Tree Genetics & Genomes (2019) 15:39
https://doi.org/10.1007/s11295-019-1344-7

ORIGINAL ARTICLE

Predicted functions, subcellular localizations, and expression patterns of genes encoding secretory proteins associated with pine pollen germination

Arnold M. Salazar[1] · Danilo D. Fernando[1]

Received: 25 January 2019 / Revised: 2 April 2019 / Accepted: 14 April 2019

Abstract

Pollen germination represents the transition from mature to germinated pollen, which is a critical event in seed plant reproduction. Cell wall–related processes represent many of the cellular activities during this transition but information on the genes involved is limited, particularly in gymnosperms. Yeast secretion trap (YST) is employed in the current study to isolate cDNAs encoding secretory proteins associated with in vitro germinated pollen of loblolly pine (*Pinus taeda*). YST is a functional screen and when coupled with computational prediction provides insights on the diversity of populations, subcellular localizations, and functions of the encoded proteins. Based on 100 confirmed YST clones, our results identified 21 known, 4 unknown, and 10 hypothetical genes encoding secretory proteins, which are mostly predicted to be localized in the plasma membrane or extracellular space. Based on the known sequences, pollen germination involves genes associated with cell wall degradation, biosynthesis and remodeling, stress and defense responses, signaling, and protein processing. This study characterizes further 10 highly expressed cDNAs based on their temporal (mature pollen vs germinated pollen) and spatial (germinated pollen, young stem, and needle tissue) expression patterns, as well as sequence features such as the presence of transmembrane α-helix, glycosylphosphatidylinositol anchor, and conserved domains. It appears that pine pollen germination involves genes with static and dynamic expression profiles including those with germination-specific expressions. This study confirms the distinct expression profiles of mature and germinated pollen, and expands our understanding of the likely molecular players and processes associated with pine pollen germination, particularly pollen tube wall formation.

Keywords Conifers · *Pinus taeda* · Gymnosperms · Pollen tube · Cell Wall · Extracellular matrix

Introduction

The pollen grain is one of the innovations that played a major role in the success of seed plants, particularly in the colonization of dry lands. It represents the reduced male gamete-producing phase of seed plants' life cycle, which is well adapted for long-distance dispersal. The evolution of the pollen grain allowed sexual reproduction to proceed without the reliance on freestanding water for fertilization, and together with at least the evolution of the ovule, these traits likely propelled the domination of seed plants in most terrestrial ecosystems since the Permian (Crane and Leslie 2014). Gymnosperm pollen represents a different line of evolutionary adaptation in seed plant reproduction which is characterized by many structural, physiological, and developmental traits including generally slow pollen germination and tube growth (Fernando et al. 2010). These and other traits differentiate them from the remarkably faster germination and tube growth of angiosperm pollen, which has evolved under different evolutionary time and environmental conditions (Williams 2012).

Gymnosperm pollen commonly lands on the pollination drop at the tip of the naked ovule, and depending on its orientation, the pollen usually floats on or sinks in the pollination drop to arrive on the nucellus inside the ovule (Owens et al. 1998; Fernando et al. 2010). On the nucellus, the pollen

Communicated by A. M. Dandekar

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-019-1344-7) contains supplementary material, which is available to authorized users.

✉ Danilo D. Fernando
dfernando@esf.edu

[1] Department of Environmental and Forest Biology, State University of New York College of Environmental Science and Forestry, One Forestry Drive, Syracuse, NY 13210, USA

Published online: 02 May 2019

germinates and in most species, becomes stationary or dormant within a week, resumes pollen tube growth only after several months, and then releases sperm for fertilization (Fernando et al. 2010). In angiosperms, pollen lands on the stigma, usually germinates within a few minutes or hours; pollen tube grows through the style and enters the micropyle of an ovule to accomplish fertilization (Lora et al. 2015). Thus, interactions between pollen and female reproductive structures also differ between these two major groups of seed plants. Furthermore, pollination, pollen germination, pollen tube growth, and fertilization all occur within the same season in angiosperms (Williams 2012), whereas in gymnosperms, pollination and pollen germination occur in one season while pollen tube growth and fertilization occur in another season (Fernando et al. 2010). Pollen germination is characterized by increased in cellular metabolism, the establishment of polarity, and cell wall formation. It is one of the critical developmental events in seed plant reproduction but perhaps, even more so in gymnosperms since they exhibit a long stationary phase that makes them vulnerable to biotic and abiotic influences (Crane and Leslie 2014; Fernando 2014). Our knowledge of the developmental and molecular mechanisms behind pollen formation, germination, and tube growth continues to advance at a tremendous pace in angiosperms, particularly in *Arabidopsis*, rice, maize, and tobacco (Rutley and Twell 2015; Zhang et al. 2017), while this area of research remains understudied in gymnosperms.

