

## Article

# Transcriptomic Insight into the Pollen Tube Growth of *Olea europaea* L. subsp. *europaea* Reveals Reprogramming and Pollen-Specific Genes Including New Transcription Factors

Amanda Bullones <sup>1,2</sup>, Antonio Jesús Castro <sup>3</sup>, Elena Lima-Cabello <sup>3</sup>, Noe Fernandez-Pozo <sup>2</sup>,  
Rocío Bautista <sup>4</sup>, Juan de Dios Alché <sup>3,5</sup> and Manuel Gonzalo Claros <sup>1,2,6,7,\*</sup>

- <sup>1</sup> Department of Molecular Biology and Biochemistry, Universidad de Málaga, 29010 Malaga, Spain; amandabullones@uma.es
- <sup>2</sup> Institute for Mediterranean and Subtropical Horticulture “La Mayora” (IHSM-UMA-CSIC), 29010 Malaga, Spain; noe.fernandez.pozo@csic.es
- <sup>3</sup> Plant Reproductive Biology and Advanced Imaging Laboratory (BReMAP), Estación Experimental del Zaidín (EEZ-CSIC), 18008 Granada, Spain; antoniojesus.castro@eez.csic.es (A.J.C.); elena.lima@eez.csic.es (E.L.-C.); juandedios.alche@eez.csic.es (J.d.D.A.)
- <sup>4</sup> Plataforma Andaluza de Bioinformática, Supercomputing and Bioinnovation Center (SCBI), Universidad de Málaga, 29590 Malaga, Spain; rociobm@uma.es
- <sup>5</sup> University Institute of Research on Olive Grove and Olive Oils (INUO), Universidad de Jaén, 23071 Jaen, Spain
- <sup>6</sup> CIBER de Enfermedades Raras (CIBERER) U741, 29071 Malaga, Spain
- <sup>7</sup> Institute of Biomedical Research in Málaga (IBIMA), IBIMA-RARE, 29010 Malaga, Spain
- \* Correspondence: claros@uma.es or gonzalo.claros@ihsm.uma-csic.es; Tel.: +34-952-137-284

**Abstract:** The pollen tube is a key innovation of land plants that is essential for successful fertilisation. Its development and growth have been profusely studied in model organisms, but in spite of the economic impact of olive trees, little is known regarding the genome-wide events underlying pollen hydration and growth in this species. To fill this gap, triplicate mRNA samples at 0, 1, 3, and 6 h of in vitro germination of olive cultivar Picual pollen were analysed by RNA-seq. A bioinformatics R workflow called *RSeqFlow* was developed contemplating the best practices described in the literature, covering from expression data filtering to differential expression and clustering, to finally propose hub genes. The resulting olive pollen transcriptome consisted of 22,418 reliable transcripts, where 5364 were differentially expressed, out of which 173 have no orthologue in plants and up to 3 of them might be pollen-specific transcription factors. Functional enrichment revealed a deep transcriptional reprogramming in mature olive pollen that is also dependent on protein stability and turnover to allow pollen tube emergence, with many hub genes related to heat shock proteins and F-box-containing proteins. Reprogramming extends to the first 3 h of growth, including processes consistent with studies performed in other plant species, such as global down-regulation of biosynthetic processes, vesicle/organelle trafficking and cytoskeleton remodelling. In the last stages, growth should be maintained from persistent transcripts. Mature pollen is equipped with transcripts to successfully cope with adverse environments, even though the in vitro growth seems to induce several stress responses. Finally, pollen-specific transcription factors were proposed as probable drivers of pollen germination in olive trees, which also shows an overall increased number of pollen-specific gene isoforms relative to other plants.

**Keywords:** germination; olive; pollen; pollen tube; RNA-seq; transcriptomics; bioinformatics; reprogramming



**Citation:** Bullones, A.; Castro, A.J.; Lima-Cabello, E.; Fernandez-Pozo, N.; Bautista, R.; Alché, J.d.D.; Claros, M.G. Transcriptomic Insight into the Pollen Tube Growth of *Olea europaea* L. subsp. *europaea* Reveals Reprogramming and Pollen-Specific Genes Including New Transcription Factors. *Plants* **2023**, *12*, 2894. <https://doi.org/10.3390/plants12162894>

Academic Editor: Andrew McCubbin

Received: 13 July 2023

Revised: 1 August 2023

Accepted: 2 August 2023

Published: 8 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Flowering plants generate pollen grains that germinate on the stigma and produce a pollen tube (PT)—an evolutionary innovation that is one of the fastest-growing plant-cell-derived structures—to securely deliver sperm cells into the maternal reproductive tissues

toward the ovules [1,2]. Pollen tube growth is, therefore, essential for successful fertilisation, becoming of major importance from an agronomical viewpoint. Its development starts once the pollen grain lands on the stigma, where both surfaces are recognised by means of glycosylated proteins. When both surfaces are compatible, reactive oxygen species (ROS) of the stigma are decreased to promote pollen hydration [3] and PT emergence through the pollen grain aperture after 30–45 min [4,5]. The PT grows very rapidly through the style towards the ovules in an energy-demanding process where apoplasmic sucrose accumulated in mature pollen is consumed without any need for sucrose uptake [6]. Only the PT tip enlarges [7], depending on cytoskeleton reorganisation for organelle and vesicle trafficking [8] in a profilin-modulated assembly of actin microfilaments [9] with cell wall biosynthesis at the tip, generation of ROS and nitric oxide, pH homeostasis, energy metabolism, protein folding and degradation (reviewed in [1]),  $\text{Ca}^{2+}$  gradient and other ion fluxes, together with aquaporins-dependent water transport, to avoid osmotic stress in coordination with ion transporters in the trans-Golgi network and plasma membrane proton pumps [10]. An increase in autophagy due to the high metabolic demand of PT growth was recently reported [11]. Flavonoids participate in successful pollen germination by increasing the phenylpropanoid pathways [12], ROS scavenging and maintenance of turgor and normal morphology [13,14]. Morphology of PTs depends on ions and water filling the new volume [15,16] and the cell wall that sustains turgor pressure [17,18].

The rapid growth of PT provokes the misfolding of many proteins after their fast synthesis in ribosomes, which requires (i) additional expression of heat shock proteins (HSPs) to stabilise them and, for failing cases, (ii) an increase in some ubiquitination and proteasome components undetectable in mature pollen grains [19]. HSPs are also shielding climate change effects that are exacerbating the intensity, frequency and extremity of heat events, with non-reversible consequences for crops due to the substantial decrease of pollen development, germination, and PT elongation [20]. Plants produce many more small HSPs compared to other eukaryotic organisms, including the HSP20 proteins (presenting a significant homology to eye lens  $\alpha$ -crystallins) responsible for heat stress endurance, desiccation tolerance, and lipid stabilisation at the plasma membrane level [21,22]. Interestingly HSP20s accumulate in dry seeds, rapidly disappearing upon germination [22]. F-box proteins (FBPs) belong to one of the families of ubiquitin-ligating enzymes E3, one of the largest and most diverse protein superfamilies found in plants [23]. For example, nearly 700 FBPs have been predicted in *Arabidopsis* [23,24], a pool approximately 10 times larger than in the human genome. They recognise different substrates for ubiquitination, exerting direct control over a wide range of cellular processes [25,26]. Most FBPs are located in pollen, although their function does not seem to be related to it [24] except the pollen-pistil self-incompatibility mechanism to augment genetic variability [27].

Pollen germination requires an overall increase in gene transcription [4,19] that has attracted scarce research, in contrast to the abundant information about pollen development and maturation. Genomics approaches reported that at least 30% of the plant genome could be involved in this process, with a higher proportion of selectively expressed genes than vegetative tissues [28] as a result of its high specialisation. Transcription, ubiquitin-mediated proteolysis (involving up to 54% of the FBP genes), cell cycle progression, and defence/stress response were critical in the transition from mature pollen to the germinated pollen [28,29]. Moreover, PT growth depends on protein synthesis but is relatively independent of transcription [29], which is now explained by the enucleation of the apical region by the callose plug that drives growth progression without de novo transcription using persistent transcripts [30].

Most pollen research has been conducted on model systems, but there is still a place for non-model systems such as olive tree (*Olea europaea* L. subsp. *europaea*), where research on floral biology has been principally horticultural in orientation [31], and relevant and high-throughput technologies have been applied recently [32–37]. Olive is an emblematic and dominant crop in the Mediterranean basin whose relevance is related to the high agricultural and nutraceutical value of the fruits (olives), the landscape shaping, and more

recently, its use as a source for renewable biomaterials and fuel [38]. In fact, it is the only species within the genus *Olea* that produces edible fruits [39]. Genome sequencing confirmed that the olive tree has a medium-sized genome that ranges from 1.3 to 1.68 Gb, distributed in 23 chromosomes, but with different gene content (compiled in [40,41]). In our laboratory, we developed ReprOlive (the first database of olive reproductive tissues [32]) that was the precursor of our recent OliveAtlas gathering all transcriptomic experiments performed on Picual cultivar published to date [37].

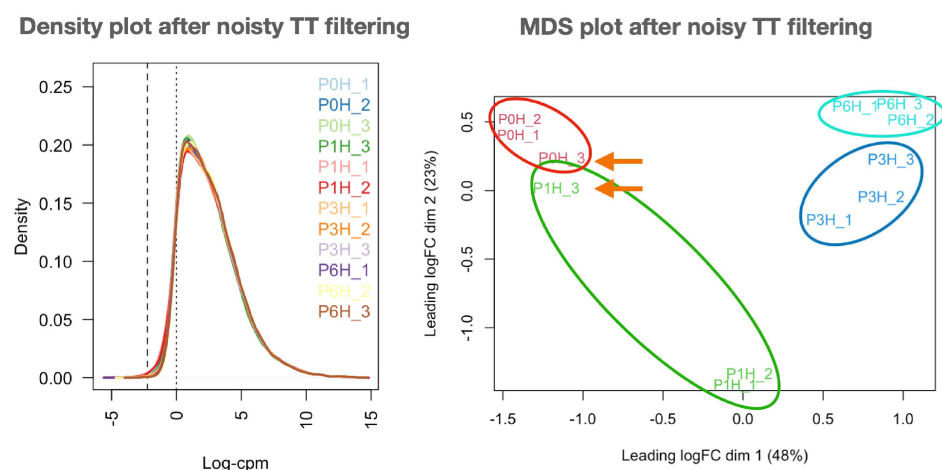
The olive pollen belongs to the big group of angiosperms that undergo mitosis II after anthesis and produces a less differentiated, longer-lived bicellular pollen, whose generative cell divides mitotically in the PT [31,42]. This division represents an important developmental transition, and it is expected to be associated with the removal of a significant part of ribosomes and transcripts, as described for other plants [1]. Messenger RNA breakdown is likely to be initiated by mRNA deadenylation accompanied by effective protein turnover and the accumulation of transcripts associated with the ubiquitin-proteasome pathway (such as FPBs of the E3 ligases mentioned above) [12]. Gene expression changes during PT growth and pollination are governing the fruit set and olive oil production [43], but the only high-throughput approach to the olive pollen was focused on its interaction with stigma [44]. To gain insights into the biological processes involved in pollen germination and PT growth, here we present the first comprehensive study of the *in vitro* transcriptome dynamics of olive PT growth during four sequential developmental stages using the genome draft of cultivar Picual [45] as a reference. The automated R pipeline *RSeqFlow* was developed, revealing a pollen transcriptome of 22,418 genes, with 5364 differentially expressed genes (DEGs) during the process. New pollen-specific genes were identified, including transcription factors, as well as proofs of cell reprogramming by transcription regulation and protein modifications mediated by HSPs and F-box proteins. Changes in responses to stresses and the final male gamete development at 6 h were also observed.

## 2. Results

### 2.1. Filtered Datasets Revealed the Olive Pollen Transcriptome

RNA from samples in triplicate from germinated pollen at 0, 1, 3, and 6 h of culture were sequenced, providing 38.1–47.7 million paired-reads (samples P3H\_2 and P6H\_3, respectively) that, once pre-processed, provided a high quantity of useful reads (83.47–95.67%; Appendix A.1). Each library was then mapped to the 81,484 transcripts of the ‘Picual’ genome (Oleur v0.61) and to the 50,681 transcripts of the *Olea europaea* var. *sylvestris* transcriptome (Oeu v1). Results in Appendix A.1 show that mapping onto the ‘Picual’ transcriptome is, as expected, better (69.61–76.06%) than to the wild olive.

Mapping counts were then loaded into *RSeqFlow* to (i) filter unexpressed and low-expressed genes, (ii) normalise the reliable genes, and (iii) inspect the replicates. Up to 24,667 transcripts (30%) were never expressed in any pollen sample, 33,708 transcripts (41%) were somewhat expressed in all samples, and the consistent olive pollen transcriptome comprised 22,418 (28%) reliably expressed transcripts (three replicates with  $CPM > 1$ ; their IDs and expression counts are in `filteredData- $\{$ DATETIME $\}$ .tsv` (Appendix A.2). Libraries devoid of noisy transcripts follow a nearly identical gauss-shaped profile, and samples were grouped as expected (Figure 1; compared to raw data (see *Load data* → *Inspect raw data* of `Report.html` file in Appendix A.2) even though samples P0H\_3 and P1H\_3 appeared mixed in raw data. The normalisation of filtered datasets maintains the same sample grouping (see plots in *Normalisation* → *Sample similarities* → *Sample MDS and PCA* and *Raw sample similarities* → *Grouping raw samples* of `Report.html` file in Appendix A.2) as reflected by the decrease of dispersion: BCV (biological coefficient of variation) decreased from 0.1792 in raw data to 0.1475 in normalised data. The MDS plot also revealed the closeness of samples collected at 3 and 6 h of growth. Consequently, filtered, normalised data are appropriate for further gene expression studies.



**Figure 1.** Improved distribution of libraries devoid of noisy transcripts. All libraries follow a similar Gaussian profile (**density plot** at (**left**)) and replicates can be grouped as expected (**MDS plot** at (**right**)). Both plots are provided in Appendix A.2 and modified for convenience. Sample names in plots are as in Appendix A.1. Samples from different replicates that are surprisingly close are marked by an arrow.

## 2.2. Differential Gene Expression during Olive Pollen Germination

Heteroscedasticity of the 22,418 useful transcripts comprising the reliable olive pollen transcriptome was then removed in *RSeqFlow* (saved as `normHomoscedCPM- $\{DATETIME\}$ .tsv` in Appendix A.2) before applying a generalised linear model in the seek of differentially expressed genes (DEGs). Although the Bayesian approach with *eBayes* function is the most common method, *RSeqFlow* is designed to use the *treat* function that yields more biologically relevant genes and produces fewer false positives. Table 1 confirmed that *treat* provided 2.63–7.9 fold fewer DEGs than *eBayes* since the latter includes additional 9120 genes (see the Venn diagram of *Differential Expression* → *Saving DEGs* → *Per contrasts* in Report.html file in Appendix A.2). Up to 54.6% of pollen transcripts may be DEGs using *eBayes*, suggesting an oversize issue.

**Table 1.** Number of DEGs per contrast depending on *eBayes* and *treat* determinations.

DEGs	0→1 h	1→3 h	3→6 h	1→6 h	0→3 h	0→6 h
<i>eBayes</i> <sup>a</sup>	6715	4776	42	10,348	4015	7087
down	3662	2297	4	4716	1830	3115
up	3053	3479	38	5632	2185	3972
<i>treat</i> <sup>b</sup>	850 (76)	910 (115)	10 (2)	3929 (429)	948 (133)	2397 (305)
down	561 (49)	119 (13)	0 (0)	1144 (71)	280 (28)	745 (85)
up	289 (27)	791 (102)	10 (2)	2785 (358)	668 (105)	1652 (220)
Ratio <i>eBayes/treat</i>	7.90	5.25	4.2	2.63	4.23	2.96

<sup>a</sup> All *eBayes* DEGs can be found at file `AllGenes_allContrast_eB_ $\{P-cutoff\}$ _ $\{FC-cutoff\}$ _ $\{DATETIME\}$ .tsv` in Appendix A.2. <sup>b</sup> The numbers within parentheses are differentially expressed TFs. All *treat* DEGs can be found at file `AllGenes_allContrast_TREAT_ $\{P-cutoff\}$ _ $\{FC-cutoff\}$ _ $\{DATETIME\}$ .tsv` in Appendix A.2.

