bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Genome-wide DNA mutations in *Arabidopsis* plants after multigenerational exposure to high temperature

Zhaogeng Lu[1, 2], Jiawen Cui[1], Li Wang[1], Nianjun Teng[3], Shoudong Zhang[4], Hon-Ming Lam[4], Yingfang Zhu[5], Siwei Xiao[7], Wensi Ke[7], Jinxing Lin[6], Chenwu Xu[2]*, Biao Jin[1]*

[1] College of Horticulture and Plant Protection, Yangzhou University, Yangzhou, China

[2] Agricultural College, Yangzhou University, Yangzhou, China

[3] College of Horticulture, Nanjing Agricultural University, Nanjing, China

[4] Center for Soybean Research of the State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong Special Administrative Region, China

[5] Institute of Plant Stress Biology, State Key Laboratory of Cotton Biology, Department of Biology, Henan University, Kaifeng, China

[6] College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China

[7] Wuhan Frasergen Bioinformatics Co, Wuhan, China

***Correspondence**:

Biao Jin, e-mail: bjin@yzu.edu.cn

Chenwu Xu, e-mail: cwxu@yzu.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Email:

Zhaogeng Lu: d160068@yzu.edu.cn

Jiawen Cui: DX120180104@yzu.edu.cn

Li Wang: liwang@yzu.edu.cn

Nianjun Teng: njteng@njau.edu.cn

Shoudong Zhang: shoudongzhang@cuhk.edu.hk

Hon-Ming Lam: honming@cuhk.edu.hk

Yingfang Zhu: zhuyf@henu.edu.cn

Siwei Xiao: xiaosiwei@frasergen.com

Wensi Ke: kewensi@frasergen.com

Jinxing Lin: linjx@ibcas.ac.cn

Chenwu Xu: cwxu@yzu.edu.cn

Biao Jin: bjin@yzu.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Abstract

Background: Elevated temperatures can cause physiological, biochemical, and molecular responses in plants that can greatly affect their growth and development. Mutations are the most fundamental force driving biological evolution. However, how long-term elevations in temperature influence the accumulation of mutations in plants remains unknown.

Results: Here we report that multigenerational exposure of *Arabidopsis* to extreme heat and moderate warming resulted in significantly increased mutation rates in single-nucleotide variants (SNVs) and small indels. We observed distinctive mutational spectra under extreme and moderately elevated temperatures, with significant increases in transition (C:G→T:A) and transversion (A:T→T:A) frequencies. Mutation occurred more frequently in intergenic regions, coding regions (especially nonsynonymous mutations), and transposable elements (TEs). At elevated temperatures, more mutations accumulated in genes associated with defense responses, DNA repair, and signaling, including the transcriptional response–related genes *HSP70* and *HSFA1A*. Methylation was observed more frequently at mutation sites, indicating that it contributed significantly to the mutation process at elevated temperatures. Moreover, the mutations in lines and populations grown under elevated temperatures were significantly biased toward low gene density regions, special trinucleotides (GC context), tandem repeats, and adjacent simple repeats. Additionally, 24% (n = 64) of SNVs and 43% (n = 40) of indels found in all mutation accumulation lines overlapped significantly with genetic variations reported in 1001 Genomes,

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

suggesting a non-uniform distribution of de novo mutations through the genome.

Conclusion: Collectively, our results suggest that elevated temperatures can accelerate the accumulation, and alter the molecular profiles, of DNA mutations in plants, thus providing significant insight into how environmental temperatures fuel plant evolution.

Keywords: *Arabidopsis thaliana*, heat, warming, mutation accumulation, mutation rate, mutation spectrum, mutation bias, molecular evolution, transposable element

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Background

Mutations are the ultimate source of genetic variations within and among all species, and act as the fundamental driving force of biological evolution [1-2]. DNA sequence mutations result from single-base substitutions, small insertions and deletions (indels), and large duplications/deletions (dels), leading to changes in the genomic sequence composition. An accurate assessment of the rates, types, and molecular spectra of these mutations within a species is critical to advancing understanding of diverse problems in evolutionary biology, such as species divergence time [3], population genetic diversity and effective population size [4-6], intra-organism mutation rate variation [7], and adaptation to specific environments [8-11]. However, mutation rates are difficult to estimate because mutations are often subject to purifying selection and drift in natural populations [2, 12]. The prevailing mutation accumulation (MA) experiments involve establishing multiple independent lines from a single progenitor and repeatedly propagating these lines for many generations, such that the effect of selection can be greatly reduced [13]. Given the effectiveness of this experimental system, MA experiments combined with whole-genome sequencing technology have been extensively conducted to investigate spontaneous mutation rates and spectra in multiple eukaryotic species, including *Chlamydomonas reinhardtii* [14, 15], *Drosophila melanogaster* [16-18], *Caenorhabditis elegans* [19, 20], *Escherichia coli* [1, 21], *Saccharomyces cerevisiae* [22], and *Arabidopsis thaliana* [23]. These MA studies have provided important reference data for estimating overall spontaneous mutation rates and the patterns underlying the molecular evolution of various species.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

96 In nature, however, organisms often encounter various environmental stresses,
97 and the effect of these stresses (such as elevated temperatures) on mutations has
98 begun to receive increasing attention. In *E. coli*, accumulated genetic changes under
99 high temperature stress were characterized in a long-term evolution experiment with
100 replicate populations [8]. A recent MA study on yeast grown under high temperatures
101 also revealed accumulated mutations associated with thermotolerance [24]. However,
102 these studies focused on unicellular organisms; no study has examined accumulated
103 mutations in multicellular organisms under high temperature conditions through MA
104 experiments.

105 Global warming is accelerating [25], resulting in frequent occurrences of
106 extremely high temperatures and prolonged warming. Such changes have profound
107 impacts on plant growth and development, productivity, and adaptation [26-28]. As
108 sessile organisms, plants have evolved sophisticated mechanisms to cope with the
109 elevated temperatures. Over the last decade, mechanisms of heat stress response and
110 the regulation thereof, as well as heat memory and priming in plants, have been well
111 elucidated [29-32]. Moreover, phenological and life-history traits underlying adaptive
112 evolution in response to climate warming have also been explored in plant
113 populations [27, 33]. However, most studies have focused on the short-term (within a
114 generation) responses to heat stress and phenological changes due to warming; little is
115 known about the longer-term evolutionary genetic consequences of exposure of plants
116 to high temperatures and mild warming over successive generations.

117 *A. thaliana*, as a model plant, is an ideal organism for investigating MA at the

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

genomic level. Previous MA studies have characterized DNA sequence mutations accumulating in *A. thaliana* lineages grown in non-stress [23] and high-salinity soil stress [34] environments. Although elevated temperatures significantly affect plant growth and development, how elevated temperatures induce MA in plants is still unknown. To address this question, we combined long-term MA experiments on lines and populations with deep whole-genome sequencing to determine the genome-wide mutation rates and profiles of *A. thaliana* grown under extreme heat, moderate warming, and control conditions for 10–22 successive generations. We found significantly increased mutation rates, along with distinct mutation spectra, mutated genes, and mutation bias, under elevated temperature conditions (moderate and extreme), providing insights into plant molecular evolution under environmental warming.

Results

MA experiments and whole-genome sequencing

We conducted long-term MA experiments on *A. thaliana* in both single-seed descent lineages and populations grown under Control (day 23℃ / night 18℃), Heat (day 32℃ / night 27℃), and Warming (day 28℃ / night 23℃) conditions (Fig. 1a, b) (see "Methods"). The elevated temperature treatments, especially Heat (32℃), resulted in various stress symptoms such as significantly decreased leaf size, shorter siliques (Fig. 1c, d), and shorter generation times. We sequenced 35 *A. thaliana* genomes, including 15 plants from MA lines at generation 10 (G10; five plants from each treatment) and

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

15 plants from MA populations [five plants each from G16 (Control, A16), G19 (Warming, C19), and G22 (Heat, B22)], spanning 10–22 successive generations, as well as their ancestor genomes (five individual plants from G0). In total, approximately 165 Gb of clean reads (30 libraries) from 30 genomes of progenies, and 25 Gb of clean reads from five libraries (see "Methods") representing the genetic background of the ancestor (Additional file 1: Table S1) were obtained. For all MA lines and populations, an average of 99.68% of sequenced reads was mapped to the *A. thaliana* reference genome, with average depths of 52.5×, 49.7×, 47.4×, 42.7×, 37.4×, and 36.3× per individual in D10 (Control), E10 (Heat), F10 (Warming), A16 (Control), B22 (Heat), and C19 (Warming), respectively. Accordingly, an average of 116 Mb (96.9%) of the reference genome was accessible for variant calling (Additional file 1: Table S1). To obtain sufficient coverage of the genetic background of the ancestor, the five G0 libraries (average coverage 37.1×) were combined. This sequencing depth/coverage and number of accessible reference sites allowed for precise detection of mutations at the whole-genome level.

Accumulated mutations and mutation rates in MA lines and populations under elevated temperatures

We obtained a total of 211 homozygous de novo mutations from MA lines under three temperature treatments (Fig. 2a and Additional file 1: Tables S2-S3), including 39 mutations (31 single-nucleotide variants [SNVs] and 8 indels) in D10 (Control), 98 mutations (69 SNVs and 29 indels) in E10 (Heat), and 74 mutations (54 SNVs and 20 indels) in F10 (Warming). Most (85.9%) of the 57 indels in the MA lines were short

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

(1–3-bp) deletions (dels) and insertions (ins) (Fig. 2c). Furthermore, the indels of E10 (25 dels vs. 4 ins) and F10 (17 dels vs. 3 ins) showed strong biases toward dels. In addition, we also detected 376 homozygous de novo mutations in MA populations, including 70 mutations (60 SNVs and 10 indels) in A16 (Control), 183 mutations (130 SNVs and 53 indels) in B22 (Heat), and 123 mutations (88 SNVs and 35 indels) in C19 (Warming) (Fig. 2b and Additional file 1: Tables S2-S3). Similar to the indels identified from MA lines, most indels in the MA populations were short (1–3-bp) and biased toward dels in B22 (42 dels vs. 11 ins) and C19 (22 dels vs. 13 ins) (Fig. 2d). Moreover, we found no novel transposable element (TE) insertion event in any MA line or population.

We further estimated the accuracy of the mutation calling pipelines using two simulation tests [14, 35]. For the first test, we simulated 600 random SNVs using six copies of reference genomes (see Methods). After read mapping and SNV filtering against the mutated reference genomes, our pipeline recovered 588 (98%) of 600 expected SNVs (Additional file 1: Table S4). For the second simulation test, we introduced homozygous SNVs and performed heterozygous SNV filtering, resulting in the recovery of 71–91% homozygous SNVs (Additional file 1: Table S5). To confirm our mutation calls, we experimentally examined all SNVs and indels from MA lines by Sanger sequencing. In total, 205 of 211 mutations were confirmed (6 mutations were identified as PCR failures) (Additional file 1: Table S6).

We estimated the SNV mutation rate (μ_{SNV}) and indel mutation rate (μ_{indel}) per site per generation in the MA lines. Mutigenerational growth of *A. thaliana* under

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Heat conditions caused significant increases relative to Control D in the average rates of SNVs ($\mu_{\text{E-SNV}} = 1.17 \times 10^{-8}$ vs. $\mu_{\text{D-SNV}} = 5.28 \times 10^{-9}$; two-sample *t* test, $p = 1.0 \times 10^{-3}$) and indels ($\mu_{\text{E-indel}} = 4.94 \times 10^{-9}$ vs. $\mu_{\text{D-indel}} = 1.36 \times 10^{-9}$, $p = 6.3 \times 10^{-4}$; Fig. 3a). Similarly, warming also increased the SNV ($\mu_{\text{F-SNV}} = 9.21 \times 10^{-9}$, $p = 9.9 \times 10^{-3}$) and indel ($\mu_{\text{F-indel}} = 3.41 \times 10^{-9}$, $p = 0.04$) mutation rates. Furthermore, the mutation rates of dels were more than 5-fold higher than those of ins in both Heat E ($\mu_{\text{E-del}} = 4.26 \times 10^{-9}$ vs. $\mu_{\text{E-ins}} = 6.82 \times 10^{-10}$) and Warming F ($\mu_{\text{F-del}} = 2.89 \times 10^{-9}$ vs. $\mu_{\text{F-ins}} = 5.12 \times 10^{-10}$), in contrast to the lack of difference observed in Control D. The overall MA mutation rates (SNVs and indels) of the Heat and Warming lines were 1.67×10^{-8} ($\mu_{\text{E-total}}$) and 1.26×10^{-8} ($\mu_{\text{F-total}}$) per site per generation, approximately 2.5-fold ($p = 8.6 \times 10^{-6}$) and 1.9-fold ($p = 4 \times 10^{-3}$) higher than the Control ($\mu_{\text{D-total}} = 6.65 \times 10^{-9}$), respectively (Fig. 3a). In addition, we observed significantly higher rates of total mutations and SNVs in Heat E compared to Warming F ($p < 0.05$), whereas the difference in indel rates was not significant ($p = 0.1$).

In parallel, we also estimated the average mutation rates in MA populations (Fig. 3b). The plants grown under Heat conditions had significantly higher SNV rates than Control plants, such as $\mu_{\text{B-SNV}}$ (Heat) $= 1.03 \times 10^{-8}$ vs. $\mu_{\text{A-SNV}}$ (Control) $= 6.53 \times 10^{-9}$ ($p = 1.6 \times 10^{-4}$), whereas no significant difference was observed between Warming ($\mu_{\text{C-SNV}} = 8.08 \times 10^{-9}$) and Control conditions ($p = 0.2$). However, the total mutation rates of Heat B and Warming C were 1.45×10^{-8} ($\mu_{\text{B-total}}$) and 1.13×10^{-9} ($\mu_{\text{C-total}}$), nearly 2.0- and 1.5-fold ($p < 0.05$) higher than in Control A ($\mu_{\text{A-total}} = 7.61 \times 10^{-9}$), respectively (Fig. 3b). Additionally, compared to Warming C, Heat B had significantly

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

higher total mutation and SNV rates ($p < 0.05$).

