

RESOURCE

A transcriptome atlas of zygotic and somatic embryogenesis in Norway spruce

Katja Stojković¹, Camilla Canovi², Kim-Cuong Le³, Iftikhar Ahmad¹, Ioana Gaboreanu¹, Sofie Johansson¹, Nicolas Delhomme¹, Ulrika Egertsdotter^{1,4,*}  and Nathaniel R. Street^{2,5,*} 

¹Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-901 87 Umeå, Sweden,

²Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, SE-901 87 Umeå, Sweden,

³G.W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA,

⁴Renewable Bioproducts Institute, Georgia Institute of Technology Atlanta, Georgia, USA, and

⁵SciLifeLab, Department of Plant Physiology, Umeå University, SE-901 87 Umeå, Sweden

Received 14 April 2024; revised 4 October 2024; accepted 8 October 2024; published online 27 October 2024.

*For correspondence (e-mail nathaniel.street@umu.se and ulrika.egertsdotter@slu.se).

SUMMARY

Somatic embryogenesis (SE) is a powerful model system for studying embryo development and an important method for scaling up availability of elite and climate-adapted genetic material of Norway spruce (*Picea abies* L. Karst). However, there are several steps during the development of the somatic embryo (Sem) that are suboptimal compared to zygotic embryo (Zem) development. These differences are poorly understood and result in substantial yield losses during plant production, which limits cost-effective large-scale production of SE plants. This study presents a comprehensive data resource profiling gene expression during zygotic and somatic embryo development to support studies aiming to advance understanding of gene regulatory programmes controlling embryo development. Transcriptome expression patterns were analysed during zygotic embryogenesis (ZE) in Norway spruce, including separated samples of the female gametophytes and Zem, and at multiple stages during SE. Expression data from eight developmental stages of SE, starting with pro-embryogenic masses (PEMs) up until germination, revealed extensive modulation of the transcriptome between the early and mid-stage maturing embryos and at the transition of desiccated embryos to germination. Comparative analysis of gene expression changes during ZE and SE identified differences in the pattern of gene expression changes and functional enrichment of these provided insight into the associated biological processes. Orthologs of transcription factors known to regulate embryo development in angiosperms were differentially regulated during Zem and Sem development and in the different zygotic embryo tissues, providing clues to the differences in development observed between Zem and Sem. This resource represents the most comprehensive dataset available for exploring embryo development in conifers.

Keywords: Norway spruce, *Picea abies*, embryogenesis, embryo, transcriptome, differential expression.

INTRODUCTION

The gymnosperm and angiosperm lineages diverged about 300 million years ago, resulting in different morphological and functional pathways for key steps of development, including embryogenesis (La Torre et al., 2019). The high similarity between zygotic embryogenesis (ZE; Cairney & Pullman, 2007) and somatic embryogenesis (SE) in conifers renders SE a powerful model system to study

embryogenesis, for example, to elucidate gene expression regulatory mechanisms (reviewed in Trontin, Klimaszewska, Morel, Hargreaves, & Lelu-Walter, 2016). A major strength of SE as a model system is the ability to produce an almost unlimited number of somatic embryos (Sem) of the same genotype that can be sampled at distinct, synchronised stages of the developmental process. SE is particularly effective in *Picea* species as the process works

efficiently across many genotypes (Mamun, Aidun, & Egertsdotter, 2018). This strong potential for clonal multiplication additionally makes SE a valuable tool for the forestry industry to obtain clonally propagated conifer plants derived from elite varieties where clonal propagation by cuttings or other techniques is limited or absent (Egertsdotter, Ahmad, & Clapham, 2019, Rosvall, Bradshaw, Egertsdotter, Ingvarsson, & Wu, 2019).

In *P. abies*, zygotic embryo (Zem) development starts in the same year after pollination and continues until a fully mature, and desiccation-tolerant embryo has formed 8–12 weeks later. For most conifers, embryo development progresses through three stages: proembryogeny, early embryogeny and late embryogeny (Singh, 1978). Proembryo formation is characterised by a free nuclear stage ending when the proembryo elongates into the corrosion cavity of the female gametophyte (FG), followed by the further elongation of the suspensor mass and development of a generative root meristem during early embryogeny. Root and shoot meristem development is concluded during late embryogeny (Singh, 1978). During seed development, programmed cell death of FG cells adjacent to the embryo also provides nutrients for the growing embryo (Durzan, 2012). FG tissue further supports the onset of germination by allocating storage reserves to the developing Zem and remains active until seed imbibition (Vuosku et al., 2009).

The *in vitro* SE process in conifers is typically started from an immature Zem that will produce early stage somatic embryos in response to auxin and cytokinin. Recently, a low frequency of Sem initiation from shoot buds of *P. abies* was, however, demonstrated from four- to six-year-old plants that had originated from Sem (Varis, Klimaszewska, & Aronen, 2018). The induced early stage somatic embryos then continue to multiply and form pro-embryogenic masses (PEMs) that can be captured for continued multiplication. Maturation can then be induced by replacing the cytokinin and auxin with abscisic acid (ABA) and an increased osmotic potential. The fully mature somatic embryos are often subjected to a period of drying by exposure to a relatively drier gas phase to become partially desiccated before the onset of germination. After germination, embryos showing root and shoot development can be acclimated to *ex vitro* conditions under controlled humidity and temperature (Egertsdotter, 2018).

The SE process in conifers is slow and there can be loss of genotypes in addition to limitations in the number of embryos produced at each step. Efforts to optimise SE protocols have rendered some improvements but further optimisations are required to enable cost-effective SE plant production across genotypes. In particular, the formation of fully mature, desiccation-tolerant embryos is still a limiting factor in most conifer species. Compared to the Zem, the Sem lacks the FG supporting tissue and, arguably,

components of the FG are missing from the SE culture medium. The composition and metabolism of the FG tissue have been analysed to inform attempts to improve the SE processes by amendment of the *in vitro* culture medium to more closely resemble the FG environment of the Zem (reviewed by Pullman & Bucalo, 2014).

While gene expression during embryo development has been extensively studied in angiosperm model systems, efforts to analyse global transcript abundance during conifer embryo development were previously hampered by a lack of reference genome assemblies due to the large genome sizes of gymnosperms, which average 200 times the size of the Arabidopsis genome (Mackay et al., 2012). However, it has been demonstrated that gene families active during embryo development in angiosperms are also active in conifers, with differential expression (DE) of epigenetic regulators and transcription factors (TFs) implicated as developmental regulators in angiosperms being detected by microarray analysis in Zem of *Pinus pinaster* (De Vega-Bartol et al., 2013), and RNA-Sequencing of Zem and FG respectively of *Pinus sylvestris* (Merino et al., 2016) and whole seeds of *Picea mongolica* (Yan, Buer, Wang, Zhula, & Bai, 2021). Efforts to analyse global transcript abundance during conifer embryo development have otherwise mainly focused on SE material due to the easy access of embryos at different developmental stages and the possibilities to analyse embryos under different growth conditions and from cell lines with varying degrees of developmental arrest (reviewed in Trontin et al., 2016). SE cultures composed of early stage embryos (PEMs) also offer the advantage of being amenable to genetic transformation to test the function of candidate genes with regulatory functions during embryo development (e.g. Vetrici, Yevtushenko, & Misra, 2021) and have recently been shown to be amenable to genomic modifications via CRISPR-Cas, permitting more in-depth analyses of developmental regulators of embryo development in *Pinus radiata* (Poovaiah et al., 2021) and *Picea glauca* (Cui et al., 2021).

Transcript profiling of conifer zygotic and somatic embryos at corresponding developmental stages has also been conducted with the aim of identifying discrepancies between the respective developmental stages. The SE process for scalable plant production in conifers is suboptimal in most and blocked in some, species. As such, learning from the natural processes during ZE could help identify approaches to improve SE plant production methods. This approach was explored early on by using a 326 cDNA fragment microarray in *Pinus taeda* to compare the most mature stage of Sem to the full series of developmental stages of Zem. From this comparison of a limited number of gene fragments it appeared that the most mature stage reached by the Sem *in vitro* did not correspond to the same level of maturity as Zem, suggesting that Sem

development in *P. taeda* was blocked at a pre-mature stage (Cairney, Xu, Mackay, & Pullman, 2000; Pullman, Johnson, Peter, Cairney, & Xu, 2003). More recently, gene expression in Zem and FG at three different developmental stages during seed development was compared in two cell lines with different Sem maturation competencies in *Araucaria angustifolia* (Eibl et al., 2015). The results from the comparison of gene expression indicated that the cell line with blocked embryo maturation had a disturbed auxin distribution pattern.

In the present study, a comprehensive developmental time series of *P. abies* Zem and their associated FG were profiled by RNA-Sequencing (RNA-Seq) at biweekly intervals. Nine time points were covered, starting 2 weeks after pollination and ending with a fully mature and desiccated embryo. Furthermore, eight developmental stages of somatic embryos were profiled by RNA-Seq and tentatively aligned with corresponding stages of zygotic embryos based on embryo morphologies and differential gene expression results. Extensive modulation of the transcriptome at specific points during SE and ZE development was detected, with DE analysis conducted to identify candidate genes and biological processes involved in the *P. abies* embryo development transcriptional program. There was extensive DE of genes at the major developmental stage transitions, with the largest changes occurring between the early and mid-stage maturing somatic embryos and at the transition of desiccated embryos to germination for zygotic embryos. Overall, major differences in gene expression patterns were detected between Zem and Sem development, revealing differences in the two developmental processes. Furthermore, differences in expression patterns of TF orthologs with key functions in angiosperms were detected during Zem and Sem development. The data presented here have been integrated into PlantGenIE.org (Sundell et al., 2015) to serve as a community resource for fundamental studies into the mechanisms of conifer embryo development.

