Received: 17 December 2021 | Accepted: 4 January 2022

DOI: 10.1111/pce.14268

ORIGINAL ARTICLE

Plant, Cell & Environment | WILEY

Genome survey sequencing of wild cotton (*Gossypium robinsonii*) reveals insights into proteomic responses of pollen to extreme heat

Farhad Masoomi-Aladizgeh[1] | Karthik Shantharam Kamath[2] | Paul A. Haynes[3] | Brian J. Atwell[1]

[1]School of Natural Sciences, Macquarie University, North Ryde, New South Wales, Australia

[2]Australian Proteome Analysis Facility, Macquarie University, North Ryde, New South Wales, Australia

[3]School of Natural Sciences, Macquarie University, North Ryde, New South Wales, Australia

Correspondence
Brian J. Atwell, School of Natural Sciences, Macquarie University, North Ryde, NSW 2109, Australia.
Email: brian.atwell@mq.edu.au

Funding information
The Linnean Society of NSW; The North Shore Group of the Australian Plants Society; BioBam (OmicsBox)

Abstract

Heat stress specifically affects fertility by impairing pollen viability but cotton wild relatives successfully reproduce in hot savannas where they evolved. An Australian arid-zone cotton (*Gossypium robinsonii*) was exposed to heat events during pollen development then mature pollen was subjected to deep proteomic analysis using 57 023 predicted genes from a genomic database we assembled for the same species. Three stages of pollen development, including tetrads (TEs), uninucleate microspores (UNs) and binucleate microspores (BNs) were exposed to 36°C or 40°C for 5 days and the resulting mature pollen was collected at anthesis (p-TE, p-UN and p-BN, respectively). Using the sequential windowed acquisition of all theoretical mass spectra proteomic analysis, 2704 proteins were identified and quantified across all pollen samples analysed. Proteins predominantly decreased in abundance at all stages in response to heat, particularly after exposure of TEs to 40°C. Functional enrichment analyses demonstrated that extreme heat increased the abundance of proteins that contributed to increased messenger RNA splicing via spliceosome, initiation of cytoplasmic translation and protein refolding in p-TE40. However, other functional categories that contributed to intercellular transport were inhibited in p-TE40, linked potentially to Rab proteins. We ascribe the resilience of reproductive processes in *G. robinsonii* at temperatures up to 40°C, relative to commercial cotton, to a targeted reduction in protein transport.

KEYWORDS
crop wild relatives, heat stress, metabolic down-regulation, pollen development, SWATH-MS, tetrads, thermotolerance

1 | INTRODUCTION

Attempts to develop crops resilient to heatwaves are increasingly important as we experience changes in the Earth's climate, particularly global average temperatures that are predicted to increase up to 2°C during 21st century (IPCC, 2021). Extreme temperatures are considered to be detrimental to vegetative and reproductive processes in crops, ultimately diminishing yield. The global yield of major crops is likely to diminish by approximately 3%–7% for each degree-Celsius increase in average temperature (Zhao et al., 2017). Detailing the cellular responses to heat stress at precise developmental stages of germ cell lines will help us define the genetic markers required for molecular plant breeding of thermotolerant crops.

For plants to reproduce sexually, egg cells in the ovary must be successfully fertilized by sperm cells from the pollen tube. The reproductive phase of plants, particularly male reproduction, is highly vulnerable to heat stress (De Storme & Geelen, 2014; Müller & Rieu, 2016; Zinn et al., 2010). Microsporogenesis, including the development of tetrads (TEs) and unicellular (UN) microspores following meiotic divisions, is considered the most sensitive stage of pollen development to heat stress in plants (Begcy et al., 2019; Endo et al., 2009; Masoomi-Aladizgeh et al., 2020; Sato et al., 2002), with gametogenesis that follows possibly less heat sensitive, as reported in the binucleate stage of development in cotton (Masoomi-Aladizgeh et al., 2020). Molecular mechanisms that confer heat tolerance when the meiotic stage of pollen development is exposed to 40°C could be revealed by investigating related wild cotton species endemic to hot dry semi-deserts.

Wild species are valuable sources of novel genes, with unique adaptive mechanisms that enable survival in inhospitable environments. Cotton (*Gossypium*) is an important crop that is commonly grown for natural fibre production across the world. The genus *Gossypium* includes eight groups of diploid species ($n = 13$) divided into multiple genomes (A–G and K), of which C-genome species (*Gossypium robinsonii* and *G. sturtianum*) are endemic to Australia (Wendel & Cronn, 2003). Wild cotton species host many unique traits for resistance to abiotic stresses such as heat, drought and salinity (Mammadov et al., 2018) but after early work on the cytology and taxonomy of the genus there are no reports on the mechanism(s) of thermotolerance in these species. Hence, the applied goal of this study is to identify novel genes from distantly related species and introgress them into cultivated cotton without a yield penalty.

We sequenced genomic DNA from *G. robinsonii*, an Australian wild cotton C-genome species ($2n = 2x = 26$), and developed a draft genome for this species. This enabled us for the first time to use genome survey sequencing (GSS) of a wild cotton (*G. robinsonii*), thereby translating predicted gene sequences for deep analysis of our *G. robinsonii* sequential windowed acquisition of all theoretical mass spectra (SWATH) proteomic data. Mature pollen was collected for proteomic analysis after pollen cell lines of *G. robinsonii* had been exposed to moderate (36°C) and extreme heat (40°C). Observations that *G. robinsonii* plants set seed even at 40°C led to the hypothesis that heat stress during three stages of pollen development would elicit developmentally distinct proteomic patterns in mature pollen and that while some proteins will reflect impaired metabolism as a result of heat stress, others play a direct role in heat stress tolerance.

2 | MATERIALS AND METHODS

2.1 | Plant materials

G. robinsonii ($2n = 2x = 26$; Australian wild cotton) was grown in glasshouse conditions at 30/22°C day/night temperature and 12/12 h light/dark photoperiod under natural light. The light intensity during the day cycle was maintained at a minimum of 600 µmol m^{-2} s^{-1} using supplemented light (Philips LED grow lights). Plants were fertilized weekly using Aquasol® soluble fertilizer (Yates, Australia) at the rate of 4 g per 5 L and watered daily. For GSS, young leaves grown under control conditions were collected. For proteomics, three stages of pollen development in *G. robinsonii* including TEs (5–5.5 mm), uninucleate microspores (UNs; 7–10 mm) and binucleate microspores (BNs; 13–24 mm) were exposed to 36/25°C (moderate heat) or 40/30°C (extreme heat) for 5 days and corresponding mature pollen grain samples were collected and analysed in biological triplicate. Figure 1 shows the workflow of GSS on leaf and proteomics analysis of mature pollen heated at early stages of development.

2.2 | DNA isolation

Total genomic DNA was extracted from approximately 100 mg young leaf tissue using a combination of a method described in Masoomi-Aladizgeh et al. (2016) and a DNeasy Plant Mini Kit (Qiagen). Briefly, a lysis buffer for DNA extraction was prepared as described in Masoomi-Aladizgeh et al. (2016). The lysis buffer (1200 µl) in addition to β-mercaptoethanol (50 µl) was added to the ground leaf and samples were incubated at 65°C for 10 min. Chloroform (600 µl) was added to each tube and the mixture was vortexed vigorously, followed by centrifugation at 14 000 rpm at room temperature for 10 min. The supernatant (650 µl) was transferred into the QIAshredder Mini spin column to continue the isolation processes according to the manufacturer's instructions. Isolated DNA was treated with 2 µl RNase, and the quality and quantity of the sample were assessed using agarose gel electrophoresis and a fluorometric procedure (Qubit Fluorometer; Invitrogen) by the Beijing Genomics Institute (BGI).

2.3 | Whole-genome sequencing

A paired-end library with insert size of 350 base pairs (bp) was constructed from randomly fragmented genomic DNA in the BGI. A BGISEQ-500 sequencer was utilized to generate sequencing data with a read length of 150 bp. The quality of the reads was controlled and a stringent filtering process was carried out to obtain clean data for subsequent analyses as described by Li, Fan et al. (2010). These reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive database and are available under BioProject accession number PRJNA592601. The clean data were used for *K*-mer analysis to estimate the size of the genome, repetitive sequences and heterozygosis (Li, Zhu et al., 2010). Short-read sequences were assembled to contigs and scaffolds using SOAPdenovo (Li, Zhu et al., 2010; Potato Genome Sequencing et al., 2011). Five-kilobase nonoverlapping sliding windows along the assembled sequence was used to calculate the guanine plus cytosine (GC) content and average sequencing depth among the windows.