After contrasting all-against-all samples (Table 1), the first notable finding was that the 0→1 h transition involved a remarkable down-regulation, while up-regulation occurred for the remaining transitions. In addition, the closeness of samples P0H\_3 and P1H\_3 has not introduced any concerning distortion in DEGs, probably due to the results of the *treat* function. The negligible number of DEGs (10, all up-regulated) for the 3→6 h transition is consistent with the MDS plot in Figure 1 and suggests that most gene expression changes occurred within 3 h of culture. Three genes were DEGs in all contrasts only when *eBayes* was used (see `ubiquitousDEGs_ $\{DATETIME\}$ .tsv` file in Appendix A.2), corresponding to

Oleur061Scf3360g05017.1 (a poly(A) binding protein), Oleur061Scf3133g07005.1 (a cyclin-like protein) and Oleur061Scf0430g00024.1 (a probable aquaporin).

Tentative transcription factors (TFs) were identified within DEGs as candidates for regulators of the studied process. Using the 2741 predicted TFs in 57 families for *Olea europaea* var. *sylvestris* in PlantTFDB v5.0, 12,095 tentative TFs were identified in 'Picual' (Appendix A.3), 608 of them (5%, see Appendix A.4) distributed in the different contrasts (numbers within parentheses in Table 1). Because *A. thaliana* only contains 1726 TFs in PlantTFDB, it is clear that the olive tree may contain more TFs, even though the number of TFs in cv. Picual could be an overestimation due to the presence of paralogues or alleles as separate transcripts. However, it is worth mentioning that (i) the STAT family was not represented in wild olive, (ii) the putative 608 TFs in DEGs are distributed in 46 TF families (Appendix A.4), with no members in families BBR-BPC, GeBP, HRT-like, LBD, LFY, RAV, S1Fa-like, SAP, SRS, YABBY and ZF-HD, and (iii) the first TF families in Appendix A.4 having >5.0% of its members within DEGs can be considered enriched relative to the total number of TFs in cv. Picual.

### 2.3. New PT-Specific DEGs

The TAIR10 annotations of cv. Picual transcriptome v0.61 was extended (Appendix A.5) so that 93.7% of DEGs were annotated, 91.6% with a functionally characterised orthologue (Appendix A.5). However, 173 DEGs (Appendix A.5) were not annotated in any reference database, which makes them candidates for olive genes specifically involved in pollen hydration and PT growth. The possibility of being artefactual transcripts was discarded since (i) 44 of them were already identified as pollen-specific genes in ReprOlive [32] (19 of them with a distant orthologue in UniProt); (ii) 12 were present in other reproductive tissues; and (iii) three of them were likely TFs orthologous to the *Olea europaea* var. *sylvestris* TFs predicted in PlantTFDB v5.0 (Appendix A.6). Two of these new TFs, Oleur061Scf0040g05012.1 (up-regulated in the 1→6 h transition) and Oleur061Scf2615g01010.1 (up-regulated in the 0→1 h transition), belong to DEGs identified in ReprOlive pollen transcriptome, corresponding to members of the CPP and bHLH families, respectively. The third one, Oleur061Scf1950g00010.1 (down-regulated in the 1→6 h transition), is somewhat homologous to the CO-like family but was not present in the ReprOlive database.

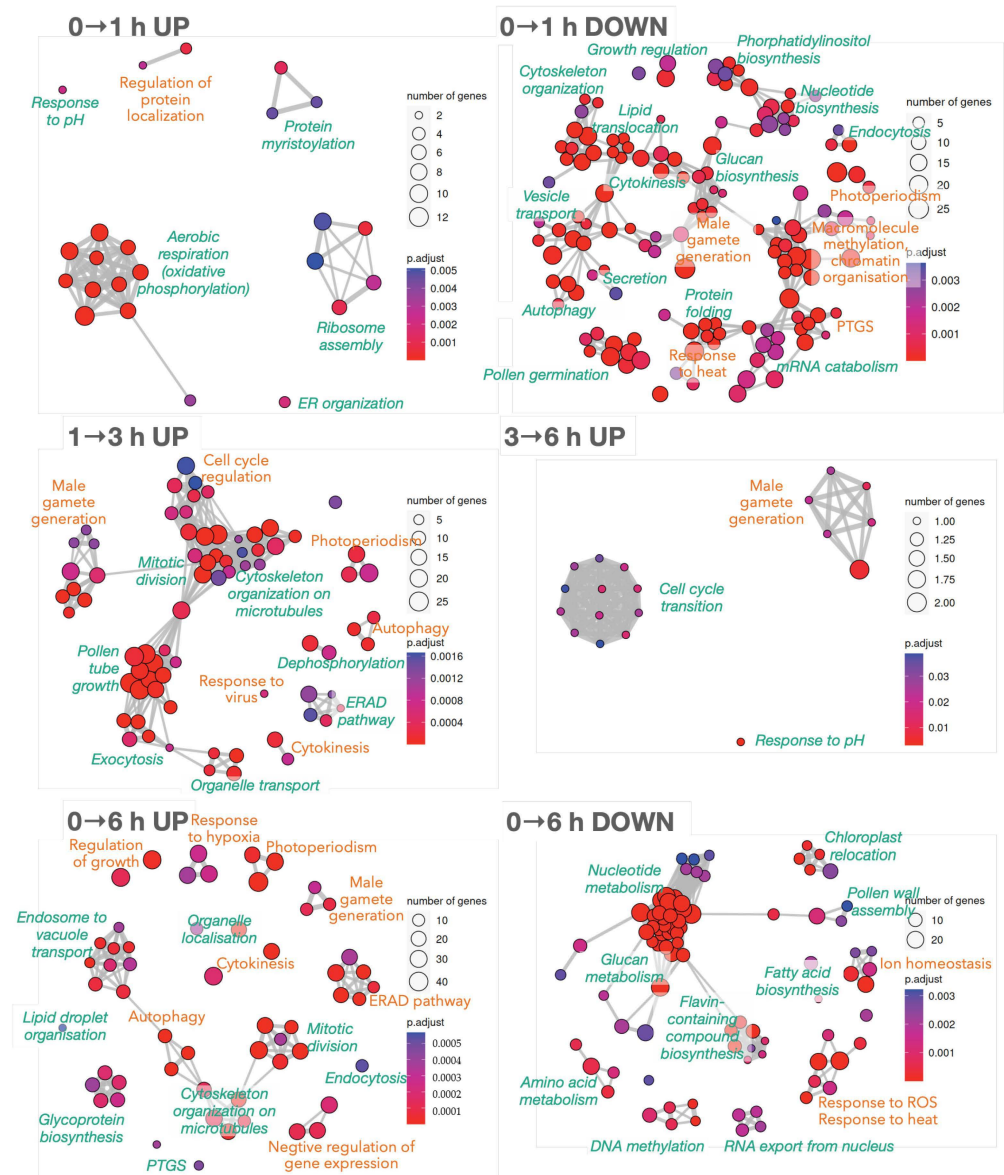
### 2.4. Functional Enrichment during PT Growth

The relevant cellular functions affected in 0→1 h, 1→3 h, 3→6 h and 0→6 h transitions in Table 1 were then analysed using the *A. thaliana* TAIR10 orthologue in Appendix A.5. To circumvent a possible confounding result caused by the different number of genes between both species, TAIR10 IDs present both in up- and down-regulated DEGs for the same contrast (2 DEGs in the 0→1 h transition, 3 DEGs in the 1→3 h transition and 25 DEGs in the 0→6 h transition) were removed before functional analyses. GO enrichment focused on the Biological Process aspect was represented as a network (Figure 2) where processes are nodes and the edges are supported by shared DEGs pointing to that process. Identification of functional modules (green and orange labels in Figure 2) was easy because mutually overlapping gene sets tend to cluster together.

Because many processes in Figure 2 do not contain TFs, transcriptional and post-translational regulations must be concomitantly involved in PT growth. When TFs were inspected in the same contrasts (Appendix A.7), NAC, MYB, and, to a lesser extent, ERF, HSF, and bHLH seemed relevant because they are families containing abundant TFs (Appendix A.4). Over-represented families were, for example, bZIP, C3H, and C2H2.

Deep expression changes were observed in the 0→1 h transition (0→1 h UP panel of Figure 2; already advanced in Table 1). Up-regulated processes are consistent with the activation of 'oxidative phosphorylation' and 'mRNA surveillance' KEGG pathways (Appendix A.9) and 'ribosome synthesis' and 'oxidative phosphorylation' pathways decrease in the 1→6 h transition (BP-0E\_p\_6H.p\_1H\_down.tsv file in Appendix A.8). Down-regulated processes within the first hour of pollen culture, such as 'PTGS (post-transcriptional

gene silencing), 'autophagy', 'endocytosis', 'cytoskeleton organisation', 'cytokinesis', or 'male gamete generation' (0→1 h DOWN panel in Figure 2) will be recovered in later stages, corroborating the deep reprogramming. Other specific processes disappear within the first hour of hydration ('pollen germination', 'mRNA catabolism', 'demethylation' of histones and DNA), together with related KEGG pathways ['biosynthesis of amino acids and cofactors', 'carbon metabolism' including 'glycolysis/gluconeogenesis', 'endocytosis', 'protein processing at endoplasmic reticulum', and 'inositol phosphate pathway' and its signalling system (Appendix A.9)].



**Figure 2.** Biological Process enrichment analysis for the DEGs obtained for 0→1 h, 1→3 h, 3→6 h and 0→6 h transitions of germinating pollen to reveal the main changes from mature pollen to full PT growth in vitro. UP refers to processes up-regulated from the first time to the second, and DOWN refers to down-regulated processes when advancing from the first time to the second. Green labels close to clusters identify the functional module; labels in orange mark processes containing TFs (Appendix A.7). Node sizes reflect the number of DEGs supporting the biological process. Blue-to-red colour reflects the significance as an adjusted  $p$ -value. The complete list of biological processes can be found in Appendix A.8.

The 1→3 h transition step is a pure PT-growth stage, and the functional analysis only showed up-regulated processes (1→3 h UP panel in Figure 2), including many of the processes previously down-regulated in the 0→1 h transition mentioned above. Others, such as 'mitotic division' and 'cytokinesis', 'male gamete generation', 'pollen tube growth', and 'cell (cycle) regulation', may be specifically required for PT growth and differentiation of male gametes. The fact that only 10 up-regulated DEGs in the last stage (3→6 h transition; Table 1) are related to 'male gamete generation', 'response to pH' (that was already activated in the 0→1 h transition period), and the 'cell cycle transition' involving phosphorylation and cyclin-dependent kinases (3→6 h UP panel in Figure 2) suggests that further growing will depend only on persistent transcripts. This is confirmed by the up-regulation of 'RNA degradation' and 'mRNA surveillance' KEGG pathways (also up-regulated in the 0→1 h transition), also indicating that the required set of transcripts was nearly constant after the 3 h of pollen germination. Moreover, the last mitotic division to obtain the reproductive male gametophytes and their transport on microtubules seems to occur in the 1→3 h transition stage of PT growth.

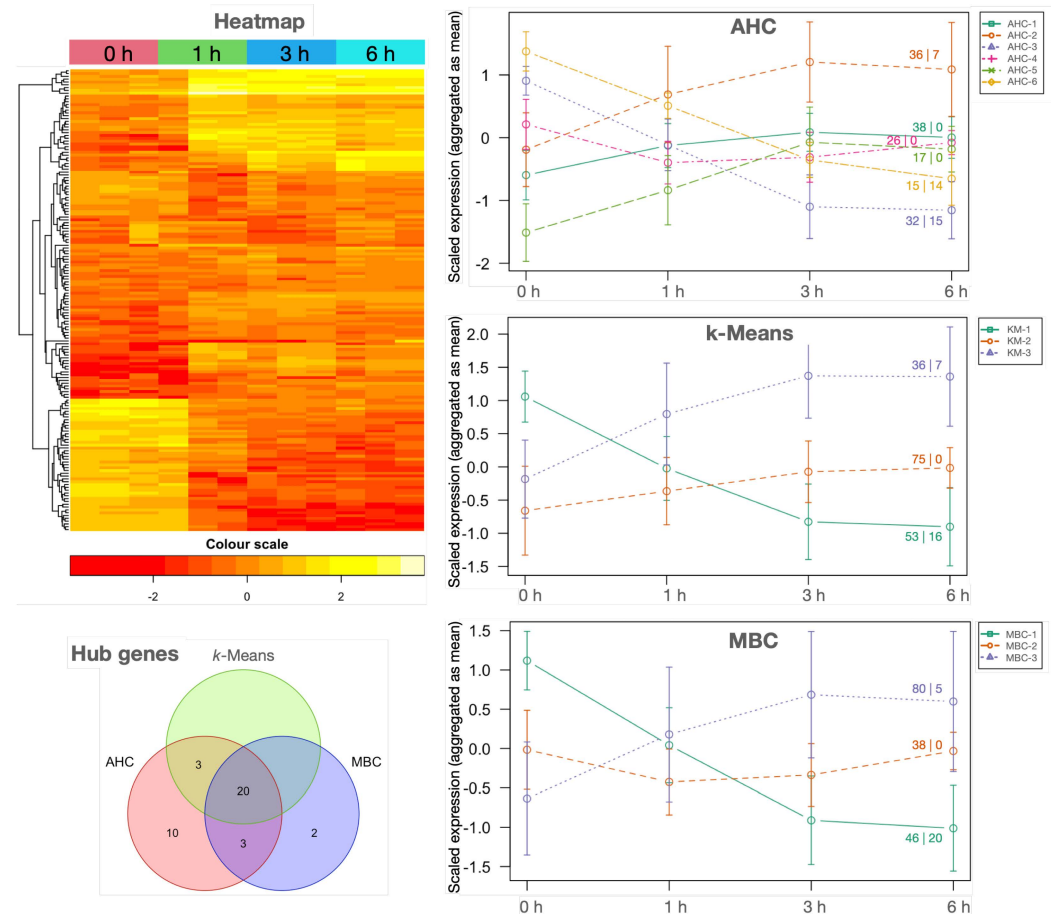
To obtain an overview of the process, the 0→6 h germination period was inspected (0→6 h UP and DOWN panels at the bottom of Figure 2). It was confirmed the up-regulation of the main expected processes of PT growth ('endocytosis', 'endosome to vacuolar transport', 'organelle localisation', 'cytoskeleton organisation on microtubules', 'autophagy', 'PTGS', 'mitotic division', 'cytokinesis', 'male gamete generation', 'regulation of growth', 'protein glycosylation', 'lipid droplet organisation', and 'response to hypoxia'). More interestingly, the negative regulation of gene expression supports the down-regulation of KEGGs pathways related to the biosynthesis of amino acids, cofactors, nucleotide, fatty acids, and glucans, in agreement with down-regulated biological processes in the 0→6 h DOWN panel in Figure 2. The functional enrichment at the 0→6 h transition also showed down-regulation of processes that are no longer essential for pollen tube growth, such as 'chloroplast relocation', 'DNA methylation', 'RNA export from nucleus', 'nucleotide metabolism', 'amino acid metabolism', 'fatty acid biosynthesis', 'pollen wall assembly', 'response to heat', and 'response to ROS', combined with the increased response to other stresses (hypoxia, salt, light) and the proteasome pathway.