RESULTS AND DISCUSSION

Overview of gene expression during ZE

RNA-Sequencing at nine stages of ZE, from the seed two weeks after pollination to a fully desiccated seed, was performed to assay the transcriptome during ZE (Figure 1a). Whole seeds (SD) were sampled at the first three profiled time points (Z1–Z3) while at subsequent time points (Z4–Z9) embryos (Zem) were isolated from seeds and separated from the female gametophytes (FG). We note that samples were not 'staged' at each sampling point. Each sample point occurred a specific number of days after controlled pollination and each biological replicate represented a pool of numerous seeds, Zem or FG samples (see methods for details). As such, there would have been

some degree of heterogeneity in developmental progression among the individual seeds used to form each biological replicate pool at each sampled stage, with the pool representing the average state of development at the sampled time. Previous work in *Picea engelmannii* × *glauca* following development of the FG and the post-fertilisation events described by ultrastructural analyses revealed that fertilisation takes place around 2 weeks after pollination (Owens & Molder, 1984) suggesting that the first developmental stages of seeds sampled in this study 2 weeks after pollination would correspond to seeds around the time when fertilisation occurred, followed by proembryo development. The second sample would then have contained proembryos represented by the free-nuclear stage of the zygote. Proembryogeny lasts for about a week in both *P. glauca* and *P. engelmannii* (Owens & Molder, 1984) but has been observed to last for around 2 weeks in *P. abies* (Hakman, 1993) and ends when the embryonal cells are pushed into the corrosion cavity by the suspensor cells to start early embryogeny (Singh, 1978). The appearances of the female gametophytes (FGs) of the seed sampled for this study were strikingly different between the two first developmental stages and the third stage. At stage 3 (ZE3), the FG was considerably larger and had changed from a transparent appearance to opaque, whitish colouring. This supports the assumption that for this present study, the first two stages corresponded to proembryogeny and that by the third collection time, seeds had started to transition to early embryogeny. Further support comes from the observation of the Zem resembling the typical early embryo stages with a distinctive embryo—proper at the fourth collection time (Figure 1a). The phase of early embryogeny then continues until late embryogeny starts about 8 weeks later (Owens & Molder, 1979). The appearances of the sampled FGs remained the same from the third stage to the last stage, which contained the fully mature and desiccated embryo. Examination of a principal component analysis (PCA) plot revealed that samples from FG and Zem were clearly separated by tissue type and showed a time-dependent pattern until stage Z6 (Figure 1b). Samples from the latest stages (Z7–Z9) clustered together, indicating that no major transcriptional changes occurred after stage Z7 suggesting that these stages corresponded to the desiccated and dormant seed. Samples from the first two stages (Z1 and Z2) were more similar to each other than to the third stage (Z3), which clustered closer to the FG samples. All three stages (Z1–Z3) represented the whole seed before it was possible to separate FG from Zem. While clear sample clustering was seen, principal components 1 and 2 explained only 47% (26% and 21%, respectively) of total variance.

Across all sampled ZE tissue types (seed, zygotic embryo and female megagametophyte) 36 718 genes had detectable expression of which 40.0% (14686) were

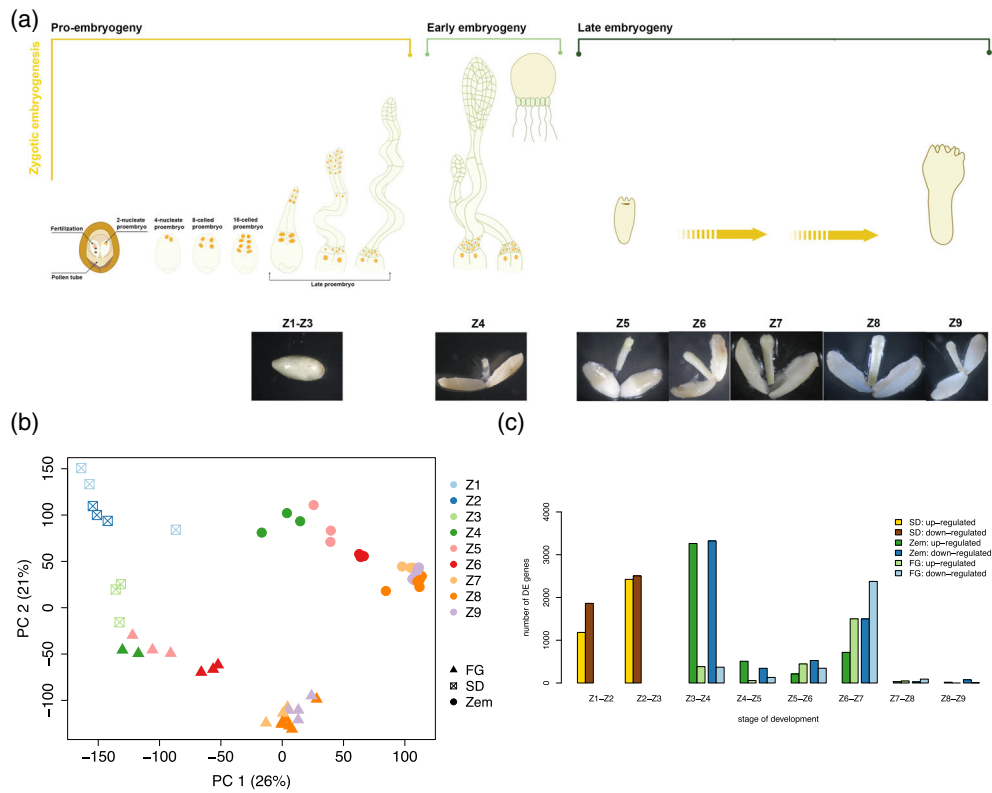


Figure 1. (a) Diagrammatic representation of the stages sampled at each sampling time during zygotic embryogenesis. Representative photos of each sampled stage are shown below the diagram. Each sample consisted of a pool of a minimum of 100 seeds/embryos and three replicate pools were sampled at each sampling point. Pollination was performed on 19 June 2017. The first samples were collected 2 weeks after pollination and every 2 weeks subsequently until 18 weeks after pollination.

(b) Principal component analysis plot of the RNA-Sequencing data showing the first two principal components. Different symbols indicate sample type and colours indicate sample stage.

(c) Bar graph representation of the number of differentially expressed genes (FDR-adjusted P value < 0.01 and $|\log_2$ fold change ≥ 0.5). Up-regulated genes are shown in dark (Zygotic embryo; Zem) or light (Female gametophyte; FG) green. Down-regulated genes are shown in dark (Zem) or light (FG) blue. Up-regulated genes in the sampled seed (SD) stages are shown in yellow and down-regulated genes in the SD stages are shown in brown.

differentially expressed (DE) in at least one stage transition in one of the tissues. Differentially expressed genes (DEGs) at any time in the seed, Zem and FG represented 19.7% (7230 genes), 23.8% (8749 genes) and 13.8% (5050 genes), respectively, of all ZE expressed genes. The largest numbers of DEGs in ZE (Figure 1c) were observed firstly in the early stages of seed development, from Z1 to Z4 (3046, 4928, 7048 DEGs between Z1–Z2, Z2–Z3 and Z3–Z4, respectively), which is when the transition from proembryogeny to early embryogeny is expected to take place (Owens & Molder, 1984). In *Picea abies*, the optimal time for SE initiations from the Zem is around stage Z4, suggesting that some of the observed DEGs play a role in the dedifferentiation process when PEMs are formed. The next stage showing the largest numbers in DEGs was between Z6 and Z7 (4940 DEGs), which is around the time when the switch from early embryogeny to late embryogeny occurs (Owens & Molder, 1984). There were 792 and 212 DEGs common to the two-stage transitions with the largest number of DEGs (Z3–Z4 and Z6–Z7) in Zem and FG, respectively.

The FG supports embryo development by allocating storage reserves to the developing embryo until the onset of germination and stays active until seed imbibition, as observed in *Picea glauca* (He & Kermodé, 2003) and *Pinus silvestris* (Vuosku et al., 2009). This is in agreement with the results from the functional enrichment analysis in the present study (discussed below). There were 5807 DEGs between Zem and FG considering all stage comparisons, with more than 1500 DEGs between the two tissue types at every time point. When comparing separated tissues of the seed from stage Z4 to the combined seed sample (Z3), a larger number of DEGs was observed for the Zem than for FG. The large number of DEGs at Z3–Z4 likely results from comparing the combined sample types (Z3) to each in isolation (Z4) and suggests that before Z4 the assayed transcriptome was dominated by FG with the FG showing little change in the transcriptome during these stages. In contrast, at the time of the second biggest change in the number of DEGs in the Zem (Z6–Z7), presumably corresponding to accumulation of storage compounds in

