



High mutational rates of large-scale duplication and deletion in *Daphnia pulex*

Nathan Keith, Abraham E Tucker, Craig E Jackson, et al.

Genome Res. published online October 30, 2015

Access the most recent version at doi:[10.1101/gr.191338.115](https://doi.org/10.1101/gr.191338.115)

P<P Published online October 30, 2015 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Comprehensive immune receptor profiling.
Discover the **DriverMap™ AIR Assay** difference.

LEARN
MORE



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

High mutational rates of large-scale duplication and deletion in *Daphnia pulex*

Nathan Keith^{*1 †}, Abraham E. Tucker^{‡1}, Craig E. Jackson^{*}, Way Sung[§], José Ignacio Lucas Lledó[¶], Daniel R. Schrider^{**}, Sarah Schaack^{‡‡}, Jeffrey L. Dudycha^{¶¶}, Matthew Ackerman[§], Andrew J. Younge^{***}, Joseph R. Shaw^{§§2 †}, and Michael Lynch^{§2 †}

^{*}School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana 47405;

[‡]Biology Department, Southern Arkansas University, Magnolia, Arkansas 71753;

[§] Department of Biology, Indiana University, Bloomington, Indiana 47405;

[¶]Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany;

^{**}Department of Genetics, Rutgers University, Piscataway, NJ 08854;

^{‡‡}Biology Department, Reed College, Portland, OR 97202;

^{¶¶}Department of Biological Sciences, University of South Carolina, Columbia, SC 29208;

^{§§}School of Biosciences, University of Birmingham, Birmingham UK B15 2TT;

^{***}School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405;

¹These authors contributed equally to this work

²These authors jointly supervised this work

[†]To whom correspondence should be addressed: E-mail: rnkeith@umail.iu.edu, joeshaw@indiana.edu, milynch@indiana.edu

Running Title: The mutational processes of *Daphnia pulex*

Key words: Copy Number Variation, Mutation Rate, Mutator Phenotype

Abstract

Knowledge of the genome-wide rate and spectrum of mutations is necessary to understand the origin of disease and the genetic variation driving all evolutionary processes. Here, we provide a genome-wide analysis of the rate and spectrum of mutations obtained in two *Daphnia pulex* genotypes via separate mutation-accumulation (MA) experiments. Unlike most MA studies that utilize haploid, homozygous, or self-fertilizing lines, *D. pulex* can be propagated ameiotically while maintaining a naturally-heterozygous, diploid genome, allowing the capture of the full spectrum of genomic changes that arise in a heterozygous state. While base-substitution mutation rates are similar to those in other multicellular eukaryotes ($\sim 4 \times 10^{-9}$ per site per generation), we find that the rates of large-scale (>100 kb) *de novo* copy-number variants (CNVs) are significantly elevated relative to those seen in previous MA studies. The heterozygosity maintained in this experiment allowed for estimates of gene-conversion processes. While most of the conversion tract lengths we report are similar to those generated by meiotic processes, we also find larger tract lengths that are indicative of mitotic processes. Comparison of MA lines to natural isolates reveals that a majority of large-scale CNVs in natural populations are removed by purifying selection. The mutations observed here share similarities with disease-causing complex, large-scale CNVs, thereby demonstrating that MA studies in *D. pulex* serve as a system for studying the processes leading to such alterations.

Introduction

All aspects of evolutionary change ultimately depend on the rate of origin and molecular composition of mutations. An accurate characterization of the different types of mutations is therefore critical to our understanding of evolutionary rates, the origin of genetic disorders, and how signatures of natural selection are interpreted (Halligan and Keightley 2009). Mutation accumulation (MA) experiments combined with whole-genome sequencing have been shown to provide a powerful tool for surveying genome-wide mutation patterns. In an MA experiment, repeated bottlenecking of initially genetically identical lines reduces the efficiency of selection to a minimum, thereby enabling the capture of nearly all *de novo* mutations, with the exception of the small minority of mutations that cause immediate lethality and/or sterility (Kibota and Lynch 1996).

Although this type of experiment has enhanced our understanding of the rate and spectrum of spontaneous mutational changes across the Tree of Life (Denver et al. 2004; Lynch et al. 2008b; Denver et al. 2009; Haag-Liautard et al. 2008; Lynch 2010b; Conrad et al. 2011; Ness et al. 2012; Sung et al. 2012a), previous MA studies are limited by an inability to characterize the *de novo* rates of large-scale insertions and deletions that have mechanisms of origin and repair distinct from base-substitution mutations (Hastings et al. 2009). Furthermore, most MA experiments in eukaryotes have been confined to either haploid asexuals or obligately sexual species maintained as homozygous lines by extreme inbreeding, conditions that might influence various aspects of genome stability. For example, mutation assays involving inbred sexual lines reduce or eliminate the possibility of capturing genomic changes that are extremely deleterious or lethal in the homozygous state. Inbred lines of normally outcrossing species may also exhibit unusual mutational features owing to the exposure of segregating recessive factors influencing replication fidelity (Lynch 2008; 2010a). Finally, although the mutational processes from an asexual,

diploid, homozygous MA experiment have been investigated (Zhu et al. 2014) the lack of heterozygosity limited the ability to detect gene-conversion and/or internal genome homogenizing effects and the influence of these processes on genome composition (Nagylaki 1983; Lamb 1984).

One way to avoid these issues is to perform MA experiments with natural isolates of diploid species capable of asexual reproduction that can therefore maintain normal levels of heterozygosity. Lines propagated by single, clonally-produced offspring allow recessive deleterious mutations to accumulate in a heterozygous background. To this end, we performed deep-coverage, genome-wide next-generation sequencing on two long-term MA experiments using genotypes of the microcrustacean *Daphnia pulex* directly derived from natural environments and permanently harboring naturally occurring levels of heterozygosity. Most *D. pulex* alternate between bouts of sexual and asexual reproduction in nature (cyclical parthenogenic reproduction), but can still be maintained in culture in an exclusively clonal fashion (Hebert and Ward 1972; Lynch and Gabriel 1983). As a consequence of annual phases of sex, cyclical parthenogenic populations are generally in Hardy-Weinberg equilibrium like most other obligately sexual species (Lynch 1983).

A second unique aspect of *D. pulex* biology convenient for mutation analyses is that some natural isolates have completely lost the ability to reproduce sexually, hereafter referred to as obligate asexuals (Innes and Hebert 1988; Paland et al. 2005; Lynch et al. 2008a; Tucker et al. 2013). Because such genotypes accumulate mutations in nature in a perpetually clonal manner, a direct comparison between MA experiments and patterns of variation in natural asexual lineages has the potential to yield insight into the role of selection in either eliminating or promoting various classes of mutation (Tucker et al. 2013).

In this study, we provide joint, genome-wide estimates of the rates and spectra of base-substitution mutations, gene-conversion-like processes, and *de novo* copy number variants (CNVs), for both cyclically parthenogenetic and obligately asexual *D. pulex* genotypes. By comparing the asexual MA experiment to patterns of polymorphism found in asexual populations in a previous study, we also give insight into how selection is operating on *de novo* CNVs.

Results

To estimate mutation rates in *D. pulex*, we performed deep-coverage, next-generation sequencing on two separate MA experiments (average sequencing coverage of 36×, Table S1), whereby successive genetic bottlenecks of isogenic clonal sublines were performed in a naturally-occurring, heterozygous state: one experiment founded by an obligately asexual genotype from Linwood Lake, Ontario (ASEX) consisting of 4 sublines propagated for an average of 170 generations, and the other founded by a cyclically parthenogenetic genotype from Slimy Log Pond, Oregon (CYC), consisting of 3 sublines propagated an average of 85 generations.

Base-substitution mutation. After filtering low coverage, repeat, and gap regions from the whole-genome alignments, and applying strict criteria for base quality and alignment scores (Materials and Methods), we were able to screen 30.1% and 37.2% of the *D. pulex* genome for base-substitution mutations in the ASEX and CYC MA lines, respectively. We identified 733 base-substitution mutations, yielding genome-wide base-substitution mutation rates of 1.02×10^{-8} (SE=0.80) for ASEX and 4.33×10^{-9} (SE=0.34) / site / generation for CYC (Table 1, Fig. 1).

It is important to note the higher rate of base-substitution mutation in the ASEX experiment is driven by a possible mutator phenotype in one of the four ASEX sublines (ASEX-12, Table S1). The base-substitution mutation rate in ASEX-12 was 9.2× greater than the average of the other ASEX sublines. For ASEX-12, we analyzed the functional categories (COG database, Tatusov et al. 2003) for genes in which non-synonymous mutations occurred in the experiment, with the reasoning that a subline-specific mutation(s) may be driving the increased base-substitution rate. After surveying 27 genes with non-synonymous mutations, only one gene was regarded as having possible influence in DNA replication and repair, the replication factor C (*RFC*) large subunit (Table S3, Fig. S1). Replication factor C, a DNA clamp-loader protein essential for DNA replication and repair, catalyzes the opening of the DNA clamp, PCNA, allowing PCNA to encircle the DNA strand and recruit mismatch repair (MMR)-associated proteins (McAlear et al. 1996). We observe two non-synonymous mutations in the large subunit of *RFC*.

A knock-out study of the *RFC* large subunit in yeast resulted in mutator phenotypes (Xie et al. 1999). Although these two mutations are not in the DNA or PCNA binding domains (Fig. S1), the observed hydrophilic (P) → hydrophobic (Q) replacement and a hydrophilic (E) → positively charged (K) replacement occurring in close proximity (11 amino acids) could lead to loss or reduced function, thereby providing an explanation for the increased mutation rate of ASEX-12. If we remove this outlier subline, the ASEX rate of mutation is $3.34 \text{ (SE = 0.49)} \times 10^{-9} \text{ / site / generation}$, which is lower than the observed rate in CYC ($4.53 \times 10^{-9} \text{ / site / generation}$).

