Bioinformatic analysis of RNASeq data from Heat Stressed *Arabidopsis thailana* buds

Name: Mekiran Srividhyadharsan

Project Supervisor: Prof. Zoe Wilson, Dr. Alison Tidy

Keywords: *Arabidopsis thaliana*, Heat Stress, Tapetum, Pollen,

Project write-up submitted to The University of Nottingham in partial fulfilment of the requirements for the degree of Bachelor of Science with Biotechnology

Total length of abstract: 316 words

Total length of text: 5987 words

1. Abstract

Heat stress can severely affect pollen development by disrupting tapetum functioning, particularly its coordinated degradation. This leads to the incorrect formation of microspores and male sterility. Bioinformatic analysis was carried out on genes expressed from heat-stressed Arabidopsis thaliana Ler buds. The objective was to see how heat stress affects tapetum development, with growth stage of the buds and time of treatment being dependent variables. Data was based on buds that were heat treated for six hours, or buds that were heat treated for twenty-four hours. Buds were either prior to microspore stage or buds were from microspore stage to pollen mitosis. Genomic and proteomic analysis was carried out to study the genes, in terms of: what they are; what they express; why they were induced by heat stress and do they link to tapetum development. The frequency of significant genes varied between datasets, with T24 having many more genes expressed than T6, and more genes upregulated than downregulated in general. GO analysis on the datasets found that T6 seemed to be more limited to less complex and simpler GO terms, with T24 having many more and a broader scope. The regulation of tapetum development by specific transcription factors and their respective proteins led to the identification of proteins in our datasets that could be specifically induced by that network as a response to heat stress. Using databanks and creating stress-focused correlation networks that compared protein-protein interactions, some thermotolerant mechanisms and respective genes from the datasets were analysed. Buds that were heat treated for six hours and were from microspore stage to pollen mitosis growth stage lacked genes that seemed to be key factors in thermotolerance, and present in other buds. This could be due to resource competition, but when the heat stress persists, the buds switch priorities to thermotolerance, which is why genes lacking in this stage was found expressed after twenty-four hours of heat treatment.

2. Introduction

Climate change will cause a rise in the global average temperature which will reduce global yields of many important crops, such as wheat, rice, maize and soybeans (Zhao et al., 2017). The seeds and fruits from many flowering plants contribute heavily to the human diet, e.g. two-thirds of human caloric intake comes from those four crops (Rieu, Twell and Firon, 2017; Zhao et al., 2017). The frequency and intensity of hot days and heat waves are predicted to increase in many temperate regions (Rieu, Twell and Firon, 2017). High temperatures will put the plant under heat stress and can cause: cellular injury; inactivation of enzymes, chloroplast and mitochondria; loss of membrane integrity; protein denaturation and aggregation and inhibition of protein synthesis (Cha et al., 2013).

Pollen development and functioning are heat-sensitive processes, and exposure to high temperature matches reproductive phase in the plant's life cycle (Rieu, Twell and Firon, 2017). Pollen injury causes aberrations in tapetum development or morphology, and alters timing of tapetum degeneration (Rieu, Twell and Firon, 2017). Tapetum is one of the four nonreproductive layers found in anther lobes, and the correct functioning of tapetum and its coordinated degradation is essential for pollen development (Ma, 2005; Rieu, Twell and Firon, 2017). Tapetum tissue provides developing microspores with carbohydrates, nutrients, enzymes compounds for outer pollen wall formation (Rieu, Twell and Firon, 2017). Tapetum functioning, particularly its coordinated degradation shortly after microspores are released from the tetrad, is essential for pollen development (Ma, 2005; Rieu, Twell and Firon, 2017. Heat stress tapetal defects strongly affects the progression of male gametogenesis, resulting in the incorrect formation of microspore cells causing male sterility (Giorno et al., 2013).

Datasets for an RNA-sequencing (RNASeq) experiment focusing on how heat stress affects tapetum development was produced by Yang Song.

Arabidopsis thaliana genotype Ler buds were heat stressed to 32^0C under continuous lights, and after six hours, samples were collected and labelled 'T6`, and after twenty-four hours, samples were collected and labelled 'T24`. Buds were either prior to polarised microspore stage and labelled 'Young`, or they were buds from polarised microspore stage to pollen mitosis and labelled 'Old`. The four RNASeq datasets are split into 'Young T6`, 'Old T6`, 'Young T24`, and 'Old T24`. Abiotic stress can differ based on factors such as species, genotypes, age, timing of stress and intensity (Le Gall et al, 2015). The aim of this bioinformatic analysis is to study how the effect of heat stress differs based on the time of heat treatment and the growth stage of the buds, with a particular focus on tapetum development.

3. Materials and Methods

3.1 AgriGO

Gene Ontology (GO) is the standard in gene functionality description and widely used in functional annotation and enrichment analysis (Du et al., 2010). AgriGO is an integrated, web-based GO analysis toolkit and data repository containing 38 agricultural species composed of 274 data types (Du et al., 2010). There is a search option where cross-comparisons between data sets can be explored and visualised (Du et al., 2010). Also, there is a customisable, analysis approach that uses Gene Set Enrichment strategy (Du et al., 2010). Four tools are integrated into the toolkit to meet different demands: SEA (Singular enrichment analysis); PAGE (Parametric Analysis of Gene set Enrichment); BLAST4ID (Transfer IDs by BLAST) and SEACOMPARE (Cross comparison of SEA) (Du et al., 2010). AgriGo was used on the four datasets in Table 1 to identify significant GO terms and create text trees based on the three GO term categories/root terms: biological process, molecular function and cellular component.

3.2 BAR Classification SuperViewer Tool with Bootstrap

BAR's Classification SuperViewer Tool with Bootstrap provides an alternate display of GO, GO slim or MapMan classifications for a list of genes using a bar code scheme (Waese and Provart, 2016). Instead of finding significant GO terms, overrepresented GO terms represent how common the term is relative to the overall GO database (Waese and Provart, 2016). The tool outputs an overview table that highlights enriched categories in bold if the p-value is 0.05 or lower based on a hypergeometric test (Waese and Provart, 2016). Overrepresented GO terms have been identified for the four datasets from Table 1.

3.3 Venny

Venny is a venn diagram tool used to compare up to four lists of elements and identify any overlap between the lists (Oliveros, 2015). The four lists were the upregulated or downregulated genes in each dataset, so genes presented exclusively in the dataset and common genes found in more than one dataset have been recorded. The frequency of these genes is detailed in Table 2.

3.4 CORNET

CORrelation NETworks (CORNET) is a data mining and integration tool that can be used to construct networks based on multiple gene inputs or proteins found in *Arabidopsis thaliana* (De Bodt et al., 2012). Three tools can be used individually or in combination, and they are the coexpression tool, the protein-protein interaction (PPI) tool and the transcription factor (TF) tool (De Bodt et al., 2012). TF and PPI cannot be used at the same time. Genes can be returned with specific selection criteria based on multiple microarray datasets. The method is based on Pearson or Spearman correlation coefficient and the P-Value can be manually set. There are three types of correlations, pairwise correlations, correlation of query genes with neighbours and correlation between neighbours (De

Bodt et al., 2012). The edge label could be based on the mean, maximum or minimum correlation coefficient. For this investigation, the default guidelines were used, and all correlations was selected, but only one could be shown at a time and saved. The networks produced shows pairwise correlations based on protein-protein interactions and the localisation of genes, using the selection criteria dataset called Stress (abiotic and biotic) TAIR10. All protein-protein interactions databases were selected which were: IntAct, TAIR, BAR, ArathReactome, AtPID, BioGRID, DIP, MINT, BIND, DeBodt (predicted-low stringency; filtered – high stringency), Ath Interactome Mapping Consortium, AraNet (gene-gene associations), MIND0.5, G-protein interactome, EVEX binding, STRING and BORDER. The network is visualised using Java Web Start Launcher, but in order to see them offline, Cytoscape can be used as a comprehensive visualisation of the networks generated (De Bodt, 2012). 20 networks were produced in total and the query genes produced and their interactions with other genes have been considered to identify the select genes that have been analysed in the results section.