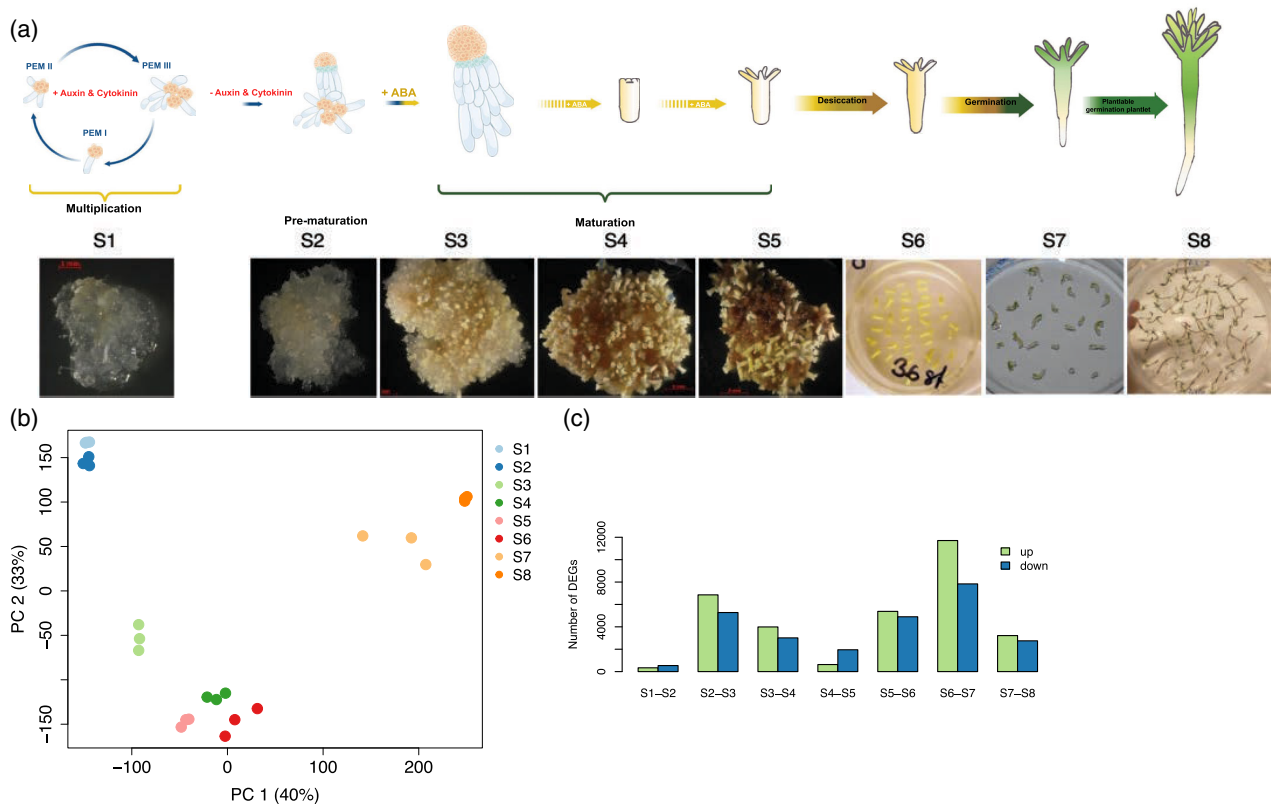


Figure 2. (a) Diagrammatic representation of the sampled stages during somatic embryogenesis. Representative photos of each sampled stage are shown below the diagram. Each sample consisted of a pool of >100 (50 mg) of pro-embryogenic masses (PEMs; S1 and S2) that start developing into embryos (from S3). Three replicate pools were sampled at each sampling stage. S1 corresponds to pro-embryogenic masses (PEMs) collected from proliferation medium; S2 was PEMs collected after 1 week on pre-maturation medium; S3, S4 and S5 were samples from the SE culture with maturing embryos collected after three, six and eight weeks on maturation medium; S6 was mature isolated embryos collected after 3 weeks of desiccation; S7 was mature, desiccated embryos germinated for 1 week to show early root formation, and S8 correspond to germinants ready to be planted and be established in compost (Figure 1). (b) Principal component analysis plot of the RNA-Sequencing data showing the first two principal components. Colours indicate sample stage. (c) Bar graph representation of the number of differentially expressed genes (FDR-adjusted P value <0.01 and $|\log_2$ fold change| ≥ 0.5). Down-regulated genes are coloured blue and up-regulated genes green.

preparation for desiccation, there were more DEGs in the FG. Functional enrichment analysis (Table S1) of the 5807 DEGs between the FG and Zem at any point of the experiment identified significant enrichment for cellular nitrogen compound metabolic process, multiple terms connected to chromatin organisation and terms connected to transcription and translation. Functional enrichments of DEGs at the major points of transition in both tissue types were also examined to gain insight into the key biological processes active (Table S1). For Zem there was significant enrichment of terms associated with chromosome organisation, nucleotide phosphorylation, biosynthetic process, rRNA metabolic process and DNA packaging at the Z3–Z4 transition, suggesting active transcription, changes to organisation of chromatin and biosynthesis. At Z6–Z7 there was enrichment for lipid metabolic process, organophosphate metabolic process and carbohydrate metabolic process. In the FG, the most marked change in the transcriptome occurred at Z6–Z7, with a clear bias towards down-regulation of

transcripts. These DEGs were enriched for carbohydrate metabolic process, cell wall organisation or biogenesis and external encapsulating structure organisation.

Overview of gene expression during SE

RNA-Sequencing to profile gene expression was performed at eight stages of SE in Norway spruce, from the earliest embryo stage in culture, composed of PEMs, to an established growing plant (Figure 2a). PEMs are primarily made up of early stage somatic embryos composed of an embryonic region associated with vacuolated, elongated suspensor cells (Filonova, Bozhkov, & Von Arnold, 2000).

The relationship among sampled stages was examined using PCA, in which the first two principal components (PCs) together explained 73% of variation in the data (Figure 2b). The first PC (40% variance explained) separated plant (S7 and S8) and embryogenic stages of the profiled development series, while the second PC (33% variance explained) separated embryonic stages. Stages

S1 and S2 clustered together, indicating high similarity in the transcriptome of samples at these two stages, which is expected as the two stages represented PEMs only distinguished by stage S2 being collected after 1 week on pre-maturation medium (a medium without growth regulators). Stages S3–S6, which represented embryo maturation, grouped together and formed a cluster apart from the earlier two stages. Stage S3 was the stage where the PEMs were expected to first respond to the maturation treatment, which included a higher osmoticum and ABA, by stopping multiplication and initiating maturation; stages S4 and S5 represent a gradual maturation process until formation of a fully mature embryo. During desiccation of the mature somatic embryos, between stages S5 and S6, many processes related to dormancy in the seed are expected to take place. The third observed cluster consisted of samples from stages S7 and S8, collected after the start of germination (stage S7) and further development to plantlets that were ready to be planted into soil (stage S8). As such, the PCA revealed two major transitions in the composition of the transcriptome, the first representing a shift from proliferation to maturation (stage S2 to S3) and the second representing transition from desiccated Sem to germination (stage S6 to S7).

A total of 40 281 genes were detected as expressed in SE samples of which 69.5% (27 995 genes) were DE at some point during the experiment, with the number of DEGs varying substantially between consecutive developmental stages (Figure 2c).

Taken together, the PCA and DE analyses revealed extensive modulation of the transcriptome during SE between stages S2 and S3 and stages S5 and S6 with a later, and expected, additional change resulting from germination and the establishment of photosynthesis. As expected, at the later transition (S6–S7) there were many significantly enriched GO categories associated with photosynthesis (Table S2). At the earlier transition we found groups of over-represented genes, but the role of these was not clearly identifiable due to the general nature of the enriched GO terms (the five most enriched terms were carbohydrate metabolic process, cellular macromolecule metabolic process, external encapsulating structure organisation, cell wall organisation or biogenesis and generation of precursor metabolites and energy). This is a common challenge faced when working with non-model organisms lacking high-quality functional annotations, in particular in the current case where extensive divergence from Arabidopsis makes ortholog-based annotation problematic.

Comparison of gene expression during ZE and SE

One of the aims of the study was to identify differences in the regulation of gene expression changes during SE and ZE development. To perform this analysis, we performed stage-wise differential expression tests separately in each

sampled developmental series and then contrasted the pattern of expression change between the two developmental series. We did not make any direct stage comparisons between sampled stages of the SE and ZE developmental series as these were not equivalently staged. As such, even small differences in developmental timing at the sampled stages would negate making like-for-like developmental stage comparisons between SE and ZE stages, which would result in erroneous biological interpretations. Our analysis, therefore, focused on comparison of the pattern of stage-wise changes in gene expression during development between SE and ZE.

There were similar numbers of genes detected as expressed in total and at each sampled stage in both datasets. As such, there was no clear indication of a general activation or transcriptome-wide mis-regulation of expression in SE compared to ZE. While the number of genes detected as expressed was similar in the two sample sets, with a relatively constant number of genes detected as expressed at all sampled stages (Figure S1), there was a clear difference in the number of DEGs, with more extensive DE between consecutive stages in SE (27 995 non-redundant DEGs) than in ZE (14 686 non-redundant DEGs).

The intersection of sets of genes DE between any consecutive time points in each tissue (Figure 3a; Figure S2) revealed DEGs unique to each of the examined tissues and that Sem and Zem shared more DEGs than did FG with either type of embryo. Uniquely DEGs in SE, regardless of when that DE occurred, were examined. These 16 258 DEGs were notably enriched (Table S3) for genes involved in stress related responses including GO:0006952—defence response (padj = 0.0241), GO:0050896—response to stimuli (padj = 0.000161), GO:0055114—oxidation–reduction (padj = 3.36E-06), GO:0016491—oxidoreductase activity (padj = 0.000259), GO:0016209—antioxidant activity (padj = 0.00105), GO:0072593—reactive oxygen species metabolic processes (padj = 0.000441). The importance of stress during SE is well established (Fehér, 2015). Furthermore, DEGs reflecting the in vitro environment of exposure to plant growth regulators; GO:0008144—drug binding (padj = 3.81E-05), GO:0009636—response to toxic substances (padj = 0.0224), GO:0010817—regulation of hormone level (padj = 0.0254) were also present. While these genes that were uniquely DE in SE were not classified as DE in the ZE samples, examination of their expression suggested most were expressed in ZE with variable expression in different stages and tissues (Figure S3). The extent of overlap of DEGs at the most marked stage transitions in each sample set was examined. At the first major transcriptome transition, 2611 DEGs were in common to the S2–S3 and Z3–Z4 Zem stage comparisons, representing ~40% of DEGs at Z3–Z4 Zem, and 753 (~48%) between S2–S3 and Z3–Z4 FG. Similarly, at the second major transcriptome transition,

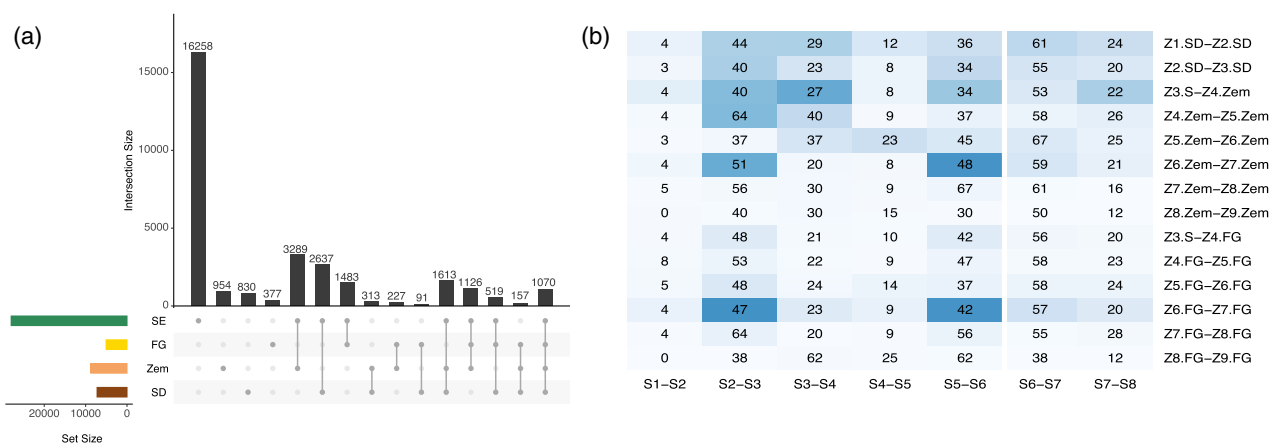


Figure 3. (a) An UpSet plot to show the intersection of differentially expressed genes from all sample types at all stage transitions. Only intersections with non-zero values are shown. SE, somatic embryo; Zem, zygotic embryo; FG, female gametophyte; SD, seed.