Proteomic analyses that compared both mature and germinated pollen in the same study have revealed several interesting differentially expressed and pollen germination/tube growth-specific proteins in gymnosperms (Fernando 2005; Chen et al. 2006, 2012) and angiosperms (Dai et al. 2007; Zou et al. 2009; Zhu et al. 2011). It is interesting to note that only a few of the proteins identified in these studies are related to pollen tube (or cell) wall formation and information on their subcellular localization is lacking. To obtain a better understanding of proteins in germinating pollen, the identification of genes encoding proteins associated with subcellular compartments such as plasma membrane (Han et al. 2010; Yang and Wang 2017) or apoplast (Ge et al. 2011; Hafidh et al. 2016) has been done for angiosperms. These studies have put forward interesting candidate proteins involved in pollen germination, pollen tube growth, and pollen tube guidance into the ovule. To date, no published information is available on genes encoding proteins related to the secretory pathway during pollen germination in gymnosperms.

The primary goal of this study is the isolation, identification, and characterization of genes encoding secretory proteins associated with germinated pine pollen to contribute to our understanding of the mechanisms involved in pollen germination. Cell wall–related processes represent many of the cellular activities in germinating pollen, and the utilization of a gene isolation approach designed for their capture is necessary. Yeast secretion trap (YST) is a gene cloning strategy and a functional screen to identify secreted proteins based on their subcellular destinations (Lee et al. 2006). This approach has been successfully used in the isolation of large numbers of proteins from the plasma membrane, cell wall, and apoplast from various plant tissues (Yamane et al. 2005; Yeom et al. 2011; Lee et al. 2014).

This study aims to (1) isolate cDNAs encoding secretory proteins from in vitro germinated pine pollen using the high-throughput yeast secretion trap; (2) obtain insights on the diversity, subcellular localizations, functions, and amino acid sequence characteristics through computational analysis; and (3) determine spatio-temporal expression patterns of representative cDNAs through Real-Time PCR. This study highlights representative genes encoding secreted proteins that are associated with pollen germination and provides support to the distinct expression profiles of mature (ungerminated) versus germinated pollen. Overall, this study furthers our understanding of the molecular components that likely constitute the network of proteins involved in pollen germination in gymnosperms, with loblolly pine (*Pinus taeda*) pollen as the experimental system. Loblolly pine is one of the most ecologically and economically important forest tree species in the Eastern United States. It is a significant biological resource not only for the timber, pulp, and paper industries, but also as a non-food lignocellulosic biofuel feedstock and in mitigating climate change through carbon sequestration (Daystar et al. 2014). Thus, an understanding of the mechanism behind pollen germination will help facilitate fertilization that may result in increased seed production and creation of novel hybrids.

Materials and methods

Plant material and culture conditions

Pollen cones of loblolly pine (*Pinus taeda*) were obtained from trees growing in the campus of North Carolina State University, Raleigh, NC. Procedures for the extraction of sterile (microbe contamination-free) mature pollen grains from pollen cones, pollen storage, and culture conditions were those of Fernando et al. (1997). Mature pollen is defined as the stage of pollen at the time that they are released from the male cone for dispersal, whereas germinated pollen is defined as the stage when the pollen tube is about as long as the diameter of the pollen grain. In conifers, germinated pollen usually occurs at approximately 48 h under in vitro conditions (Fernando et al. 1997). At this stage, the pollen tubes are mostly only about as long as the diameter of the pollen grains and no pollen tubes are more than twice the size of the pollen grains (Fig. 1). Thus, the developmental stage examined in this study represents a narrow window so as to increase the chance of capturing more germination-related proteins.