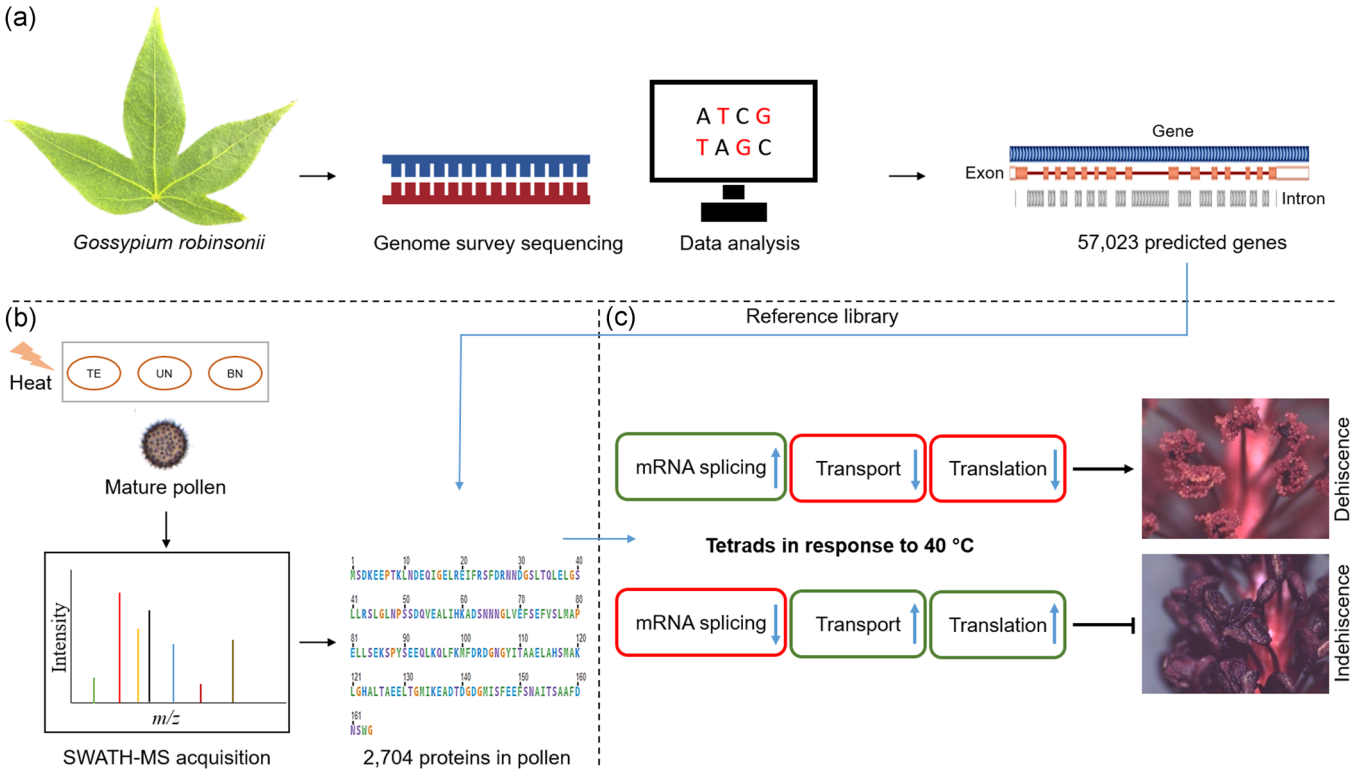


FIGURE 1 (a) Genome survey sequencing of *Gossypium robinsonii* resulting in a total of 57 023 genes. (b) Exposure of three developmental stages of pollen development (TEs, UN and BN) to 36 or 40°C for 5 days and proteomics analysis of the resulting mature pollen (p-TE, p-UN and p-BN). (c) The proposed thermotolerance mechanism of tetrads in *G. robinsonii*. BN, binucleate microspores; SWATH-MS, sequential windowed acquisition of all theoretical mass spectra; TE, tetrad cells; UN, uninucleate microspores

2.4 | Phylogenetic analysis of *G. robinsonii*

The cotton internal transcribed spacer (ITS) regions (ITS1, 5.8S ribosomal RNA and ITS2) sequences were downloaded from the NCBI and compared with the ITS of *G. robinsonii* in this study. The alignment of these sequences was performed using ClustalW following default parameters, and a phylogenetic tree was constructed using the MEGA X software with maximum likelihood inference. Data were analysed using a general time-reversible model with gamma-distributed rate variation across five categories and invariant sites (Kumar et al., 2018). A total of 500 bootstrap replicates were run for this analysis.

2.5 | Gene prediction and annotation

For de novo gene prediction, scaffolds of larger than 1000 bp were processed in OmicsBox ver. 1.4.11. Repeat elements in the genome were identified and masked by searching the RepBase database (Bao et al., 2015) using RepeatMasker ver. 4.0.9 and Dfam ver. 3.0 (Hubley et al., 2016; Smit et al., 2013). Augustus was employed to predict genes from the repeat-masked genome using the following parameters trained on *Arabidopsis thaliana*: ab initio as gene finding method, Qmap threshold of 30, minimum read alignment of 11, minimum intron length of 32, minimum exon length of 300 and depth coverage of 20 (Hoff & Stanke, 2013). For gene annotation, BLASTx alignment was performed between the predicted genes and common databases including UniProtKB and Swiss-Prot (*E* value <1*e* − 3). The predicted genes were also aligned to GO mapping (Götz et al., 2008) and InterProScan (Hunter et al., 2009) to obtain functional annotation.

2.6 | Protein isolation and peptide preparation

Proteins were extracted by a phenol-based protocol previously described by Masoomi-Aladizgeh et al. (2020). The concentration of proteins was measured by BCA protein assay (Pierce, Thermo Fisher Scientific). A total of 85 µg proteins from each sample was digested in-solution with trypsin, which was then followed by desalting the resulting peptides using stage tips packed in-house with SDB-RPS (styrenedivinylbenzene-reversed-phase sulphonate) (3M) as described in Hamzelou, Pascovici et al. (2020). The peptide concentration was then measured using the Micro BCA assay (Pierce, Thermo Fisher Scientific) and an equal amount of peptide sample from each individual sample type was used for the subsequent analysis.

2.7 | NanoLC-MS/MS

To determine whether the draft genome of *G. robinsonii* assembled in this study would provide sufficient information for further proteomics

analyses, an initial shotgun proteomic assessment was performed. Nonfractionated peptides were analyzed by nanoflow liquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS) using a Q-Exactive, Orbitrap mass spectrometer (Thermo Fisher Scientific) coupled to an Easy nLC 1000 nanoflow liquid chromatography system as described in Hamzelou, Kamath, et al. (2020).

2.8 | High pH (HpH) reversed-phase fractionation

For the generation of an ion library, a small fraction of peptides from all samples was pooled and fractionated using a Pierce High pH Reversed-Phase Peptide Fractionation Kit (Pierce, Thermo Fisher Scientific). Briefly, 60 µg of the pooled peptides was reconstituted in 300 µl of 0.1% trifluoroacetic acid. The mixture was transferred into a high pH reversed-phase spin column and the flow-through fraction was retained after centrifugation at 3000 *g* for 3 min. The same procedure was repeated with Milli-Q water to collect the wash fraction. Peptides were sequentially eluted using a total of eight gradient fractions (5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20% and 50%) of acetonitrile and triethylamine (0.1%). The peptide fractions were then vacuum-dried and resuspended in 24 µl of 0.1% formic acid (FA).

2.9 | Spectral library creation and SWATH-MS acquisition

Samples were analyzed in two stages. First, in a data-dependent mode for ion-library generation, followed by a data-independent acquisition on SWATH mode for peptide quantification. Both were performed using LC-MS/MS on a Triple TOF 6600 mass spectrometer (Sciex) equipped with an Eksigent nanoLC 400 liquid chromatography system (Sciex).

2.9.1 | Information-dependent acquisition (IDA)

HpH fractionated peptides were injected onto a reversed-phase trap (Halo-C18, 160 Å, 2.7 µm, 150 µm × 3.5 cm) for preconcentration and desalted with loading buffer. The peptide trap was then switched in line with the analytical column (Halo-C18, 160 Å, 2.7 µm, 200 µm × 20 cm). Peptides were eluted from the column using a linear solvent gradient of 5%–35% of mobile phase B (0.1% FA, 99.9% acetonitrile) over 60 min at a flow rate of 600 nl min^{-1}. The reversed-phase nano-LC eluent was subject to positive ion nano-flow electrospray analysis in an IDA mode. In IDA mode, a time-of-flight mass spectrometry (TOF-MS) survey scan was acquired (*m/z* 350–1500, 0.25 s) with the 20 most intense multiply charged ions (2+ to 4+; exceeding 200 counts s^{-1}) in the survey scan being sequentially subjected to MS/MS analysis. MS/MS spectra were accumulated for 100 ms in the mass range *m/z* 100–1800 using rolling collision energy.

2.9.2 | Data-independent acquisition by SWATH

For SWATH-MS, peptides of the individual samples were separated over RP linear gradient using the same LC and MS instruments as specified above with positive ion nanoflow electrospray mode. In SWATH mode, first a TOF-MS survey scan was acquired (*m/z* 350–1500, 0.05 s) and then the 100 predefined *m/z* ranges were sequentially subjected to MS/MS analysis. A total of 100 variable windows were selected based on the intensity distribution of precursor *m/z* in IDA data. MS/MS spectra were accumulated for 40 ms in the mass range *m/z* 350–1500 with rolling collision energy optimized for lower *m/z* in *m/z* window +10%. To minimize sample carryover, blank injections were performed between every sample injection. Additionally, sample data were acquired in a randomized order to avoid batch-effect biases.

2.10 | LC-MS/MS data analysis

The raw data acquired on Q-Exactive were searched against two reference proteome files including the available *G. hirsutum* protein sequences in UniProtKB (79 242 entries, April 2021) and the protein sequences obtained from *G. robinsonii* in the present study (57 023 entries) using MaxQuant 1.6.17.0 for peptide-spectrum matching (Cox & Mann, 2008). The reference proteome of *G. robinsonii* was selected for subsequent proteomics analyses. For ion library generation, data in the IDA mode were processed in ProteinPilot ver 5.0.1 (Sciex) using the Paragon algorithm with default parameters against the reference proteome of *G. robinsonii*. The obtained library was then imported into PeakView (ver. 2.2; Sciex) as a reference ion library and matched against individual SWATH data files. The following matching criteria were applied: top six most intense fragment ions for each peptide, a maximum number of peptides of 100, 75 ppm mass tolerance, 99% peptide confidence threshold, 1% false discovery rate threshold and a 5-min retention time extraction window. The peak areas for peptides were extracted by summing the area under curve values of the corresponding fragment ions using PeakView. The summed peak areas of the peptides were normalized against the total abundance values of respective samples and then used for protein quantification. To assess differentially expressed proteins (DEPs), a comparison of protein abundances across respective sample groups was performed using two-sample Student's *t* tests. Proteins with $p < 0.05$ and an expression fold-change ±1.5 were considered significantly changed between treatments (Wu et al., 2016).