As in numerous previous experiments directly measuring mutation rates in various eukaryotes, we observed an appreciable number of single nucleotide mutations separated by <50 bp (Schrider et al. 2011; Campbell et al. 2012; Michaelson et al. 2012; Schrider et al. 2013; Zhu et al. 2014; Peters et al. 2015). In particular, 9 mutations within the CYC lines were found within 50 bp of another mutation (representing 13% of base changes in these lines), and 76 such mutations were found in the ASEX lines (32% of ASEX mutations). These clusters represent multinucleotide mutations (MNMs; Schrider et al. 2011), in which multiple bases are changed simultaneously, perhaps by an error-prone translesion polymerase (Harris and Nielsen 2014). The majority of these clusters (80%) are found in ASEX-12, suggesting they may be another consequence of the mutator allele we propose is present in that line. In ASEX, only 16% of mutations are MNMs when excluding ASEX-12. We also note that while the most commonly observed MNM is a pair of adjacent mutations, only one such pair is present in our data, and therefore most MNMs we observe are similar to what we observed in the *RFC* large subunit of ASEX-12 (Fig. S1, Table S22). Thus, it is unclear whether these putative MNMs were created by the same process as in other species, or if a different mechanism is at play.

Gene-conversion-like processes. We report gene-conversion processes, which we define as all genome-homogenizing processes with the exception of deletions that result in loss-of-heterozygosity (LOH). We estimated the rates of LOH via gene-conversion-like processes to be $1.18 (\text{SE}=0.39) \times 10^{-5}$ and $1.14 (0.010) \times 10^{-6}$ / site / generation in the ASEX and CYC lines, respectively (Table S4). We assigned transmission to either A/T or G/C at each converted AT/GC heterozygous position, and observed that converted sites were significantly biased toward GC alleles (one-sample proportion test, as described in Lesecque et al. 2013). Genome-wide, 54.2% ($P < 10^{-6}$, 95% CI [0.525, 0.560]) of AT/GC heterozygous

sites were converted to G or C, which is slightly higher than a genome-wide *de novo* estimate in yeast (51.3% GC-bias, $P < 10^{-3}$, Mancera et al. 2008). We utilized a maximum likelihood model to estimate average conversion tract lengths, and found average tract lengths for ASEX were 1358 bp (95% CI [1358, 1590]), 236 bp (95% CI [187, 302]), 197 bp (95% CI [166, 234]), and 28.3 bp (95% CI [28.2, 28.4]), for ASEX-2, ASEX-12, ASEX-41, and ASEX-48, respectively (Fig. 2A, 2B, S2). For the CYC experiment tract length estimates were 84.2 bp (95% CI [83.5, 85.0]), 48.1 (95% CI [48.0, 48.4]), and 65.6 bp (95% CI [65.1, 66.0]) for CYC-17, CYC-237, and CYC-97, respectively (Table S9).

***de novo* CNV analysis.** We also observed many genomic duplication and deletion mutations. The rates of LOH (per heterozygous “site”) resulting from spontaneous deletion were $4.06 \text{ (SE=1.33)} \times 10^{-5}$ and $1.16 \text{ (SE=0.99)} \times 10^{-5}$ /site/generation for ASEX and CYC, respectively, comparable to a previous study in *D. pulex* (2-30 kb = 6.7×10^{-5} /site/generation, Xu et al. 2011). We also observed many large-scale *de novo* CNVs (which we define as indels > 0.1 Mb, Figs. 3, 4, Tables S5). In ASEX there were seven large-scale duplications, ranging from 0.10 to 1.35 Mb, and thirteen large-scale deletions, ranging from 0.10 to 1.01 Mb. In contrast in the CYC MA lines we observed only one large-scale deletion of 0.40 Mb resulting from a highly complex mutational event (Fig. 5), and a single large-scale duplication of 0.35 Mb. It is important to note that CYC was propagated for fewer generations, which results in only a four-fold difference in deletion rate between ASEX and CYC despite the presence of 12 more large-scale CNVs in ASEX. We also found a significant enrichment of these mutations on asexual-conferring Chromosomes XIII and IX in the ASEX MA experiment (8/20 *de novo* CNVs > 100 kb on 13 scaffolds that comprise Chromosomes XIII and IX, out of 200 scaffolds total; $P < 4.0 \times 10^{-5}$, Fisher’s exact test). Twelve large-scale

CNVs were randomly selected for validation by real-time quantitative PCR and all were subsequently verified (Tables S6, S7).

De novo CNV rate analysis. In the ASEX MA experiment, copy-number change was observed for 722 single-copy genes, via 328 gene duplications and 394 gene deletions, yielding gene duplication and deletion rates of $3.27 \text{ (SE=1.03)} \times 10^{-5}$ and $3.71 \text{ (SE=1.09)} \times 10^{-5}$ / single-copy gene / generation, respectively (see Materials and Methods). In the CYC MA experiment, copy-number change was observed for 153 single copy genes, via 58 duplications and 95 deletions, leading to gene duplication and deletion rates of $1.06 \text{ (SE=0.87)} \times 10^{-5}$ and $1.75 \text{ (1.43)} \times 10^{-5}$ / single-copy gene / generation, respectively (Table 2). We also looked at the rates of duplication and deletion for indels greater than 3kb (duplication or deletion / bp or 'site' / generation). The rates of spontaneous deletion are 7.6× higher in ASEX, and the rate of duplication is 5.0× higher in ASEX relative to CYC (Table S10).

Copy-number variation and selection. Asexual *D. pulex* lineages provide an opportunity for studying genomic consequences of natural selection because a specific, asexual-linked Chromosome XIII and IX haplotype confers obligate asexuality when introgressed into a cyclical parthenogenic background by asexual males that undergo sporadic meiosis (Tucker et al. 2013). Therefore, as this shared haplotype that confers the origin of asexual lineages accumulates mutations in a lineage-specific fashion, it provides an estimate of the ages of asexual lineages (Tucker et al. 2013). We applied the same method (See Supplemental Materials and Methods) to re-estimate the ages of extant asexual lineages from across North America with our revised mutation rate estimates, and used these revised estimates of lineage age to estimate the total number of base pairs we would expect to undergo CNV in

the absence of selection. We then compared this null estimate to the observed number of base pairs subject to CNV on these chromosomes in natural populations. Tucker *et al.* (2013) concluded that base-substitutions were accumulating in a neutral fashion on the asexual haplotype (d_N/d_S , or the ratio of nonsynonymous base-substitutions per site to the synonymous base-substitutions per site = 0.94, with a ratio of 1.0 implying neutral evolution). However, after comparing the total accumulation of *de novo* CNVs (> 3 kb) in natural populations to the null expectation derived from the MA experiments (CNVs > 3 kb) there are 19× less base pairs subject to *de novo* CNVs than would be expected in the absence of selection (Table S8). Although the disparity in the rates that CNVs are accumulating between natural populations and ASEX could be explained by a single ASEX subline with many large-scale CNVs (and none in the sublines) it is important to note that we observed these mutations in all sublines for ASEX (Table S5). These results suggest that selection is acting differently on these two different classes of mutations.

We further found that while large-scale *de novo* CNVs (i.e. indels > 0.1 Mb) commonly accumulate on XIII and IX in the ASEX MA experiment, no large-scale *de novo* CNVs were found on this haplotype in natural populations (Supplemental Methods; Table S4), suggesting that purifying selection against these large-scale mutations is in part leading to the discrepancy between observed and expected *de novo* CNV.

Discussion

MA studies have substantially increased our knowledge of the rate and molecular spectrum of mutations across cellular life, but the *de novo* rates of certain mutational classes still remain a mystery. This study allows the ascertainment of classes of mutations that until now have been unobservable, most notably the *de novo* rates of large-scale CNV and gene-conversion-like processes. While the base-substitution mutation rates in *Daphnia* are

similar to those reported for other species, we discovered specific mutations that potentially underlie a mutator phenotype. We observe high rates of gene duplication and deletion driven by large-scale *de novo* CNVs (>100 kb). We also report average conversion-like tract lengths for all sublines. Finally, in regions where lineage-specific *de novo* CNVs can be studied in asexual populations, we find that CNVs are accumulating at lower rates in natural populations when compared to estimates derived from this MA experiment.

The *D. pulex* base-substitution mutation rate (avg. $\approx 3.8 \times 10^{-9}$ /site/generation) is similar to those reported for *Caenorhabditis elegans* (Denver et al. 2009), *Arabidopsis thaliana* (Ossowski et al. 2010), and *Drosophila melanogaster* (Keightley et al. 2009; Schrider et al. 2013), but higher than estimates in unicellular eukaryotes (Lynch et al. 2008b; Ness et al. 2012; Sung et al. 2012b; Schrider et al. 2013; Zhu et al. 2014) and lower than estimates in humans (Campbell et al. 2012) (Table 1). Transitions arise more frequently than transversions (1.52:1), as observed in previous MA studies. If we consider the conditional mutation rates, i.e. the number of each type of base-substitution mutation normalized by the genome-wide G+C and A+T composition (Lynch 2007) (Fig. 1, Table S2), the rate of G/C $\rightarrow$ A/T mutations is 2.5-fold higher than A/T $\rightarrow$ G/C mutations, a bias observed in previous studies (Denver et al. 2004; Lynch et al. 2008b; Keightley et al. 2009; Campbell et al. 2012; Sung et al. 2012b; Zhu et al. 2014). Based on this A+T bias, we would expect an equilibrium A+T composition of 72%. With an observed A+T composition of 59%, the genome composition of *D. pulex* cannot be explained by base-substitution forces alone, and opposing forces such as GC-biased gene conversion may be countering the observed mutational pressure to A+T. Taking into account the GC-bias of conversion we report (54.2%, $P < 10^{-6}$, 95% CI [0.525, 0.560]), the expected equilibrium A+T composition is *D. pulex* is actually 2% A+T. Even if we consider the lower bound of the confidence interval for G/C bias, we would still expect an equilibrium A+T composition of 11%.

In previous MA studies the A/T bias we report has been attributed to oxidation of 8-oxo-guanine (Duncan and Miller 1980) or the spontaneous deamination of 5-methylcytosine, which is a main source of C:G → T:A transitions (Friedberg 2003). As 5-methylcytosine is often found in CpG islands, we investigated whether CpG context is influencing the GC/AT bias via C:G → T:A transitions by comparing CpG context mutations vs. all other genomic contexts (as detailed in Sung et al. 2015, Tables S13-S21). If methylation at CpG islands was driving the observed mutational bias via deamination, we would expect an enrichment of 5'-X[C/G→X]G/C-3' and 5'-G/C[C/G→X]X-3' mutations (i.e., CpG context mutations) relative to the contexts of other mutation triplets (where X is either A,T, G or C). Overall, we observe that both ASEX and CYC are not significantly enriched for CpG context mutations (Fisher's exact test, $P < 0.05$, Tables S13, S19). However, the ASEX experiment was only marginally non-significant ($P = 0.06$ for ASEX compared to $P = 0.46$ for CYC). We therefore investigated CpG context in each of the seven ASEX and CYC sublines, and determined that there was one outlier that was significantly enriched with CpG context mutations, the mutator phenotype ASEX-12 ($P = 0.04$, Table S15).