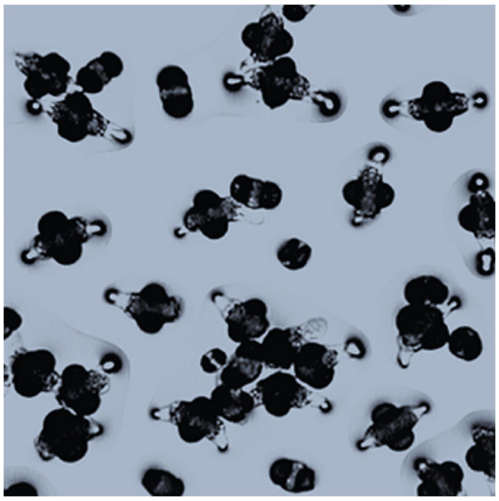

Fig. 1 Stage of pollen germination in loblolly pine (*Pinus taeda*) examined in the study

Isolation of secretory proteins

The premise of the yeast secretion trap (YST) is that any yeast cell transformed with a cDNA that encodes a secretory protein has the potential to secrete such polypeptide as an invertase fusion protein, which results in the reconstitution of the extracellular invertase activity and thus, survival of the mutant yeast (Lee et al. 2006). YST was utilized to isolate genes encoding secreted proteins from germinated loblolly pine pollen following the protocol of Lee et al. (2006). Briefly, mRNA was purified from total RNA extracted from in vitro grown germinated loblolly pine pollen (48 h in culture) using Oligotex mRNA Mini Kit (Qiagen, Valencia, CA). First- and second-strand cDNAs were synthesized (from 5 μg purified mRNA) using a linker-primer with *Not*I site then ligated *Eco*RI adapters to the blunted cDNAs and phosphorylated. cDNAs were purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA) and subjected to restriction digests and fragments were gel fractionated using 1% agarose. Fragments with sizes corresponding to 0.3 to 0.5 kb were sliced out from the gel, purified using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA), and ligated into the *Eco*RI and *Not*I sites of the YST vector. Quality and transformation efficiency of the cDNA library were assessed through the transformation of TOP10 electrocompetent *E.coli* strain (Invitrogen, Carlsbad, CA). A large quantity of stable cDNA library was produced through pooling of cDNA library plasmids from *Escherichia coli* transformation. Quantity of the library plasmid DNA was evaluated using ND-1000 spectrophotometer and 50 μg was used for yeast (*Saccharomyces cerevisiae* DBYα2445) transformation with sucrose as a sole selection agent. As the transformants show up on the selection medium, they were individually re-streaked on a freshly prepared medium to verify the secretion of the clones under the YST system. The first 100 verified YST clones were used for further analysis. Plasmid DNAs were isolated from each of the 100 YST clones, resuspended in TE buffer and stored at − 80 °C until use. Each plasmid DNA was cloned in *E. coli*, plasmid DNA was isolated and sent to the Biotechnology Resource Center at Cornell University for sequencing using the AHD1 promoter sequence primer of the YST vector (5′-TCCTCGTCATTGTTCTCGTTCC-3′).

Computational analysis of YST clones

Inserts within the plasmid DNAs were identified by locating the *Eco*RI and *Not*I sites using the programs 4Peaks and EnzymeX (www.mekentosj.com) followed by performing BLASTx (www.blast.ncbi.nlm.nih.gov) search using default parameters. ExPASY (www. expasy.org) was used to translate DNA sequences into amino acid sequences and determine the correct reading frame. The amino acid sequences were used as inputs in SignalP 4.1 (www.cbs.dtu.dk/services/SignalP/) to predict the presence of N-terminal signal peptides and their cleavage sites. The cutoff value for the presence of signal peptides was set to $D = 0.5$ instead of the default value of $D = 0.45$ to increase the sensitivity of the predictions. The *D* score (degree of confidence) instead of the *C*, *S*, or *Y* score in the previous SignalP versions has been shown to be the more accurate in an evaluation study of signal peptide predictors (Klee and Ellis 2005). BUSCA (Bologna Unified Subcellular Annotator) developed by Savojardo et al. (2018) was used to predict the subcellular localizations of the proteins and re-checked of the presence of signal peptides and other localization sequences. Data from NCBI (www.ncbi.nlm.nih.gov), UniProt (www.uniprot.org), and Gene Ontology (www.geneontology.org) were used to classify the proteins into their various functional categories.

