Supplementary Text 1. Sources of uncertainty in generation number in mutation accumulation experiments and its potential impacts in this study.

Uncertainty in generation number in mutation accumulation experiments has two major types of sources, which may be loosely termed technical and conceptual. Technical sources refer to uncertainty due to the way generations are counted in a given experiment, while conceptual sources refer to what is considered a generation in an experiment. While uncertainty in generation number is typically ignored in MA experiments, it has the potential to impact comparisons within studies and to confound comparisons across studies when different standards for determining generation number are used. As for other types of uncertainty, failing to consider uncertainty in generation number may lead to reporting parameter estimates (e.g. – mutation rates) that are more precise than justified by the data, biased, or both.

Single-celled organisms

In studies of single-celled organisms, it is almost impossible to directly count the number of generations passed for at least two reasons. First, their size typically precludes identifying and handling individual cells, at least at the scales required. Second, generation times are typically on the orders of hours and thus per-generation handling is impractical. The typical strategy is to propagate these organisms on solid medium via single colony transfer (e.g. - Long *et al.* 2016), which ensures a bottleneck to a single cell per transfer period. The number of cells per colony pre-transfer (N) is estimated via a range of approaches, and the number of generations (G) is that period estimated via the equation G = logN/log2; this estimate is usually made for a subset of transfers and extrapolated to all transfers. This assumes both strict exponential growth and synchronicity of divisions among cells in a colony, as well as allowing some degree of selection to occur within colonies which may underestimate true mutation rates (Halligan and Keightley 2009). If a portion of cells die during colony growth, true generation number of surviving cells will be underestimated. Where quantifying the number of generations per transfer is impossible or impractical, mutation rates can be expressed in units of absolute time, such as mutations per day, which still allows comparisons within experiments (e.g. – treatment vs control) but limits comparisons between studies as it is unlikely that demographics are replicated across environments and organisms (Shibai *et al.* 2017).

While estimating generation number in experiments with single-celled organisms may be technically difficult, such experiments do not suffer from the conceptual issues associated with the number of cell divisions per generation (below) due to the fact that one cell division is one generation in such organisms.

Multi-cellular organisms

In contrast to single-cellular organisms, determining generations passed in multicellular organisms is typically technically more straight-forward, as they can usually be seen and handled individually. From a technical perspective, in principle, the exact number of generations accumulated for each MA line can be known if investigators count and track transfers. Where the individual representing the current generation of a MA line (the ‘focal’ individual) either dies or displays sterility, a backup must be used to continue propagating that MA line. These back ups are typically from the previous generations of the same line (Halligan and Keightley 2009) and at the time of use can contain individuals from the previous generation (the parent of the current focal individual, i.e. - previous focal individual), but also from the current generation (siblings of the focal individual) and from the next generation (offspring of siblings of the focal individual). It is therefore possible that the backup selected as replacement may not come from the presumed generation. Since there is only one individual from the previous generation and potentially many from other generations, mistakes likely involve the later generations. This may lead to either over- or underestimating the number of generations accumulated depending on whether the experimenter intends to replace the focal individual with one from the same generation (leading to underestimation) or to advance the experiment by selecting the offspring of one of its siblings (leading to overestimation). The extent to which this may cause uncertainty increases with the frequency to which backups are employed and with the difficulty in identifying individuals of different generations. Where there is little variability in age of first reproduction and lines are therefore expected to propagate relatively synchronously (e.g. – *C. elegans*), some authors forgo tracking the generation of each MA line separately in favour of an experiment-wide mutation number (e.g. - Baer *et al.* 2005). Such estimates should be considered maximum generation numbers (GMAX following the notation of Baer *et al.* (2005)), since failure to reproduce in time for a transfer / generation advance and incorrect identification of generations during back up use will reduce the actual number of generations accumulated in a specific MA line. Again, the amount of uncertainty increases with the use of backups and the difficulty in identifying individuals of different generations.