2.11 | Functional enrichment analysis

The DEPs were functionally annotated in OmicsBox (ver. 1.4.1). Using Fisher's exact test, the lists of up- and down-regulated proteins at each condition were compared with the annotated *G. robinsonii* proteome. Significantly enriched gene ontologies (GOs, $p < 0.05$)

were selected for subsequent analysis. To demonstrate the tendency towards increased or decreased abundance under heat stress, the common enriched GOs were displayed in polar plots using GOplot (Walter et al., 2015).

3 | RESULTS

3.1 | Genome sequencing and assembly of short reads

To generate an initial draft genome of *G. robinsonii* (2*n* = 2*x* = 26, CC), DNA isolated from leaves was used for GSS. A total of 133.53 Gb clean reads with sequence coverage of approximately 70-fold were generated from the small-insert (350 bp) library after low-quality reads were filtered. These data were used for 17-mer analysis from which the estimated genome size was 1.91 Gb (Figure 2a and Table 1). Likewise, 17-mer analysis was used to calculate the heterozygous and repeat rates of the genome, which were 0.36% and 84.2%, respectively.

Clean reads were used as input to the SOAPdenovo program to create a de novo assembly. This assembly included a total of 50 428 contigs with an N50 of 6837 bp. These contigs were subsequently assembled, resulting in 48 804 scaffolds with an N50 of 7008 bp. Among all scaffolds, a total of 1 292 294 scaffolds were equal to or longer than 100 bp and 146 656 scaffolds were equal to or longer than 2000 bp. The longest scaffold was 292 647 bp (Table 2). The average GC content of the *G. robinsonii* genome was 36.78% (Figure 2b).

3.2 | Repetitive elements analysis

A total of 1 292 294 scaffolds (≥100 bp) with an overall length of 1 498 229 442 bp were input into RepeatMasker to identify repeats in the assembled genome using Repbase. The results indicated that 54.8% of the assembled genome included repetitive sequences. The most abundant repetitive classes accounted for the long terminal repeat (LTR) elements consisting of 51.52% of the assembled genome, followed by DNA transposons (1.53%), simple repeats (1.06%), low complexity DNA sequences (0.37%), long interspersed nuclear elements (0.33%) and small RNA (0.03%). Gypsy/DIRS1 was the most common repeat among LTR elements, consisting of 47.99%, followed by Ty1/Copia (3.47%).

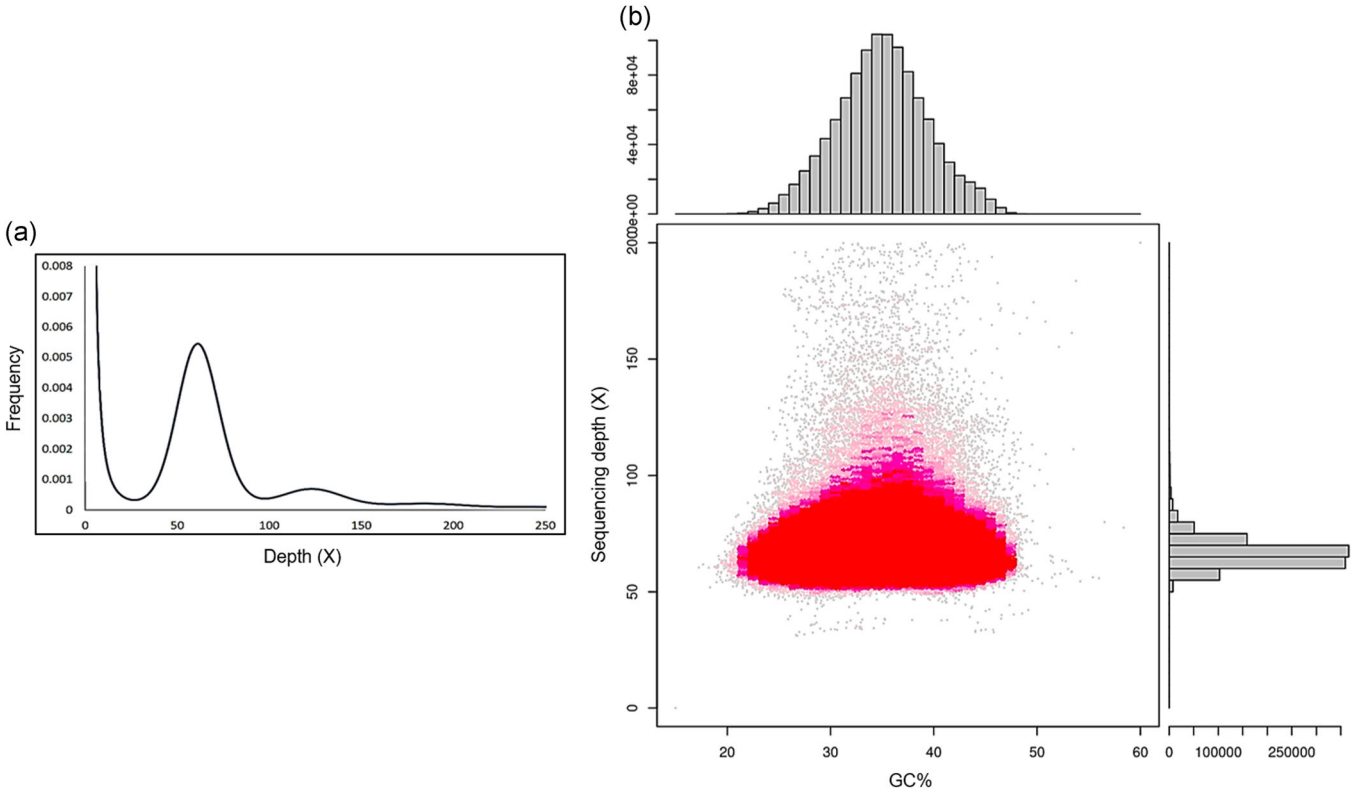


FIGURE 2 Genome characteristics of *Gossypium robinsonii*. (a) *K*-mer (*K* = 17) analysis to estimate the genome size of *Gossypium robinsonii* by the following formula: Genome size = *K*-mer number/peak depth. The *x*-axis is depth, whereas the *y*-axis indicates the proportion that represents the frequency at that depth divided by the total frequency of all depths. (b) The correlation analysis of guanine plus cytosine (GC) content and sequencing depth. The *x*-axis indicates the GC content, whereas the *y*-axis represents the sequence depth. The distribution of sequence depth is shown on the right side, while the distribution of GC content is at the top [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Statistical data from the 17-mer analysis

K	*K*-mer number	Peak depth	Genome size (Mb)	Used bases	Used reads	Coverage
17	116 452 740 243	61	1 909 061 315	13 352 764 170	890 184 278	69.94

TABLE 2 Statistics of de novo assembly results

	Contig		Scaffold	
	Size (bp)	Number	Size (bp)	Number
N50	6837	50 428	7008	48 804
Longest	205 105	-	292 647	-
Total size	1 497 361 583	-	1 498 229 442	-
Total number (≥100 bp)	-	1 310 827	-	1 292 294
Total number (≥2 kb)	-	147 733	-	146 656

3.3 | Confirmation of phylogenies in the genus *Gossypium*

The *ITS* region was used to establish phylogenetic clusters and species relationships for all 44 *Gossypium* species (Table S1). Specifically, this enabled an assessment of the general reliability of the genomic data assembled above for *G. robinsonii* by establishing that it clustered among other Australian species that are believed to form a distinct taxonomic grouping. Cluster C grouped *G. robinsonii* with another C-genome Australian species (*G. sturtianum*) and the G-genome species, *G. australe*, *G. bickii* and *G. nelsonii*. The whole-genus phylogeny revealed six clusters (Figure 3), including cluster A that included 12 Australian species with the K-genome. Surprisingly, the *ITS* gene analysis suggested that *G. robinsonii* was more similar to the G-genome species in cluster C than its C-genome congener, *G. sturtianum*. Moreover, the substitution rate was higher in *G. robinsonii* compared with other Australian species, indicating a high rate of mutation over generations.

3.4 | Gene prediction and functional annotation

Based on the assembled genome of *G. robinsonii* with 242 945 scaffolds (>1000 bp), we predicted a total of 57 023 genes, which were assembled into a reference proteome sequence file. DNA fragments ranged from 102 to 21 615 bp in length with an average of 916.15 bp, all of which were retrieved from only 36 258 scaffolds. The putative genes were searched against UniProtKB and Swiss-Prot databases using BLASTx, resulting in 30 847 (54%) functionally annotated genes.

3.5 | Extreme heat (40°C) during the tetrad stage partially impaired dehiscence

The consequence of heat stress on reproductive success was investigated through monitoring dehiscence in p-TE, p-UN and p-BN. Dehiscence was successful at all stages of development when exposed to 36°C, and p-UN and p-BN underwent complete dehiscence in response to 40°C. Extreme heat, however, resulted in partial failure of dehiscence in p-TE. Of 17 squares heated to 40°C at the tetrad stage, only 53% still underwent complete dehiscence while a further 18% partially dehisced and the remaining 29% were sterile.

3.6 | Proteome profile of *G. robinsonii* for SWATH-MS analyses of pollen

Before SWATH-MS analysis, the reference proteome sequences of *G. hirsutum* and *G. robinsonii* were employed for the peptide to spectrum matching in trial experiments, resulting in the identification of approximately 50% more unique peptides, from the same amount of starting material, when the *G. robinsonii* sequence was used. Hence, the *G. robinsonii* proteome was used for subsequent proteomics analyses. A proteome library was generated by pooling all samples analyzed, including p-TE36/40, p-UN36/40, p-BN36/40 and p-AN. Using the annotated *G. robinsonii* sequence as a reference, we obtained a total of 13 588 peptides representing 2704 proteins from mature pollen (Table S2). Previously, 868, 1103 and 1604 proteins in cotton (Masoomi-Aladizgeh et al., 2020), tomato (Chaturvedi et al., 2013) and tobacco (Ischebeck et al., 2014) dehisced pollen, respectively, were reported using shotgun proteomics. The functional annotations of these proteins, with associated biological process, cellular process and molecular function are shown in Figure 4 (only 30 top enriched categories). This SWATH ion library is available through the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2019) with the data set identifier PXD027097.