Computational analysis of representative secretory proteins

SuperScript III Reverse Transcriptase (GeneRacer Kit, Invitrogen, Grand Island, NY) was used to generate the templates from the 10 well-represented cDNAs to obtain full-length sequences. Rapid amplifications of cDNA ends were done using Taq PCR kit (New England Biolabs, Ipswich, MA), gene-specific 5′ primers (Supplement 1), and GeneRacer 3′ primer (5′-GCTGTCAACGATACGCTACG TAACG-3′). Using full-length sequences, the presence of signal peptides and cleavage sites, subcellular localizations, and protein homologies and functions were predicted using the same programs described above. The BUSCA program integrates many different tools to predict localization-related features including ENSEMBLE3.0 (recognition and topology prediction of all α-membrane proteins; Martelli et al. 2003), PredGPI (prediction of GPI anchors in proteins, Pierleoni et al. 2008), and TPpred3 (prediction of organelle-targeting peptides, Savojardo et al. 2015).

Thus, the availability of full-length sequences allowed the prediction of the presence of transmembrane α-helixes (TMH) and glycosylphosphatidylinositol (GPI) anchos. It also allowed the identification of conserved structural and/or functional domains using NCBI's Conserved Domain Database (CDD). Full-length cDNA sequences from the 10 representative genes have been submitted to GenBank (www.ncbi.nlm.nih.gov/WebSub/?tool=genbank).

Temporal and spatial expression analysis

Total RNA was extracted using TRIzol Reagent (Life technologies, Grant Island, NY) from three biological replicates of mature and germinated loblolly pine pollen, where each replicate set of mature and germinated pollen was obtained from three different individual trees collected from the same location as described previously. SuperScript III Reverse Transcriptase from GeneRacer kit (Invitrogen, Grand Island, NY) was used to synthesize cDNAs. Gene-specific 5′ primers (Supplement 2) for each of the 10 representatives were designed from consensus regions of cDNA sequences after homology search and multiple alignments through NCBI. RACE-PCR was performed using Taq PCR kit (New England Biolabs, Ipswich, MA), and PCR products were purified using Zymo DNA Clean & Concentrator kit (Irvine, CA). Initial tests were conducted for all primers including the reference gene 18s rRNA through PCR. The specificity of the semi-qReal-Time PCR protocol and gene-specific primers used in the expression study was evaluated through melting curve analysis. Transcript expression analyses were also done using cDNA templates generated from young stems and needles of loblolly pine in triplicates from the same three individuals as indicated above. Significance of expression levels was determined using the two-tailed *t* test (Excel, Microsoft Corp).

Results

Sequence analysis of YST clones

The first step in the identification of encoded secreted proteins from the 100 verified YST clones or sequences was based on the presence of N-terminal signal peptides using SignalP (Tables 1 and 2). Of the 100 YST sequences, 91 were predicted to have signal peptides and most (i.e., 83) have *D* scores of at least 0.5 (Tables 1 and 2), which indicates a high degree of confidence for the presence of these signal peptides. Prediction of the subcellular locations of the 91 sequences based on BUSCA software indicates localization to the plasma membrane (PM), extracellular space (ECS), chloroplast (CHL), mitochondria (MIT), and nucleus (NUC) (Tables 1 and 2). The eight sequences with signal peptide *D* score lower than 0.5 were those predicted to be localized in CHL, MIT, and NUC, but these sequences have *D* scores higher than 0.5 for their localizations (Table 2). PSORT (https://psort.hgc.jp/) analysis indicates that none of the 91 sequences contained KDEL/HDEL motifs for retention in the ER or Golgi apparatus and so overall, their sequence characteristics indicate that they encode conventionally secreted proteins. Four of the 100 YST clones that did not contain signal peptides. Five of the 100 YST clones contained signal peptides, but the rest of the encoded proteins were too short (only 3–5 amino acids long) to be confidently identified and so they were not included for further analysis.

Based on BLASTx sequence homology, the 91 sequences were represented by 21 known (homologous to genes with known function), 4 unknown (homologous to genes with no known function), and 10 hypothetical (no sequence homology in the database) protein-encoding genes (Tables 1 and 2). Overall, these numbers indicate a total of 35 non-redundant nucleic acid sequences. Our results show that while some of these cDNAs were represented in our library by only one YST clone, others were represented by two to eight YST clones (Tables 1 and 2). These results indicate redundancy in the cloned cDNAs, but it also implies that these cDNAs are strongly expressed (Link and Voegele 2008). Of the 21 known genes encoding secreted proteins, 8 were represented by at least three YST clones (Table 1) indicating their relative abundance in the cDNA library from germinated pine pollen and thus, they were chosen for further analysis. These eight genes encoding secretory proteins include ALLW1950, arabinogalactan protein, desiccation-related protein, expansin A, extensin-like protein, CHALLAH-like2 like protein, serine carboxypeptidase, and thaumatin-like protein. Sequences from seven YST clones matched with four different unknown genes which were represented by one to three YST clones (Table 2), and the most abundant sequence (represented by Yeast Colony # 24, 34, and 87) was chosen as a representative for the set of unknown genes for further analysis and we refer to this sequence as YC24. Twenty-seven YST clones were represented by 10 different sequences that did not match with any gene in the NCBI database. We also chose the sequence that is most abundant (i.e., YC54) as a representative of the hypothetical genes. In total, 10 genes encoding different secretory proteins were chosen for further analysis and Table 3 shows their various sequence characteristics including the presence of signal peptide cleavage sites. The gene encoding thaumatin-like protein is the only one that has no cleavage site among the 10 representatives (Table 3), as well as among all the other known, unknown, and hypothetical encoded proteins examined in this study.