We also found that the enrichment of CpG context mutations in ASEX-12 is driven by a significant enrichment of C:G → T:A transitions relative to the other mutational outcomes. We discovered that 45/71 mutations are C:G → T:A transitions and these transitions represent 1/3 of possible mutations from a CpG context ($\chi^2=27.5$, $N = 71$, $P = 1.56 \times 10^{-7}$). Although, taken together, these results suggest germline methylation of cytosine at CpG contexts results in C:G → T:A mismatches that are normally repaired by MMR processes, future studies are needed to investigate the germline state of methylation in *D. pulex*.

Previous *D. pulex* studies have suggested that LOH reduces genomic diversity at much higher rates than the input of variation by base-substitution mutation possibly contributing to the demise of obligately asexual lineages (Omilian et al. 2007; Tucker et al.

2013). In ASEX, which has a level of heterozygosity that is representative of North American *D. pulex* (Tucker et al. 2013), LOH via gene-conversion processes occur ~17-fold more often than gain-of-heterozygosity via base-substitutions. When also considering LOH due to deletion, LOH arises ~75-fold more often than gain-of-heterozygosity, supporting previous claims that LOH generates substantially more genomic input than base-substitution mutations in *D. pulex*.

We also report average tract lengths for both the ASEX and CYC experiment. Meiotic conversion tracts are relatively small (>1 kb) compared to mitotic conversion tracts, which are generally 1 to many kb (Jeffreys and May 2004; Chen et al. 2007). In *Daphnia*, parthenogenesis occurs via abortive meiosis (Hiruta et al. 2010) and conversion-like processes could be expected to show a mixture of tract lengths due to both meiotic and mitotic processes. We provide evidence suggesting both meiotic-like and mitotic-like conversion processes are operating. While for the ASEX experiment the average tract length is similar to that expected from meiotic conversion processes (median = 221 bp, Fig. 2B), ASEX-2 has a significantly higher average tract length (~1.5 kb, Fig. S2, Table S9) compared to the other ASEX and CYC sublines. The majority of ASEX-2 tracts (293 out of 339 gene-conversion-like events) reside in three large-scale regions of concentrated conversion. These concentrated regions of conversion tracts are ~300 kb (scaffold 13), ~125 kb (scaffold 14), and ~400 kb (scaffold 51), and are directly flanked by large-scale duplications (e.g. Fig. 2C). These regions are not observed in the other ASEX, or CYC sublines. While in the ASEX experiment the overall median tract length is 221 bp, the median tract lengths in these regions are 314 bp (scaffold 13), 1381 (scaffold 14), and 1446 (scaffold 51), suggesting these complex regions of larger conversion tracts are the result of mitotic and/or replication-based repair processes that are driving the increased average tract length in ASEX-2.

The combined rate of gene copy number change by either duplication or deletion in *D. pulex* (average of $\sim 5.4 \times 10^{-5}$ / single-copy gene / generation) is higher than direct estimates from previous MA experiments: *Drosophila melanogaster* (1.1×10^{-6} /gene/generation (Schrider et al. 2013)), *Caenorhabditis elegans* (3.4×10^{-7} /gene/generation (Lipinski et al. 2011)), and *Saccharomyces cerevisiae* (5.5×10^{-6} dup./gene/generation (no gene deletion estimate) (Lynch et al. 2008b)). It is important to note that our study is limited to duplication and deletions of single-copy genes to eliminate biases that arise when reads map to multiple locations and to control for genetic differences between isolates. Genes residing in clusters of tandem duplicates, which account for half of the reported gene duplicates in the *D. pulex* assembly (Colbourne et al. 2011), may undergo higher rates of copy number change due to an increased probability of unequal crossing over during recombination (Read et al. 2004). Thus, the total rate of gene duplication and deletion in *D. pulex* may be higher than the estimates we provide.

The difference in mutation rates between ASEX and CYC is also of interest. Relative to the CYC lines, in the ASEX experiment, the base-substitution mutation rate is 2.4-fold greater (although the rates are similar if the candidate mutator phenotype, ASEX-12, is removed from the mutation-rate calculation), the combined rate of gene duplication and deletion is ~ 3 -fold higher, duplications (> 3 kb) arise 5 $\times$ more often, deletions (> 3 kb) arise 7 $\times$ more often, and the rate of gene-conversion processes is ~ 10.4 -fold greater. However, we are hesitant to conclude that there are differences in mutation rates between all asexual and cyclically parthenogenetic *D. pulex*, as the small sample size of this experiment ($N=1$ for asexuals and sexuals) does not justify such a conclusion, and the differences we observe are well within the variation observed among individuals within the same species (Conrad et al. 2011; Schrider et al. 2013). Future studies are needed to determine if mutation rates are higher in asexual *D. pulex* populations.

As observed in a previous, yeast MA experiment (Zhu et al. 2014), by maintaining *D. pulex* asexually and in a heterozygous state, we are able to study large-scale CNVs that are typically lost due to complementation of recessive lethal mutations in experiments that utilize sexual reproduction. We find that the rate of large-scale CNVs is 1.5×10^{-2} events per haploid genome per generation, higher than the rates observed in an asexually propagated yeast MA study (3.2×10^{-6} events per haploid genome per generation, Zhu et al. 2014), the only previous MA study reporting such *de novo* mutations. When considering the number of nucleotides in the two genomes (events per bp per generation), large-scale *de novo* CNVs arise $\sim 5000\times$ more often in *D. pulex*.

The over-representation of CNVs on the asexual-conferring haplotype (Chromosomes XIII and IX) is intriguing, as this haplotype was originally introgressed from a divergent sister-species, *D. pulicaria* (origin of asexuality $\sim 1,250$ years before present, Tucker et al. 2013). Because of this foreign origin, there is $2\times$ the silent-site heterozygosity on Chromosomes XIII and IX compared to all other genomic regions in asexuals (synonymous-site heterozygosity, or $\Pi_s = 0.0442$ vs. 0.0223 per site, Tucker et al. 2013), suggesting that chromosomal divergence is contributing to the increased mutation rates in this region. It has been observed that mutations accumulate at higher rates in F1 hybrids (Yang et al. 2015; Scavetta and Tautz 2010), and repeat regions are more likely to change in length if the homologous sequence has a drastically divergent length (Amos et al. 1996). However, perhaps most pertinent to this study is that mutations have been shown to arise at a higher rate in heterozygous regions of genomes that have interspersed homozygous and heterozygous domains (Yang et al. 2015). Future studies are needed to fully understand the mechanistic phenomena leading to the observed changes on *D. pulex* Chromosomes XIII and IX in asexual lines.

Because large changes in the order, orientation, and copy-number of genomic elements can affect transcriptional regulators and gene dosage, large-scale CNVs can be highly deleterious and may be subject to higher levels of purifying selection than single base mutations (Itsara et al. 2010). Furthermore, large-scale CNVs have been associated with debilitating neurodevelopmental disorders and genetic diseases in humans (Zhang 2003; Sebat et al. 2007; Stone et al. 2008; Girirajan et al. 2011). Thus, while Tucker *et al.* reported no purifying selection operating on base-substitutions in young asexual lineages (based on the ratio of non-synonymous base-substitution mutations / non-synonymous site to synonymous base-substitutions / synonymous site, $d_N/d_S=0.94$; with a ratio of 1.0 corresponding to neutral base-substitution accumulation), we find that large-scale CNVs are being removed in asexual *Daphnia* populations (Table S8). If we consider that on average each large-scale CNV (> 100 kb) reported in this MA experiment changes the copy-number of 15.87 genes, their observed removal in the natural populations we examined is consistent with the idea that large-scale CNVs have a detrimental effect on the fitness of an organism, and that they are therefore often removed by purifying selection. Taken together, these findings show the importance of large-scale CNVs on genome stability, and that *Daphnia* is an ideal model to study the rate and consequences of all mutational classes.

Materials and Methods

MA line propagation, and DNA sequencing procedures. *D. pulex* MA lines were initiated from two independent genotypes, one from Linwood, Ontario (ASEX) and the other from Slimy Log Pond, Oregon (CYC). For both experiments, a single individual was grown up into an isogenic (clonal) population, separated into sublines, each maintained under the following conditions: 100 ml of filtered lake water, stored at 20°C, 12 hour on:12 hour off, light:dark environment, and fed *ad libitum* with *Scenedesmus obliquus*. Every generation, a

randomly chosen daughter was selected to propagate the next generation, and back-ups were maintained to restore lineages in the event of mortality, production of males only, overproduction of ephippia, or operator error. Over the course of the experiment, each replicate MA line was propagated through single-progeny descent, reducing the effective population size to approximately one, minimizing the effect of natural selection, and accumulating all but the most deleterious mutations (Kibota and Lynch 1996). At the end of the experiment, a total of 8 randomly selected MA lines (four from each of the two genotypes) were chosen for whole-genome sequencing from the remaining sublines.

DNA was extracted using CTAB buffer and purified using phenol/chloroform in phase-lock tubes. In order to collect enough DNA for sequencing (1-10 µg), a single female from each of the randomly chosen MA lines was clonally propagated to hundreds of individuals. Standard Illumina paired-end libraries were prepared at the Center for Genomic and Bioinformatics at Indiana University (CYC) and BGI (ASEX), sequenced, and mapped to the *D. pulex* genome reference assembly using Novoalign (Novocraft Technologies). The reference genome alignments from each of the eight lines were converted to SAM format, and analyzed using SAMtools (Li et al. 2009). Final sequence-coverage depth for mapped reads ranged from 18 to 55× (Table S5) across the 8 lines.

Identifying *de novo* copy number variation and gene duplication and deletion.

Our study is limited to duplications and deletions of single-copy genes to eliminate biases that arise when reads map to multiple locations and to control for genetic differences between isolates. Genes residing in clusters of tandem duplicates, which account for half of the reported gene duplicates in the *D. pulex* assembly (Colbourne et al. 2011), may undergo higher rates of copy number change due to an increased probability of unequal crossing over during recombination (Read et al. 2004). Therefore, we restricted our analysis of the

rates of gene duplication and deletion to uniquely-mapped sequences within the 200 largest scaffolds of the reference. This approach is able to analyze duplication and deletion across 10,521 single-copy (Colbourne et al. 2011) genes in the reference.