Molecular spectra of mutations in plants under elevated temperatures

Base substitution mutation spectra varied after multigenerational growth of *A. thaliana* under Heat, Warming, and Control conditions. We found a strong C:G → T:A bias (driven by C→T and G→A) in six mutational spectra that commonly occurred in MA lines under all three temperature treatments (Fig. 3c); however, C:G → T:A mutations under elevated temperatures (Heat and Warming) had much higher rates compared to Control. Furthermore, compared to Heat E ($\mu_{\text{E-C:G→T:A}} = 1.47 \times 10^{-8}$ per site per generation), Warming F exhibited a lower C:G → T:A mutation rate ($\mu_{\text{F-C:G→T:A}} = 1.23 \times 10^{-8}$). In addition, in Heat E and Warming F, the second most frequent substitution was A:T → T:A (mutation rate, $\mu_{\text{E-A:T→T:A}} = 3.20 \times 10^{-9}$); however, this differed from Control D, in which the second most frequent substitution was A:T → G:C ($\mu_{\text{E-A:T→G:C}} = 2.39 \times 10^{-9}$). In general, the mean rate of mutations occurring at C:G sites was nearly 3-fold higher than at A:T sites in Heat E and Warming F (Fig. 3c), in contrast to ~2-fold in Control D. In MA populations, we observed similar results under Heat and Warming (Fig. 3d); for example, the most frequent substitutions in Heat B and Warming C were also biased toward C:G → T:A ($\mu_{\text{B-C:G→T:A}} = 1.50 \times 10^{-8}$; $\mu_{\text{C-C:G→T:A}} = 1.54 \times 10^{-8}$) and were higher than in Control A ($\mu_{\text{A-C:G→T:A}} = 1.01 \times 10^{-9}$). The second most frequent substitutions (mutation rate) in Heat B occurred at A:T → T:A ($\mu_{\text{B-A:T→T:A}} = 2.37 \times 10^{-9}$) sites, similar to Heat MA lines. By contrast, the second most frequent substitutions in Warming C were A:T → G:C ($\mu_{\text{C-A:T→G:C}} = 2.16 \times 10^{-9}$), somewhat different from Warming MA lines.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

We calculated the transition and transversion frequencies (per genome per generation) for the three treatments. In MA lines, the transition (Ts) and transversion (Tv) frequencies in Heat E (0.84 and 0.54, respectively) and Warming F (0.68 and 0.40, respectively) showed obvious increases compared to Control D (0.46 and 0.16, respectively; Fig. 3e-f), resulting in significantly decreased Ts/Tv ratios at both elevated temperatures (Fig. 3g and Additional file 1: Table S7). Moreover, compared to Heat E, Warming F had a higher Ts/Tv ratio, which can be attributed to its lower frequencies of Ts and Tv. In MA populations, significantly decreased Ts/Tv ratios were found in Heat B (1.82) compared to Control A (2.53; Fig. 3g). Nevertheless, the Warming population showed a high Ts/Tv ratio due to its higher transition and lower transversion rates relative to those of the Heat and Control populations.

Mutation frequency distribution across different genomic regions in plants under elevated temperatures

We annotated the mutations and estimated their frequencies across different genomic regions in MA lines and populations. All MA lines showed higher mutation frequencies in intergenic regions than in genic regions. Heat E and Warming F showed >50% increases in mutation frequencies in both genic (1.4–2.2-fold increases) and intergenic (0.5–1-fold increases) regions compared to Control D ($p < 0.05$; Fig. 4a). Notably, within genic regions of Heat E and Warming F, higher mutation frequencies occurred in coding regions than in noncoding regions, different from Control D. The predominance of variants in coding regions of Heat E and Warming F was attributed to the disproportionate occurrence of nonsynonymous mutations

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

(Additional file 1: Table S3). For example, the nonsynonymous mutations in Heat E (0.26) were significantly more frequent than in Control D (0.02; $p < 0.05$). In addition, Heat E showed higher mutation frequencies in intergenic and genic regions than Warming F, but this difference was not significant ($p > 0.05$). In noncoding regions, the frequencies of intronic and untranslated region (UTR) mutations were highest in Heat E. Interestingly, more mutations occurred in transposable elements (TEs) under the Heat treatment, with a significantly increased frequency in Heat E compared to Control D ($p = 0.02$). Finally, we calculated SNV rates within intergenic, genic, and TE regions (Fig. 4a), all of which increased with temperature.

In MA populations, we found that the mutation frequencies of intergenic regions and TEs in Heat B were significantly higher than those in Control A ($p < 0.01$; Fig. 4b). By contrast, the frequency of nonsynonymous mutations in Heat B was lower than that in Warming C, probably due to a stronger selection pressure in populations under the Heat treatment. Consistently, the SNV rates of genic regions in Heat B were lower than those in Warming C (Fig. 4b).

Nonsynonymous SNVs, gains and losses of stop codons, and indels within coding regions are likely to affect fitness [36]. Therefore, we estimated the rates of diploid genomic mutations affecting fitness under different treatments. These rates were 0.48 (± 0.1) and 0.36 (± 0.2) per generation in Heat E and Warming F, respectively, and these values were higher than those in Control D (0.13 ± 0.1; Heat *vs*. Control, $p = 0.005$; Warming *vs*. Control, $p = 0.16$). In the comparison among lines, lower mutation rates affecting fitness were found in MA populations. Genomic

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

mutation rates affecting fitness were significantly higher in Heat B (0.16) and Warming C (0.24) than in Control A (0.05; $p < 0.003$).

Mutations in functional genes of *A. thaliana* under elevated temperatures

To investigate the accumulated mutations in functional genes that may be involved in various biological processes underlying high temperature responses, we performed Gene Ontology (GO) functional analysis of 29, 46, and 55 genes with mutations from the Control, Warming, and Heat MA lines and populations, respectively (Additional file 1: Table S3). We found that these genes were enriched in multiple related terms, including the cellular process, metabolite process, cell part, membrane, binding, and catalytic activity (Fig. 5a). In contrast, elevated temperatures resulted in the enrichment of more genes associated with the "response to stimulus", "reproductive process", "development process" and "biological regulation" terms. Kyoto Encyclopedia of Genes and Genomes (KEGG) functional analysis showed enrichment of common pathways, including "signaling transduction", "development" and "replication and repair" at elevated temperatures (Additional file 2: Fig. S1).

To further determine whether these mutations occurred at genes involved in the transcriptional response to heat and warming, we used a previously obtained RNA-seq dataset to identify potential temperature-responsive (significantly differentially expressed) transcripts among the Heat, Warming, and Control treatments (see "Methods"). Interestingly, 9 (16%) of 55 genes from Heat MA samples showed significantly differential expression between the Control and Heat treatments, and 10 (22%) of 46 genes from Warming MA samples were differentially expressed between

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Control and Warming (Fig. 5b). In particular, mutations occurred in two genes encoding heat shock protein 70-17 (*HSP70-17*) and heat stress transcription factor A-1a (*HSFA1A*), which were upregulated under Heat treatment and were identified in Heat E and B, respectively.

We further focused on genes with nonsynonymous, frameshift, stop-gain SNVs or indels (Fig. 5c). In Heat lines, in addition to *HSP70-17*, described above, a nonsynonymous mutation was found in a gene encoding fumarate hydratase 2 (*FUM2*), which is associated with respiratory metabolism. Interestingly, a mutation occurred in the gene encoding a proliferating cell nuclear antigen (PCNA) domain-containing protein (*AT4G17760*), which is associated with DNA repair. Moreover, a frameshift del and a nonsynonymous SNV in the defense-related (i.e., disease resistance) protein Toll interleukin 1 receptor-like nucleotide binding leucine-rich repeat (TIR-NB-LRR; *AT5G48770* and *AT4G10780*) were identified in the Heat E and warming F lines, respectively. In contrast to the accumulated mutations in MA lines, we found many exonic mutations distributed in genes associated with development and signal transduction, such as those encoding the calcium-dependent lipid-binding family protein (*AT1G48090*) and WD40 repeat-like superfamily protein (*AT3G54190*), in MA populations (Fig. 5c). Given the experimental design applied to MA populations, we speculate that these common exonic mutations probably originated from a parental individual in the same generation (not the ancestor plant), suggesting that some genetic variants are more likely to spread in populations under selective pressure over multiple generations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

315 **Interaction between methylation and TE annotation**

316 We conducted whole-genome bisulfite sequencing of MA lines and identified more
317 methylated cytosines (mCs) in Heat E (10.54%) and Warming F (10.44%) than in
318 Control D (9.78%; Additional file 1: Table S8). mCs in CG, CHG, and CHH (where H
319 refers to A, T, or G) contexts are summarized in Supplemental Table 8. Spontaneous
320 deamination of methylated cytosine (mC) to thymine is known to be a major source of
321 mutations, resulting in elevated mutation rates at methylated sites [37]. We thus
322 focused on mutations in the Control, Heat, and Warming MA lines in three contexts
323 methylated and nonmethylated contexts. In the Heat treatment, the proportions of
324 methylation at mutated bases were much greater than the genome-wide occurrence of
325 methylation in the CG (Fisher's exact test, p = 4.58e-8), CHG (Fisher's exact test, p =
326 1.92e-21), and CHH (Fisher's exact test, p = 1.63e-3) contexts (Fig. 6b). High
327 frequencies of methylation at mutation sites were also found in the Warming (Fisher's
328 exact test: CG, p = 3.36e-4; CHG, p =1.92E-15; CHH, p = 0.02) and Control (Fisher's
329 exact test: CG, p = 3.56e-12; CHG, p = 2.27e-21; CHH, p = 0.08) treatments (Fig. 6a,
330 c). Because methylation and TEs correlate significantly [35, 38], we further tested the
331 main effects of methylation and TE position (two-way analysis) on mutation rates
332 under elevated temperatures using a logistic regression model. The methylated sites
333 and TE regions were associated positively with mutations in the Control, Heat, and
334 Warming MA lines (Additional file 1: Table S9). In general, methylated sites within
335 and outside TEs had higher mutation rates than did non-methylated sites in MA lines
336 (Additional file 2: Fig. S2). Compared with those in Control lines, methylated and

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

337 nonmethylated sites in the Heat and Warming lines showed higher mutation rates
338 regardless of location (within or outside TEs); Heat E had the highest mutation rate on
339 methylated sites outside TEs (Fig. 6d). In addition, we observed that nonmethylated
340 sites within TEs had a higher rate in Warming F than in Heat E, but this difference
341 was not significant.

342 **Mutational bias and context effects of *A. thaliana* under elevated temperatures**

343 We estimated the relative contributions of various genomic properties to mutation
344 frequency, including gene density and GC content. In both MA lines and populations,
345 we found that fewer mutations in high versus low gene density regions in all three
346 treatments (Fig. 7a). For MA lines, the mutation rate was significantly biased toward
347 low versus high gene density region in Heat E (*t* test, $p = 0.009$; Fig. 7b). By contrast,
348 mutation propensity was similar between Warming F and Control D, and no
349 significant difference in mutation rate was observed between the high and low gene
350 density regions ($p = 0.35$ in Control, $p = 0.089$ in Warming). This suggests that plants
351 exposed to extremely high temperatures (Heat) over multiple generations have a
352 stronger ability to induce DNA sequence mutations toward low gene density regions
353 than plants under moderate warming or ambient temperatures. Moreover, we also
354 estimated whether GC content (per 1-kb window) affected local mutation rates in the
355 MA lines and populations of each treatment. For all MA lines and populations, the GC
356 contents and observed mutation rates were not well correlated (Additional file 2: Fig.
357 S3), suggesting that GC content did not affect the mutation rate in our MA
358 experiments.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

We evaluated the effect of local sequence context on the mutation rates of A/T and G/C positions flanked by different nucleotides at either site, regardless of DNA strand orientation (for example, AAG and its complement CTT both contribute to the mutation rate in the central AT position under the category AAG). As expected, GC bases had significantly higher mutation rates than did AT bases in all treatments (*t* test, $p < 0.03$) except Control D ($p > 0.14$; Fig. 7c–e). In general, mutation rates of AT bases in all contexts were uniform for each MA experiment (*G* test, $p > 0.15$), but mutation rates of GC bases were not ($p < 0.01$), except in the Warming treatments (Warming F, $p = 0.98$; Warming C, $p = 0.44$). Moreover, the nucleotides located one position upstream or downstream had significant effects on the mutation rate in the Heat B population (*t* test, $p < 3.32e\text{-}4$; Fig. 7d), whereas those in other MA lines and populations did not ($p > 0.05$). Of all 16 possible combinations of flanking nucleotides, GCG in Control A (two-tailed *Z* test, $p = 5.56e\text{-}13$), GCG in Heat B ($p = 7.40e\text{-}14$), and CCG in Warming C ($p = 3.97e\text{-}6$) had significantly higher mutation rates than did other GC contexts. In contrast to MA populations, the Control D, Heat E, and Warming F lines showed significantly higher mutation rates in the CCC ($p = 3.03e\text{-}4$), CCG ($p = 0.02$), and GCT ($p = 6.24e\text{-}5$) contexts than in other GC contexts (Fig. 7g-h). However, the trinucleotides CCG (or GGC) and GCG (or CGC) appeared to have strong mutation abilities in all MA groups, regardless of temperature treatment. In addition, we observed that almost all indels within the Heat (E10 and B22), Warming (F10 and C19), and Control groups (D10 and A16) either occurred near simple repeats or involved tandem-repeat dels and ins (Additional file 1: Table

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

S10), suggesting that the occurrence of indels is strongly biased toward repeat sequences in *A. thaliana*.

Comparison of de novo mutations with natural genetic variations

The 1001 Genomes Consortium (2016) reported 10,707,430 single-nucleotide polymorphisms (SNPs) and 1,424,879 indels (≤40 bp) in 1135 natural accessions of *A. thaliana*. To compare de novo mutations with natural variations, we merged the mutations from all MA lines and populations into 263 unique SNVs and 93 indels. We found that 64 (24%) of 263 total SNV sites coincide with biallelic SNPs in the 1001 Genomes dataset, and 50 (19% of the total) of these 64 shared SNVs are identical (Fig. 8). Among the 93 indels identified in MA experiments, 40 (43%) overlap with indels from the 1001 Genomes population, 12 (13% of the total) of which are identical. These identical sites (86% of shared SNVs, 75% of shared indels) are derived mainly from the Heat and Warming lines. Compared with the expected overlap (based on a random distribution of mutations and polymorphisms), the overlap between polymorphisms in all of our MA lines and populations with those of natural variants is highly significant (Fisher's exact test: SNV, p = 2e–24; indel, p = 1e–12; Fig. 8). To determine whether the SNVs identified in our MA results were biased toward conserved or substitution sites in *A. thaliana*, we compared them with the 219,909 ancestral variants (SNPs occurring at substitution sites in *A. thaliana*) and 1,799,125 derived variants (SNPs occurring at conserved sites) from the 1001 Genomes biallelic SNP dataset (see Methods). Among all SNVs identified in our MA results, only four SNVs (1% of SNVs from Warming C and F and Heat E) overlap with ancestral

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

403 variants and one SNV (from Heat B) is shared with derived variants, indicating a low
404 frequency of de novo SNVs (identified from elevated temperature) at the conserved
405 sites.