### 2.5. Coordinated Processes Revealed by Co-Expressed Genes

Because genes under the same regulatory control tend to be functionally related, DEGs were submitted to correlation pattern detection in *RSeqFlow*. Because 5364 DEGs outnumber the 12 samples by far, DEGs having little variant expression profiles ( $CV < 0.2$ ) were removed to avoid computational constraints, resulting in 164 DEGs (see the heatmap in Figure 3). Correlation matrices were calculated using three different, well-established clustering algorithms based on a different method to obtain the optimal number of clusters: (i) AHC (agglomerative hierarchical clustering) combined with dynamic cut-off, (ii) *k*-means included the most repeated *k* after calculating 26 metrics, and (iii) MBC (model-based clustering) exploited dynamic clustering (AHC, *k*-Means and MBC plots in Figure 3). Using AHC, three clusters described increased expression and another three a decreased one, while *k*-means and MBC gather up-regulated DEGs in one cluster, down-regulated DEGs in another cluster, and transcripts in the third only slightly increased (Figure 3). Expression profiles of MBC and *k*-means were slightly different because DEGs in each cluster were not exactly the same. Cluster profiles also corroborate that main up-and-down expression changes were consolidated within the first 3 h of growth.

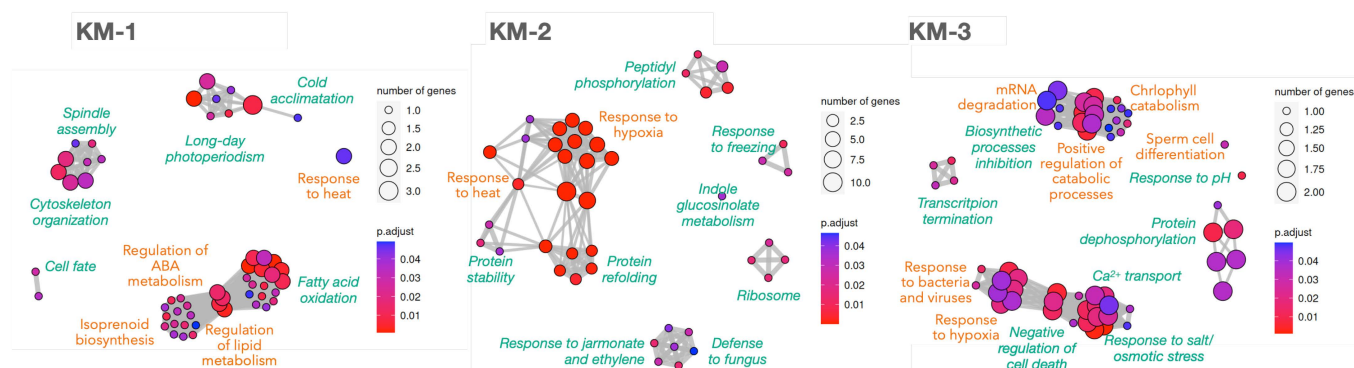
Functional enrichment analysis was performed again for the *k*-means clusters (Figure 4) because this algorithm is usually considered faster and more reliable than the others, and is preferred for large datasets. Similarly, in Figure 2, not all coordinated processes depend on TFs, corroborating that post-translational modifications may also be important to coordinate the processes. Cluster KM-1 gathers already known down-regulated processes (Figure 2) as well as new ones, such as 'ABA metabolism' to regulate lipid metabolism, 'cell fate' due to changes in the regulatory networks, and 'cytoskeleton organisation' related

to spindle assembly. KM-2 contains low-expressed DEGs (scaled mean below  $-0.5$  in the 'k-Means' plot in Figure 3) that slightly increased their levels over the PT growth period and correspond to responses to stresses mentioned above, supported by the KEGG pathway 'protein processing in ER' (Appendix A.10). It is also revealing the slight increase of ribosomes and peptidyl phosphorylation, and that defence to fungus shares DEGs with response to jasmonate and ethylene.



**Figure 3.** Individual and clustered expression profiles of the 164 highly variable DEGs along PT growth stages. Individual expression in each sample is represented as a **heatmap** (top-left plot). Scaled DEG expression profiles as aggregated medians after clustering using **AHC**, **k-Means** and **MBC** algorithms with standard deviation as error bars are the right plots; the first number on the cluster profile lines corresponds to the number of DEGs aggregated in each point of the cluster and the number after the pipe '|' indicates how many hub DEGs were identified in that cluster. A Venn diagram illustrating the coincidences of the hub DEGs is at the bottom-left (**Hub genes**). The complete list of highly variable DEGs per cluster is available in file `ClustersCTF-{DATETIME}.tsv` and the putative hubs in each cluster in `OutstandingGenes-{DATETIME}.tsv` at Appendix A.2.

Finally, up-regulated processes (KM-3 in Figure 4) confirm the findings in Figure 2, and also reveals the 'inhibition of biosynthetic processes/positive regulation of catabolism', 'Ca<sup>2+</sup> transport', 'transcript termination' (together with the 'mRNA surveillance' KEGG pathway), 'chlorophyll catabolism', and 'negative regulation of cell death'. Hence, coordinated processes based on correlated DEGs are consistent with previous knowledge, with special emphasis on response to stresses, catabolism and signalling.



**Figure 4.** Biological processes enrichment analysis for the 164 highly variable DEGs along PT growth stages after clustering using *k*-means algorithm. The cluster names **KM-1**, **KM-2** and **KM-3** correspond to the same clusters in Figure 3. Green labels close to clusters identify the functional module; labels in orange mark processes containing TFs. Node sizes reflect the number of genes supporting the biological process, as indicated in every legend. Blue-to-red colour reflects the significance as an adjusted *p*-value. The complete list of biological processes and the TAIR10 IDs supporting them can be found in Appendix A.10.

## 2.6. Hub Genes

Because there are correlated genes that can be organised in coordinated processes (Figure 4), some kind of regulation can be inferred, with especial interest on unknown and uncharacterised genes (Appendices A.5 and A.6). Key regulators should be extracted from hub gene detection implemented in *RSeqFlow* based on networks, where nodes are genes, and the edges are correlations as surrogates of co-expression (see Report .html, section ‘Networking’ at Appendix A.2), and hub genes are highly connected, central nodes. A total of 36 hub genes were proposed from AHC, 23 from *k*-means and 25 from MBC, resulting in 38 putative, unique hub genes, 20 of them being common for the three clustering algorithms (‘Hub genes’ in Figure 3). Because TFs, kinases, HSPs and FBPs are relevant groups of hub genes (Table 2), transcriptional and post-translational regulations seem to be involved in the control of olive PT growth.

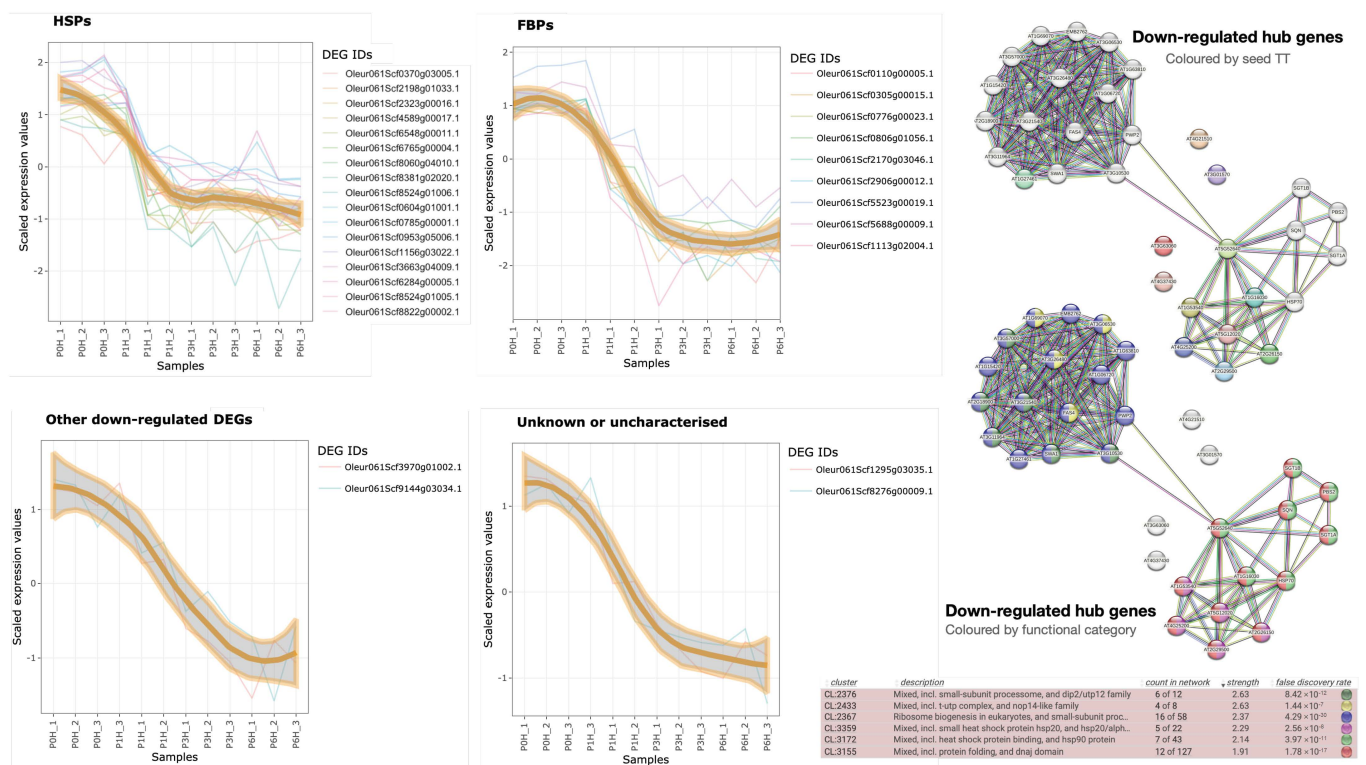
There are four down-regulated families of hub genes: HSPs, FBPs, other genes, and unknown/uncharacterised genes (Table 2 and Figure 5). The most populated one is HSPs with 16 olive genes (Table 2), most of them (13) orthologous to members of the HSP20 family, as well as two HSP90 and one HSP70. The elevated number of olive orthologues to the same HSP20 suggests that the olive genome contains more paralogues of this class of proteins than *A. thaliana*. The expression level of these olive genes falls after 1 h of germination (‘HSPs’ panel in Figure 5), suggesting their involvement in maintaining the viability of mature pollen. Because one of the members (Oleur061Scf0370g03005.1) is orthologue to a heat stress TF, it is tempting to hypothesise that this hub gene may directly or indirectly control the other members of the group.

The next populated group are FBPs (Table 2) with eight orthologues to the same F-box protein SKIP27 (a component of SCF [ASK-cullin-F-box]: an E3 ubiquitin ligase mediating ubiquitination and subsequent proteasomal degradation of target proteins), and one to EDL3, a protein involved in SCF-dependent proteasomal ubiquitin-dependent proteolysis. The parallelism with HSPs is not only evident regarding the prominence of one TAIR10 orthologue but also in the expression profile: FBP expression becomes negligible after 1 h of germination, which limits their influence to the very early stages of pollen germination.

**Table 2.** Putative hub genes detected after networking analyses of the 164 highly variable DEGs.

<b>'Picual' v0.61 ID <sup>a</sup></b>	<b>TAIR10 AC#</b>	<b>Description in OliveTreeDB v0.61</b>
<b><i>Heat shock proteins (HSPs); down-regulated</i></b>		
<sup>1</sup> Oleur061Scf2323g00016.1, <sup>1</sup> Oleur061Scf6765g00004.1, <sup>1</sup> Oleur061Scf8060g04010.1, <sup>1,3</sup> Oleur061Scf8524g01006.1, <sup>1,2</sup> Oleur061Scf0785g00001.1, <sup>1,2,3</sup> Oleur061Scf2198g01033.1, <sup>1,2,3</sup> Oleur061Scf0953g05006.1, <sup>1,2,3</sup> Oleur061Scf8524g01005.1	AT1G53540	HSP20-like (class I) chaperone
<sup>1</sup> Oleur061Scf8822g00002.1	AT2G29500	HSP20-like (class I) chaperone
<sup>1,2</sup> Oleur061Scf3663g04009.1, <sup>1,2,3</sup> Oleur061Scf6548g00011.1	AT5G12020	17.6 kDa class II heat shock protein
<sup>1,2,3</sup> Oleur061Scf4589g00017.1, <sup>1,2,3</sup> Oleur061Scf8381g02020.1	AT4G25200	Heat shock 22 kDa protein, mitochondrial
<sup>1,2,3</sup> Oleur061Scf1156g03022.1	AT1G16030	Heat shock 70 kDa protein
<sup>1</sup> Oleur061Scf6284g00005.1, <sup>1,3</sup> Oleur061Scf0604g01001.1	AT5G52640	Heat shock protein Hsp90 family
<sup>1</sup> Oleur061Scf0370g03005.1	AT2G26150	Heat shock transcription factor A2 with winged helix-turn-helix DNA-binding domain
<b><i>Proteins with F-box domain (FBPs); down-regulated</i></b>		
<sup>1</sup> Oleur061Scf5523g00019.1, <sup>3</sup> Oleur061Scf1113g02004.1, <sup>1,3</sup> Oleur061Scf0305g00015.1, <sup>1,2,3</sup> Oleur061Scf0776g00023.1, <sup>1,2,3</sup> Oleur061Scf0806g01056.1, <sup>1,2,3</sup> Oleur061Scf2170g03046.1, <sup>1,2,3</sup> Oleur061Scf2906g00012.1, <sup>1,2,3</sup> Oleur061Scf0110g00005.1	AT4G21510	F-box protein SKIP27
<sup>1,2,3</sup> Oleur061Scf5688g00009.1	AT3G63060	EID1-like F-box protein 3
<b><i>Other down-regulated DEGs; down-regulated</i></b>		
<sup>1</sup> Oleur061Scf3970g01002.1	AT4G37430	Cytochrome P450, family 91, subfamily A, polypeptide 2
<sup>1</sup> Oleur061Scf9144g03034.1	AT3G01570	Oleosin family protein
<b><i>Unknown or uncharacterised; down-regulated</i></b>		
<sup>1</sup> Oleur061Scf8276g00009.1, <sup>1,2,3</sup> Oleur061Scf1295g03035.1	AT1G27461	Uncharacterized protein LOC113713235
<b><i>Transcription factors (TFs); up-regulated</i></b>		
<sup>3</sup> Oleur061Scf2392g00020.1	AT2G01300	Putative Mediator of RNA polymerase II transcription subunit 18
<sup>1,2,3</sup> Oleur061Scf3825g03001.1	AT1G12440	Zinc finger A20/AN1 domain-containing stress-associated protein
<sup>1,2,3</sup> Oleur061Scf3317g04032.1	AT5G24590	NAC domain-containing protein 91
<sup>1,2</sup> Oleur061Scf9115g00013.1		NAC domain-containing protein
<b><i>Other up-regulated DEGs; up-regulated</i></b>		
<sup>1,2</sup> Oleur061Scf7673g01036.1	AT3G11760	C2 NT-type domain-containing protein
<sup>1,2</sup> Oleur061Scf1255g08007.1	AT1G53210	Sodium/calcium exchanger family protein
<sup>1,2,3</sup> Oleur061Scf3503g02024.1	AT3G24550	Proline-rich receptor-like protein kinase PERK1
<sup>1,2,3</sup> Oleur061Scf6158g01023.1	AT3G13060	YTH domain-containing family protein 2 putatively regulated by high glucose

<sup>a</sup>: Hub genes by AHC (<sup>1</sup>), *k*-means (<sup>2</sup>) and/or MBC (<sup>3</sup>) in OutstandingGenes-{DATETIME}.tsv file in Appendix A.2.