3.5 PANTHER, UniProt and TAIR

In order to analyse the genes and their respective proteins, a combination of PANTHER, UniProt and TAIR was used as needed. PANTHER provides a library of protein families and subfamilies that relate the protein sequence relationships to function relationships (Thomas et al., 2003). Microarray ID, Gene symbol, description from PANTHER and protein class have been recorded for all the genes referenced in Table 2. Some of the genes produced from the dataset will have their respective proteins grouped as as either oxidoreductases or transcription factors, but frequency of protein hits compared to data provided found in Table 3 is rather low due to proteins still needing to be characterised, particularly if they are from downregulated genes. UniProt has a large resource of protein sequences

and associated detailed annotations with over 60 million sequences, with over half a million sequences curated by experts who critically review experimented and predicted data for each protein (Apweiler et al., 2004). TAIR provides an integrated view of genomic data for *Arabidopsis thaliana* that collates information from a number of sources and provides information on genes, markers, polymorphism, maps, sequences, clones, DNA and seed stocks, gene families and proteins (Garcia-Hernandez et al., 2002). Information from all three has been collected for the analysis of many genes produced from the datasets.

4. Results

4.1 General Overview

4.1.1 Significant genes

There are 33602 genes from *Arabidopsis thaliana* for each dataset. To analyse the genes, the first task was to isolate the significant genes. To do this, genes had to have a false discovery rate less than 0.05. Then, if the fold change was greater than or equal to 2, the significant gene would be classified as upregulated, or if the fold change was less than or equal to -2, the significant gene was classified as downregulated. The frequency of the genes between datasets is described in Table 1. The proportion of significant genes is very low, the highest being 3.2% from Old T24 and the smallest being 0.562% for Young T6. Out of the significant genes, upregulated genes have a higher proportion than downregulated genes in all the datasets. Additionally, changing heat treatment from 6 hours to 24 hours caused a greater expression of significant genes, than the change from young to old buds. Taking Young T6 as an example, Old T6 had 43 more significant genes expressed than Young T6 (22.8% increase), while Young T24 had 600 more significant genes expressed than Young T6 (317% increase). Furthermore, the longer the heat treatment, the

smaller the difference between upregulated and downregulated genes becomes. This is based on the results that Old T6 had a 38% difference, while Young T24 had a smaller difference of 21.2% and Old T24 had the smallest difference of only 6.8%.

Table 1. Frequency of significant, upregulated and downregulated genes

Dataset	Significant	Upregulated	Downregulated
Young T6	189 (0.562%)	110 (58.2%)	79 (41.8%)
Old T6	232 (0.69%)	160 (69.0%)	72 (31.0%)
Young T24	789 (2.35%)	478 (60.6%)	311 (39.4%)
Old T24	1099 (3.2%)	585 (53.2%)	514 (46.8%)

The table shows the significant, upregulated and downregulated genes present in the four datasets for an RNA-sequencing (RNASeq) experiment focusing on how heat stress affects tapetum development, produced by Yang Song. Six hours of heat stress treatment = 'T6'; Twenty-four hours of heat stress treatment = 'T24'; Buds either prior to polarised microspore stage = 'Young' ; Buds from polarised microspore stage to pollen mitosis = 'Old'. Percentage values in the significant column represents significant genes over the total number of genes in the dataset, while percentage values in Upregulated and Downregulated column represent the upregulated or downregulated genes over the total number of significant genes, respectively

4.1.2 GO term analysis

AgriGO and BAR Classification SuperViewer Tool with Bootstrap was used on the new four datasets shown in Table 1, to produce significant and overrepresented GO terms, respectively. For all three GO term categories: biological process, cellular component and molecular function, there are no significant GO terms in T6 buds. For Young T24, there were no significant GO terms in cellular component and molecular function but for biological process, response to abscisic acid (ABA) stimulus was significant. For Old T24, there were no significant GO terms for molecular function. For the cellular component category, photosystem 2 was a significant GO term. For the biological process category: photosynthesis;

ABA mediated signalling pathway; response to salt stress and response to water deprivation were significant GO terms.

Overrepresented GO terms for all three GO categories are much fewer in T6 than T24. For the biological process category, three terms appeared in T6 buds, which were: DNA-dependent transcription; Cell organisation and biogenesis and Signal transduction. Both DNA-dependent transcription and Signal transduction is associated with downregulated T6, while Cell organisation and biogenesis is associated with upregulated Old T6. These three terms were present in T24 with: Response to abiotic or biotic stimulus; DNA-dependent transcription and Cell organisation and biogenesis, associated with all T24 buds. Signal transduction was linked solely to upregulated Young T24. Response to stress was linked to Young T24, but only upregulated Old T24. Developmental process is in Young T24, but only downregulated Old T24. The most specific terms are DNA or RNA metabolism, which is in upregulated Young T24, and electron transport or energy pathway, which is in upregulated Old T24.

For the molecular function category, terms are related to either binding or activity. Three terms appeared in T6 buds, which were: DNA or RNA binding; Nucleotide binding and Nucleic acid binding. DNA or RNA binding is linked to Young T6, while nucleotide binding is linked to downregulated Young T6 and upregulated Old T6. Nucleic acid binding is linked to upregulated Old T6. These three terms are present in T24 with: DNA or RNA binding; Nucleotide binding; Nucleic acid binding and Protein binding associated with all T24 buds. For the other half of molecular function, hydrolase activity is associated with Young T24 and downregulated Old T24. Transcription factor activity is linked to Young T24. Transporter activity is associated with upregulated Young T24 and downregulated Old T24. Structural molecule activity is associated with Old T24 and downregulated Young T24. The most specific terms is kinase activity, only linked to downregulated Old T24.

For the cellular component category, four terms appeared in T6 buds: which were: Cytosol, Cell wall, Plasma membrane and Nucleus. Cytosol is linked to downregulated Young T6. Cell wall is associated with upregulated Young T6 and downregulated Old T6. Plasma membrane is associated with upregulated Young T6 and downregulated Old T6. Nucleus is associated with upregulated Young T6 and downregulated Old T6. These four terms are present in T24 with: cytosol, plasma membrane, nucleus, plastids, mitochondria and ribosome associated with all T24 buds. Cell wall and Chloroplasts were associated with all Old T24 and downregulated in Young T24. The most specific term is linked to Golgi apparatus found in upregulated Old T24.

A few assumptions about our data can be made from this analysis. Most of the genes encode proteins with binding activity, that respond to abiotic or biotic stimulus and are linked to cell wall, nucleus, plasma membrane and cytosol, in some regard or another. A few proteins in T24 undergo a lot of movement as either transporters or kinases, with mitochondria providing energy. The high number of transcription factors and ribosomes could mean that new transcripts and proteins are being made in response to heat stress.

4.1.3 Exclusive and common genes

There are several heat responses that the plant undergoes to maintain physiological homeostasis during pollen development at high temperature (Rieu, Twell and Firon, 2017). Some of these responses was based on how they linked to tapetum development, and any corresponding genes was chosen for analysis with significance given to whether they appeared in the correlation networks and the validated information available. The frequency of two protein classes is shown in Table 2, and these particular proteins are linked to some of the heat responses mentioned. To know if genes were expressed in one or more buds, the frequency of the exclusive and common significant genes is shown in Table 3. Further

analysis using PANTHER and CORNET was carried out specifically on the types of data presented in Table 3.

Table 2. Frequency of two particular protein classes

	YT6 (Up)	Y6 (Down)	OT6 (Up)	OT6 (Down)	YT24 (Up)	YT24 (Down)	OT24 (Up)	OT24 (Down)
Transcription factors	2	1	5	2	16	8	15	10
Oxidoreductases	3	1	7	1	19	15	30	18

Numbers of transcription factors and oxidoreductases was based on PANTHER Functional classifications analysing the previous four datasets produced in Table 1 with the upregulated and downregulated genes split into separate columns. The upregulated columns have been filled with a light grey background to differentiate between downregulated columns. Six hours of heat stress treatment = 'T6'; Twenty-four hours of heat stress treatment = 'T24'; Buds either prior to polarised microspore stage = 'Y'; Buds from polarised microspore stage to pollen mitosis = 'O'. 'Up' = upregulated, 'Down' = downregulated.