3.7 | Protein quantification of developing pollen heated at tetrad, uninucleate and binucleate stages

Using SWATH-MS for quantitative proteomic analysis, we identified 422, 489 and 94 DEPs in p-TE, p-UN and P-BN, respectively,

FIGURE 3 Phylogenetic relationship of *Gossypium robinsonii* and the other 43 cotton species using internal transcribed spacer (ITS) regions (ITS1, 5.8S ribosomal RNA and ITS2). The phylogenetic tree was constructed using MEGA X software with maximum likelihood inference. A general time-reversible model with gamma-distributed rate variation across five categories and invariant sites was used for analyses (Kumar et al., 2018). The ITS sequences were downloaded from NCBI except for *G. robinsonii* for which the gene was retrieved from the present study. The geographic location of the species was obtained from the CottonGen database (http://www.cottongen.org). The branch lengths are drawn to scale with the bar, which indicates 0.01 nucleotide substitutions per site [Color figure can be viewed at wileyonlinelibrary.com]

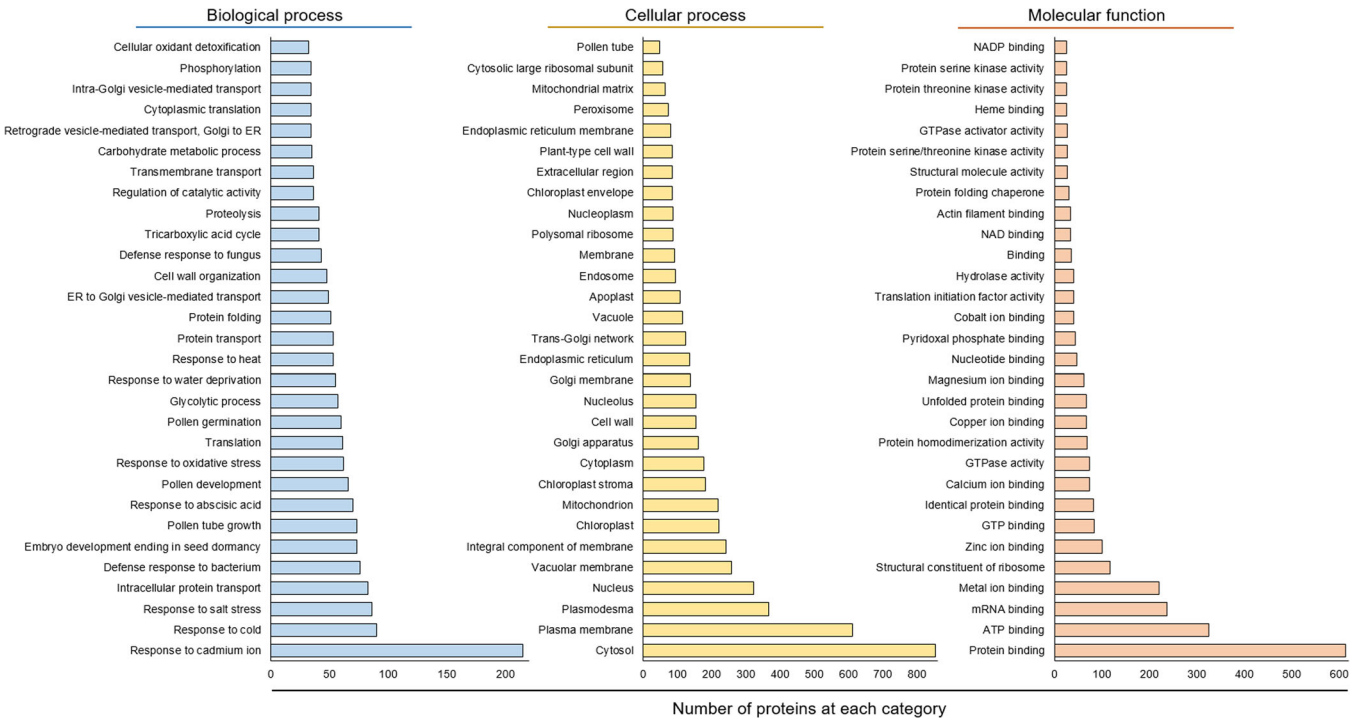


FIGURE 4 Functional gene ontology (GO) annotation of 2704 proteins identified in pollen of *Gossypium robinsonii* into 30 narrower categories that belong to biological process, cellular process and molecular function. The *x*-axis shows the number of proteins, while the *y*-axis represents functional categories in which proteins are classified according to their abundance [Color figure can be viewed at wileyonlinelibrary.com]

after exposure to 36°C (moderate heat), compared with control pollen. Extreme heat (40°C) led to the identification of only 297, 154 and 61 DEPs in p-TE40, p-UN40 and p-BN40, respectively (Table S3), indicating a diminished response compared with the 36°C treatment, and as above, a declining translational response when the heat was imposed in late stages of pollen development. To increase accuracy in our functional analysis, we considered a stringent fold-change threshold (±1.5) to filter differentially expressed proteins. For example, a 1.5-fold cutoff reduced the number of DEPs between 79% and 255% compared with the result had we used a onefold threshold. Differentially abundant proteins were visualized in an UpSet plot (Khan & Mathelier, 2017), highlighting that 196 DEPs were uniquely regulated in p-TE after exposure to extreme heat, whereas only 55 and 28 DEPs were unique to p-UN and p-BN, respectively, under the same conditions (Figure 5a,b). A total of 36 DEPs were common between the stages when plants were exposed to 36°C, while only 11 DEPs co-occurred in all stages under extreme heat (Table 3). Significantly, we found that the number of DEPs in p-BN36 and p-BN40 decreased compared with p-TE and p-UN exposed to elevated temperature, in spite of all samples being derived from mature pollen (Figure 5c,d). Most notably, extreme heat caused changes in the abundance of a large number of proteins in tetrads when compared with the later stages of development, which can be attributed to its distinctive response to high temperature and potentially its vulnerability to temperature stress.

3.8 | Functional annotation of p-TE, p-UN and p-BN in response to extreme heat

The top 20 most enriched GO categories in the biological process (Fisher's exact test, $p < 0.05$) were plotted for DEPs under extreme heat (Figure 6). A detailed list of GO enrichment analyses is available in Table S4. Enrichment analysis demonstrated that extreme heat caused an increased abundance of proteins involved in messenger RNA (mRNA) splicing and translation in p-TE40. Heat shock cognate 71 kDa protein (HSPA8/HSC70; Gr_Sca1204283G34) and glycine-rich RNA-binding protein RZ1A (RZ1A; Gr_Sca97552G39) were involved in mRNA splicing, whereas translation initiation factor 1A (Eif1a; Gr_Sca410536G29) and eukaryotic translation initiation factor 3 subunit D (Eif3d; Gr_Sca18748G18) were associated with cytoplasmic translation. All these key proteins increased in abundance after exposure of p-TE to 40°C. In addition, extreme heat increased the abundance of HSPA8/HSC70, 17.9 kDa class II heat shock protein (Gr_Sca284404G23) and cytosolic glyceraldehyde-3-phosphate dehydrogenase (Gr_Sca544237G31) in the response to hydrogen peroxide category in p-TE40 (Figure 6a). On the other hand, those

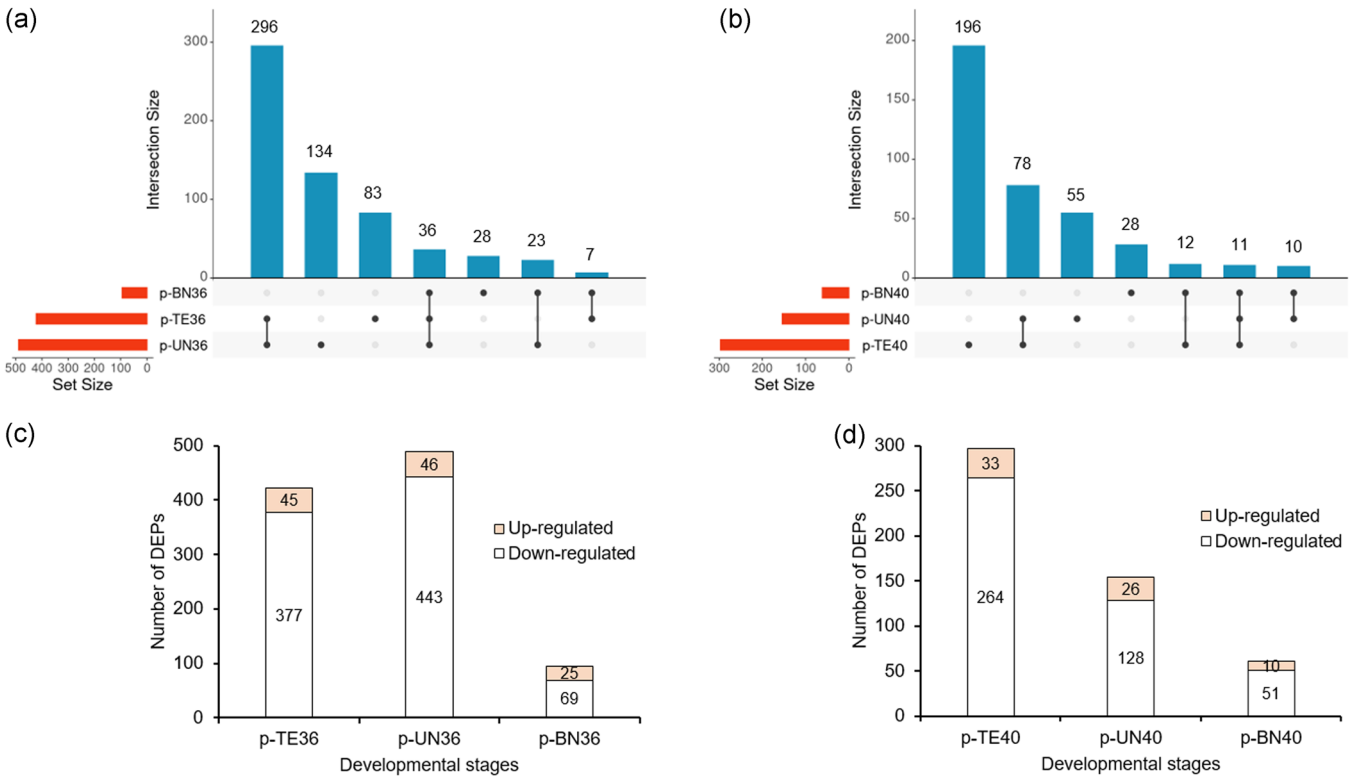


FIGURE 5 Differentially regulated proteins in mature pollen corresponding to tetrad cells (p-TE), uninucleate (p-UN) and binucleate (p-BN) microspores after exposure to 36°C and 40°C for 5 days. (a and b) Upset plot illustrating the number of DEPs that are unique to each developmental stage (p-TE, p-UN and p-BN) after 5-day exposure to 36°C and 40°C, respectively (c and d) Number of up- and down-regulated proteins in p-TE, p-UN and p-BN stage when exposed to 36°C and 40°C, respectively [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 DEPs that were common between p-TE, p-UN and p-BN when plants were exposed to 36°C or 40°C

