The effects of *de novo* mutation on gene expression and the consequences for fitness in *Chlamydomonas reinhardtii*

Eniolaye J. Balogun[1] and Rob W. Ness[1]

1. Department of Biology, William G. Davis Building, University of Toronto, Mississauga, L5L1C6, Canada

Abstract

Mutation is the ultimate source of genetic variation, the bedrock of evolution, but predicting the consequences of new mutations remains a challenge in biology. Gene expression provides a potential link between genotype and phenotype with growing interest in the influence of mutation on expression expanding our understanding of evolutionary changes and disease. Yet, the total amount of expression variation created by mutations; the aspects of buffering in the regulatory network or specific genes against mutation; and the fitness consequences of mutational changes to expression remain relatively unexplored. Here, we investigate the effects of >2600 *de novo* mutations on gene expression across the transcriptomes of 28 mutation accumulation lines derived from two independent wild-type genotypes of the green algae *Chlamydomonas reinhardtii*. We found a much larger number of differentially expressed genes than mutations implying large effect trans-acting mutations, but also show evidence for some smaller effect cis-acting mutations. The larger influence of trans-acting mutations likely eliminated any correlation between the number of mutations and the extent of expression divergence in each line. We also found that the per generation increase in expression variance is substantially higher in *C. reinhardtii* than other species, which is in part due to the lower number of mutations per generation seen in unicellular organisms. Finally, we also demonstrate that the extent of expression divergence negatively correlates with overall fitness, providing direct evidence of stabilizing selection on gene expression.

Introduction

Mutations provide both the raw variation for evolution through adaptation and contribute towards disease and aging. Unfortunately, a challenge in predicting the effects of mutations is linking the genotypic consequences of mutation to phenotypic change. Gene expression provides a potentially useful link where we can observe changes in the regulation of genes due to mutation. Moreover, there is a longstanding hypothesis that evolution may proceed primarily through alteration in gene expression rather than changes in protein structure [1]. For this to be true, mutations altering expression must occur at a higher rate or be more beneficial than protein-coding mutations. Though there are now numerous examples of regulatory changes driving both adaptive changes (e.g [2]) and harmful diseases (e.g [3]), we still lack a general description of how variation in gene expression originates from *de novo* mutation and the effects of these mutations on fitness and other important traits.

Comparative research has revealed numerous patterns in gene expression variation that provide insights into the impact of mutation on gene regulation (reviewed in [4]). In general, essential genes, genes with high expression and those that are central in molecular interaction networks tend to display reduced

expression variation within and between species [5–12]. An obvious explanation is that stabilizing selection for expression level removes deleterious variants and reduces the overall segregating expression variation. However, the distribution of effects of mutations may also influence the extent of variation in different gene classes if mechanisms have evolved that buffer essential genes from perturbations by mutation. Similarly, it has been observed that more within species variation is driven by trans-regulatory variants, while cis-regulatory mutations underlie between-species variation [13–16]. This might be explained if trans-regulatory mutations are more common, because there is a larger mutation target, and are also more detrimental because they influence numerous genes [17]. We may also expect trans-regulatory variation to only segregate at low frequencies and thus not contribute to between species divergence. Critically, the above explanations all depend on the rate, target size, expression effect and fitness impact of regulatory mutations, highlighting the importance of studying expression mutation.

Some of the most detailed information on how *de novo* mutation affects gene expression comes from experiments using single gene models where the effect of mutations on a reporter gene can be studied in detail. For example, in *S. cerevisiae* cis-acting mutations in the TDH3 promoter were shown to generally reduce expression and have a greater effect on expression than trans-acting mutations, which did not show an asymmetry in the direction of effect [18]. However, although trans-acting mutations were relatively weak, they were 265 times more likely to occur than cis-acting mutations, implying a much larger mutational target [18,19]. Despite the valuable insights gained from such precise experiments, these systems can not capture the diverse mechanisms that may govern the effects of natural mutations on expression throughout the transcriptome. For example, some genes have redundant promoters and regulatory proteins that maintain expression in the presence of mutational perturbations (reviewed in [16,20]). Network feedback may also serve to mute the effects of mutation on expression of genes, but such effects would likely be most evident when genes native to the genome in question are investigated. Additionally, there is evidence that trans-acting eQTLs may be clustered in hotspots, affecting thousands of genes [21], which may not be observable in single gene model systems. There is therefore good reason to investigate the effect of *de novo* mutation on the transcriptome as a whole.

Investigating spontaneous mutations has been constrained by the rarity with which new mutations arise. Mutation accumulation (MA) experiments overcome this limitation by evolving lines under minimal selection, therefore facilitating the build-up of spontaneous changes irrespective of their fitness [21]. MA has been used extensively in the estimation of the rate, spectrum, and fitness effects of mutations. However, there are surprisingly few studies investigating the expression changes in MA lines in a small number of model organisms including *S. cerevisiae*, *D. melanogaster* and *C. elegans* [22–27]. A key parameter in measuring how mutation affects gene expression is the contribution of mutation to

expression variation per generation i.e. mutational variance (V_m) [28]. From the studies formerly mentioned, the mutational variance of *C. elegans [22]* was up to 9-fold higher than that of *D. melanogaster* and [23] and yeast [24]. MA studies have also shown that mutation introduced far more variation in expression than would be expected based on standing expression variation in natural populations, suggesting strong stabilizing selection in nature [22,23,26,29]. We have also learned that the expression of genes likely involved in fitness-related traits and pivotal developmental stages are highly constrained [23,27,30]. Moreover, there is evidence of an inverse correlation between V_m and expression level [23]. Despite these breakthroughs [31,32], our knowledge of the global effects of mutations across the transcriptome is restricted to a small number of organisms with varied and sometimes limited technologies.