406 ## Discussion

407 MA experiments in combination with whole-genome sequencing have been used to
408 investigate spontaneous mutations in various model organisms with short generation
409 times [1, 2, 14, 15, 23, 39, 40]. In unicellular MA studies, the organisms generally
410 reproduce clonally (asexually), with each clonal population propagating for multiple
411 generations, allowing the accumulation of mutations in a nearly neutral manner
412 (minimal selection pressure) [1, 2]. By contrast, multicellular organisms usually
413 produce progeny by sexual reproduction; thus, large numbers of progeny inevitably
414 suffer from environmental selection pressure, resulting in accumulated mutations that
415 differ from those arising from clonal reproduction. To avoid such selection effects,
416 single seed descent lines have been employed to accurately reflect mutation profiles
417 under ideal conditions [23, 40]. In nature, however, plants generally occur in
418 populations rather than in groups descended from a single seed. Therefore, a
419 combination of MA populations and single seed descent MA lines could better reveal
420 how spontaneous or environmentally induced MA shapes genetic variation. In our
421 study, we employed both single-seed descent line and population MA experiments
422 with *A. thaliana* grown under ambient, extreme heat, and moderately warming
423 temperatures, and estimated their mutation profiles.

424 Environmental changes can promote genome-wide accumulation of mutations

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

425 and change the profiles of new mutations in organisms, in turn potentially
426 contributing to long-term evolutionary adaptation [8, 11, 24, 34, 41-44]. However, in
427 plants, only one environmental factor (high-salinity soil) and artificial irradiation have
428 been reported to induce MA [34, 45, 46]. In this study, we revealed that
429 multigenerational exposure of *A. thaliana* to high and moderate heat could
430 significantly increase the rates of accumulated de novo mutations in both MA lines
431 and populations, which showed overall mutation rates that were more than 1.4-fold
432 (1.49 for Warming C and 1.90 for Warming F) and 1.9-fold (1.91 for Heat B and 2.51
433 for Heat E) higher than Control groups. However, the rates of both spontaneous
434 ($5.28\text{-}6.53 \times 10^{-9}$) and temperature-driven mutations (1.18×10^{-8}) in *A. thaliana* are
435 lower than that of the long-lived perennial plants (*Populus trichocarpa*, 2.66×10^{-8};
436 Oak, $4.2\text{–}5.8 \times 10^{-8}$) [47, 48]. The majority of accumulated SNVs in trees were likely
437 to be result from long-term exposure to ultraviolet light and longer lifespan (with
438 higher number of mitoses), which is different from *A. thaliana* maintained in the
439 laboratory (lower ultraviolet light and short life cycle, 2~3 month). On the other hand,
440 the increased mutation rates under Heat and Warming temperatures were mainly
441 attributed to the accumulation of a large proportion of SNVs and small (≤3-bp)
442 deletions. This relatively high frequency of deletions at elevated temperatures
443 provides evidence that *A. thaliana* genomes are still shrinking [49]. Moreover, these
444 accumulated mutations in Heat (E, B) and Warming (C, F) treatments exhibited
445 distinct molecular spectra compared to Control, with significantly increased rates of
446 specific transversions (A:T → T:A) and transitions (C:G → T:A) under Heat

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

447 treatment in particular. These results differ from those obtained in a salt stress MA
448 study in which only the transversion rate was significantly increased [34], suggesting
449 a possible divergence in response mechanisms driven by heat and salt stresses in *A.*
450 *thaliana*. Nevertheless, the relatively low Ts/Tv ratios generated by Heat (1.56) and
451 Warming lines (1.70) were similar to those generated by salt stress and irradiation in *A.*
452 *thaliana* lineages [34, 45], further demonstrating that environment factors are
453 responsible for changes in plant molecular spectra.

454 Given that altered lifespans, increased reactive oxygen species (ROS) levels
455 (oxidative stress), and DNA repair impairments are induced by environmental stresses
456 (including heat stress) in plants [26, 50-53], the increased mutation rates in *A.*
457 *thaliana* grown under Heat and Warming temperatures may be explained by three
458 previously proposed major causes of environmental stresses: generation time,
459 metabolic rate, and DNA repair [26, 34, 54, 55].

460 Compared with MA lines, the slightly lower rates of total mutations in population
461 lineages probably resulted from our artificial selection procedure. In addition, a recent
462 MA study of *C. elegans* showed that larger population sizes (N = 10, 100) with
463 increasingly intense selection show lower spontaneous mutation rates compared to
464 small (N = 1) populations [40]. This result, together with our results regarding the
465 mutation rates of populations (N = 35) and lines (N = 1), may support the hypothesis
466 that larger population sizes show lower mutation rates under selection pressure.

467 When an organism is exposed to long-term environmental stress, the genomic
468 distribution of accumulated variants usually changes compared to that under

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

469 non-stress conditions, and such stress-induced mutations were increased in exonic
470 regions of yeast and *E. coli*, for the purposes of evolutionary adaptation [8, 24]. In the
471 present study, we observed significantly higher mutation frequencies and rates in
472 coding regions of *A. thaliana* MA lines exposed to Heat (E) and Warming (F)
473 temperatures compared to Control (D), especially nonsynonymous mutations; this
474 suggests that elevated temperatures could accelerate the accumulation of mutations in
475 coding regions. When the plants are exposed to high temperatures (especially in Heat
476 MA lines), the accelerated rate of mutations in protein-coding regions may increase
477 the load of deleterious mutations, thereby affecting fitness. Assuming that
478 nonsynonymous SNVs (including stop codon gain) and indels in coding regions affect
479 fitness, the significantly higher mutation rates affecting fitness in the Heat and
480 Warming treatments suggest that long-term exposure to high temperatures has a
481 stronger effect on fitness and may contribute to genetic variation affecting fitness. In
482 contrast to MA lines, the Heat MA population displayed lower mutation frequencies
483 in coding regions but higher frequencies in noncoding regions, indicating that
484 long-term heat stress coupled with strong selection constraints strongly restricted the
485 accumulation of potentially deleterious mutations in *A. thaliana*. The low mutation
486 rate affecting fitness estimated for the Heat MA population further suggests that the
487 negative effects on fitness in this population may be offset by selective pressures,
488 despite the increase in the genome-wide mutation rate. Overall, even in the Heat
489 treatment, the mutation rate of SNVs in coding regions remained lower than that in
490 other regions, in agreement with the predictive model of mutation rates, which

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

491 indicates that a significantly lower rate in coding regions shapes gene evolution [56].
492 In contrast, we observed the highest mutation rates in intergenic regions, coupled with
493 a significant bias toward low gene density, demonstrating that high temperature also
494 accelerates mutations in non-genic regions.

495 Transposons are mobile DNA elements that are prevalent in eukaryotic genomes,
496 and play important roles in regulating gene expression, accelerating sequence
497 mutations, genome reshaping, and adaptive variation [57-61]. Transposon activity can
498 be triggered by environmental stress, and some TEs can affect the expression of
499 nearby stress-responsive genes [62, 63]; this indicates that TEs can exhibit
500 stress-responsive transposition or movement. In our MA study, however, no novel TE
501 insertion was observed in any MA line or population subjected to high temperatures,
502 and we found higher mutation rates in TEs of the Heat and Warming groups than in
503 those of the Control, in accordance with the recently reported higher frequency of
504 SNVs in TEs [35] Given that a high mutation rate in TEs could result in their
505 transposable inactivation [64], we suggest that the growth of multiple generations at
506 high temperatures can accelerate genome-wide transposon mutations in *A. thaliana*,
507 consequently affecting stress-responsive transposition or movement of TEs.

508 Methylated cytosines contribute to changes in spontaneous deamination to
509 thymidine, thereby leading to elevated mutation rates in MA lines [23, 37]. Consistent
510 with these findings, the present study revealed that the frequency of methylation (in
511 all three contexts) at mutated bases is significantly elevated in plants exposed to high
512 temperatures. Further analysis confirmed that methylated sequences were associated

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

positively with mutations in the Heat and Warming MA lines. These results suggest that methylation contributes to the accumulation of mutations at high temperatures. In addition, the occurrence of DNA methylation was biased toward TE regions [35, 38]. Indeed, we observed higher mutation rates of methylated sites within TEs, suggesting a universal mechanism of methylated TE-biased mutations in *A. thaliana* regardless of environmental factors. However, heat treatment increased the mutation rate at methylated sites outside TEs, suggesting that methylation within and outside of TEs contributes to the high mutation rate observed after multigenerational exposure of *A. thaliana* to heat stress.

The datasets from 1001 Genomes provide abundant polymorphism resources for natural populations of *A. thaliana* [65]. The overlap between our identified de novo mutations and 1001 Genome biallelic variants was significantly greater than expected, indicating that these variants are not distributed uniformly in our MA results. Although high proportions of mutations were found in our MA lines and populations at high temperatures, very low frequencies of SNVs were observed at conserved sites in all MA lines. These results are consistent with the finding that the rate of spontaneous mutations is lower at conserved sites than at polymorphic sites [35].

Plants subjected to excess heat exhibit a characteristic set of cellular and metabolic responses, including dramatic changes in the stability of cytoskeleton structures, membranes and proteins, ROS production, and metabolic uncoupling (primary cellular affects), as well as in calcium signaling, phosphorylation, hormones, and transcriptional responses (secondary signal responses) [26, 66, 67]. A number of

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

genes involved in these responses are induced or transcriptionally activated by elevated temperatures [26, 31], such as heat shock proteins (HSPs), ROS-scavenging enzymes, and heat shock transcription factors. In this study, we observed that mutigenerational growth of *A. thaliana* at high temperatures led to greater accumulation of mutations in genes involved in responses to stimuli, reproduction, replication and repair, and development, relative to growth under control conditions. Furthermore, we observed that mutations accumulated in some genes that are transcriptionally responsive to heat and warming, including *HSP70-17* and *HSFA1A*. Additionally, some nonsynonymous and frameshift mutations of stress response-related genes (*AT5G48770* and *AT4G10780*) were identified in the Heat and Warming lines. These results suggest that mutigenerational growth of *A. thaliana* at high temperatures causes mutations in temperature response– or stress-associated genes. Given that mutations in some functional gene (*e.g.* loss-of-function) are important for plant adaptation [68, 69], thus whether these mutations of genes contribute to high-temperature adaptation of plants requires further investigation. We also found that one gene (*AT4G17760*) associated with DNA repair and DNA damage checkpoints was mutated in a Heat MA line. Under heat stress, ROS-induced DNA damage is usually repaired by both the nucleotide and base excision repair processes, during which DNA repair pathways are activated [52, 70]. These results suggest that plants alter their genes to re-adjust their own physiological and developmental processes in response to long-term high temperature stress.

Elevated temperatures, including both extreme heat and moderate warming, have

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

557 reportedly caused genetic variations in plant phenology and developmental timings
558 [26, 33, 71-73]. The genetic basis for thermal sensitivity was identified through the
559 exposure of *A. thaliana* to seasonal warming [71, 72]. In addition, the developmental
560 stage-specific quantitative trait loci in *A. thaliana* grown under short-term heat stress
561 have been revealed [74]. These investigations indicate that plants can evolve different
562 strategies to adapt to short-term warming and heat stress. However, how plants
563 respond to long-term moderate or extreme heat stress at the whole-genome level
564 remains unknown. Our results from both MA lines and populations revealed higher
565 mutation rates of accumulated SNVs and indels in *A. thaliana* exposed to
566 multigenerational heat stress (Heat E, B) compared to moderate warming (Warming F,
567 C). Moreover, almost all mutation types in both genic and intergenic regions showed
568 higher mutation frequencies under Heat compared to Warming temperatures. The
569 lower Ts/Tv ratio in Heat E (1.56) compared to Warming F (1.70) suggests that plants
570 exposed to extremely high temperatures might accumulate more deleterious variations
571 (e.g., more transversions in coding regions). Possible explanations for this include
572 greater DNA damage driven by increased amounts of free radicals under environment
573 stress [34, 54], and successive induction of DNA damage due to long-term heat stress
574 coupled with greater impairment in DNA repair mechanisms. In addition to these
575 differences, a common mutation bias toward a high rate of A:T → T:A (transversions)
576 was observed in the Heat and Warming groups, which suggests potential convergence
577 of the mutation mechanisms induced by the two elevated temperature treatments.

578 The effect of epigenetic variations (e.g., DNA methylation) on the probability of

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

579 mutation may be ubiquitous in plants, and more epigenetic modifications occur in
580 response to environmental stresses [23, 34, 75-77]. Thus, investigations into
581 epigenetic variations, and in particular epimutations across generations (non-transient
582 variations), could shed additional light on the genetic evolutionary mechanisms
583 activated in response to long-term high temperature stress.

584 ## Conclusion

585 Growth of *Arabidopsis* under elevated temperatures over multiple generations
586 significantly accelerated mutation accumulation and altered mutation profiles. The
587 increased mutation frequency occurred in intergenic regions, coding regions, and
588 transposable elements, as well as in nonsynonymous mutations in functional genes.
589 Our results reveal the fundamental mutation spectrum under environmental warming,
590 improving the understanding of genetic variations in plants under long-term heat
591 stress.

592

593

594

595

596

597

598

599

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Methods

Plant materials, experimental design, and growth conditions for the MA experiments

We used the wild-type Columbia (Col-0) strain of *A. thaliana* to conduct MA experiments; this strain has been maintained for several years in our laboratory [53]. A single ancestral Col-0 plant was planted in soil and grown in a growth chamber under a 16-h light/8-h dark photoperiod and 150 μmol m^{-2} s^{-1} of photosynthetically active radiation (PAR) provided by fluorescent tubes (Philips Electronics Trading & Services, Shanghai, China), at 23/18°C (light/dark) ambient temperature and 50/80% relative humidity (RH). In addition, the plant was alternately watered to saturation with 1/2 Murashige and Skoog (MS) solution or deionized water every week. The ancestral plant was self-pollinated, and its progeny seeds were harvested for our MA line and population experiments. The MA line and population experiments were performed under 23/18°C (light/dark), 28/23°C, and 32/27°C temperature regimes, and all detailed treatment methods are described in the Additional file 3: Supplementary detailed methods.