To detect *de novo* CNVs, defined as regions that had either been duplicated or deleted throughout the MA experiment, a sliding window approach via CNV-seq software (Xie and Tammi 2009) was utilized for all pairwise comparisons of the ASEX and CYC MA lines, independently. To identify CNVs, we required a minimum of four consecutive 500-bp windows (sliding in 250-bp increments to have a significantly different depth of coverage deviating from a 1:1 ratio, (p-value = 0.001, $\log_2$ -threshold = 0.45). After genome-wide CNV regions were identified for all pairwise comparisons, we then identified annotated *D. pulex* genes located within these CNV regions. Each gene was classified as either a deletion or duplication and compared against the average depth of coverage across all other sublines for each identified gene, and for each experiment (ASEX and CYC), independently. If a gene within a CNV region for a given line had an increased depth of coverage relative to the other lines, this gene was categorized as a gene-duplication candidate, while a relative decrease in the depth of coverage was categorized as a gene deletion candidate because of its co-occurrence with a CNV region. We then filtered out duplication and deletion candidate genes without depth of coverage ratios supporting a duplication or deletion (deletion depth of coverage ratio ≤ 0.6 , duplication depth of coverage ratio ≥ 1.4 , as using strict ratios of 0.5 for deletion and 1.5 for duplication cause many duplications and deletions to be filtered out), leaving us with high-confidence gene duplication and deletions located within CNV regions > 1000 bp. Although some previous studies (Mills et al. 2011; Alkan et al. 2011), have defined copy number variants as indels > 50 bp, we did not use a read-pair approach (as described in Mills et al. 2011; Alkan et al. 2011, Fig. 2) because of the divergence of the ASEX progenitor from the *D. pulex* reference genome. Thus, our CNV minimum length cutoff

was chosen to return high confidence CNV calls with a depth of coverage method by joining multiple 500 bp windows, with each window having a significant difference in average depth of coverage.

For each line duplication and deletion rates were estimated with $\mu_g = n/(2i \times T)$, where n is the number of gene duplications or deletions within a line, i is the total number of single-copy genes that were analyzed for copy number change, and T is the total number of generations the line was propagated. Due to differences in mapping efficiency (given the genetic divergence between ASEX and the reference), we reasoned that i could differ between ASEX and CYC, and we filtered single-copy genes with low coverage for each experiment, independently. The standard error (S.E.) of the gene duplication and deletion rates were calculated for individual sublines independently, with the formula $SE_x = \sqrt{\mu_g / 2n_g T}$, where μ_g is the rate of gene duplication or deletion, n_g is the number of analyzed single-copy genes, and T is the number of generations (Lynch et al. 2008b; Denver et al. 2009; Ossowski et al. 2010). The overall S.E. across all sublines was estimated with $SE_{pooledCNV} = s/\sqrt{N-1}$, where (s) is the standard deviation of the gene duplication or deletion rates, and N is the number of sublines.

We also calculated rates of duplication and deletion based on events large enough to duplicate or delete a *Daphnia* gene (duplications and deletions > 3 kb). The new formula is $\mu = n/(2i \times T)$, where n is the number of de novo CNVs (duplication or deletion > 3 kb), i is the total number bps analyzed, and T is the total number of generations for each subline, arriving at a rate of duplication or deletion / bp / generation.

To address the issues of non-unique and low coverage genomic regions, which hinder our ability to accurately estimate CNV lengths based on our sliding window approach by interrupting CNV events, by utilizing an agglomerative, hierarchical clustering algorithm for all CNV-seq pairwise comparisons. We then merged CNV calls less than 10 kb apart, which

accounts for the lengths of 95% of assembly gaps and low complexity (repeat) regions in the *D. pulex* reference genome. This method was applied to natural asexual populations and the ASEX and CYC experiment to assign CNVs lengths for all pairwise comparisons. To assign lineage-specific, or MA subline-specific CNVs, we used the output CNV regions from our pairwise-comparison after clustering, and computed the intersection of all pairwise CNV comparisons using BEDOPS (Neph *et al.* 2012).

CNV Validations. We performed RT-PCR validation for ten randomly-selected *de novo* CNVs for ASEX sub-lines where genomic DNA was available after sequencing. RT-PCR was performed using Roche Lightcycler 480 DNA SYBR Green I Master Kit on the LightCycler 480 (Roche) hardware, and reactions included 10 ng of DNA, 0.5 mM of each primer, and 10 ml of 1X SYBR Green I Master mix, which includes reaction buffer, MgCl₂, dNTP mix, FastStart Taq DNA polymerase, and SYBR Green I Dye.

Each Roche Lightcycler 384-well clear qPCR plate (cat# 05102430001) contained: 1) three technical replicates of two randomly selected genes residing within a candidate *de novo* CNV for the line where the CNV occurred, and three technical replicates for these two genes in a 'control' line where no candidate *de novo* CNV was observed. We used ASEX-48 as control for all CNVs, as only two *de novo* CNV were observed, with no CNV overlap with other sublines. 2) Standard curve, replicated three times per gene, with one standard curve per gene, dilution series from 0.78 – 50 ng, and 3) negative controls which consisted of no template control, no mastermix control, and no primer control.

Samples, standards, and controls were then randomly assigned to wells (<http://www.random.org/>) to avoid any spatial biases that may arise on the LightCycler plate. RT-PCR cycling conditions followed the manufacturer's recommendations and included an initial denaturation at 95°C followed by 40 cycles of denaturing at 95°C for 10

seconds, annealing for 20 seconds, and extension at 72°C for 30 seconds. Melting curve analysis was generated by running the RT-PCR amplicons through one cycle of 95°C for 5 sec, then 65°C for 1 min, followed by a ramp up to 97°C.

Data Access. The data in this manuscript has been submitted to the NCBI BioProject (<http://www.ncbi.nlm.nih.gov/bioproject>) under accession number PRJNA275628.

Acknowledgments. We thank J.F. Gout, S. Reynolds, S. Xu, S. Glaholt and M. Hahn for helpful discussions, and G. Fox for computational support and expertise. E. Williams and M. Westphal maintained the MA lines. Funding was provided by grants from National Institutes of Health to M. L. (R01GM036827 and R01GM101672-01A1), the Murdock Charitable Trust and the National Science Foundation to S.S. (MCB-1150213) and from the National Institute of Environmental Health Sciences to J. S. (R01ES019324).

The authors declare no conflict of interest.

Figure Legends

Fig. 1. Conditional base-substitution rates for both the ASEX and CYC experiments (per site per generation). For each type of base-substitution mutation, $X \rightarrow Y$ denotes the conditional mutation rate per site per generation ($\times 10^{-9}$, y-axis) from nucleotide X to Y.

Fig. 2. Fig. 2A Parallel coordinate visualization of the minimum and maximum lengths for conversion tracts. The minimum (min) and maximum (max) possible lengths for each event are shown on the x-axis, with the tract length (bp) shown along the y-axis. Fig. 2B. Frequency distribution of tract lengths. Fig. 2B visualizes the size distribution of the midpoint of the minimum and maximum tract lengths for all 879 conversion tracts in the ASEX experiment (see Supplemental Materials and Methods). Vertical bars group conversion-like tracts by size in 100 bp increments. The red vertical dashed line denotes the median tract length (221 bp). Fig 2C. Complex region of large-scale duplication directly flanking a concentrated region of gene-conversion region. Black dots (top track) represent $-\log_{10}P$ for each consecutive sliding window on ~900 kb of scaffold 51 in ASEX-2, and denotes the probability the normalized depth of sequencing coverage deviates from 1:1 by chance between the MA subline where large-scale duplication occurred and the normalized average coverage of the other sublines (See Materials and Methods). On the middle track (red and blue) the normalized, $\log_2$ coverage ratio of the subline where the *de novo* CNV mutation versus a composite of the other three subline comparisons is plotted in 500 bp windows, sliding 250 bp increments. The 'LOH Sites' track represents a density plot of LOH positions (i.e. individual nucleotide positions of LOH), while the bottom track (Gene-conversion Tracts) shows density of gene-conversion-like tracts within the region. The conversion region begins within 500 bp of the large-scale duplication. Note how the far 3' end of the conversion region has intervening regions of deletion, starting ~730 kb, indicating further mutational complexity.

Fig. 3: Large-scale duplication and deletion. Fig. 3a depicts a large-scale duplication, while Fig. 3b shows a large-scale deletion. Black dots (top track on each figure) represent $-\log_{10}P$ for each consecutive sliding window on scaffold 8 (Fig. 3a) and scaffold 17 (2b) from the *D. pulex* reference assembly, where P denotes the probability the depth of sequencing coverage deviates from 1:1 by chance between the MA subline where the *de novo* CNV occurred and other sublines (See Materials and Methods). On the middle track in 3a (blue, duplication) and 3b (red, deletion) the normalized, $\log_2$ coverage ratio of the subline where the *de novo* CNV mutation versus a composite of the other three subline comparisons is plotted for both scaffolds in 500 bp windows, sliding 250 bp increments (see Materials and Methods). The black arrow below the normalized coverage tracts denotes all *D.pulex* full-gene models nested within each duplication and deletion. Note that the separate scaffolds of these two events are not the same length, so the deletion (310 kb) and duplication (1Mb) are not to scale. Large-scale CNVs were subsequently validated by RT-qPCR.

Fig. 4: Size spectrum of *de novo* duplications and deletions (indels) > 3000 bp. Fig. 4A shows all indel sizes for both the ASEX and CYC experiments > 3 kb, while Fig. 4B shows only large-scale *de novo* indels > 100 kb, split into insertions and deletions.

Fig. 5. Complex mutation event leading to multiple duplications and deletions in the MA subline 'CYC-17'. Arrows denote the left and right Illumina read pairs (with arrows not signifying left or right, specifically), and provide evidence for complex *de novo* event on Scaffold 18 in CYC-17, including a large-scale deletion (center of figure) and multiple smaller duplications and deletions. While distance between read pairs should only be a few hundred base pairs, the arrows show read pairs with drastic differences in mapping distance, and are indicative of indel activity. The sequence coverage track represents the normalized depth of sequencing coverage, relative to the other CYC MA sublines, where lighter color represents lower relative coverage (e.g. the 401 kb deletion). Also, $-\log P$ is plotted for each consecutive sliding window (i.e. the probability of a region deviating from 1:1 coverage ratios by chance, or the probability of an indel), and higher values relative to lower values represent increasing probability of CNV (see Materials and Methods).