Here, we investigate the effect of *de novo* mutation on gene expression and fitness in *Chlamydomonas reinhardtii*. This haploid green alga is a longstanding model for numerous aspects of cell, molecular, and genome biology [33,34] and has emerged as a model for studying the evolution of mutation in eukaryotes. Over the past decade, numerous studies have explored mutation rate [35–39], base spectrum, and the predictors of mutational variation across the genome of *C. reinhardtii* and related species [37,40]. We also know that on average mutations in *C. reinhardtii* are deleterious, but evidence from the genetic mapping of mutational effects demonstrates a significant fraction of spontaneous mutations are beneficial [41,42]. However, in this system, it remains unexplored how the accumulation of mutations has altered patterns of gene expression. Here, we sequenced the transcriptomes of 28 well-studied MA lines derived from two independent wild-type genotypes with fully characterized mutations and known fitness [37,41,42]. Each MA line was grown for 800-1100 generations and carries approx. 76 mutations. Integrating new transcriptome data with detailed characterization of these lines, we address the following questions: (1) How much does *de novo* mutation influence the direction and overall variation in gene expression? (2) do mutations located near genes confer specific expression changes suggesting cis-acting effects? (3) Are there expression patterns associated with genes that play roles in regulatory and/or metabolic networks? (4) Do lines with more mutational transcriptome divergence also suffer greater fitness effects?

Methods:

Sample collection and growth

We examined gene expression in 28 mutation accumulation lines derived from two wild-type *C. reinhardtii* strains: CC-2391 (n=13) and CC-2344 (n=15) [37,41,42]. These MA lines have been

intensively studied, including characterization of the location of each mutation, the mutation rate of each line, and the distribution of fitness effects of these mutations. The MA lines were grown for 800-1100 generations and carry a mean of 76.2 mutations in CC-2344 and 112.1 mutations in CC-2931. To measure expression change from *de novo* mutation, we grew three replicates of each MA line and their unmutated ancestors synchronously on Bold's agar medium at 25ºC under standard, uniform lab conditions. Each replicate was grown until the density of colonies was sufficient for RNA isolation, then frozen at -80ºC until extraction. RNA extraction was conducted via Maxwell RSC 48 Instrument and assessed for degradation, purity, and quantity using gel electrophoresis and a Qubit fluorometer. mRNA isolation and library preparation were conducted with the NEB mRNA stranded library preparation kit and Illumina NovaSeq preparation. All samples were sequenced using Illumina 100bp paired-end sequencing with at least 20 M high-quality reads per sample. Sequencing and library preparation were performed by Genome Quebec.

Read count preparation

Sequence quality of raw reads was assessed with FastQC (*v0.11.9*) [43] and aligned using STAR aligner (*v2.7.8a*) [44]. We used the Chlamydomonas v6.0 reference genome with the mt+ mating type allele, and organelle genomes included. STAR alignment was conducted in two-pass mode, such that the first pass identified possible splice junctions (SJ), and the second pass utilized the reference gene annotation in conjunction with the SJ output. Parameters of both alignments included "--alignIntronMax 5000", which excluded known false large introns identified in other STAR-mapped BAM files. The resulting BAM files were sorted and indexed with samtools [45]. Sequence coverage and mapping quality was measured by the bamqc and rnaseq function of Qualimap quality control package (*v2.2.2a*) [46]. The tool HTSeq-count (*v0.13.5*) [47] was used to quantify transcripts per gene. Low read counts were filtered out via NOISeq (*v2.36.0) [48,49]* and normalized for gene length and GC bias with CQN (*v1.38.0)* [50].

Differential gene expression analysis

We identified differentially expressed genes (DEGs) in each MA line with respect to its unmutated ancestral line with DESeq2 (*v1.30.1*) [51]. DESeq2 uses Wald's test, adjusted *p*-value cutoff < 0.05 for DEG calling, and the BH method was used to account for multiple testing (*FDR < 0.05*).

Quantifying expression divergence contributed by mutations in DEGs

To determine the effect of *de novo* mutation on gene expression, we quantified the change of gene expression relative to their respective unmutated ancestors as $\log_2$-fold (L2F) change. We estimated both

the change in individual genes and the total expression divergence from the ancestor as the sum of expression changes across all genes. We also tested for a bias in the direction of expression divergence by separating each MA line's total expression change into their positive (up-regulation) and negative (down-regulation) components.

Additionally, the average L2F change per *de novo* mutation was calculated by dividing the sum L2F change (positive and negative components treated separately) in a MA line by the average number of mutations per strain.

The direction and distribution of expression change across FPKM percentile bins

We calculated the mutational variance as $2\sigma^2/t$ where σ^2 represents the among-line variance and t is the average generation time [28]. The among-line variance is estimated from the deviation of each MA line from its unmutated ancestor. We also estimated the increase in expression variance per *de novo* mutation by dividing the among-line variance by the average number of mutations per strain. Our estimations are likely to be understated since the average mutation count was skewed by two lines that likely sustained damage to their DNA repair system. To test whether mutation's effect on expression depends on the expression level in the ancestral genotype, the FPKM (fragments per kilobase of exon per million mapped fragments) of ancestral gene read counts was calculated and categorized into 10-percentile bins from lowest to highest FPKM. Each bin contained 1527 and 1583 genes for wild-type strains CC-2344 and CC-2931 respectively. To estimate expression change in each gene within the percentile bins, we calculated the average L2F change across all MA lines within a strain. With this, we examined the distribution of expression changes within each bin to determine whether the mean or variance in expression correlated with ancestral expression level and whether there was a bias in the direction of expression change (up- vs. down-regulation).

Gene function and network analysis

We tested whether genes with essential functions or genes with more central location in the *C. reinhardtii* metabolic network had less expression change induced by *de novo* mutation. We used the metabolic network iCre1355 to determine the positions of DEGs in the metabolic network and measure the number of connections to genes (i.e., degree) [52]. The iCre1355 network was converted to a gene centric network following [53].

We also sought to test whether DEGs tended to be closely positioned (clustered) in the Chlamydomonas co-expression network [54]. If DEGs tended to be closer in the network it may indicate that they are

controlled by common trans-acting regulators. To test for this, we calculated the shortest path in the network between each DEG and all other DEGs in that MA line using the multi_source_dijkstra function in the NetworkX python package [55]. Edges in the co-expression network represent significant correlations in expression among genes where we defined the strength of the relationship between genes as distance = correlation coefficient^{-1}. Using this network the shortest edges represent the strongest correlations. We tested our observed distribution of shortest paths in each MA line against a null generated from the shortest path distributions of 100 random samples of *n* DEGs, where *n* is the number of DEGs in a given MA line. The function multi_source_disjskra was amended to consider unconnected DEGs returning a distance of zero; these values were excluded from the analysis.

The code used for the analysis can be found at www.github.com/enibalogun/MAexpression and the raw data at the European Nucleotide Archive [PRJEB51708/ERP136361].