DNA extraction, sequencing, quality trimming and filtering, variant calling, and annotation

These processes are described in detail in the Additional file 3: Supplementary detailed methods.

Sanger sequencing validation

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

We employed Sanger sequencing to estimate the accuracy of our mutation calls. All identified mutation calls (211 mutations) including SNVs and indels from MA lines were selected for verification. Primers were designed around the mutations using Primer5 to amplify a 200–600-bp region (Additional file 1: Table S3). All but six mutations were verified through PCR as being present in MA lines but absent in the progenitor (background, G0), indicating a negligible false-positive detection rate; six mutations could not be verified through PCR amplification.

Mutation rate calculations

We calculated the mutation frequency (per genome per generation) and rate (per site per generation) using the formulas $m = n/g$ and $\mu = n/gb$ (or m/b), respectively, where n is the total number of mutations per line detected, g is the number of generations, and b is the number of bases (*A. thaliana* genome bases) analyzed. Detailed methods are illustrated in Additional file 3: Supplementary detailed methods. In addition, we also calculated the conditional mutation rates for all six substitution possibilities in each line and population using the same equations, but with m representing the number of substitutions of the focal substitution type, and n representing the total number of analyzed sites capable of producing a given mutation. For our analysis of mutation rates within chromosomes with different regions, we first divided each chromosome into multiple 100-kb bins, and then classified these bins into high and low gene density groups based on the gene distribution in each chromosome. Mutation rates in each group were measured by dividing the total number of SNVs and indels from our study (Control, Heat, and Warming) by the product of the total

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

644 number of sites in each group and the number of generations per line using the
645 following formula: $\mu = n/gb$. Although the actual mutation rate in each treatment
646 might be underestimated, this has little effect on the mutation rate comparisons (see
647 Additional file 3: Supplementary detailed methods).

648 **Comparison of genetic variations between MA and natural populations**

649 All VCF files of natural variations were downloaded from the 1001 Genome Project
650 website (http://1001genomes.org). SNP loci and allele frequencies were assessed.
651 Ancestral and derived variants were defined according to Weng et al. [35]. Briefly,
652 triple whole-genome alignment among *A. thaliana* (TAIR10), *A. lyrata*, and *Capsella*
653 *rubella* was performed, and identical sites within the alignment regions of all three
654 species were identified as conserved. Sites that were the same in *A. lyrata* and *C.*
655 *rubella*, but different in *A. thaliana* were considered to be ancestral. De novo variants
656 from MA lines or populations at conserved and ancestral sites were identified as
657 derived and ancestral mutations, respectively.

658 **DNA methylation and interaction analysis**

659 Whole-genome bisulfite sequencing (WGBS) was conducted using the same genomic
660 DNA extracts used for whole-genome sequencing. Genomic DNA was fragmented to
661 200–300 bp with a Covais sonicator, followed by end repair and adenylation. The
662 purified DNA fragments were bisulfite converted using the EZ DNA
663 Methylation-Gold™ Kit (Zymo Research). WGBS libraries were sequenced on the
664 Illumina HiSeq 2500 platform (Novogene, China). The raw paired reads were quality
665 controlled and trimmed, and then aligned to the reference genome (TAIR10) using

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Bismark software v0.16.1 [78] with the default parameters. Methylated cytosines (mCs) in each sequence were extracted from the aligned reads using the Bismark methylation extractor. The proportions of mCG, mCHG, and mCHH (where H refers to A, T, or G) contexts were calculated by dividing the number of mCs by the total number of Cs. Methylation-level analysis was performed using divided bins (bin size = 100 bp). The methylation level (ML) of each C was defined as reads (mC) / reads (mC) + reads (C). mCs with corrected p values ≤ 0.05 were used for subsequent interaction analysis. A logistic regression model was used to test the effects of C methylation and TE on the mutation probability of a given nucleotide. Analysis of interaction among cytosine methylation (mC or nonmethylated cytosine), TE (TE or non-TE), and chromosome region (arm or pericentromeric/centromeric site) was generally conducted according to the method of Weng et al. [35].

Local sequence context and GC content analyses

To investigate the effect of local sequence context on the mutation rate in each treatment, we extracted the flanking bases (1 bp upstream and downstream of the mutation site) at A/T and G/C positions surrounding the mutation sites. We calculated the rates of trinucleotide mutation in MA lines and populations under the Control, Heat, and Warming conditions. In addition, we estimated the relative contribution of the GC content to mutability through calculating the mutation rate under a given GC content, using a sliding window of 1,000 bp (bin size = 1,000 bp) and 0.01 intervals of GC content. The formula used is $\mu = n / (m \cdot g \cdot 1000)$, where μ represents the mutation rate under a given GC content, m represents the number of bins in a given

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

GC content, and *n* represents the total number of mutations within *m*.

Statistical analyses

All statistical analyses were performed in R Studio (version 1.0.143) using the Stats analysis package (R Development Core Team, Vienna, Austria) and SPSS software (version 23.0; IBM Corp., Armonk, NY, USA). *p* values < 0.05 and < 0.01 were indicative of significant and very significant results, respectively.

Declarations

Acknowledgements

We thank Jing Wang, Xiaoxue Jiang, Ming Fang, Kaige Luo, Kaibiao Ma, Zhichao Jia, Shixiong Ren that contributed to maintaining the mutigenerational growth of materials over the past ten years, and thank Frasergen and Novogene companies for assistance with genome sequencing and technical support. No conflict of interest declared.

Authors' contributions

B.J., C.X., L.J and N.T. conceived and designed the project. Z.L. and J.C. performed mutigenerational growth of materials and seed harvest, and prepared the samples for sequencing. Z.L., S.X. and W.K. designed the analytical approach and performed data analysis. Z.L. and J.C. performed the PCR validations. Z.L., L.W. and B.J. wrote the manuscript. Y.Z., S.Z. and H.-M.L. contributed to conceptual and analytical inputs of the paper.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Funding

This work was supported by the National Natural Science Foundation of China (grant no. 30870436), the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), and self-raised funds.

Availability of data

All whole-genome sequencing data from this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA548479.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

References

1. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc Natl Acad Sci U S A. 2012;109:2774–83.
2. 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;111:2310–8.
3. Kumar S. Molecular clocks: four decades of evolution. Nat Rev Genet. 2005;6:654–62.
4. Johnson T, Barton N. Theoretical models of selection and mutation on quantitative traits. Philos Trans R Soc Lond B Biol Sci. 2005;360:1411–25.
5. Lynch M. Evolution of the mutation rate. Trends Genet. 2010;26:345–52.
6. 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;109:18488–92.
7. Wang L, Ji Y, Hu Y, Hu H, Jia X, Jiang M, et al. The architecture of intra-organism mutation rate variation in plants. PLoS Biol. 2019;17:e3000191.
8. Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, et al. The molecular diversity of adaptive convergence. Science. 2012;335:457–61.
9. Bragg JG, Supple MA, Andrew RL, Borevitz JO. Genomic variation across landscapes: insights and applications. New Phytol. 2015;207:953–67.
10. Flynn JM, Chain FJ, Schoen DJ, Cristescu ME. Spontaneous mutation accumulation in *Daphnia pulex* in selection-free vs. competitive environments. Mol. Biol Evol. 2016;34:160–73.
11. Gorter FA, Derks MF, van den Heuvel J, Aarts MG, Zwaan BJ, de Ridder D, et al. Genomics of adaptation depends on the rate of environmental change in experimental yeast populations. Mol Biol Evol. 2017;34:2613–26.
12. Kondrashov FA, Kondrashov AS. Measurements of spontaneous rates of mutations in the recent past and the near future. Philos Trans R Soc Lond B Biol Sci. 2010;365:1169–76.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

754 13. Farlow A, Long H, Arnoux S, Sung W, Doak TG, Nordborg M, et al. The
755 spontaneous mutation rate in the fission yeast *Schizosaccharomyces pombe*.
756 Genetics. 2015;201:737–44.
757 14. Ness RW, Morgan AD, Colegrave N, Keightley PD, Ness RW, Morgan AD, et al.
758 Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*.
759 Genetics. 2012;192:1447–54.
760 15. Ness RW, Kraemer SA, Colegrave N, Keightley PD. Direct estimate of the
761 spontaneous mutation rate uncovers the effects of drift and recombination in the
762 *Chlamydomonas reinhardtii* plastid genome. Mol Biol Evol. 2015;33:800–8.
763 16. Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, et al.
764 Direct estimation of per nucleotide and genomic deleterious mutation rates in
765 *Drosophila*. Nature. 2007;445:82–5.
766 17. Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. Analysis
767 of the genome sequences of three *Drosophila melanogaster* spontaneous mutation
768 accumulation lines. Genome Res. 2009;19:1195–201.
769 18. Keightley PD, Ness RW, Halligan DL, Haddrill PR. Estimation of the
770 spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster*
771 full-sib family. Genetics. 2014;196:313–20.
772 19. Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, et al. A
773 genome-wide view of *Caenorhabditis elegans* base-substitution mutation
774 processes. Proc Natl Acad Sci U S A. 2009;106:16310–14.
775 20. Denver DR, Wilhelm LJ, Howe DK, Gafner K, Dolan PC, Baer CF. Variation in
776 base-substitution mutation in experimental and natural lineages of
777 *Caenorhabditis nematodes*. Genome Biol Evol. 2012;4:513–22.
778 21. Foster PL, Lee H, Popodi E, Townes JP, Tang H. Determinants of spontaneous
779 mutation in the bacterium *Escherichia coli* as revealed by whole-genome
780 sequencing. Proc Natl Acad Sci U S A. 2015;112:5990–9.
781 22. Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, et al. A
782 genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl
783 Acad Sci U S A. 2008;105:9272–7.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

784 23. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw
785 RG, et al. The rate and molecular spectrum of spontaneous mutations in
786 *Arabidopsis thaliana*. Science. 2010;327:92–4.
787 24. Huang C, Lu M, Chang Y, Li W. Experimental evolution of yeast for high
788 temperature tolerance. Mol Biol Evol. 2018;35:1823–39.
789 25. Xu Y, Ramanathan V, Victor DG. Global warming will happen faster than we
790 think. Nature. 2018;564:30–2.
791 26. Bita CE, Gerats T. Plant tolerance to high temperature in a changing environment:
792 scientific fundamentals and production of heat stress-tolerant crops. Front Plant
793 Sci. 2013;4:273.
794 27. Wilczek AM, Cooper MD, Korves TM, Schmitt J. Lagging adaptation to warming
795 climate in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 2014;111:7906–13.
796 28. Lesk C, Rowhani P, Ramankutty N. Influence of extreme weather disasters on
797 global crop production. Nature 2016;529:84–7.
798 29. Kotak S, Larkindale J, Lee U, von Koskull-Doring P, Vierling E, Scharf, KD.
799 Complexity of the heat stress response in plants. Curr Opin Plant Biol.
800 2007;10:310–16.
801 30. Bäurle I. Plant heat adaptation: priming in response to heat stress. F1000Res.
802 2016;5:694.
803 31. Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional
804 regulatory network of plant heat stress response. Trends Plant Sci.
805 2017;22:53–65.
806 32. Liu HC, Lämke J, Lin SY, Hung MJ, Liu KM, Charng YY, et al. Distinct heat
807 shock factors and chromatin modifications mediate the organ-autonomous
808 transcriptional memory of heat stress. Plant J. 2018;95:401–13.
809 33. Springate DA, Kover PX. Plant responses to elevated temperatures: a field study
810 on phenological sensitivity and fitness responses to simulated climate warming.
811 Global Change Biol. 2014;20:456–65.
812
813

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

814 34. Jiang C, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP. Environmentally
815 responsive genome-wide accumulation of de novo *Arabidopsis thaliana*
816 mutations and epimutations. Genome Res. 2014;24:1821–9.
817 35. Weng ML, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG, et al.
818 Fine-grained analysis of spontaneous mutation spectrum and frequency in
819 *Arabidopsis thaliana*. Genetics. 2019;211(2):703–14.
820 36. Rutter MT, Shaw FH, Fenster CB. Spontaneous mutation parameters for
821 *Arabidopsis thaliana* measured in the wild. Evolution. 2010;64:1825–35.
822 37. Xia J, Han L, Zhao Z. Investigating the relationship of DNA methylation with
823 mutation rate and allele frequency in the human genome. BMC Genomics.
824 2012;13:S7.
825 38. Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants.
826 Nat Rev Mol Cell Biol. 2018;19:489–506.
827 39. Keith N, Tucker AE, Jackson CE, Sung W, Lucas-Lledó JI, Schrider DR, et al.
828 High mutational rates of large-scale duplication and deletion in *Daphnia pulex*.
829 Genome Res. 2016;26:60–9.
830 40. Konrad A, Thompson O, Waterston RH, Moerman DG, Keightley PD,
831 Bergthorsson U, et al. Mitochondrial mutation rate, spectrum and heteroplasmy in
832 *Caenorhabditis elegans* spontaneous mutation accumulation lines of differing
833 population size. Mol Biol Evol. 2017;34:1319–34.
834 41. Rando OJ, Verstrepen KJ. Timescales of genetic and epigenetic inheritance. Cell.
835 2007;128:655–68.
836 42. Shor E, Fox CA, Broach JR. The yeast environmental stress response regulates
837 mutagenesis induced by proteotoxic stress. PLoS Genet. 2013;9:e1003680.
838 43. Wu Q, Han TS, Chen X, Chen JF, Zou YP, Li ZW, et al. Long-term balancing
839 selection contributes to adaptation in Arabidopsis and its relatives. Genome Biol.
840 2017;18:217.
841 44. Chain FJ, Flynn JM, Bull JK, Cristescu ME. Accelerated rates of large-scale
842 mutations in the presence of copper and nickel. Genome Res. 2019;29:64–73.
843

