Evolutionary Ecology (2022) 36:845–857
https://doi.org/10.1007/s10682-022-10210-8

ORIGINAL PAPER

Analysis of the leaf metabolome in *Arabidopsis thaliana* mutation accumulation lines reveals association of metabolic disruption and fitness consequence

Sydney Kreutzmann[1] · Elizabeth Pompa[2] · Nhan D. Nguyen[2] · Liya Tilahun[3] · Matthew T. Rutter[4] · Mao-Lun Weng[5] · Charles B. Fenster[6] · Carrie F. Olson-Manning[1]

Received: 12 November 2021 / Accepted: 15 August 2022 / Published online: 10 September 2022

Abstract

Understanding the mechanisms by which mutations affect fitness and the distribution of mutational effects are central goals in evolutionary biology. Mutation accumulation (MA) lines have long been an important tool for understanding the effect of new mutations on fitness, phenotypic variation, and mutational parameters. However, there is a clear gap in predicting the effect of specific new mutations to their effects on fitness. In an attempt to directly connect the effect of spontaneous mutations to their fitness effects, we quantified the metabolic expression of 386 known compounds in primary and secondary metabolism in *Arabidopsis thaliana* MA lines that had consistently higher and lower relative fitness than the progenitor. The high and low fitness lines do not have a difference in the average number of mutations and share the same types of metabolic pathways disrupted. However, compared to the progenitor, low fitness lines have significantly more metabolic subpathways disrupted than lines with higher fitness. These results suggest that the effect of a new mutation on fitness depends less on the specific metabolic pathways disrupted and potentially more on the number of disrupted pathways. We fail to identify any direct connection of mutations in or near well annotated genes to their effect on well-characterized biochemical pathways, possibly due to incomplete annotations of molecular function or to non-genetic variation controlling metabolic expression. Our findings indicate that organisms can explore a considerable amount of physiological space with only a few mutations.

Keywords Mutation accumulation · Spontaneous mutations · Metabolomics

✉ Carrie F. Olson-Manning
colsonmanning@augie.edu

[1] Department of Biology, Augustana University, Sioux Falls, SD 57105, USA

[2] Department of Chemistry and Biochemistry, Augustana University, Sioux Falls, SD 57105, USA

[3] Make School, Dominican University of California, San Francisco, CA 94102, USA

[4] Department of Biology, College of Charleston, Charleston, SC 29424, USA

[5] Department of Biology, Westfield State University, Westfield, MA 01086, USA

[6] South Dakota State University, Oak Lake Field Station, Brookings, SD 57007, USA

Introduction

Adaptation requires mutations that both beneficially affect a trait under selection and, importantly, do not have large deleterious effects on other traits (Fisher 1930). With the practically unlimited sequencing data now available, the challenge to predicting the effect of spontaneous mutations now rests in understanding the distribution of mutational effects and the likelihood that mutations disrupt important biological processes and fitness. Recent advances in metabolomics technology (in both separation techniques in liquid and gas chromatography and detection via mass spectroscopy (Tebani et al. 2018)) have made the qualitative and quantitative measurement of low molecular weight metabolites possible and have the potential to improve our understanding of the connection between genotypic differences and cellular physiology.

An important tool for examining the connection of genotype and phenotype, and the types and effects of new spontaneous mutations are mutation accumulation (MA) lines (Halligan and Keightley 2009). MA lines are created when sublines of an original progenitor are allowed to accumulate spontaneous mutations in the absence of selection through many generations. These lines have shed light on the distribution of phenotypic effects (Chang and Shaw 2003; Shaw et al. 2002; Weng et al. 2021), the mutational spectrum (Weng et al. 2019; Schrider et al. 2013; Denver et al. 2006; Monroe et al. 2020; Nguyen et al. 2020), and the average fitness effect of new mutations (Rutter et al. 2018; Rutter et al. 2012; Roles et al. 2016; Rutter et al. 2010) (reviewed in (Katju and Bergthorsson 2019; Eyre-Walker and Keightley 2007)).

The *Arabidopsis thaliana* MA lines (Chang and Shaw 2003; Shaw et al. 2002) offer an opportunity to connect mutation to fitness through physiology. These lines have been fully sequenced (Weng et al. 2019) and their fitness measured in field trials (Rutter et al. 2018; Rutter et al. 2012). However, there has been little use of this resource to examine the mutational effect of fitness through metabolism and physiology. With these high and low fitness lines and their associated mutations, we ask, are there qualitatively different changes in metabolic output in the low fitness lines compared to the high fitness lines? We expect high fitness lines will have less metabolic disruption or have disruption in pathways with less importance to fitness. Here, we identified a trend where high fitness lines have fewer changes to their metabolic pathways than the low fitness lines. Contrary to our initial assumptions, although we identify mutations in or flanking well-annotated genes that have known effects in well-annotated processes, we find no clear connection between the mutations and their expected effect on metabolism.

Materials and methods

Choice in plants to grow and analyze

Of the 107 total *A. thaliana* mutation accumulation (MA) lines (Chang and Shaw 2003; Shaw et al. 2002), we chose lines that showed relative fitness that was higher and lower in most environments compared to the unmutated Col-0 progenitor (Rutter et al. 2018) for metabolic analysis (Supplementary Table 1).