Protein ID	Description	p-TE36	p-UN36	p-BN36
Gr_Sca89800G8	Uncharacterized protein	43.54	39.91	50.07
Gr_Sca244142G37	V-type proton ATPase subunit G 2	1.83	1.73	1.50
Gr_Sca142526G37	Late embryogenesis abundant protein 1	1.73	1.88	1.69
Gr_Sca75553G29	Protein kinase G11A	1.70	1.69	1.51
Gr_Sca10830G29	Carbonyl reductase [NADPH] 1	1.63	−1.63	−1.63
Gr_Sca399437G33	Uncharacterized protein	1.52	1.50	1.57
Gr_Sca324712G22	Mitochondrial 2-oxoglutarate/malate carrier protein	−1.66	−1.97	−1.56
Gr_Sca3826G9	Full = *O*-fucosyltransferase 7	−1.70	−2.16	−1.56
Gr_Sca1122997G3	Protein AGENET DOMAIN (AGD)-CONTAINING P1	−1.70	−2.11	−1.76
Gr_Sca282494G21	Lysine histidine transporter 1	−1.75	−1.93	−1.54
Gr_Sca384799G1	*Trans*-resveratrol di-*O*-methyltransferase	−1.75	−1.91	−1.63
Gr_Sca23941G38	Solute carrier family 35 member E3	−1.87	−2.16	−1.55
Gr_Sca1258128G42	GDSL esterase/lipase	−1.90	−1.88	−1.55
Gr_Sca191955G1	Fasciclin-like arabinogalactan protein 14	−1.94	−1.96	−1.59
Gr_Sca550796G34	Peroxygenase	−1.96	−2.00	−1.69
Gr_Sca264066G27	Peroxidase 18	−1.98	−2.54	−1.71
Gr_Sca287154G30	Ras-related protein Rab-8B	−2.05	−2.28	−1.71
Gr_Sca107215G8	40S ribosomal protein S15a	−2.05	−1.83	−1.54
Gr_Sca638953G4	Uncharacterized protein	−2.06	−2.08	−2.31
Gr_Sca241703G9	Calcium-transporting ATPase 9, plasma membrane-type	−2.06	−2.61	−1.58
Gr_Sca268785G33	Cytidine deaminase	−2.17	−2.69	−1.51
Gr_Sca262619G32	Small COPII coat GTPase SAR1	−2.21	−2.17	−1.50
Gr_Sca1095075G13	Adenosylhomocysteinase	−2.22	−1.89	−1.64
Gr_Sca209002G25	Probable 26S proteasome non-ATPase regulatory subunit 3	−2.29	−2.98	−1.65
Gr_Sca146828G27	UDP-glucose 4-epimerase	−2.30	−2.52	−1.62
Gr_Sca131039G13	UDP-glucose 4-epimerase	−2.32	−2.31	−1.75
Gr_Sca238340G3	Fasciclin-like arabinogalactan protein 21	−2.73	−5.13	−3.00
Gr_Sca70333G21	Protein transport protein Sec. 61 subunit beta	−2.87	−3.05	−1.67
Gr_Sca171436G36	Callose synthase 10	−3.17	−3.65	−2.07
Gr_Sca557747G34	Glycogen phosphorylase, muscle form	−3.21	−3.88	−2.01
Gr_Sca23526G4	Exocyst complex component 7	−3.59	−4.35	−1.80
Gr_Sca102497G32	Sodium/calcium exchanger NCL	−3.86	−3.05	−2.04
Gr_Sca94813G23	Phospholipase ABHD3	−4.73	−5.24	−1.87
Gr_Sca572213G29	Coatomer subunit beta	−6.57	−5.25	−1.56
Gr_Sca142153G39	Fasciclin-like arabinogalactan protein 14	−8.43	−7.71	−2.07
Gr_Sca52928G8	1-3-beta-glucan endohydrolase	−8.81	−8.66	−7.23
Protein ID	**Description**	**p-TE40**	**p-UN40**	**p-BN40**
Gr_Sca1122637G16	Isocitrate dehydrogenase [NADP]	2.22	2.07	1.91
Gr_Sca248462G23	14-3-3 protein epsilon	2.00	1.97	2.02

TABLE 3 (Continued)

Protein ID	Description	p-TE40	p-UN40	p-BN40
Gr_Sca284404G23	17.9 kDa class II heat shock protein	1.78	1.53	1.62
Gr_Sca689210G35	Omega-3 fatty acid desaturase, chloroplastic	−1.62	−1.50	−1.53
Gr_Sca191955G1	Fasciclin-like arabinogalactan protein 14	−1.65	−2.53	−1.60
Gr_Sca550796G34	Peroxygenase	−2.04	−1.61	−1.66
Gr_Sca101458G58	CAAX prenyl protease 1 homolog	−2.49	−2.15	−2.11
Gr_Sca23526G4	Exocyst complex component 7	−3.04	−3.67	−1.79
Gr_Sca253694G25	Uridine kinase	−3.75	−3.44	−3.23
Gr_Sca94813G23	Phospholipase ABHD3	−4.35	−5.53	−1.72
Gr_Sca142153G39	Fasciclin-like arabinogalactan protein 14	−8.30	−7.26	−1.98

Abbreviations: BN, binucleate microspores; TE, tetrad cells; UN, uninucleate microspores.

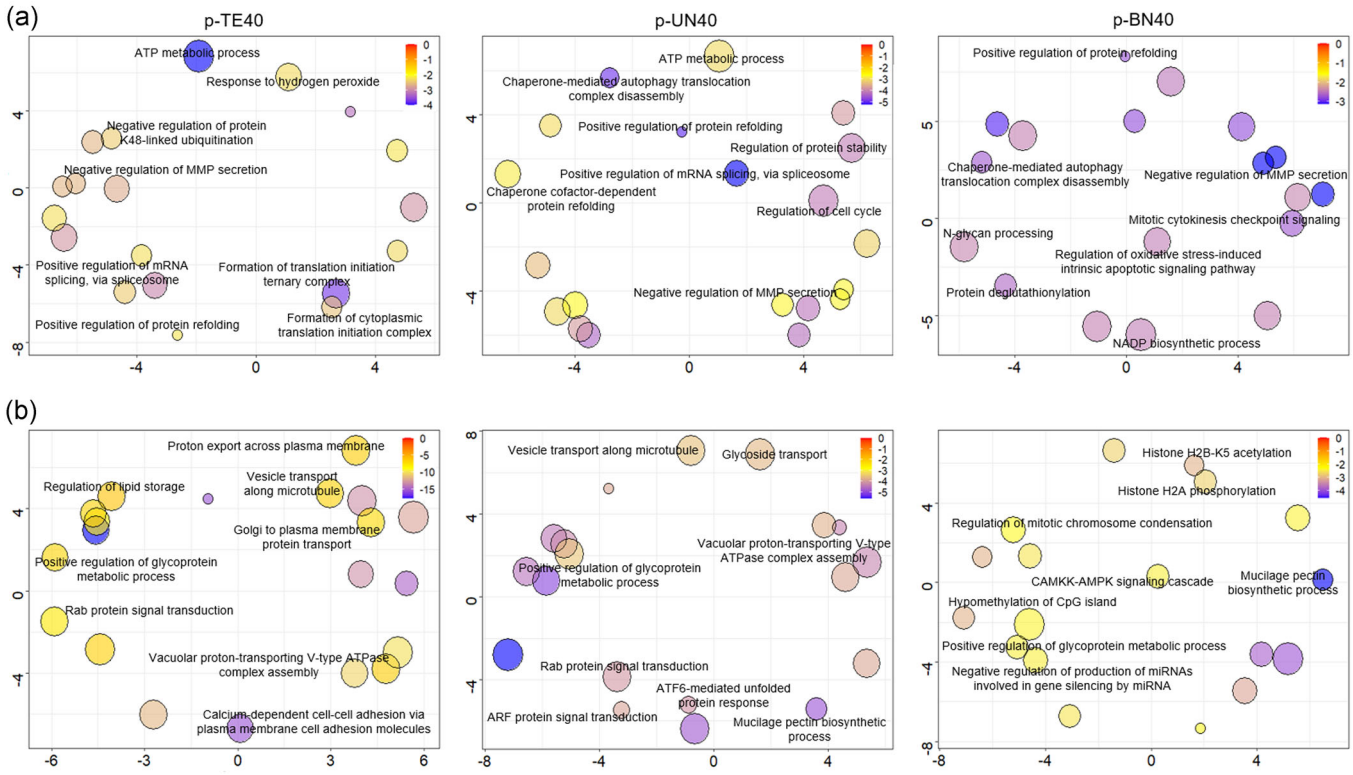


FIGURE 6 Most enriched categories ($p < 0.05$) in biological process up-regulated (a) and down-regulated (b) in TE, p-UN and p-BN in response to extreme heat. Each bubble represents a significantly enriched term in a two-dimensional space (Supek et al., 2011), with a size that is proportional to the recurrence of the gene ontology (GO) term in the whole Uniprot database. The scale indicates the log 10 p value, with red and blue representing the larger and smaller p value, respectively. Of the 20 most enriched GOs, only key categories were labelled in each plot. BN, binucleate microspores; TE, tetrad cells; UN, uninucleate microspores [Color figure can be viewed at wileyonlinelibrary.com]