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

844 45. Belfield EJ, Gan X, Mithani A, Brown C, Jiang C, Franklin K, et al.
845 Genome-wide analysis of mutations in mutant lineages selected following
846 fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. Genome Res.
847 2012;22:1306–15.
848 46. Li G, Chern M, Jain R, Martin JA, Schackwitz WS, Jiang L, et al. Genome-wide
849 sequencing of 41 rice (*Oryza sativa* L.) mutated lines reveals diverse mutations
850 induced by fast-neutron irradiation. Mol Plant. 2016;9:1078–81.
851 47. Schmid-Siegert E, Sarkar N, Iseli C, Calderon S, Gouhier-Darimont C, Chrast J,
852 et al. Low number of fixed somatic mutations in a long-lived oak tree. Nat Plants.
853 2017;3:926–9.
854 48. Hofmeister BT, Denkena J, Colomé-Tatché M, Shahryary Y, Hazarika R,
855 Grimwood J, et al. A genome assembly and the somatic genetic and epigenetic
856 mutation rate in a wild long-lived perennial *Populus trichocarpa*. Genome Biol.
857 2020;21:259.
858 49. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J, Clark RM, et al. The *Arabidopsis*
859 *lyrata* genome sequence and the basis of rapid genome size change. Nat Genet.
860 2011;43:476–81.
861 50. Sairam RK, Srivastava GC, Saxena DC. Increased antioxidant activity under
862 elevated temperatures: a mechanism of heat stress tolerance in wheat genotypes.
863 Biol Plant. 2000;43:245–51.
864 51. Bray CM, West CE. DNA repair mechanisms in plants: crucial sensors and
865 effectors for the maintenance of genome integrity. New Phytol. 2005;168:511–28.
866 52. Tuteja N, Ahmad P, Panda BB, Tuteja R. Genotoxic stress in plants: shedding
867 light on DNA damage, repair and DNA repair helicases. Mutat Res.
868 2009;681:134–49.
869 53. Jin B, Wang L, Wang J, Jiang KZ, Wang Y, Jiang XX, et al. The effect of
870 experimental warming on leaf functional traits, leaf structure and leaf
871 biochemistry in *Arabidopsis thaliana*. BMC Plant Biol. 2011;11:35.
872 54. Baer CF, Miyamoto MM, Denver DR. Mutation rate variation in multicellular
873 eukaryotes: causes and consequences. Nat Rev Genet. 2007;8:619–31.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

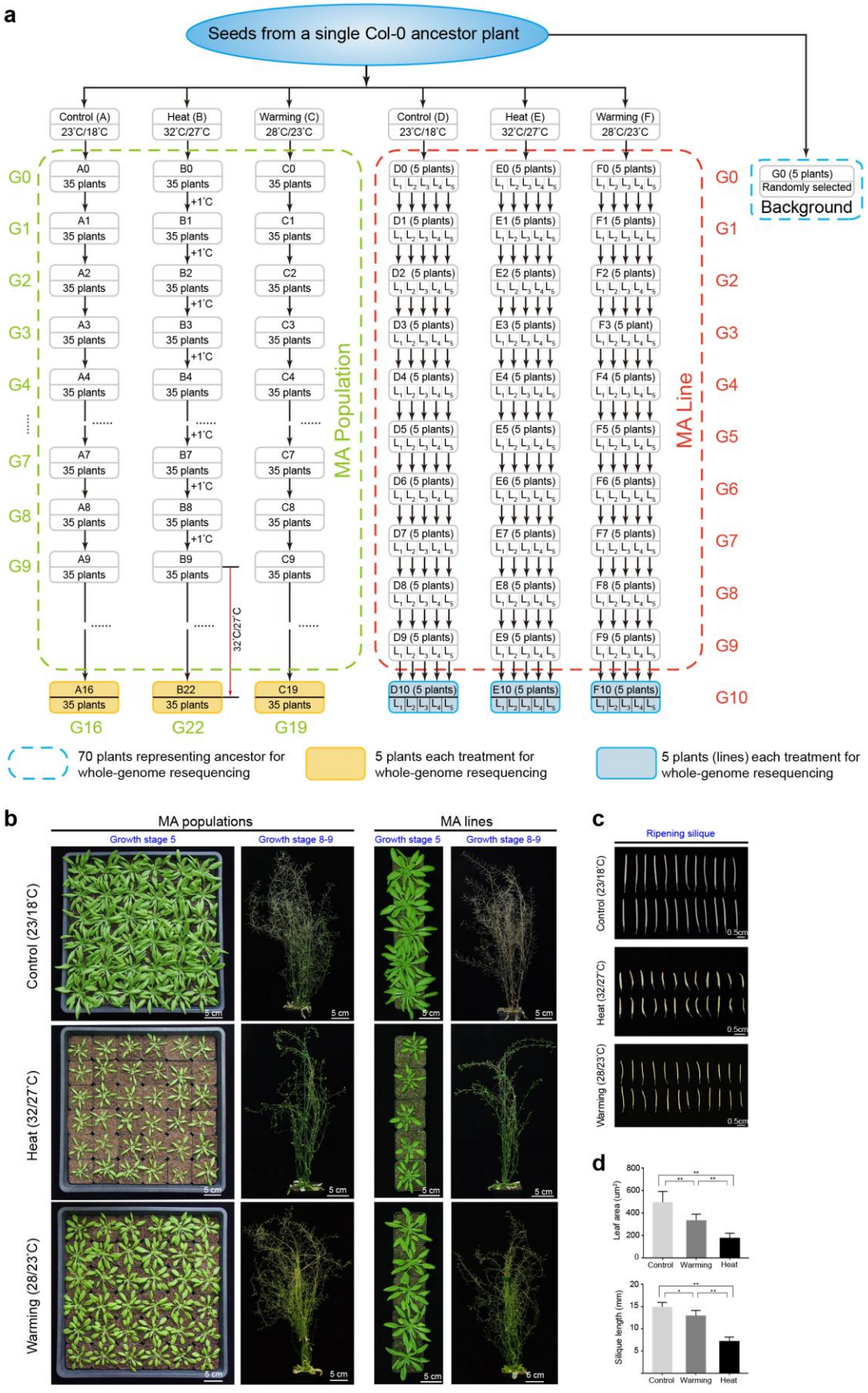
874 55. Dowle EJ, Morgan-Richards M, Trewick SA. Molecular evolution and the
875 latitudinal biodiversity gradient. Heredity. 2013;110:501.
876 56. Monroe JG, Srikant T, Carbonell-Bejerano P, Exposito-Alonso M, Weng M,
877 Rutter MT, et al. Mutation bias shapes gene evolution in *Arabidopsis thaliana*.
878 2020; https://doi.org/10.1101/2020.06.17.156752.
879 57. Kazazian HH. Mobile elements: drivers of genome evolution. Science
880 2004;303:1626–32.
881 58. Lisch D. How important are transposons for plant evolution? Nat Rev Genet.
882 2012;14:49–61.
883 59. Underwood CJ, Henderson IR, Martienssen RA. Genetic and epigenetic variation
884 of transposable elements in Arabidopsis. Curr Opin Plant Biol. 2017;36:135–41.
885 60. Quadrana L, Etcheverry M, Gilly A, Caillieux E, Madoui MA, Guy J, et al.
886 Transposition favors the generation of large effect mutations that may facilitate
887 rapid adaption. Nat Commun. 2019;10:3421.
888 61. Niu XM, Xu YC, Li ZW, Bian YT, Hou XH, Chen JF, et al. Transposable
889 elements drive rapid phenotypic variation in *Capsella rubella*. Proc Natl Acad Sci
890 USA. 2019;116:6908–13.
891 62. Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Scheid OM. How a
892 retrotransposon exploits the plant's heat stress response for its activation. PLoS
893 Genet. 2014;10:e1004115.
894 63. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, et al.
895 Transposable elements contribute to activation of maize genes in response to
896 abiotic stress. PLoS Genet. 2015;11:e1005566.
897 64. Rouzic AL, Boutin T, Capy P. Long-term evolution of transposable elements.
898 Proc Natl Acad Sci USA 2007;104:19375–80.
899 65. 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of
900 polymorphism in *Arabidopsis thaliana*. Cell. 2016;166:481–91.
901 66. Potters G, Pasternak TP, Guisez Y, Palme KJ, Jansen MAK. Stress-induced
902 morphogenic responses: growing out of trouble? Trends Plant Sci.
903 2007;12:98–105.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

67. Zhu JK. Abiotic stress signaling and responses in plants. Cell. 2016;167:313–24.
68. Xu YC, Niu XM, Li XX, He W, Chen JF, Zou YP, et al. Adaptation and phenotypic diversification in Arabidopsis through loss-of-function mutations in protein-coding genes. Plant Cell. 2019;31:1012–25.
69. Xu YC, Guo YL. Less is more, natural loss-of-function mutation is a strategy for adaptation. Plant Comm. 2020;1:100103.
70. Hu Z, Cools T, De Veylder L. Mechanisms used by plants to cope with DNA damage. Annu Rev Plant Biol. 2016;67:439–62.
71. Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 2010;107:21199–204.
72. Li Y, Cheng R, Spokas KA, Palmer AA, Borevitz JO. Genetic variation for life history sensitivity to seasonal warming in *Arabidopsis thaliana*. Genetics. 2014;196:569–77.
73. Wang L, Ma K, Lu Z, Ren S, Jiang H, Cui J, et al. Differential physiological, transcriptomic and metabolomic responses of Arabidopsis leaves under prolonged warming and heat shock. BMC Plant Biol. 2020;20:86.
74. Bac-Molenaar JA, Fradin EF, Becker FF, Rienstra JA, van der Schoot J, Vreugdenhil D, et al. Genome-wide association mapping of fertility reduction upon heat stress reveals developmental stage-specific QTLs in *Arabidopsis thaliana*. Plant Cell. 2015;27:1857–74.
75. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell. 2008;133:523–36.
76. Chinnusamy V, Zhu JK. Epigenetic regulation of stress responses in plants. Curr Opin Plant Biol. 2009;12:133–39.
77. Ueda M, Seki M. Histone modifications form epigenetic regulatory networks to regulate abiotic stress response. Plant Physiol. 2020;182:15–26.
78. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–72.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Figure legends



935

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Fig. 1 Schematic illustration and morphological comparison of *A. thaliana* grown under Control, Heat, and Warming conditions. **a** Schematic illustration of *A. thaliana* mutation accumulation (MA) lines and populations. Two MA experiments were conducted in this study (see Methods). For MA population experiments, seeds from the same ancestor plant as the MA lines were divided into three groups (~35 seedlings per group) and planted under Control conditions (A) for 16 generations, Warming conditions (C) for 19 generations, or Heat conditions (B) for 22 generations [the first 9 generations grown under gradual warming, i.e., increase of 1℃ per generation (from 24/18℃ to 32/27℃ [day/night]); the following 13 generations were grown at constant 32/27℃]. Five 10^{th} generation (Generation 10, G10) plants (five MA lines) from each treatment (D10, E10, F10) were used for individual whole-genome sequencing. Five 16^{th}, 22^{th}, 19^{th} generation plants from each treated population were also randomly selected for sequencing. For MA line experiments, seeds from a single Col-0 ancestor plant were grown independently under Control (D), Heat (E), or Warming (F) conditions for 10 successive generations. To maximize coverage and provide progenitor background genetic information (reference genome sequence) for MA experiments, five individuals (G0) were combined for sequencing. Genome-sequenced plants from MA lines and populations are highlighted in yellow- and grey-shaded (blue outline) boxes, respectively; see also Additional file 1: Table S1. **b** Growth status of MA plants exposed to Control, Heat, and Warming conditions at stage 5 (bolting) and stage 8–9 (silique ripening and senescence). Leaves at stage 5 (major axis ≤ 1 cm) were sampled for DNA extraction and sequencing. Scale bar, 5 cm. **c** Ripened siliques from the Control, Heat, and Warming treatments. Scale bar, 0.5 cm. **d** Phenotypic statistics of leaf area and silique length under different temperature treatments. Leaves at stage 5 (bolting) and siliques at stage 9 were measured. The experiments were repeated three times and the data are presented as means ± standard errors of the mean (SEMs; n = 30). Significant differences were revealed using analysis of variance (ANOVA) with post-hoc tests (*p < 0.05, **p < 0.01 vs. Control or Warming).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

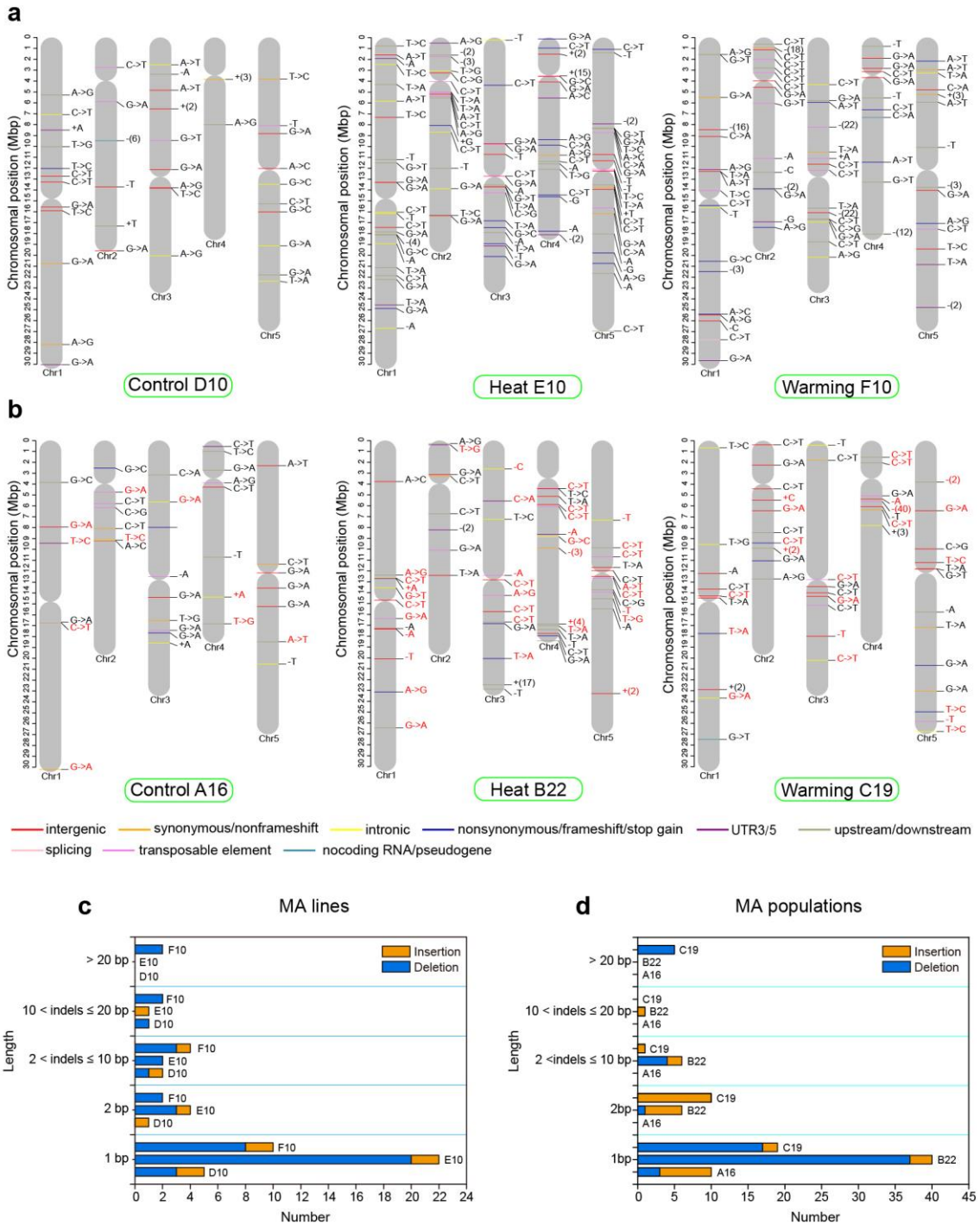


Fig. 2 Distribution across chromosomes of de novo mutations [single-nucleotide variants (SNVs) and small insertions and deletions (indels)] detected in genomes of *Arabidopsis* from the Heat, Warming, and Control MA lines and populations. **a, b** Labels indicate the type of mutation; colors indicate the functional class or predicted consequence. Single base insertions (ins) and deletions (dels) are indicated by base letters preceded by a plus and minus sign, respectively. Large ins and dels are