There is abundant evidence that de novo mutations in *A. thaliana* are frequently beneficial. One greenhouse study with *A. thaliana* MA lines showed an average decline of fitness

(Schultz et al. 1999), but did not report the full distribution of MA line performance relative to the founder. In contrast, four other greenhouse studies demonstrated significant MA line effects but no average decline in fitness among the MA lines compared to the founder; thus, beneficial mutations must be of high frequency (R. G. Shaw et al. 2000; MacKenzie et al. 2005; Rutter et al. 2010). Fifteen field studies with *A. thaliana* found that MA line average performance and founder performance were not significantly different (Rutter et al. 2018; Rutter et al. 2010; Weng et al. 2021; Roles et al. 2016), or that under stressful conditions a significant minority of MA lines outperformed the founder (Stearns and Fenster 2016; Weng et al. 2021).

The focal MA lines in this study (line numbers 44, 53, 61, 73, 75, and 119) are numbered based on the original propagation line numbers (Chang and Shaw 2003; Shaw et al. 2002). Briefly, in a previous study, Rutter et al. (2018), all MA lines were grown in four temporal environments at a single field site and plant fitness of the MA lines was compared to the unmutated progenitor. While most lines displayed genotype-by-environment dependent fitness, for this study we chose lines that consistently ranked among the highest or lowest in relative fitness compared to the progenitor (Supplementary Table S1). As defined in this manuscript, the "high fitness lines" (numbers 53, 61, and 119) had higher relative fitness than the progenitor in at least three of four environments and the "low fitness lines" (lines 44, 73, and 75) had lower relative fitness than the progenitor in at least three of four environments (Rutter et al. 2018). These lines have accumulated an average of 17 mutations over 30 generations of mutation accumulation (Weng et al. 2019).

GO Term Enrichment comparison in high and low fitness lines

To identify whether there were gene ontology terms that might be over- or under-represented in the high or low fitness lines, we entered the *A. thaliana* gene model names for mutations (or the gene model names surrounding intergenic mutations) into the Protein Analysis Through Evolutionary Relationships (PANTHER) resource, GO Ontology database DOI: 10.5281senodo.4495804 Released 2021-02-01 (Mi et al. 2021) (gene names found in Supplementary Table S2). We compared the set of mutations in the high and low fitness lines for biological processes, molecular function complete, and cellular component complete with the Fisher's exact test with Bonferroni correction for multiple testing. The analyses were performed separately on the set of three high and three low fitness lines of focus in this study.

We also conducted GO term analysis for pairs of genes with intragenic mutations between them. We repeated the analysis and submitted only genes either containing mutations with a mutation within 1500 bp of the start codon or 500 bp downstream of the stop codon (Korkuc et al. 2014). Finally, we conducted the same GO term analysis for the additional 15 lines that had average relative fitness that was higher and 8 with relative fitness lower than the progenitor (Rutter et al. 2018). When we repeated the GO term analysis on the mutations in each line, none had significant enrichment. Here we report all of the significant enrichment results we identified.

Plant growth and tissue harvest

Individual seeds from the 25th generation of mutation accumulation were sown on moistened soil (Promix BX with Osmocote fertilizer added per manufacturer's instructions), randomized in a 24-cell pallet and allowed to imbibe. The seeds were placed in the dark

for 3 days at 4 °C to stratify and overcome dormancy. Plants were grown for 3 weeks under long day conditions (16 h light) at 18 °C. The plant rosettes at time of harvest were vigorous and did not display any obvious signs of stress or growth defects. To control for the developmental stage, all above-ground tissue was harvested before the plant began bolting and was immediately frozen on liquid nitrogen. Each of the high and low fitness lines had four biological replicates and the progenitor had six. The tissue was stored at -80 C until further processing. Frozen leaves were pulverized on liquid nitrogen and added to a frozen collection tube and stored at -80 C until shipment. Samples were shipped on dry ice to Metabolon Inc. (North Carolina). Frozen samples were lyophilized and each sample was standardized with dry weight.

Metabolon sample extraction and metabolite identification

Briefly, Metabolon Inc. extracted the samples with the MicroLab STAR System (Hamilton Company) with the internal standards. Each extract was divided into five fractions: two were analyzed using acidic positive ion conditions (one optimized for hydrophilic compounds, one for hydrophobic), two used basic negative ion conditions (one optimized for hydrophilic compounds, one for hydrophobic), and one used for a backup. Metabolomic analyses were performed using Ultra High Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC MS/MS). These methods utilized a Waters Aquity UPLC and Q-Exactive high resolution mass spectrometer (Thermo Scientific) with a heated electrospray ionization (HESI-II) source. The MS analysis used dynamic exclusion, alternating between MS and data-dependent MS^n, and varied between 70 and 1000 m/z. The MS/MS scores were derived from a comparison of the ions in the experimental samples and the ions in a known library spectrum (within the Metabolon LIMS system. The MS/MS scores (forward and reverse), retention time/index, and molecular mass (m/z) were all processed using Metabolon's software and compared to purified standards within the internal library to identify each metabolite. Peaks were quantified by integrating the area-under-the-curve. Samples analyzed over multiple days were corrected in run-day blocks and normalization based on internal standards.

