³ and Joseph R. Shaw¹

¹O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, USA

²Department of Biology, Indiana University, Bloomington, Indiana, USA

³Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA

BACKGROUND: Germline mutations provide the raw material for all evolutionary processes and contribute to the occurrence of spontaneous human diseases and disorders. Yet despite the daily interaction of humans and other organisms with an increasing number of chemicals that are potentially mutagenic, precise measurements of chemically induced changes to the genome-wide rate and spectrum of germline mutation are lacking.

OBJECTIVES: A large-scale *Daphnia pulex* mutation-accumulation experiment was propagated in the presence and absence of an environmentally relevant cadmium concentration to quantify the influence of cadmium on germline mutation rates and spectra.

RESULTS: Cadmium exposure dramatically changed the genome-wide rates and regional spectra of germline mutations. In comparison with those in control conditions, *Daphnia* exposed to cadmium had a higher overall A:T → G:C mutation rates and a lower overall C:G → G:C mutation rate. *Daphnia* exposed to cadmium had a higher intergenic mutation rate and a lower exonic mutation rate. The higher intergenic mutation rate under cadmium exposure was the result of an elevated intergenic A:T → G:C rate, whereas the lower exon mutation rate in cadmium was the result of a complete loss of exonic C:G → G:C mutations—mutations that are known to be enriched at 5-hydroxymethylcytosine. We experimentally show that cadmium exposure significantly reduced 5-hydroxymethylcytosine levels.

DISCUSSION: These results provide evidence that cadmium changes regional mutation rates and can influence regional rates by interfering with an epigenetic process in the *Daphnia pulex* germline. We further suggest these observed cadmium-induced changes to the *Daphnia* germline mutation rate may be explained by cadmium's inhibition of zinc-containing domains. The cadmium-induced changes to germline mutation rates and spectra we report provide a comprehensive view of the mutagenic perils of cadmium and give insight into its potential impact on human population health. <https://doi.org/10.1289/EHP8932>

Introduction

Chemical stressors can induce DNA mutations (Ames et al. 1973), thereby changing the rate of disease occurrence in human populations and shaping genetic diversity within ecosystems. Chemical induced changes to the rate of DNA mutation are therefore problematic because the rate of chemical pollution has increased dramatically since the beginning of widespread industrialization, especially within the past 70 y (Landrigan et al. 2018). During this period more than 140,000 novel chemicals have been synthesized and introduced into the market (Landrigan et al. 2018). Yet despite the increased synthesis, development, and use of chemicals, and mounting evidence of their extreme effects on public health (Landrigan et al. 2018), precise measurements of chemicals altering the rate and genome-wide distribution of germline mutations are lacking.

Understanding the influence of chemical pollutants on genome-wide patterns of germline mutation is needed to more accurately predict the rate of disease and disorder occurrence in human populations. The analysis of chemical stressor-induced genome-wide mutation signatures in tumors has been important for linking chemicals to their perturbation of specific polymerases and DNA repair pathways, genome-wide (Alexandrov et al. 2016; Supek and Lehner 2017). However, these studies are limited in their ability to provide

understanding of spontaneous mutation rates in the germline that are critical for determining inheritance, penetrance, and ultimately population-level impact, because mutations observed in cancer genomes occur in the presence of disease- and tumor-specific evolutionary processes in the soma. Additionally, extrapolating mutational observations from somatic genomes is not ideal, due to inherent physiological and functional genomic differences between somatic and germline cellular environments (Kimmins and Sassone-Corsi 2005).

We completed a long-term mutation-accumulation (MA) experiment with *Daphnia pulex* to investigate the effects of chronic, continuous cadmium (Cd) stress on the rate and spectrum of germline mutations. MA experiments provide an opportunity to directly measure the chemical influences on the rate and spectrum of germline mutation. These experiments remove natural selection via strict genetic bottlenecks after organismal reproduction each generation (Halligan and Keightley 2009; Keith et al. 2016; Sung et al. 2012a) and therefore preserve all germline mutations except the extreme minority that cause immediate lethality or sterility. When propagated for thousands of generations and subjected to whole-genome sequencing, MA experiments provide precise measurements of germline mutational rates and spectra.

As an established regulatory toxicity test species (Shaw et al. 2008), *D. pulex* serves as a tractable model organism for MA experiments designed to measure mutational processes in the absence and presence of a constant chemical exposure. The ability to propagate *D. pulex* clonally in MA experiments allows mutations to accumulate in a diploid, naturally heterozygous genome, thus avoiding the complementation of recessive lethal mutations that occurs in MA experiments that use sibling mating each generation (See “Introduction” in Keith et al. 2016). Because uncharacterized chemicals are potentially highly mutagenic, which would therefore increase the occurrence of recessive lethal mutations and lead to extinction/loss of MA experiments, the ability of *D. pulex* MA experiments to avoid complementation of recessive lethal mutations also makes it an ideal eukaryotic model for measuring the impact of toxicants on germline mutation rate (Chain et al. 2019; Keith et al. 2016).

Address correspondence to Nathan Keith, 717 Potter St., Berkeley, CA 94710 USA. Email: rnkeith@lbl.gov; Joseph R. Shaw, 702 N. Walnut Grove St., Bloomington, IN 47405 USA. Email: joeshaw@iu.edu

Supplemental Material is available online (<https://doi.org/10.1289/EHP8932>).
The authors declare they have no actual or potential competing financial interests.

Received 8 January 2021; Revised 15 September 2021; Accepted 21 September 2021; Published 8 October 2021.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

A previous *D. pulex* MA study demonstrated that *Daphnia*'s directly measured germline mutation rates and spectra of mutation classes are similar to those observed in the human germline (Keith et al. 2016). Further, the well-annotated *D. pulex* genome also enables the identification of the genomic mutation contexts and elements where mutations arise (Colbourne et al. 2011; Ye et al. 2017). Characterization of the enrichment of *de novo* mutations at a specific mutation context(s) is of utmost importance for understanding the rate of occurrence of human disease in toxicant exposure because many disease- and disorder-linked toxicants interfere with specific DNA repair-associated pathways and polymerases (Jin et al. 2003; Shin et al. 2019). The utility of *Daphnia* MA experiments for studying all classes of mutations along with the genome-wide patterns of mutational contexts makes it an ideal model organism for understanding how toxicant exposure influences germline mutational processes in the human germline.

Cd was selected as a model MA chemical stressor because it is a recognized chemical of concern by the World Health Organization (WHO) for its effects on human health (IARC 1993), and Cd is a major by-product of fossil fuel combustion and mining of nonferrous ores (Hutton and Symon 1986), making it prevalent in the environment (IARC 1993). Cd is mutagenic, yet it does not directly react with DNA (IARC 1993), and it interferes with multiple DNA repair pathways (Filipic et al. 2006; Hartmann and Speit 1994; Jin et al. 2003; Lynn et al. 1997). Due to the long biological half-life of Cd, it accumulates in tissues and therefore readily moves through food webs (Guan and Wang 2006). Cd induces oxidative stress and can replace zinc (Zn) in Zn-containing protein domains, which can change protein structure and function (Li and Manning 1955).

This study directly compared results from an MA experiment propagated in the absence and presence of Cd (Figure 1), providing to our knowledge the first direct measure of chronic Cd stress on germline mutation. For both control and Cd exposure conditions, we analyzed the overall genome-wide mutation rate, the conditional mutation rate of the six single-nucleotide mutation classes, and the mutation rates in annotated genome regions [e.g., intergenic, promoters, exons, splice junctions, introns, and 3'-untranslated regions (UTRs)], and we link our findings to their molecular causes.

Methods

Experimental Design and Maintenance of MA Lines

A single female *Daphnia* genotype sampled from Buck Lake, Dorset, Ontario, Canada, was used to produce a population of isogenic female offspring that were used to initiate this MA experiment (Figure 1). MA experiments use strict genetic bottlenecks, which reduce the effective population size to approximately one, enabling the measurement of DNA mutational rates in a laboratory setting where selection is minimized. For each subline, after clonal reproduction each generation we randomly selected one offspring within 24 h after clutch release from the brood pouch to serve as the clonal mother for the next generation. We also randomly selected two other offspring (for each subline) to serve as backups in the event of failed reproduction/lethality of the originally selected clonal mother. Exposure water (control and Cd) was completely changed every generation (14–21 d) and Cd was periodically measured (Dartmouth Trace Element Analysis Core) throughout the MA experiment to ensure nominal concentrations were as expected.