From Table 3, the results show that exclusive T24 genes have a very high frequency, with exclusive downregulated Old T24 genes having more than half of all downregulated genes. In contrast, the frequency of the common genes rarely exceeds 3%, except for the common genes between Young and Old T24. Table 2 shows that T24 will have more proteins than T6, and proteins from upregulated genes will surpass the proteins from downregulated genes, although that table is limited to only two protein classes. By looking at specific genes and their link to tapetum development, more useful conclusions can be surmised on the effects of heat stress.

**Table 3.** Frequency of exclusive and common genes

Type of data	Upregulated	Downregulated
Exclusive YT6	61 (5.46%)	58 (7.16%)
Exclusive OT6	94 (8.41%)	47 (5.80%)
Exclusive YT24	326 (29.2%)	232 (28.6%)
Exclusive OT24	435 (38.9%)	436 (53.8%)
Common in YT6, OT6 and OT24	3 (0.268%)	1 (0.123%)
Common in YT6, YT24 and OT6	4 (0.358%)	0 (0.00%)
Common in YT6, YT24 and OT24	6 (0.537%)	3 (0.370%)
Common in YT24, OT6 and OT24	1 (0.0894%)	0 (0.00%)
Common in YT6 and OT6	13 (1.16%)	4 (0.494%)
Common in YT6 and YT24	13 (1.16%)	10 (1.23%)
Common in YT6 and OT24	10 (0.894%)	3 (0.370%)
Common in OT6 and YT24	0 (0,00%)	8 (0.988%)
Common in OT6 and OT24	24 (2.15%)	0 (0.00%)
Common in YT24 and OT6	21 (1.88%)	8 (0.988%)
Common in YT24 and OT24	107 (9.57%)	0 (0.00%)

The table shows the frequency of exclusive and common genes, split into genes that are upregulated and downregulated. Exclusive means that significant gene is present in any other dataset. Common means that significant gene is present in more than one datasets. Six hours of heat stress treatment = 'T6'; Twenty-four hours of heat stress treatment = 'T24'; Buds either prior to polarised microspore stage = 'Young'; Buds from polarised microspore stage to pollen mitosis = 'Old'. Percentage values compare how many genes in that specific element over the total number of either upregulated or downregulated genes. This splitting of data was produced using Venny.

4.2 Processes affected by Heat Stress and linked to Tapetum Development

4.2.1 Genetic regulatory network for tapetum development

In *Arabidopsis thailana,* four transcription factors DYT1, TDF1, AMS, MS188 transcription factors form regulates tapetum development and pollen wall formation (Li et al., 2017). DYT1 regulates early tapetum development via E3 ubiquitin ligases (Li et al., 2017). TDF1 regulates redox and cell degradation (Li et al., 2017). AMS regulates lipid transfer proteins and other E3 ubiquitin ligases (Li et al., 2017). MS188 regulates cell-wall related genes, involved in tapetum cell wall degradation and pollen wall formation (Li et al., 2017). The expression of all four transcription factors is shown in Figure 1. All four are located in the tapetum but have slightly different expressions: DYT1 is expressed strongly in anther stage 15 and later. TDF1 expression begins moderately in anther stage 12-14 and more strongly in the anther stages 15-18. AMS expression starts at anther stage 6 and increases tills its strongly expressed in anther stages 12-18 stages, and in the stamen. MS188 is expressed in 12-14 anther stages and less strongly in anther stages 15-18, and in the stamen. Tapetum cells are identified at anther stage 5 and undergoes programmed cell death (PCD) in stage 10 (Li et al, 2017). Apart from AMS, most of these transcription factors are not expressed in the same anther stage. Even AMS has a low expression in its early stages. Each transcription factor regulates many genes that are involved in tapetum development that work downstream from the transcription factor (Li et al., 2017), but the effects of heat stress on those genes was not tested. The four transcription factors are not present in the dataset.
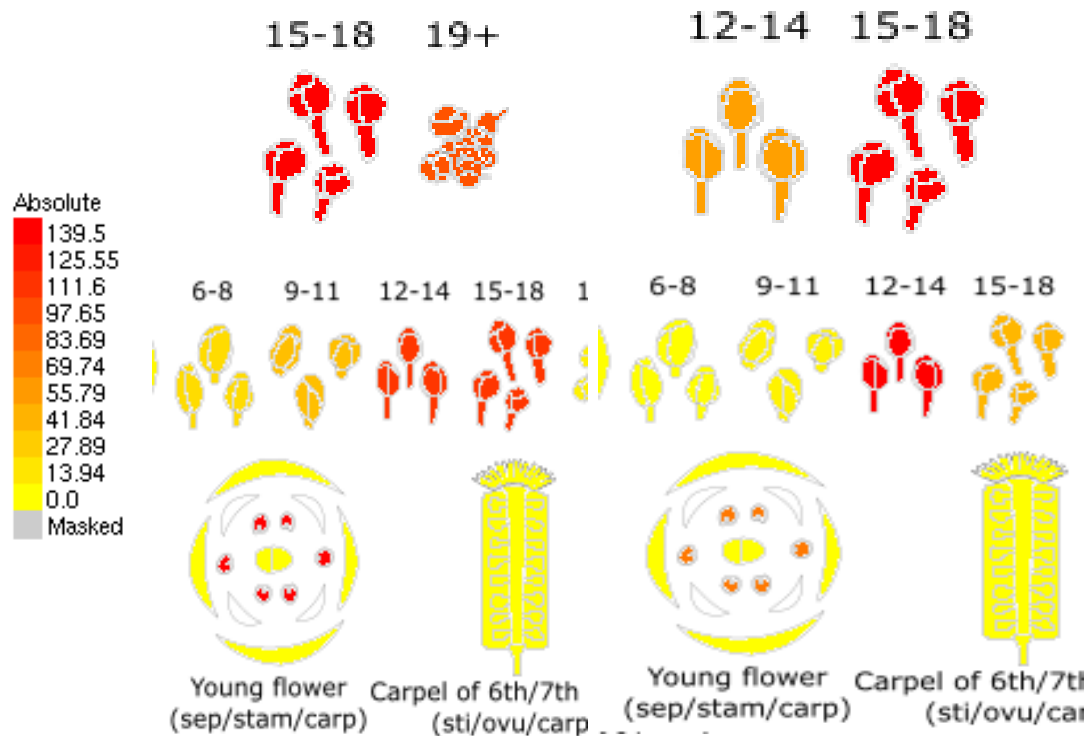
Figure 1. Expression of DYT1-TDF1-AMS-MS188

These images were taken and edited from TAIR Data Source, Klepikova Arabidopsis Atlas efp Browser, which shows the expression of four transcription factors, **DYT1 (top left), TDF1 (top right), AMS (bottom left) and MS188 (bottom right)** on a high resolution map of the *Arabidopsis thalina* development developmental transcriptome based on RNA-seq profiling. The bar on the left shows the level of expression and the colour gradient, with red being the highest and yellow being the lowest. DYT1 is expressed strongly in anther stage 15 and later. TDF1 expression begins moderately in anther stage 12-14 and more strongly in the anther stages 15-18. AMS expression starts at anther stage 6 and increases tills its strongly expressed in anther stages 12-18 stages, and in the stamen. MS188 is expressed in 12-14 anther stages and less strongly in anther stages 15-18, and in the stamen

Apart from those transcription factors, E3 ubiquitin ligases, cytochrome P450 and polygalacturonases are all involved in regulating tapetum and pollen development (Li et al., 2017). Some of these proteins are in the datasets and is shown in Table 4.