575

576

577

578

579

580

581

582

583

584

585

586

587

Figures

Fig. 1

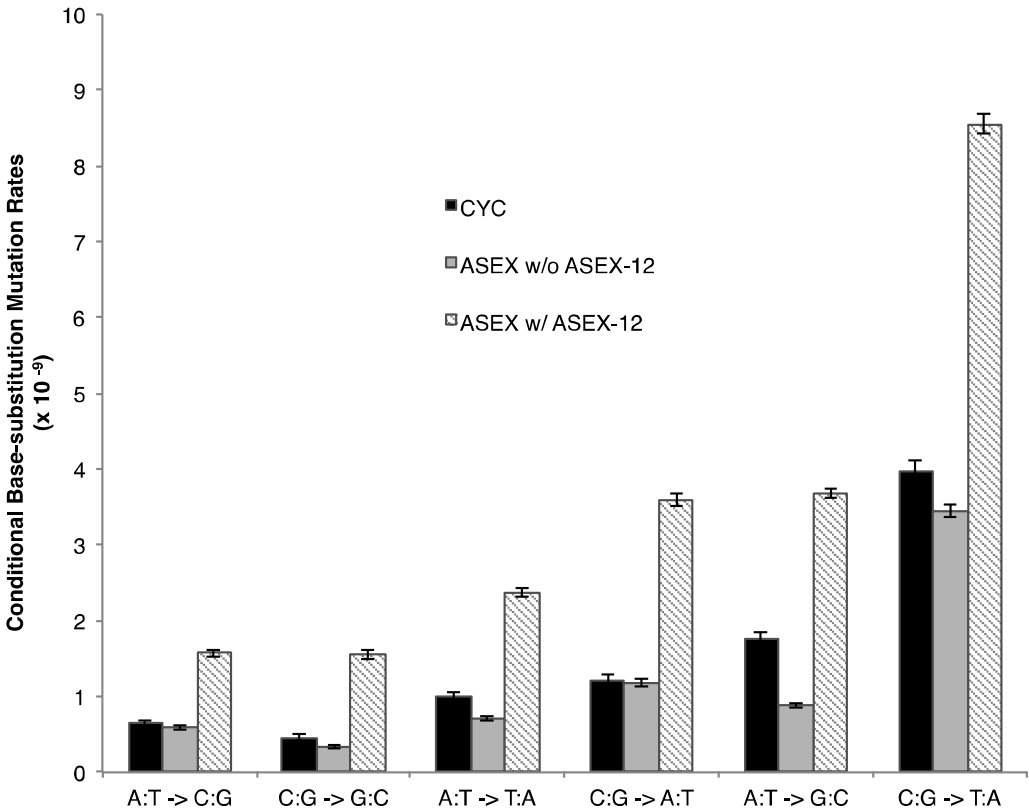


Fig. 2

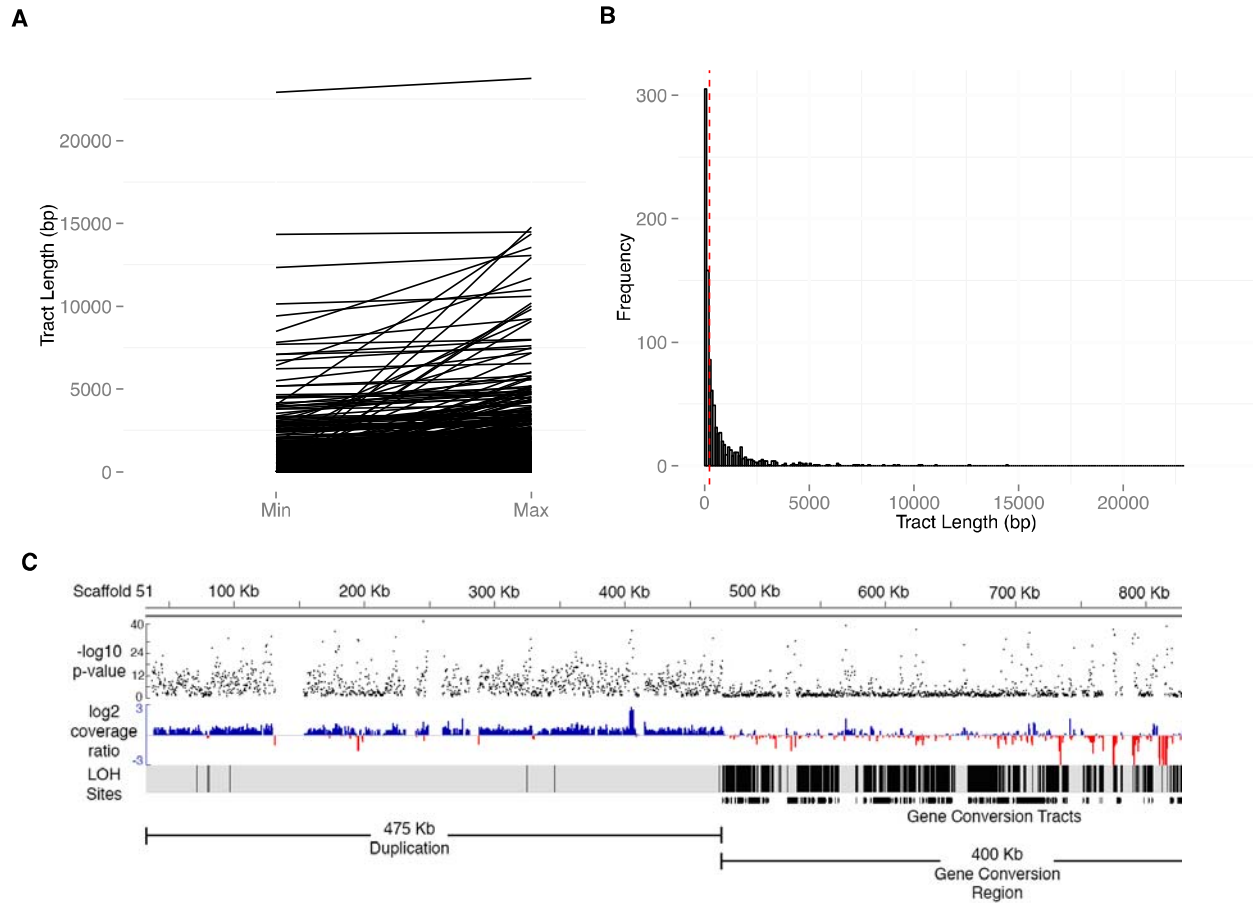


Fig. 3

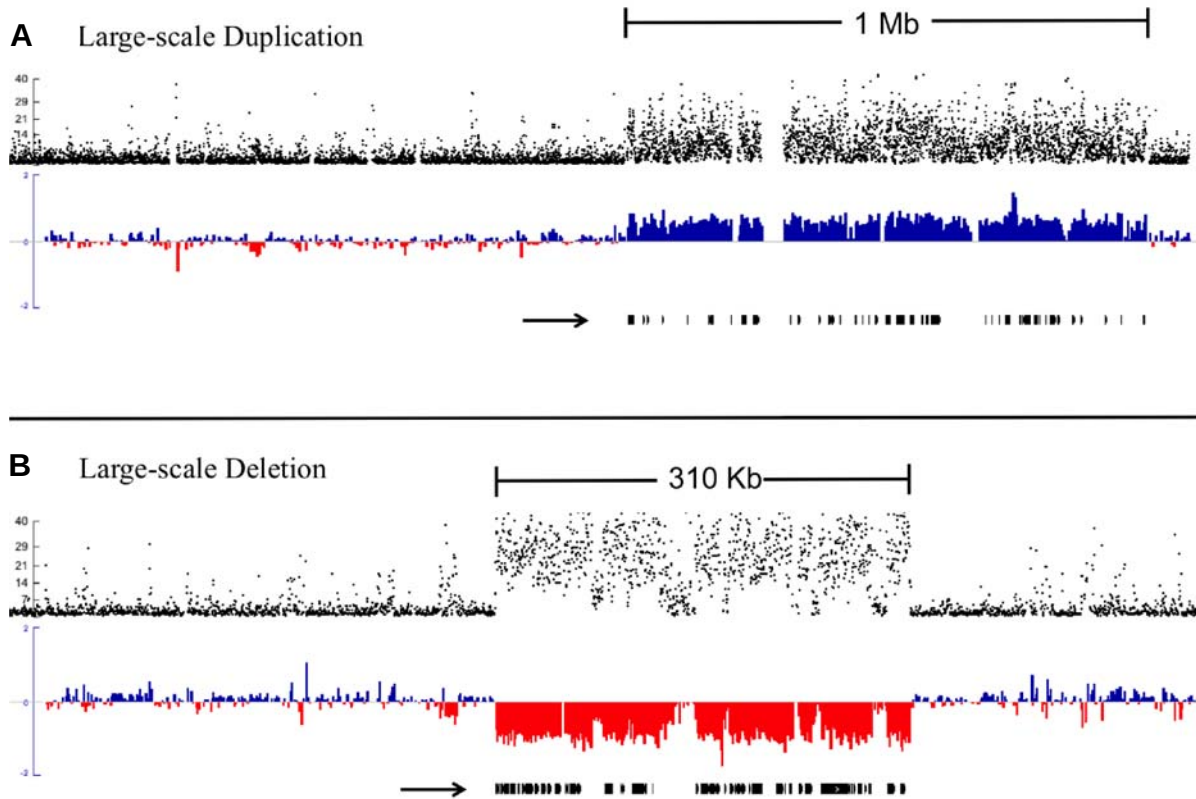


Fig. 4

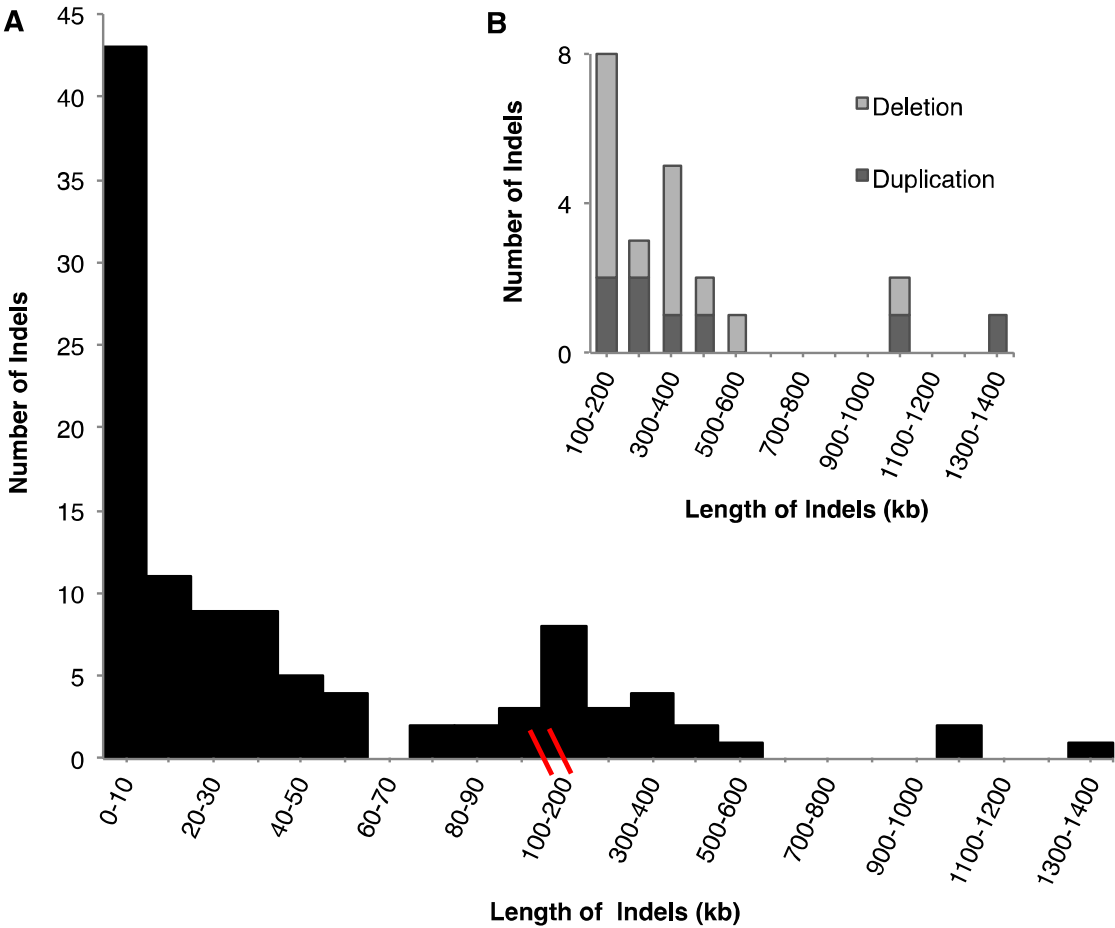


Fig. 5

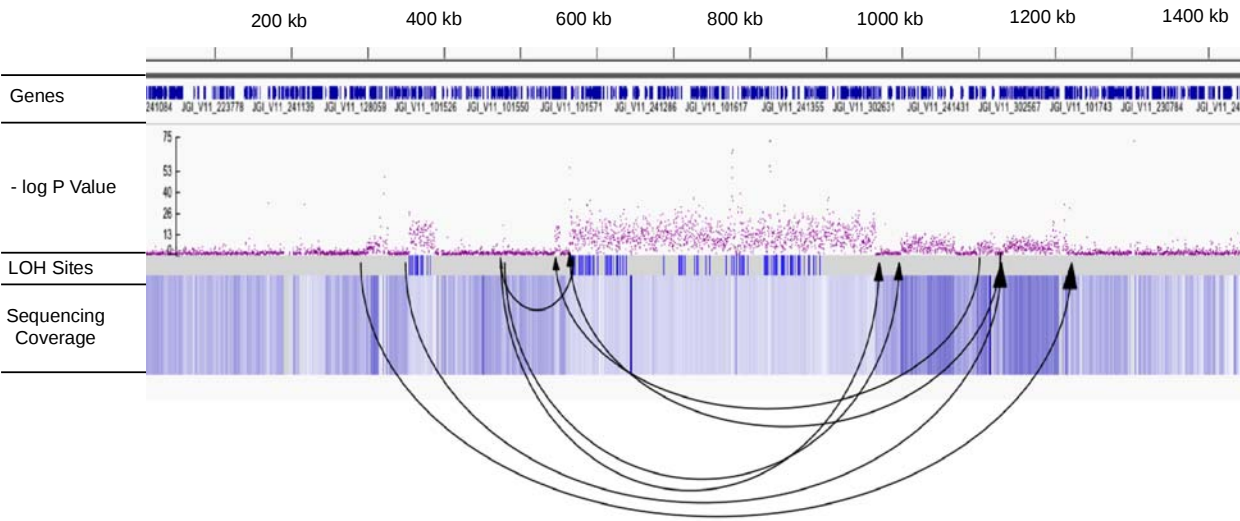


Table Legends

Table 1. Base-substitution mutation statistics for *D. pulex* and previous MA studies. Comparison of base-substitution mutation rates from this and previous MA estimates. AT:GC represents the ratio of G/C → A/T to A/T → G/C base-substitution mutations observed in this study (non-conditional), while Ts:Tv represents the transition to transversion ratio (non-conditional). For *D. pulex* the base-substitution mutation rates are shown for both ASEX and CYC, respectively. Both the Ts:Tv and AT:GC ratios omit the mutator phenotype ASEX-12. Standard errors are in parentheses. Citations: (1) Denver et al. (2009); (2) Keightley et al. (2009); (3) Schrider et al. (2013); (4) Lynch et al. (2008b); (5) Campbell et al. (2012); (6) Ossowski et al. (2010); (7) Sung et al. (2012b).

Table 2. Summary of gene duplication and deletion rates for both the CYC and ASEX experiments. Standard errors are in parentheses.

Tables

Table 1

Species	Base-substitution Mutation Rate ($\times 10^{-9}$ /site/generation)	AT:GC	Ts:Tv
<i>C. elegans</i> (1)	1.93 (0.47)	2.24	0.45
<i>D. melanogaster</i> (2, 3)	4.58 (1.12)	1.51	0.98
<i>S. cerevisiae</i> (4)	0.33 (0.08)	3.74	0.64
<i>H. sapiens</i> (5)	14.01 (1.46)	1.73	1.70
<i>A. thaliana</i> (6)	5.90 (0.60)	3.63	2.40
<i>P. tetraurelia</i> (7)	0.02 (3.60×10^{-3})	5.20	0.73
<i>D. pulex</i>	7.17 (3.28); 4.53 (0.15)	1.95	1.58

Table 2
733

	Subline	Gene duplications	Gene deletions	Gene duplication rate (/single-copy gene/generation) ($\times 10^{-5}$)	Gene deletion rate (/single-copy gene/generation) ($\times 10^{-5}$)
737	CYC-17	56	92	3.08	5.05
738	CYC-237	0	3	0	0.19
	CYC-97	2	0	0.11	0
739	CYC Average			1.06 (0.87)	1.75 (1.43)
740	ASEX-2	116	68	2.76	1.62
	ASEX-12	129	125	7.76	7.52
741	ASEX-48	45	13	1.50	0.43
	ASEX-41	38	188	1.07	5.28
742	ASEX Average			3.27 (1.03)	3.71 (1.09)

759

760 **Sources**

- 761 Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping.
762 *Nat Rev Genet* **12**: 363-376.
- 763 Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC. 1996. Microsatellites show mutational bias
764 and heterozygote instability. *Nat genet* **13**: 390-391.
- 765 Campbell CD, Chong JX, Malig M, Ko A, Dumont BL, Han LD, Vives L, O'Roak BJ, Sudmant PH,
766 Shendure J et al. 2012. Estimating the human mutation rate using autozygosity in a