Results and Discussion

In this study, we examine how >2600 mutations across 28 MA lines have altered gene expression in the transcriptome of *C. reinhardtii*. We estimated the rate at which the transcriptome diverged by calculating the fold-change in expression from the ancestor added each generation due to mutation. On average each MA line had 177 differentially expressed genes (DEGs), with more in MA lines derived from strain CC-2344 (mean 219 DEGs/MA line) than CC-2931 (mean 130 DEGs/MA line). The median mutational variance (V_m), which was calculated using the $\log_2$ fold-change per gene across MA lines within a strain, was 3.6×10^{-3} (95% CI: [2.81×10^{-3}, 3.79×10^{-3}) and 2.1×10^{-3} (95% CI: [1.59×10^{-3}, 2.13×10^{-3}) for CC-2344 and CC-2931 respectively. It is noteworthy that the transcriptome of CC-2344 had a V_m 1.7× higher than that of CC-2931, indicative of a genetic difference in their vulnerability to mutational change (ANOVA, $p < 0.05$; Bootstrapping, $p < 10^{-4}$). It is difficult to directly compare V_m between different organisms due to past studies' use of different experimental designs and technologies (e.g, microarray vs RNASeq) but our estimates of V_m were on average 57× higher than past estimates in Drosophila [23,25,26], 6× that of *C. elegans* [22] and 13× that of *S. cerevisiae* [24]. Our results demonstrate that the transcriptome of *C. reinhardtii* is particularly prone to perturbation by spontaneous mutation, which we explore in more detail below.

The effect of *de novo* mutation on expression is not additive

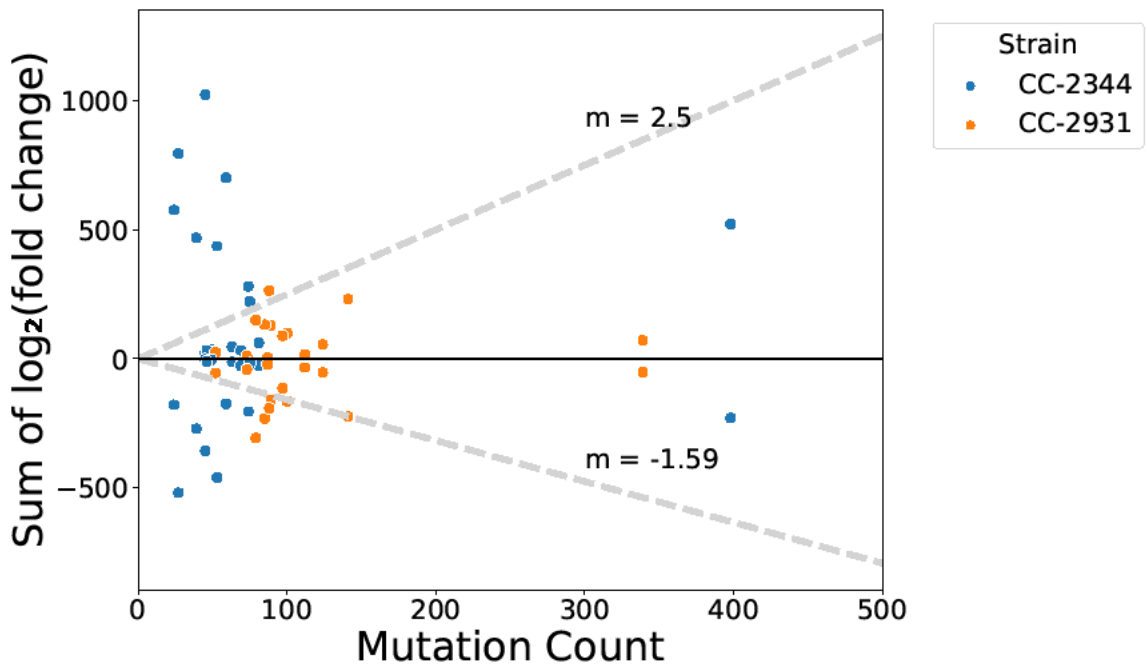

Fig 1. No correlation between total expression change and mutation count.

Sum of total expression change of significantly differentially expressed genes per MA line against the total number of mutations carried by that line. Total positive and negative $\log_2$fold changes (coloured dots) were summed separately and represented above and below the horizontal line. The dashed line represents the average rate of fold-2 expression change per mutation.

The large disparity between V_m in *C. reinhardtii* and other species could be driven by different mutation counts per generation among species. V_m is generally measured per generation but amongst the few species we have data for, the number of expected mutations per generation varies from 1 mutation in 285 generations in yeast [56–59], to 1.7 mutations per generation in *D. melanogaster* [26,60–62]. Using these expected mutation counts per generation and our own detailed information on the mutations in each MA line, we were able to calculate the variance contributed by each mutation, V_{MM}. The median V_{MM} was 2.2 x 10^{-2} (95% CI: [1.14 x 10^{-2}, 3.41 x 10^{-2}) for CC-2344 and 9.7 x 10^{-3} (95% CI: [5.78 x 10^{-3}, 1.18 x 10^{-2}) for CC-2931, which is on average 264× higher than that of Drosophila, 8× that of *C. elegans* and 0.52× that of *S. cerevisiae*. Although the taxonomic representation is extremely limited, viewing V_{MM} in this context may indicate that multicellular organisms, with differentiated cell types may experience less expression change per mutation, but without more comprehensive sampling the observed trend is mostly speculative.

Contrary to the simple prediction that more mutations would cause more expression divergence, we found no correlation between total absolute $\log_2$-fold (L2F) expression change and the number of mutations in a MA line (R^2 = -0.074, Fig 1). CC-2344, which had fewer mutations, had more DEGs and a larger absolute fold expression change (absolute = 7760, positive-sum = 5264, negative-sum = -2496) than CC-2931(absolute = 2932, positive-sum = 1281, negative-sum = -1651) (Fig 1). Because measurements of fold-change may be skewed by large relative changes in low expression genes, we also tested whether the sum of expression level (FPKM) differences relative to the unmutated ancestor or the number of DEGs per line was related to mutation count (fig S1), but similarly found no correlations. One possible explanation for the lack of relationship between mutation count and expression divergence is that many of the mutations may have had little effect on expression variance, as was noted by McGuigan et al. [25]. If the effects of individual mutations on expression are highly variable, then the correlation between mutation count and expression might be expected to be weak. Variable mutational effects and a weak correlation might be expected if mutations that influence expression are mostly trans-acting, where a small number of mutations affect many genes while others have no influence at all, as is seen in eQTL hotspots (reviewed in [63]).