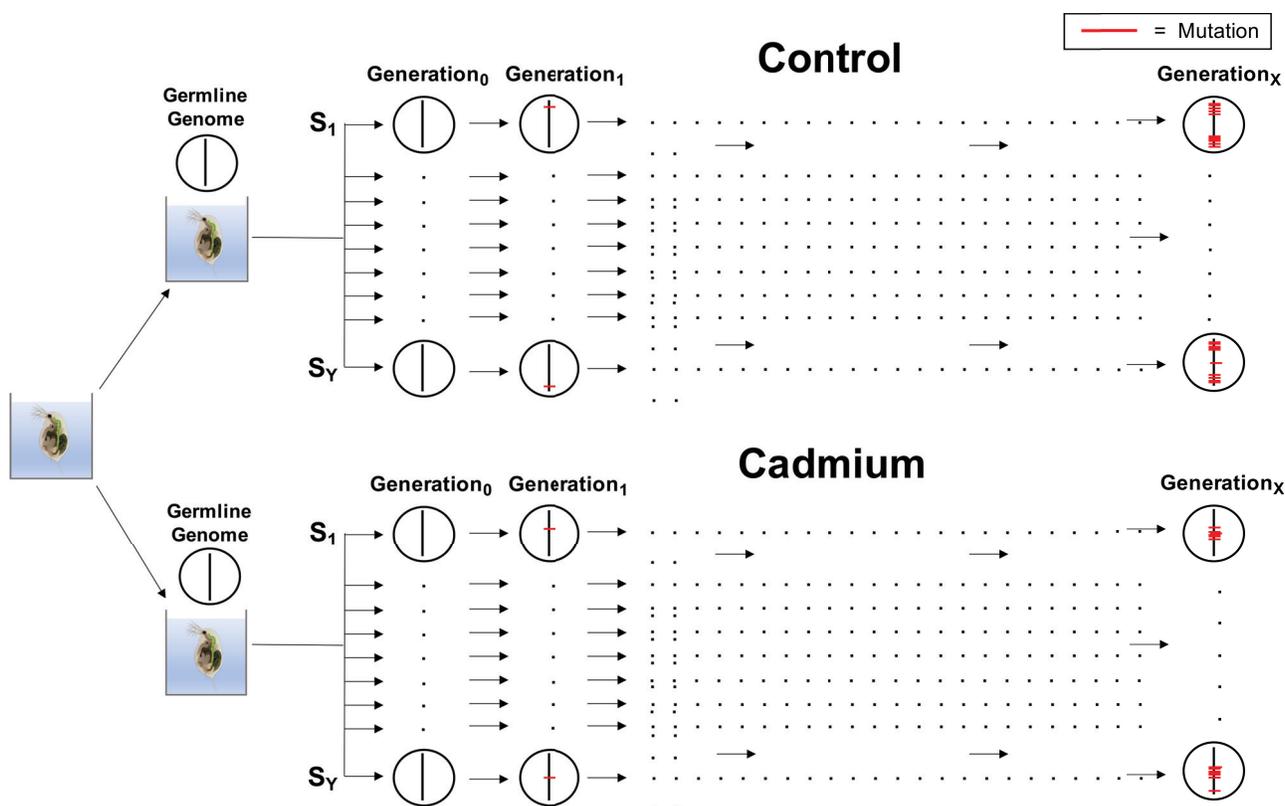


Figure 1. Overview of mutation-accumulation (MA) experiment. A single *Daphnia pulex* genotype from Buck Lake, Dorset, Ontario, Canada, was the progenitor of the sublines for both the control conditions and Cd exposure. “S” refers to individual sublines. Circles containing a vertical line represent the germline genome. Horizontal red dashes in the “germline genome” represent mutations. Whole-genome sequencing with Illumina was undertaken at “Generation_X” for all sublines (Table 1).

Sublines were maintained in laboratory culture using methods originally described by Shaw et al. (2007). Sublines were fed daily with *Ankistrodesmus falcatus* (75,000 cells/mL) and maintained under constant light:dark laboratory conditions (20°C; 12 h light: 12 h dark). The control condition was maintained in 50-mL beakers of modified COMBO media, lacking the addition of nitrogen and phosphorus (Kilham 1998), whereas the Cd exposure sublines were maintained in modified COMBO media with the addition of 0.25 µg Cd/L (Cadmium Chloride, ACS/analytical grade; Fischer Scientific) (Shaw et al. 2007). Primary stocks were made annually by dissolving CdCl₂ (analytical grade, Fischer Scientific) in ultra-pure water. Test Cd concentrations were verified annually using a magnetic sector inductively coupled plasma/mass spectrometer (ELEMENT; ThermoElectron) fitted with a standard liquid sample introduction system [microconcentric nebulizer (MCN-2; CETAC) and cooled Scott-type spray chamber] at the Dartmouth Trace Element Analysis Core (Shaw et al. 2019). To ensure the efficacy of our working stocks and culturing conditions of our *Daphnia* throughout the duration of the MA experiment, working stocks were tested more frequently (monthly) using slightly modified chronic American Society for Testing and Materials (ASTM) toxicity tests (ASTM 1990) on *Daphnia pulex*. This 21-d life table assay measures the reproductive response of single *Daphnia* (10 reps per treatment) exposed low-level concentration of a stressor (in this case, Cd) or control conditions (see below for more details). These standard tests allowed us to verify consistency in our working stock concentrations as well as culturing conditions of our animals.

Two criteria were used to select the concentration of Cd used in these studies: *a*) environmental relevance to Cd concentrations experienced by *D. pulex* living in polluted Ontario lakes (Rajotte and Couture 2002; Stephenson and Mackie 1988), and *b*) concentrations resulting in minimal health effects in the *Daphnia* (i.e., no effects on survival or reproductive fitness). To meet the second criteria and to ensure Cd responses, we set the exposure concentration to the lowest effects concentration based on U.S. Environmental Protection Agency (U.S. EPA) *Daphnia* chronic effect data (Eaton and Gentile 2001). Early in the MA experiments (generation 6), the test concentration was validated using the slightly modified chronic life table tests following ASTM methods (ASTM 1990) and detailed in Shaw et al. (2019). In these experiments, net reproductive rate (R_0) was determined by summing the daily reproductive [as described in Chen and Folt (1996)] output of each of the five clonal replicates for the 90 different isolates in both control (0 µg Cd/L) and Cd-exposed conditions (0.25, 0.5, and 1 µg Cd/L) over the duration of the experiment [methods detailed in Shaw et al. (2007)]. However, the length of the test was modified from the traditional 21 d to 30 d to adjust for the life-history dynamics of these isolates, ensuring measurements were collected across three broods. The results were analyzed using an analysis of variance (ANOVA) followed by a post hoc *t*-test, confirmed the concentration of Cd (0.5 µg Cd/L) had only a slight but nonsignificant effect (control, 11.2 ± 2.14 and Cd-exposed, 10.9 ± 4.16) on net reproductive rate (R_0). The same life table experiments were conducted for each MA line around generation 40 to show how mutation rates affect fitness (see “Results” section).

Library Preparation and Whole-Genome Sequencing

For both the control and Cd exposures, we randomly selected 12 sublines for next-generation sequencing after an average of 55 generations of genetic bottlenecks per subline. For each of these sublines, we randomly selected a single female and allowed her to clonally reproduce until there were >25 isogenic offspring to provide >1 µg genomic DNA for library preparation. DNA

isolation was performed with standard Trizol DNA extraction methods. Whole-genome, paired-end sequencing libraries were prepared using NEBNext Ultra DNA Library Prep Kit for Illumina (Catalog #E7370) and the NEBNext Multiplex Oligos for Illumina. Paired-end libraries were then sequenced on the Illumina HiSeq 2500 platform. For the Cd exposure, 2 out of 12 sublines were lost via failed library preparation, and another was a mutator phenotype with a mutation rate >200-fold higher than the other sublines and was therefore excluded from mutation rate analyses (Table S1).