Identification of disrupted metabolites and pathways

A total of 386 known compounds were identified in this dataset. Missing data points were given the minimum observed value for each compound and all data were log transformed. The dry weight of each starting tissue was also used for normalization. Each metabolite was categorized into one of nine super pathways (e.g. Amino Acid, Carbohydrate, Peptide, Secondary Metabolite) and into one of 58 subpathways (e.g. Serine family amino acid, TCA cycle, Dipeptide, Benzenoids) (Supplementary Table 3) (Kanehisa and Goto 2000).

To identify whether there was significant among-line variance in metabolites, we ran a linear mixed-effects models with the log normalized Metabolite (MetaboliteValue) as a fixed effect and both Line and the metabolite identity (MetaboliteIdentity) as random effects (MetaboliteValue ~ 1 + (1|Line) + (1|MetaboliteIdentity)) compared to a second model without line as a random effect (MetaboliteValue ~ 1 + (1|MetaboliteIdentity)) (lmer() in the *lme4* version 1.1–29 package in R version 4.1.1). We then compared the models with ANOVA is significant (anova(), stats package version 4.1.1).

Biological replicate measures of each metabolite were compared to the value in the progenitor with Welch's two-sample t-test. A metabolite was judged to be different from

the progenitor in a certain line if its value was significantly different from the value in the progenitor based on the t-test at an alpha value of 0.05.

We took two approaches to identify enrichment in metabolic subpathway disruption to balance Type I and Type II errors. With a pathway enrichment metric (Xia and Wishart 2011), the number of significant ($p \leq 0.05$) compounds were calculated for individual metabolic subpathways. The Enrichment score (Xia and Wishart 2011) compares the ratio of significant metabolites in a pathway (k) to total detected metabolites in that pathway (m), and standardized it by considering the total number of significant metabolites in the entire dataset (n) and the total metabolites detected in the dataset (N) (Eq. 1) (results averages in Table 1 and calculations in Supplementary Table S4).

Equation 1 Enrichment score = (k/m) / ((n-k)/(N-m)).

Due to the complex nature of biological interactions, and the small sample size used in this study, the traditional enrichment score is likely conservative (Nguyen et al. 2019). The traditional enrichment score ignores individual pathways with many metabolites that approach significance in the Welch's two-sample t-test ($p < 0.1$). As a less conservative test, we also calculated an expanded enrichment score that includes metabolites that approach significance (Table 1, Supplementary Table S4). This test allowed us to identify metabolic subpathways that have sets of metabolites that are disrupted and share common biological function. The trend that lines with relatively low fitness have more disruption is consistent across the traditional enrichment analysis, the extended analysis, and the number of metabolites.

Comparison of up and down regulation of metabolic pathways

We compared the number of metabolites that were significantly up- or down-regulated as compared to the Col-0 progenitor with a Chi-square statistic. The null expectation is that spontaneous mutations will cause an equal number of up- and down-regulation in metabolites.

Table 1 The number of enriched metabolic subpathways and individual metabolites calculated in a comparison between a specific *A. thaliana* MA line and the progenitor Col-0 that met the 0.05 significance threshold and those that met the 0.1 threshold

Line number	Relative fitness	Enriched subpathways $p < 0.05$	Enriched subpathways $p < 0.1$*	Total metabolites $p < 0.05$	Total metabolites $p < 0.1$
44	Low	10	21	13	37
73		20	29	32	60
75		12	25	13	39
53	High	14	17	16	25
61		10	14	18	35
119		7	15	9	25

Significant difference between the number of enriched subpathways that reach the $p < 0.1$ threshold denoted with * ($p < 0.05$ unpaired t-test)

Results and discussion

The ways in which mutations affect organismal fitness is of great interest in evolutionary biology and molecular evolution. Mutations have been used to identify metabolic pathways for over eighty years (Beadle and Tatum 1941). However, we lack an empirical and general understanding relating mutations to alterations of metabolic pathways and in turn these effects on individual fitness. In our analyses, we attempted to quantify effect of mutations on organismal metabolism and to identify the composition of spontaneous mutations in *A. thaliana* MA lines with high and low relative fitness (measured under field conditions).

Spontaneous mutations affect metabolism in varied ways

The effect of new spontaneous mutations on metabolism is varied (Figs. 1 and 2) and suggests that the mutations accumulated in these lines are idiosyncratic with respect to effect on metabolic pathways. In a comparison of a linear mixed-effects model we confirmed there was significant among-line variance with a nested ANOVA of metabolite identity and line as random effects compared to a model with only line as a random effect ($p < 2.2e\text{-}16$), In a comparison of all measured metabolites in the seven lines (the Col-0 progenitor and the three high and three low fitness lines) with PCA (Fig. 2) the high and low fitness lines are intermixed suggesting that there are not consistent changes to the types of metabolic disruption that led to high or low fitness. These results are also consistent with the heatmap generated from the log-fold data for all metabolites (Supplementary Fig. S1). These results suggest that the spontaneous mutations were indeed random with respect to their effect on metabolism (Fig. 3).