Evidence of significant trans-regulatory mutations

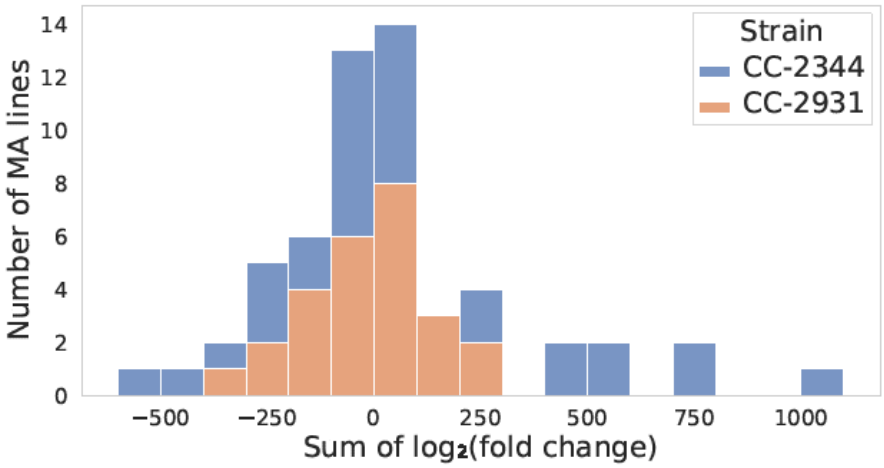

Fig 2. Nearly symmetrical distribution of summed expression changes across MA lines.

Stacked bar plot representing the number of MA lines with a given amount of total positive or negative expression change ($\log_2$-fold). Negative and positive values are summed separately to display up- and down-regulation, such that each MA line is represented twice.

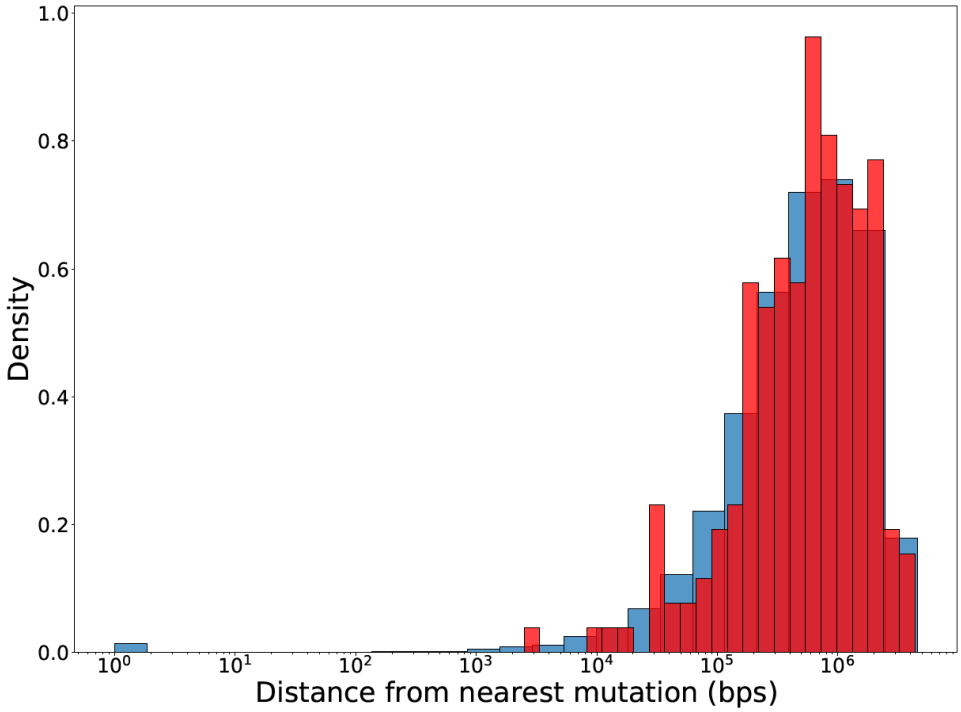

Fig 3. Evidence of cis-regulatory mutations in representation plot.

Overlapping histogram of representative plot displaying the distance between a given DEG and its nearest mutation. The observed distribution for one representative MA line is shown in red, and the expected distribution based on 10000 trials where the positions of DEGs in the genome are randomized is shown in blue. The excess of DEGs co-localizing with mutations is apparent in the excess of observed data on the left of the distribution.

Further evidence of mostly trans-acting regulatory mutations comes from the simple fact that there are more DEGs than there are mutations in each line. The ratio of DEGs to mutations was 219/76.2 for MA lines from CC-2344 and 130/116.8 for those from CC-2931. The actual ratio of DEGs to expression-changing mutations is likely even higher given that many mutations will have no effect on expression. It has also previously been shown that while cis-regulatory mutations tend to reduce expression, trans-acting mutations have smaller effects with no bias in direction. However, the trans- acting mutations that did increase expression tended to have a larger effect size than those that reduced expression [17,18,64]. We observed this expected pattern in CC-2931 (Fig 2). Additionally, we see that MA lines from CC-2344 had more expression differentiation skewed towards more upregulated genes, while MA lines from CC-2931 had a more even distribution of expression changes, with only a slight bias towards down-regulation (Fig

2). It is worth noting that these differences between MA lines from different backgrounds are not completely consistent, with some individual MA lines displaying bias in the opposite direction from their strain mean (fig S2).

We also tested whether there was a signal of cis-acting mutations in our data. We reasoned that if there was a substantial number of cis-acting mutations, we would see an over-representation of DEGs with nearby mutations. To generate a null distribution of the expected distance of a DEG to its nearest mutation, we conducted 10000 simulations where the relative positions of mutations and DEGs were randomized by assigning random genes as DEGs, which maintains the non-uniform distribution of mutations. When we compared the observed to null distribution, we found an enrichment of DEGs with mutations within 100bp of the gene flanks in 7 of 28 lines (Permutations test, $p < 0.05$; fig S3) with another five lines that were only marginally insignificant ($0.05 < p < 0.1$). In Fig 3 we present a representative sample from one MA line (CC-2931 L15) showing significant evidence of cis-regulatory mutations indicated by the higher fraction of observed mutations near DEGs than expected if DEGs were randomized in the genome. Of the mutations that are in DEGs, they were twice as likely to be found in UTRs as would be expected based on the amount of UTR sequence. This may not be too surprising as UTRs are enriched with promoters/enhancers and known to regulate RNA transcription, stability and translation [65]. MA lines with many DEGs were less likely to show the enrichment of nearby mutations potentially due to the presence of one or more trans-acting mutations with multiple regulatory targets that masked the signals of any cis-acting mutations that only effect one gene. Lastly, though previous experimental results suggest that cis-acting mutations tend to down-regulate genes [18], we did not see a bias in the direction of expression change in DEGs that had mutations in them.