More fundamentally than the above technical issues, mutation accumulation experiments involving multicellular organisms suffer from conceptual uncertainties associated with possible variation in the number of cell-divisions between generations. Such issues are well-highlighted by Latta *et al.* (2013), in which MA lines of the microcrustacean *Daphnia pulicaria* of 10 different genotypes were propagated using two strategies, either consistently using the second clutch (brood) of offspring (rapid propagation) or the last possible clutch (slow propagation). Across a range of traits, mutational decline was ~3.3 times greater per generation in the slow propagation treatment. The explanation offered is in the slow propagation treatment an increased number of cell divisions occurred per generation, increasing the number of errors associated with DNA replication (i.e. – mutations) (Drake *et al.* 1998). Such an interpretation is supported by so-called “male mutation bias” in mammals, where per generation mutation rate increases with the parent’s age for fathers but not for mothers. These effects are explained by the number of cell-divisions leading to sperm formation increases with age in males but the number of cell-divisions leading to egg formation is constant across age in females (Ellegren 2000). Curiously, per cell division mutation rates can arise even when the mutations are not strictly associated with replication errors. For example, Gao *et al.* (2016) demonstrated that when an equilibrium level of premutations (DNA damage that results in incorrect replication of DNA if not repaired prior to replication) occur in cells as result of exogenous factors producing mutations and cellular machinery repairing premutations, incorporation of uncorrected premutations into daughter cells during DNA replication can produce per cell-division mutation rates without invoking replication errors *per se*.

Although Latta *et al.* (2013) presents an extreme case in which differences in per-generation cell-divisions were maximized, similar effects can be expected on a much smaller scale unless each MA line is held to same demographic standard, such as always using offspring from the first reproductive episode with individuals mated (if appropriate) at the exact same life history stage. Such a system is unlikely to be practical given lines would need to be assessed every day over a several year experiment, and it is unclear if the reduction in variability in cell-divisions per generation would be greater than noise introduced via additional environmental noise associated with intensive handling. Where demographics can potentially differ between lines, e.g. – such as in our experiment where lines are given an additional chance to reproduce, correcting for this will be difficult since number-of-cell divisions per generation will likely vary with age in a complex non-linear way (Gao *et al.* 2016). Given this, we recommend authors ensure their findings are robust to some uncertainty in generation number, and that authors appreciate that reported size of confidence intervals around mutation rates are likely underestimates since MA studies typically don’t consider uncertainty in generation number. Below we show that the results from our experiment are robust against realistic levels of uncertainty in generation number.

Effects of uncertainty in generation number in our experiment

 Since we count generations for each MA line and try to select siblings of the focal individual when using backups, our generation numbers are minimum generation numbers (GMIN,by analogy to GMAX in Baer *et al.* (2005)). They may be underestimates due to incorrectly selecting the offspring of a sibling rather than a sibling when using backups, and because we consider cases where an individual is given up to two additional chances to reproduce (“carry-overs”) as a single generation, despite the fact a greater-than-usual number of cell-divisions are possible in such time, and they thus may represent more than one “typical” generation.

Although we did not record the full per-line history of backup use and carry overs per line, we do know 4.8 % of all transfers were backups and 5.8 % were carry-overs. Assuming 25 % of all backups involved a misidentification (a very high estimate since reproduction and growth tends to slow in our backups due to crowding and the resource dependence of growth and reproduction in *Daphnia* (Lynch 1989), and thus generations typically are non-overlapping in backups), this indicates the generation number of the average MA line may be underestimated by around 7%. However, the potential for underestimation is higher in lines with fewer progressed generations since they involved more backups and carry overs. To combine these factors, we calculated the theoretical maximum number of generations (GMAX) a line could have accumulated as the number of the days of the experiment was running at the time of the measurement or tissue sample collection divided by 12.7, the average number of days between transfers. The difference between GMAX and recorded generation number (GMIN) is the range of potential degrees of underestimation. For each line we correct the generation number as GMIN + (a \* (GMAX – GMIN)), where ‘a’ is a value drawn from a normal distribution with mean 0.07 and standard deviation 0.02. We generated 1000 such generation-corrected data sets in R (R Core Team 2015), and repeat several analyses described in the main text to quantify the robustness of our results to uncertainty in generation number (low uncertainty simulations). We also repeated this with ‘a’ drawn from a normal distribution with mean 0.5 and standard deviation 0.2 as an extreme case (high uncertainty simulations).