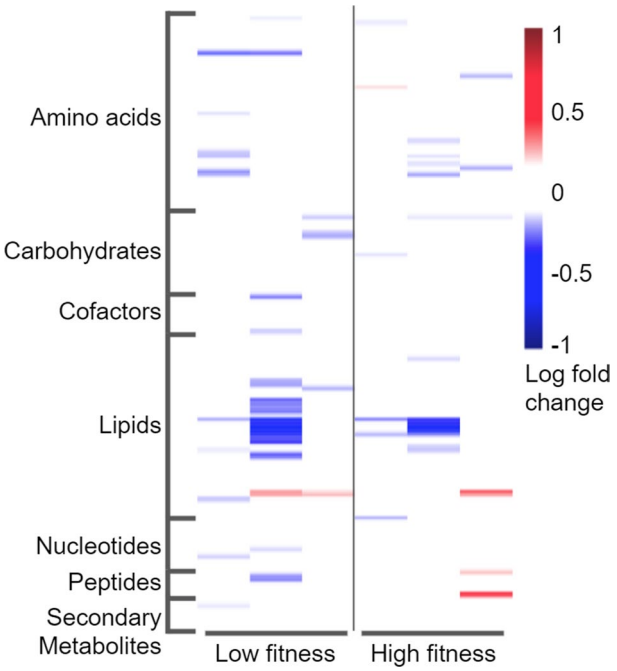

Fig. 1 Log fold change of all significant and near significant metabolites. The columns from left to right are the low fitness lines (44, 73, 75) and high fitness lines (53, 61, 119). The rows are ordered based on superpathway (full data in Supplemental Table 3)

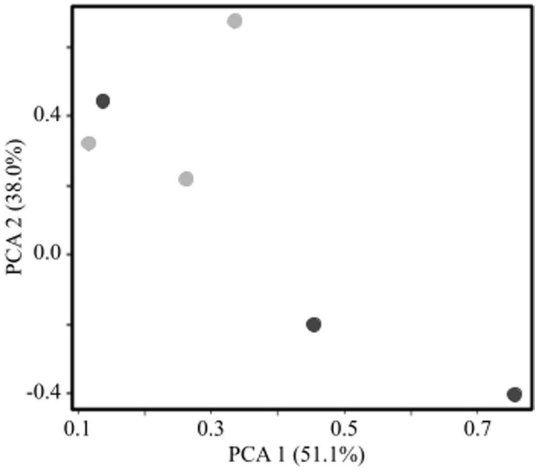

Fig. 2 Principal component analysis (PCA) of the fold-change metabolomics data with the percentage of variance explained by each principal component noted in parentheses. The lines with high relative fitness (dark gray points) and the lines with low fitness (light gray points) do not group according to fitness

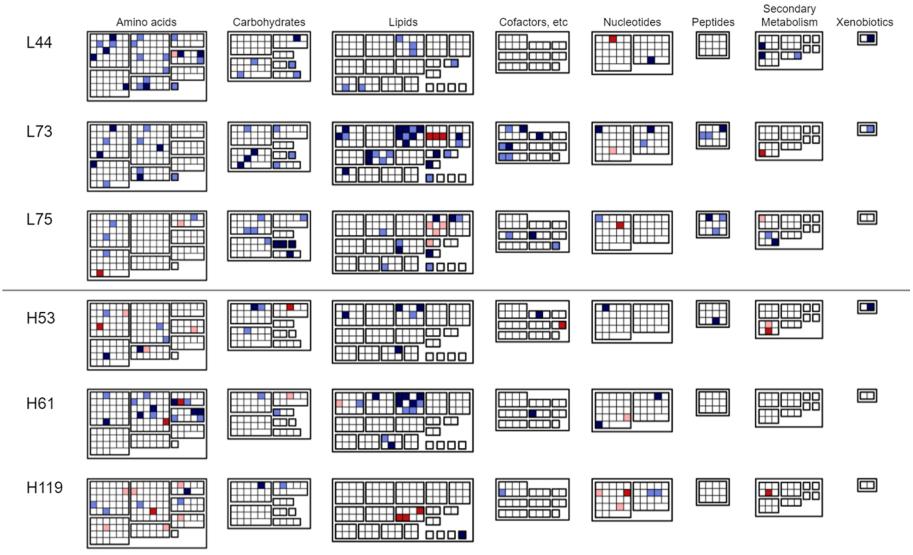

Fig. 3 Graphical depiction of disruption in the low and high fitness lines. Up-regulation (light and dark red) and down-regulation (light and dark blue) if metabolites (small boxes) in their super pathways context (labeled on top of each large enclosing box) with subpathways (small boxes in large enclosing boxes) for *A. thaliana* MA lines (low fitness lines 44, 73, 75; high fitness lines 53, 61, 119). Lighter colored boxes represent metabolites that approach significance ($p < 0.1$) and darker colors for $p < 0.05$ following a Welch's two-sample t-test. We excluded hormone metabolism and partially characterized molecules metabolism from this figure that did not show any significant or near-significant differences. A key of this metabolite map for each metabolite is found in Supplementary Fig. S2

Spontaneous mutations cause excess down-regulation of metabolic pathways

We found that overall, mutations caused significantly more down-regulation to metabolic pathways than up-regulation in both the high and low fitness lines when compared

to the progenitor (Fig. 1, blue indicates down-regulation, red is up-regulation, Chi-square $p < 0.007$). Down-regulated metabolites are 4-times more common than up-regulated metabolites and there was no significant difference in the number of up- and down-regulation in a comparison of the high and low fitness lines. It is possible that mutations are more likely to lead to down-regulation because mutations tend to disrupt steps in metabolic pathways consistent with the general thought that mutations are deleterious (Keightley and Lynch 2003).