(b) Size of intersection of differentially expressed genes at each stage transition in the zygotic and somatic embryogenesis datasets. The values indicate the percentage of differentially expressed genes detected between consecutive stages of zygotic embryogenesis that were also differentially expressed in comparison with consecutive stages of somatic embryogenesis. To aid visual exploration in identifying points of high similarity, cells are shaded based on $\log_{10}(P)$ value from white (not significant) to dark blue (highly significant) to indicate the probability of seeing the intersect only by chance due to the size of the groups in comparison. FG, female gametophyte; SD, seed; SE, somatic embryogenesis; S1-S8, sampled stages of somatic embryogenesis; Z1-Z9, sampled stages of zygotic embryogenesis; Zem, zygotic embryo.

represented by S5-S6 and Z6-Z7, there were 1151 and 1071 DEGs in common between S5-S6 and Z6-Z7 Zem and Z6-Z7 FG, respectively, representing ~50% of DEGs in Z6-Z7 Zem and ~30% in Z6-Z7 FG. As such, DEGs from both ZE tissue types were DE in the SE samples despite the lack of FG tissue in SE.

There were 2949 DEGs unique to ZE regardless of when the DE occurred. These DEGs were enriched for the GO terms (Table S3) gene expression (GO:0009058, $\text{padj} = 3.47e-2$) and biosynthetic process (GO:0010467, $\text{padj} = 3.72e-2$) and represented in all three tissue types (the early stage developing seed (SD), FG, Zem); 377 DEGs were unique to SD, 954 DEGs to Zem, 830 DEGs to FG and 788 in more than one tissue type (most often shared between the SD and Zem) (Figure 3a). DEGs unique to ZE could be important for explaining differences in SE development and could indicate which processes are crucial to achieve better performance of the SE protocol.

The percentage of DEGs detected in ZE that were also differentially expressed in SE and the probability of observing such intersection by chance indicated that the transition from S5 to S6, where desiccation occurred, was most similar to the transition from Z6 to Z7 in ZE (Figure 3b). In ZE, these genes were distinctly DE between Z6 and Z7 and generally remained highly expressed in subsequent sampling points. Examining the set of all Zem DEGs that were also DE in SE (Figure S4a) or that were not also DE in SE (Figure S4b) revealed examples of highly contrasting expression profiles between the two sets of samples. A number of DEGs from Zem were up-regulated across many consecutive stages but were up-regulated at only a single

transition in SE (e.g. the second cluster from the top as indicated in Figure S4a, comprising genes that were DE at Z6-Z7). As such, there were clear contrasts in the regulation of DEGs during desiccation between the two developmental processes. Genes specifically DE in desiccated Zem and not during SE included ribosomal elements known to be associated with the storage and regulation of mRNA translation during seed dormancy (Bai et al., 2020), genes associated with mitochondrial activity related to the TCA cycle important during germination (Zhang & Fernie, 2023) and expression of genes related to the onset of germination as a morphogenic event involving cellular processes that require utilisation of lipids, carbohydrates and starch (results files are in the folder 'DiffExp SE vs. ZE' at the FigShare resource). Previous analysis of quantitative data on water content, carbohydrates and total protein collected from the later stages of Zem development and Sem development in *Pinus pinaster* also indicated that the most advanced mature stage that the somatic embryos could reach in culture corresponded to the fresh, non-desiccated stage of Zem development, suggesting that the Sem maturation does not get fully completed (Morel et al., 2014). The addition of sugar alcohols at a late stage of Sem maturation has previously been shown to increase the quality of Sem, possibly by decreasing the water content in the embryo thus mimicking the desiccation process in the seed (Morel et al., 2014).

As TFs play essential regulatory roles and can have major impacts on transcriptional programmes, the sets of DE TFs in the different sample types were examined (Figure 4a). Among the set of TF families containing

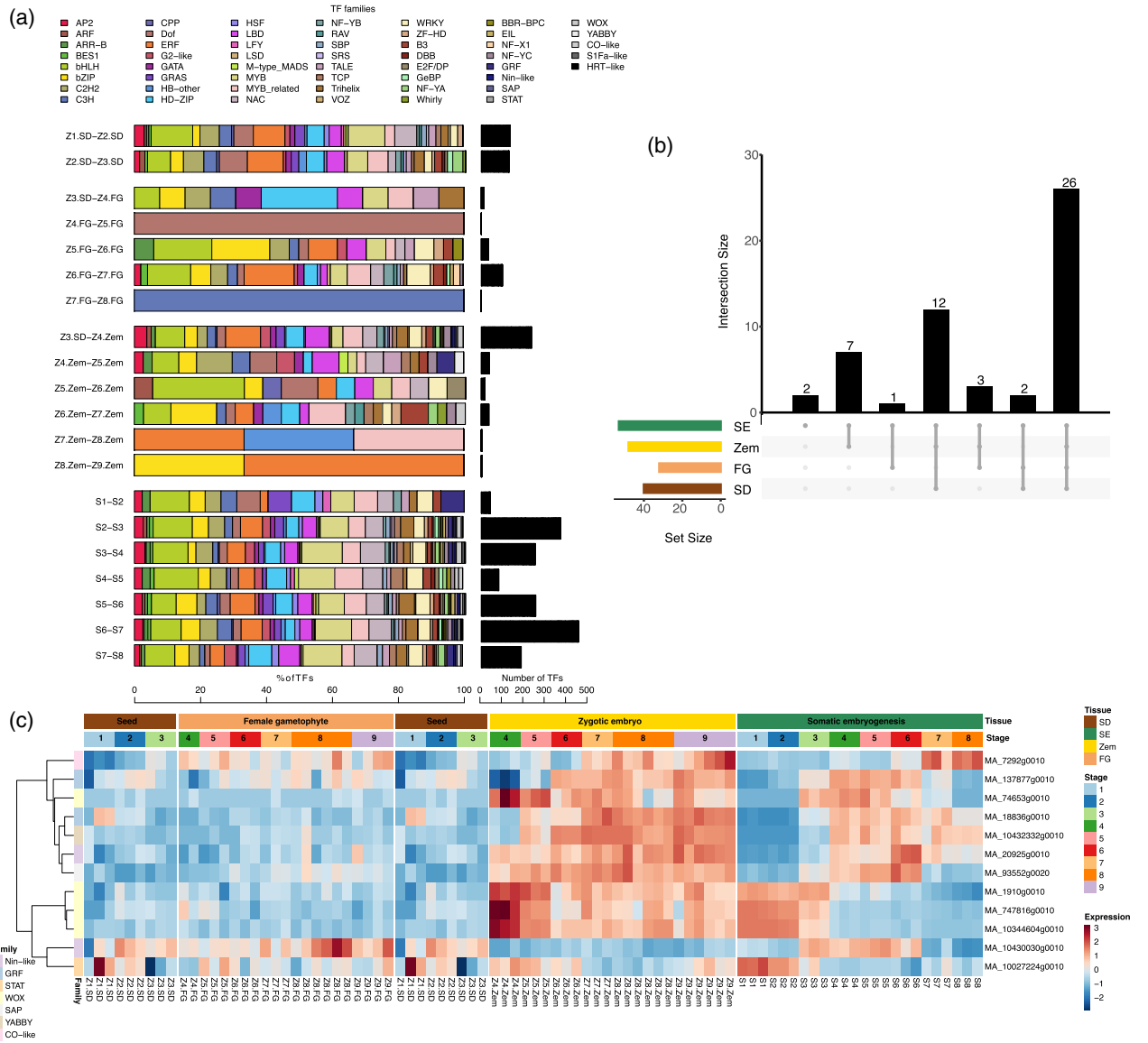


Figure 4. Differential expression of transcription factor genes during development of seed tissue, zygotic embryos, female gametophytes and somatic embryos. (a) The percentage representation of differentially expressed transcription factor genes in transcription factor families (left panel, coloured bars coloured according to family as indicated in the key above the bars) and corresponding total number of differentially expressed transcription factor genes (right panel, black bars) at all sample transitions. (b) An UpSet plot of the intersection of transcription factor families with differentially expressed genes in any of the sampled stages from all sample types. Only intersections with non-zero values are shown. SE, somatic embryo; Zem, zygotic embryo; FG, female gametophyte; SD, seed. (c) A heatmap representation of the 12 transcription factor genes that were uniquely differentially expressed in zygotic embryo. Coloured bars above the heatmap represent sample type and sample stage, and to the left transcription factor family, as indicated in the colour legend to the right of the heatmap. The data is row-scaled (z-scores). FG, female gametophyte; SD, seed; SE, somatic embryogenesis; S1–S8, sampled stages of somatic embryogenesis; Z1–Z9, sampled stages of zygotic embryogenesis; Zem, zygotic embryo.

differentially expressed TFs in any of the samples, all TF families DE in the ZE samples (of all types) were DE in SE and half of these were commonly DE in SD, Zem and FG (Figure 4b). Considering only ZE samples, there were more families represented in Zem than FG and genes from seven TF families were DE only in Zem samples. Families common to SD and Zem may actually be Zem specific, but it is not possible to determine which tissue(s) of the SD

samples those genes were expressed in. Within the ZE dataset, a single TF family, Nuclear Transcription Factor, X-Box Binding 1 (NF-X1), was uniquely present in FG. *AtNFXL1* is component of a regulatory mechanism that has been shown to improve physiological status and to support growth and survival under stress (Müssig, Schröder, Usadel, & Lisso, 2010). There were two families (S1Fa-like and Hordeum Repressor Transcription HRT-like)

uniquely DE in SE. Overexpression of *S1Fa* genes has been linked to improved drought tolerance in *Populus trichocarpa* (Zhao, Niu, Dong, Jia, & Wang, 2021). Relatively little is known about *HRT-like* genes but work in barley indicated that HRT can repress the α -amylase promoters (Raventós et al., 1998) and analysis of promoter binding motifs in *Eragrostis tef* indicated that HRT may be involved in the control of hormones (ABA, Gibberellic acid), sugars (zein, α -amylase) and abiotic stress (Mulat & Sinha, 2020). In general, the representation of TF families at different stage transitions was far more even in SE than in ZE and there were more DE TFs at more stage transitions in SE than in ZE (Figure 4a). In the overall comparison across development, the pattern of expression for different TF families appeared more uniformly across SE than ZE with almost all families showing DE at all SE developmental transitions. In contrast, in ZE some transitions had only one (Z4.FG – Z5.FG and Z7.FG – Z8.FG), two (Z8.Zem – Z9.Zem) or three (Z7.Zem – Z8.Zem) TF families represented. This reflected the low numbers of DEGs at these stages (Figure 1a). It is also notable that in ZE, the greatest number of DEGs was identified at the transitions from proembryogeny to early embryogeny (between Z1–Z2, Z2–Z3 and Z3–Z4, respectively), and later early to late embryogeny (Z6–Z7), which was also reflected in the extent of DE of TFs. In SE, there was no corresponding peak in DE at the transition from proembryogeny to early embryogeny (Figure 1c and Figure 2c), indicating that prior developmental processes required for subsequent processes may not be correctly activated by the culture conditions present at the earlier stage. The importance of abiotic factors early in embryo development on the later stages of development was also recently shown for *Abies nordmanniana* (Nielsen, Hansen, Hansen, Johansen, & Egertsdotter, 2022), suggesting that efforts to optimise SE culture conditions should consider the interrelation between developmental and culture conditions at steps of the SE protocol. While there were no TF families uniquely DE in ZE, there were 12 TF genes that were uniquely DE in Zem (Figure 4c). While these 12 genes were not classified as significantly DE in SE they did have clear evidence of expression level variation during SE within specific stages, indicative of expression regulation. The expression profiles of these genes were distinctly different during SE than ZE and, as such, these represent interesting candidate TFs that could explain differences in the regulation of gene expression during the two developmental processes. Taken together, this suggests that the regulation of many TFs differs during the two developmental processes.