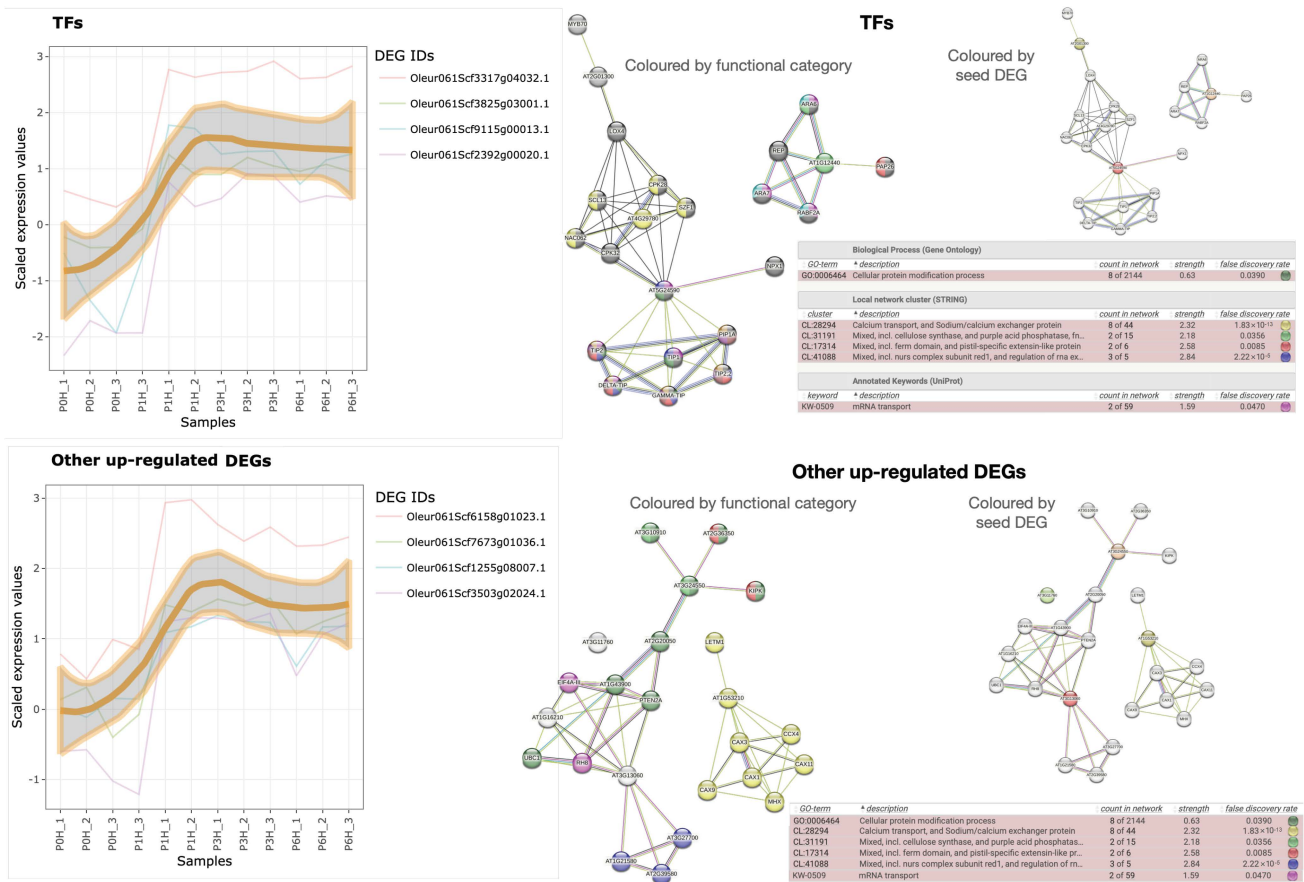
**Figure 5.** Expression profiles of the four main groups of down-regulated hub genes based on their annotation. The 12 samples were ordered in the abscissa, and the ordinate refers to the scaled transcription level. The functional network of the 30 genes, as revealed by STRING-db, is shown on the right, with input nodes in colours (**top**) or nodes coloured by relevant function (**bottom**).

Another tentative down-regulated hub genes (Table 2) are a cytochrome P450 (involved in the metabolism of glucosinolates), an oleosin (necessary to stabilise lipid bodies to prevent their coalescence), and two uncharacterised genes orthologous to the same TAIR10 ID (AT1G27461). Because their expression profiles at 3 h are higher than HSPs and FBPs, their relevance can be extended to the initial steps of pollen germination.

Using the 30 TAIR10 orthologues to down-regulated hub genes in olive (Table 2) as input for STRING-db, the significant functional network ( $p < 1.0 \times 10^{-16}$ ; ‘Down-regulated hub genes’ networks in Figure 5) indicates that oleosin, cytochrome P450, and the two TAIR10 orthologues to FBPs are not linked to any other process. However, the tentative HSPs are tightly linked to each other, with the HSP90 (AT5G52640) linking the protein folding process with the highly connected ribosome biogenesis network established by AT1G27461, the TAIR10 orthologue to both ‘unknown or uncharacterised’ genes. HSPs were also related to the 18S pre-rRNA processome.

There are only eight up-regulated olive hub genes that reach the maximum after 1 h of germination, while their transcription level is negligible in mature pollen, which strongly relates them to the PT growth. They are classified among ‘transcription factors’ and ‘other up-regulated DEGs’ (last two blocks of Table 2, and Figure 6). Unfortunately, the four TFs are orthologue to uncharacterised TFs in *A. thaliana*. One has a Zn-finger domain, two probably contain NAC domains, and the fourth may be one subunit of the Mediator co-activation complex. Using the TAIR10 orthologues to seed STRING-db (‘TFs’ panel in Figure 6), they seem significantly involved in calcium transport and  $\text{Ca}^{2+}/\text{Na}^{+}$  exchange ( $p < 1.83 \times 10^{-13}$ ), and regulation or RNA export from nucleus ( $p < 2.22 \times 10^{-5}$ ). Additionally, it can be observed that (i) the putative Mediator subunit may be related to auxin response due to its association with MYB70, as well as to pest resistance and senescence or responses to wounding by means of LOX4; (ii) the Zn-finger-containing protein might be involved in GTPase-mediated signal transduction located in

endocytotic vesicles; and (iii) the TAIR orthologue to the NAC domain-containing protein *Oleu061Scf3317g04032.1* is expressed in response to pathogens and hypoxia based on their GO terms, and the STRING-db also relates it to the TF NPX1 controlled by ABA, with plausible participation in water transport to central vacuole mediated by TIP and PIP proteins (that may also be related to *Oleu061Scf0430g00024.1*, the DEG corresponding to the aquaporin TIP2-1). Hence, these uncharacterised TFs may be controlling the some processes of pollen germination.



**Figure 6.** Expression profiles of the main two groups of up-regulated hub genes based on their annotation. The 12 samples are ordered in *x* axis, and the *y* axis refers to the scaled transcription level. Separate functional networks of the 4 TFs and the other up-regulated 4 DEGs, as revealed by STRING-db, are shown on the right, with nodes coloured by relevant functions (big, middle network) or input nodes in colours (small, right network).

The last four olive genes (‘Other up-regulated DEGs’ panel in Table 2 and Figure 6) did not show an apparent functional relation, and their inspection with STRING-db revealed the same functions as above with significant *p*-value. In particular, the TAIR10 orthologue of *Oleu061Scf7673g01036* was uncharacterised, but OliveTreeDB v0.61 describes it as a putative C2 NT-type domain-containing protein in trEMBL, corresponding to PROSITE entry PDOC51840. *Oleu061Scf1255g08007.1* may be a vacuolar sodium/calcium exchanger with calcium-binding EF-hand that participates in the maintenance of calcium homeostasis in exchange with Na, as reflected by the separate yellow network in ‘Other up-regulated DEGs’ panel in Figure 6, and is also involved in diurnal rhythm (‘photoperiodism’ in Figures 2 and 4) and salt stress response [46]. *Oleu061Scf3503g02024.1* may be a proline-rich receptor-like protein kinase PERK1 that appears associated in the STRING-db network with other proteins involved in ion binding, but not transport and also to cellulose synthase (together with the RING/U-box superfamily protein AT3G10910) and other protein kinases involved in cell growth and pistil-specific extensins (KIPK and AT2G36350).

Finally, Oleur061Scf6158g01023.1 seems to contain an RNA-binding YTH domain like ECT5 or CPSF30 (there are 12 *A. thaliana* proteins with this domain in [PANTHER 17.0](#)) that has RNA-binding capacity that decreases the mRNA stability, making it susceptible of degradative processes mediated by NURS complex (InterPro ID [IPR039278](#)), in agreement with the STRING-db network in [Figure 6](#). Among the proteins containing this domain, those suppressing meiosis-specific genes during gametogenesis or general silencing as indicated by PFAM ID [PF04146](#) and InterPro ID [CD21134](#) may be related to PT growth. Interestingly, these processes are already seen in [Figures 2 and 4](#).

### 3. Discussion

#### 3.1. RSeqFlow Pipeline Is Reliable and Expandable

Bioinformatics analyses, excluding the functional enrichment, were performed with our R pipeline *RSeqFlow*, easily configurable in one single file. This enables the claimed repeatability and reproducibility in published papers [\[47\]](#). Every result is saved within a folder accompanied by an HTML report containing a detailed explanation of results and the rationale of each analytical step, including relevant bibliography (see the Report .html file in [Appendix A.2](#)). This approach facilitates (i) interpretation, (ii) repeating the analysis as many times were necessary with different parameters and data, and (iii) retaining all executions for further comparisons. *RSeqFlow* uses the best practices described in the literature [\[48–70\]](#) concerning removal of noisy genes ([Figure 1](#)), resilience for unequal library sizes, TMM (trimmed mean of M-values) normalisation for differential expression, CTF (counts adjustment with TMM factors) normalisation for correlation analyses, heteroscedasticity removal, best linear models for less than 10 replicates, use of *treat* instead of *eBayes* ([Table 1](#)) to increase the biological relevance, use of several clustering methods with objective criteria for the optimal number of clusters, building networks and communities to extract relevant biological functions and remove spurious connections, avoid memory overflow when calculating networks, and select putative hub transcripts as a surrogate of regulators. This is why the number of DEGs in our study is not comparable to previous works [\[19,71,72\]](#), while the functional interpretation is consistent, which agrees with the guidelines that claim for the interpretation of biological processes instead of genes [\[73,74\]](#). The tables of TAIR10 orthologues ([Appendix A.5](#)) and TFs ([Appendix A.3](#)) are useful for further functional enrichment analyses ([Figures 2 and 4](#)). Authors and co-workers are using *RSeqFlow* for a wide range of applications in the olive tree and other plants, which illustrates its suitability in any field of research where differentially expressed genes are required, and further correlations/co-expressions between them are desired, especially in non-model organisms.

#### 3.2. The Olive Pollen Transcriptome Contains Pollen-Specific Genes

The mature pollen plus germinating pollen transcriptome was estimated in 22,418 reliable genes ([Appendix A.2](#)), with a certain level of redundancy as deduced from annotations in [Appendix A.5](#) and [Table 2](#), in agreement with results reported in [\[37\]](#). The number of genes identified in this work is very close to the 24,672 identified in *Nicotiana tabacum* PT for similar culture times [\[71\]](#), the 25,062 identified in rice [\[28\]](#), as well as the 21,394 in *Pyrus bretschneideri* [\[75\]](#) or the 19,868 of *Lilium longiflorum* [\[76\]](#), all of them higher than the 13,017 genes of poplar [\[29\]](#) and the 4172 genes described for *A. thaliana* [\[77\]](#). In jasmine, another member of the Oleaceae family, 178,098 gene products were assembled when pollen-pistil interactions were analysed [\[72\]](#).

The extended annotation of olive pollen transcripts presented here ([Appendix A.5](#)) implied that most DEGs (93.7%) were annotated, the majority (91.6%) with a known function. Within the 5364 DEGs ([Table 1](#)), 310 were uncharacterised PT-specific genes, including a subset of 173 DEGs without any orthologue in the databases. However, 56 of them were previously sequenced in the reproductive tissues of cv. Picual [\[32,34\]](#), with up to three DEGs being putative TFs ([Appendix A.6](#)). These findings confirmed the existence of

unnoticed pollen-specific genes, as previously reported [19,29,71], which should stimulate the molecular study of pollen germination in more detail [34].

### 3.3. Olive Mature Pollen Is Reprogrammed for PT Growth within the First Hour of Germination

The 0→1 h transition is the only comparison with more down-regulated and less up-regulated DEGs (Table 1) resulting in many enriched processes (Figure 2). This behaviour is consistent with previous findings about a global reduction of transcription after pollen germination [71] due to the deep gene reprogramming necessary for the PT emergence after hydration [4,5]. Pollen reprogramming in vitro to acquire totipotency to produce doubled haploid plants has also been reported [78], where hormonal and epigenetic mechanisms involving DNA and histone demethylation and some histone acetylation, as well as autophagy, cell death and cell wall remodelling, have been observed [78]. Particularly, chromatin reorganisation implies an initial decrease of DNA methylation with a subsequent increase in this activity [79], which is consistent with the decreased ‘macromolecule methylation’ and ‘chromatin organisation’ observed in 0→1 h DOWN and 0→6 h DOWN panels in Figure 2. Autophagy in pollen reprogramming towards embryogenesis [78] is also compatible with ‘autophagy’ processes in 0→1 h DOWN, 1→3 h UP and 0→6 h UP panels in Figure 2. Cell wall remodelling was supported by the increase of ‘glycoprotein biosynthesis’ in the 0→6 h UP and the decrease of ‘pollen wall assembly’ in the 0→6 h DOWN panel in Figure 2. Another proof reinforcing the hypothesis of pollen reprogramming is the decrease of PTGS within the first hour of germination (0→1 h DOWN panel in Figure 2), followed by an increase after 6 h of culture (0→6 h UP panel in Figure 2), which is consistent with previous observations [71] indicating that miRNA regulation is involved in PT growth. Since TFs in Figure 6 are clearly necessary after 1 h of germination, they are good candidates to directly or indirectly participate in the required reprogramming. In conclusion, it is strongly suggested that a deep chromatin reorganisation is required to reprogram olive mature pollen towards the PT emergence and that this reprogramming involves both protein and miRNA regulation of chromatin.

### 3.4. Olive Pollen May Express Specific Isoforms

Focusing on the proposed hub genes in Table 2, many TAIR10 genes have several orthologues in olive pollen, suggesting that, for example, olive tree contains more HSP and FBP proteins than *A. thaliana* [23,24]. The same idea can be extracted from repeated TAIR10 IDs in up- and down-regulated DEGs.

Expression profiles of HSPs and FBPs in Figure 5 clearly suggest their involvement in maintaining the viability of pollen grain before landing on stigma due to their rapid decrease to negligible values after 1 h of culture. The fact that many HSP20 genes appear in this process would extend to pollen the findings described in plant seeds [22], where this protein family provides desiccation tolerance, decreasing their expression upon germination. Regarding hub FBPs, their expression is as expected for many pollen FBPs [24,28] and may be more related to pollen-pistil self-incompatibility [27] or removal of unnecessary proteins after pollen maturation [12] than with pollen protein stability [26]. In conclusion, these results also strengthen the importance of the identification of gene isoforms (alleles, paralogues, alternative splicing, etc.) throughout different developmental stages to discover those specifically involved in olive PT growth.

### 3.5. RNA Surveillance during Olive PT Growth

The olive pollen undergoes the second mitotic division in the PT [42,80] with the elimination of a significant part of ribosomes and transcripts [1], a process that also involves the FBPs of the ubiquitin-proteasome pathway [12], as discussed above. RNA expression appears to be a key process during the first 3 h of PT growth, since ‘mRNA surveillance’—detection and degradation of abnormal mRNAs—as biological process and as KEGG pathway appears upregulated in both 0→1 and 1→3 h transitions (Figure 2) and in KM-3 cluster (Figure 4). In addition, the ‘RNA export from nucleus’ is clearly decreased

in the 3→6 h transition, ‘nucleotide metabolism’ is down-regulated in the 0→6 h comparison, and ‘transcription termination’ and ‘RNA degradation’ are accompanying mRNA surveillance in KM-3 (Figure 4). Pollen-specific activities of transcription termination, including 3′ processing and polyadenylation (Appendix A.10), were recently described as up-regulated in PT during growth and guidance in *A. thaliana*, these activities are required for full acquisition of fertilisation competence [81]. Oleur061Scf3360g05017.1 (a polyadenylate-binding protein 7)—a ubiquitous DEG initially down-regulated but then up-regulated during all stages of PT growth involved in the mRNA degradation process of KM-3 (Figure 4)—supports the increasing importance of mRNA termination in olive PT growth. It can be then inferred that after 3 h of PT growth, the pool of olive pollen mRNAs is already formed, and there is no longer de novo transcription onwards [30]. Therefore, the later part of PT growth has to be conducted with persisting transcripts, and can also explain why ribosomes synthesis was initially up-regulated in the 0→1 h transition but appears in the co-expression group KM-2 (Figure 4) with a slight increase (Figure 3). The role of persisting transcripts in later pollen tube growth has been recently demonstrated in *A. thaliana* [30].