Effects of mutation accumulation on intrinsic rate of increase and on body size at maturity

 For the uncorrected data (no uncertainty in generation number, i.e. – that reported in the main text of this manuscript), for intrinsic rate of increase (r), there was no significant effect of metal treatment on mutational bias (ΔM) (F2, 24 = 2.05, P = 0.15), and ΔM across all treatments was not significantly different from zero at -0.0079 ± 0.0062 (standard error) % per generation (t29 = 1.29, P = 0.21). For low uncertainty simulations, there was no significant effect of metal treatment on mutational bias (ΔM) in any simulation (F2, 24 ≤ 2.71, P ≥ 0.087), and ΔM across all treatments ranged from -0.0080 ± 0.0061 % per generation (t29 = 1.31, P = 0.20) to -0.0078 ± 0.0061 % per generation (t29 = 1.28, P = 0.21). For high uncertainty simulations, there were significant effects of metal treatment on mutational bias (ΔM) in 81 of 1000 simulations (F2, 24 ≥ 3.47, P ≤ 0.05), and ΔM across all treatments ranged from -0.0083 ± 0.0057 % per generation (t29 = 1.45, P = 0.16) to -0.0069 ± 0.0056 % per generation (t29 = 1.22, P = 0.23).

 For uncorrected data, for body size, there was a significant effect of metal treatment on ΔM (F2,24 = 7.97, P = 0.002, Figure 1B, Table 1) and ΔM across all treatments significantly different from zero at -0.0035 ± 0.0062 % per generation (t29 = 0.57, P = 0.56). For low uncertainty simulations, there was a significant effect of metal treatment on mutational bias (ΔM) in all simulation (F2, 24 ≥ 4.83, P ≤ 0.017), and ΔM across all treatments ranged from -0.0036 ± 0.0062 % per generation (t29 = 0.59, P = 0.56) to -0.0035 ± 0.0061 % per generation (t29 = 0.56, P = 0.58). For high uncertainty simulations, there was a significant effect of metal treatment on mutational bias (ΔM) in 180 of 1000 simulation (F2, 24 ≥ 3.41, P ≤ 0.05), and ΔM across all treatments ranged from -0.0044 ± 0.0057 % per generation (t29 = 0.76, P = 0.45) to -0.0028 ± 0.0058 % per generation (t29 = 0.48, P = 0.63).

 Thus for our life-history experiments, our findings are robust to uncertainty in generation number in cases where realistic levels of uncertainty are introduced and at most, slightly underestimate the width of true confidence intervals. Higher levels of uncertainty did produce situations were results qualitatively changed.

Genomic analyses - SNMs

For uncorrected data, overall, there was no evidence that metal treatment affected processes producing SNMs. There was no significant effect of metal treatment on the number of SNMs per bp per generation (one-way ANOVA, F3,51 = 0.994, P = 0.40). The 644 SNMs detected across all lines corresponds to an overall SNM rate of 1.61 (± 0.08, standard error) \* 10-9 bp / gen. For low uncertainty simulations, there was no significant effect of metal treatment on the number of SNMs per bp per generation in any simulation (one-way ANOVA, F3,51 ≤ 1.07, P ≥ 0.38), and SNM rates ranged from 1.578 ± 0.076 \* 10-9 bp / gen to 1.585 ± 0.077 \* 10-9 bp / gen. For high uncertainty simulations, there was no significant effect of metal treatment on the number of SNMs per bp per generation in any simulation (one-way ANOVA, F3,51 ≤ 1.84, P ≥ 0.15), and SNM rates ranged from 1.41 ± 0.07 \* 10-9 bp / gen to 1.47 ± 0.08 \* 10-9 bp / gen.

Thus for the reported SNM rates, our findings are robust to uncertainty in generation number in cases where realistic levels of uncertainty are introduced. At most, our results slightly underestimate the width of true confidence intervals and lead to a slightly upwardly biased estimate because our reported generation numbers are minimums and therefore likely slight underestimates of true mutation number.

References

Baer, C. F., F. Shaw, C. Steding, M. Baumgartner, A. Hawkins *et al.*, 2005 Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. Proceedings of the National Academy of Sciences 102**:** 5785-5790.

Drake, J. W., B. Charlesworth, D. Charlesworth and J. F. Crow, 1998 Rates of spontaneous mutation. Genetics 148**:** 1667-1686.

Ellegren, H., 2000 Heterogeneous mutation processes in human microsatellite DNA sequences. Nature Genetics 24**:** 400-402.

Gao, Z., M. J. Wyman, G. Sella and M. Przeworski, 2016 Interpreting the Dependence of Mutation Rates on Age and Time. PLOS Biology 14**:** e1002355.

Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. Annual Review of Ecology, Evolution, and Systematics 40**:** 151-172.

Latta, L. C., K. K. Morgan, C. S. Weaver, D. Allen, S. Schaack *et al.*, 2013 Genomic background and generation time influence deleterious mutation rates in *Daphnia*. Genetics 193**:** 539-544.