Processing and Mapping of Sequencing Reads

For each subline, paired-end read trimming was performed with Trimmomatic software (Trimmomatic v.0.36, USADEL Lab) (Bolger et al. 2014) with the following parameters: Paired-end mode ILLUMINACLIP:2:30:10 LEADING:28 TRAILING:28 MINLEN:50. Paired-end reads were mapped with BWA-MEM (BWA-MEM v0.7.17, Heng Li Lab) (Li and Durbin 2009) to the *Daphnia pulex* (TCO genotype) reference assembly V1.0 (GCA_0001878751.1) with paired-end mode using default parameters. We generated two SAMtools (Samtools v1.9, Wellcome Sanger Institute) (Li et al. 2009) mpileup files that reported only sites with high-confidence genotype calls (parameters -ABx -min-BQ 30 -min-MQ 30): one file containing the control sublines and the other containing the sub-lines exposed to Cd.

Defining Analyzable Sites and Single-Nucleotide Mutation (SNM) Identification

To identify single-nucleotide mutations (SNMs), we designed and developed a SNM identification Python script that is useful for future MA experiments that also use deep-coverage, whole-genome sequencing of clonally reproducing diploid species that use heterozygous genomes (Mutation_Caller.py; Supplemental File 1). The following criteria were used to identify analyzable sites from control and Cd exposure mpileup files independently: *a*) For a site to be analyzed we required the minimum proportion of mapped reads to be 0.9 for homozygous sites, and between 0.3 for and 0.7 for heterozygous sites. If a given site met either of these criteria across all sublines, and if all sublines had the same genotype, then this shared genotype became the “consensus” genotype for the site. Further, if all sublines but one shared a given genotype, then the unique genotype was regarded as a potential mutation. *b*) We required a minimum sequencing depth of coverage of 12× while excluding sites with depth of coverage over 45×. *c*) Sites within 20 bp of indels (i.e., insertions and deletions between our genotype and the *Daphnia pulex* reference) were excluded to eliminate genotyping errors due to misalignment. *d*) To eliminate false genotype calls via PCR artifacts from library preparation, we required a minimum of two reads mapped in both orientations supporting each genotype call. *e*) To eliminate the issue of false positives that can arise in repetitive regions, we removed reads that map to multiple loci.

SNM Rate Calculations

We calculated mutation rates according to the methods outlined by Keith et al. (2016). For each subline in control and Cd exposure, we independently calculated the genome-wide base-substitution rate with $\mu_{bs} = m_i / (2n_i T)$ (Lynch et al. 2008; Sung et al. 2012b), where μ_{bs} is the genome-wide rate of SNM, m is the total number of base-substitutions, $2n_i$ is the total number of sites analyzed (we used “2n” because each line is maintained in a heterozygous state), and T is the number of generations for the subline. For each subline, the standard error (SE) was calculated with $SE_x = \sqrt{\mu_{bs} / 2nT}$, where μ_{bs} is the base-substitution mutation

rate for a given subline, T is number of generations, and n is number of analyzed sites. Independently for control and Cd exposure, the overall genome-wide rate of base-substitution mutation was calculated as the average of the subline genome-wide base-substitution rates. The overall SE for control and cadmium exposure was calculated with $SE_{pooled} = s/\sqrt{N-1}$, where s is the standard deviation of the base-substitution mutation rate for all sublines, and N is the total number of analyzed sublines.

For each subline, the conditional mutation rates for each of the six classes of base-substitution mutations were calculated with $\mu_{cond} = m_i/(2n_iT)$, but with $n_{b,i}$ instead of n_i , and $m_{b->d,i}$ instead of m_i , where $n_{b,i}$ is the number of ancestral sites of nucleotide type b ($b = A, T, G, \text{ or } C$) in an MA line i , and $m_{b->d,i}$ is the number of mutations from nucleotide type b to any of the three possible base pairs to which it can mutate (e.g., $A:T \rightarrow G:C$). For control and Cd exposure, the overall conditional mutation rate for each of the six classes was calculated as the average of each class across all sublines. For each base-substitution class, the SE was calculated according to $SE_{cond} = \sqrt{\mu_{cond}/LnT}$, where μ_{cond} is the rate for a specific class, L is the number of lines analyzed in either control and Cd exposure, T is the average number of generations, and n is the number of analyzed sites.

Genomic Regional Mutation Rate Analyses

We investigated *de novo* mutations in our *D. pulex* sublines using the recently defined, high-quality *D. pulex* genome annotations from Ye et al. (2017). In addition, we analyzed regional mutation patterns using data from a previous *D. pulex* MA experiment—whose regional mutation spectrum had not been investigated (genotype “ASEX”; Keith et al. 2016). The regions investigated were intergenic regions, promoters, exons, splice-site junctions, introns, and 3'-UTRs. For all genic regions, i.e., all regions besides intergenic, we used annotated genes from the most recent *Daphnia pulex* reference genome assembly (Ye et al. 2017), which reduced the gene count from 30,097 (Colbourne et al. 2011) to 18,440 after gene prediction and annotation methods were updated to current methods. Because of the lack of promoter characterization in *D. pulex*, promoter regions were conservatively designated as the 500 base pairs directly upstream from translation start sites. Additionally, splice-site junctions were defined as the 20 base pairs around intron/exon junctions because these 20 bps are evolutionarily conserved in *D. pulex*, suggesting these regions are important for spliceosome function (Lynch et al. 2017).

The Exact Binomial Test was used with R (version 3.3.0; R Development Core Team) to identify significant differences in the proportion of total mutations within each genomic region category (exon, intron, UTR, promoter, intron-exon junction, and intergenic regions) relative to the expectation if mutations were evenly distributed throughout the genome. These tests were performed for control conditions, Cd exposure, and ASEX (Keith et al. 2016), independently. Fisher's exact test was used to compare the proportions of total mutations within each genomic region category to identify significant mutation spectra differences between control conditions and Cd exposure.

5-Hmc Quantification

5-Hydroxymethylation levels were measured in the “NONA” *D. pulex* genotype, which is the genotype used to initiate the MA experiments. NONA *Daphnia* were exposed to Cd chronically (14-d; 0.25 $\mu\text{g Cd/L}$) or acutely (1-d; 20 $\mu\text{g Cd/L}$), or to control conditions, and 5-hydroxymethylation was measured using the Abcam Hydroxymethylated DNA Quantification Kit (Abcam, Cat. No. 117130) according to manufacturer recommendations.

Specifically, a Metertech M965 microplate reader set to excitation/emission spectra of 535/587 nm, along with proprietary PC-AccuMate software was used for colorimetric quantification. Four replicates consisting of five adult *Daphnia* each were assayed per experimental condition. Data were then imported into Excel, and differences were estimated using one-way ANOVA followed by Tukey's test.

Results

The isogenic offspring (sublines) of a *D. pulex* genotype (genotype ID “NONA”) sampled from Buck Lake, Dorset, Ontario, were subjected to an MA experiment independently propagated under control conditions and an environmentally relevant, continuous Cd exposure (Figure 1). The experiment was conducted for an average of 55 generations of genetic bottlenecks per subline (effective population size (N_e) = 1). The control condition sublines were propagated for 632 total generations, and the sublines constantly exposed to Cd (0.25 $\mu\text{g Cd/L}$) were propagated for 491 total generations (see “Methods” section under “Experimental Design and Maintenance of MA Lines”; Table 1). Prior to sequencing, life table experiments were repeated on each subline. Again, there was no effect on survivorship, but there was a significant ($p < .05$) effect of Cd on R_0 (control, 9.8 ± 3.68 and Cd-exposed, 7.6 ± 4.11). Although R_0 was lower in the Cd-exposed group, the tails of the distribution for each group remained similar, ranging from 2.6 to 19.6 for the controls and 2 to 28.2 for the Cd-exposed sublines, thus indicating that the overall effects of our Cd concentration exposure on *Daphnia* health were small.

After 1,123 total generations of MA in laboratory conditions that minimize selection, the sublines from both the control and Cd exposure were subjected to deep coverage, whole-genome sequencing (average sequence coverage of $\sim 25 \times$ per subline; Table 1), which allowed for the analysis of *de novo* SNMs (hereafter referred to simply as “mutations”) that met our stringent criteria for analysis (see “Methods” section under “Defining Analyzable Sites and Single-Nucleotide Mutation (SNM) Identification”). We analyzed *de novo* mutations that occurred across 47,522,803 and 51,842,909 sites for each control condition and Cd exposure subline, respectively (Table 1).