In concordance with maintaining PT growth without transcription, it was detected an up-regulation of the ‘negative regulation of cell death’ function (KM-3 in Figure 4). The slow increase of the different pathways to maintain the correct protein folding (KM-2 in Figure 4 and up-regulation of ‘ERAD pathway’ in Figure 2) also confirms that olive PT can grow even though from persistent transcripts, in concordance with previous findings in pear [75].

### 3.6. Olive PT Growth In Vitro Is Subjected to Stress and/or Redox Signalling

Functional networks in Figures 2 and 4 contain many biological processes related to responses to stress and/or redox signalling, which have been already reported in other plants [28,29]. Those processes down-regulated within the first hour of culture (responses to ROS, heat, misfolded proteins and mechanical stimuli) can be hypothesised to be present in the mature pollen to prepare it against adverse environmental conditions, while the up-regulated functions (‘responses to hypoxia’, ‘ERAD proteasome pathway’, and ‘salt, light, and biotic stimuli’), might appear to afford the adverse environment during the in vitro PT growth. In fact, the use of in vitro pollen growth instead of an in vivo approach is a limitation of the present work because it misses the complex dialogue between pollen and pistil occurring from the earliest stages of pollen germination all the way to the signalling that culminates in double fertilization. It is likely that this dialogue triggers profound changes in gene expression in germinating pollen and growing pollen tubes. Therefore, several transcriptional responses shown here, which were detected at 3 and 6 h of growth, could be interpreted as stress responses to the in vitro environment, although they might also be related to the pollen tube getting ready for the required ROS signalling between pollen tube and the pistil. This must be further tested in the olive tree by using in vivo conditions. It is known that flavonoids are important for maintaining PT morphology and ROS modulation [12,14] and are probably involved in these responses to hypoxia and ROS to maintain olive pollen viability, turgor, and shape, as well as contributing to signalling [13,14]. Because ROS response is decreasing, it can explain why the flavin-containing compounds are also globally decreased (0→6 h DOWN panel in Figure 2) as well as the ‘isoprenoid biosynthesis’ in KM-1 (Figure 4) in coordination with the metabolism of other lipids.

It must be mentioned the prevalent presence of ‘protein refolding’ and ‘ER organisation’ processes observed in the 0→1 h transition step required for further ‘ERAD pathway’ for misfolded/unfolded proteins observed at 3 h of germination (0→1 UP and 1→3 UP panels in Figure 2, and in KM-2 in Figure 4) accompanied by proteins of the proteasome pathway (Figure 2). HSPs and FPBs may be guiding the ubiquitination and proteasome pathways that are known to regulate virtually all aspects of plant biology [24], including pollen recognition and rejection [25]. It is probable that the rapid synthesis of proteins for the fast

olive PT growth could produce incorrectly folded peptides that may require refolding or even specific turnover, as reported in *A. thaliana* [19]. Taken together, those results suggest that mature, desiccated pollen seems to be equipped with a complete set of transcripts that prepares it against the stressing environment before its landing on the stigma; once the PT starts growing throughout the style (in our case, on the in vitro plate), transcripts related to the protective responses disappear, together with unnecessary biosynthetic processes. In fact, the only protective process that must be retained seems to be related to the cell's turgor pressure and mechanical obstacles that keep the wall under continuous tension [16], as will be discussed below.

### 3.7. Last Mitotic Division of Generative Cells

Olive pollen undergoes mitosis II after anthesis and produces a less differentiated, longer-lived bicellular pollen whose generative cell divides mitotically in the PT [42,80]. The already discussed decrease of nucleotide, protein and metabolite biosynthesis at 6 h is specifically supported by the 3→6 h transition, where (i) only 10 DEGs were detected (Table 1) since their transcriptional profiles at 3 and 6 h of germination are very similar (Figure 1), and (ii) expression levels of up- and down-regulated clusters in Figure 3 are constants at both culture times. Cyclin-dependent kinases and microgametogenesis are the main processes involved (3→6 h UP panel in Figure 2), and 'cell fate' and 'cytoskeleton organisation' down-regulation in KM-1 together with 'negative regulation of cell death' in KM-3 in Figure 4 suggest that the last mitotic division of microsporogenesis [82] leading to male gamete maturation in olive PTs was probably achieved at 6 h of PT growth and that new transcripts are no longer required, even though the pollen tube is not expected to arrive to the ovary in nature before 24 h of growth [31].

### 3.8. Other Relevant Processes during Olive PT Growth

Pollen hydration is the elicitor of PT emergence [4] and, indirectly, may also drive the well-known biological processes of ribosome assembly for intensive protein synthesis, oxidative phosphorylation to afford energy demands from early stages, and the already discussed mRNA surveillance and endoplasmic reticulum reorganisation (to export proteins and mRNA and control misfolded proteins and unprocessed mRNAs). After 1 h germination, the pollen grain is transcriptionally ready to face PT growth by up-regulation of cytoskeleton reorganisation for anterograde/retrograde transport of vesicles and organelles, vesicular trafficking,  $\text{Ca}^{2+}$  and other ion fluxes, distribution of reactive oxygen species, pH control, energy metabolism, and protein folding and degradation [1,4,8,9,12,19,72,75,83], as was also supported results in Figures 2 and 4. This may explain why down-regulation of 'endocytosis', 'secretion' and 'vesicle transport' in the 0→1 h transition is followed by up-regulation of 'organelle transport', 'exocytosis', 'endosome to vacuole transport', 'reorganisation of cytoskeleton', probably due to the retrograde and anterograde movement of organelles and vesicles, including the gametes [7], that coordinate exocytosis and endocytosis for a fast pollen tube tip growth [2,8].

$\text{Ca}^{2+}$  was known to be a key messenger for PT growth [1] sensitive to inositol signalling [10], and both processes appeared during PT growth. Olive PT  $\text{Ca}^{2+}$  transport was up-regulated (KM-3 cluster in Figure 4) while phosphatidylinositol biosynthesis is significantly down-regulated in the 0→1 h transition (0→1 h DOWN panel in Figure 2 and the corresponding KEGG pathways in Appendix A.9). The presence of 'autophagy', ' $\text{Ca}^{2+}$  transport' and 'phosphatidylinositol biosynthesis' in Figures 2 and 4 is consistent with recent findings where autophagy is required for the high metabolic demand of PT growth, an activation that depends on  $\text{Ca}^{2+}$  increase and inositol decrease [11]. Ion fluxes and pH control are very active during PT growth as reflects the increase of 'response to pH' and response to salt/osmosis processes at 6 h of culture and the decrease of 'ion homeostasis' (3→6 h UP and 0→6 h DOWN panels in Figures 2 and KM-3 cluster in Figure 4). This may explain the presence of a putative aquaporin TIP2-1 (Oleur061Scf0430g00024.1, DELTA-TIP) among the ubiquitous genes that can be regulated by cyclin-B1-2 (Oleur061Scf3133g07005.1)

which is also a ubiquitous gene. Ion balance would be required in olive pollen to maintain polarity through microfilament skeleton, signalling, and oxidative status of male gametes, as described in other species [10,75], as well as to preserve the cylindrical shape and osmolarity of growing PT [84–86], because ions and water are used to fill the new volume and sustain turgor pressure [15].

### 3.9. Putative Regulators

A total of 608 TFs distributed in 46 TF families were DEGs during pollen hydration and PT growth (Appendix A.4). It seems that 11 TF families are not related to PT growth, while other 22 can be considered over-represented, extending the pollen-specific TF families already reported in rice [28]. Because pollen expresses a high proportion of pollen-specialised genes [28,29], uncharacterised pollen genes (Table 2) were extracted following the ‘guilt-by-association’ approach [61,64,87,88]. Down-regulation of HSPs and FPBs, mainly involved in protein folding, refolding or degradation, can explain why the protein refolding and the responses to stresses in Figure 2 appear down-regulated. It is tempting to propose that this mechanism could be directly or indirectly governed by Oleur061Scf0370g03005.1, a heat-shock TF containing a winged helix-turn-helix DNA-binding domain. Additionally, the uncharacterised genes Oleur061Scf1295g03035.1 and Oleur061Scf8276g00009.1 were involved in the ribosome biogenesis, a process that was up-regulated from the first hour of culture onwards (Figures 2 and 4).

Activated hub genes, mainly those considered TFs (Table 2), may be controlling some relevant processes. Because their functional annotation was transferred by homology, it can be suggested that they were never studied before because their expression might be PT-specific. Transcription factors belonging to bHLH, WRKY and MYB are usually found in pollen [72] since they are very large families of TFs, but our co-expression approach also found TFs from other important families. Particularly, Oleur061Scf9115g00013.1 has no orthologue but is described as a NAC domain-containing protein with low identity in trEMBL, while NACs appear relevant gene up-regulators during PT growth in the present study (Appendix A.7). Its real participation in PT growth is then unknown, and it is likely a PT-specific TF whose functions and expression would merit future research. Another NAC domain-containing hub gene, Oleur061Scf3317g04032.1, seems related to hypoxia, response to pathogens, and water transport to vacuole probably via the aquaporin Oleur061Scf0430g00024.1 because pollen-specific aquaporins in the vacuole have been described important for pollination and PT-growth [89]. In other words, it is tempting to suggest that both NAC domain-containing genes, together with the Zn-finger-containing protein transcript Oleur061Scf3825g03001.1 are good, uncharacterised candidates to coordinate the reprogramming of olive genes during PT growth.

Oleur061Scf7673g01036.1 contains a C2 NT-type domain typical of proteins involved in the intracellular trafficking of chloroplasts and endosomal vesicles; this suggests a putative implication in vesicle and/or organellar movement during PT growth as already suggested in olive [34].

The  $\text{Na}^+/\text{Ca}^{2+}$  exchanger nature of Oleur061Scf1255g08007.1 can be deduced from the yellow network in ‘Other up-regulated DEGs’ panel in Figure 6 and may be involved in the  $\text{Ca}^{2+}$  and ion homeostasis in olive PT, in agreement with *A. thaliana* [46].

Oleur061Scf3503g02024.1 is a putative receptor that may transduce signals to different processes. Our data suggest its relation to ion binding and cellulose biosynthesis, even though this last function was not found to be significantly up- or down-regulated.

Finally, functional annotation of Oleur061Scf6158g01023.1 may suggest its involvement in suppressing meiosis-specific genes during gametogenesis by its ability to bind RNA and send it to NURS-based degradation [90]; this may not only relate this gene with PT growth but also with the second mitotic division of the generative cell.

#### 4. Conclusions

The R pipeline *RSeqFlow* contemplating the best practices described in the literature provided relevant transcriptional information about olive pollen germination. It revealed a transcriptome of 22,418 genes that underwent a deep cell reprogramming (entailing DNA reorganisation) from mature pollen to 6 h growth pollen tube, where many HSPs and F-box proteins were key proteins. For example, *Oleur061Scf3360g05017.1* (a polyadenylate-binding protein 7) is differentially down-regulated in the 0→1 h transition but differentially up-regulated in the remaining growing stages. Besides the well-known processes involved in pollen germination, protein synthesis and oxidative phosphorylation were from very early stages, and autophagy for the later ones, together with an overall reduction of the biosynthetic metabolism. Mature, desiccated pollen seems to be equipped with protecting transcripts that disappear upon germination. Transcription is very active before the 3 h of pollen germination, but after that, the last mitotic division of microsporogenesis seems to occur together with a transcription arrest, making PT growth dependent on persistent transcripts. Many TFs were identified belonging to 22 different families, with two NAC members (*Oleur061Scf9115g00013.1* and *Oleur061Scf3317g04032.1*) and one Zn-finger protein *Oleur061Scf3825g03001.1* having a central role. Aquaporins, proteins with C2 NT-type domain or YTH domain, and  $\text{Na}^+/\text{Ca}^{2+}$  exchangers also appeared as relevant hub genes in this study. Hence, the olive tree displays a wider genetic diversity arising from its variate origin, as well as its domestication, giving rise to different alleles, paralogues, alternative splicing, etc., that merit future characterisation.

#### 5. Materials and Methods

##### 5.1. In Vitro Pollen Germination, RNA Extraction and Sequencing

Samples from mature rehydrated pollen (0 h growth), as well as pollen germinated at 1, 3, and 6 h of in vitro culture, were prepared in triplicate. The procedures to germinate olive pollen and to extract RNA extract were previously described in [43]. The integrity, quality, and quantity of RNA were assessed using the Agilent Bioanalyser 2100 and RNA Nano 6000 Labchip kit (Agilent Technologies, Palo Alto, CA, USA). RNA was converted to cDNA and sequenced by the High-Throughput Sequencing Unit of the University of Malaga, which yielded  $2 \times 75$  bp reads in their NextSeq 550, generating a total of 40–50 million reads per replicate. Raw reads were deposited as Bioproject PRJEB59024.

##### 5.2. Bioinformatic Procedures

Homology, as well as pre-processing and mapping, was conducted on a SUSE Linux Enterprise Server 11SP2 with Slurm queue system and Infiniband network (54/40 Gbps) consisting of 216 nodes with Intel E5-2670 2.6 GHz cores for a total of 3 456 cores and 8.4 TB of RAM.

###### 5.2.1. Closest Homologue

The closest homologue for transcripts in the olive (cv. Picual) transcriptome v0.61 (<https://genomaolivar.dipujaen.es/downloads/Sequences> (accessed on 2 July 2023)) against the *Olea europaea* var. *sylvestris* v1 transcriptome ([https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/fasta/olea\\_europaea\\_sylvestris/cdna/Olea\\_europaea\\_sylvestris.O\\_europaea\\_v1.cdna.all.fa.gz](https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/fasta/olea_europaea_sylvestris/cdna/Olea_europaea_sylvestris.O_europaea_v1.cdna.all.fa.gz) (accessed on 2 July 2023)), and the extension of *Arabidopsis thaliana* TAIR10 orthologues were performed using reciprocal comparison based on the fast and sensitive protein aligner *Diamond BLASTp* v2.0.14 [91] using default parameters and `-max-target-seqs 1` in `-ultra-sensitive` mode to get the best hit.

###### 5.2.2. Pre-Processing and Mapping

Raw reads were analysed with *FastQC* v0.11.9 and *MultiQC* v1.11 [92] to inspect if raw data merit the analysis. They were then pre-processed for low-quality regions, contaminant sequences, adapters, etc., using *SeqTrimBB* v2.1.8 (<https://github.com/rafnunser/seqtrimbb> (accessed on 2 July 2023); an evolution of *SeqTrim-Next* [93]) applying the standard

parameters for Illumina paired reads, a minimal length of 40 bases and a contaminant minimal ratio of 0.65. Clean reads were submitted again to *FastQC/MultiQC* to verify sequence artefact removal. The resulting clean reads from every sample were mapped to the *Olea europaea* cv. Picual transcriptome (cDNA v0.61 with 81,484 transcripts) using *Bowtie* v2.4.4 [94] with default parameters and `-no-mixed -no-discordant` options to only retain appropriately mapped paired-end reads in the SAM file. *SAMtools* [95] helped in sorting and indexing of SAM files to facilitate read counting per transcript with *sam2counts* (<https://github.com/vsbuffalo/sam2counts> (accessed on 2 July 2023)). This results in a count matrix where rows are transcripts and columns are samples.