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

973 indicated by a plus (with the number of inserted base pairs) and minus sign (with the
974 number of deleted base pairs), respectively. Individual colors indicate intergenic
975 region (red), intron (yellow), synonymous/non-frameshift (orange),
976 nonsynonymous/frameshift/stop gain (blue), UTR3/5 (purple), upstream/downstream
977 (green), splicing (pink), transposable element (violet), and noncoding/pseudogene
978 (lake blue) mutations. Red labels in each MA population indicate the same mutations
979 detected in at least two sequenced samples. **c**, **d** Frequencies and categories of ins and
980 dels were determined based on their indel lengths (see also Additional file 1: Table
981 S3**)**.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

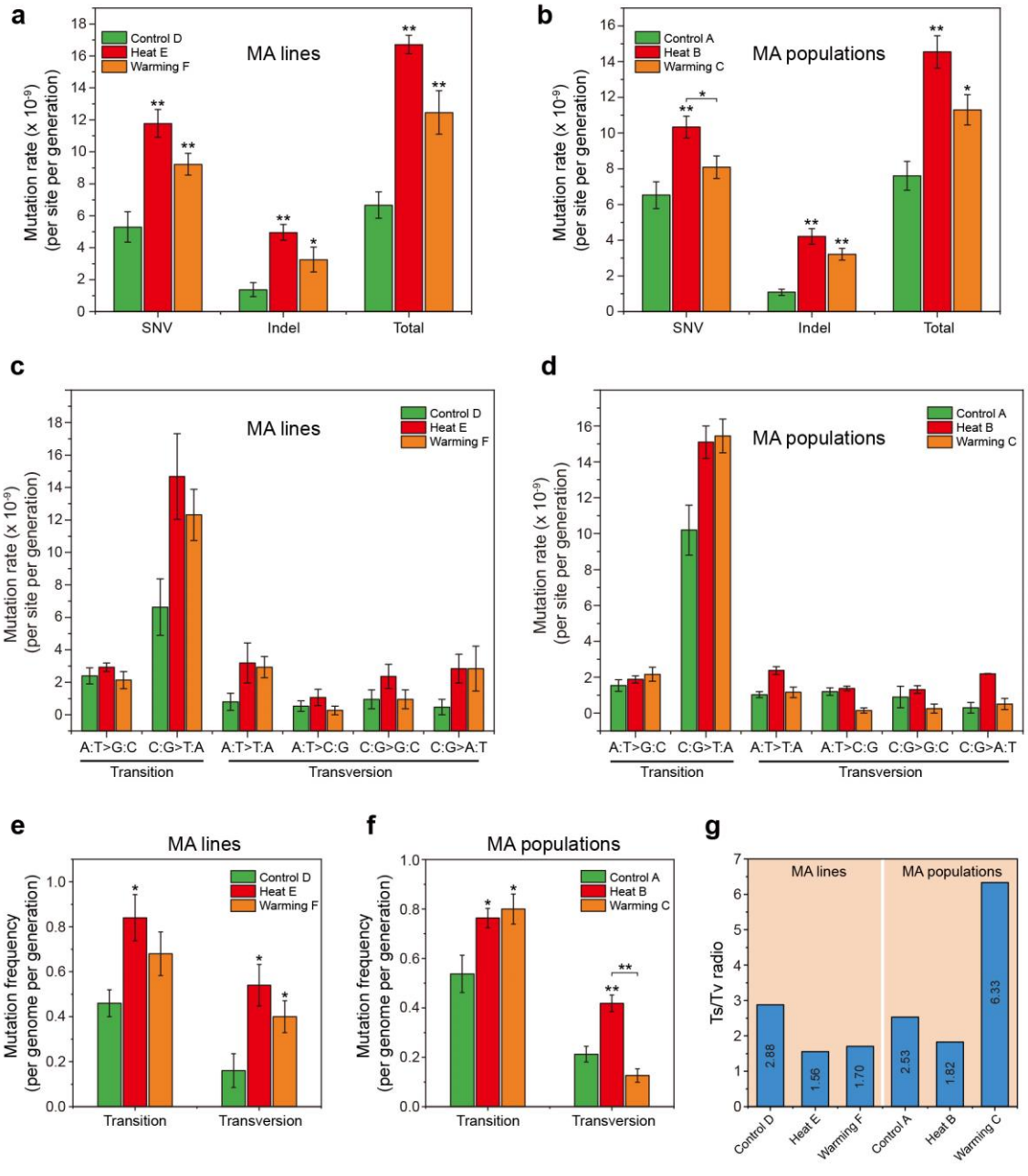


Fig. 3 Estimation of mutation rates of observed mutations (SNVs, indels) and molecular spectra in Control, Warming, and Heat MA lines and populations. **a**, **b** SNV, indel, and total mutation rates (per site per generation) of de novo mutations in MA lines and populations subjected to different temperature treatments. Significant differences were revealed using a two-tailed Student's *t*-test (*$p < 0.05$, **$p < 0.01$ compared to the Control or Warming treatments). **c**, **d** Mutation rates of different mutation types in MA lines and populations subjected to different temperature treatments. Conditional rates of each mutation type per site per generation were estimated by dividing the number of observed mutations by the number of analyzed

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

sites capable of producing a given mutation and the number of generations of MA in each Control, Warming, and Heat lineage and population lineage. Error bars indicate SEM. **e**, **f** Mutation frequencies (per genome per generation) of transition and transversion mutations accumulated in MA lines and populations subjected to different temperature treatments. Significant differences were revealed using a two-tailed Student's *t*-test (*$p < 0.05$, **$p < 0.01$ compared to the Control or Warming treatments). **g** Transition/transversion ratios (Ts/Tv) of SNVs accumulated in MA lines and populations subjected to different temperature treatments.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

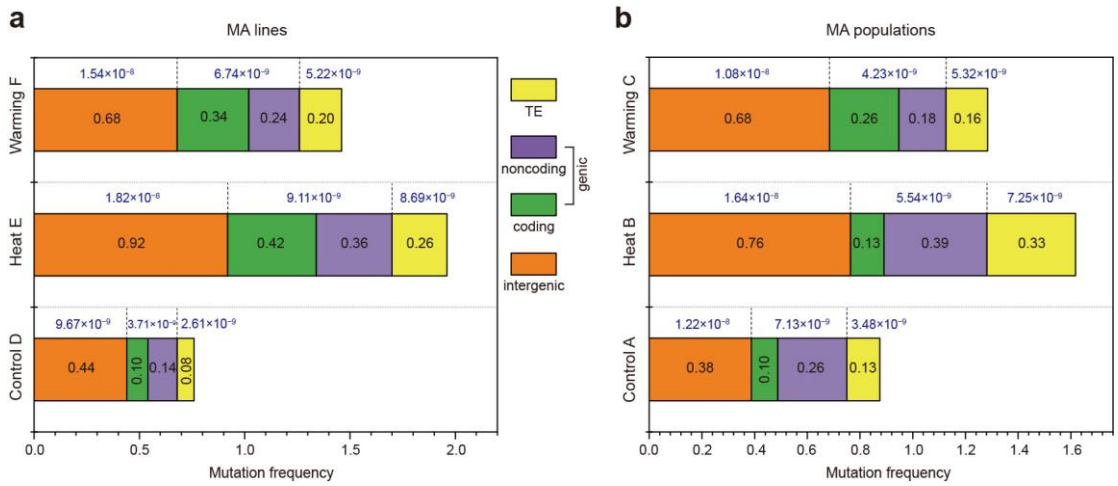


Fig. 4 Comparison of mutation frequencies in various genomic regions among the Control, Heat, and Warming lines (**a**) and populations (**b**). The numerical values in the stacked bar chart indicate the frequencies of total mutations (SNVs and indels) in the genomic regions. The values above the bars indicate the SNV rates in the genomic regions.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

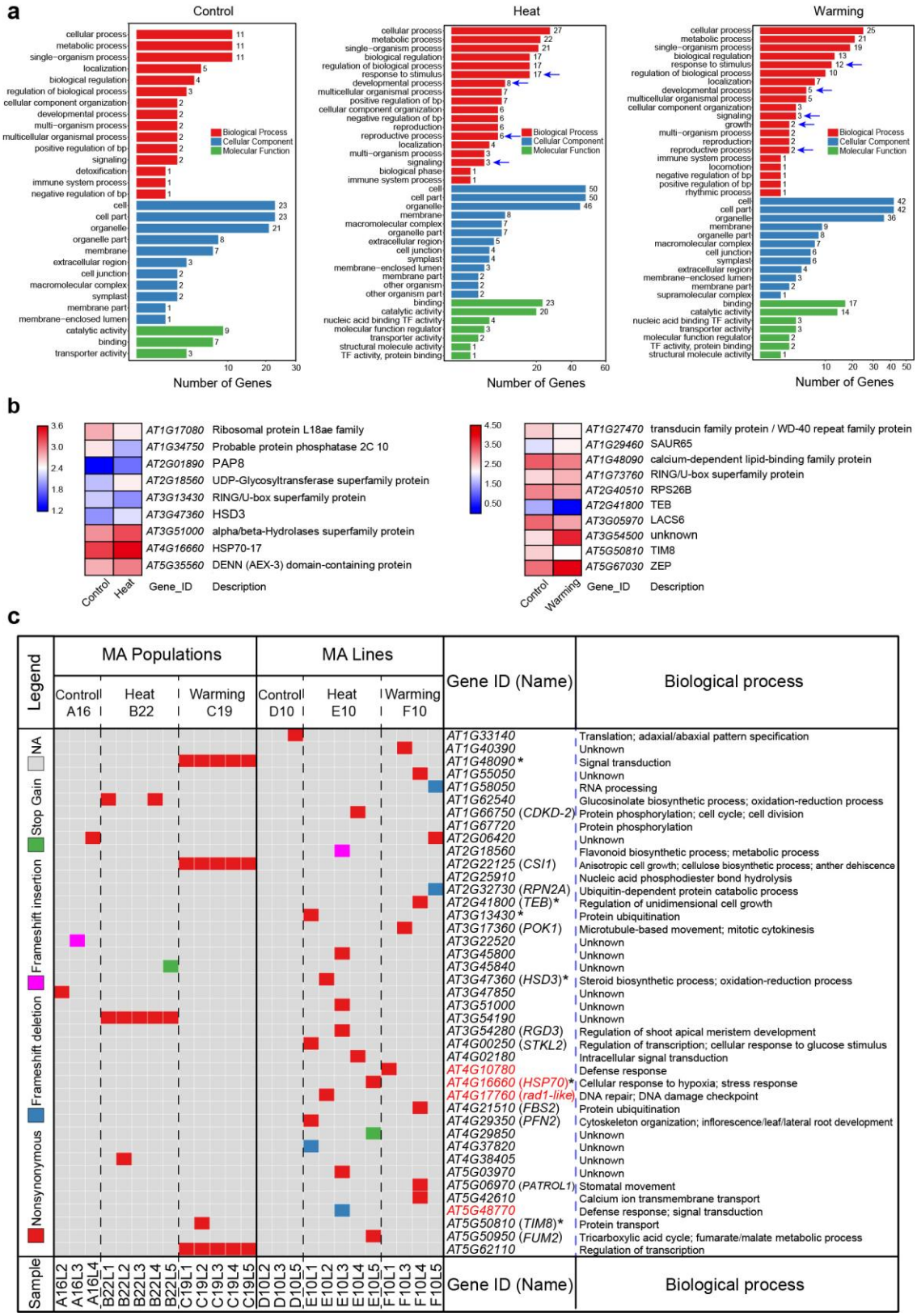


Fig. 5 Functional enrichment of mutated genes in MA lines and populations. **a** GO enrichment of mutated genes in the Control (A and D), Heat (B and E), and Warming (C and F) treatments. The arrow indicates an important biological process. **b**

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Expression levels of mutated genes under Heat and Warming differed significantly from the Control (expression dataset obtained from NCBI GSE118298). **c** Mutations (nonsynonymous, frameshift, and stop gain) in gene coding regions of MA lines and populations. Each gene involved in a putative biological process is shown. Defense response– and DNA repair–associated genes are marked in red, and asterisks indicate the differentially expressed genes shown in **b**.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