Next, using the location of structural mutations (SMs) in eight CC-2931 MA lines (i.e., L1, L2, L6, L9, L11, L13, L14, L15) [66], we tested if SMs were associated with changes in gene expression, possibly due to changes in cis regulatory sequences or the gene sequences itself. Surprisingly, the 238 SMs examined overlapped or were adjacent to a total of 2717 genes, but on average only < 2% of those genes were differentially expressed. Due to the lack of significant expression change in genes affected by SMs, we can only conclude that trans-factors are still able to effectively regulate genes after these large mutational events.

Lastly, we tested whether DEGs are clustered in the co-expression network [54] by examining the distribution of shortest path lengths among genes. An excess of short paths compared to the null expectation was found in all 28 MA lines when considering the proportion of DEGs that were a single step away from another DEG, indicating significant clustering in the network (Permutations test, $p <$

0.05). In Fig 4, we present a plot of a representative MA line (CC-2344 L1). The fact that DEGs tend to be found near to one another in the network is further evidence of trans-acting mutation where the regulator of modules or pathways has altered multiple genes.

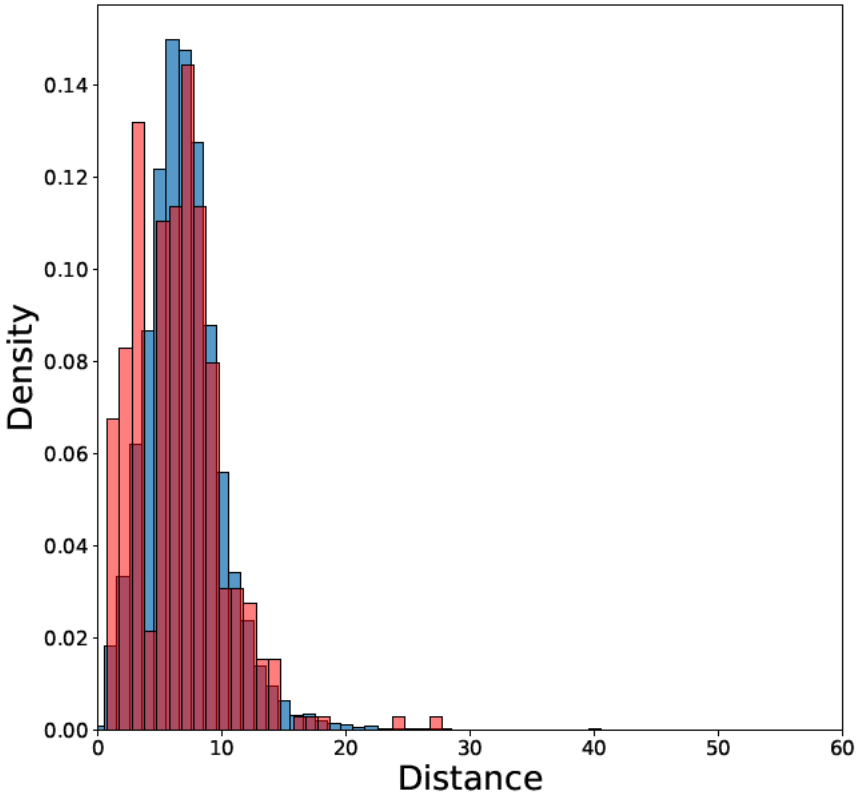

Fig 4. Distribution of shortest paths between DEGs in a coexpression network reveals signs of clustering.

Overlaid distributions of shortest path lengths of observed DEGs (red) compared to the null (blue) expectation. The plot is a single representative MA line to demonstrate the clustering of DEGs in the coexpression network. The excess of observed short paths among DEGs at the left of the plot is indicative of an excess of clustering (Permutations test, $p < 0.05$). The null distribution was created by calculating the shortest paths among a random sample of genes from the coexpression network.

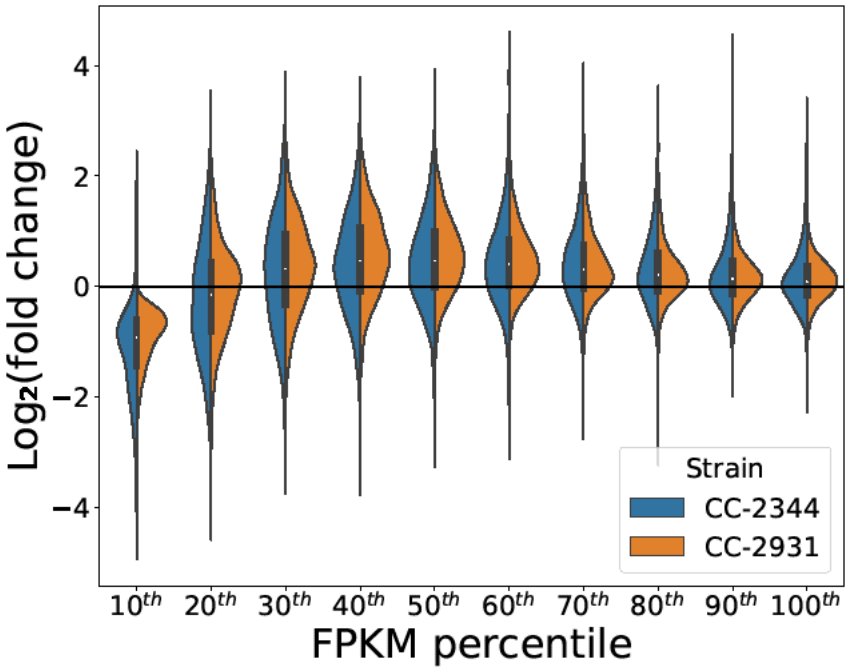

Fig 5. Distribution of $\log_2$ fold expression changes across increasing FPKM percentile bins.