### 5.2.3. RSeqFlow Pipeline

The previously obtained count matrix is the input for our *RSeqFlow* where configuration, filtering, differential expression, correlation/co-expression and networking analyses are conducted automatically. This workflow can be performed on a supercomputer as well as on laptop computers based on Linux, macOS 12.0 or higher, or even Windows. The complete *RSeqFlow* code can be cloned from GitHub (<https://github.com/mgclaros/RSeqFlow> (accessed on 2 July 2023)). The results presented here were performed using Bioconductor v 3.16 (BiocManager v 1.30.19) and R 4.2.2 (31 October 2022). It creates the folder `RSeqFlow1_results_{DATETIME}` (from now on, `{DATETIME}` is the date and time of the execution; see Appendix A.2) besides the analysed dataset containing tables of correlations, clusters, normalisation, etc., and a comprehensive HTML report explaining the analysis and plots. Customisation of the `configure_wf.R` file is compulsory (refer to the GitHub Readme file).

**Pre-processing and differential expression:** Within *RSeqFlow*, read counts from cv. Picual cDNAs v0.61 were converted to CPM (counts per million reads) as described in [58] since only inter-sample comparisons of the same gene will be performed. Those transcripts having  $CPM > 1$  (a configurable threshold) in at least three replicates were considered reliable and saved at `filteredData-{DATETIME}.tsv` file in Appendix A.2. Reliable transcripts were then normalised using the TMM (Trimmed Mean of M-values) method implemented in BioConductor *edgeR* v3.39.6 [50] library to remove systematic technical effects that occur in the data to ensure that technical bias has minimal impact on the results [96]. TMM was selected due to its established robustness [53,59]. Consistency of filtered and normalised data was confirmed by the decrease of dispersion by means of BCV (biological coefficient of variation) [51]. Normalised counts per gene and sample can be found at `TMMnormalisedCounts-{DATETIME}.tsv` and `TMMnormalisedCPMs-{DATETIME}.tsv` files, respectively, in Appendix A.2. Sample comparisons (0→1, 1→3, 3→6, 0→3, 1→6 and 0→6) were conducted by the highest incubation time relative to the lower incubation time in order to interpret up-regulation as the increase of gene expression over the time. Differentially expressed genes (DEGs) were obtained using a generalised linear model (GLM) adapted to binomial distributions based on *edgeR* v3.39.6 and *limma* v3.52.4 libraries. Since this analysis requires that the standard deviation of transcripts is not dependent on the expression level [56], heteroskedasticity was removed using the *limma::voom* function. The combination of *limma::voom* was reported to provide fewer false positives than other [69] and deals better with outliers and unequal library sizes [54]. Another improvement in *RSeqFlow* is that the fitted model was not adjusted with the classical empirical Bayes moderation of *limma::eBayes*, but the *limma::treat* function [48] that is more rigorous and conservative for few replicates. Its rationale is not based on fold-change thresholds, but using a variably increased fold-change depending on the consistency of the gene expression, resulting in more biologically relevant DEGs [48,56,58]. This is why the minimal fold-change (FC) was set to 1.2 (it is configurable). Concerning the adjusted *p*-value, to avoid the misuse of the statistical significance [63,65], the minimal cutoff was set to 0.1 (it is configurable). All DEGs with *limma::eBayes* and *limma::treat* methods can be found at files `AllGenes_allContrast_TREAT_{P-cutoff}_{FC-cutoff}_{DATETIME}.tsv` and `AllGenes_allContrast_eB_{P-cutoff}_{FC-cutoff}_{DATETIME}.tsv`, respectively,

in Appendix A.2. DEGs in each contrast have also been saved in a separate file as `DEGs_{CONTRAST}_{method}_{P-cutoff}_{FC-cutoff}_{DATETIME}.tsv` in Appendix A.2. Genes that result in differential expression in all contrast using the `limma::treat` approach (or the `limma::eBayes` approach when no DEG was obtained with `limma::treat`) were saved in the `ubiquitousDEGs_{DATETIME}.tsv` in Appendix A.2 for further analyses.

**Clustering:** *RSeqFlow* also includes gene clustering to infer co-expression or co-regulation using R functions of `base`, `stats`, `cluster`, `mclust` and `dendextend` libraries. Expression data (counts) were normalised using the CTF (count adjustment with TMM factors) because it is described as more appropriate for correlation studies [68]. CTF-normalised counts were then standardised (scaled) using the `base::scale` function to reduce spurious variations and expression level differences and obtain comparable expression data. CTF-normalised counts per gene and sample were saved at `CTFnormalisedCPMs_{DATETIME}.tsv` file in Appendix A.2. Although only DEGs were considered for clustering, computational constraints due to the oversize of genes versus samples were circumvented by the elimination of stable genes defining a coefficient of variance (CV) threshold of  $CV > 0.2$  (configurable). The required dissimilarity matrix based on Euclidean distances was derived from Pearson's correlation  $r$  using the `stats::cor` function to obtain distances as  $1 - r$ . The matrix was mined using Agglomerative Hierarchical Clustering (AHC),  $k$ -means and model-based clustering (MBC) [97] using functions in libraries `stats`, `NbClust` v 3.0.1 [55], `dynamicTreeCut` v 1.63.1 [98], and `mclust` v 6.0.0 [60]. The optimal linkage method for genes was calculated using `cluster::agnes` function, the optimal number of clusters was statistically inferred using up to 26 different methods with `NbClust::NbClust`, `dynamicTreeCut::cutreeDynamic` and `mclust::Mclust` functions, the latter exploiting the dynamic clustering based on the Bayesian information criterion [99]. The minimal number of genes per cluster was set to 10 (it is configurable) to strengthen the functional interpretation. Gene profiles of clusters were obtained using the `stats::aggregate` function on the scaled and CPM expression values for each gene. The assignments of DEGs to AHC,  $k$ -means, and MBC are saved in `ClustersCTF_{DATETIME}.tsv` file in Appendix A.2. As a help for users, the optimal cluster for the three methods was analysed with the `cluster.stats` function of the `fpc` v 2.2.9 library to decide which clustering result is preferable.

**Networks:** The previously-used Pearson's correlation  $r$  was also used to construct gene networks that will facilitate the study of biological functions [64,70] because genes under the same regulatory control tend to be functionally related [61] based on a 'guilt-by-association' approach [87]. A cutoff score of  $|r| > 0.75$  was established so that the coefficient of determination  $R^2$  is  $>50\%$  ( $r^2 = 0.75^2 = 0.56 = R^2$ ) meaning that genes exhibiting such a score would have more than the half of their variance in common, leaving only a 44% [ $(1 - 0.56) \times 100$ ] to randomness [49]. Relevant correlations were then filtered by the adjusted  $p$ -value to decrease the number of correlations without biological (or causation) sense. Meaningful gene pairs were then used to construct undirected networks as described by Provart et al. [52]. The best correlation of clusters was saved in separate tsv files starting by `BestCorrelations_` and the cluster name in Appendix A.2. Because biological networks have been observed to be highly modular, where tightly-connected genes represent organised communities that are expected to conduct common biological functions [57], and maybe are responsible for a common phenotype, networks were then characterised by the node degree and betweenness, Kleinberg's centrality score [100], hub score, and other network properties, as well as possible communities (modules), using the corresponding functions from the `igraph` v1.3.5 library. Cluster, network and communities are saved as an R object in file `List_of_clusters_{DATETIME}.Rds` in Appendix A.2.

**Candidate hub genes:** Tentative hub genes were proposed from 'ubiquitous' DEGs (DEGs in all contrast saved above; in many cases, there are no ubiquitous DEG) and 'outstanding' DEGs (DEGs having a Kleinberg's centrality score  $> 0.9$  and a degree  $> 10$ —a note that links with at least other 10 nodes in the network; both values are configurable—) as generally recommended [62]. Candidate hub genes were saved as tsv files starting by

OutstandingGenes- and ubiquitousDEGs\_ in Appendix A.2 for further analyses as hub gene candidates.

#### 5.2.4. Functional Interpretations

Functional enrichment analyses of gene lists are crucial to interpret the biological implications of DEGs in high-throughput studies. Since the olive tree genome has been recently published [45], its functional annotation is incomplete, requiring the use of *Arabidopsis thaliana* orthologues instead (Appendix A.5). The R library *clusterProfiler* v4.6 [101] was used, running on default parameters (in all cases,  $FDR < 0.05$  and processes with at least 10 DEGs). Input files for the analysis were any of the `tsv` files starting by `AllGenes_allContrast_` or alternatively `List_of_clusters-{DATETIME}.Rds` (Appendix A.2). The final report also includes graphs of the most significant GO terms focusing on biological processes (Appendix A.8) and KEGG pathways (Appendix A.9).

Unique TAIR10 orthologues of putative hub genes were subjected to *STRING-db* v 11.5 (<https://string-db.org/> (accessed on 2 July 2023)) analysis [67] to know if correlation-linked genes share other functionally related associations. This analysis was also used to determine if any biological processes, locations or pathways were specifically highlighted. Default values were used but asked for up to 20 interactions in the second shell. Functional classification of proteins deduced from genes was also conducted in *PANTHER* 17.0 (<http://www.pantherdb.org> (accessed on 2 July 2023)) [102]. Putative transcription factors in olive tree were inspected using the TF prediction tool of PlantTFDB v5.0 [103] based on their DNA-binding domain predictions, as well as comparisons to transcription factors for wild olive tree contained in the extended PlantTFDB ([http://plantfdb.gao-lab.org/index\\_ext.php?sp=Oeu](http://plantfdb.gao-lab.org/index_ext.php?sp=Oeu) (accessed on 2 July 2023)).

**Author Contributions:** Conceptualization, M.G.C. and J.d.D.A.; methodology, M.G.C., N.F.-P., E.L.-C., R.B. and A.J.C.; software, A.B. and M.G.C.; validation, A.B. and N.F.-P.; formal analysis, A.B., R.B. and M.G.C.; investigation, A.B. and E.L.-C.; resources, E.L.-C. and J.d.D.A.; data curation, A.B. and E.L.-C.; writing—original draft preparation, M.G.C.; writing—review and editing, J.d.D.A., A.J.C. and N.F.-P.; visualization, A.B. and M.G.C.; supervision, A.J.C., J.d.D.A. and M.G.C.; project administration, J.d.D.A. and M.G.C.; funding acquisition, J.d.D.A., M.G.C. and N.F.-P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by MCIN/AEI/ERDF-UE grants TED2021-130015B-C21, TED2021-130015B-C22, PID2020-113324GB-I00, RED2022-134072-T and RYC2020-030219-I, and Consejería de Economía, Conocimiento, Empresas y Universidad, Junta de Andalucía/ERDF-UE grants P18-RT-1577 and UMA20-FEDERJA-029.

**Data Availability Statement:** RSeqFlow is available at GitHub <https://github.com/mgclaros/RSeqFlow> (accessed on 2 July 2023). Raw reads are available at Bioproject [PRJEB59024](https://bioproject.ncbi.nlm.nih.gov/submitter/study.cgi?study_id=PRJEB59024). Appendix A data are available in the FigShare project [https://figshare.com/projects/Olive\\_pollen\\_tube\\_growth\\_analysis/171015](https://figshare.com/projects/Olive_pollen_tube_growth_analysis/171015) (accessed on 2 July 2023).

**Acknowledgments:** The authors thank the Supercomputing and Bioinnovation Center (SCBI) of the University of Malaga for their provision of computational resources (<https://www.scbi.uma.es> (accessed on 2 July 2023)) as well as the bioinformatic support provided by the Plataforma Andaluza de Bioinformática of the University of Malaga.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

#### Abbreviations

The following abbreviations are used in this manuscript: ABA: Abscisic acid; AHC: Agglomerative hierarchical clustering; CPM: Count per million reads; CTF: Count adjustment with TMM factors; CV: Coefficient of variance; cv.: cultivar; DEG: Differentially expressed gene; ERAD: Endoplasmic-reticulum-associated protein degradation; FBP: F-box protein; FC: Fold-change; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; MBC: Model-based clustering; MDS: Multidi-

mensional scaling; PCA: Principal component analysis; PT: Pollen tube; PTGS: Post-transcriptional gene silencing; ROS: Reactive oxygen species; SCF: ASK-cullin-F-box; TMM: Trimmed mean of *M*-values; TF: Transcription factor.

## Appendix A

Data available in a publicly accessible repository (FigShare project [https://figshare.com/projects/Olive\\_pollen\\_tube\\_growth\\_analysis/171015](https://figshare.com/projects/Olive_pollen_tube_growth_analysis/171015) (accessed on 2 July 2023)) with the accompanying DOI.

### Appendix A.1. Pre-Processing and Mapping of Olive (*cv. Picual*) Pollen Libraries

DOI: <https://doi.org/10.6084/m9.figshare.23586372> (accessed on 2 July 2023)

*Preprocessing-and-mapping.tsv*: Results about library contents, read pre-processing and mapping (onto two different transcriptomes) of pollen (*cv. Picual*) replicates at different times of pollen culture. Data are presented as a tab separated values (tsv) that list sample replicates, number of raw reads, number of clean reads (after pre-processing), percent of useful reads, and the percent of mapping onto the transcriptomes of *cv. Picual* v0.61 and wild olive *Oeu* v1.

### Appendix A.2. RSeqFlow Folder Gathering the Results

DOI: <https://doi.org/10.6084/m9.figshare.23587032> (accessed on 2 July 2023)

*RSeqFlow1\_results\_2023-02-09\_16.03.33.zip*: Compressed file containing the output folder created by *RSeqFlow* that includes the report and results mentioned in the present work. Files with extension *.tsv* contain tab separated values. Taking into account that {DATETIME} is 2023-02-09\_16.03.33 (date and time when it was executed), the following files can be found, in alphabetical order:

- *Report.html*: The complete report explaining the analysis, including figures, analysis rationale, bibliography and explanation of each saved file.
- *AllGenes\_allContrast\_TREAT-P-0.1\_FC-1.2\_{DATETIME}.tsv*: Average expression, coefficient (LogFC), *t* statistic, *p* value, adjusted *p* value, *F* statistic, and DEG result for all genes in each of the contrasts using the *treat* method and the *p* and *FC* indicated in the filename.
- *AllGenes\_allContrast\_eB\_P-0.1\_FC-1.2\_{DATETIME}.tsv*: Average expression, coefficient (LogFC), *t* statistic, *p* value, adjusted *p* value, *F* statistic, and DEG result for all genes in each of the contrasts using the *eBayes* method and the *p* and *FC* indicated in the filename.
- *AllGenes\_{CONTRAST}\_TREAT\_{DATETIME}.tsv*: LogFC, average expression, *p* value, and adjusted *p* value, for all genes in the contrast indicated in {CONTRAST} using the *treat* method.
- *BestCorrelations\_{METHOD}\_{CLUSTER}\_{DATETIME}.tsv*: Correlation (*r*), *p* value, and adjusted *p* value for the pair of genes (Item1 and Item2) in each {CLUSTER} obtained with the indicated {METHOD}.
- *ClustersCTF\_{DATETIME}.tsv*: Averaged CTFs of the DEGs present in any cluster, indicating the number of the cluster and method where it appears.
- *CTFnormalisedCPMs\_{DATETIME}.tsv*: Normalised CPMs for each gene (rows) in each sample replicate (columns) using the CTF algorithm.
- *DEGs\_{CONTRAST}\_TREAT\_P-0.1\_FC-1.2\_{DATETIME}.tsv*: LogFC, average expression, *t* statistic, *p* value, and adjusted *p*-value, for all DEGs (rows) obtained for the {CONTRAST} using the *treat* method and the *p* and *FC* indicated in the filename.
- *DEGs\_{CONTRAST}\_eB\_P-0.1\_FC-1.2\_{DATETIME}.tsv*: LogFC, average expression, *t* statistic, *p* value, adjusted *p*-value, and *B* statistic for all DEGs (rows) for the {CONTRAST} using the *eBayes* method and the *p* and *FC* indicated in the filename.
- *filteredData\_{DATETIME}.tsv*: Raw counts in all sample replicates (columns) for genes (rows) that presented a reliable expression.
- *List\_of\_clusters\_{DATETIME}.Rds*: R object containing all the information about clusters obtained with the three clustering methods AHC, *k*-means and MBC. It must be read with the R function *readRDS* to inspect or use its contents.
- *normHomoscedCPM\_{DATETIME}.tsv*: TMM-normalised and homeoscedastic counts of reliable genes (rows) in all sample replicates (columns).
- *OutstandingGenes\_{DATETIME}.tsv*: Highly linked genes in the different clusters that can be considered hub genes.
- *TMMnormalisedCounts\_{DATETIME}.tsv*: Normalised counts for each gene (rows) in each sample replicate (columns) using the TMM algorithm.