While it might be reasonable to assume that general degradation of metabolic function is deleterious, there are many instances of loss of function alleles contributing to adaptation (Stower 2013; Xu et al. 2019; Monroe et al. 2020; Monroe et al. 2018). In *A. thaliana*, loss-of-function alleles are common in coding genes, and an estimated 1% of loss-of-function alleles are under positive selection (Xu et al. 2019). Therefore, without further measure of allele frequency changes in the MA lines following selection, it is not possible to speculate on the evolutionary importance in an overall reduction in metabolic activity and it is likely that some of the down-regulation has adaptive potential.

Metabolic disruption in high and low fitness lines

We find a consistent trend of more metabolic subpathways disruption in the low fitness lines than in the high fitness lines. While the high and low fitness lines have no average difference in the number of mutations outside of transposable elements (average of 8.7 for high fitness and 8 for low fitness lines, two-tailed t-test, $p = 0.84$, Supplementary Table S1), there is considerable variance in the number of mutations among the lines. There is no relationship between the number of mutations and the number significantly disrupted metabolic subpathways (F-statistic 0.017, $p = 0.9026$). On average, each mutation affects multiple metabolic pathways. Mutations in four of the six lines affect 1.5–3.5 metabolic subpathways per mutation (Supplementary Table S1). High fitness line 119 has an average of fewer than one metabolic subpathways per mutation.

The low fitness lines have an average of 25 disrupted subpathways compared to an average of 15.3 disrupted subpathways in the high fitness lines when we consider all metabolites that approach significance (two-sided unpaired t-test, p-value = 0.02) (Table 1, Supplementary Table S3). This pattern of a higher average number of disrupted metabolites in low fitness lines is consistent at the super pathway, subpathway, and individual metabolite levels, regardless of whether we include only metabolites that reach a the typical significance threshold $p \leq 0.05$ or metabolites that approached significance ($0.05 < p < 0.10$ (Table 1, Supplementary Table S4). Overall, these results suggest that the mutations in the low fitness lines change flux through more metabolic pathways than the mutations in high fitness lines.

Based on these data, it appears that *A. thaliana* can explore a considerable amount of physiological space with only a few mutations. Our analyses indicate overall erosion of metabolic function and previous research could indicate that a side-effect of metabolic degradation could be specialization to specific environments. In *E. coli*, spontaneous mutations contribute to metabolic specialization due to the erosion of unneeded pathways (Leiby and Marx 2014) with genetic drift strengthening this effect (Aguilar-Rodríguez et al., 2019). In heterogeneous temperature environments, *E. coli* have impaired growth on all carbon resources at low temperatures, but little or no growth defects at higher temperatures (Chu and Zhang 2021), suggesting that the average effect of mutations depends on the specific environment.

High fitness lines are enriched in mutations in or near genes with transcription regulator activity

In a Gene Ontology (GO) term analysis for molecular function (Mi et al. 2021) we found genes in the high fitness lines submitted for metabolomics in this study have a 7.7-fold enrichment (Fisher's Exact Test with Bonferroni correction, $p < 0.01$) in transcription regulatory activity (8 of the 16 mapped gene IDs and represent between 25 and 50% of the mutations in each line). In this initial analysis, if a mutation were intergenic, we included both flanking genes in this analysis. Our logic was that if the mutation was impacting the expression of only one of the flanking genes, including both in the GO analysis would most likely dilute any patterns and make it less likely for us to detect an effect. Thus, identifying the enrichment suggests that mutations in the high fitness lines are more often in or in the intergenic region around genes with transcriptional regulation.

We repeated this same GO term analysis for all 18 high fitness MA lines (Supplementary Table S2) and found that together they also have a greater than 3.1-fold enrichment for transcription regulator activity ($p < 0.001$, 21 of 114 mapped gene IDs have transcription regulator activity). We identified no other significant enrichment in the genes in the high fitness lines. The same analyses on all 11 low fitness lines revealed no enrichment in any activity and only one out of 32 genes with mutations in the low fitness lines has transcriptional regulatory activity.

Although most gene regulatory elements are within 0.5 to 2 kb of the transcription start site (and *A. thaliana* has a more restricted 500 bp upstream and 250 bp downstream), it is well documented that regulatory elements far from the promoter can impact gene expression (Benfey and Chua 1989). When we modify our GO analysis, only including genes in the analysis if the mutation lies within 1500 bp of the transcription start site, we retain only half of the genes with transcription regulatory activity and the enrichment in transcriptional regulatory activity does not pass a false discovery rate correction.