Table 4. Genes that encodes proteins from cytochrome P450s, E3 ubiquitin ligases and polygalacturonases families

Dataset	Protein family
Young T6	**Cytochrome P450s -** *Upregulated*: CYP81D7 **E3 Ubiquitin ligases -** *Upregulated*: RHF1A **Polygalacturonases -** *Upregulated*: PGIP1
Old T6	**Cytochrome P450s -** *Upregulated*: CYP71B20, P450**;** *Downregulated*: CYP96A5 **E3 Ubiquitin ligases -** *Upregulated*: MYC6.13, ATTL4
Young T24	**Cytochrome P450s -** *Upregulated*: CYP71B20, CYP72A7, CYP71B3, CYP78A8**;** *Downregulated*: CYP705A24, CYP96A1, CYP98A8, CYP94C1 **E3 ubiquitin ligases -** *Upregulated*: RGLG1, RHF1A, LUC1
Old T24	**Cytochrome P450s -** *Upregulated*: CYP709B2, CYP89A4, CYP71B10, CYP76C2**;** *Downregulated*: CYP71A16, CYP78A6, CYP7A7, CYP71A18 **E3 ubiquitin ligases -** *Upregulated*: MAJ23.10, SINAT5, ORTHL, RHA1B, MPSR1; *Downregulated*: BRG3 **Polygalacturonases -** *Downregulated*: PGL3

Proteins from cytochrome P450s, E3 ubiquitin ligases and polygalacturonases family of proteins are split into the datasets that they are expressed from. Six hours of heat stress treatment = 'T6`; Twenty-four hours of heat stress treatment = 'T24`; Buds either prior to polarised microspore stage = 'Young`; Buds from polarised microspore stage to pollen mitosis = 'Old

Cytochrome P450s are heme-containing proteins catalysing nicotinamide adenine dinucleotide phosphate hydrogen (NADPH) and superoxide (O_2^-) dependent monooxygenase (Li et al., 2017). E3 ubiquitin ligases degrade proteins in plant development (Li et al., 2017). Polygalacturonases are typical cell wall hydrolytic enzyme (Li et al., 2017). More analysis should be carried out on these genes to see why these responded significantly to heat stress compared to other members in their protein families.

4.2.2 Heat shock response

Heat stress affects protein function by changing their structure and creating misfolded proteins (Rieu, Twell and Firon, 2017). Furthermore, proteins can become cytotoxic if more hydrophobic regions were exposed, eventually leading to cell death (Rieu, Twell and Firon, 2017). In response to heat stress, elevated levels of Heat shock proteins (HSPs) establish protection and thermotolerance mechanisms in male reproductive cells (Giorno et al., 2013). Acquired thermotolerance is the capacity of plants to cope with repeated heat stimulations (Meiri et al, 2009). HSPs function as molecular chaperones that can fold and refold proteins during meiosis and tetrad formation when pollen develops (Giorno et al., 2013). The 90-kDA HSPs family of proteins (HSP90) are highly conserved molecular chaperones, and HSP90-1 exhibits a strong expression pattern distinct from all other HSP90 genes found in *Arabidopsis thaliana* (Cha et al., 2013; Swindell et al, 2007).

HSP90-1 is localised to the cytosol, and under normal conditions, it is mainly found in cotyledons and the tip of roots, but this tissue specificity is lost when plants are heat shocked (Prasinos et al., 2005). After heat shock treatment, HSP90-1 promoter activity was low throughout the early stages of embryo development and high before embryo maturation (Prasinos et al., 2005). HSP90-1 promoters were found in pollen grains and was active during pollen germination (Prasinos et al., 2005). HSP90-1 is one of the more common gene expressed in Young T6, Young T24 and Old T24. From Figure 2, we can see that HSP90-1 has protein-protein interactions with 14 proteins and dimerises. One of these proteins, FKBP65 is a heat stress modulator that functions in negative feedback regulation of HSDA2 (Meiri et al, 2009). FKBP65 interacts with HSP90-1, by translocating to the nucleus and forming heterodimers with ROF1(FKBP62), joining a complex between ROF1-HSP90.1-HSFA2, to give the plant long-term thermotolerance (Meiri et al, 2009). The protein-protein interactions of FKBP62 shown in Figure 2, confirm that FKBP65

interacts with HSP90-1 and ROF1, as well as other heat shock proteins. FKBP65 is upregulated in Young and Old T6 buds. Some of the proteins that HSP90-1 interacts with are other heat shock proteins from the HSP70, HSP17, HSP23 families of protein, as well as a heat stress transcription factor, HSFB2B. These protein families have other members that are present in our datasets.

Hsp17 proteins exhibit large expression responses to many types of stress in both roots, shoots and seeds (Swindell et al., 2007). HSP23 proteins are highly responsive to stress treatments in root tissue (Swindell et al., 2007). HSP70 proteins have a smaller expression response to stress but are strongly induced by heat in early times of heat treatment, mainly in shoot tissues (Swindell et al., 2017). HSP17.6A is a gene that is upregulated in Young T6 and Young T24, induced by heat stress during seed development, and demonstrates chaperone activity (Sun et al., 2001). HOP1 is upregulated in Old T24 and encodes an organising protein that mediates association of molecular chaperones from HSP70s and HSP90s. HSP23.5 is a pseudogene localised to the mitochondria upregulated in both Young and Old T24, with a currently unknown function (Waters et al., 2008).

Figure 2. Pairwise correlations of HSP90-1 and FKBP65

These images were taken from correlation networks produced focusing on common genes in Young T6, Young T24 and Old T24. Not all genes in this network came from the genes provided. Pairwise correlations are based on protein-protein interactions and the localisation of genes, using the selection criteria dataset called Stress (abiotic and biotic) TAIR10. The network is visualised using Cytoscape and the yellow gene highlighted is the query gene, **HSP90.1 (left), FKBP65 (right)**. The red lines represent correlation, and if it is dotted this means data was predicted, while straight lines means experimental evidence was used. The thickness of the line represents how much supporting evidence is available. Looping back on itself means dimerization. The blue lines and black lines show other correlations in the network, with dark blue and black meaning that more experimental evidence supports this correlation. HSP90-1 is found in Young T6, Young T24 and Old T24, and has protein-protein interactions with 14 proteins as well as dimerising. FKBP65 is found in Young T6 and Old T6, and has 10 protein-protein interactions, 2 of which are predicted interactions. Six hours of heat stress treatment = 'T6'; Twenty-four hours of heat stress treatment = 'T24'; Buds either prior to polarised microspore stage = 'Young'; Buds from polarised microspore stage to pollen mitosis = 'Old

Heat stress transcription factors (HSFs) play crucial roles in heat shock response, by regulating expression of stress-responsive genes as the terminal component in the signal transduction chain (Guo et al., 2016). HSFA2 is upregulated in Young T24 and is thought to have a central role in plant heat stress response, accumulating in the nucleus and forming heterodimers with HSFA1 resulting in synergistic, transcriptional activation of heat stress genes (Guo et al., 2016). HSFA2 is induced during the process of callus formation and growth from root explants (Guo et al., 2016). HSFA1E is downregulated in Young T24 and is a key regulator of the heat stress signalling network (Nishizawa-Yokoi et al., 2011). Overexpression of HSFA1E lowers the expression of HSFA2, which fits with our results given that it is the only downregulated HSF found in Young T24 (Nishizawa-Yokoi et al., 2011). HSFA9 is exclusively expressed in late stages of seed development and is found upregulated in Old T24 (Guo et al., 2016). Similarly, HSFB2a is upregulated in Old T24 and is involved in vegetative and gametophytic development as without HSFB2a, there is an impaired biomass production and reduction in female gametophytes (Wunderlich, Groß-Hardt and Schöffl, 2014). Most of the HSFs do not appear under normal conditions, but rather respond to heat stress and provide some regulation to assist development in the later growth stages of the plant. This explains their lack of presence in young buds, which is also supported by the lower number of transcription factors found in T6 compared to T24 shown in Table 2.

## 4.2.3 Reactive Oxygen Species

Reactive oxygen species (ROS) are produced by aerobic metabolism from cellular components like mitochondria, chloroplasts, peroxisomes and apoplast (Rieu, Twell and Firon, 2017). ROS have important roles in modulating cell survival, cell death, differentiation, cell signalling and inflammation-related factor production (Abdal Dayem et al., 2017). There are many types of ROS elements including free radicals like singlet

oxygen, hydroxyl and alkoxyl, and nonradicals like hydrogen peroxide, nitric oxide and organic peroxides (Abdal Dayem et al., 2017). Cells possess ROS scavenging and detoxification machinery (Rieu, Twell and Firon, 2017). These could be enzymes like superoxide dismutases, catalase, ascorbate peroxidase, and antioxidants like ascorbic acid, flavonoids, vitamin E and glutathione (Abdal Dayem et al., 2017). The usual balance is disturbed by high temperature as heat leads to an accumulation of ROS (Rieu, Twell and Firon, 2017). Disturbance of the reduction-oxidation homeostasis caused by ROS generation and neutralisation, can lead to interference in cell-signalling mechanisms and result in oxidative damage to biomolecules like lipids, proteins and nucleic acids (Abdal Dayem et al., 2017).