- `TMMnormalisedCPMs-{DATETIME}.tsv`: Normalised CPMs for each gene (rows) in each sample replicate (columns) using the TMM algorithm.
- `ubiquitousDEGs_{DATETIME}.tsv`: List of genes that are DEGs in all the contrast performed in the analysis, including if it is up-regulated (1) or down-regulated (−1) in each contrast.

#### Appendix A.3. ‘Picual’ Orthologues against Wild Olive TFs

DOI: <https://doi.org/10.6084/m9.figshare.23587089> (accessed on 2 July 2023)

`Picual-TFs-orthologous-to-wildoliveTFs.tsv`: Tab separated list containing ‘Picual’ v0.61 IDs and their orthologue in *Olea europaea* var. *sylvestris* transcription factors in PlantTFDB v5.0 at [http://planttfdb.gao-lab.org/index\\_ext.php?sp=Oeu](http://planttfdb.gao-lab.org/index_ext.php?sp=Oeu) (accessed on 2 July 2023) together with the corresponding TF family.

#### Appendix A.4. Members of TF Families in Arabidopsis, Wild Olive and cv. Picual

DOI: <https://doi.org/10.6084/m9.figshare.23587125> (accessed on 2 July 2023)

`TF-families-comparison.tsv`: Tab separated list of the number of members of each transcription factor family as in PlantTFDB v5.0 for *A. thaliana* TAIR10, wild olive tree, cv. Picual v0.61, the ‘Picual’ DEGs determined in the current study, and the percent of enrichment of each TF family. The table is ordered by the last column.

#### Appendix A.5. Annotation Extension for ‘Picual’ DEGs with More TAIR10 Orthologues

DOI: <https://doi.org/10.6084/m9.figshare.23587128> (accessed on 2 July 2023)

- `DEG-annot-extension.pdf`: Discussion about the results obtained with the extended annotation of DEGs using DIAMOND to assign a TAIR 10 orthologue to unannotated ‘Picual’ transcripts.
- `orthology-Picual-TAIR10-extended-230608.tsv`: Tab separated list of cv. Picual v0.61 IDs with TAIR10 IDs of *A. thaliana* using the orthology downloaded from OliveTreeDB ([https://genomaolivar.dipujaen.es/downloads/Annotations/Olea\\_europaea\\_v061\\_annotations.xlsx](https://genomaolivar.dipujaen.es/downloads/Annotations/Olea_europaea_v061_annotations.xlsx) (accessed on 2 July 2023)) extended for the TAIR10 IDs found using DIAMOND.

#### Appendix A.6. What Is Known about Unknown ‘Picual’ DEGs

DOI: <https://doi.org/10.6084/m9.figshare.23587194> (accessed on 2 July 2023)

`unk173-ReprOlive-OeuTFs.tsv`: Tab separated list of the 173 unknown ‘Picual’ v0.61 IDs with their putative orthologue with a pollen transcript in ReprOlive, a transcript in reproductive tissues of ReprOlive, or to the tentative transcription factors predicted for wild olive (*Olea europaea* var. *sylvestris*) in PlantTFDB v5.0 at [http://planttfdb.gao-lab.org/index\\_ext.php?sp=Oeu](http://planttfdb.gao-lab.org/index_ext.php?sp=Oeu) (accessed on 2 July 2023).

#### Appendix A.7. Distribution of TFs per Family in the Contrasts Shown in Figure 2

DOI: <https://doi.org/10.6084/m9.figshare.23587209> (accessed on 2 July 2023)

`TFs-in-functions-of-contrasts.pdf`: Number of TFs per family observed in the same contrasts of Figure 2. Only TF families with at least one member in one contrast are presented in the ordinate. The number of TF is in the abscissa. Each plot corresponds to the UP or DOWN set of DEGs in 0→1 h, 1→3 h, 3→6 and 0→6 h transitions.

#### Appendix A.8. Enriched Biological Processes in Time Contrasts

DOI: <https://doi.org/10.6084/m9.figshare.23587254> (accessed on 2 July 2023)

`OE-BPs.zip`: A compressed file containing the biological processes up- and down-regulated in each time contrast as a tab separated list. Columns in each file correspond to:

- *ID*: the GO code.
- *Description*: explanation of the *ID*.
- *GeneRatio*: ratio of DEGs annotated in the *ID* term relative to the total number of DEGs in this contrast.
- *BgRatio*: ratio of all genes that are annotated in this *ID* term relative to the total number of genes in the organism (i.e., *A. thaliana*).
- *pvalue*: reliability of enrichment.
- *p.adjusted*: Benjamini and Hochberg’s multitasking adjustment of *p*-values.
- *qvalue*: John Storey’s estimation of the false discovery rate by the R package *qvalue*.
- *GenID*: TAIR10 code of *A. thaliana* orthologues for DEGs belonging to the *ID* term.

- *Count*: number of TAIR codes in the *ID* term.
- *TF*: family of transcription factors present in the *ID* term.

The *fold enrichment* can be easily calculated by dividing *geneRatio* by *BgRatio*.

#### Appendix A.9. Enriched KEGG Pathways for Time Contrasts

DOI: <https://doi.org/10.6084/m9.figshare.23587257> (accessed on 2 July 2023)

OE-KEGG-pathways.zip: A compressed file containing the KEGG pathways up- and down-regulated in each time contrast as a tab separated list. Columns in each file are the same as in Appendix A.8, but the *ID* column corresponds now to the *A. thaliana* pathway. No column for TFs is present.

#### Appendix A.10. Enrichment Results for *k*-Means Clusters

DOI: <https://doi.org/10.6084/m9.figshare.23587269> (accessed on 2 July 2023)

OE-kMeans.zip: A compressed file containing the biological processes and KEGG pathways for each of the three clusters obtained using *k*-means as tab separated lists. Columns in each file are the same as in Appendix A.8 for biological processes (files starting by BP-) and Appendix A.9 for KEGG pathways (files starting by KEGG-).

## References

- Hafidh, S.; Honys, D. Reproduction Multitasking: The Male Gametophyte. *Annu. Rev. Plant Biol.* **2021**, *72*, 581–614. [[CrossRef](#)] [[PubMed](#)]
- Weng, X.; Shen, Y.; Jiang, L.; Zhao, L.; Wang, H. Spatiotemporal organization and correlation of tip-focused exocytosis and endocytosis in regulating pollen tube tip growth. *Plant Sci.* **2023**, *330*, 111633. [[CrossRef](#)] [[PubMed](#)]
- Zhang, M.J.; Cui, J.J.; Wang, Z.M.; Dong, Y.X.; Gao, X.Q. GR1 and NTRA involved in pollen tube growth in the stigma of *Arabidopsis*. *Planta* **2023**, *258*, 1. [[CrossRef](#)]
- Wang, H.J.; Huang, J.C.; Jauh, G.Y. Pollen Germination and Tube Growth. In *Advances in Botanical Research*; Kader, J.C., Delseny, M., Eds.; Academic Press: Cambridge, MA, USA, 2010; Volume 54, pp. 1–52. [[CrossRef](#)]
- Vogler, F.; Konrad, S.S.A.; Sprunck, S. Knockin' on pollen's door: Live cell imaging of early polarization events in germinating *Arabidopsis* pollen. *Front. Plant Sci.* **2015**, *6*, 246. [[CrossRef](#)] [[PubMed](#)]
- Seitz, J.; Reimann, T.M.; Fritz, C.; Schröder, C.; Knab, J.; Weber, W.; Stadler, R. How pollen tubes fight for food: The impact of sucrose carriers and invertases of *Arabidopsis thaliana* on pollen development and pollen tube growth. *Front. Plant Sci.* **2023**, *14*, 1063765. [[CrossRef](#)]
- Hepler, P.K.; Winship, L.J. The pollen tube clear zone: Clues to the mechanism of polarized growth. *J. Integr. Plant Biol.* **2015**, *57*, 79–92. [[CrossRef](#)]
- Zhang, R.; Xu, Y.; Yi, R.; Shen, J.; Huang, S. Actin Cytoskeleton in the Control of Vesicle Transport, Cytoplasmic Organization and Pollen Tube Tip Growth. *Plant Physiol.* **2023**, kiad203. [[CrossRef](#)]
- Morales, S.; Jiménez-López, J.C.; Castro, A.J.; Rodríguez-García, M.I.; Alché, J.D. Olive pollen profilin (Ole e 2 allergen) co-localizes with highly active areas of the actin cytoskeleton and is released to the culture medium during in vitro pollen germination. *J. Microsc.* **2008**, *231*, 332–341. [[CrossRef](#)]
- Zhou, Z.; Zheng, S.; Haq, S.I.U.; Zheng, D.; Qiu, Q.S. Regulation of pollen tube growth by cellular pH and ions. *J. Plant Physiol.* **2022**, *277*, 153792. [[CrossRef](#)]
- Yan, H.; Zhuang, M.; Xu, X.; Li, S.; Yang, M.; Li, N.; Du, X.; Hu, K.; Peng, X.; Huang, W.; et al. Autophagy and its mediated mitochondrial quality control maintain pollen tube growth and male fertility in *Arabidopsis*. *Autophagy* **2023**, *19*, 768–783. [[CrossRef](#)]
- Yuan, T.L.; Huang, W.J.; He, J.; Zhang, D.; Tang, W.H. Stage-Specific Gene Profiling of Germinal Cells Helps Delineate the Mitosis/Meiosis Transition. *Plant Physiol.* **2018**, *176*, 1610–1626. [[CrossRef](#)]
- Wu, H.M.; Xie, D.J.; Jia, P.F.; Tang, Z.S.; Shi, D.Q.; Shui, G.H.; Wang, G.D.; Yang, W.C. Homeostasis of flavonoids and triterpenoids most likely modulates starch metabolism for pollen tube penetration in rice. *Plant Biotechnol. J.* **2023**, early view. [[CrossRef](#)]
- Daryanavard, H.; Postiglione, A.E.; Mühlemann, J.K.; Muday, G.K. Flavonols modulate plant development, signaling, and stress responses. *Curr. Opin. Plant Biol.* **2023**, *72*, 102350. [[CrossRef](#)]
- Cheng, J.; Wang, Z.; Yao, F.; Gao, L.; Ma, S.; Sui, X.; Zhang, Z. Down-Regulating CsHT1, a Cucumber Pollen-Specific Hexose Transporter, Inhibits Pollen Germination, Tube Growth, and Seed Development. *Plant Physiol.* **2015**, *168*, 635–647. [[CrossRef](#)]
- Kapoor, K.; Geitmann, A. Pollen tube invasive growth is promoted by callose. *Plant Reprod.* **2023**, *36*, 157–171. [[CrossRef](#)]
- Xie, B.; Deng, Y.; Kanaoka, M.M.; Okada, K.; Hong, Z. Expression of *Arabidopsis* callose synthase 5 results in callose accumulation and cell wall permeability alteration. *Plant Sci.* **2012**, *183*, 1–8. [[CrossRef](#)]
- Xia, Z.; Wen, B.; Shao, J.; Zhang, T.; Hu, M.; Lin, L.; Zheng, Y.; Shi, Z.; Dong, X.; Song, J.; et al. The transcription factor PbrbZIP52 positively affects pear pollen tube longevity by promoting callose synthesis. *Plant Physiol.* **2023**, *191*, 1734–1750. [[CrossRef](#)]
- Wang, Y.; Zhang, W.Z.; Song, L.F.; Zou, J.J.; Su, Z.; Wu, W.H. Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiol.* **2008**, *148*, 1201–1211. [[CrossRef](#)]