767 founder population. *Nat genet* **44**: 1277-1281.
- 768 Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion:
769 mechanisms, evolution and human disease. *Nat Rev Genet* **8**: 762-775.
- 770 Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts
771 A, Arnold GJ, Basu MK et al. 2011. The ecoresponsive genome of *Daphnia pulex*.
772 *Science* **331**: 555-561.
- 773 Conrad DF, Keebler JEM, DePristo MA, Lindsay SJ, Zhang YJ, Casals F, Idaghdour Y, Hartl CL,
774 Torroja C, Garimella KV et al. 2011. Variation in genome-wide mutation rates within
775 and between human families. *Nat genet* **43**: 712-U137.
- 776 Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, Howe DK, Lewis SC, Okamoto K,
777 Thomas WK, Lynch M et al. 2009. A genome-wide view of *Caenorhabditis elegans*
778 base-substitution mutation processes. *Proc Natl Acad Sci USA* **106**: 16310-16314.
- 779 Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of
780 insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**: 679-682.
- 781 Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature*
782 **287**: 560-561.
- 783 Friedberg EC. 2003. DNA damage and repair. *Nature* **421**: 436-440.
- 784 Girirajan S, Campbell CD, Eichler EE. 2011. Human copy number variation and complex
785 genetic disease. *Annu Rev Genet* **45**: 203-226.
- 786 Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B,
787 Keightley PD. 2008. Direct estimation of per nucleotide and genomic deleterious
788 mutation rates in *Drosophila*. *Nature* **453**: 128-128.
- 789 Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in
790 evolutionary genetics. *Annu Rev Ecol Evol S* **40**: 151-172.
- 791 Harris K, Nielsen R. 2014. Error-prone polymerase activity causes multinucleotide
792 mutations in humans. *Genome Res* **24**: 1445-1454.
- 793 Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy
794 number. *Nat Rev Genet* **10**: 551-564.
- 795 Hebert PDN, Ward RD. 1972. Inheritance during parthenogenesis in *Daphnia magna*.
796 *Genetics* **71**: 639.
- 797 Hiruta C, Nishida C, Tochinal S. 2010. Abortive meiosis in the oogenesis of parthenogenetic
798 *Daphnia pulex*. *Chromosome Res* **18**: 833-840.
- 799 Innes DJ, Hebert PDN. 1988. The origin and genetic-basis of obligate parthenogenesis in
800 *Daphnia pulex*. *Evolution* **42**: 1024-1035.
- 801 Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE. 2010. De novo rates
802 and selection of large copy number variation. *Genome Res* **20**: 1469-1481.
- 803 Jeffreys AJ, May CA. 2004. Intense and highly localized gene conversion activity in human
804 meiotic crossover hot spots. *Nat genet* **36**: 427-427.