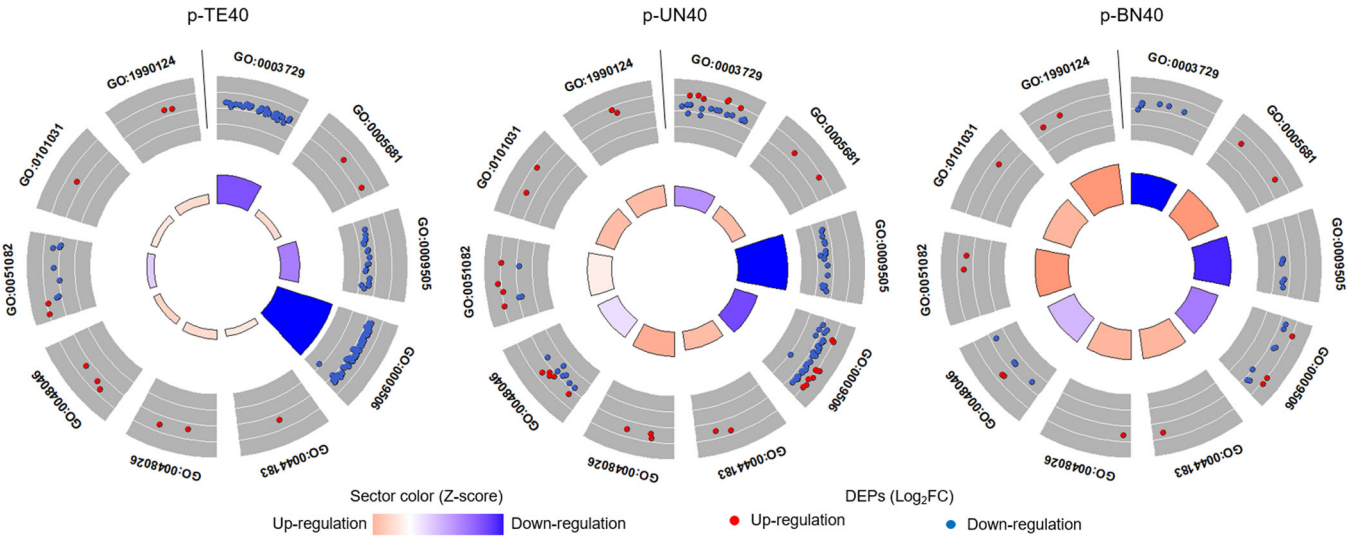
functions mainly contributing to transport, such as proton export across the plasma membrane, vesicle transport along microtubules, and Golgi to plasma membrane protein transport, were inhibited in response to extreme heat in p-TE. Importantly, Rab proteins (GTPases), which are key regulators of intracellular vesicle protein transport (Bhuin & Roy, 2014), decreased in abundance in p-TE40, indicating that intracellular transport might be down-regulated as part of the response of p-TE to extreme heat. These proteins included Ras-related protein Rab-2A (Gr_Sca778179G17 and Gr_Sca17313G6), Ras-related protein Rab-1B (Gr_Sca3776G55, Gr_Sca898430G34 and Gr_Sca790589G39) and Ras-related protein Rab-18 (Gr_Sca270469G20) (Figure 6b).

In p-UN, extreme heat increased the abundance of two isoforms of heat shock cognate 71 kDa protein (Gr_Sca1204283G34 and Gr_Sca438049G25) and Gr_Sca97552G39, which play a role in mRNA splicing and protein folding. Ankyrin-3 (Gr_Sca258696G41), 40S ribosomal protein S8 (Gr_Sca354645G21), 14-3-3 protein epsilon (Gr_Sca248462G23) and NEDD8-activating enzyme E1 catalytic subunit (Gr_Sca161791G12) also increased in abundance in p-UN40, all of which were involved in the regulation of cell cycle (Figure 6a). As for p-TE40, proteins associated with transport were inhibited in p-UN after exposure to extreme heat. Rab proteins including Gr_Sca270469G20, Gr_Sca17313G6 and Gr_Sca790589G39 involved in transport were also down-regulated in p-UN40. Moreover, ADP-ribosylation factor 1 isoforms (Gr_Sca53462G15, Gr_Sca93217G3 and Gr_Sca18109G34), which are associated with ARF protein signal transduction and intracellular protein trafficking decreased in abundance in p-UN40 (Figure 6b).

In p-BN, Gr_Sca438049G25 and glucosidase 2 subunit beta (Gr_Sca421233G26) were up-regulated in response to 40°C, both being involved in protein refolding. NADP dependent isocitrate dehydrogenase (Gr_Sca1122637G16), protein disulphide-isomerase (Gr_Sca49083G17) and Gr_Sca248462G23, which were enriched in NADP biosynthetic process, protein deglutathionylation and mitotic cytokinesis checkpoint signalling, respectively, also increased in abundance in p-BN40 (Figure 6a). Extreme heat, however, decreased the abundance of CAAX prenyl protease 1 homolog (Gr_Sca101458G58), which contributes to histone H2B-K5 acetylation, hypomethylation of CpG island, histone H2A phosphorylation and the CAMKK-AMPK signalling cascade. Additionally, proteins related to the regulation of gene expression decreased in abundance in p-BN after exposure to extreme heat, including 60S ribosomal protein L5 (Gr_Sca687372G12), cold-inducible RNA-binding protein (Gr_Sca144232G10), SRSF protein kinase 1 (Gr_Sca381564G50), Gr_Sca101458G58 and programmed cell death protein 5 (Gr_Sca326285G39) (Figure 6b).

3.9 | p-TE decreased protein transport, plant cell wall synthesis and m-RNA binding after severe heat

Commonly enriched GOs across the pollen developmental stages (p-TE, p-UN and p-BN) were plotted to highlight the effect of extreme heat on the expression pattern of each category (Figure 7). The data indicated a notable decrease in abundance in mRNA binding (GO: 0003729), plant-type cell wall (GO: 0009505) and plasmodesma (GO: 0009506) after exposure of TE and UN to 40°C. Importantly, all proteins belonging to p-TE40 enriched in these categories decreased in abundance, indicating the importance of reduced protein expression in responding to extreme heat in the vulnerable stage of pollen development. Transmembrane



ID	GO term	ID	GO term
GO:0003729	mRNA binding	GO:0048026	positive regulation of mRNA splicing, via spliceosome
GO:0005681	spliceosomal complex	GO:0048046	apoplast
GO:0009505	plant-type cell wall	GO:0051082	unfolded protein binding
GO:0009506	plasmodesma	GO:0101031	chaperone complex
GO:0044183	Protein folding chaperone	GO:1990124	messenger ribonucleoprotein complex

FIGURE 7 Expression patterns of ten commonly enriched gene ontologies (GOs) ($p < 0.05$) shared among three stages of pollen development (p-TE, p-UN and p-BN) after exposure to 40°C for 5 days. The outer ring shows the log 10 fold-change of up- and down-regulated proteins enriched in a specific GO category. The sector size in the inner ring is proportional to the statistical significance (adjusted p value) of each GO category, whereas the colour shows the tendency of the GO to increase or decrease in abundance in response to extreme heat according to its z-score, following this formula: ($z = \frac{\text{up - regulated} - \text{down - regulated}}{\sqrt{\text{total DEPs}}}$). BN, binucleate microspores; DEPs, differentially expressed proteins; TE, tetrad cells; UN, uninucleate microspores [Color figure can be viewed at wileyonlinelibrary.com]

9 superfamily member 1 (Gr_Sca1268022G12, Gr_Sca610067G50, Gr_Sca302002G49 and Gr_Sca117972G35) and member 4 (Gr_Sca501547G17, Gr_Sca205722G14, Gr_Sca610558G59, Gr_Sca11-31572G34 and Gr_Sca69313G26) associated with plasmodesma and plant-type cell wall categories were less abundant in p-TE40 in *G. robinsonii*. Ribosomal proteins such as 40S ribosomal proteins S3 (Gr_Sca61122G18), S4 (Gr_Sca627331G15) and S9 (Gr_Sca85229G45), as well as 60S ribosomal proteins L3 (Gr_Sca997729G33), L6 (Gr_Sca873671G24), L7 (Gr_Sca847687G38, Gr_Sca70333G25 and Gr_Sca75178G19), L18 (Gr_Sca49010G31) and L19 (Gr_Sca43243G17) were involved in plasmodesma and mRNA binding functions, all of which decreased in abundance in P-TE40. The GO category of unfolded protein binding (GO: 0051082) was also mainly down-regulated in p-TE after exposure of *G. robinsonii* to 40°C. Proteins involved in this functional category became less abundant in response to heat, including calreticulin (Gr_Sca87704G38 and Gr_Sca296822G15), endoplasmin (Gr_Sca229-620G7) and UDP-glucose:glycoprotein glucosyltransferase (UGGT; Gr_Sca121716G5, Gr_Sca121716G6 and Gr_Sca121716G7).