While these results are conflicting, the possible enrichment of genes with mutations in or in the intergenic region flanking genes encoding transcriptional regulatory proteins make up a substantial portion of the mutational spectrum in MA lines with high relative fitness and are less prevalent in lines with low fitness. Proteins with transcriptional regulatory activity assemble into varied complexes consisting of other proteins and cis-regulatory sequences in a combinatorial fashion to tightly control the timing and location of gene expression (Brkljacic and Grotewold 2017). In this way, a small handful of transcription factors can control the expression of a wide variety of genes. The vast majority of the mutations in our mutation accumulation lines are intergenic, and thus these mutations may affect the regulation of the identified proteins with transcription regulatory activity (the mutations are in the cis-regulatory regions or in trans enhancers or repressors. The combinatorial nature of these regulatory networks have led many to hypothesize that mutations that affect gene regulation could play an outsized role in adaptive evolution (Jeong et al. 2008; Wray et al. 2003). The evolutionary genetics literature has many examples implicating mutations in proteins with transcriptional regulatory activity in adaptations, including flower color evolution (Streisfeld et al. 2011; Wessinger and Rausher 2013; Lin and Rausher 2021; Quattrocchio et al. 1999; Whittall et al. 2006), adaptation to high temperature (Koini et al. 2009), cold acclimation (Buskirk et al., 2006), drought (Haake et al. 2002; Leng and Zhao 2020; Jan et al. 2019), Zinc deficiency (Inaba et al. 2015), the interface between stress and growth response (Danisman 2016) and many more. Given the purported role of proteins with transcriptional regulatory activity (Wagner and Lynch 2008), and their

prevalence in the mutational spectrum in lines with high relative fitness, these mutations are readily available for adaptation.

The potential and limitations of metabolomics studies in understanding the effects of spontaneous mutations

In this study, we used metabolomic measurements to try to close the gap between knowing the identity of a mutation and understanding and predicting its effect on fitness (Eyre-Walker and Keightley 2007). We find that although our results are consistent across all metrics of metabolic change, our sample size is small and we base our connections between metabolic disruption and fitness based primarily on consistent trends in the number of disruptions (and not on strong statistical or large fold-change effects) in the comparison of the lines with high relative fitness and low relative fitness. In retrospect, this is unsurprising given each mutation accumulation line has only a small number of mutations and many are intergenic. We expected to find larger effects given the consistent trends in relative fitness observed across environments in these lines (Supplemental Table S1). Nonetheless, we believe these results are a valuable step toward integrating metabolomics experiments into evolutionary analyses. The cost of metabolomics experiments still limits the number of biological replicates and developmental time points that can be measured for a large number of samples. Thus applying these untargeted metabolomics analyses to all MA lines or to large genome-wide studies is a large investment. As prices continue to decline, this methodology is poised to provide hundreds of physiological measures that can link questions of fitness effects to the abundance of available sequencing data.

Even with hundreds of metabolites measured in this study and many mutations in well-known pathways in the well-annotated *A. thaliana* genome (Lamesch et al. 2012) we failed to infer any direct connections between mutations either in the coding region or in the intergenic region flanking genes that code for enzymes with an effect on metabolism (full list of AT TAIR numbers and functional annotation available in Supplementary Table S2). There could be several reasons for this pattern. For example, epimutations measured in mutation accumulation lines contribute substantial variation and their effects can outweigh mutational variance (Becker et al. 2011; Daves et al. 2016). Further studies could evaluate the contribution of epimutations to metabolomic variance. Additionally, many of the mutations in the MA lines in this study are intergenic and may have direct or indirect effects on the expression on distal genes. That the small number of mutations failed to disrupt any known pathways or pathways measured in this study in a predictable way suggests that our annotations of gene function are woefully incomplete.

This study measures the metabolic effects in leaf tissue at one time point on only six of the 107 *A. thaliana* MA lines, so the conclusions should be viewed in that light. We also chose lines that showed consistent relative fitness in field trials (Rutter et al. 2018); however, most of the lines had strong genotype-by-environment effects. Additionally, studies on the effect of new, spontaneous mutations on transcriptional expression have shed light on some of the biases of the mutational spectrum (Zalts and Yanai 2017; Huang et al. 2016), but lack fitness measures to connect genotypic changes to fitness effects. Future work that includes transcriptional studies combining metabolic effects with sampling of additional tissues and an analysis of more lines under many conditions will further elucidate the effect of spontaneous mutations on gene expression, fitness, and physiology. This more comprehensive metabolomics data, we could begin to understand whether there are higher levels

of polymorphism in regions with lower disruption of the metabolome or if mutations under positive selection are overall less disruptive of cellular physiology.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10682-022-10210-8.

Acknowledgements We thank Detlef Weigel for thoughtful comments and for the generation of the sequencing data. This work was supported by the National Science Foundation (DEB 2017485 to C.F.O.-M, OIA EPSCoR 1920954, DEB 1257902 to C.B.F., DEB 0844820 to C.B.F., DEB 1258053 to M.T.R., and DEB 0845413 to M.T.R.); and the National Institute of Health (INBRE 2P20GM103443-19). A previous version of this paper was available on BioRXiv (Kreutzmann, Sydney, Elizabeth Pompa, Nhan Nguyen, Liya Tilahun, Matthew T. Rutter, Mao-Lun Weng, Charles B. Fenster, and Carrie F. Olson-Manning. 2021. "Pleiotropy Is Associated with Fitness Consequences of New Mutations in Mutation Accumulation Lines." bioRxiv. https://doi.org/10.1101/2021.06.28.450192.)