Genes of known importance for embryogenesis

A set of genes known to be important in embryogenesis in the processes of somatic embryo initiation, desiccation and nitrogen utilisation was identified from the literature

(Table S4) and their expression patterns (Figure S5) and differential expression between developmental stages (Table S4) were examined within the current ZE and SE data. In addition, specific TFs of known importance during embryo development and with contrasting expression between the two developmental processes were further considered.

The only TF family active in FG at the transition Z4.FG – Z5.FG represented the TCP (*Zea mays* TEOSINTE BRANCHED1 (TB1), *Antirrhinum majus* CYCLOIDEA (CYC), *Oryza sativa* PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR1/2) TF family that have been implicated as regulators during plant growth, development and abiotic stress (Nie et al., 2022). TCP TFs affect growth directly via the cell cycle and indirectly via influencing plant hormonal signalling and the circadian clock (Danisman, 2016). In Arabidopsis AtTCP14 has been shown to specifically regulate embryonic growth development during germination. It is expressed primarily in the embryo with limited expression in the endosperm (Tatematsu, Nakabayashi, Kamiya, & Nambara, 2008). This contrasts with the observed expression pattern in the current data, where the highest sequence similarity homologue was the only DE TF at the Z4.FG – Z5.FG transition. Forty-three TCP genes were identified from *Pinus tabuliformis* with several being co-expressed with clock genes and during seasonal oscillations (Nie et al., 2022). These observed expression patterns could reflect function in FG during dormancy and before onset of germination.

The Ethylene Response Factor (ERF) family was one of two (Z8.Zem – Z9.Zem) or three (Z7.Zem – Z8.Zem) TF families that increasingly dominated the DE set of TFs towards the end of the Zem developmental process (Figure 4a). Functional enrichment results suggested that AP2/ERFs were transcribed in the FG (Results file available at FigShare). There was generally little DE of TFs at these stages, which is perhaps not surprising given that the seeds were dormant. Desiccation is an essential part of the later stages of seed development, preparing for dormancy until germination is triggered. Ethylene is known to release dormancy in a number of plant species (Corbineau, Xia, Bailly, & El-Maarouf-Bouteau, 2014) and several ERFs also bind to dehydration-responsive elements (Müller & Munné-Bosch, 2015). During the corresponding stages of SE, where stage S6 represented somatic embryos desiccated for 3 weeks and S7 embryos after 1 week in germination, there was no substantial representation of DE ERFs. This may be interpreted as an indication of another step of the SE process that deviates from Zem development and suggests further studies are warranted into expanding culture condition optimisations to include the gas phase of the culture containers during SE germination. Maturation success rates of Sem have been associated with production of ethylene in a number of conifer species, e.g. *Pinus*

sylvestris (Lu, Vahala, & Pappinen, 2011), *Araucaria angustifolia* (Jo, Santos, Bueno, Barbosa, & Floh, 2014), *Picea glauca* (El Meskaoui, Desjardins, & Tremblay, 2000) and *Picea mariana* (El Meskaoui & Tremblay, 2001). However, further studies into the effect of ethylene during germination are currently lacking. Overall, the less distinct regulation of the *ERF* family during SE may reflect missing effects of ethylene signalling during development due to Sem development occurring inside a culture container surrounded by a relatively large volume of air compared to Zem development, which is contained within the seed coat.

It is notable that the WUSCHEL-related homeobox (*WOX*) family, being the best-studied TF family in conifer embryogenesis based on its function during the formation of the embryogenic shoot meristem in Arabidopsis and other plants (Hassani, Trontin, Raschke, Zoglauer, & Rupps, 2022), was scarcely represented in the set of TF families containing DEGs in both ZE and SE samples (Figure 4a). However, when analysing the expression patterns of the twelve TF genes that were DE specifically in Zem and in no other tissue, the most significant changes in expression occurred for three *WOX* genes expressed in Zem during the transition from early to late embryogeny. In Sem, the expression was not enhanced during the corresponding stages of S3 and S4 (pre-maturation and early maturation) (Figure 4c). In *Pinus pinaster*, *WOX2* was mostly expressed during proliferation of Sem (Hassani et al., 2022) and low during maturation and germination which also agrees with the three *WOX* genes in the present study.

A notable discrepancy in the expression pattern of the DE TFs in ZE and SE was observed for three putative baby boom (*BBM*) genes that in other plant species have been demonstrated to have a significant effect on induction of SE (MA_86195g0010, MA_98095g0010, MA_121578g0010 in Figure S5 and Table S4). In *Larix decidua*, a conifer species typically showing high success rates for SE plant formation, expression of *LdBBM* increased from proliferation to fully mature embryos both during ZE and SE (Rupps, Raschke, Rümmler, Linke, & Zoglauer, 2016). In the present study, Zem showed a slight increase in expression over maturation similar to the observed patterns in *Larix* Zem. In contrast, expression in Sem did not follow this pattern (Figure S5), as was similarly observed in *Larix* Sem.

The signal transduction pathway involved in controlling ABA levels during embryo development involves *Viviparous-1* (*VP1*) and *VP2* (Kermode, 2005). The expression of *VP1* and *VP2* increased from early embryogeny until mid-late embryogeny similarly during Zem and Sem development (S4; MA_66505g0010 and MA_66505g0020). This is in agreement with previous findings for both ZE where the endogenous levels of ABA during Zem development in *Picea glauca* were shown to increase from

relatively low levels at the onset of embryo development to peak mid-stage embryo maturation (Kong, Attree, & Fowke, 1997) and the expression levels of *VP1* during *P. abies* SE (Uddenberg et al., 2011).

SE is increasingly utilised for propagation of commercially important conifers; mostly spruces in the northern hemisphere (*Picea abies*, *Picea glauca*) and pine in the southern (*Pinus radiata*). The method has potential for cost-effective, large-scale propagation through automation of mainly the *in vitro* steps (*ex vitro* steps are already automated); however, the process from initiation of PEMs to the plant growing in the nursery is still far from being as efficient as the zygotic seed developmental process. To improve the survival rates and quality of embryos through development, efforts have been made to understand the natural conditions for development of the Zem by analysing the composition of the FG and then to duplicate the seed nutritional, hormonal and environmental conditions *in vitro* (reviewed by Pullman & Bucalo, 2014). The data presented here cover more developmental stages than any previous study, providing a powerful resource for elucidating new knowledge of gene expression events during the very first stages of seed development, key events such as desiccation and the differences between ZE and SE. Exploration of the data suggests more extensive differences in DEGs between the SE process and FG than between Zem and FG, with these contrasting examples of developmental regulation representing candidate genes that may link to targets for protocol improvement such as modifications to culture media and controlling gas composition during the gas phase.

Conclusion

The comparison of ZE and SE presented here highlights that while there is extensive commonality in transcriptional regulation during ZE and SE, there are also numerous differences in the sets of genes classified as significantly differentially regulated at stage transitions during ZE and SE development, in addition to differences in the timing and duration of expression changes. Genes with contrasting regulation include many TFs, which are important targets for understanding the regulatory mechanisms underlying differences in ZE and SE. Importantly, genes with contrasting regulation between ZE and SE can help identify associated biochemical, hormonal and physiological influences and these can direct future optimisation of SE protocols. For example, there were obvious differences in the timing and duration of changes in gene expression associated with desiccation. Future exploration of the known factors determining regulation of these genes could identify specific targets for optimisation such as changes in the timing of application or composition of hormones in culture media or changes in environmental conditions to induce physiological responses. Such differences can help

explain differences in the *in vitro* and *in planta* embryo developmental processes, which helps provide greater understanding of which regulatory events are important in the natural process of Zem in conifers. Exploration of this data resource identified essential factors in the natural seed maturation process, as summarised by the following findings: (1) further evidence to support the hypothesis that the desiccation process in Sem is incomplete compared to this process during maturation in Zem as indicated by the differences in DEG between the corresponding developmental stages before and after desiccation in ZE and SE (Fig. S4a,b). These findings point to the importance of the interrelation of abiotic conditions during embryo development for the completion of the developmental process and provide information useful for directing optimisations of SE protocols: (2) the expression of BBM may be used as an indicator of successful Sem and, as such, may represent a key regulatory mechanism in embryogenesis; (3) specific transcription factors with contrasting expression were identified between Zem and Sem development, which represent candidates for future characterisation studies to more fully determine their role in regulating conifer embryogenesis. This data resource, which is explorable within expression visualisation tools at the [PlantGenIE.org](https://plantgenie.org) resource, provides a reference gene expression atlas for conifer embryo development, enabling exploration of gymnosperm embryo development, understanding of which is currently lagging behind that of angiosperm species. Furthermore, the resource enables community utilisation to explore discrepancies between Sem and Zem development to improve understanding of observed discrepancies between somatic and zygotic embryogenesis. Such knowledge should facilitate development of improved methods to obtain high yields of SE plants for research and reforestation purposes.