Distribution of the $\log_2$ fold expression change of genes relative to their respective unmutated ancestor by ancestral expression level. The FPKM percentile bins are based on the ancestral read counts; each bin contains on average 1526 genes in CC-2344 and 1487 genes in CC-2931.

Does expression level influence the properties of expression change (i.e., direction, magnitude, variance)?

We hypothesized that genes with higher expression may have less expression divergence from mutation if they have evolved mechanisms that buffer them from perturbation by mutation. To test this, we binned genes by expression level (FPKM) based on the read counts of the unmutated ancestors. We found a clear trend of up-regulation in all but the bottom 20% of lowest expression genes (Brunner Munzel test, $p < 0.05$, Fig 5). The bottom 20 percent of genes tended to be down-regulated, with median L2F values that ranged from -1.074 to -0.323 in CC-2344 and -0.809 to -0.032 in CC-2931. For the rest of the genes, median values were skewed towards up-regulation, but declined to near zero in the most highly expressed genes. Consistent with our prediction, we also found a negative correlation between the variance in L2F change and FPKM percentile bins. Quantifying genes with notably low read counts is prone to error and noise due to a lower statistical power of detection. Thus, the accuracy of the expression change of low

expression genes should be treated cautiously. Ignoring genes in the lowest percentile bin, we note that genes with higher expression bins were more resistant to change (fig S4, Fligner-Killeen test, $p < 0.05$). We again note this same observation when comparing the mutational variance (i.e., V_{MM}) of genes in the highest percentile bin to that in the lowest. High expression genes had a comparatively lower mutational variance (One-thousand bootstrap test, $p < 0.001$; Brunner Munzel, $p < 0.001$; Fig 6, DEGs only in fig S5) consistent with findings in MA lines of *D. melanogaster* [23] and natural expression divergence between humans and mice [67]. It has been shown that genes whose expression is critical for fitness may have evolved redundant regulatory mechanisms to ensure stable expression in the face of mutations in their regulatory apparatus (reviewed in [16,20]). If genes in the highest expression categories are essential to the survival of an organism, this could explain why they have been shown to have the lowest overall expression change amongst natural lines.

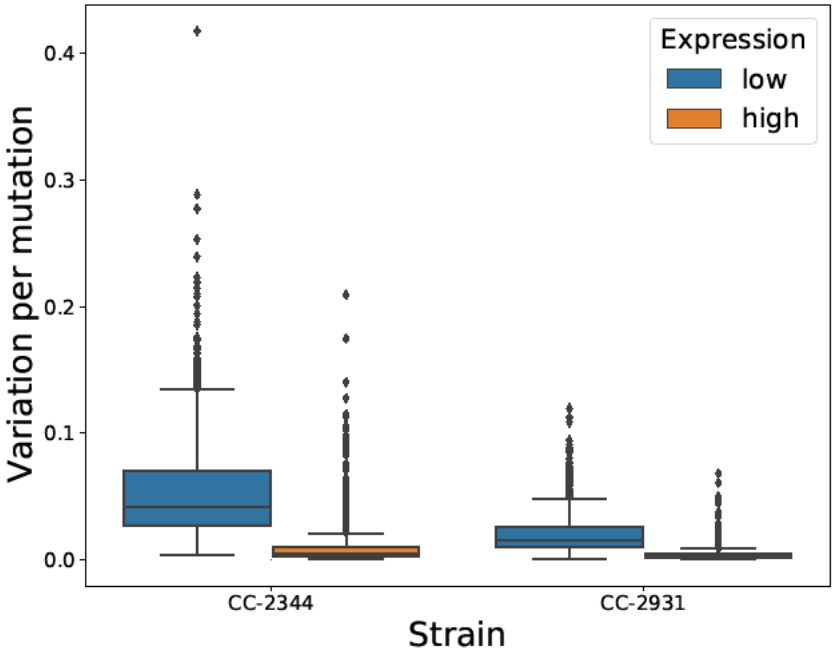

Fig 6. High expression genes are more robust to expression change due to mutation.

Comparison of the distribution of mutational variance (V_{MM}) in high (90th FPKM percentile) and low (10th FPKM percentile) expression genes across both strains. Each FPKM bin contains on average 1526 and 1487 genes for CC-2344 and CC-2931 respectively. High expression genes have significantly lower expression variation due to mutation than that of low expression genes (Brunner Munzel test, $p < 0.001$).

Degree of connectedness in metabolic genes is not indicative of expression change

The centrality-lethality rule describes the concept that the genes with the most connections to other genes in an interaction network are more likely to be critical for fitness [6–11]. There is clear support for this idea from experimental and comparative genomics [5,12,68–71]. It is also reasonable that the expression level of central genes is also subject to strong stabilizing selection, and that mechanisms may have evolved to protect central genes from mutations that alter their expression. We examined the extent of expression change induced by mutation in the most and least central genes as measured by degree and betweenness centrality in the *C. reinhardtii* metabolic network iCre1355. We failed to find a difference in the distribution of the fold-2 change and mutational variance of the metabolic genes with the highest and lowest centrality (Brunner Munzel, $p > 0.05$). One possible explanation for this is that so many of the genes in the metabolic network are constrained that they would all demonstrate mutational stability. However, there was no significant difference in the expression change of genes in the metabolic network and those genes not included in the network (Brunner Munzel, $p > 0.05$). Additionally, we found that metabolic genes had significantly less non-synonymous genes than non-metabolic genes (Brunner Munzel, $p < 0.05$). Our results could imply that metabolic genes have not developed any additional robustness to mutation than other non-metabolic genes or there is a smaller subset of more functionally important metabolic genes.

Do lines with more expression changes suffer greater reductions in fitness?