ROS plays a key role in tapetal programmed cell death (PCD) in *Arabidopsis thaliana* (Yu et al., 2017). ROS levels peaks during tapetum degradation and pollen maturity in anthers (Rieu, Twell and Firon, 2017). Not removing ROS in the tapetum, results in premature tapetal PCD, resulting in male sterility (Yu et al., 2017). Reducing ROS levels during anther development by mutagenesis of genes encoding NADPH oxidases delays tapetal PCD (Yu et al., 2017). In both cases, pollen development is extremely impaired, so tight regulation of dynamic ROS levels for tapetal PCD and pollen development is essential for *Arabidopsis thaliana* (Yu et al., 2017).

The data supports these ideas as from Table 2, oxidoreductases are present in all datasets and is both upregulated and downregulated. As the time of heat treatment increases, more oxidoreductases are produced to deal with the increased accumulation of ROS. Glutathione is one of the antioxidants mentioned and glutathione transferases (GST) are present throughout the datasets. GSTF2 is upregulated in Young T6 and GSTU5 is upregulated in Old T24. GSTU17, GSTU18 and GSTU25 are downregulated in Old T24 and GSTT1 is downregulated in Old T6. As for peroxidase genes, PER63 is upregulated in Young T24, and both PER40 and PER17

are upregulated in Old T6. PER9, PER44 and PER43 genes are all downregulated in Old T24. All peroxidases are involved in removal of hydrogen peroxide (H_2O_2), a highly stable, non-radical ROS element with a long lifespan (Abdal Dayem et al., 2017). Hydrogen peroxide (H_2O_2) is an essential component of the heat stress signalling pathway, specifically being one mechanism for HSF activation during the early phase of heat stress (Volkov et al, 2006). The fact that more peroxidases are expressed in Old buds supports this. One gene in particular is GPX3 which encodes a glutathione peroxidase that has dual and distinctive roles in H_2O_2 homeostasis. It is expressed ubiquitously in all tissues and found in the mitochondria (Milla et al., 2003). GPX3 functions as both a scavenger and a redox transducer relaying the H_2O_2 oxidative signal in ABA and drought stress signalling (Miao et al., 2006). Our data shows that it also responds to heat stress, specifically upregulated in Young T24. GPX3 dimerises and interacts with three proteins, PUM5; AGAL3 which regulates leaf development and functions in cell wall loosening, and CBSX3 which is involved in cell redox homeostasis as well, as shown in Figure 3.
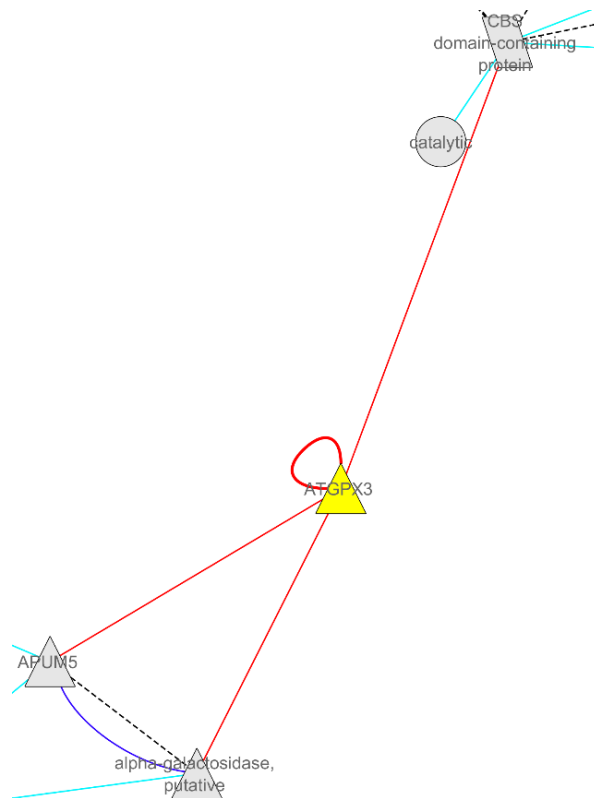
Figure 3. Pairwise correlations of GPX3

This image was taken from correlation networks produced focusing on exclusive genes in Young T24. Not all genes in this network came from the genes provided. Pairwise correlations are based on protein-protein interactions and the localisation of genes, using the selection criteria dataset called Stress (abiotic and biotic) TAIR10. The network is visualised using Cytoscape, the yellow gene highlighted is the query gene, GPX3. The red lines represent correlation. Looping back on itself means dimerization. GPX3 interacts with three proteins and dimerises. The blue lines and black lines show other correlations in the network, with dark blue and black meaning that more experimental evidence supports this correlation. Buds either prior to polarised microspore stage = 'Young'; Twenty-four hours of heat stress treatment = 'T24';

4.2.4. Phytohormone Ethylene and Lipid stress signalling

Phytohormones can be involved in stress signalling and programmed cell death. Jasmonic acid (JA), salicylic acid (SA), abscisic acid (ABA) and ethylene regulate the expression of HSFs and since they share some common elements, HSFs could act as a cross-point or node connecting

several abiotic and phytohormone signalling pathways (Guo et al., 2015). Lipid signalling can be triggered by changes in membrane fluidity caused by heat stress (Bokszczanin, 2013). Phosphatidylinositol signalling pathway is part of signal transduction, and signalling lipids play critical roles in plant growth and stress responses (Zheng et al., 2011). SAL2 is one of the proteins involved in this pathway and is found upregulated in Old T6 (Gil-Mascarell et al., 1999). A few specific phospholipids react within minutes of a sudden temperature increase and is important for a plant to maintain homeostasis to survive high temperatures (Mishkind et al., 2009). They are phosphatidylinositolphosphate kinase (PIPK), phosphatic acid (PA), phosphatidylinositol 4,5-biphosphate (PIP_2) and PLD (Mishkind et al, 2009). Heat stress activates PLD and PIPK which induces PA and PIP_2 which are key mediators of signalling pathways, membrane dynamics and cytoskeletal organisation (Mishkind et al., 2009). PIP5K6 gene was upregulated in Young T24, and PIP2-6 gene was upregulated in T24 buds.

Ethylene signalling could be involved in tapetal development and PCD (Parish and Li, 2010). Delayed PCD of tobacco tapetum cells occurred when a melon receptor was constitutively expressed by an anther-specific promoter (Parish and Li, 2010). Rice tapetum express ethylene signalling genes, especially during tetrad and unicellular stages (Parish and Li, 2010).

Ethylene production could be an acclimation factor, since the ethylene-insensitive tomato mutant, Never ripe (Nr) is more sensitive to chronic mild heat stress (Rieu, Twell and Firon, 2017). Pollen tolerance can be improved by chemically inducing ethylene production before a short heat stress treatment and the application of ethylene inhibitor can reduce this pollen tolerance (Rieu, Twell and Firon 2017). The overexpression of ethylene response factors (ERFs) is thought to regulate heat stress responses (Müller and Munné-Bosch, 2015).

The results in the data support this as 4 ERFs in Young T24 (ERF5, ERF8, CRF5, ERF055) and 2 ERFS in Old T24 (ERF113, ERF104) are upregulated. There is an ETR1 gene that encodes an ethylene receptor in Young T6 that is also upregulated. However, not all ERF genes in the dataset are upregulated, two ERF genes, RAP2-2 and ANT, are downregulated in Old T24. For RAP2-2, the downregulation could be due to the fact that RAP2-2 is induced by darkness, which is limited by the continuous light used in the heat stress treatment (Hinz et al., 2010). ANT gene encodes an ERF that is involved in ovule development; initiation and growth of floral organs and some processes in flower development (Krizek, Prost and Macias, 2000). ANT is a negative regulator of an AGAMOUS (AG) gene (Krizek, Prost and Macias, 2000), and AG genes are necessary to specify reproductive organs in the early steps of flower development (Ito et al., 2004). AGL70 gene is upregulated in Old T24 and encodes an AMAGOUS MADS-BOX protein that prevents premature flowering by negatively regulating flowering time (Ratcliffe et al., 2003). The downregulation of ANT and the upregulation of AGL70 suggest this is an acclimation response for Old T24 to prevent flowering in unfavourable conditions.