20. Lohani, N.; Singh, M.B.; Bhalla, P.L. High temperature susceptibility of sexual reproduction in crop plants. *J. Exp. Bot.* **2020**, *71*, 555–568. [\[CrossRef\]](#)
21. Bourguine, B.; Guihur, A. Heat Shock Signaling in Land Plants: From Plasma Membrane Sensing to the Transcription of Small Heat Shock Proteins. *Front. Plant Sci.* **2021**, *12*, 710801. [\[CrossRef\]](#)
22. Waters, E.R.; Vierling, E. Plant small heat shock proteins—Evolutionary and functional diversity. *New Phytol.* **2020**, *227*, 24–37. [\[CrossRef\]](#)
23. Hua, Z.; Vierstra, R.D. The cullin-RING ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* **2011**, *62*, 299–334. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Li, L.; Wang, K.; Zhou, Y.; Liu, X. Review: A silent concert in developing plants: Dynamic assembly of cullin-RING ubiquitin ligases. *Plant Sci.* **2023**, *330*, 111662. [\[CrossRef\]](#)
25. Lechner, E.; Achard, P.; Vansiri, A.; Potuschak, T.; Genschik, P. F-box proteins everywhere. *Curr. Opin. Plant Biol.* **2006**, *9*, 631–638. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Nguyen, K.M.; Busino, L. The Biology of F-box Proteins: The SCF Family of E3 Ubiquitin Ligases. *Adv. Exp. Med. Biol.* **2020**, *1217*, 111–122. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Abd-Hamid, N.A.; Ahmad-Fauzi, M.I.; Zainal, Z.; Ismail, I. Diverse and dynamic roles of F-box proteins in plant biology. *Planta* **2020**, *251*, 68. [\[CrossRef\]](#)
28. Wei, L.Q.; Xu, W.Y.; Deng, Z.Y.; Su, Z.; Xue, Y.; Wang, T. Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. *BMC Genom.* **2010**, *11*, 338. [\[CrossRef\]](#)
29. Zhao, L.J.; Yuan, H.M.; Guo, W.D.; Yang, C.P. Digital Gene Expression Analysis of *Populus simonii* × *P. nigra* Pollen Germination and Tube Growth. *Front. Plant Sci.* **2016**, *7*, 825. [\[CrossRef\]](#)
30. Motomura, K.; Sugi, N.; Takeda, A.; Yamaoka, S.; Maruyama, D. Possible molecular mechanisms of persistent pollen tube growth without de novo transcription. *Front. Plant Sci.* **2022**, *13*, 1020306. [\[CrossRef\]](#)
31. Suárez, C.; Castro, A.J.; Rapoport, H.F.; Rodríguez-García, M.I. Morphological, histological and ultrastructural changes in the olive pistil during flowering. *Sex. Plant Reprod.* **2012**, *25*, 133–146. [\[CrossRef\]](#)
32. Carmona, R.; Zafra, A.; Seoane, P.; Castro, A.J.; Guerrero-Fernández, D.; Castillo-Castillo, T.; Medina-García, A.; Cánovas, F.M.; Aldana-Montes, J.F.; Navas-Delgado, I.; et al. ReprOlive: A database with linked data for the olive tree (*Olea europaea* L.) reproductive transcriptome. *Front. Plant Sci.* **2015**, *6*, 625. [\[CrossRef\]](#)
33. Jiménez-Ruiz, J.; Leyva-Pérez, M.d.l.O.; Schilirò, E.; Barroso, J.B.; Bombarely, A.; Mueller, L.; Mercado-Blanco, J.; Luque, F. Transcriptomic Analysis of L. Roots during the Early Infection Process. *Plant Genome* **2017**, *10*, plantgenome2016-07. [\[CrossRef\]](#)
34. Zafra, A.; Carmona, R.; Traverso, J.A.; Hancock, J.T.; Goldman, M.H.S.; Claros, M.G.; Hiscock, S.J.; Alche, J.D. Identification and Functional Annotation of Genes Differentially Expressed in the Reproductive Tissues of the Olive Tree (*Olea europaea* L.) through the Generation of Subtractive Libraries. *Front. Plant Sci.* **2017**, *8*, 1576. [\[CrossRef\]](#)
35. Fernández-Parras, I.; Ramírez-Tejero, J.A.; Luque, F.; Navarro, F. Several Isoforms for Each Subunit Shared by RNA Polymerases are Differentially Expressed in the Cultivated Olive Tree (*Olea europaea* L.). *Front. Mol. Biosci.* **2021**, *8*, 679292. [\[CrossRef\]](#)
36. Ramírez-Tejero, J.A.; Jiménez-Ruiz, J.; Leyva-Pérez, M.d.l.O.; Barroso, J.B.; Luque, F. Gene Expression Pattern in Olive Tree Organs (*Olea europaea* L.). *Genes* **2020**, *11*, 544. [\[CrossRef\]](#)
37. Bullones, A.; Castro, A.J.; Lima-Cabello, E.; Alché, J.d.D.; Luque, F.; Claros, M.G.; Fernandez-Pozo, N. OliveAtlas: A Gene Expression Atlas Tool for *Olea europaea*. *Plants* **2023**, *12*, 1274. [\[CrossRef\]](#)
38. Conde, C.; Delrot, S.; Gerós, H. Physiological, biochemical and molecular changes occurring during olive development and ripening. *J. Plant Physiol.* **2008**, *165*, 1545–1562. [\[CrossRef\]](#)
39. Iaria, D.L.; Chiappetta, A.; Muzzalupo, I. A De novo Transcriptomic Approach to Identify Flavonoids and Anthocyanins “Switch-Off” in Olive (*Olea europaea* L.) Drupes at Different Stages of Maturation. *Front. Plant Sci.* **2015**, *6*, 1246. [\[CrossRef\]](#)
40. Julca, I.; Marcet-Houben, M.; Cruz, F.; Gómez-Garrido, J.; Gaut, B.S.; Díez, C.M.; Gut, I.G.; Alioto, T.S.; Vargas, P.; Gabaldón, T. Genomic evidence for recurrent genetic admixture during the domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC Biol.* **2020**, *18*, 148. [\[CrossRef\]](#)
41. Skodra, C.; Titeli, V.S.; Michailidis, M.; Bazakos, C.; Ganopoulos, I.; Molassiotis, A.; Tanou, G. Olive Fruit Development and Ripening: Break on through to the “-Omics” Side. *Int. J. Mol. Sci.* **2021**, *22*, 5806. [\[CrossRef\]](#)
42. Rodríguez-García, M.I.; Alaoui, M.M.; Díaz, J.D.L.F.; del Carmen Fernández, M. Observations on Microtubules and Nuclei Motility in the Pollen Tube of Olive (*Olea europaea* L.). *Acta Biol. Cracoviensia* **2003**, *45*, 97–101.
43. Hernández, M.L.; Lima-Cabello, E.; Alché, J.d.D.; Martínez-Rivas, J.M.; Castro, A.J. Lipid Composition and Associated Gene Expression Patterns during Pollen Germination and Pollen Tube Growth in Olive (*Olea europaea* L.). *Plant Cell Physiol.* **2020**, *61*, 1348–1364. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Iaria, D.; Chiappetta, A.; Muzzalupo, I. De Novo Transcriptome Sequencing of *Olea europaea* L. to Identify Genes Involved in the Development of the Pollen Tube. *Sci. World J.* **2016**, *2016*, 4305252. [\[CrossRef\]](#)
45. Jiménez-Ruiz, J.; Ramírez-Tejero, J.A.; Fernández-Pozo, N.; Leyva-Pérez, M.d.l.O.; Yan, H.; Rosa, R.d.l.; Belaj, A.; Montes, E.; Rodríguez-Ariza, M.O.; Navarro, F.; et al. Transposon activation is a major driver in the genome evolution of cultivated olive trees (*Olea europaea* L.). *Plant Genome* **2020**, *13*, e20010. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Wang, P.; Li, Z.; Wei, J.; Zhao, Z.; Sun, D.; Cui, S. A Na<sup>+</sup>/Ca<sup>2+</sup> exchanger-like protein (AtNCL) involved in salt stress in *Arabidopsis*. *J. Biol. Chem.* **2012**, *287*, 44062–44070. [\[CrossRef\]](#) [\[PubMed\]](#)

47. Boekel, J.; Chilton, J.M.; Cooke, I.R.; Horvatovich, P.L.; Jagtap, P.D.; Käll, L.; Lehtiö, J.; Lukasse, P.; Moerland, P.D.; Griffin, T.J. Multi-omic data analysis using Galaxy. *Nat. Biotechnol.* **2015**, *33*, 137–139. [[CrossRef](#)]
48. McCarthy, D.J.; Smyth, G.K. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* **2009**, *25*, 765–771. [[CrossRef](#)]
49. Usadel, B.; Obayashi, T.; Mutwil, M.; Giorgi, F.M.; Bassel, G.W.; Tanimoto, M.; Chow, A.; Steinhäuser, D.; Persson, S.; Provart, N.J. Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats. *Plant Cell Environ.* **2009**, *32*, 1633–1651. [[CrossRef](#)]
50. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)]
51. McCarthy, D.J.; Chen, Y.; Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **2012**, *40*, 4288–4297. [[CrossRef](#)]
52. Provart, N. Correlation networks visualization. *Front. Plant Sci.* **2012**, *3*, 240. [[CrossRef](#)]
53. Dillies, M.A.; Rau, A.; Aubert, J.; Hennequet-Antier, C.; Jeanmougin, M.; Servant, N.; Keime, C.; Marot, G.; Castel, D.; Estelle, J.; et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.* **2013**, *14*, 671–683. [[CrossRef](#)]
54. Sonesson, C.; Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinform.* **2013**, *14*, 91. [[CrossRef](#)]
55. Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* **2014**, *61*, 1–36. [[CrossRef](#)]
56. Law, C.W.; Chen, Y.; Shi, W.; Smyth, G.K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **2014**, *15*, R29. [[CrossRef](#)]
57. Fortunato, S.; Hric, D. Community detection in networks: A user guide. *Phys. Rep.* **2016**, *659*, 1–44. [[CrossRef](#)]
58. Law, C.W.; Alhamdoosh, M.; Su, S.; Dong, X.; Tian, L.; Smyth, G.K.; Ritchie, M.E. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* **2016**, *5*, 1408. [[CrossRef](#)]
59. Lin, Y.; Golovnina, K.; Chen, Z.X.; Lee, H.N.; Negron, Y.L.S.; Sultana, H.; Oliver, B.; Harbison, S.T. Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*. *BMC Genom.* **2016**, *17*, 28. [[CrossRef](#)]
60. Scrucca, L.; Fop, M.; Murphy, T.B.; Raftery, A.E. mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **2016**, *8*, 289–317. [[CrossRef](#)]
61. Emamjomeh, A.; Saboori Robot, E.; Zahiri, J.; Solouki, M.; Khosravi, P. Gene co-expression network reconstruction: A review on computational methods for inferring functional information from plant-based expression data. *Plant Biotechnol. Rep.* **2017**, *11*, 71–86. [[CrossRef](#)]
62. Contreras-López, O.; Moyano, T.C.; Soto, D.C.; Gutiérrez, R.A. Step-by-Step Construction of Gene Co-expression Networks from High-Throughput Arabidopsis RNA Sequencing Data. *Methods Mol. Biol.* **2018**, *1761*, 275–301. [[CrossRef](#)]
63. Amrhein, V.; Greenland, S.; McShane, B. Scientists rise up against statistical significance. *Nature* **2019**, *567*, 305–307. [[CrossRef](#)] [[PubMed](#)]
64. Parsana, P.; Ruberman, C.; Jaffe, A.E.; Schatz, M.C.; Battle, A.; Leek, J.T. Addressing confounding artifacts in reconstruction of gene co-expression networks. *Genome Biol.* **2019**, *20*, 94. [[CrossRef](#)] [[PubMed](#)]
65. Wasserstein, R.L.; Schirm, A.L.; Lazar, N.A. Moving to a World Beyond “ $p < 0.05$ ”. *Am. Stat.* **2019**, *73*, 1–19. [[CrossRef](#)]
66. Arbet, J.; Zhuang, Y.; Litkowski, E.; Saba, L.; Kechris, K. Comparing Statistical Tests for Differential Network Analysis of Gene Modules. *Front. Genet.* **2021**, *12*, 630215. [[CrossRef](#)] [[PubMed](#)]
67. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [[CrossRef](#)] [[PubMed](#)]
68. Johnson, K.A.; Krishnan, A. Robust normalization and transformation techniques for constructing gene coexpression networks from RNA-seq data. *Genome Biol.* **2022**, *23*, 1. [[CrossRef](#)]
69. Li, Y.; Ge, X.; Peng, F.; Li, W.; Li, J.J. Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biol.* **2022**, *23*, 79. [[CrossRef](#)]
70. Sia, J.; Zhang, W.; Jonckheere, E.; Cook, D.; Bogdan, P. Inferring functional communities from partially observed biological networks exploiting geometric topology and side information. *Sci. Rep.* **2022**, *12*, 10883. [[CrossRef](#)]
71. Conze, L.L.; Berlin, S.; Le Bail, A.; Kost, B. Transcriptome profiling of tobacco (*Nicotiana tabacum*) pollen and pollen tubes. *BMC Genom.* **2017**, *18*, 581. [[CrossRef](#)]
72. Wang, H.; Qi, X.; Chen, S.; Feng, J.; Chen, H.; Qin, Z.; Deng, Y. An integrated transcriptomic and proteomic approach to dynamically study the mechanism of pollen-pistil interactions during jasmine crossing. *J. Proteom.* **2021**, *249*, 104380. [[CrossRef](#)]
73. Lu, W.; Zhou, Q.; Chen, Y. Impact of RNA degradation on next-generation sequencing transcriptome data. *Genomics* **2022**, *114*, 110429. [[CrossRef](#)]
74. Spies, D.; Renz, P.F.; Beyer, T.A.; Ciaudo, C. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Brief. Bioinform.* **2019**, *20*, 288–298. [[CrossRef](#)]

75. Zhou, H.; Yin, H.; Chen, J.; Liu, X.; Gao, Y.; Wu, J.; Zhang, S. Gene-expression profile of developing pollen tube of *Pyrus bretschneideri*. *Gene Expr. Patterns* **2016**, *20*, 11–21. [[CrossRef](#)]
76. Lang, V.; Usadel, B.; Obermeyer, G. De novo sequencing and analysis of the lily pollen transcriptome: An open access data source for an orphan plant species. *Plant Mol. Biol.* **2015**, *87*, 69–80. [[CrossRef](#)]
77. Loraine, A.E.; McCormick, S.; Estrada, A.; Patel, K.; Qin, P. RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. *Plant Physiol.* **2013**, *162*, 1092–1109. [[CrossRef](#)]
78. Testillano, P.S. Microspore embryogenesis: Targeting the determinant factors of stress-induced cell reprogramming for crop improvement. *J. Exp. Bot.* **2019**, *70*, 2965–2978. [[CrossRef](#)]
79. Ibáñez, S.; Carneros, E.; Testillano, P.S.; Pérez-Pérez, J.M. Advances in Plant Regeneration: Shake, Rattle and Roll. *Plants* **2020**, *9*, 897. [[CrossRef](#)]
80. Zienkiewicz, K.; Zienkiewicz, A.; Rodríguez-García, M.I.; Castro, A.J. Characterization of a caleosin expressed during olive (*Olea europaea* L.) pollen ontogeny. *BMC Plant Biol.* **2011**, *11*, 122. [[CrossRef](#)]
81. Ramming, A.; Kappel, C.; Kanaoka, M.M.; Higashiyama, T.; Lenhard, M. Poly(A) polymerase 1 contributes to competence acquisition of pollen tubes growing through the style in *Arabidopsis thaliana*. *Plant J.* **2023**, *114*, 651–667. [[CrossRef](#)]
82. Moser, M.; Kirkpatrick, A.; Groves, N.R.; Meier, I. LINC-complex mediated positioning of the vegetative nucleus is involved in calcium and ROS signaling in *Arabidopsis* pollen tubes. *Nucleus* **2020**, *11*, 149–163. [[CrossRef](#)]
83. Vidali, L.; McKenna, S.T.; Hepler, P.K. Actin polymerization is essential for pollen tube growth. *Mol. Biol. Cell* **2001**, *12*, 2534–2545. [[CrossRef](#)] [[PubMed](#)]
84. Çetinbaş-Genç, A.; Conti, V.; Cai, G. Let's shape again: The concerted molecular action that builds the pollen tube. *Plant Reprod.* **2022**, *35*, 77–103. [[CrossRef](#)] [[PubMed](#)]
85. Franklin-Tong. Signaling and the modulation of pollen tube growth. *Plant Cell* **1999**, *11*, 727–738. [[CrossRef](#)] [[PubMed](#)]
86. Bushart, T.J.; Roux, S.J. Conserved features of germination and polarized cell growth: A few insights from a pollen-fern spore comparison. *Ann. Bot.* **2007**, *99*, 9–17. [[CrossRef](#)] [[PubMed](#)]
87. Beiki, H.; Nejati-Javaremi, A.; Pakdel, A.; Masoudi-Nejad, A.; Hu, Z.L.; Reecy, J.M. Large-scale gene co-expression network as a source of functional annotation for cattle genes. *BMC Genom.* **2016**, *17*, 846. [[CrossRef](#)] [[PubMed](#)]
88. Zainal-Abidin, R.A.; Harun, S.; Vengatharajuloo, V.; Tamizi, A.A.; Samsulrizal, N.H. Gene Co-Expression Network Tools and Databases for Crop Improvement. *Plants* **2022**, *11*, 1625. [[CrossRef](#)]
89. Pérez Di Giorgio, J.A.; Barberini, M.L.; Amodeo, G.; Muschietti, J.P. Pollen aquaporins: What are they there for? *Plant Signal. Behav.* **2016**, *11*, e1217375. [[CrossRef](#)]
90. Egan, E.D.; Braun, C.R.; Gygi, S.P.; Moazed, D. Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *RNA* **2014**, *20*, 867–881. [[CrossRef](#)]
91. Buchfink, B.; Reuter, K.; Drost, H.G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **2021**, *18*, 366–368. [[CrossRef](#)]
92. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [[CrossRef](#)]
93. Falgueras, J.; Lara, A.J.; Fernandez-Pozo, N.; Canton, F.R.; Perez-Trabado, G.; Claros, M.G. SeqTrim: A high-throughput pipeline for preprocessing any type of sequence reads. *BMC Bioinform.* **2010**, *11*, 38. [[CrossRef](#)]
94. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)]
95. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCftools. *Gigascience* **2021**, *10*, giab008. [[CrossRef](#)]
96. Liu, X.; Li, N.; Liu, S.; Wang, J.; Zhang, N.; Zheng, X.; Leung, K.S.; Cheng, L. Normalization Methods for the Analysis of Unbalanced Transcriptome Data: A Review. *Front. Bioeng. Biotechnol.* **2019**, *7*, 358. [[CrossRef](#)]
97. Fraley, C.; Raftery, A.E. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Stat. Assoc.* **2002**, *97*, 611–631. [[CrossRef](#)]
98. Langfelder, P.; Zhang, B.; Horvath, S. Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **2008**, *24*, 719–720. [[CrossRef](#)]
99. McNicholas, P.D. Model-Based Clustering. *J. Classif.* **2016**, *33*, 331–373. [[CrossRef](#)]
100. Kleinberg, J.M. Authoritative Sources in a Hyperlinked Environment. *J. ACM* **1999**, *46*, 604–632. [[CrossRef](#)]
101. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)]
102. Mi, H.; Muruganujan, A.; Huang, X.; Ebert, D.; Mills, C.; Guo, X.; Thomas, P.D. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **2019**, *14*, 703–721. [[CrossRef](#)]
103. Jin, J.; Tian, F.; Yang, D.C.; Meng, Y.Q.; Kong, L.; Luo, J.; Gao, G. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2017**, *45*, D1040–D1045. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.