Regional Mutation Rates in *Daphnia* MA Sublines Maintained under Control Conditions and Cd Exposure

We independently analyzed *de novo* mutations in annotated genome regions (Ye et al. 2017) in controls and Cd exposures (Table 1; see “Methods” section under “Genomic Regional Mutation Rate Calculations”). After comparing the proportion of total mutations in each region to the expected proportion if mutations were distributed randomly across the genome, we found that in control conditions, mutations occurred more often than expected in exons (exact binomial test, $p = 3 \times 10^{-4}$). However, no differences from random expectations were observed in all other genomic regions (i.e., intergenic sites, promoters, introns, splice-site junctions, and 3'-UTRs; Figure 2A; Table S2). To help validate these results, we repeated our analyses on data from a different *D. pulex* genotype from a previous MA experiment conducted under control conditions (Keith et al. 2016) and also observed more exonic mutations than the random expectation (exact binomial test, $p = 4 \times 10^{-7}$; Figure S1 and Table S3).

In contrast to controls, we did not observe a difference in the proportion of exon mutations in Cd exposure in comparison with the random expectation (exact binomial test, $p = 0.53$ Figure 2B; Table S2). As was observed in control conditions, in Cd exposure there were also no differences from the expectation in all other genome regions (Figure 2B; Table S2).

Table 1. Summary of mutation results for control and Cd MA sublines.

Genotype/Condition/ID	No. of generations	Depth of sequencing coverage	No. of SNMs	Tot. sites analyzed	No. of transitions	No. of transversions	Ts/Tv ratios	Subline SNM rates \pm SE
								($\times 10^{-9}$)
NONA Control 1	54	23.7	14	47,522,803	8	6	1.3	2.73 \pm 0.73
NONA Control 10	52	23.9	5	47,522,803	2	3	0.7	1.01 \pm 0.45
NONA Control 11	51	18.8	5	47,522,803	3	2	1.5	1.03 \pm 0.46
NONA Control 12	53	23.2	4	47,522,803	2	2	1.0	0.79 \pm 0.4
NONA Control 2	52	24.1	6	47,522,803	3	3	1.0	1.21 \pm 0.50
NONA Control 3	53	25.6	5	47,522,803	2	3	0.7	0.99 \pm 0.44
NONA Control 4	52	24.1	13	47,522,803	5	8	0.6	2.63 \pm 0.73
NONA Control 5	54	24.4	5	47,522,803	1	4	0.3	0.97 \pm 0.44
NONA Control 6	52	23.1	8	47,522,803	5	3	1.7	1.62 \pm 0.57
NONA Control 7	52	15.9	16	47,522,803	8	8	1.0	3.24 \pm 0.81
NONA Control 8	53	17.9	7	47,522,803	1	6	0.2	1.39 \pm 0.53
NONA Control 9	54	21.8	6	47,522,803	4	2	2.0	1.17 \pm 0.48
Control summary	632*	22.2 (Avg.)	94*	570,273,636*	44*	50*	0.9	1.57 \pm 0.23 (Avg.)
NONA Cd 11	52	25.0	9	51,842,909	5	4	1.3	1.67 \pm 0.56
NONA Cd 12	55	24.2	14	51,842,909	10	4	2.5	2.45 \pm 0.66
NONA Cd 2	56	25.2	7	51,842,909	5	2	2.5	1.21 \pm 0.46
NONA Cd 3	54	22.8	12	51,842,909	6	6	1.0	2.14 \pm 0.62
NONA Cd 4	55	23.4	11	51,842,909	8	3	2.7	1.93 \pm 0.58
NONA Cd 5	54	23.4	9	51,842,909	3	6	0.5	1.61 \pm 0.54
NONA Cd 6	54	27.6	2	51,842,909	2	0	—	0.36 \pm 0.25
NONA Cd 7	55	25.2	11	51,842,909	5	6	0.8	1.93 \pm 0.58
NONA Cd 8	56	30.3	11	51,842,909	7	4	1.8	1.89 \pm 0.57
Cd summary	491*	25.2 (Avg.)	86*	466,586,181*	51*	35*	1.5	1.69 \pm 0.20 (Avg.)

Note: Summary statistics of mutation results for sublines in control and cadmium conditions. “NONA” is the name of the genotype used in the experiment. The “NONA Cd” has three fewer sublines because of two failed libraries, and sequenced subline was a mutation phenotype that was excluded from downstream analyses. “*” in the summary rows denotes a value as being additive. The average SNM rate (or “mutation rate”) across all sublines for each experimental condition is shown in the Summary row for the “Subline SNM Rates” column. Cd, cadmium; SE, standard error; MA, mutation-accumulation; SNMs, single-nucleotide mutations (which are referred to as “mutations” in the main text). Ts/Tv ratio, the ratio of transition mutations to transversion mutations.

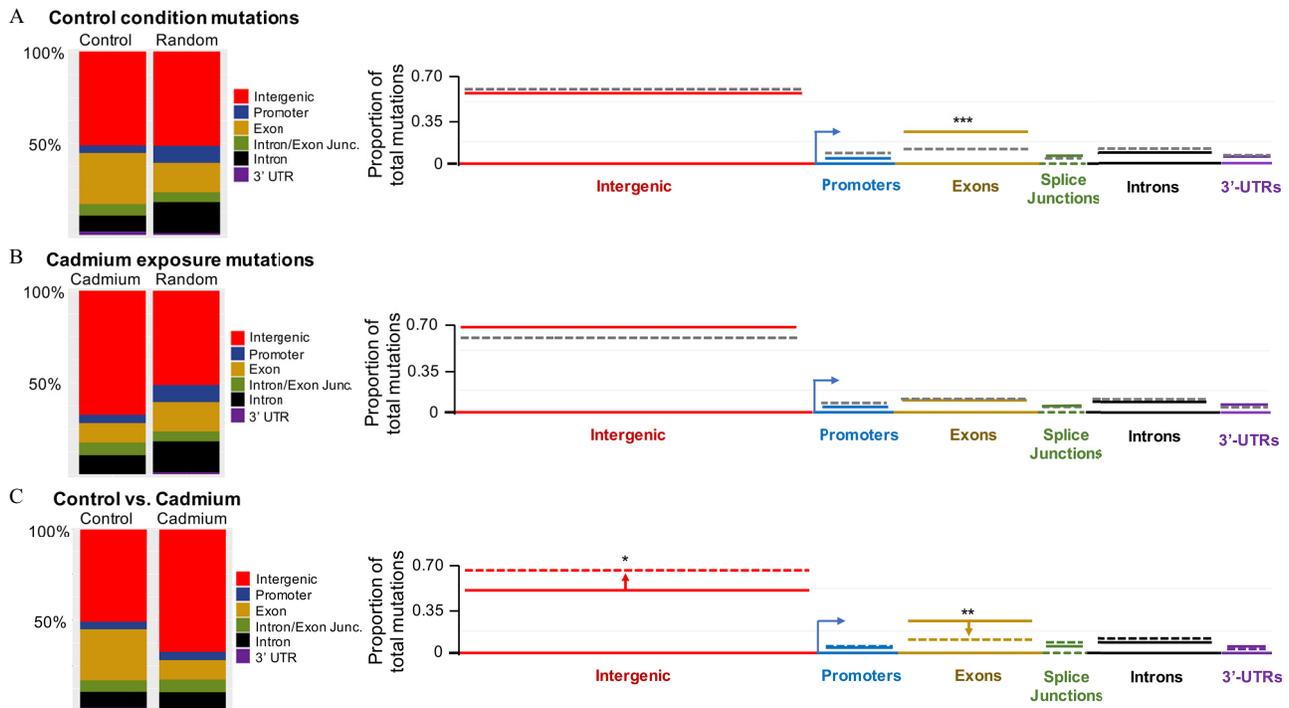


Figure 2. (A, B) Comparison of the proportion of total mutations in each genome region to the random expectation (i.e., if mutations were randomly distributed throughout the genome). Gray dashed lines represent the proportion of total mutations in each region that are expected if mutations are randomly distributed across the genome. Solid lines represent observed mutation proportions in each genome region for control conditions (A) and Cd exposure (B). *** corresponds to $p < 5.0 \times 10^{-4}$, Exact Binomial Test. (C) Comparison of the proportion of total mutations in each genome region between control conditions (solid lines) and Cd exposure (dashed lines). Arrows (2C) indicate the direction regional mutation rate were significantly different between Cd exposure and control conditions. * Corresponds to $p < 5.0 \times 10^{-2}$; ** corresponds to $p < 5.0 \times 10^{-3}$; Fisher’s Exact Test. Summary data, including the expected and observed proportions, for the mutation rates can be found in Table S2 (control and Cd). Actual p -values for control vs. Cd can be found in Table S4.