One particular gene, MBF1c encodes a multiprotein bridging factor that accumulates rapidly and is localised to nuclei during heat stress (Suzuki et al., 2008). MBF1c is a key regulator of thermotolerance and is not required for the expression of transcripts encoding different Hsp including HSFA2 (Suzuki et al., 2008). MBF1c controls a tightly, coordinated heat stress-response network that involves trehalose, SA and ethylene signalling pathways (Suzuki et al., 2018). MBF1c dimerises and has protein-protein interactions with 11 proteins and is predicted to interact with HSP90-1 and FKBP65, as shown in Figure 4. MBF1c is upregulated in Young T6, Young T24 and Old T24.
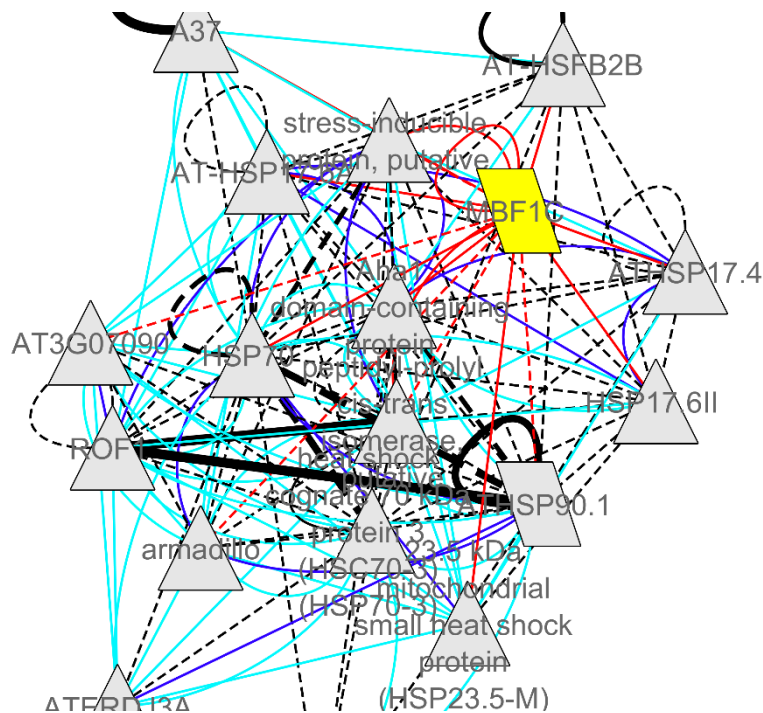
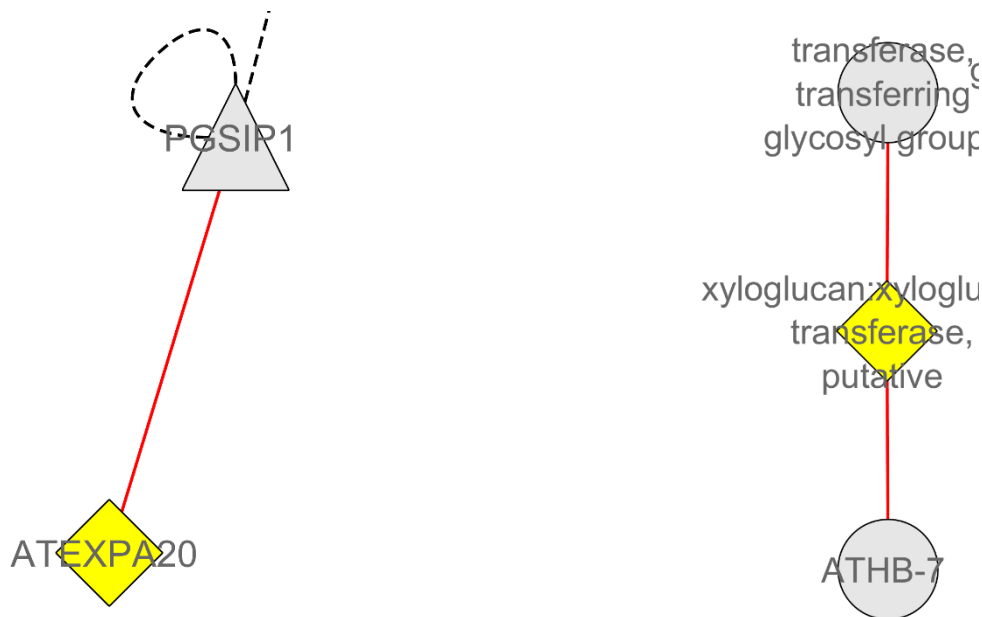**Figure 4**. Pairwise correlations of MBF1c

These images were taken from correlation networks produced focusing on common genes in Young T6, Young T24 and Old T24. Not all genes in this network came from the genes provided. Pairwise correlations are based on protein-protein interactions and the localisation of genes, using the selection criteria dataset called Stress (abiotic and biotic) TAIR10. The network is visualised using Cytoscape and the yellow gene highlighted is the query gene, MBF1c**.** The red lines represent correlation, and if it is dotted this means data was predicted, while straight lines mean experimental evidence was used. The thickness of the line represents how much supporting evidence is available. Looping back on itself means dimerization. The blue lines and black lines show other correlations in the network, with dark blue and black meaning that more experimental evidence supports this correlation. MBF1c dimerises and has protein-protein interactions with 11 proteins. Six hours of heat stress treatment = 'T6'; Twenty-four hours of heat stress treatment = 'T24'; Buds either prior to polarised microspore stage = 'Young'; Buds from polarised microspore stage to pollen mitosis = 'Old'

4.2.5 Cell wall hydrolytic enzymes

Cell wall remodelling is one important method plants employ to respond to heat stress to maintain overall function and growth (Wu, Bulgakov and Jinn, 2018). While it is difficult to identify a common pattern of stress response in cell wall maintenance, one that is known is an increased level

of xyloglucan endotransglucosylase/hydrolase (XTH) and expansin proteins (Gal et al., 2015). Both types of proteins increase rhamnogalacturonan I branching maintaining cell wall plasticity (Gal et al., 2015). XTH proteins is one type of cell wall hydrolytic enzymes (Li et al., 2017). Two expansin proteins are present in the data, EXPA20 is upregulated in Young T6, Young T24 and Old T24, and EXLB1 is upregulated in Old T24 (encodes an expansin-like protein). EXPA20 encodes an expansin protein that disrupts non-covalent bonding between cellulose microfibrils and matrix glucans to loosen and extend plant cell walls. The protein-protein interactions of EXPA20 is shown in Figure 5, and EXPA20 interacts with GUX1, a gene that encodes a glycosyltransferase that adds glucuronic and 4-O- methylglucuronic acid branches to xylan in stem cell walls (Mortimer et al., 2010).

XTH proteins are present in the data, XTH11 is upregulated in Old T6, and both XTH22 and XTH24 are upregulated in Old T24. XTH proteins are able to cleave and relegates xyloglucan polymers, an essential constituent of the primary cell wall, making it vital for cell wall construction of growing tissues. The protein-protein interactions of XTH11 is shown in Figure 4, and XTH11 interacts with ATHB-7, which encodes a transcriptional activator that acts a growth regulator in response to water deficit, and a F14P13.8 gene (Soderman, Mattsson and Engstrom, 1996). Important to note that for XTH11, the catalytic motif is atypical and lacks proton donor sites, so it may not be functional in vivo, unlike other XTH proteins. The breakdown of the cell wall, stage 7 in *Arabidopsis thaliana,* is one of the initial events in tapetal PCD (Parish and Li, 2010). Cellulase and XTH are induced as during PCD forming aerenchyma in maize roots (Parish and Li, 2010).