Author Contributions All authors contributed to the writing, discussion, and revisions of this manuscript. Author C.F.O.-M., M-L.W. and C.B.F. contributed to experimental design and took part in every contribution listed below. Authors S.K., E.P, N.D.N., and L.T. contributed to data collection and analysis. Author M.T.R. contributed to data analysis.

Funding The authors have not disclosed any funding.

Data Availability All data used in this study (including raw metabolic measures) is represented in the supplementary tables. The datasets will also be uploaded to the Dryad repository upon acceptance.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aguilar-Rodríguez J, Fares MA, Wagner A (2019) Chaperonin overproduction and metabolic erosion caused by mutation accumulation in *Escherichia coli*. FEMS Microbiol Lett. https://doi.org/10.1093/femsle/fnz121

Beadle GW, Tatum EL (1941) Genetic control of developmental reactions. Am Nat. https://doi.org/10.1086/280939

Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. Nature 480(7376):245–249

Benfey PN, Chua NH (1989) Regulated genes in transgenic plants. Science 244(4901):174–181

Brkljacic J, Grotewold E (2017) Combinatorial control of plant gene expression. Biochim Biophys Acta 1860(1):31–40

Chang S-M, Shaw RG (2003) The contribution of spontaneous mutation to variation in environmental response in Arabidopsis thaliana: responses to nutrients. Evol Int J Org Evol 57(5):984–994

Chu X-L, Zhang Q-G (2021) Consequences of mutation accumulation for growth performance are more likely to be resource-dependent at higher temperatures. BMC Ecol Evol 21(1):109

Danisman S (2016) TCP transcription factors at the interface between environmental challenges and the plant's growth responses. Front Plant Sci. https://doi.org/10.3389/fpls.2016.01930

Davies SK, Leroi A, Burt A, Bundy JG, Baer CF (2016) The mutational structure of metabolism in *Caenorhabditis* elegans. Evol Int J Org Evol 70(10):2239–2246

Denver DR, Feinberg S, Steding C, Durbin MD, Lynch M (2006) The relative roles of three DNA repair pathways in preventing *Caenorhabditis elegans* mutation accumulation. Genetics 174(1):57–65

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. Nat Rev Genet 8(8):610–618

Fisher RA (1930) The genetical theory of natural selection. Oxford University Press, London

Monroe GJ, Srikant T, Carbonell-Bejerano P, Exposito-Alonso M, Weng M-L, Rutter MT, Fenster CB, Weigel D (2020) Mutation bias shapes gene evolution in *Arabidopsis thaliana*. Cold Spring Harbor Lab. https://doi.org/10.1101/2020.06.17.156752

Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. Plant Physiol 130(2):639–648
Halligan DL, Keightley PD (2009) Spontaneous mutation accumulation studies in evolutionary genetics. Annu Rev Ecol Evol Syst 40(1):151–172
Huang W, Lyman RF, Lyman RA, Carbone MA, Harbison ST, Magwire MM, Mackay TF (2016) Correction: spontaneous mutations and the origin and maintenance of quantitative genetic variation. eLife. https://doi.org/10.7554/eLife.22300
Inaba S, Kurata R, Kobayashi M, Yamagishi Y, Mori I, Ogata Y, Fukao Y (2015) Identification of putative target genes of bZIP19, a transcription factor essential for Arabidopsis adaptation to Zn deficiency in roots. Plant J Cell Mol Biol 84(2):323–334
Jan S, Abbas N, Ashraf M, Ahmad P (2019) Roles of potential plant hormones and transcription factors in controlling leaf senescence and drought tolerance. Protoplasma 256(2):313–329
Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB (2008) The evolution of gene regulation underlies a morphological difference between two drosophila sister species. Cell 132(5):783–793
Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30
Katju V, Bergthorsson U (2019) Old trade, new tricks: insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-throughput genomic approaches. Genome Biol Evol. https://doi.org/10.1093/gbe/evy252
Keightley PD, Lynch M (2003) Toward a realistic model of mutations affecting fitness. Evol Int J Org Evol 57:683–685
Koini MA, Alvey L, Allen T, Tilley CA, Harberd NP, Whitelam GC, Franklin KA (2009) High temperature-mediated adaptations in plant architecture require the bHLH transcription factor PIF4. Curr Biol CB 19(5):408–413
Korkuc P, Schippers JHM, Walther D (2014) Characterization and identification of Cis-regulatory elements in *Arabidopsis* based on single-nucleotide polymorphism information. Plant Physiol 164(1):181–200
Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R et al (2012) The Arabidopsis information resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40(Database issue):D1202–D1210. https://doi.org/10.1093/nar/gkr1090
Leiby N, Marx CJ (2014) Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. PLoS Biol 12(2):e1001789
Leng P, Zhao J (2020) Transcription factors as molecular switches to regulate drought adaptation in maize. Theor Appl Genet 133(5):1455–1465. https://doi.org/10.1007/s00122-019-03494-y
Lin R-C, Rausher MD (2021) R2R3-MYB genes control petal pigmentation patterning in *Clarkia Gracilis* Ssp. Sonomensis (Onagraceae). New Phytol 229(2):1147–1162
MacKenzie JL, Saadé FE, Le QH, Bureau TE, Schoen DJ (2005) Genomic mutation in lines of *Arabidopsis thaliana* exposed to ultraviolet-B radiation. Genetics 171(2):715–723
Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD (2021) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res 49(D1):D394-403
Monroe JG, Powell T, Price N, Mullen JL, Howard A, Evans K, Lovell JT, McKay JK (2018) Drought adaptation in *Arabidopsis thaliana* by extensive genetic loss-of-function. eLife. https://doi.org/10.7554/eLife.41038
Nguyen T-M, Shafi A, Nguyen T, Draghici S (2019) Identifying significantly impacted pathways: a comprehensive review and assessment. Genome Biol 20(1):203
Nguyen DT, Baojun Wu, Xiao S, Hao W (2020) Evolution of a record-setting AT-rich genome: indel mutation, recombination, and substitution bias. Genome Biol Evol. https://doi.org/10.1093/gbe/evaa202
Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R (1999) Molecular analysis of the anthocyanin2 gene of *Petunia* and its role in the evolution of flower color. Plant Cell. https://doi.org/10.2307/3870973
Roles AJ, Rutter MT, Dworkin I, Fenster CB, Conner JK (2016) Field measurements of genotype by environment interaction for fitness caused by spontaneous mutations in *Arabidopsis thaliana*. Evol Int J Org Evol 70(5):1039–1050
Rutter MT, Shaw FH, Fenster CB (2010) Spontaneous mutation parameters for *Arabidopsis thaliana* measured in the wild. Evol Int J Org Evol 64(6):1825–1835
Rutter MT, Roles A, Conner JK, Shaw RG, Shaw FH, Schneeberger K, Ossowski S, Weigel D, Fenster CB (2012) Fitness of Arabidopsis Thaliana Mutation accumulation lines whose spontaneous mutations are known. Evol Int J Org Evol 66(7):2335–2339
Rutter MT, Roles AJ, Fenster CB (2018) Quantifying natural seasonal variation in mutation parameters with mutation accumulation lines. Ecol Evol 8(11):5575–5585