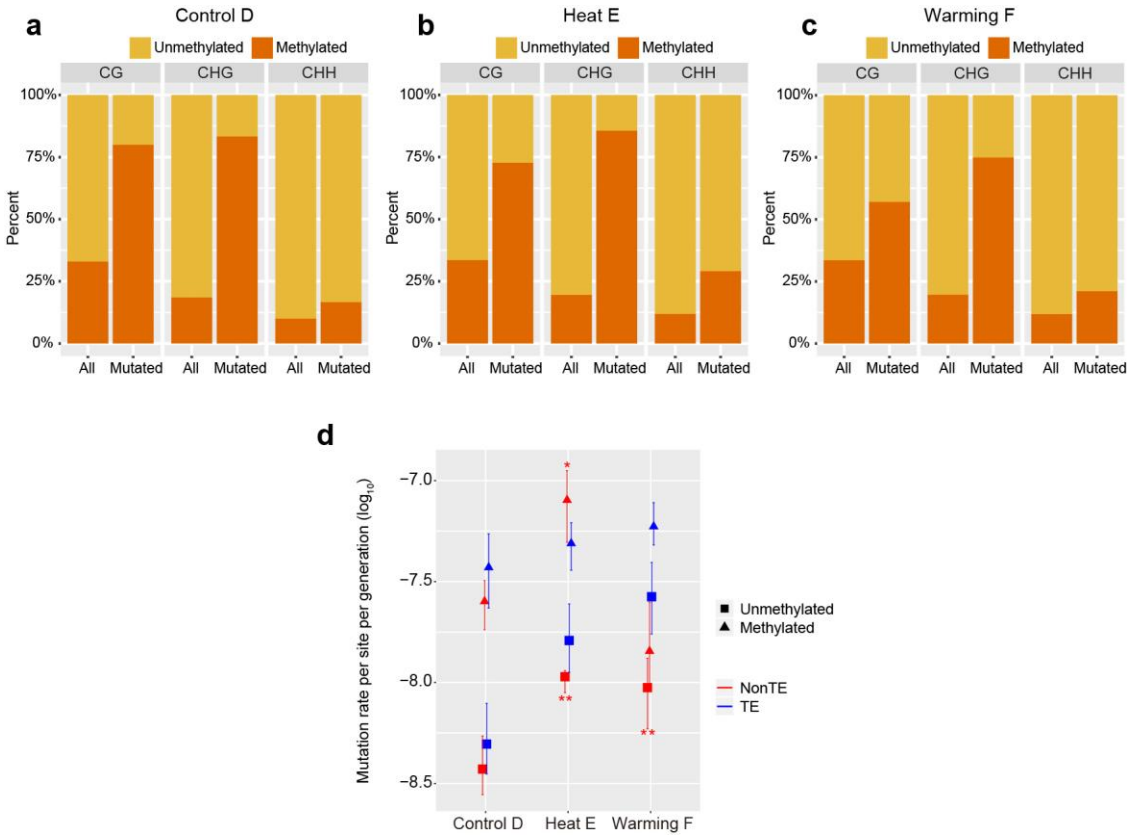
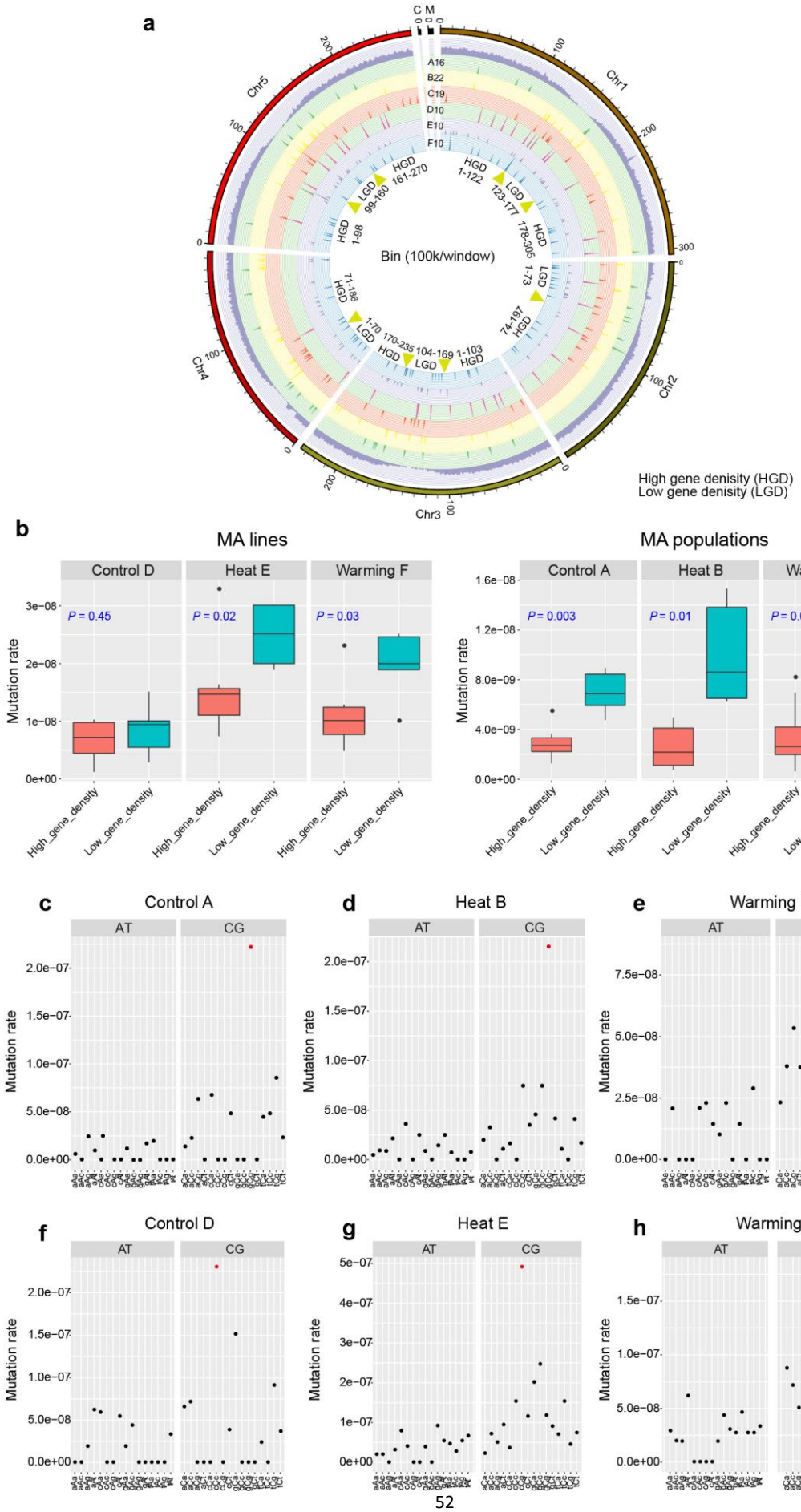


Fig. 6 Estimation of the effects of cytosine methylation, TE, and chromosome region on mutation rates in the Control, Heat, and Warming MA lines. **a**–**c** Comparison of cytosine methylation percentages at all bases in the genome and mutated bases in the Control D (A), Heat E (B), and Warming F (C) lines. H refers to A, T, or G. The methylation percentage is much higher at mutated bases than the corresponding genome-wide occurrence for all three contexts: CG (Fisher's exact test, p = 4.58e-8), CHG (Fisher's exact test, p = 1.92e-21), and CHH (Fisher's exact test, p = 1.63e-3). **d** Effects of cytosine methylation and TE region on mutation rates in the Control D (D), Heat E (E), and Warming F (F) lines. The x axis shows log-transformed (log10) mutation rates per site per generation. Mutation rates for non-TE and TE positions are marked in red and blue, respectively. Mutation rates for nonmethylated and methylated CG positions are indicated with triangles and squares, respectively. Differences in mutation rates among Control D, Heat E, and Warming F lines were assessed using Student's t test. Error bars indicates SEMs. Asterisks indicate significant differences from Control D at $p < 0.05$ (*) and $p < 0.01$ (**), respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Fig. 7 Mutational biases of Control, Heat, and Warming lines and populations. Analysis of correlations between gene density and mutation rates across chromosomes in Control, Heat, and Warming lines and populations. **a** Distribution of mutations across *A. thaliana* chromosomes shown in a Circos plot. From outer circle to inner circle, the plot shows the chromosomes, genes (purple bars), and mutations in A16 (green bars), B22 (yellow bars), C19 (red bars), D10 (pink bars), E10 (purple bars), and F10 (blue bars). Each chromosome is divided into multiple bins (bin size = 100 kb), which are grouped into high and low gene density regions. **b** Comparison of mutation rates between regions with high gene density and those with low gene density. Significant differences were revealed using two-tailed Student's *t* test ($p < 0.05$, high vs. low gene density regions). **c**–**h** Neighbor-dependent mutation rates at AT and GC bases estimated for the Control, Heat, and Warming MA lines (**c**–**e**) and populations (**f**–**h**). The trinucleotide context-dependent mutation rate is shown for each treatment. The *x* axis shows the focal nucleotides (uppercase, mutation site) and immediate flanking nucleotides (lowercase), regardless of strand orientation (e.g., the tAt class includes the overall mutation rate at tAt and aTa sites). For each treatment, the mutation rates of G/C bases were generally elevated relative to those of A/T bases. Red dots indicate significantly elevated mutation rates.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

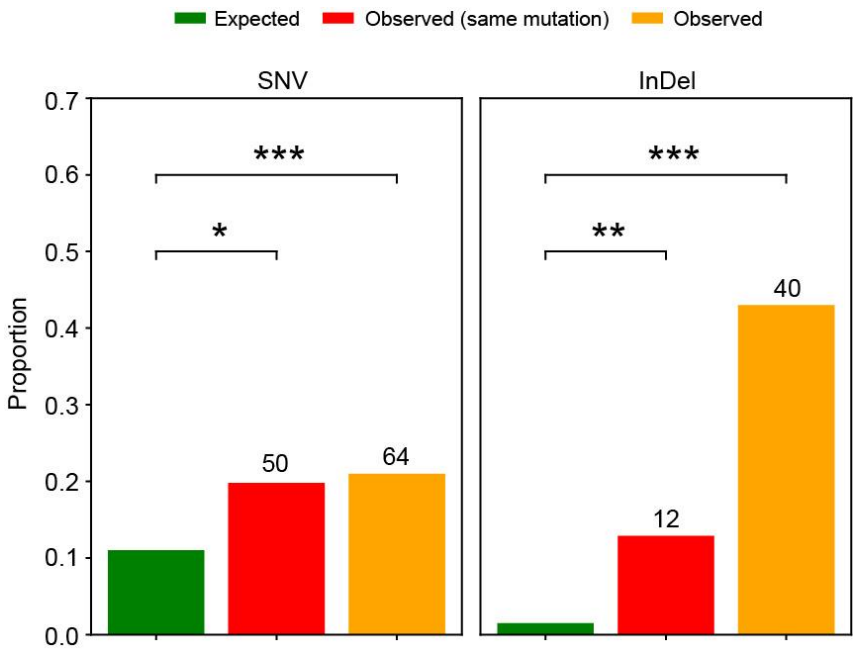


Fig. 8 Overlap between mutations identified in MA lines/populations (SNVs) and variants detected in the 1001 Genomes population (SNPs). Comparison of expected and observed proportions of SNVs and indels that overlap the SNPs and indels in the 1001 Genomes dataset. Numbers at the tops of the bars are absolute overlap values. Asterisks indicate $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) based on Fisher's Exact test with Bonferroni correction.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Supplementary information

Additional file 1: Table S1. Summary of whole-genome sequencing datasets of DNA samples.

Additional file 1: Table S2. Number of de novo mutations identified from MA lines and MA populations grown under Control, Heat, and Warming conditions.

Additional file 1: Table S3 List of individual sequence mutations identified in Control, Heat, and Warming MA lines and populations.

Additional file 1: Table S4. Simulated reference mutations recovered from six randomly selected MA lines.

Additional file 1: Table S5. Simulated sequence read mutations recovered from six randomly selected MA lines.

Additional file 1: Table S6. Validation of mutations detected in MA lines through conventional Sanger sequencing.

Additional file 1: Table S7. Numbers and frequencies of transitions (ts) and transversions (tv) identified in MA lines and populations grown under Control, Heat, and Warming conditions.

Additional file 1: Table S8. Whole-genome bisulfite sequencing and the numbers of methylated cytosine positions in genomic DNA samples.

Additional file 1: Table S9. Logistic regression analysis of the effects of cytosine methylation and TE regions on the likelihood of a given nucleotide being mutated in the Control, Heat, and Warming MA lines.

Additional file 1: Table S10. Genomic locations and affected bases of indels under Control, Heat, and Warming conditions.

Additional file 2: Fig. S1. KEGG enrichment of mutated genes in the Control (A and D), Heat (B and E), and Warming (C and F) treatments.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Additional 2: Fig. S2. Overlap between mutations identified in MA lines or populations (SNVs) and variants detected in the 1001 Genomes population (SNPs).

Additional file 2: Fig. S3. Relative mutation rates by local GC content across the *A. thaliana* genome of MA lines (Control D, Heat E, and Warming F) and populations (Control A, Heat B, and Warming C). A 1-kb bin size was used.

Additional file 3: Supplementary detailed methods.

Supplementary information: detailed methods

Temperature treatments for MA line and population experiments

To investigate the accumulation of mutations in *A. thaliana* grown under different temperature conditions over successive generations, we designed three temperature treatments for MA lines (single-seed descent maintained under laboratory conditions) and populations (simulation experiments) in growth chambers: 23/18°C (light/dark), 28/23°C, and 32/27°C (one growth chamber per temperature regime) (Fig. 1a). The 23/18°C (ambient temperature) regime was used as the Control treatment (groups D and A), because most studies use 22–23°C as the baseline temperature to investigate the effects of temperature on *Arabidopsis* [1, 2]. The 28/23°C (elevated 5°C relative to Control) regime was used as the Warming treatment (groups F and C) to simulate global warming according to a report that the global mean temperature is likely to warm by 1.5–4°C by the end of this century [3]; this treatment also represents an intermediate temperature between the Control and Heat treatments. The 32/27°C (elevated 9°C relative to control) regime was used as the Heat treatment (groups E and B), which is the maximum temperature under which *A. thaliana* can propagate for multiple generations. Because plants grown under 32/27°C have reached the upper limits for sexual reproduction, with greatly reduced seed-set rates, short siliques, fewer seeds, decreased leaf size (Fig. 1b–d), etc., plants grown at temperatures ≥ 33°C would not produce enough seeds for the next generation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Considering the gradual change in mean temperature seen in nature, for the Heat MA population (group B), we simulated a gradual warming through stepwise temperature increases in several preceding generations until the temperature reached 32℃. In detail, the temperature was elevated by 1℃ per generation for the first nine generations [from 24/19℃ (G1), to 25/20℃ (G2), 26/21℃ (G3), 27/22℃ (G4), 28/23℃ (G5), 29/24℃ (G6), 30/25℃ (G7), 31/26℃ (G8), and finally 32/27℃ (G9)], and then maintained at 32/27℃ for the subsequent 13 generations (Fig. 1a). To ensure the survival of seedlings in Heat (32/27℃) treatments (groups E and B), after cotyledons fully opened (i.e., stage 1.0) [4], we increased the temperature from 28/23℃ to 32/27℃ over a span of 3–4 days, then maintained at the temperature at 32/27℃.