Figure 4. Pairwise correlations of EXPA20 and XTH11

The left image is taken from a correlation network produced on common genes in Young T6, Young T24 and Old T24. The right image is taken from a correlation network produced on exclusive Old T6. Not all genes in this networks came from the genes provided. Pairwise correlations are based on protein-protein interactions and the localisation of genes, using the selection criteria dataset called Stress (abiotic and biotic) TAIR10. The network is visualised using Cytoscape and the yellow gene highlighted is the query gene, **EXPA20 (Left), XTH11 (Right).** EXPA20 interacts with one protein. XTH11 interacts with two proteins. The red lines represent correlation. The thickness of the line represents how much supporting evidence is available. Six hours of heat stress treatment = 'T6'; Twenty-four hours of heat stress treatment = 'T24'; Buds either prior to polarised microspore stage = 'Young'; Buds from polarised microspore stage to pollen mitosis = 'Old

Expansin and expansin-like proteins act synergistically with cellulases to perform hydrolysis (Duan et al., 2018). The breakdown products of cell walls may activate apoptosis-like PCD pathways, such as binding to arabinogalactan proteins at the plasma membrane-cell wall interface (Parish and Li, 2010).

4.2.6. Summary

All the genes explained so far are summarised in Table 5, split into their dataset and their respective processes affected by heat stress.

Table 5. Genes from processes affected by heat stress and linked to tapetum development

Dataset	Process and respective genes
Young T6	**Heat shock response** *Upregulated*: HSP90-1, FKBP65, HSP17.6A **Reactive oxygen species** *Upregulated*: GSTF2 **Phytohormone Ethylene and lipid signalling** *Upregulated*: MBF1c **Cell wall hydrolytic enzymes** *Upregulated*: EXPA20
Old T6	**Heat shock response** *Upregulated*: FKBP65 **Reactive oxygen species** *Upregulated*: GSTT1, PER40, PER17 **Phytohormone Ethylene and lipid signalling** *Upregulated*: SAL2 **Cell wall hydrolytic enzymes** *Upregulated*: XTH11
Young T24	**Heat shock response** *Upregulated*: HSP90-1, HSP17.6A, HSP23.5, HSFA2; *Downregulated*: HSFA1E **Reactive oxygen species** *Upregulated*: GPX3, PER63 **Phytohormone Ethylene and lipid signalling** *Upregulated*: ERF5, ERF8, CRF5, ERF055, MBF1c **Cell wall hydrolytic enzymes** *Upregulated*: EXPA20
Old T24	**Heat shock response** *Upregulated*: HSP90-1, HOP1, HSP23.5, HSFA9, HSFB2a *Downregulated*: HSFA1E **Reactive oxygen species** *Upregulated*: GSTU5, *Downregulated*: GSTU17, GSTU18, GSTU25, PER9, PER44, PER43 **Phytohormone Ethylene and lipid signalling** *Upregulated*: AGL70, MBF1c; *Downregulated*: RAP2-2, ANT **Cell wall hydrolytic enzymes** *Upregulated*: EXPA20, EXLB1, XTH22, XTH24

The genes and their respective process linked to the buds that they were expressed in, is shown above. These genes were chosen based on the validated information available, their importance to their respective process, and some were chosen based on their appearance in stress-focused correlation network. Six hours of heat stress treatment = 'T6'; Twenty-four hours of heat stress treatment = 'T24'; Buds either prior to polarised microspore stage = 'Young'; Buds from polarised microspore stage to pollen mitosis = 'Old'.

5. Discussion

5.1 Old T6

One explanation for why more significant genes were found in Old buds than Young buds is that while sensitivity of heat varies over the course of pollen development, peak sensitivity occurrs from meiosis to pollen mitosis, the microspore stage (Rieu, Twell and Firon, 2017). After this stage, there is a decrease in sensitivity and relative heat tolerance (Rieu, Twell and Firon, 2017). Even so, time of heat treatment has a much bigger impact to the number of significant genes expressed than the growth stage of the buds affected by heat. This is based on the high expression of genes found in T24 buds compared to T6 buds and more genes involved in thermotolerance is expressed in T24 than T6.

Genes with common thermotolerance mechanisms found in T24 buds were more likely to be in Young T6 than Old T6. This is based on HSP90-1, HSP17.6A (heat shock response), MBF1c (Ethylene) and EXPA20 (Cell wall hydrolytic enzyme), all being present in Young T6 and in T24 buds, but not Old T24. The only XTH genes found in Old T6, XTH11, may not even be functional. However, Old T6 surpasses Young T6 in prioritising tapetum development, as no peroxidases were found in Young T6, while Old T6 have PER17 and PER40 genes. Also, Old T6 did have more proteins in the cytochrome P450s, E3 ubiquitin ligases, than Young T6, but the significance of the proteins is unknown, while the importance of the genes found common in the other three datasets has been highlighted. Mild heat stress can result in defective pollen development, and microspores compete for nutrients in the anther locule under suboptimal growth conditions (Rieu, Twell and Firon 2017). Resource scarcity could limit microspores from prioritising thermotolerance mechanism (Rieu, Twell and Firon, 2017). It could be that in the first six hours of heat stress, the plant is competing for resources, rather than expressing genes related to thermotolerance. As the heat treatment increases to twenty-four hours,

the plant switches priorities, from competition to survival, explaining why Old T24 has specific genes that Old T6 lacks. This could link to Old T6 having the biggest difference in upregulated and downregulated genes compared to the other datasets. To test this, an RNA-Sequencing experiment can be carried out on heat-stressed Old buds for every hour in the first six hours, and test whether resource competition was occurring based on gene expression.

4.2 Further experimentation

In a similar manner, experiments should be done on the genes specified in the Table 6. While the functioning of the proteins and their respective interactions has been highlighted, further testing could be done to see how the expression of proteins changes in the buds depending on different factors. Changing the intensity of heat stress or the duration of the heat treatment could reveal new insights on the way those genes affect tapetum development.

Some future aims are mentioned below:

1) There may be a transcription factor involved in the regulation of HSFA9 expression during seed development (Guo et al., 2016). One study suggested that ABSCISIC ACID-INSENSITIVE3 (AIB3) gene could encode this transcription factor due to the presence of an essential binding factor (Kotak et al, 2007), but that gene is not present in our data. However, there are two similar genes also found in Old T24, ABI5 and ABF4, so there may be a link between HSFA9 and one of these genes. There needs to be a test on the molecular similarities of ABI3 to AIB5 and ABF4, and whether it is involved in regulation of HSFA9.

2) While the regulatory network for tapetum development has been documented (Li et al., 2017), the effects of heat stress on those downstream genes was not tested or focused on. Only a few

proteins from three important families of proteins involved in regulation of the tapetum is present in our dataset, shown in Table 4. Studies into these genes and why these were specifically expressed under heat stress, can provide more insight into the effects of heat on the tapetum.

3) The peroxidase genes for each dataset has been identified but the impact they had on regulating ROS is unclear. An experiment focusing on the effects of peroxidase removal of H_2O_2 and how that may vary between the different heat-stressed buds, could give us an understanding of the effectiveness of plants expressing peroxidases.

4) Since it is known that ethylene responds to heat stress and ethylene may have a role in tapetum PCD, experiments could be done on the ERFs identified and whether externally applying ethylene changes the timing of tapetal PCD or the expression of ERFs.

5) Since MS188 is responsible for cell wall related genes, comparing the genes in that network associated with our expansin and XTH proteins will allow us to see if the heat stress induced these proteins specifically.

5. References

Abdal Dayem, A., Hossain, M., Lee, S., Kim, K., Saha, S., Yang, G.-M., Choi, H. and Cho, S.-G. (2017). The Role of Reactive Oxygen Species (ROS) in the Biological Activities of Metallic Nanoparticles. *International Journal of Molecular Sciences*, **18**, 120.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N. and Yeh, L.-S.L. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic acids research*, **32**, 115-119.

Bokszczanin, K. (2013). Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Frontiers in Plant Science*, **4**.

Cha, J.-Y., Ahn, G., Kim, J.Y., Kang, S.B., Kim, M.R., Su'udi, M., Kim, W.-Y. and Son, D. (2013). Structural and functional differences of cytosolic 90-kDa heat-shock proteins (Hsp90s) in Arabidopsis thaliana. *Plant Physiology and Biochemistry*, **70**, 368–373.

De Bodt, S., Hollunder, J., Nelissen, H., Meulemeester, N. and Inzé, D. (2012). CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations*. New Phytologist*, **195**, 707–720

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, **38**, 64–70.

Duan, Y., Ma, Y., Zhao, X., Huang, R., Su, R., Qi, W. and He, Z. (2018). Real-time adsorption and action of expansin on cellulose. *Biotechnology for Biofuels*, **11**.