3.10 | p-TE increased the activity of molecular chaperones in response to severe heat

The data obtained from commonly enriched GOs across p-TE, p-UN and p-BN (Figure 7) indicated that spliceosome complex (GO: 0005681), protein folding chaperone (GO: 0044183), positive regulation of mRNA splicing, via spliceosome (GO: 0048026), chaperone complex (GO: 0101031) and messenger ribonucleoprotein complex (GO: 1990124) increased in p-TE as well as other stages in response to heat stress. Polyadenylate-binding protein (Gr_Sca18981G14), RZ1A and HSPA8/HSC70 were associated with the above-mentioned functional categories; these proteins increased significantly in abundance in tetrads (p-TE) after exposure to severe heat.

4 | DISCUSSION

Heat stress during male gametophyte development results in dramatically reduced fruit and grain production across a wide range of crops (Jagadish, 2020). Our recent investigation of a commercially cultivated cotton (Sicot 71) revealed that the meiotic stage of pollen development leading to tetrad cells is extremely vulnerable to 40°C (Masoomi-Aladizgeh et al., 2020). We also speculated that up-regulation of non-essential pathways might exacerbate sensitivity of tetrad cells to heat stress in cotton. By inference, the progress of tetrad cells to dehiscent pollen would be promoted if most pathways were down-regulated and only essential proteins were actively translated as an adaptive response to extreme heat (Masoomi-Aladizgeh et al., 2021). With this hypothesis in mind, we turned to *G. robinsonii*, a wild cotton species from the arid zone of north-western Australia, to discover whether this heat-adapted wild relative has evolved protein expression patterns during meiosis (tetrad cells) that are qualitatively distinct from what is seen in commercial cotton exposed to heat.

4.1 | Higher mRNA splicing in mature pollen corresponding to tetrads exposed to 40°C

A higher requirement for expression of heat-responsive genes has been reported when plants are exposed to abiotic stress; however, defects in mRNA splicing of these genes may occur due to insufficient splicing machinery (Cui & Xiong, 2015). Moreover, increasing evidence points to the importance of splicing regulators as key plant stress mediators and suggests that any defects in splicing factors would impair plant response to environmental fluctuations (Laloum et al., 2018). Defective splicing machinery could be the reason for the accumulation of unprocessed HSP70 in maize pollen in response to 45°C, as reported previously (Hopf et al., 1992). Masoomi-Aladizgeh et al. (2021) also attributed the sensitivity of tetrad cells to potential defects in splicing machinery after exposure of cotton pollen to 38°C. They found that glycine-rich RNA-binding proteins, involved in pre-mRNA splicing, were down-regulated after tetrads were exposed to 38°C in *G. hirsutum*. This was also evident in mature pollen in response to elevated temperature, reflecting potential failure in spliceosome activity. In wild cotton (*G. robinsonii*), however, the functional analysis demonstrated that the abundance of proteins associated with mRNA splicing increased in response to 40°C. For instance, exposure of wild cotton to 40°C during the tetrad development (p-TE40) increased the abundance of RZ1A and HSPA8/HSC70, indicating a possible role of these proteins in splicing efficacy under heat stress. This may be ascribed to greater thermotolerance of tetrads in the wild cotton relative, enabling the development of viable mature pollen at daytime temperatures that render commercial cotton flowers sterile.

HSPA8/HSC70, a constitutively expressed chaperone protein, is known to be associated with many cellular processes under normal and stress conditions (Liu et al., 2012; Stricher et al., 2013). Glycine-rich RNA-binding proteins in *Arabidopsis* are involved in negative autoregulation at the posttranscriptional level, by which mRNA containing premature stop codons are degraded via the nonsense-mediated mRNA decay pathway (Schöning et al., 2008). The up-regulation of RZ1A in *G. robinsonii* may be also involved in this pathway to improve splicing efficacy, thereby mitigating heat stress. Proteins such as HSPA8/HSC70 and RZ1A, which we have identified in wild cotton, are potential candidates for breeding improved thermotolerance in cultivated cotton.

4.2 | Lower translation in mature pollen arising from tetrads exposed to 40°C

Downshift of global protein synthesis is an adaptive response to external stimuli, notwithstanding the strategic expression of key genes that is required for acclimation. The majority of differentially regulated proteins was down-regulated in p-TE (89%) in response to 40°C, indicating that the repression of translation plays a part in thermotolerance. A qualitatively different pattern of translation was reported by Masoomi-Aladizgeh et al. (2021) in cultivated cotton

where translation was higher in response to 38°C in tetrads rather than being suppressed. For example, calreticulin, endoplasmin and UGGT involved in the unfolded protein binding were down-regulated in p-TE40 in *G. robinsonii*. These proteins, however, were up-regulated in *G. hirsutum* when tetrad cells were exposed to 38°C (Masoomi-Aladizgeh et al., 2021). Moreover, proteins such as 40S ribosomal proteins S3 and S9 as well as 60S ribosomal proteins L3, L6 and L18 were down-regulated in p-TE40 in *G. robinsonii*, while such proteins became abundant in *G. hirsutum* in tetrad cells after heat (Masoomi-Aladizgeh et al., 2021). This might reflect the decreasing activity of translational processes in *G. robinsonii* to combat the stress. Even though proteomics was conducted on samples collected at two different stages of pollen development in these two studies, namely, tetrads and mature pollen, tetrad cells were specifically exposed to heat stress in both experiments. Thus, in both cases, proteomics reflects molecular responses of tetrads to heat.

The use of the wild cotton relative indicates that a strategic response to heat is required for thermotolerance, with only key genes translated upon stress. For instance, we identified Eif1a and Eif3d up-regulated in p-TE40, both of which are associated with translation initiation. Eif1a is involved in 'ribosome scanning' to initiate translation and is reported to increase tolerance to NaCl and osmotic stress in plants (Rausell et al., 2003; Yang et al., 2017). Eif3d also plays a key role in stress-induced translation and is involved in a mechanism underlying the selective translation of essential genes required for cell survival (Lamper et al., 2020). Our finding is supported by Peredo and Cardon (2020), who highlighted that desiccation-tolerant algae massively down-regulated cellular metabolism, unlike the sensitive taxon, suggesting an energy-saving strategy under stress. Likewise, a study on grapes (*Vitis vinifera*) indicated that the abundance of proteins involved in the metabolic process decreased in response to 42°C (George et al., 2015). Repression of general protein synthesis in the meiotic stage of pollen development under extreme heat accounts for thermotolerance in the wild cotton relative, providing that translation of essential genes is maintained.

4.3 | Transport decreased substantially in mature pollen resulting from tetrads exposed to 40°C

Protein transport was reduced substantially when p-TE was exposed to extreme heat. In our previous study on cultivated cotton, we found that proteins associated with plasmodesma (GO:0009506) were highly up-regulated in tetrad cells in response to 38°C, implying a high level of cell–cell communication that might lead to consumption of scarce resources, and consequently the vulnerability of tetrad cells (Masoomi-Aladizgeh et al., 2021). In sharp contrast, *G. robinsonii* suppressed cell–cell communication in p-TE40, demonstrated by down-regulation of proteins associated with plasmodesma. For instance, transmembrane 9 superfamily member proteins were up-regulated in tetrads when *G. hirsutum* was subjected to high temperature (Masoomi-Aladizgeh et al., 2021), whereas transmembrane 9 superfamily member 1 and member 4 were down-regulated in p-TE40 in *G. robinsonii*. Importantly, Rab proteins regulating intercellular transport were key proteins down-regulated in p-TE40. This family of proteins is associated with a wide range of roles in plants, such as intracellular trafficking and responses to abiotic stress (Nielsen, 2020; Tripathy et al., 2021). For instance, it has been reported that mutations in NtRab2 blocked the localization of Golgi-resident, plasmalemma and secreted proteins, and also inhibited pollen tube growth in tobacco plants (Cheung et al., 2002). This accords with our suggestion that the down-regulation of Rab-2A, Rab-1B and Rab-18 led to inhibition of intercellular transport in p-TE40. It appears that inhibition of protein transport is required for tetrads of cotton to tolerate high temperature and that Rab proteins, along with other key proteins involved in transport, could be potential tools in biotechnological approaches to improve heat tolerance.

5 | CONCLUSION

In this study, we generated an initial draft genome of *G. robinsonii*, an Australian wild cotton, and predicted a total of 57 023 genes that were assembled into a reference proteome sequence file. We found that GSS analysis of this obscure wild relative enabled us to identify 50% more unique peptides than would be achieved using sequences obtained from the cultivated species. Proteomic analysis of the wild cotton indicated that the thermotolerance of tetrads may be related to the repression of translation of nonessential pathways, which resulted in decreased abundance of proteins in response to 40°C. Highly specific functions such as splicing were maintained as a survival mechanism. In addition, *G. robinsonii* suppressed cell–cell communication, which may conserve scarce energy resources that can be directed to essential processes throughout development. The application of mechanisms that are unique to this arid-zone wild species to cultivated cotton could increase the pollen thermotolerance, particularly through genetic intervention in meiotic processes.

ACKNOWLEDGEMENTS

Farhad Masoomi-Aladizgeh acknowledges support from Macquarie University in the form of iMQRES and MQRES scholarships. Farhad Masoomi-Aladizgeh acknowledges support from BioBam team (OmicsBox) in the form of a student scholarship for functional analysis of genes and proteins. Support from the Joyce W. Vickery Scientific Research Fund from the Linnean Society of NSW and the Val Williams Scholarship in Botany from the Australian Plants Society (NSW) was seminal in initiating DNA sequencing and analysis. We appreciate the use of the Plant Growth Facility (PGF), Australian Proteome Analysis Facility (APAF) and Macquarie University Faculty of Science and Engineering Microscope Facility (MQFoSE MF) and their valuable support. We also acknowledge the assistance of Yasmin Asar with phylogenetic analysis. Open access publishing facilitated by Macquarie University, as part of the Wiley - Macquarie University agreement via the Council of Australian University Librarians.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Farhad Masoomi-Aladizgeh and Brian J. Atwell conceived the original research plan. Farhad Masoomi-Aladizgeh cultivated and treated the plants, collected samples, performed the genomics and proteomics experiments, interpreted the results and prepared the original draft of the manuscript. Karthik Shantharam Kamath assisted with SWATH-MS and proteomic data analysis. Paul A. Haynes and Brian J. Atwell contributed to data interpretation and preparation of the final version of the manuscript. All contributed to the revision of the manuscript.