EXPERIMENTAL PROCEDURES

Plant material

Material for the ZE samples was obtained from cones collected from one replicate clonal copy of Norway spruce (*Picea abies* (L.) Karst) genotype Z4006 (as used in Nystedt et al., 2013) that was control pollinated with a known mixed pollen source at the Swedish Forest Research Institute (Skogforsk) in Sävar, Sweden. Samples were collected starting from 2 weeks after pollination on 4 June 2017 then every 2 weeks until full cone maturity at 18 weeks after pollination on 11 October 2017 (Figure S6). The cones were stored at +4°C until sample extraction, which occurred within 1 week from collection.

Embryogenic cultures were initiated and captured in 2014 from seeds from Norway spruce genotype Z4006. Cultures from cell line K14-03 were proliferated for embryo maturation and germination by standard methods (Von Arnold & Clapham, 2008). Briefly, proliferation was stimulated using liquid half-strength LP medium supplemented with auxin (2,4-D; 2.21 mg l⁻¹), N6-benzyladenine (BA; 1 mg l⁻¹) and sucrose (10 g l⁻¹). After

4–6 weeks cultures were transferred to liquid pre-maturation medium devoid of 2,4-D and BA for 1 week, then to liquid maturation medium supplemented with ABA (16 mg l⁻¹) and sucrose (30 g l⁻¹) for 6 weeks. Cultures were kept in darkness at 20°C. Mature embryos were dried for 3 weeks under high humidity in darkness and subsequently germinated in 15 cm Petri dishes on solid medium containing minerals and vitamins supplemented with sucrose (30 g l⁻¹) and casein hydrolysate (0.5 g l⁻¹), solidified with 3.5% Gelrite. The pH was adjusted to 5.8 ± 0.1 prior to autoclaving. Germinants were grown for 3 weeks in continuous red light (wavelength: 660 nm; TL-D 18 W/15, Philips, Stockholm, Sweden) at 5 μmol m⁻² s⁻¹ at 20°C, then moved under continuous white fluorescent tubes (Fluora L 18 W/77, Osram, Johannehov, Sweden) at 100–150 μmol m⁻² s⁻¹ at 20°C (Kvaalen & Appelgren, 1999). Germinants with actively growing shoot and root were planted in peat moss, acclimatised and plant growth established under the same light conditions.

Sampling

From the three earliest cone-collection times, the whole female gametophyte (FG) with the seed coat removed was sampled as no zygotic embryo (Zem) could be isolated. The two earliest collection times (Z1-2) had distinctly smaller FG compared to collection three (Z3). The transition from proembryogeny to early embryogeny (when an early stage embryo composed of an embryonic region and suspensor cells could first be observed) occurred in the Z2-Z3 transition. No embryo structures were observed in the Z2 samples, but these were observed in the Z3 samples. Additionally, the appearance of the FG was translucent in Z1 and Z2 but turned opaque in Z3, which is also indicative of the transition. From the fourth collection time on July 31 (Z4) until ninth collection time (Z9), samples were separated into zygotic embryos (Z4.Zem) and FG (Z4.FG). A total of 1618 seeds were collected during the duration of the sampling period whereof 696 whole FGs were collected from Z1-3 and 922 FG and 838 Zem from stages Z3-Z10. A minimum of 100 samples were collected from each collection date. The samples were mixed and split into separate Eppendorf tubes with approximately 25 samples of zygotic embryos with FG for stages Z1-2, and subsequently zygotic embryos or FGs, snap-frozen in liquid nitrogen and stored at -80°C until further processed.

Due to insect damage of cones, limited numbers of samples were collected on Oct 11 (Zem8/9and FG8/9). In addition to the sample collection from the controlled pollinated tree, dried seeds collected in 2016 from a related tree were sampled and used as additional Zem and FG replicates at stages 8 and 9 (see associated metadata at ENA).

Samples were collected from eight different stages of somatic embryo development from pro-embryogenic masses to SE germinants established in compost as follows: S1 corresponds to pro-embryogenic masses (PEMs) collected from proliferation medium; S2 was PEMs collected after 1 week on pre-maturation medium; S3, S4 and S5 were samples from the SE culture with maturing embryos collected after three, 6 and 8 weeks on maturation medium; S6 was mature isolated embryos collected after 3 weeks of desiccation; S7 was mature, desiccated embryos germinated for 1 week to show early root formation, and S8 correspond to germinants ready to be planted and be established in compost (Figure 1). Samples were collected in 2014. For each sampling point, 50 mg of somatic embryo tissue (PEMs, mature embryos, germinants or plants) were collected from each of three different parallel petri plates for every developmental stage. Samples from each developmental stage were mixed and split into

four Eppendorf tubes, snap-frozen in liquid nitrogen and stored at -80°C until further processed.

RNA isolation and sequencing

Total RNA was extracted from the ground samples with an RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's guidelines. Nucleotide quantity and purity were assessed using a NanoDrop 2000C Spectrophotometer (Thermo Scientific, Waltham, MA, USA) and quality of total RNA using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). RNA libraries for sequencing were prepared using TruSeq Stranded mRNA sample prep kit with 96 dual indexes (Illumina, San Diego, CA, USA) according to the manufacturer's instructions with the following changes: RNA was fragmented 5' and the protocols were automated using an Agilent NGS workstation (Agilent Technologies) with purification steps described by Lundin, Stranneheim, Pettersson, Klevebring, and Lundeberg (2010) and Borgström, Lundin, and Lundeberg (2011).

Transcriptome sequencing and data pre-processing

RNA isolated from Zem and FG collected at eight-time points during seed development (with the ZE and FG processed together for the first three collection times) and eight developmental stages during somatic embryogenesis (SE) was sequenced on Illumina HiSeq 2500 at SciLifeLab using 2×126 bp paired-end reads to an average read number of 24.1 ± 2.1 M reads per sample. Three biological replicates were sequenced for each time point/developmental stage.

The quality of the raw sequence data was assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), v0.11.4. Sequence reads originating from ribosomal RNAs (rRNA) were identified and removed using SortMeRNA (v2.1; Kopylova, Noé, & Touzet, 2012; settings `--log --paired_in --fastx--sam --num_alignments 1`) using the rRNA sequences provided with SortMeRNA (rfam-5 s-database-id98.fasta, rfam-5.8 s-database-id98.fasta, silva-arc-16 s-database-id95.fasta, silva-bac-16 s-database-id85.fasta, silva-euk-18 s-database-id95.fasta, silva-arc-23 s-database-id98.fasta, silva-bac-23 s-database-id98.fasta and silva-euk-28 s-database-id98.fasta). Data were then filtered to remove adapters and trimmed for quality using Trimmomatic (v0.36; Bolger *et al.* 2014; settings `TruSeq3-PE-2.fa:2:30:10 SLIDINGWINDOW:5:20 MINLEN:50`). After both filtering steps, FastQC was run again to ensure that no technical artefacts were introduced. Read counts were obtained using Salmon (v. 0.11.2; Patro, Duggal, Love, Irizarry, & Kingsford, 2017) using the *P. abies* v1.0 (Nystedt *et al.*, 2013) transcript sequences as a reference (retrieved from the PlantGenIE resource; Sundell *et al.*, 2015).

Differential expression analysis of genes

The salmon abundance values were imported into R (v4.0.0; R Core Team 2021) using the Bioconductor (v3.11; Gentleman *et al.*, 2004) tximport package (v.1.16.1; Sonesson, Love, & Robinson, 2015). For the data quality assessment (QA) and visualisation, the read counts were normalised using a variance stabilising transformation (VST) as implemented in the Bioconductor DESeq2 package (v1.28.1; Love, Huber, & Anders, 2014). The biological relevance of the data—e.g. biological replicates similarity—was assessed by principal component analysis (PCA) and other visualisations (e.g. heatmaps), using custom R scripts, which identified one sample as an outlier, which was therefore removed from further analysis. Statistical analysis of gene differential expression (DE) between consecutive stages of the

experiment was performed in R using the Bioconductor DESeq2 package. FDR adjusted p values (by the Benjamini–Hochberg procedure) were used to assess significance; a common adjusted threshold of 1% was used throughout. All expression results were generated in R, using custom scripts. Formulas used in DESeq2 included stages and DEGs between consecutive stages of the experiment were extracted using the function 'results', provided with option 'filter = rowMedians(counts(dds))'. DEGs considered for further analysis were filtered by fold change $|\log_2\text{FC}| \geq 0.5$ and *P* value adjusted for multiple testings $\text{padj} < 0.01$. The number of genes in the intercept whose *padj* value was different from NA was considered to be the number of all the expressed genes in the experiment. When comparing DEGs found in consecutive stages of ZE and SE the supertest function from the R package SuperExactTest (Wang, Zhao, & Zhang, 2015) was used to calculate the statistical significance of intersection sizes. The background set of genes used for the test was the union of the genes expressed in SE and ZE.

Functional enrichment analysis

Gene Ontology (Ashburner *et al.*, 2000) functional enrichment of differentially expressed genes (DEGs) and clusters of co-expressed genes at $P < 0.05$ was analysed using an in-house implementation of the parent-child test (Grossmann, Bauer, Robinson, & Vingron, 2007) and a Benjamini–Hochberg multiple testing correction. The background set of genes used for the test were all the genes expressed in the experiment or union of the genes expressed in SE and ZE when comparing groups of DEGs from different experiments.

ACCESSION NUMBERS

PRJEB72619.

ACKNOWLEDGEMENTS

Katja Stojković was supported by a grant from the Kempe Foundation (SMK1340). Nathaniel Street and Ulrika Egertsdotter are supported by the Trees for the Future (T4F) project. This work was supported by grants from the Knut and Alice Wallenberg Foundation. The authors acknowledge support from the National Genomics Infrastructure in Genomics Production Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure and the Umeå Plant Science Centre bioinformatics facility for support.

CONFLICT OF INTEREST

N Street is a shareholder in Woodhead AB. Woodheads AB is a shareholder in SweTree Technologies, which has a commercial interest in somatic embryogenesis of Norway spruce.

DATA AVAILABILITY STATEMENT

Raw sequencing data is available from the European Nucleotide Archive (ENA) as accession PRJEB72619.R scripts used to perform bioinformatic analysis are available in the GitHub repository (<https://github.com/Emkago/spruce-embryogenesis-mRNA>). Gene expression data have been integrated at PlantGenIE.org into the available

expression visualisation tools. All results of differential expression test, the normalised gene expression values used to perform the presented analyses and all functional enrichment test results are available at the Science for Life (SciLife) data centre FigShare resource at <https://doi.org/10.17044/scilifelab.25315867>.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Number of genes with any aligned sequencing read and classified as expressed at different expression thresholds (increasing variance stabilising transformation values) in the samples of zygotic embryogenesis (a) and somatic embryogenesis (b).