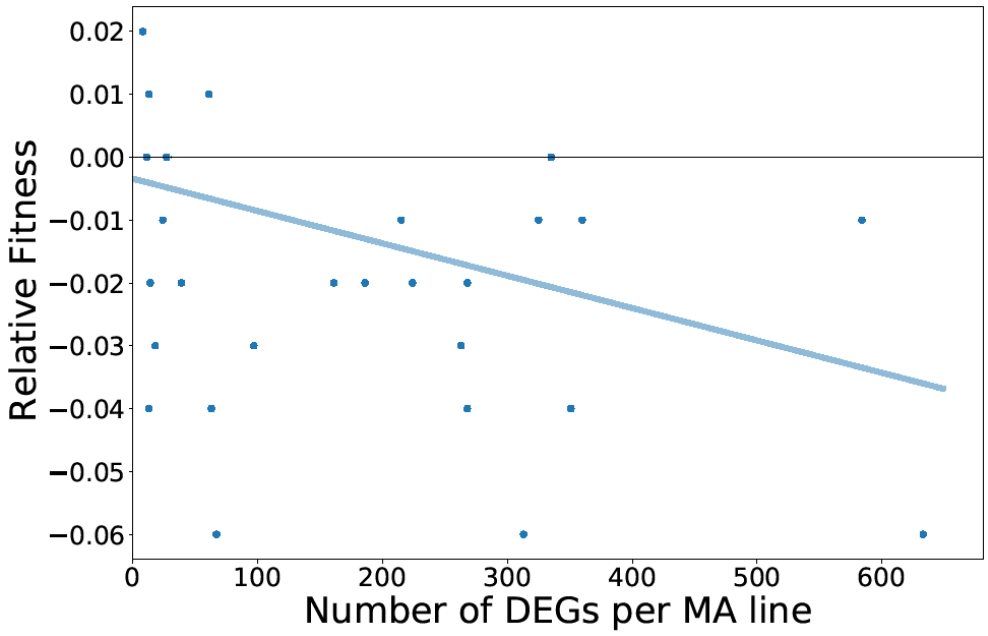

Fig 7. Evidence of fitness decline in lines with more differentially expressed genes.

Linear regression modelling the relationship between the number of significant differentially expressed per MA line and fitness relative to the unmutated ancestor ($p < 0.05$, $R^2 = 0.243$)

It has been shown that expression is often under stabilizing selection [22,23,26], thus we would expect MA lines with more total expression change to equally experience more fitness decline. Consistent with this prediction, we found a negative correlation between the number of DEGs per MA line and fitness (Fig 7), and a similar relationship between the sum of absolute L2F change and fitness. The association between declining fitness and increasing number of DEGs implies the deleterious nature of large changes in the transcriptome due to mutation. Our results support other lines of evidence that show stabilizing selection is acting on gene expression [22,23,26]. The fact that we found no correlation between mutation count and expression divergence similar to previous examinations of these MA lines where little to no correlation of mutation count or annotation with fitness were found [42] demonstrates the difficulty of predicting the effect of mutation from DNA sequences. However, the connection between expression divergence and fitness that we see provides a link in predicting the fitness effects of mutations via their effects on expression. Our results are in line with a recent study [17] which showed that mutations with trans-regulatory effects in yeast have more deleterious pleiotropic effects on fitness. In combination with our results, which describe the effects of a random set of *de novo* mutations, there is clear evidence that cells are vulnerable to harmful mutations that disrupt the regulation of gene expression.

Literature cited

1. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975;188: 107–116.

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

3. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet. 2005;76: 8–32.

4. Hill MS, Vande Zande P, Wittkopp PJ. Molecular and evolutionary processes generating variation in gene expression. Nat Rev Genet. 2021;22: 203–215.

5. Liao B-Y, Scott NM, Zhang J. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. Mol Biol Evol. 2006;23: 2072–2080.

6. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 2005;22: 803–806.

7. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature.

2001;411: 41–42.

8. Yu H, Greenbaum D, Xin Lu H, Zhu X, Gerstein M. Genomic analysis of essentiality within protein networks. Trends Genet. 2004;20: 227–231.

9. He X, Zhang J. Why do hubs tend to be essential in protein networks? PLoS Genet. 2006;2: e88.

10. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol. 2007;3: e59.

11. Zotenko E, Mestre J, O'Leary DP, Przytycka TM. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. PLoS Comput Biol. 2008;4: e1000140.

12. Pál C, Papp B, Hurst LD. Highly expressed genes in yeast evolve slowly. Genetics. 2001;158: 927–931.

13. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet. 2008;40: 346–350.

14. Graze RM, McIntyre LM, Main BJ, Wayne ML, Nuzhdin SV. Regulatory divergence in Drosophila melanogaster and D. simulans, a genomewide analysis of allele-specific expression. Genetics. 2009;183: 547–61, 1SI–21SI.

15. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430: 85–88.

16. Signor SA, Nuzhdin SV. The Evolution of Gene Expression in cis and trans. Trends Genet. 2018;34: 532–544.

17. Zande PV, Hill MS, Wittkopp PJ. Pleiotropic effects of trans-regulatory mutations on fitness and gene expression. Science. 2022;377: 105–109.

18. Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ. Contrasting Frequencies and Effects of *cis*- and *trans*-Regulatory Mutations Affecting Gene Expression. Mol Biol Evol. 2016;33: 1131–1146.

19. Gruber JD, Vogel K, Kalay G, Wittkopp PJ. Contrasting properties of gene-specific regulatory, coding, and copy number mutations in Saccharomyces cerevisiae: frequency, effects, and dominance. PLoS Genet. 2012;8: e1002497.

20. Payne JL, Wagner A. Mechanisms of mutational robustness in transcriptional regulation. Front Genet. 2015;6: 322.

21. Muller HJ. The Measurement of Gene Mutation Rate in *Drosophila*, Its High Variability, and Its Dependence upon Temperature. Genetics. 1928;13: 279–357.

22. Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. Nat Genet. 2005;37: 544–548.

23. Rifkin SA, Houle D, Kim J, White KP. A mutation accumulation assay reveals a broad capacity for

rapid evolution of gene expression. Nature. 2005;438: 220–223.

24. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Genetic properties influencing the evolvability of gene expression. Science. 2007;317: 118–121.

25. McGuigan K, Collet JM, McGraw EA, Ye YH, Allen SL, Chenoweth SF, et al. The nature and extent of mutational pleiotropy in gene expression of male *Drosophila serrata*. Genetics. 2014;196: 911–921.

26. Huang W, Lyman RF, Lyman RA, Carbone MA, Harbison ST, Magwire MM, et al. Correction: Spontaneous mutations and the origin and maintenance of quantitative genetic variation. Elife. 2016;5. doi:10.7554/eLife.22300

27. Zalts H, Yanai I. Developmental constraints shape the evolution of the nematode mid-developmental transition. Nat Ecol Evol. 2017;1: 113.

28. Lynch M, Walsh B. Genetics and analysis of quantitative traits. 1998. Available: http://www.invemar.org.co/redcostera1/invemar/docs/RinconLiterario/2011/febrero/AG_8.pdf

29. Lemos B, Meiklejohn CD, Hartl DL. Regulatory evolution across the protein interaction network. Nat Genet. 2004;36: 1059–1060.