Long, H., S. F. Miller, C. Strauss, C. Zhao, L. Cheng *et al.*, 2016 Antibiotic treatment enhances the genome-wide mutation rate of target cells. Proceedings of the National Academy of Sciences 113**:** e2498-e2505.

Lynch, M., 1989 The life history consequences of resource depression in *Daphnia pulex*. Ecology 70**:** 246-256.

R Core Team, 2015 R: A language and environment for statistical computing. R Foundation for Statistical Computing., pp., Vienna, Austria.

Shibai, A., Y. Takahashi, Y. Ishizawa, D. Motooka, S. Nakamura *et al.*, 2017 Mutation accumulation under UV radiation in Escherichia coli. Sci Rep 7**:** 14531.

Supplementary Text 2.Detailed methods for the life-history assays

Life-history assays were performed under the same environmental conditions as for the propagation of the control MA lines (constant 18 °C and 70 % humidity, a 12h : 12h light : dark cycle and standard medium described in Celis-Salgado *et al.* (2008))except the medium contained the algae *Scenedesmus sp.* at 300 000 cells / mL (no supplemental feeding occurred), and all individuals were transferred to new medium every second day throughout the experiment.

For each of the 30 MA lines assayed (10 control lines, 10 Cu lines, 10 Ni+Cu lines), two or three newborn offspring were selected and placed in separate vials. To obtain the 10 non-MA lines assayed, the non-MA population tank was briefly stirred and 10 individuals were haphazardly chosen and placed in separate vials. For both MA and non-MA lines, second-clutch offspring of these individuals were used to establish up to 7 sub-lines per line, with all offspring taken from a single mother where possible. All sub-lines underwent two generations under the common test conditions with propagation by single individual descent using the second clutch to minimize maternal effects. All individuals were transferred to fresh medium and vials every second day. Following this, two newborn individuals per sub-line were retained – one to be measured for body size at first clutch release and the other to be measured for all other life history parameters, i.e. age at release for the first four clutches, number of progeny per clutch for the first four clutches, and longevity. These individuals were checked daily for death or reproduction. Clutch sizes were counted immediately after transfer or the following day. Individuals measured for body size were placed in 100 % ethanol upon release of their first clutch while other individuals were continually transferred until they died. To determine their size, body size individuals were photographed using a LEICA MC120 HD camera mounted on a LEICA S6D microscope and measured to the nearest 0.001 mm using ImageJ v1.49 (Schneider *et al.* 2012) by comparison to 1.00 and 2.00 mm calibration slides.

Vials were distributed across six trays such that all sublines from a given line occurred on the same tray and that lines within treatments were spread across trays as evenly as possible. Transferred individuals were randomly redistributed within their tray, and trays were randomly repositioned within the growth chamber used to minimize local environmental effects at each transfer. Before further statistical analyses, we tested for tray effects by comparing single-factor ANOVAs with and without tray as a covariate relating mean intrinsic rate of increase (r, Supplementary Text 2) and mean body size per line to treatment using the R package lme4 (Bates *et al.* 2015). Since including tray did not improve model fit in either case (χ21 = 0.00, P = 0.99; χ21 = 0.71, P = 0.40 respectively), tray effects were not considered further.

References

Bates, D., M. Mächler, B. Bolker and S. Walker, 2015 Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67**:** 48.

Celis-Salgado, M. P., A. Cairns, N. Kim and N. D. Yan, 2008 The FLAMES medium: a new, soft-water culture and bioassay medium for Cladocera, pp. 265-271 in *International Association of Theoretical and Applied Limnology, Vol 30, Pt 2, Proceedings*, edited by J. Jones and J. Faaborg.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri, 2012 NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9**:** 671-675.

Supplementary Text 3.Determination of intrinsic rate of increase

To summarize across correlated traits in a biologically meaningful way, we determined intrinsic rate of increase (r) for each line by calculating partial life-history tables across sublines for each line and using simulations to determine r from these tables. Sublines that did not survive long enough to be included in the life-history assay were represented as dying on day three in the life history table (before any reproduction) to account for reduced viability of these lines. Our simulations are equivalent to the integrative approaches sometimes used. Simulations were performed using a custom R script (Supplementary File 1) which modeled populations beginning with a single individual growing over 1500 days with a given life-history table; on each day deaths occurred before surviving individuals reproduced. Intrinsic rate of increase (r) was calculated per day as population size at day n divided by population size at day n-1, and typically converged on a stable value within a few hundred days. Mean r over the last x days was used in further analyses, where x is the maximum longevity observed for that line.