Table 1 Known genes encoding putative secreted proteins from germinated pine pollen

No. of YST clones	Identity	E-value	Accession #	Species	Function	SignalP *D* score	BUSCA *D* score
1	Alpha-D-galactosidase	1e-70	PIN24325.1	*Handroanthus impetiginosus*	Cell wall degradation	0.973	ECS (0.68)
8	ALLW1950	1e-29	AAQ89160.1	*Homo sapiens*	Signaling	0.781	ECS (0.87)
8	Arabinogalactan protein, putative	0..015	AAF75827.1	*Pinus taeda*	Cell wall biosynthesis	0.900	ECS (0.83)
1	Aspartyl protease family protein At5g10770-like	2e-26	XP_020082781.1	*Ananas comosus*	Stress and defense	0.881	PM (0.57)
2	Beta-1,3-galactosyltransferase 7	2e-14	XP_002269415.1	*Vitis vinifera*	Cell wall biosynthesis	0.596	PM (0.86)
2	Calreticulin	6e-80	AAC49696.1	*Arabidopsis thaliana*	Signaling	0.757	ECS (0.99)
1	Cellulose synthase-like protein	0.002	O23386.1	*Arabidopsis thaliana*	Cell wall biosynthesis	0.834	PM (0.79)
1	Cellulose synthase A catalytic subunit 3, probable	4.7	XP_004499618.1	*Cicer arietinum*	Cell wall biosynthesis	0.565	PM (0.79)
4	CHALLAH-like2 like protein	4e-15	PSS33841.1	*Actinidia chinensis var. chinensis*	Signaling	0.669	ECS (0.89)
1	Chaperonin-containing TCP1	4e-22	NP_172392.1	*Capsaspora owczarzaki*	Cell wall biosynthesis	0.744	ECS (0.99)
2	Clathrin assembly protein 2	6e-04	XP_002270803	*Vitis vinifera*	Protein processing	0.839	ECS (0.820)
3	Desiccation-related protein PCC13–62	1e-114	XP_006844727.1	*Amborella trichopoda*	Stress and defense	0.755	ECS (1.0)
7	Expansin A	1e-117	ABO32367.1	*Lichi*	Cell wall remodeling	0.887	ECS (0.9)
3	Extensin-like protein	9e-97	ADM76653.1	*Picea sitchensis*	Cell wall remodeling	0.893	ECS (1.0)
1	Glucan endo-1,3-B-glucosidase precursor (PR2)	2.9	XP_002530862.1	*Ricinus communis*	Stress and defense	0.574	ECS (0.94)
2	Pathogenesis-related protein (PR1)	7e-21	XP_007014137.1	*Theobroma cacao*	Stress and defense	0.835	ECS (0.98)
1	Proline-rich protein	1	AAF75827.1	*Pinus taeda*	Cell wall remodeling	0.741	ECS (0.99)
1	Purple acid phosphatase	2.3	ACR23328.1	*Triticum aestivum*	Cell wall biosynthesis	0.863	ECS (1.0)
3	Serine carboxypeptidase, peptidase S10	1e-75	ABR16596.1	*Picea sitchensis*	Signaling	0.829	PM (0.87)
2	Subtilisin-related protein, peptidase S8	6e-06	PWA90026.1	*Artemisia annua*	Protein processing	0.843	ECS (1.0)
3	Thaumatin-like protein 11	4e-87	ABM91761.1	*Taxus x media*	Cell wall degradation	0.546	PM (0.77)