We then compared the regional mutation proportions in cadmium exposure with those of control conditions to identify regions where sublines raised in Cd exposure were different from those raised in control with regard to the mutation rate. As expected, based on comparisons to the random distribution, the exon mutation rate was significantly lower in Cd in comparison with controls (Fisher's exact test, $p = 4.4 \times 10^{-3}$; Figure 2C; Table S4).

Notably, the overall genome-wide mutation rate did not significantly differ between conditions (t -test, $p = 0.70$; average mutation rates are 1.6×10^{-9} and 1.7×10^{-9} site⁻¹ generation⁻¹ for control and Cd exposure, respectively; Table 1). We therefore reasoned that the lower proportion of total mutations in exons in Cd must be balanced by a higher proportion in another region(s). Indeed, the intergenic mutation rate was significantly higher under Cd exposure relative to control conditions (Fisher's exact test, $p = 0.03$; Figure 2C; Table S4), whereas no significant differences were observed in other regions.

Evaluation of Genome-Wide Mutation Biases in Sublines Propagated under Control Conditions and Cd Exposure

We observed a mutation bias in the G/C → A/T direction in both experimental conditions. Although there was no considerable difference between Cd exposure and control with regard to overall, genome-wide mutation rate, there was markedly lower G/C → A/T bias in Cd. G/C → A/T mutations arose 2.0× more frequently in controls than those in the opposite direction (i.e., A/T → G/C) compared to only being 1.2× higher in the Cd-exposed lines (Table 1; Table S5 and S6).

Evaluation of A/T → G/C Transitions

To explain the lower G/C → A/T bias in Cd-exposed lines, we reasoned that a mutation class or classes must be more prevalent in Cd-exposed lines at A/T sites (i.e., A/T → G/C, the opposing direction of the mutation bias). We observed a significantly higher conditional rate of A:T → G:C transitions in Cd exposure relative to control conditions, which to our knowledge has not been previously reported (Figure 3A; Tables S5 and S6). Because the intergenic mutation rate was higher under Cd exposure relative to control conditions (Figure 2C), we also reasoned that the higher overall conditional A:T → G:C mutation rate was specific to intergenic regions. In Cd, the intergenic, conditional mutation rate of A:T → G:C was significantly higher than that of control, which was not observed in genes (Fisher's exact test, $p < 0.05$; Figure 3B; Table S7).

Because a difference in the A:T → G:C mutation rate has not been previously reported, we investigated the 5' and 3' nucleotides to each mutation to determine whether there is a specific context(s) enriched in the Cd exposure that has been linked to a causative A:T → G:C mutational mechanism. Among significantly enriched contexts we observed a Cd-specific enrichment of GTT context mutations (Fisher's exact test, $p = 0.02$; Figure 3D; Table S8).

Evaluation of C:G → G:C Transversions

Contrary to a higher conditional mutation rate of A:T → G:C mutations under Cd exposure, the rate of C:G → G:C mutations was lower in Cd (Fisher's exact test, $p = 0.02$, $p = 0.02$; Figure 3A). In fact, under Cd exposure, no C:G → G:C mutations were observed in genes (Figure 3B).

C:G → G:C mutations have been shown to be enriched at 5-hydroxymethylcytosine (5-hmC) positions (Supek and Lehner 2017), and Cd has been shown to decrease 5-hmC levels in *Daphnia* (Streptkaitė et al. 2015). To examine whether or not

Cd was involved with the lack of C:G → G:C mutations we therefore measured 5-hmC levels in both control conditions and chronic exposure to 0.25 μg Cd/L, the same Cd concentration used for this MA experiment (see "Methods" section under "5-hmC Quantification"). Lines exposed to Cd had significantly lower 5-hmC levels relative to control conditions (one-way ANOVA with Tukey's test, $p = 0.01$; Figure 4A; Table S9).

Evaluation of G:C → T:A Transversions

In comparison with those in the control media, lines exposed to Cd did not differ in overall genome-wide G:C → T:A mutation rates: a mutation that occurs if 8-oxoguanine is not repaired (Avkin and Livneh 2002). However, although there was no difference in the rate of G:C → T:A when considering all mutations genome-wide (Figure 3A), there was a higher proportion of G:C → T:A mutations within mutation clusters (i.e., more than two mutations within 50 bp) under Cd exposure relative to control conditions (Figure 3C; Table S10).

Discussion

Here, to the best of our knowledge we provide the first analysis of the genome-wide patterns of Cd-induced mutations in the germline. The results presented here were derived from the measurement of mutational rates via an MA experiment propagated independently, in both control and chronic exposure to Cd. The Cd exposure concentration in this experiment was carefully chosen because of its minimal effects on *D. pulex* reproductive fitness (see "Methods" section under "Experimental Design and Maintenance of MA Lines"). Relative to control conditions, the intergenic mutation rate was higher in Cd exposure and the exonic mutation rate was lower in Cd. These regional differences were linked to differences in specific mutation classes. Specifically, the higher intergenic rate was the result of a higher rate of A:T → G:C in Cd-exposed lines and the lower genic rate was the result of the complete loss of genic C:G → G:C mutations in Cd-exposed lines.

A recent study of human *POLE*-mutant colorectal tumor genomes reported a lower-than-expected rate of mutations within exons and suggested this result could be extrapolated to the germline (Frigola et al. 2017). Here, we report that in control conditions the exon mutation rate was higher than the random expectation in the *Daphnia* germline and the observed higher levels of exonic mutations relative to introns. The observation of more mutations in exons was confirmed (Figure S1 and Table S3) by analyzing the regional mutation patterns using data from a previous MA experiment that used *D. pulex* genotype, which was also propagated under control laboratory conditions (exact binomial test, $p = 3.51 \times 10^{-7}$; Genotype ID "ASEX"; Keith et al. 2016). Additionally, in the ASEX genotype the proportion of intergenic mutations was significantly lower than expected (exact binomial test, $p = 2.08 \times 10^{-6}$; Figure S1) and the proportion of 3'-UTR mutations was uniquely elevated in ASEX (Figure S1). In the control conditions here, we also observed a lower proportion of mutations in intergenic regions relative to the random expectation although the difference marginally nonsignificant (exact binomial test, $p = 0.07$; Table S2; Figure 2). These results suggest that *a*) exon mutations arise more often than expected in the *D. pulex* germline, and *b*) that regional *D. pulex* germline mutation patterns, in at least some genome regions, may differ between genotypes, which was reported in a *D. pulex* MA experiment (Keith et al. 2016).

Growing evidence suggests that C:G → G:C mutations are concentrated at 5-hydroxymethylcytosine positions (5-hmC), which are the result of the oxidation of 5-methylcytosine (5-mC) by TET proteins (He et al. 2011; Ito et al. 2011; Kriaucionis and