In four cases, clutches were accidentally discarded before counting, resulting in missing data. To allow our model to run, we used the mean clutch size for that clutch from other sublines within that line for these missing values. These four cases correspond to 0.47 % of all clutch counts.

Supplementary Text 4.Detailed bioinformatic methods

Reads were trimmed to remove adapters and overlapping read-pairs were merged with SeqPrep (https://github.com/jstjohn/SeqPrep) then aligned to the *Daphnia arenata* genome(Colbourne *et al.* 2011) using BWA v0.7.10 (Li and Durbin 2009). Duplicate reads were removed with Picard tools v1.123 (https://broadinstitute.github.io/picard/) and local realignment around putative INDELs was performed with GATK v.3.3.0 (DePristo *et al.* 2011). This resulted in a filtered BAM file for each sequenced line. Base calling was performed using HaplotypeCaller in GATK v.3.3.0 to assign putative genotypes within lines, with GenotypeGVCFs in GATK v.3.3.0 used to refine genotypes by considering information from all lines. SelectVariants in GATK v.3.3.0 was then used to retain only high-quality SNM and INDEL variants using the GATK recommended parameters (SNMs - QD > 2.0, FS < 60.0, MQ > 40.0, QUAL > 50.0, MQRankSum > -12.5, ReadPosRankSum > -8.0; INDELS - QD > 2.0, FS > 200.0, QUAL > 50.0, ReadPosRankSum > -20.0). We discarded sites withread depth greater than twice the average depth across samples (total depth > 1375) as we considered these likely mapping errors. We then excluded scaffolds corresponding to the mitochondrial genome, as our focus is on nuclear mutations, and scaffolds corresponding to ribosomal DNA and sites annotated as repetitive regions in the reference genome, which are prone to mapping errors.

To identify mutations, we performed additional filtering on the putative variant sites called with GATK based on the underlying read distribution in BAM files. We first used mpileup in samtools v0.1.19 (Li 2011) to extract the read coverage from BAM files for each putative variant site in all lines and excluded reads having mapping quality < 20 and/or base quality < 10. Binomial tests were then used to determine if allele frequencies within each line at each site were consistent with heterozygosity or homozygosity; sites with frequencies not supporting either were discarded as probable mapping errors. We retained sites where only a single line had a mutant genotype, as mutations should be unique in an MA design, where the putative mutant had at least six reads, where there was no more than one other read supporting the mutant allele across all other lines, and where all other lines at least one read and the same inferred genotype. We rationalize that if the putatively mutant allele was actually present in the ancestor, the likelihood it would not have been sampled in the other 54 lines even with a single read each is 5.55 \* 10-17 (0.554) for each such site. We did not require 6 reads in all lines due to the small fraction of the genome recovered under such requirements (20,016,581 bp or 11.12 % of the reference genome) and avoid problems associated with false negatives by determining the number of sites potentially callable as mutant on a per line basis when calculating mutation rates (below). We also discarded sites where the ancestral genotype differed from that of the reference genome, sites immediately adjacent to INDELs, as mutations with these characteristics were previously shown to be caused by mapping errors (Flynn *et al.* 2017). Finally we discarded sites putatively mutated from one heterozygous state to another heterozygous state (Het-Het sites,throughout we use ‘het’ to refer to heterozygous sites, ‘hom’ to refer to homozygous sites, and ‘-‘ to the transition between states) due to an inability to successfully validate these sites (details below, 0 of 3 successfully validated, Table S3). Based on our observed mutation rate at homozygous sites (assuming no difference in mutation rate between homozygous and heterozygous sites), the fraction of the genome recovered, and 0.7 % per-site heterozygosity in the ancestor (Flynn *et al.* 2017), we would expect to be able to observe only four Het-Het SNM across the whole genomes of all our MA lines. Exclusions of these sites should only cause a slight (~0.6 %) underestimation of the true SNM rate

References

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker *et al.*, 2011 The ecoresponsive genome of *Daphnia pulex*. Science 331**:** 555-561.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43**:** 491-498.

Flynn, J. M., F. J. J. Chain, D. J. Schoen and M. E. Cristescu, 2017 Spontaneous mutation accumulation in *Daphnia pulex* in selection-free vs. competitive environments. Molecular Biology and Evolution 34**:** 160-173.

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

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25**:** 1754-1760.