- 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.
- Kibota TT, Lynch M. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* **381**: 694-696.
- Lamb BC. 1984. The properties of meiotic gene conversion important in its effects on evolution. *Heredity* **53 (Pt 1)**: 113-138.
- Lesecque Y, Mouchiroud D, Duret L. 2013. GC-biased gene conversion in yeast is specifically associated with crossovers: Molecular mechanisms and evolutionary significance. *Mol Biol Evol* **30**: 1409-1419.
- Lipinski KJ, Farslow JC, Fitzpatrick KA, Lynch M, Katju V, Bergthorsson U. 2011. High spontaneous rate of gene duplication in *Caenorhabditis elegans*. *Curr Biol* **21**: 306-310.
- Lynch M. 1983. Ecological genetics of *Daphnia pulex*. *Evolution* **37**: 358-374.
- Lynch M. 2007. *The origins of genome architecture*. Sinauer Associates, Sunderland, Mass.
- Lynch M. 2008. The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* **180**: 933-943.
- Lynch M. 2010a. Evolution of the mutation rate. *Trends Genet* **26**: 345-352.
- Lynch M. 2010b. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci USA* **107**: 961-968.
- Lynch M, Gabriel W. 1983. Phenotypic Evolution and Parthenogenesis. *Am Nat* **122**: 745-764.
- Lynch M, Seyfert A, Eads B, Williams E. 2008a. 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, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL et al. 2008b. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* **105**: 9272-9277.
- Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**: 479-U471.
- McAlear MA, Tuffo KM, Holm C. 1996. The large subunit of replication factor C (*Rfc1p/Cdc44p*) is required for DNA replication and DNA repair in *Saccharomyces cerevisiae*. *Genetics* **142**: 65-78.
- Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, Jian M, Liu G, Greer D, Bhandari A et al. 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* **151**: 1431-1442.
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK et al. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* **470**: 59-65.
- Nagylaki T. 1983. Evolution of a finite population under gene conversion. *Proc Natl Acad Sci USA* **80**: 6278-6281.
- Ness RW, Morgan AD, Colegrave N, Keightley PD. 2012. Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*. *Genetics* **192**: 1447.
- Omilian AR, Cristescu MEA, Dudycha JL, Lynch M. 2007. Aneiotic recombination in asexual lineages of *Daphnia*. *Proc Natl Acad Sci USA* **104**: 2554-2554.
- Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**: 92-94.
- Paland S, Colbourne JK, Lynch M. 2005. Evolutionary history of contagious asexuality in *Daphnia pulex*. *Evolution* **59**: 800-813.

- Peters BA, Kermani BG, Alferov O, Agarwal MR, McElwain MA, Gulbahce N, Hayden DM, Tang YT, Zhang RY, Tearle R et al. 2015. Detection and phasing of single base de novo mutations in biopsies from human in vitro fertilized embryos by advanced whole-genome sequencing. *Genome Res* **25**: 426-434.
- Read LR, Raynard SJ, Ruksc A, Baker MD. 2004. Gene repeat expansion and contraction by spontaneous intrachromosomal homologous recombination in mammalian cells. *Nuc acid res* **32**: 1184-1196.
- Scavetta RJ, Tautz D. 2010. Copy number changes of CNV regions in intersubspecific crosses of the house mouse. *Mol Biol Evol* **27**: 1845-1856.
- Schrider DR, Hourmozdi JN, Hahn MW. 2011. Pervasive multinucleotide mutational events in eukaryotes. *Curr Biol* **21**: 1051-1054.
- Schrider DR, Houle D, Lynch M, Hahn MW. 2013. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* **194**: 937-954.
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J et al. 2007. Strong association of de novo copy number mutations with autism. *Science* **316**: 445-449.
- So M, Ohtsuki H, Makino W, Ishida S, Kumagai H, Yamaki KG, Urabe J. 2015. Invasion and molecular evolution of *Daphnia pulex* in Japan. *Limnol Oceanogr* **60**: 1129-1138.
- Stone JL, O'Donovan MC, Gurling H, Kirov GK, Blackwood DHR, Corvin A, Craddock NJ, Gill M, Hultman CM, Lichtenstein P et al. 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**: 237-241.
- Sung W, Ackerman MS, Gout JF, Miller SF, Williams E, Foster PL, Lynch M. 2015. Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. *Mol Biol Evol* **32**: 1672-1683.
- Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012a. Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci USA* **109**: 18488-18492.
- 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**: 19339-19344.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN et al. 2003. The COG database: an updated version includes eukaryotes. *Bmc Bioinformatics* **4**: 41.
- Tucker AE, Ackerman MS, Eads BD, Xu S, Lynch M. 2013. Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. *Proc Natl Acad Sci USA* **110**: 15740-15745.
- Xie Y, Counter C, Alani E. 1999. Characterization of the repeat-tract instability and mutator phenotypes conferred by a Tn3 insertion in *RFC1*, the large subunit of the yeast clamp loader. *Genetics* **151**: 499-509.
- Zhang JZ. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol* **18**: 292-298.
- Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci USA* **111**: E2310-2318.

Fig. 1

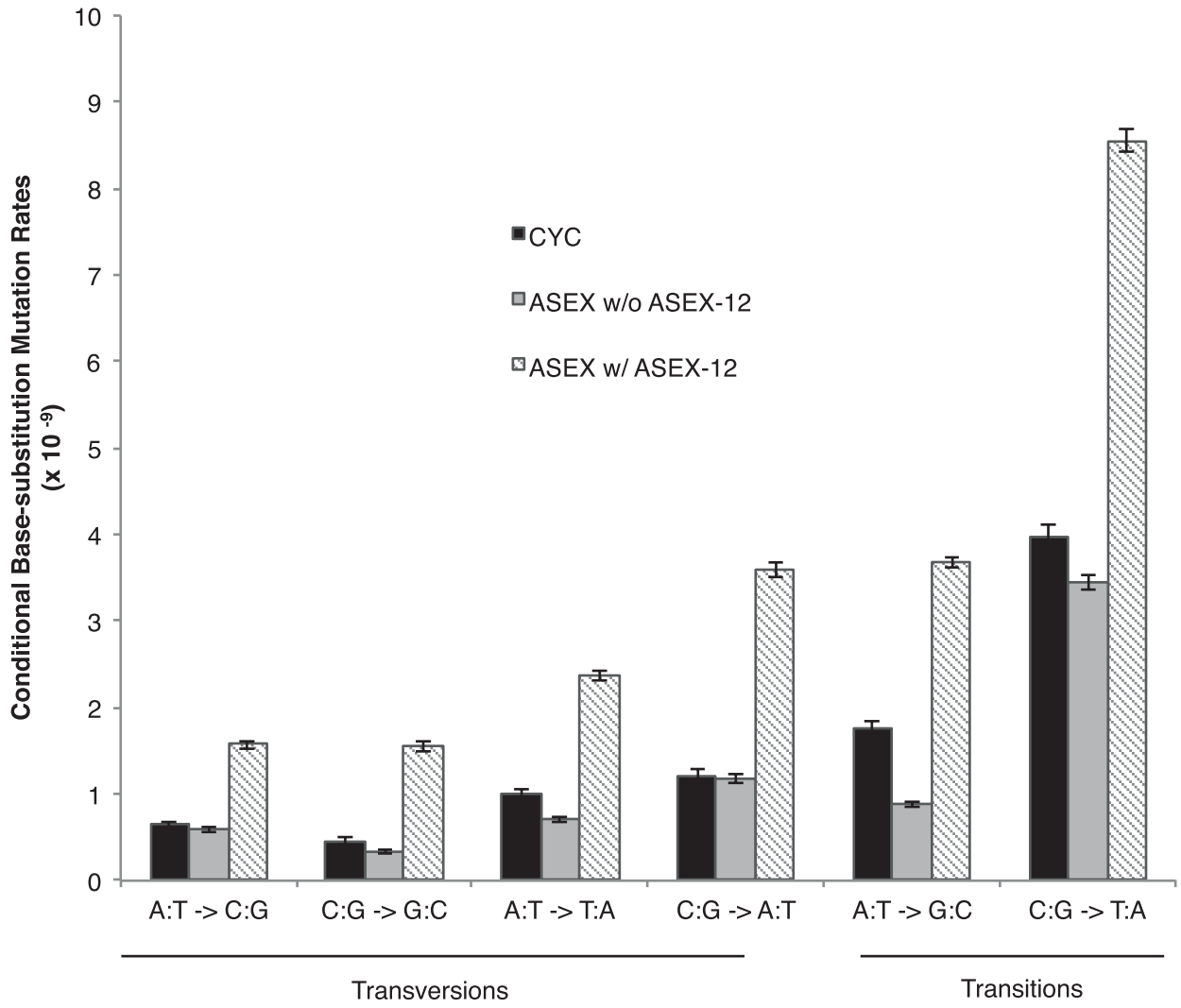
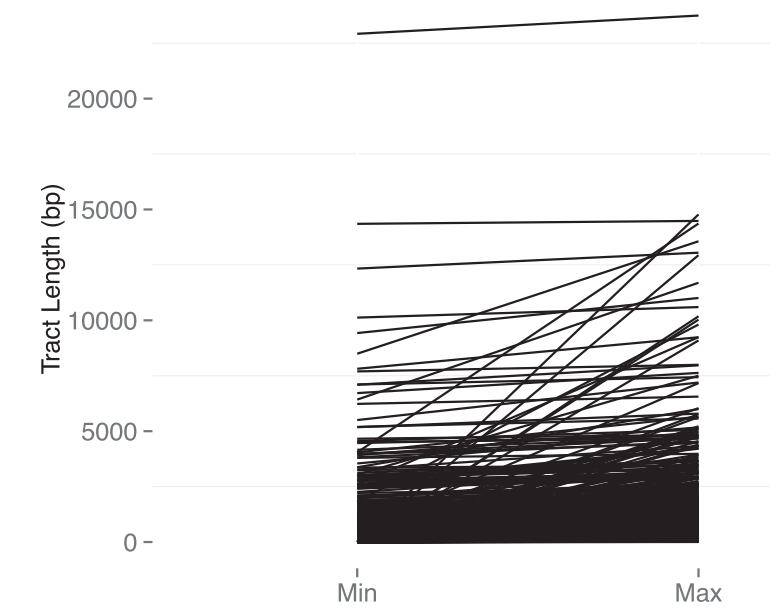
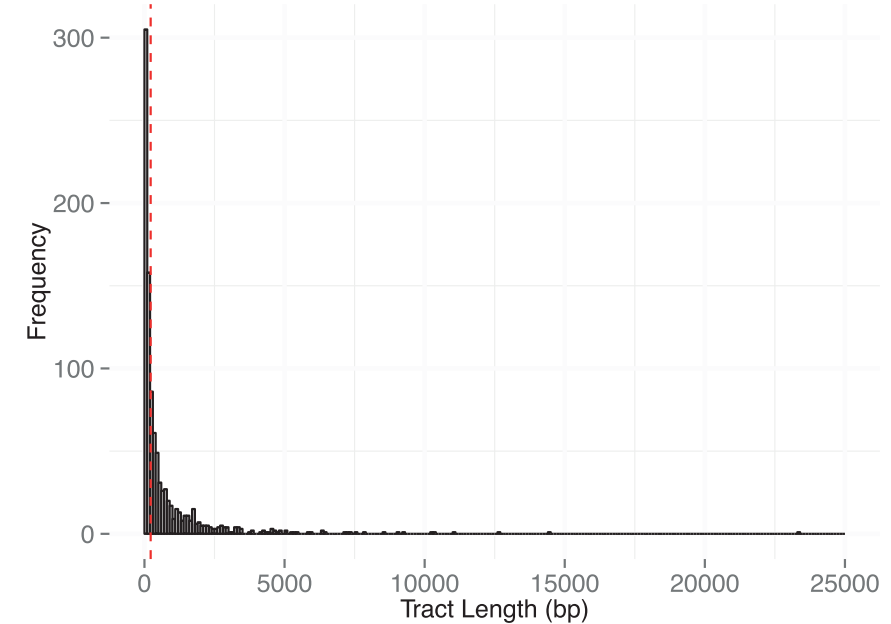