30. Zhang L, Li W-H. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol Biol Evol. 2004;21: 236–239.

31. Hine E, Runcie DE, McGuigan K, Blows MW. Uneven Distribution of Mutational Variance Across the Transcriptome of *Drosophila serrata* Revealed by High-Dimensional Analysis of Gene Expression. Genetics. 2018;209: 1319–1328.

32. Hodgins-Davis A, Duveau F, Walker EA, Wittkopp PJ. Empirical measures of mutational effects define neutral models of regulatory evolution in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 2019;116: 21085–21093.

33. Harris EH. *Chlamydomonas* as a model organism. Annu Rev Plant Physiol Plant Mol Biol. 2001;52: 363–406.

34. Grossman AR, Harris EE, Hauser C, Lefebvre PA, Martinez D, Rokhsar D, et al. *Chlamydomonas reinhardtii* at the crossroads of genomics. Eukaryot Cell. 2003;2: 1137–1150.

35. Ness RW, Morgan AD, Colegrave N, Keightley PD. Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*. Genetics. 2012;192: 1447–1454.

36. Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. Drift-barrier hypothesis and mutation-rate evolution. Proc Natl Acad Sci U S A. 2012;109: 18488–18492.

37. Ness RW, Morgan AD, Vasanthakrishnan RB, Colegrave N, Keightley PD. Extensive *de novo* mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*. Genome Res. 2015;25: 1739–1749.

38. Ness RW, Kraemer SA, Colegrave N, Keightley PD. Direct Estimate of the Spontaneous Mutation Rate Uncovers the Effects of Drift and Recombination in the *Chlamydomonas reinhardtii* Plastid Genome. Mol Biol Evol. 2016;33: 800–808.

39. Flynn JM, Lower SE, Barbash DA, Clark AG. Rates and Patterns of Mutation in Tandem Repetitive DNA in Six Independent Lineages of *Chlamydomonas reinhardtii*. Genome Biol Evol. 2018;10: 1673–1686.

40. López-Cortegano E, Craig RJ, Chebib J, Samuels T, Morgan AD, Kraemer SA, et al. *De Novo* Mutation Rate Variation and Its Determinants in *Chlamydomonas*. Mol Biol Evol. 2021;38: 3709–3723.

41. Morgan AD, Ness RW, Keightley PD, Colegrave N. Spontaneous mutation accumulation in multiple strains of the green alga, *Chlamydomonas reinhardtii*. Evolution. 2014;68: 2589–2602.

42. Böndel KB, Kraemer SA, Samuels T, McClean D, Lachapelle J, Ness RW, et al. Inferring the distribution of fitness effects of spontaneous mutations in *Chlamydomonas reinhardtii*. PLoS Biol. 2019;17: e3000192.

43. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.

44. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21.

45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.

46. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32: 292–294.

47. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2014;31: 166–169.

48. Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res. 2011;21: 2213–2223.

49. Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res. 2015;43: e140.

50. Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics. 2012;13: 204–216.

51. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

52. Imam S, Schäuble S, Valenzuela J, López García de Lomana A, Carter W, Price ND, et al. A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses. Plant J. 2015;84: 1239–1256.

53. Chaiboonchoe A, Ghamsari L, Dohai B, Ng P, Khraiwesh B, Jaiswal A, et al. Systems level analysis of the Chlamydomonas reinhardtii metabolic network reveals variability in evolutionary co-conservation. Mol Biosyst. 2016;12: 2394–2407.

54. Strenkert D, Schmollinger S, Gallaher SD, Salomé PA, Purvine SO, Nicora CD, et al. Multiomics resolution of molecular events during a day in the life of Chlamydomonas. Proceedings of the

National Academy of Sciences. 2019;116: 2374–2383.

55. Hagberg A, Swart P, S Chult D. Exploring network structure, dynamics, and function using networkx. Los Alamos National Lab. (LANL), Los Alamos, NM (United States); 2008 Jan. Report No.: LA-UR-08-05495; LA-UR-08-5495. Available: https://www.osti.gov/biblio/960616

56. Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, et al. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci U S A. 2008;105: 9272–9277.

57. Nishant KT, Wei W, Mancera E, Argueso JL, Schlattl A, Delhomme N, et al. The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. PLoS Genet. 2010;6: e1001109.

58. Serero A, Jubin C, Loeillet S, Legoix-Né P, Nicolas AG. Mutational landscape of yeast mutator strains. Proc Natl Acad Sci U S A. 2014;111: 1897–1902.

59. Zhu YO, Siegal ML, Hall DW, Petrov DA. Precise estimates of mutation rate and spectrum in yeast. Proc Natl Acad Sci U S A. 2014;111: E2310–8.

60. Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. 2009;19: 1195–1201.

61. Schrider DR, Houle D, Lynch M, Hahn MW. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. Genetics. 2013;194: 937–954.

62. Sharp NP, Agrawal AF. Low Genetic Quality Alters Key Dimensions of the Mutational Spectrum. PLoS Biol. 2016;14: e1002419.

63. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 2008;24: 408–415.

64. Metzger BPH, Wittkopp PJ. Compensatory *trans*-regulatory alleles minimizing variation in *TDH3* expression are common within *Saccharomyces cerevisiae*. Evol Lett. 2019;3: 448–461.

65. Steri M, Idda ML, Whalen MB, Orrù V. Genetic variants in mRNA untranslated regions. Wiley Interdiscip Rev RNA. 2018;9: e1474.

66. López-Cortegano E, Craig RJ, Chebib J, Balogun EJ, Keightley PD. Rates and spectra of de novo structural mutation in Chlamydomonas reinhardtii. bioRxiv. 2022. p. 2022.05.23.493040. doi:10.1101/2022.05.23.493040

67. Liao BY, Zhang J. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. Mol Biol Evol. 2006;23: 1119–1128.

68. Zhang J, Yang J-R. Determinants of the rate of protein sequence evolution. Nat Rev Genet. 2015;16: 409–420.

69. Rocha EPC, Danchin A. An analysis of determinants of amino acids substitution rates in bacterial proteins. Mol Biol Evol. 2004;21: 108–116.

70. Zhang J, He X. Significant impact of protein dispensability on the instantaneous rate of protein evolution. Mol Biol Evol. 2005;22: 1147–1155.

71. Subramanian S, Kumar S. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics. 2004;168: 373–381.