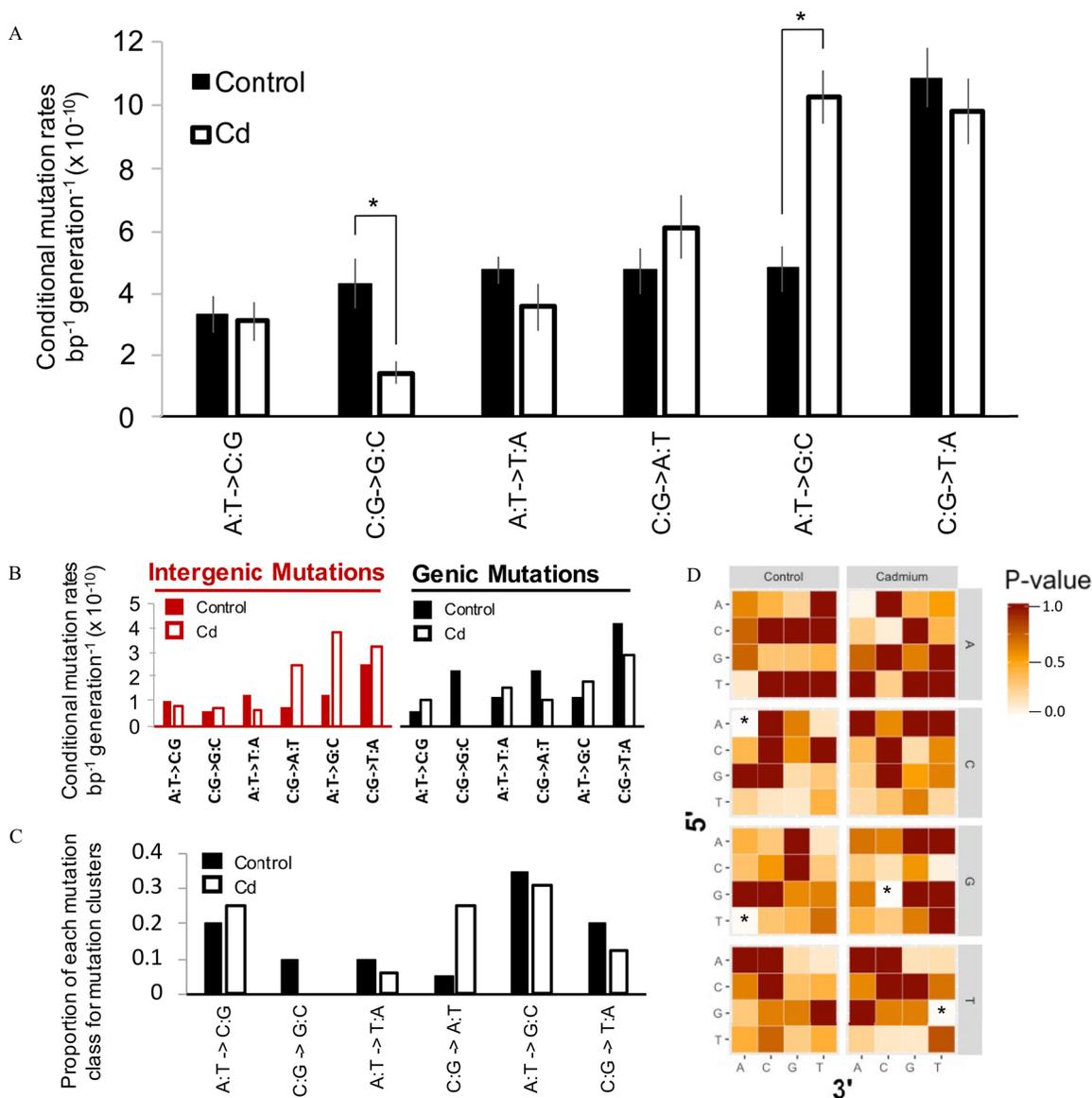


Figure 3. (A) Genome-wide conditional mutation rates of control conditions and Cd exposure. For control (solid black boxes) and Cd exposure (white boxes), the rates of each of the six base-substitution classes are plotted. Error bars are included (SE; gray). Significant differences between control and Cd are denoted by asterisks. The data for 3A are located in Tables S5–S6. Values represent 12 (control) and 9 (Cd) sublines (B) conditional base-substitution rates of intergenic and genic regions. The conditional mutation rates are plotted independently for intergenic and genic regions for the six mutation classes. Solid boxes represent control condition results, and white boxes represent Cd exposure results. The data for 3B are in Table S7 (C) Mutation cluster analysis. The proportion of total mutations for each mutational class are plotted for control conditions and Cd exposure. The data for 3C are in Table S10. (D) Heat map of *p*-values from Fisher's exact test for all possible mutation contexts. The Fisher's exact test for Cd exposure is in Table S8, and the results for control conditions are in Table S11. *p*-Values are plotted for control conditions and Cd exposure, independently. The z-axis is the pre-mutation nucleotide. The y-axis is the nucleotide that is 5' adjacent to the mutation site. The x-axis is the nucleotide that is 3' adjacent to the mutation site. Note: Cd, cadmium.

Heintz 2009; Tahiliani et al. 2009; Supek et al. 2014). In *D. pulex*, 5-hmC sites were recently shown to be concentrated in genes (Streptkaić et al. 2015). Our results suggest that the lower overall rate of C:G → G:C in Cd exposure is the result of the interference of 5-mC → 5-hmC conversion, which eliminated C:G → G:C mutations in genes where 5-hmC was concentrated in *D. pulex* (Figure 4B). To our knowledge we report the first evidence that the interference of hydroxymethylation by a toxicant may influence C:G → G:C mutational outcomes in the germline.

It was recently reported that Cd inhibits 5-mC → 5-hmC conversion by reducing the enzymatic activity of TET proteins in mouse ES cells (Xiong et al. 2017), although the mechanism of inhibition has not been described. TET proteins contain two Zn-finger domains, and the structure of these domains is essential for

stabilizing TET above DNA, thereby allowing TET to convert 5-mC to 5-hmC via oxidation (Hu et al. 2013). Given the propensity for Cd to replace Zn in Zn-finger domains (Li and Manning 1955), we propose that Cd inhibits the conversion of 5-mC to 5-hmC by interfering with the Zn-finger domains of TET proteins, and the corresponding reduction of 5-mC sites results in fewer C:G → G:C mutations (Figure 4B). Indeed, we measured lower 5-mC levels in *D. pulex* exposed to Cd.

In Cd exposure, the significantly elevated GTT mutation context we observed was recently linked to Pol η-induced A:T → G:C mutations in certain lymphomas (Supek and Lehner 2017), providing evidence that the elevated A:T → G:C rate in Cd is linked to this polymerase. Pol η is recruited to bypass 8-oxoG lesions (Rodriguez et al. 2013), and translesion synthesis (TLS)

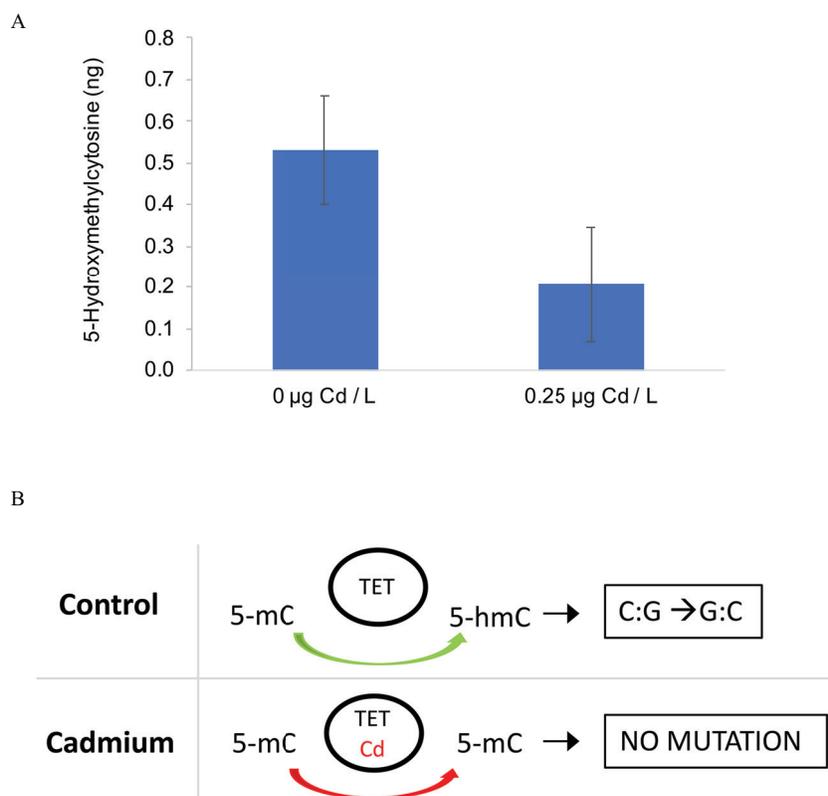


Figure 4. (A) Measurement of 5-hydroxymethylation levels in control conditions and Cd exposure. The mean and corresponding SEM are plotted. The data plotted correspond to three to four replicates for each experimental condition. The measurements used to generate Figure 4A are provided in Table S9. (B) Proposed mechanism for lower C:G → G:C mutation rate in Cd exposure. 5-mC = 5-methylcytosine. 5-hmC = 5-hydroxymethylcytosine. “TET” is TET protein, which converts 5-mC to 5-hmC. Note: Cd, cadmium; SEM, standard error of the mean.

is activated by the binding of the Pol η ubiquitin-binding Zn-finger domain (UBZ) with an available monoubiquitin of PCNA (Bienko et al. 2010). Disruption of the amino acids that coordinate Zn binding in the UBZ reduces TLS efficiency when bypassing 8-oxoG lesions (Plosky et al. 2006). Given the propensity for Cd to replace Zn in Zn-finger domains (Li and Manning 1955), the elevated A:T → G:C rate in Cd exposure is possibly the result of Cd inhibiting the UBZ in Pol η , thereby decreasing the fidelity of Pol η and further increasing the number of A:T → G:C mutations. However, an alternative explanation for the increased A:T → G:C mutation rate in Cd exposure is that error-prone Pol η TLS is used more often because of an increased occurrence of oxidative stress-induced 8-oxoG lesions.