Fig. 2

A



B



C

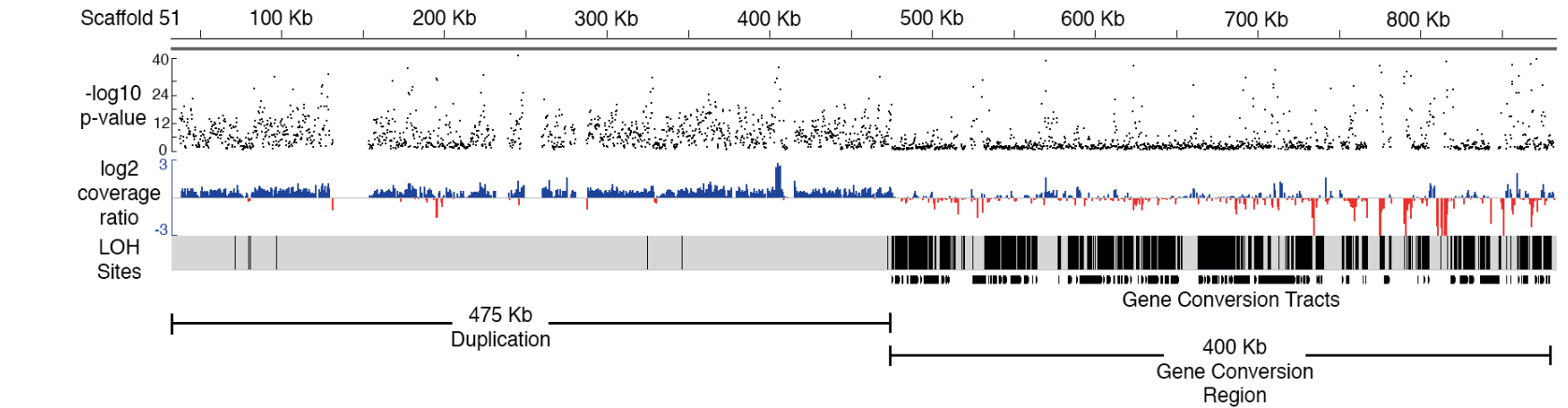


Fig. 3

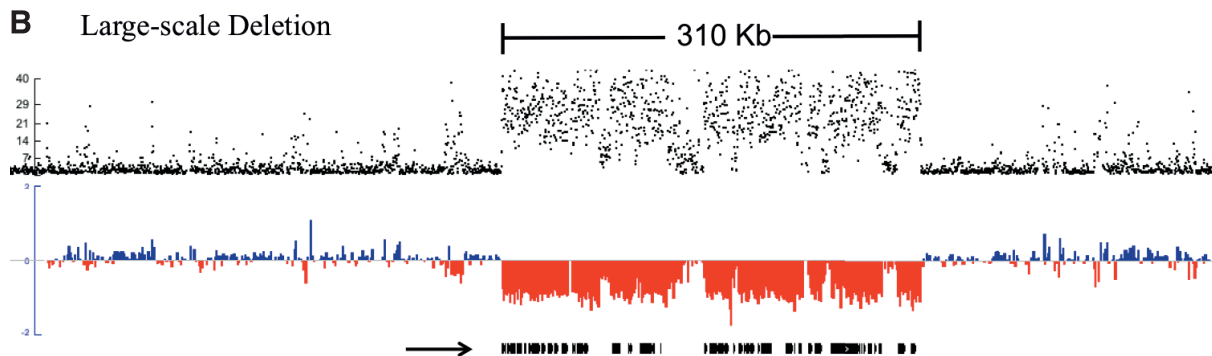
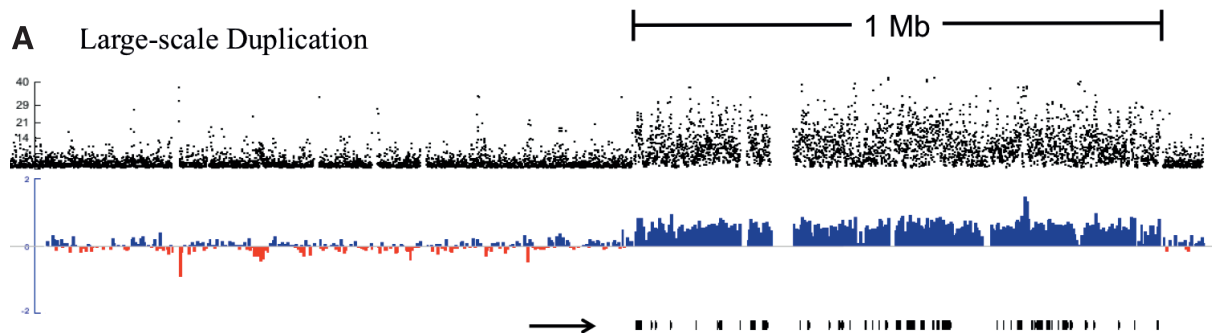


Fig. 4

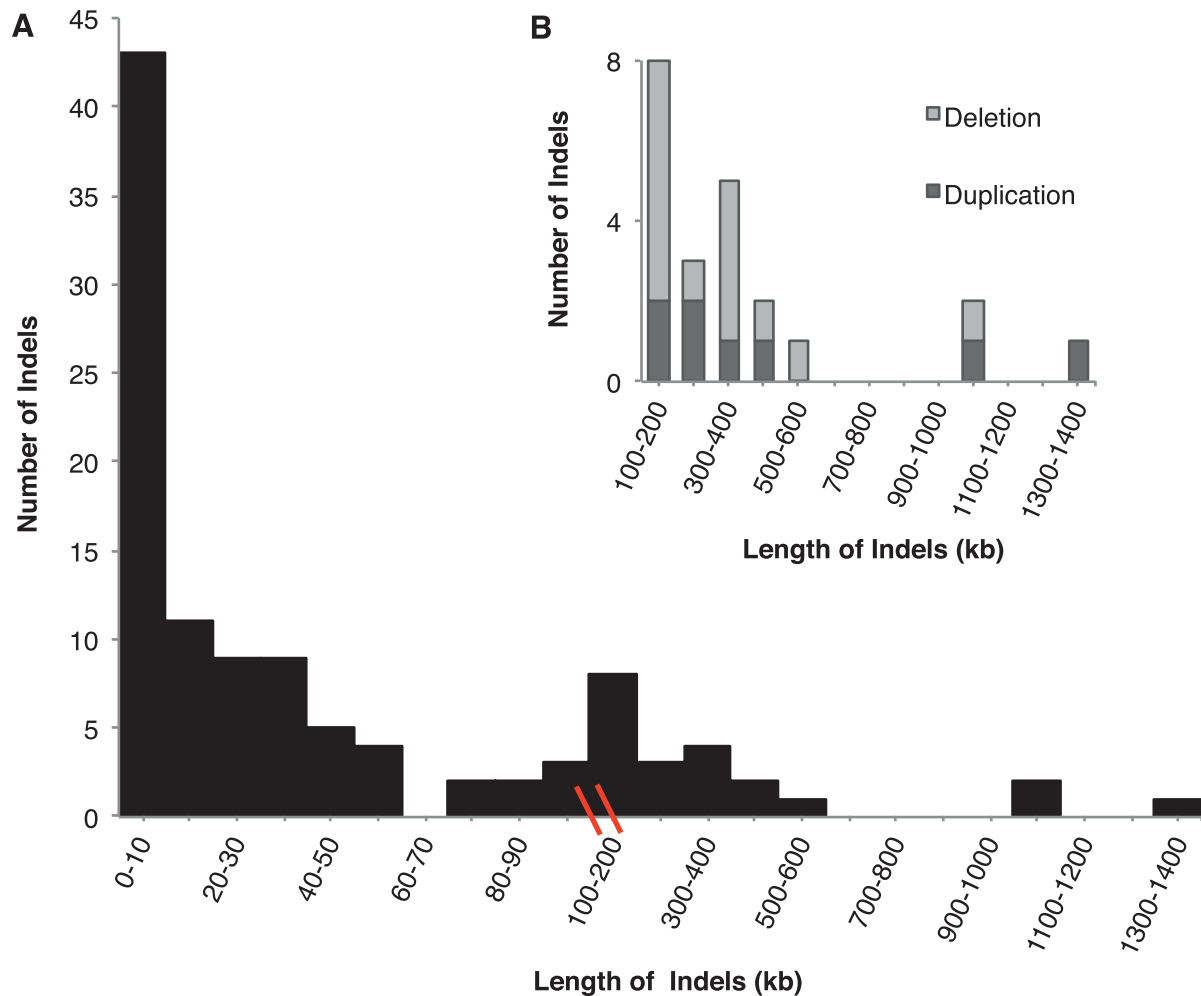


Fig. 5