Figure S2. Intersection of genes differentially expressed between all consecutive stages from each tissue analysed during somatic and zygotic embryogenesis. FMG, female gametophyte.

Figure S3. Expression of genes in zygotic embryogenesis samples that were uniquely differentially expressed during somatic embryogenesis. Expression values for each gene were Z-score normalised across samples. FG, female gametophyte; SD, seed; Z1–Z9 sampling stages from zygotic embryogenesis. Each sample from Z1 to Z3 is represented twice for technical reasons. Zem, zygotic embryo.

Figure S4. (a) Expression of genes differentially expressed in both somatic and zygotic embryogenesis during desiccation of the zygotic embryo (stage Z6–Z7). (b) Expression of genes differentially expressed at transition Z6–Z7 in zygotic embryos but not differentially expressed in somatic embryogenesis. Expression values for each gene were Z-score normalised across samples. FG, female gametophyte; SD, seed; SE, Somatic embryogenesis; Z1–Z9, sampled stages during zygotic embryogenesis; S1–S8, sampled stages during somatic embryogenesis; Zem, zygotic embryo.

Figure S5. Expression of genes known to be important in embryogenesis in the processes of somatic embryo initiation, desiccation and nitrogen utilisation shown in somatic and zygotic embryogenesis samples. Expression values for each gene were Z-score normalised across samples. FG, female gametophyte; SD, seed; SE, Somatic embryogenesis; S1–S8, sampled stages during somatic embryogenesis; Z1–Z9, sampled stages during zygotic embryogenesis; Zem, zygotic embryo. Details of the genes are available in Table S4.

Figure S6. Female gametophytes (FG) and zygotic embryos (Zem) were collected from immature cones from two weeks after pollination (WAP) on 4 June 2017 to fully mature, desiccated cones 18 WAP on Oct 11. Separated samples of Zem and FGs could only be obtained from the fourth collection time (Z4) and later as the embryos were not detectable in the earlier samples. The photos show the variation in development among embryos at each collection time. Seeds from all cones from each time point were collected and the samples therefore contained embryos at slightly different developmental stages since embryo development is not synchronised within each cone and among cones.

Table S1. Functional enrichment of differentially expressed genes during zygotic embryogenesis.

Table S2. Functional enrichment of differentially expressed genes during somatic embryogenesis.

Table S3. Functional enrichment of differentially expressed genes found only in somatic or zygotic embryogenesis.

Table S4. Differential expression in zygotic and somatic embryogenesis stage transitions of genes known to be important in embryogenesis.

REFERENCES

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. et al. (2000) Gene ontology: tool for the unification of biology. *Nature Genetics*, **25**, 25–29. Available from: <https://www.eurekaselect.com/162338/article>
- Bai, B., van der Horst, S., Cordewener, J.H.G., America, T.A.H.P., Hanson, J. & Bentsink, L. (2020) Seed-stored mRNAs that are specifically associated to Monosomes are translationally regulated during germination. *Plant Physiology*, **182**, 378–392. Available from: <https://academic.oup.com/plphys/article/182/1/378-392/6116138>
- Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
- Borgström, E., Lundin, S. & Lundeberg, J. (2011) Large scale library generation for high throughput sequencing M. N. Nitabach, ed. *PLoS One*, **6**, e19119. Available from: <https://doi.org/10.1371/journal.pone.0019119>
- Cairney, J. & Pullman, G.S. (2007) The cellular and molecular biology of conifer embryogenesis. *The New Phytologist*, **176**, 511–536.
- Cairney, J., Xu, N., Mackay, J. & Pullman, J. (2000) Special symposium: in vitro plant recalcitrance transcript profiling: a tool to assess the development of conifer embryos. *In Vitro Cellular and Developmental Biology—Plant*, **36**, 155–162. Available from: <https://doi.org/10.1007/s11627-000-0031-5>
- Corbineau, F., Xia, Q., Bailly, C. & El-Maarouf-Bouteau, H. (2014) Ethylene, a key factor in the regulation of seed dormancy. *Frontiers in Plant Science*, **5**, 1–13. Available from: <https://doi.org/10.3389/fpls.2014.00539/abstract>
- Cui, Y., Zhao, J., Gao, Y., Zhao, R., Zhang, J. & Kong, L. (2021) Efficient multi-site genome editing and plant regeneration via somatic embryogenesis in *Picea glauca*. *Frontiers in Plant Science*, **12**, 1–12. Available from: <https://doi.org/10.3389/fpls.2021.751891/full>
- Danisman, S. (2016) TCP transcription factors at the Interface between environmental challenges and the Plant's growth responses. *Frontiers in Plant Science*, **7**, 1–13. Available from: <https://doi.org/10.3389/fpls.2016.01930/full>
- De Vega-Bartol, J.J., Simões, M., Lorenz, W.W., Rodrigues, A.S., Alba, R., Dean, J.F.D. et al. (2013) Transcriptomic analysis highlights epigenetic and transcriptional regulation during zygotic embryo development of *Pinus pinaster*. *BMC Plant Biology*, **13**, 123.
- Durzan, D. (2012) Second International Conference of 2.09.02 – Somatic embryogenesis and other vegetative propagation technologies: integrating vegetative propagation, biotechnologies and genetic improvement for tree production and sustainable forest management; Brno, Czech Republic; 25–28 June 2012. *Proceedings*. Available from: <https://www.iufro.org/fileadmin/material/publications/proceedings-archive/20902-brno12.pdf>
- Egertsdotter, U. (2018) Plant physiological and genetical aspects of the somatic embryogenesis process in conifers. *Scandinavian Journal of Forest Research*, **34**, 360–369. Available from: <https://doi.org/10.1080/02827581.2018.1441433>
- Egertsdotter, U., Ahmad, I. & Clapham, D. (2019) Automation and scale up of somatic embryogenesis for commercial plant production, with emphasis on conifers. *Frontiers in Plant Science*, **10**, 109. Available from: <https://doi.org/10.3389/fpls.2019.00109/full>
- El Meskaoui, A., Desjardins, Y. & Tremblay, F.M. (2000) Kinetics of ethylene biosynthesis and its effects during maturation of white spruce somatic embryos. *Physiologia Plantarum*, **109**, 333–342. Available from: <https://doi.org/10.1034/j.1399-3054.2000.100315.x>
- El Meskaoui, A. & Tremblay, F.M. (2001) Involvement of ethylene in the maturation of black spruce embryogenic cell lines with different maturation capacities. *Journal of Experimental Botany*, **52**, 761–769. Available from: <https://doi.org/10.1093/jexbot/52.357.761>
- Elbl, P., Lira, B.S., Andrade, S.C.S., Jo, L., Santos, A.L.W.D., Coutinho, L.L. et al. (2015) Comparative transcriptome analysis of early somatic embryo formation and seed development in Brazilian pine, *Araucaria angustifolia* (Bertol.) Kuntze. *Plant Cell, Tissue and Organ Culture*, **120**, 903–915.
- Fehér, A. (2015) Somatic embryogenesis — stress-induced remodeling of plant cell fate. *Biochimica et Biophysica Acta, Gene Regulatory*