We conducted this study using three temperature regimes, with five replicates per regime at the line level and 30 individuals at the population level (five individuals were selected for the next generation). These lines and populations have been maintained through many generations in 15 growth chambers over 10 years. This study involved tremendous labor, facilities, and time; thus, we are unable to grow more lines or populations through similar multigenerational cultivation.

For MA lines, ~100 seeds from the ancestral plant were isolated after 2 d of stratification at 4°C, and about 75 germinated seedlings were sown and finally thinned to 15 seedlings to establish the single-seed line system (15 lines; one seedling per line) (Fig. 1a, b). The 15 lines were divided into three groups (five lines per group) for the Control, Warming, and Heat treatments. They were planted in sterilized substrate [1:1 (v/v) mixture of vermiculite and peat] in individual pots, and the growth conditions were consistent with those of their ancestral plant (described above), except for the temperature settings. The resultant “Generation 0 (G0)” (Fig. 1a) plants from the three groups were grown and self-pollinated under Control, Warming, and Heat conditions in six independent champers, respectively. To offset any possible effect of position within each chamber, and the effects of the different chambers, the pots were randomly rearranged within each chamber, and interchanged among chambers

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1173 periodically. After G0 seeds were collected separately, five "Generation 1 (G1)"
1174 seedlings from each line within different temperature treatments were sown, while a
1175 single plant (after seedlings emerged) was randomly chosen to propagate the next
1176 generation. The remaining seeds from each line were stored in microcentrifuge tubes
1177 with desiccants. This process was repeated for nine subsequent generations under the
1178 Control, Warming, and Heat conditions, leading to 15 independent MA lines within
1179 the three treatments (five lines per treatment) spanning 10 successive generations.

1180 **For MA populations**, ~300 seeds from the same ancestral plant were initially isolated
1181 and then divided into three groups (35 seedlings per group) for the Control, Warming,
1182 and Heat treatments, respectively. Each population (group) was propagated in each
1183 generation using mixed seeds from five individual vigorous plants chosen from
1184 among the 35 seedlings (Fig. 1a, b). Except for the temperature treatment of the B
1185 group (described above), all growth conditions and management procedures within
1186 the chambers were identical to those in the MA lines. The MA populations grown
1187 under the three different treatments were propagated for more successive generations
1188 than the MA lines, with 16 generations for Control (A), 22 generations for Heat (B),
1189 and 19 generations for Warming (C). The accumulated generations for each
1190 population treatment differed, because warmer temperatures resulted in shorter
1191 generation times; thus, more generations grew under the Heat and Warming
1192 conditions than under the Control conditions within the same time period.

1193 To maximize coverage of the genetic background of the ancestral plant, ~100 seeds of
1194 the ancestral plant were sown to generate 100 individuals grown under the same
1195 conditions as the progenitor. Among these 100 individuals, five plants were randomly
1196 selected and sampled as the progenitors to establish the genetic background of the MA
1197 lines and populations.

1198 At stage 5 (bolting) and stage 8–9 (silique ripening), randomly selected leaves (~30)
1199 and siliques (~30) from 35 plants per treatment (Control, Heat, and Warming) were
1200 photographed using a digital camera, and the leaf area and silique length were
1201 measured using ImageJ software (NIH, Bethesda, MD, USA). Three replicates from

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

each treatment were analyzed.

DNA extraction and sequencing

The 15 MA lines (D10, E10, F10) at G10, and 15 plants from MA populations (five plants per population) at G16 (Control, A16), G22 (Heat, B22), and G19 (Warming, B19) (Fig. 1a), were chosen for sampling and sequencing. When each plant was grown to the bolting stage [leaves were fully expanded and matured in this stage; Fig. 1B (growth stage 5 according to [4])], the leaves were sampled and immediately frozen in liquid nitrogen. Similarly, the leaves of five plants representing the progenitor were also collected separately. Genomic DNA was extracted and purified from each plant using a MiniBEST Plant Genomic DNA Extraction Kit (TaKaRa, Dalian, China). To obtain DNA samples representing the progenitor (G0) (Figure 1A), genomic DNA samples from five independent plants (G0-1–G0-5) were used to construct five separate libraries. These five libraries were combined to constitute the genetic background of the progenitor. All DNA sequencing (paired-end) projects were performed by Frasergan Genomics Institute (Wuhan, China) and Novogene (Beijing, China) according to the manufacturer's specifications (Illumina, San Diego, CA, USA). Finally, the Illumina HiSeq 2500 and HiSeq X Ten platforms were employed to generate raw sequences with 125-bp (MA populations and G0) and 150-bp (MA lines) read lengths, respectively.

Sequence quality checking and filtering

Raw reads were cleaned by removing the following types of reads: 1) reads containing adaptor sequences; 2) reads with Phred quality t20; and 3) reads with length $\leq$ 50 bp. The remaining paired-end reads were used for variant calling. Consequently, we obtained approximately 190 Gb of high-quality sequences with an average of 43.3× genome coverage, which were used for subsequent analyses (Additional file 1: Table S1).

Sequence alignment, variant calling, and annotation

We aligned the high-quality reads from each sample to the *A. thaliana* genome (TAIR10) using the BWA-MEM (version 0.7.16a) algorithm [5]. To obtain

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

high-quality SNVs and indels from each sample, we applied the Genome Analysis Toolkit (GATK, version 4.0.4.0) [6] procedures to perform SNV and indel callings. In the recommended GATK procedure, after removing the duplicate reads with Picard MarkDuplicates, we called the SNVs and indels using the HaplotypeCaller method [7]. Subsequently, the overlapping or common variant calls from GATK were further filtered with the following parameters: 1) for SNVs, QD < 2.0 || FS > 60.0 || MQ < 20.0 || MQRankSum< –12.5 || ReadPosRankSum < –8.0, DP < 8; and 2) for indels, QD < 2.0 || FS > 200.0 || MQ < 20.0 || ReadPosRankSum < –20.0, DP < 5. After filtering, the preliminary high-quality SNVs and indels from each sample were retained for base quality score recalibration (BQSR), which was performed using the BAM files generated through the GATK calling method described above.

To ensure the accuracy of de novo mutations called in the MA lines (G10) and populations (G16 for Control, G22 for Heat, G19 for Warming), we employed the following stringent criteria: 1) for MA and ancestor (background) samples, SNVs and indels must be called in at least eight reads and five reads, respectively; 2) the called SNVs must include both the forward and reverse reads; 3) in each MA sample, the mutation candidates were compared to the background (G0-1–G0-5) to discard spurious candidates resulting from variation between the reference genome sequence and the genome of the background; 4) owing to alignment difficulties in the vicinity of indels, SNVs located around indels (< 10 bp on each side) in each MA sample were removed; (5) called indels with a ≤ 20 bp interval between them were discarded; 6) called indels that were repeated multiple times at that site in the reference genome were subjected to further validation through Sanger sequencing to ensure the accuracy of the indels. In addition, we further filtered the variant sites (SNVs and indels) that were heterozygous in G0 samples, resulting in the identification of homozygous SNVs and indels in each MA sample. Finally, all filtered variants remaining in the MA lines and populations were inspected manually using Integrated Genome Viewer software [8]. Accordingly, we obtained a list of reliable new homozygous variants in our MA lines and populations. De novo TE insertions in each sample were detected

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

with the Jitterbug software package [9].

To validate the reliability of our mutation calling pipeline, we conducted two types of simulation. For the first simulation, 600 random point mutations from throughout the reference genomes of six selected MA lines and populations (with different reference genome coverage depths, 100 mutations per reference) were simulated to test whether our pipeline could recover them [10, 11]. The simulated mutation positions remained independent among the six reference genomes. Twenty mutations per chromosome were simulated. Each clean read from these MA lines and populations was mapped separately to one of the mutated reference genomes, and then their genotypes were called using GATK. The number of simulated mutations recovered and the corresponding accessible nonvariant sites were reported. For the second simulation, homozygous mutations were introduced to the sequencing reads in the BAM files for the same MA lines and populations; for detailed simulation methods, refer to [11].

We then used the ANNOVAR software [12] to annotate the identified SNVs and indels, and divided them into groups of variations occurring in intergenic regions, exonic regions (overlapping a coding exon), intronic regions (overlapping an intron), splicing sites, upstream or downstream regions (within a 1-kb region upstream from the transcription start site or downstream from the transcription stop site), UTR regions, noncoding RNAs, and pseudogenes, on the basis of *A. thaliana* genome annotation information (TAIR10). SNVs in coding regions were further divided into nonsynonymous SNVs (those that caused amino acid changes) or synonymous SNVs (those that did not cause amino acid changes); frameshift, stop-gains, and stop-losses caused by mutations were also classified in coding regions. In addition, an Araport11 genes and transposons GFF file [13] was used to further annotate mutations occurring in TEs. In addition, previous studies have suggested that 66% of mutations in coding regions could be purged through natural selection. According to the method of [14], we estimated the genomic mutation rate (diploid) affecting the fitness of each MA line by multiplying our mutation rate (per site per generation) in the line's coding region

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

by the mean selective constraint of 0.66, and then doubling that value to obtain the total number of coding sites. Finally, the average genomic mutation rates affecting fitness in the Control, Heat, and Warming groups were calculated.

To identify the mutations occurring in potential temperature-responsive transcripts in MA lines and population, we firstly retrieved the previous transcriptome datasets (deposited in GSE118298; [11]) of *A. thaliana* leaves grown at prolonged warming, heat shock and control. Then, differentially expressed genes (control *vs* heat shock; control vs prolonged warming) occurred in the mutated gene of MA lines and populations were considered as temperature-responsive transcripts.

Mutation rate estimation

The mutation frequency (per genome per generation) and rate (per site per generation) in each treatment were estimated by the formulas $m = n/g$ and $\mu = n/gb$ (or $\mu = m/b$). For example, among the single-seed descent MA lines, Heat E10L1 had 15 SNVs (n = 15) and g = 10; therefore, the mutation frequency m = 15/10 = 1.5 SNVs per generation per line. Similarly, the SNV frequencies of the Heat E10L2, E10L3, E10L4, and E10L5 lines were calculated as 1.4 (14/10), 1.6 (16/10), 1.4 (14/10), and 1.0 (10/10), respectively, and the average mutation frequency of Heat E was 1.4 ($\sum m/5$) mutations per generation. Based on the calculated mutation frequency, we further estimated the per site per generation mutation rate using the formula $\mu = n/gb$ or (m/b), where b is the number of accessible reference sites analyzed. Taking Heat E10L1 as an example, the mutation rate μ_{SNV} = 1.5/ 117,293,204 (the reference sites accessible for variant calling) = 1.28e-08, and correspondingly, the final SNV mutation rate of Heat E calculated by averaging the individual mutation rates was μ = 1.18e-08.

However, due to the limited number of generations in our MA experiments, actual mutation rates may be underestimated using the above calculations [14, 15]. The reason for the potential underestimation of mutation rates has been analyzed in detail by [16]. In brief, the putative formula $g\mu + 1/2\tau\ (g - 2 + (1/2)^{g-1})$ was used to show the total probability of accumulated homozygous mutations over g generations, where μ and τ represent the probabilities of a new homozygous and new heterozygous

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

mutation per site per generation, respectively. As a consequence, after g generations, the count of mutations was $N \cdot g \cdot (\mu + 1/2\tau)$, where N is the number of sites. In the present case, g = 10, and the mutation rate (per site per generation) obtained by dividing the count of identified homozygous mutations by the number of generations and the number of sites is $\mu + 4\tau/10$, which is lower than the expected homozygous mutation rate, that is, $\mu + 5/10\tau$. Because of the difficulty in estimating μ and τ during plant development [14, 15], we are not able to accurately correct the underestimation. However, this underestimation (20% underestimation of the actual mutation rate) had a limited effect on our results and conclusion, similar to analyses performed by [14] and [16]. Since we applied the same approach to estimate mutation rates in MA lines and populations under Control, Heat, and Warming conditions, the underestimation should have little effect when comparisons of mutation rates are performed.

References

1. Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, and Ahn JH. Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. Genes Dev. 2007;21:397–402.
2. Jin B, Wang L, Wang J, Jiang KZ, Wang Y, Jiang XX, et al. The effect of experimental warming on leaf functional traits, leaf structure and leaf biochemistry in *Arabidopsis thaliana*. BMC Plant Biol. 2011;11:35.
3. Flato G, Marotzke J, Abiodun B, Braconnot P, Chou SC, Collins W, et al. Evaluation of climate models, in: *Climate Change 2013*: *The Physical Science Basis*. *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Stocker T.F., et al. Cambridge Univ. Press, Cambridge, U. K. and New York, USA, 2013. pp. 741–866.
4. Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, et al. Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. Plant Cell. 2001;13:1499–510.
5. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1348 transform. Bioinformatics. 2009;25:1754–60.
1349 6. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al.
1350 The Genome Analysis Toolkit: a MapReduce framework for analyzing
1351 next-generation DNA sequencing data. Genome Res. 2010;20:1297–1303.
1352 7. Depristo MA, Banks E, Poplin R, et al. A framework for variation discovery and
1353 genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43,
1354 491–8.
1355 8. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV):
1356 high-performance genomics data visualization and exploration. Brief Bioinform.
1357 2013;14:178–92.
1358 9. Hénaff E, Zapata L, Casacuberta JM, Ossowski S. Jitterbug: somatic and germline
1359 transposon insertion detection at single-nucleotide resolution. BMC Genomics.
1360 2015;16:768.
1361 10. Ness RW, Morgan AD, Colegrave N, Keightley, P.D., Ness RW, Morgan AD, et al.
1362 Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*.
1363 Genetics. 2012;192:1447–54.
1364 11. Weng M, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG et al.
1365 Fine-grained analysis of spontaneous mutation spectrum and frequency in
1366 *Arabidopsis thaliana*. Genetics, 2019;211: 703–14.
1367 12. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic
1368 variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
1369 13. Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD.
1370 (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference
1371 genome. Plant J. 89:789–804.
1372 14. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG,
1373 et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis*
1374 *thaliana*. Science. 2010;327:92–4.
1375 15. Hoffman PD, Leonard JM, Lindberg GE, Bollmann SR, Hays JB. Rapid
1376 accumulation of mutations during seed-to-seed propagation of
1377 mismatch-repair-defective *Arabidopsis*. Genes Dev. 2004;18:2676–85.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.17.423344; this version posted December 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

16. Jiang C, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. Genome Res. 2014;24:1821–9.