Cd did not change the overall, genome-wide G:C → T:A mutation rate: a mutation that occurs if 8-oxoguanine is not repaired (Avkin and Livneh 2002). This result was surprising, considering the well-established history of Cd inducing oxidative stress, which causes oxidative DNA damage (Avkin and Livneh 2002). Cd interference with the BER glycosylase MutY provides a possible explanation for the increased prevalence of G:C → T:A clustered mutations in Cd exposure. To repair 8-oxoG:A lesions, MutY recognizes and excises adenine (Fromme et al. 2004; Lu et al. 1996), and cytosine is added opposite 8-oxoG by gap filling polymerases (van Loon and Hübscher 2009). MutY contains a Zn linchpin motif, where Zn is bound by evolutionarily conserved cysteines that reside in an interdomain connector (IDC) (Engstrom et al. 2014). Changes to IDC Zn-binding cysteines in *E. coli* have been shown to increase the rate of mutation by up to 12-fold (Engstrom et al. 2014), indicating the IDC structure is needed for adenine excision. Further, a study of an *E. coli* MutY knockout genotype resulted in an elevated rate of clustered G:C → T:A mutations

(Pearson et al. 2004). These findings are consistent with our observations in Cd exposure and suggests that the higher number of G:C → T:A mutations observed in clusters is the result of Cd preferentially binding within the MutY IDC domain.

Because many inherited diseases in humans are caused by pathogenic mutations in specific genome regions, any chemical that changes the rate of germline mutation in individual genome regions influences the accuracy of disease occurrence predictions. A growing number of neurological disorders are linked to pathogenic mutations in intergenic regions (Short et al. 2018), and any pollutant that increases the intergenic mutation rate, such as we report here for Cd, has the potential to influence the rate of occurrence of these disorders. Additionally, because many diseases and disorders are caused by mutations that change the structure and function of proteins (Antonarakis and Beckmann 2006), it is likely that uncharacterized pollutants that specifically elevate the mutation rate in coding regions exist, thereby increasing the probability of diseases caused by changes to protein structure and function. The results reported here, combined with the knowledge that changes to the germline mutational spectrum can influence disease occurrence, shows that a more concentrated effort is needed for understanding how the more than 140,000 anthropogenic chemicals affect genome-wide patterns of germline mutation.

Acknowledgments

The authors thank F. Supek, W. Sung, and T. Leuthner for many helpful discussions about germline mutation, which greatly improved this manuscript.

Funding for this project was provided by the National Institutes of Environmental Health Sciences to J.R.S. (R01ES019324).

N.K. and J.R.S. designed all analyses. N.K., C.J., M.L., and J.R.S. analyzed next-generation sequencing results. K.Y. and C.G. maintained the MA lines. M.L. provided laboratory space for the MA lines. All authors contributed edits to the final manuscript.

The data from this study have been submitted to the NCBI BioProject under accession PRJNA522033.