Garcia-Hernandez, M., Berardini, T., Chen, G., Crist, D., Doyle, A., Huala, E., Knee, E., Lambrecht, M., Miller, N., Mueller, L., Mundodi, S., Reiser, L., Rhee, S., Scholl, R., Tacklind, J., Weems, D., Wu, Y., Xu, I., Yoo, D., Yoon, J. and Zhang, P. (2002). TAIR: a resource for integrated Arabidopsis data. *Functional & Integrative Genomics*, **2**, 239–253.

Gil-Mascarell, R., Lopez-Coronado, J.M., Belles, J.M., Serrano, R. and Rodriguez, P.L. (1999). The Arabidopsis HAL2-like gene family includes a novel sodium-sensitive phosphatase. *The Plant Journal*, **17**, 373–383.

Giorno, F., Wolters-Arts, M., Mariani, C. and Rieu, I. (2013). Ensuring Reproduction at High Temperatures: The Heat Stress Response during Anther and Pollen Development. *Plants*, **2**, 489–506.

Guo, M., Liu, J.-H., Ma, X., Luo, D.-X., Gong, Z.-H. and Lu, M.-H. (2016). The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. *Frontiers in Plant Science*, **7**.

Hinz, M., Wilson, I.W., Yang, J., Buerstenbinder, K., Llewellyn, D., Dennis, E.S., Sauter, M. and Dolferus, R. (2010). Arabidopsis RAP2.2: An Ethylene Response Transcription Factor That Is Important for Hypoxia Survival. *PLANT PHYSIOLOGY*, **153**, 757–772

Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L. and Meyerowitz, E.M. (2004). The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCYTELESS. *Nature*, **430**, 356–360.

Krizek, B.A., Prost, V. and Macias, A. (2000). AINTEGUMENTA promotes petal identity and acts as a negative regulator of AGAMOUS. *The Plant cell*, **12**, 1357–66.

Le Gall, H., Philippe, F., Domon, J.-M., Gillet, F., Pelloux, J. and Rayon, C. (2015). Cell Wall Metabolism in Response to Abiotic Stress. *Plants*, **4**, 112–166.

Li, D.-D., Xue, J.-S., Zhu, J. and Yang, Z.-N. (2017). Gene Regulatory Network for Tapetum Development in Arabidopsis thaliana. *Frontiers in Plant Science*, **8**

Ma, H. (2005). Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annual review of plant biology*, **56**, 393–434.

Meiri, D., Tazat, K., Cohen-Peer, R., Farchi-Pisanty, O., Aviezer-Hagai, K., Avni, A. and Breiman, A. (2009). Involvement of Arabidopsis ROF2 (FKBP65) in thermotolerance. *Plant Molecular Biology*, **72**, 191–203.

Miao, Y., Lv, D., Wang, P., Wang, X.-C., Chen, J., Miao, C. and Song, C.-P. (2006). An Arabidopsis Glutathione Peroxidase Functions as Both a Redox Transducer and a Scavenger in Abscisic Acid and Drought Stress Responses. *Plant Cell*, **18**, 2749–2766.

Milla, M.A.R., Maurer, A., Huete, A.R. and Gustafson, J.P. (2003). Glutathione peroxidase genes in Arabidopsis are ubiquitous and regulated by abiotic stresses through diverse signaling pathways. *The Plant Journal*, **36**, 602–615.

Mishkind, M., Vermeer, J.E.M., Darwish, E. and Munnik, T. (2009). Heat stress activates phospholipase D and triggers PIP2accumulation at the plasma membrane and nucleus. *The Plant Journal*, **60**, 10–21.

Mortimer, J.C., Miles, G.P., Brown, D.M., Zhang, Z., Segura, M.P., Weimar, T., Yu, X., Seffen, K.A., Stephens, E., Turner, S.R. and Dupree, P. (2010). Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. *Proceedings of the National Academy of Sciences*, **107**, 17409–17414.

Müller, M. and Munné-Bosch, S. (2015). Ethylene Response Factors: A Key Regulatory Hub in Hormone and Stress Signaling. *Plant Physiology*, **169**, 32–41.

Nishizawa-Yokoi, A., Nosaka, R., Hayashi, H., Tainaka, H., Maruta, T., Tamoi, M., Ikeda, M., Ohme-Takagi, M., Yoshimura, K., Yabuta, Y. and Shigeoka, S. (2011). HsfA1d and HsfA1e Involved in the Transcriptional Regulation of HsfA2 Function as Key Regulators for the Hsf Signaling Network in Response to Environmental Stress. *Plant and Cell Physiology*, **52**, 933–945.

Parish, R.W. and Li, S.F. (2010). Death of a tapetum: A programme of developmental altruism. *Plant Science*, **178**, 73–89.

Prasinos, C., Krampis, K., Samakovil, D., Hatzopoulos, P. (2005). Tight regulation of expression of two Arabidopsis cytosolic Hsp90 genes during embryo development. *Journal of Experimental Botany*, 56, 633–644.

Ratcliffe, O.J. (2003). Analysis of the Arabidopsis MADS AFFECTING FLOWERING Gene Family: MAF2 Prevents Vernalization by Short Periods of Cold. *Plant Cell*, **15**, 1159–1169.

Rieu, I., Twell, D. and Firon, N. (2017). Pollen Development at High Temperature: From Acclimation to Collapse. *Plant Physiology*, **173**, 1967–1976

Soderman, E., Mattsson, J. and Engstrom, P. (1996). The Arabidopsis homeobox gene ATHB-7 is induced by water deficit and by abscisic acid. *The Plant Journal*, **10**, 375–381.

Sun, W., Bernard, C., van de Cotte, B., Van Montagu, M. and Verbruggen, N. (2001). At-HSP17.6A, encoding a small heat-shock protein in Arabidopsis, can enhance osmotolerance upon overexpression. *The Plant journal : for cell and molecular biology*, **27**, 407–15.

Suzuki, N., Bajad, S., Shuman, J., Shulaev, V. and Mittler, R. (2008). The Transcriptional Co-activator MBF1c Is a Key Regulator of Thermotolerance inArabidopsis thaliana. *Journal of Biological Chemistry*, **283**, 9269–9275.

Swindell, W.R., Huebner, M. and Weber, A.P. (2007). Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC Genomics*, **8**,125.

Thomas, P.D. (2003). PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research*, **13**, 2129–2141.

Volkov, R.A., Panchuk, I.I., Mullineaux, P.M. and Schöffl, F. (2006). Heat stress-induced H2O2 is required for effective expression of heat shock genes in Arabidopsis. *Plant Molecular Biology*, **61**, 733–746.

Waese, J. and Provart, N.J. (2016). The Bio-Analytic Resource for Plant Biology. *Methods in Molecular Biology*, 119–148

Waters, E.R., Nguyen, S.L, Eskandar, R., Behan, J., Sanders-Reed, Z. (2008). The recent evolution of a pseudogene: diversity and divergence of a mitochondria-localized small heat shock protein in *Arabidopsis thaliana. Genome,* **51**, 177-186

Wu, H.-C., Bulgakov, V.P. and Jinn, T.-L. (2018). Pectin Methylesterases: Cell Wall Remodeling Proteins Are Required for Plant Response to Heat Stress. *Frontiers in Plant Science*, **9**.

Wunderlich, M., Groß-Hardt, R. and Schöffl, F. (2014). Heat shock factor HSFB2a involved in gametophyte development of Arabidopsis thaliana and its expression is controlled by a heat-inducible long non-coding antisense RNA. Plant Molecular Biology, **85**, 541–550.

Yu, S.-X., Feng, Q.-N., Xie, H.-T., Li, S. and Zhang, Y. (2017). Reactive oxygen species mediate tapetal programmed cell death in tobacco and tomato. *BMC Plant Biology*, **17**

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I.A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., Peng, S., Peñuelas, J., Ruane, A.C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z. and Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences*, **114**, 9326–9331.

Zheng, S.-Z., Liu, Y.-L., Li, B., Shang, Z.- lin, Zhou, R.-G. and Sun, D.-Y. (2011). Phosphoinositide-specific phospholipase C9 is involved in the thermotolerance of Arabidopsis. *The Plant Journal*, **69**, 689–700.

Websites and non-scientific refernces

Oliveros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html