DATA AVAILABILITY STATEMENT

The proteomics data were deposited in PRIDE (https://www.ebi.ac.uk/pride) under the accession number PXD027097. The genome sequence reads obtained by BGISEQ-500 are available in the NCBI Sequence Read Archive (SRA) at https://www.ncbi.nlm.nih.gov/sra. The Bioproject accession number is PRJNA592601 and the Biosample accession number is SAMN17125411.

ORCID

Farhad Masoomi-Aladizgeh https://orcid.org/0000-0003-4270-7461
Karthik Shantharam Kamath https://orcid.org/0000-0002-5641-0709
Paul A. Haynes https://orcid.org/0000-0003-1472-8249
Brian J. Atwell http://orcid.org/0000-0002-0466-1688

REFERENCES

Bao, W., Kojima, K.K. & Kohany, O. (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6, 11.

Begcy, K., Nosenko, T., Zhou, L.Z., Fragner, L., Weckwerth, W. & Dresselhaus, T. (2019) Male sterility in maize after transient heat stress during the tetrad stage of pollen development. *Plant Physiology*, 181, 683–700.

Bhuin, T. & Roy, J.K. (2014) Rab proteins: the key regulators of intracellular vesicle transport. *Experimental Cell Research*, 328, 1–19.

Chaturvedi, P., Ischebeck, T., Egelhofer, V., Lichtscheidl, I. & Weckwerth, W. (2013) Cell-specific analysis of the tomato pollen proteome from pollen mother cell to mature pollen provides evidence for developmental priming. *Journal of Proteome Research*, 12, 4892–4903.

Cheung, A.Y., Chen, C.Y.H., Glaven, R.H., De Graaf, B.H.J., Vidali, L., Hepler, P.K. et al. (2002) Rab2 GTPase regulates vesicle trafficking between the endoplasmic reticulum and the Golgi bodies and is important to pollen tube growth. *Plant Cell*, 14, 945–962.

Cox, J. & Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26, 1367–1372.

Cui, P. & Xiong, L. (2015) Environmental stress and pre-mRNA splicing. *Molecular Plant*, 8, 1302–1303.

Endo, M., Tsuchiya, T., Hamada, K., Kawamura, S., Yano, K., Ohshima, M. et al. (2009) High temperatures cause male sterility in rice plants with transcriptional alterations during pollen development. *Plant & Cell Physiology*, 50, 1911–1922.

George, I.S., Pascovici, D., Mirzaei, M. & Haynes, P.A. (2015) Quantitative proteomic analysis of cabernet sauvignon grape cells exposed to thermal stresses reveals alterations in sugar and phenylpropanoid metabolism. *Proteomics*, 15, 3048–3060.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J. et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36, 3420–3435.

Hamzelou, S., Kamath, K.S., Masoomi-Aladizgeh, F., Johnsen, M.M., Atwell, B.J. & Haynes, P.A. (2020) Wild and cultivated species of rice have distinctive proteomic responses to drought. *International Journal of Molecular Sciences*, 21, 5980.

Hamzelou, S., Pascovici, D., Kamath, K.S., Amirkhani, A., McKay, M., Mirzaei, M. et al. (2020) Proteomic responses to drought vary widely among eight diverse genotypes of rice (Oryza sativa). *International Journal of Molecular Sciences*, 21, 363.

Hoff, K.J. & Stanke, M. (2013) WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research*, 41, W123–W128.

Hopf, N., Plesofsky-Vig, N. & Brambl, R. (1992) The heat shock response of pollen and other tissues of maize. *Plant Molecular Biology*, 19, 623–630.

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W. et al. (2016) The Dfam database of repetitive DNA families. *Nucleic Acids Research*, 44, D81–D89.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D. et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Research*, 37, D211–D215.

IPCC. (2021) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S. et al. (Eds.). Cambridge University Press. In Press.

Ischebeck, T., Valledor, L., Lyon, D., Gingl, S., Nagler, M., Meijón, M. et al. (2014) Comprehensive cell-specific protein analysis in early and late pollen development from diploid microsporocytes to pollen tube growth. *Molecular & Cellular Proteomics*, 13, 295–310.

Jagadish, S.V.K. (2020) Heat stress during flowering in cereals—effects and adaptation strategies. *New Phytologist*, 226, 1567–1572.

Khan, A. & Mathelier, A. (2017) Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics*, 18, 287.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547–1549.

Laloum, T., Martín, G. & Duque, P. (2018) Alternative splicing control of abiotic stress responses. *Trends in Plant Science*, 23, 140–150.

Lamper, A.M., Fleming, R.H., Ladd, K.M. & Lee, A.S.Y. (2020) A phosphorylation-regulated eIF3d translation switch mediates cellular adaptation to metabolic stress. *Science*, 370, 853–856.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J. et al. (2010) The sequence and de novo assembly of the giant panda genome. *Nature*, 463, 311–317.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z. et al. (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, 20, 265–272.

Liu, T., Daniels, C.K. & Cao, S. (2012) Comprehensive review on the HSC70 functions, interactions with related molecules and involvement in clinical diseases and therapeutic potential. *Pharmacology & Therapeutics*, 136, 354–374.

Mammadov, J., Buyyarapu, R., Guttikonda, S.K., Parliament, K., Abdurakhmonov, I.Y. & Kumpatla, S.P. (2018) Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. *Frontiers in Plant Science*, 9, 886.

Masoomi-Aladizgeh, F., Jabbari, L., Khayam Nekouei, R. & Aalami, A. (2016) A simple and rapid system for DNA and RNA isolation from diverse plants using handmade kit. *Protocol Exchange*. Available from: https://doi.org/10.21203/rs.2.1347/v2

Masoomi-Aladizgeh, F., Mckay, M.J., Asar, Y., Haynes, P.A. & Atwell, B.J. (2021) Patterns of gene expression in pollen of cotton (*Gossypium hirsutum*) indicate down-regulation as a feature of thermotolerance. *The Plant Journal*. Available from: https://doi.org/10.1111/tpj.15608

Masoomi-Aladizgeh, F., Najeeb, U., Hamzelou, S., Pascovici, D., Amirkhani, A., Tan, D.K.Y. et al. (2020) Pollen development in cotton (*Gossypium hirsutum*) is highly sensitive to heat exposure during the tetrad stage. *Plant Cell & Environment*, 44, 2150–2166.

Müller, F. & Rieu, I. (2016) Acclimation to high temperature during pollen development. *Plant Reproduction*, 29, 107–118.

Nielsen, E. (2020) The small GTPase superfamily in plants: a conserved regulatory module with novel functions. *The Annual Review of Plant Biology*, 71, 247–272.

Peredo, E.L. & Cardon, Z.G. (2020) Shared up-regulation and contrasting down-regulation of gene expression distinguish desiccation-tolerant from intolerant green algae. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 17438–17445.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J. et al. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research*, 47, D442–D450.

Rausell, A., Kanhonou, R., Yenush, L., Serrano, R. & Ros, R. (2003) The translation initiation factor eIF1A is an important determinant in the tolerance to NaCl stress in yeast and plants. *The Plant Journal*, 34, 257–267.

Sato, S., Peet, M.M. & Thomas, J.F. (2002) Determining critical pre- and post-anthesis periods and physiological processes in *Lycopersicon esculentum* Mill. exposed to moderately elevated temperatures. *Journal of Experimental Botany*, 53, 1187–1195.

Schöning, J.C., Streitner, C., Meyer, I.M., Gao, Y. & Staiger, D. (2008) Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in *Arabidopsis*. *Nucleic Acids Research*, 36, 6977–6987.

Smit, A., Hubley, R. & Green, P. (2013) *RepeatMasker Open-4.0*. Retrieved 2018 at: http://www.repeatmasker.org

De Storme, N. & Geelen, D. (2014) The impact of environmental stress on male reproductive development in plants: biological processes and molecular mechanisms. *Plant, Cell & Environment*, 37, 1–18.

Stricher, F., Macri, C., Ruff, M. & Muller, S. (2013) HSPA8/HSC70 chaperone protein: structure, function, and chemical targeting. *Autophagy*, 9, 1937–1954.

Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, 6, e21800.

Tripathy, M.K., Deswal, R. & Sopory, S.K. (2021) Plant RABs: role in development and in abiotic and biotic stress responses. *Current Genomics*, 22, 26–40.

Walter, W., Sánchez-Cabo, F. & Ricote, M. (2015) GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*, 31, 2912–2914.

Wendel, J.F. & Cronn, R.C. (2003) Polyploidy and the evolutionary. *Advances in Agronomy*, 78, 139–186.

Wu, J.X., Song, X., Pascovici, D., Zaw, T., Care, N., Krisp, C. et al. (2016) SWATH mass spectrometry performance using extended peptide MS/MS assay libraries. *Molecular & Cellular Proteomics*, 15, 2501–2514.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D. et al. (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, 475, 189–195.

Yang, G., Yu, L., Wang, Y., Wang, C. & Gao, C. (2017) The translation initiation factor 1A (TheiF1A) from *Tamarix hispida* is regulated by a dof transcription factor and increased abiotic stress tolerance. *Frontiers in Plant Science*, 8, 513.

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B. et al. (2017) Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 9326–9331.

Zinn, K.E., Tunc-Ozdemir, M. & Harper, J.F. (2010) Temperature stress and plant sexual reproduction: uncovering the weakest links. *Journal of Experimental Botany*, 61, 1959–1968.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Masoomi-Aladizgeh, F., Kamath, K.S., Haynes, P.A. & Atwell, B.J. (2022) Genome survey sequencing of wild cotton (*Gossypium robinsonii*) reveals insights into proteomic responses of pollen to extreme heat. *Plant, Cell & Environment*, 45, 1242–1256. https://doi.org/10.1111/pce.14268