- Mechanisms*, 1849, 385–402. Available from: <https://doi.org/10.1016/j.bbagr.2014.07.005>
- Filonova, L.H., Bozhkov, P.V. & Von Arnold, S. (2000) Developmental pathway of somatic embryogenesis in *Picea abies* as revealed by time-lapse tracking. *Journal of Experimental Botany*, 51, 249–264. Available from: <https://doi.org/10.1093/jexbot/51.343.249>
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5, R80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15461798>
- Grossmann, S., Bauer, S., Robinson, P.N. & Vingron, M. (2007) Improved detection of overrepresentation of gene-ontology annotations with parent-child analysis. *Bioinformatics*, 23, 3024–3031. Available from: <https://academic.oup.com/bioinformatics/article/23/22/3024/208216>
- Hakman, I. (1993) Embryology in Norway spruce (*Picea abies*). An analysis of the composition of seed storage proteins and deposition of storage reserves during seed development and somatic embryogenesis. *Physiologia Plantarum*, 87, 148–159. Available from: <https://doi.org/10.1111/j.1399-3054.1993.tb00137.x>
- Hassani, S.B., Trontin, J.-F., Raschke, J., Zoglauer, K. & Rupps, A. (2022) Constitutive overexpression of a conifer WOX2 homolog affects somatic embryo development in *Pinus pinaster* and promotes somatic embryogenesis and organogenesis in Arabidopsis seedlings. *Frontiers in Plant Science*, 13, 1–19. Available from: <https://doi.org/10.3389/fpls.2022.838421/full>
- He, X. & Kermodé, A.R. (2003) Nuclease activities and DNA fragmentation during programmed cell death of megagametophyte cells of white spruce (*Picea glauca*) seeds. *Plant Molecular Biology*, 51, 509–521. Available from: <https://doi.org/10.1023/A:1022319821591>
- Jo, L., Santos, A.L.W.D., Bueno, C.A., Barbosa, H.R. & Floh, E.I.S. (2014) Proteomic analysis and polyamines, ethylene and reactive oxygen species levels of *Araucaria angustifolia* (Brazilian pine) embryogenic cultures with different embryogenic potential. *Tree Physiology*, 34, 94–104. Available from: <https://doi.org/10.1093/treephys/tpt102>
- Kermodé, A.R. (2005) Role of Abscisic acid in seed dormancy. *Journal of Plant Growth Regulation*, 24, 319–344. Available from: <https://doi.org/10.1007/s00344-005-0110-2>
- Kong, L., Attree, S.M. & Fowke, L.C. (1997) Changes of endogenous hormone levels in developing seeds, zygotic embryos and megagametophytes in *Picea glauca*. *Physiologia Plantarum*, 101, 23–30. Available from: <https://doi.org/10.1111/j.1399-3054.1997.tb01815.x>
- Kopylova, E., Noé, L. & Touzet, H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28, 3211–3217. Available from: <https://academic.oup.com/bioinformatics/article/28/24/3211/246053>
- Kvaalen, H. & Appelgren, M. (1999) Light quality influences germination, root growth and hypocotyl elongation in somatic embryos but not in seedlings of Norway spruce. *Vitr. Cell. Dev. Biol.*, 35, 437–441.
- De La Torre, A.R., Piot, A., Liu, B., Wilhite, B., Weiss, M. & Porth, I. (2019) Functional and morphological evolution in gymnosperms: a portrait of implicated gene families. *Evolutionary Applications*, 13, 210–227. Available from: <https://doi.org/10.1111/eva.12839>
- Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550. Available from: <https://doi.org/10.1186/s13059-014-0550-8>
- Lu, J., Vahala, J. & Pappinen, A. (2011) Involvement of ethylene in somatic embryogenesis in scots pine (*Pinus sylvestris* L.). *Plant Cell, Tissue and Organ Culture*, 107, 25–33. Available from: <https://doi.org/10.1007/s11240-011-9952-4>
- Lundin, S., Stranneheim, H., Pettersson, E., Klevebring, D. & Lundeberg, J. (2010) Increased throughput by parallelization of library preparation for massive sequencing. *PLoS One*, 5, e10029.
- Mackay, J., Dean, J.F.D., Plomion, C., Peterson, D.G., Cánovas, F.M., Pavy, N. *et al.* (2012) Towards decoding the conifer giga-genome. *Plant Molecular Biology*, 80, 555–569. Available from: <https://doi.org/10.1007/s1103-012-9961-7>
- Mamun, N.H.A., Aidun, C.K. & Egertsdotter, U. (2018) Improved and synchronized maturation of Norway spruce (*Picea abies* (L.) H.Karst.) somatic embryos in temporary immersion bioreactors. *In Vitro Cellular and Developmental Biology—Plant*, 54, 612–620. Available from: <https://doi.org/10.1007/s11627-018-9911-4>
- Merino, I., Abrahamsson, M., Sterck, L., Craven-Bartle, B., Canovas, F. & Von Arnold, S. (2016) Transcript profiling for early stages during embryo development in scots pine. *BMC Plant Biology*, 16, 255. Available from: <https://doi.org/10.1186/s12870-016-0939-5>
- Morel, A., Trontin, J.-F., Corbineau, F., Lomenech, A.M., Beaufour, M., Raymond, I. *et al.* (2014) Cotyledonary somatic embryos of *Pinus pinaster* Ait. Most closely resemble fresh, maturing cotyledonary zygotic embryos: biological, carbohydrate and proteomic analyses. *Planta*, 240, 1075–1095. Available from: <https://doi.org/10.1007/s00425-014-2125-z>
- Mulat, M.W. & Sinha, V.B. (2020) First report for availability of HRT-like genes in Eragrostis tef and in silico analysis for elucidating their potential functions. *Plant Gene*, 23, 100230. Available from: <https://doi.org/10.1016/j.plgene.2020.100230>
- Müller, M. & Munné-Bosch, S. (2015) Ethylene response factors: a key regulatory hub in hormone and stress signaling. *Plant Physiology*, 169, 32–41. Available from: <https://academic.oup.com/plphys/article/169/1/32-41/6113870>
- Müssig, C., Schröder, F., Usadel, B. & Lisso, J. (2010) Structure and putative function of NFX1-like proteins in plants. *Plant Biology*, 12, 381–394. Available from: <https://doi.org/10.1111/j.1438-8677.2009.00303.x>
- Nie, Y., Han, F., Ma, J., Chen, X., Song, Y., Niu, S.-H. *et al.* (2022) Genome-wide TCP transcription factors analysis provides insight into their new functions in seasonal and diurnal growth rhythm in *Pinus tabulaeformis*. *BMC Plant Biology*, 22, 167. Available from: <https://doi.org/10.1186/s12870-022-03554-4>
- Nielsen, U.B., Hansen, C.B., Hansen, U., Johansen, V.K. & Egertsdotter, U. (2022) Accumulated effects of factors determining plant development from somatic embryos of *Abies nordmanniana* and *Abies bornmuelleriana*. *Frontiers in Plant Science*, 13, 1–19. Available from: <https://doi.org/10.3389/fpls.2022.989484/full>
- Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G. *et al.* (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497, 579–584. Available from: <https://doi.org/10.1038/nature12211>
- Owens, J.N. & Molder, M. (1979) Bud development in *Larix occidentalis*. II. Cone differentiation and early development. *Canadian Journal of Botany*, 57, 1577–1572. Available from: <https://doi.org/10.1139/b79-194>
- Owens, J.N. & Molder, M. (1984) *The reproductive cycle of Interior spruce*. Province of British Columbia, Ministry of Forests, Information Services Branch: Victoria.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14, 417–419. Available from: <https://www.nature.com/articles/nmeth.4197>
- Poovalah, C., Phillips, L., Geddes, B., Reeves, C., Sorieul, M. & Thorlby, G. (2021) Genome editing with CRISPR/Cas9 in *Pinus radiata* (D. Don). *BMC Plant Biology*, 21, 363. Available from: <https://doi.org/10.1186/s12870-021-03143-x>
- Pullman, G.S. & Bucalo, K. (2014) Pine somatic embryogenesis: analyses of seed tissue and medium to improve protocol development. *New Forest*, 45, 353–377. Available from: <https://doi.org/10.1007/s11056-014-9407-y>
- Pullman, G.S., Johnson, S., Peter, G., Cairney, J. & Xu, N. (2003) Improving loblolly pine somatic embryo maturation: comparison of somatic and zygotic embryo morphology, germination, and gene expression. *Plant Cell Reports*, 21, 747–758. Available from: <https://doi.org/10.1007/s00299-003-0586-9>
- R Core Team. (2021) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raventós, D., Skriver, K., Schlein, M., Karnahl, K., Rogers, S.W., Rogers, J.C. *et al.* (1998) HRT, a novel zinc finger, transcriptional repressor from barley. *The Journal of Biological Chemistry*, 273, 23313–23320. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0021925819749928>
- Rosvall, O., Bradshaw, R.H.W., Egertsdotter, U., Ingvarsson, P.K. & Wu, H. (2019) Using Norway spruce clones in Swedish forestry: introduction. *Scandinavian Journal of Forest Research*, 34, 333–335. Available from: <https://doi.org/10.1080/02827581.2018.1562565>
- Rupps, A., Raschke, J., Rümmler, M., Linke, B. & Zoglauer, K. (2016) Identification of putative homologs of *Larix decidua* BABYBOOM (BBM), LEAFY COTYLEDON1 (LEC1), WUSCHEL-related HOMEBOX2 (WOX2) and SOMATIC EMBRYOGENESIS RECEPTOR-like KINASE (SERK) during

- somatic embryogenesis. *Planta*, **243**, 473–488. Available from: <https://doi.org/10.1007/s00425-015-2409-y>
- Singh, H.** (1978) Embryology of gymnosperms. In: Zimmermann, W., Carlquist, Z., Ozenda, P. & Wulff, H.D. (Eds.) *Handbuch der Pflanzenanatomie (Encyclopedia of Plant Anatomy)*. Berlin: Gebrüder Borntraeger, pp. 187–241.
- Soneson, C., Love, M.I. & Robinson, M.D.** (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, **4**, 1521. Available from: <https://f1000research.com/articles/4-1521/v1>
- Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y.C., Sjödin, A. et al.** (2015) The plant genome integrative explorer resource: PlantGenIE.Org. *The New Phytologist*, **208**, 1149–1156.
- Tatematsu, K., Nakabayashi, K., Kamiya, Y. & Nambara, E.** (2008) Transcription factor AtTCP14 regulates embryonic growth potential during seed germination in *Arabidopsis thaliana*. *The Plant Journal*, **53**, 42–52. Available from: <https://doi.org/10.1111/j.1365-3113X.2007.03308.x>
- Trontin, J.-F., Klimaszewska, K., Morel, A., Hargreaves, C. & Lelu-Walter, M.-A.** (2016) Molecular Aspects of Conifer Zygotic and Somatic Embryo Development: A Review of Genome-Wide Approaches and Recent Insights. In: Germana, M.A. & Lambardi, M. (Eds.) *Methods in Molecular Biology*. New York: Springer Science & Business Media, pp. 167–207. Available from: https://doi.org/10.1007/978-1-4939-3061-6_8
- Uddenberg, D., Valladares, S., Abrahamsson, M., Sundström, J.F., Sundås-Larsson, A. & Von Arnold, S.** (2011) Embryogenic potential and expression of embryogenesis-related genes in conifers are affected by treatment with a histone deacetylase inhibitor. *Planta*, **234**, 527–539. Available from: <https://doi.org/10.1007/s00425-011-1418-8>
- Varis, S., Klimaszewska, K. & Aronen, T.** (2018) Somatic embryogenesis and plant regeneration from primordial shoot explants of *Picea abies* (L.) H. Karst. Somatic trees. *Frontiers in Plant Science*, **9**, 1–8. Available from: <https://doi.org/10.3389/fpls.2018.01551/full>
- Vetrici, M.A., Yevtushenko, D.P. & Misra, S.** (2021) Douglas-fir LEAFY COTYLEDON1 (PmLEC1) is an active transcription factor during zygotic and somatic embryogenesis. *Plant Direct*, **5**, 1–22. Available from: <https://doi.org/10.1002/pld3.333>
- Von Arnold, S. & Clapham, D.** (2008) In: Suárez, M.F. & Bozhkov, P.V. (Eds.) *Spruce Embryogenesis BT—Plant Embryogenesis*. Totowa, NJ: Humana Press, pp. 31–47. Available from: https://doi.org/10.1007/978-1-59745-273-1_3
- Vuosku, J., Sarjala, T., Jokela, A., Sutela, S., Sääskilähti, M., Suorsa, M. et al.** (2009) One tissue, two fates: different roles of megagametophyte cells during scots pine embryogenesis. *Journal of Experimental Botany*, **60**, 1375–1386.
- Wang, M., Zhao, Y. & Zhang, B.** (2015) Efficient test and visualization of multi-set intersections. *Scientific Reports*, **5**, 16923.
- Yan, J., Buer, H., Wang, Y.P., Zhula, G. & Bai, Y.** (2021) Transcriptomic time-series analyses of gene expression profile during zygotic embryo development in *Picea mongolica*. *Frontiers in Genetics*, **12**, 1–13. Available from: <https://doi.org/10.3389/fgene.2021.738649/full>
- Zhang, Y. & Fernie, A.R.** (2023) The role of TCA cycle enzymes in plants. *Advanced Biology*, **7**, 1–12. Available from: <https://doi.org/10.1002/adb.202200238>
- Zhao, H., Niu, Y., Dong, H., Jia, Y. & Wang, Y.** (2021) Characterization of the function of two S1Fa-like family genes from *Populus trichocarpa*. *Frontiers in Plant Science*, **12**, 1–15. Available from: <https://doi.org/10.3389/fpls.2021.753099/full>