References

- Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincoren I, Nik-Zainal S, et al. 2016. Mutational signatures associated with tobacco smoking in human cancer. *Science* 354(6312):618–622, PMID: 27811275, <https://doi.org/10.1126/science.aag0299>.
- Ames BN, Durston WE, Yamasaki E, Lee FD. 1973. Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc Natl Acad Sci U S A* 70(8):2281–2285, PMID: 4151811, <https://doi.org/10.1073/pnas.70.8.2281>.
- Antonarakis SE, Beckmann JS. 2006. Mendelian disorders deserve more attention. *Nat Rev Genet* 7(4):277–282, PMID: 16534515, <https://doi.org/10.1038/nrg1826>.
- ASTM (American Society for Testing and Materials). 1990. Standard Guide for Conducting *Daphnia magna* Life-Cycle Toxicity Tests. ASTM E1193–20. West Conshohocken, PA: ASTM International.
- Avkin S, Livneh Z. 2002. Efficiency, specificity and DNA polymerase-dependence of translesion replication across the oxidative DNA lesion 8-oxoguanine in human cells. *Mutat Res-Fund Mol M* 510(1–2):81–90, [https://doi.org/10.1016/S0027-5107\(02\)00254-3](https://doi.org/10.1016/S0027-5107(02)00254-3).
- Bienko M, Green CM, Sabbioneda S, Crosetto N, Matic I, Hibbert RG, et al. 2010. Regulation of translesion synthesis DNA polymerase eta by monoubiquitination. *Mol Cell* 37(3):396–407, PMID: 20159558, <https://doi.org/10.1016/j.molcel.2009.12.039>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120, PMID: 24695404, <https://doi.org/10.1093/bioinformatics/btu170>.
- Chain FJJ, Flynn JM, Bull JK, Cristescu ME. 2019. Accelerated rates of large-scale mutations in the presence of copper and nickel. *Genome Res* 29(1):64–73, PMID: 30487211, <https://doi.org/10.1101/gr.234724.118>.
- Chan K, Resnick MA, Gordenin DA. 2013. The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. *DNA Repair (Amst)* 12(11):878–889, PMID: 23988736, <https://doi.org/10.1016/j.dnarep.2013.07.008>.
- Chen CY, Folt CL. 1996. Consequences of fall warming for zooplankton overwintering success. *Limnol Oceanogr* 41(5):1077–1086, <https://doi.org/10.4319/lo.1996.41.5.1077>.
- Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, et al. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science* 331(6017):555–561, PMID: 21292972, <https://doi.org/10.1126/science.1197761>.
- Eaton J, Gentile J. 2001. 2001 Update of Ambient Water Quality Criteria for Cadmium. Washington, DC: U.S. Environmental Protection Agency.
- Engstrom LM, Brinkmeyer MK, Ha Y, Raetz AG, Hedman B, Hodgson KO, et al. 2014. A zinc linchpin motif in the MUTYH glycosylase interdomain connector is required for efficient repair of DNA damage. *J Am Chem Soc* 136(22):7829–7832, PMID: 24841533, <https://doi.org/10.1021/ja502942d>.
- Filipic M, Fatur T, Vudrag M. 2006. Molecular mechanisms of cadmium induced mutagenicity. *Hum Exp Toxicol* 25(2):67–77, PMID: 16539211, <https://doi.org/10.1191/0960327106ht590a>.
- Frigola J, Sabinathan R, Mularoni L, Muiños F, Gonzalez-Perez A, López-Bigas N. 2017. Reduced mutation rate in exons due to differential mismatch repair. *Nat Genet* 49(12):1684–1692, PMID: 29106418, <https://doi.org/10.1038/ng.3991>.
- Fromme JC, Banerjee A, Huang SJ, Verdine GL. 2004. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* 427(6975):652–656, PMID: 14961129, <https://doi.org/10.1038/nature02306>.
- Guan R, Wang WX. 2006. Comparison between two clones of *daphnia magna*: effects of multigenerational cadmium exposure on toxicity, individual fitness, and biokinetics. *Aquat Toxicol* 76(3–4):217–229, PMID: 16289344, <https://doi.org/10.1016/j.aquatox.2005.10.003>.
- Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu Rev Ecol Syst* 40(1):151–172, <https://doi.org/10.1146/annurev.ecolsys.39.110707.173437>.
- Hartmann A, Speit G. 1994. Comparative investigations of the genotoxic effects of metals in the single cells gel (SCG) assay and the sister chromatid exchange (SCE) test. *Environ Mol Mutagen* 23(4):299–305, PMID: 8013477, <https://doi.org/10.1002/em.2850230407>.
- He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, et al. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333(6047):1303–1307, PMID: 21817016, <https://doi.org/10.1126/science.1210944>.
- Hu LL, Li Z, Cheng JD, Rao QH, Gong W, Liu MJ, et al. 2013. Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* 155(7):1545–1555, PMID: 24315485, <https://doi.org/10.1016/j.cell.2013.11.020>.
- Hutton M, Symon C. 1986. The quantities of cadmium, lead, mercury and arsenic entering the UK environment from human activities. *Sci Total Environ* 57:129–150, [https://doi.org/10.1016/0048-9697\(86\)90018-5](https://doi.org/10.1016/0048-9697(86)90018-5).
- IARC (International Agency for Research on Cancer). 1993. Cadmium and cadmium compounds. *IARC Monogr Eval Carcinog Risks Hum* 58:119–237.
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333(6047):1300–1303, PMID: 21778364, <https://doi.org/10.1126/science.1210597>.
- Jin YH, Clark AB, Slebos RJC, Al-Refai H, Taylor JA, Kunkel TA, et al. 2003. Cadmium is a mutagen that acts by inhibiting mismatch repair. *Nat Genet* 34(3):326–329, PMID: 12796780, <https://doi.org/10.1038/ng1172>.
- 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(1):60–69, PMID: 26518480, <https://doi.org/10.1101/gr.191338.115>.
- Kilham SS, Kreeger DA, Lynn SG, Goulden CE, Herrera L. 1998. COMBO: a defined freshwater culture medium for algae and zooplankton. *Hydrobiologia* 377(1/3):147–159, <https://doi.org/10.1023/A:1003231628456>.
- Kimmins S, Sassone-Corsi P. 2005. Chromatin remodeling and epigenetic features of germ cells. *Nature* 434(7033):583–589, PMID: 15800613, <https://doi.org/10.1038/nature03368>.
- Kriaucionis S, Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324(5929):929–930, PMID: 19372393, <https://doi.org/10.1126/science.1169786>.
- Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu NN, et al. 2018. The Lancet Commission on pollution and health. *Lancet* 391(10119):462–430, PMID: 29056410, [https://doi.org/10.1016/S0140-6736\(17\)32345-0](https://doi.org/10.1016/S0140-6736(17)32345-0).
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760, PMID: 19451168, <https://doi.org/10.1093/bioinformatics/btp324>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079, PMID: 19505943, <https://doi.org/10.1093/bioinformatics/btp352>.
- Li NC, Manning RA. 1955. Some metal complexes of sulfur-containing amino acids. *J Am Chem Soc* 77(20):5225–5228, <https://doi.org/10.1021/ja01625a006>.
- Lu AL, Yuen DS, Cillo J. 1996. Catalytic mechanism and DNA substrate recognition of Escherichia coli MutY protein. *J Biol Chem* 271(39):24138–24143, PMID: 8798653, <https://doi.org/10.1074/jbc.271.39.24138>.
- Lynch M, Gutenkunst R, Ackerman M, Spitze K, Ye Z, Maruki T, et al. 2017. Population genomics of *Daphnia pulex*. *Genetics* 206(1):315–332, PMID: 27932545, <https://doi.org/10.1534/genetics.116.190611>.
- 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 U S A* 105(27):9272–9277, PMID: 18583475, <https://doi.org/10.1073/pnas.0803466105>.
- Lynn S, Lai HT, Kao SM, Lai J, Jan KY. 1997. Cadmium inhibits DNA strand break rejoining in methyl methanesulfonate-treated CHO-K1 cells. *Toxicol Appl Pharmacol* 144(1):171–176, PMID: 9169081, <https://doi.org/10.1006/taap.1997.8116>.
- Pearson CG, Shikazono N, Thacker J, O'Neill P. 2004. Enhanced mutagenic potential of 8-oxo-7,8-dihydroguanine when present within a clustered DNA damage site. *Nucleic Acids Res* 32(1):263–270, PMID: 14715924, <https://doi.org/10.1093/nar/gkh150>.
- Plosky BS, Vidal AE, Fernández de Henestrosa AR, McLenigan MP, McDonald JP, Mead S, et al. 2006. Controlling the subcellular localization of DNA polymerases iota and eta via interactions with ubiquitin. *EMBO J* 25(12):2847–2855, PMID: 16763556, <https://doi.org/10.1038/sj.emboj.7601178>.
- Rajotte JW, Couture P. 2002. Effects of environmental metal contamination on the condition, swimming performance, and tissue metabolic capacities of wild yellow perch (*Perca flavescens*). *Can J Fish Aquat Sci* 59(8):1296–1304, <https://doi.org/10.1139/f02-095>.
- Rodriguez GP, Song JB, Crouse GF. 2013. In vivo bypass of 8-oxodG. *PLoS Genet* 9(8):e1003682, PMID: 23935538, <https://doi.org/10.1371/journal.pgen.1003682>.
- Shaw JR, Colbourne JK, Davey JC, Glaholt SP, Hampton TH, Chen CY, et al. 2007. Gene response profiles for *Daphnia pulex* exposed to the environmental stressor cadmium reveals novel crustacean metallothioneins. *BMC Genomics* 8:477, PMID: 18154678, <https://doi.org/10.1186/1471-2164-8-477>.
- Shaw JR, Colbourne JK, Glaholt SP, Turner E, Folt CL, Chen CY. 2019. Dynamics of cadmium acclimation in *Daphnia pulex*: linking fitness costs, cross-tolerance, and hyper-induction to metallothionein. *Environ Sci Technol* 53(24):14670–14678, PMID: 31738529, <https://doi.org/10.1021/acs.est.9b05006>.
- Shaw JR, Pfrender ME, Eads BD, Klaper R, Callaghan A, Sibly RM, et al. 2008. *Daphnia* as an emerging model for toxicogenomics. In: *Advances in Experimental Biology, vol.2*. Hogstrand C, Kille P, eds. Amsterdam: Elsevier, 165–328.

- Shin N, Cuenca L, Karthikraj R, Kannan K, Colaiacovo MP. 2019. Assessing effects of germline exposure to environmental toxicants by high-throughput screening in *C. elegans*. *PLoS Genet* 15(2):e1007975, PMID: 30763314, <https://doi.org/10.1371/journal.pgen.1007975>.
- Short PJ, McRae JF, Gallone G, Sifrim A, Won H, Geschwind DH, et al. 2018. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* 555(7698):611–616, PMID: 29562236, <https://doi.org/10.1038/nature25983>.
- Stephenson M, Mackie GL. 1988. Total cadmium concentrations in the water and littoral sediments of Central Ontario Lakes. *Water Air Soil Pollut* 38(1–2):121–136, <https://doi.org/10.1007/BF00279591>.
- Strepetkaitė D, Alzbutas G, Astromskas E, Lagunavičius A, Sabaliauskaitė R, Arbačiauskas K, et al. 2015. Analysis of DNA methylation and hydroxymethylation in the genome of crustacean *Daphnia pulex*. *Genes* 7(1):1, PMID: 26729172, <https://doi.org/10.3390/genes7010001>.
- Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012a. Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci U S A* 109(45):18488–18492, PMID: 23077252, <https://doi.org/10.1073/pnas.1216223109>.
- Sung W, Tucker AE, Doak TG, Choi E, Thomas WK, Lynch M. 2012b. Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. *Proc Natl Acad Sci USA* 109(47):19339–19344, PMID: 23129619, <https://doi.org/10.1073/pnas.1210663109>.
- Supek F, Lehner B. 2017. Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell* 170(3):534–547, PMID: 28753428, <https://doi.org/10.1016/j.cell.2017.07.003>.
- Supek F, Lehner B, Hajkova P, Warnecke T. 2014. Hydroxymethylated cytosines are associated with elevated C to G transversion rates. *PLoS Genet* 10(9):e1004585, PMID: 25211471, <https://doi.org/10.1371/journal.pgen.1004585>.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324(5929):930–935, PMID: 19372391, <https://doi.org/10.1126/science.1170116>.
- van Loon B, Hübscher U. 2009. An 8-oxo-guanine repair pathway coordinated by MUTYH glycosylase and DNA polymerase lambda. *Proc Natl Acad Sci U S A* 106(43):18201–18206, PMID: 19820168, <https://doi.org/10.1073/pnas.0907280106>.
- Xiong J, Liu XN, Cheng QY, Xiao S, Xia LX, Yuan BF, et al. 2017. Heavy metals induce decline of derivatives of 5-methylcytosine in both DNA and RNA of stem cells. *ACS Chem Biol* 12(6):1636–1643, PMID: 28448110, <https://doi.org/10.1021/acscchembio.7b00170>.
- 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 (Bethesda) 7(5):1405–1416, PMID: 28235826, <https://doi.org/10.1534/g3.116.038638>.