Schrider DR, Houle D, Lynch M, Hahn MW (2013) Rates and genomic consequences of spontaneous mutational events in *Drosophila Melanogaster*. Genetics 194(4):937–954
Schultz ST, Lynch M, Willis JH (1999) Spontaneous deleterious mutation in *Arabidopsis thaliana*. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.96.20.11393
Shaw RG, Byers DL, Darmo E (2000) Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. Genetics 155(1):369–378
Shaw FH, Geyer CJ, Shaw RG (2002) A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. Evol Int J Org Evol 56(3):453–463
Stearns FW, Fenster CB (2016) The effect of induced mutations on quantitative traits in Arabidopsis Thaliana: natural versus artificial conditions. Ecol Evol 6(23):8366–8374
Stower H (2013) Adaptation by loss of function. Nat Rev Genet. https://doi.org/10.1038/nrg3557
Streisfeld MA, Liu D, Rausher MD (2011) Predictable patterns of constraint among anthocyanin-regulating transcription factors in ipomoea. New Phytol 191(1):264–274
Tebani A, Afonso C, Bekri S (2018) Advances in metabolome information retrieval: turning chemistry into biology. Part I: analytical chemistry of the metabolome. J Inherit Metab Dis 41(3):379–391
Van Buskirk HA, Thomashow MF (2006) Arabidopsis transcription factors regulating cold acclimation. Physiol Plantarum 126(1):72–80
Wagner GP, Lynch VJ (2008) The gene regulatory logic of transcription factor evolution. Trends Ecol Evol 23(7):377–385
Weng M-L, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG, Weigel D, Fenster CB (2019) Fine-grained analysis of spontaneous mutation spectrum and frequency in *Arabidopsis thaliana*. Genetics 211(2):703–714
Weng M-L, Ågren J, Imbert E, Nottebrock H, Rutter MT, Fenster CB (2021) Fitness effects of mutation in natural populations of *Arabidopsis thaliana* reveal a complex influence of local adaptation. Evol Int J Org Evol 75(2):330–348
Wessinger CA, Rausher MD (2014) Predictability and irreversibility of genetic changes associated with flower color evolution in *Penstemon Barbatus* genetics of flower color in *Penstemon Barbatus*. Evolution 68(4):1058–1070. https://doi.org/10.1111/evo.12340
Whittall JB, Voelckel C, Kliebenstein DJ, Hodges SA (2006) Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in aquilegia. Mol Ecol 15(14):4645–4657
Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 20(9):1377–1419
Xia J, Wishart DS (2011) Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. Curr Protocols Bioinform 34:14–10
Xu Y-C, Niu X-M, Li X-X, He W, Chen J-F, Zou Y-P, Qiong Wu, Zhang YE, Busch W, Guo Y-L (2019) Adaptation and phenotypic diversification in Arabidopsis through loss-of-function mutations in protein-coding genes. Plant Cell 31(5):1012–1025
Zalts H, Yanai I (2017) Developmental constraints shape the evolution of the nematode mid-developmental transition. Nat Ecol Evol 1(5):113