Table 2 Unknown and hypothetical genes encoding putative secretory proteins from germinated pine pollen (dash means not applicable)

No. of YST Clones	Unknown proteins (yeast colony #)	E-value	Accession #	Species	SignalP *D* score	BUSCA *D* score
1	9	6e-08	XP_019176861.1	*Ipomoea nil*	0.893	ECS (0.97)
3	24 (YC24), 34, 87	0.025	ABK23092.1	*Picea sitchensis*	0.765	PM (0.96)
1	35	0.005	ABR16121.1	*Picea sitchensis*	0.872	ECS (0.97)
2	50, 83	0.071	GAU46823.1	*Trifolium subterraneum*	0.866	PM (0.95)
No. of YST clones	**Hypothetical proteins (yeast colony #)**	**E-value**	**Accession #**	**Species**	**SignalP *D* score**	**BUSCA *D* score**
2	7, 8	–	–	–	0.862	ECS (1.0)
1	10	–	–	–	0.252	CHL (1.0)
4	23, 37, 76, 77	–	–	–	0.513	PM (0.6)
2	30, 63	–	–	–	0.383	NUC (1.0)
2	41, 51	–	–	–	0.365	NUC (1.0)
2	49, 79	–	–	–	0.669	CHL(0.90)
5	54 (YC54), 59, 62, 66, 68	–	–	–	0.883	ECS (0.95)
5	56, 58, 67, 93, 94	–	–	–	0.619	CHL (0.65)
1	80	–	–	–	0.662	CHL (0.83)
3	82, 88, 98	–	–	–	0.328	MIT (0.71)

Temporal and spatial expression analysis

Comparison of transcript expression levels of the 10 representative genes between mature and germinated pollen indicates that in the latter, six cDNAs were significantly upregulated and one was significantly downregulated (Fig. 2). Three of these cDNAs showed no significant difference in expression levels between mature and germinated pollen (Fig. 2). The genes encoding secreted proteins with significantly increased expressions in germinated pollen were arabinogalactan protein, desiccation-related protein, expansin A, extensin-like protein, serine carboxypeptidase, and hypothetical protein YC54. Expression of CHALLAH-like2 like protein was significantly downregulated in germinated pollen. Expansin A and serine carboxypeptidase have the highest increase in expression in germinated pollen, whereas the lowest decrease was observed in the expression of CHALLAH-like2 like protein. Three highly expressed secretory proteins but whose expressions were not significant between mature and germinated pollen were ALLW1950, thaumatin-like protein, and unknown protein YC24 (Fig. 2).

Comparison of expression levels between germinated pollen and young stem and needle tissues from loblolly pine indicated that expansin A, extensin-like protein, serine carboxypeptidase, and hypothetical protein YC54 were significantly higher in germinated pollen than in either stem or needle tissues (Fig. 3). The expression levels of extensin-like protein and hypothetical protein YC54 in the stem and needle

Table 3 Sequence characteristics of 10 representative genes encoding secreted proteins based on various computational prediction programs (dash indicates no feature detected, TMH is transmembrane α-helix, SA is signal anchor and GPI is glycosylphosphatidylinositol)

Representatives (putative identity of cDNAs)	GenBank accession number	Full-length sequence (No. of aa)	Signal peptide length (SignalP)	Cleavage site position (SignalP)	Cleavage site *D* score (SignalP)	Conserved domain identity (E-value) (NCBI) and other sequence features (BUSCA)
ALLW1950	MK388869	84	1–26	26–27	0.781	–
Arabinogalactan	MK388870	74	1–23	23–24	0.900	–
Desiccation-related protein	MK388871	325	1–26	26–27	0.755	Ferritin 2 (5.13e-45)
Expansin A	MK388872	269	1–27	27–28	0.888	Double-psi beta barrel (9.05e-114)
Extensin-like Protein	MK388873	159	1–22	22–23	0.902	Pollen Ole e I (2.28e-15)
CHALLAH-like2 like protein	MK388874	118	1–28	28–29	0.717	Epidermal patterning factor (9.58e-14)
Serine Carboxypeptidase	MK388875	232	1–21	21–22	0.512	Peptidase S10 (6.83e-39); TMH
Thaumatin-like protein	MK388876	229	1–25	–	0.440	Thaumatin (9.46e-98); SA
Unknown YC24	MK388877	64	1–24	24–25	0.768	TMH; GPI
Hypothetical YC54	MK388878	74	1–22	22–23	0.843	GPI

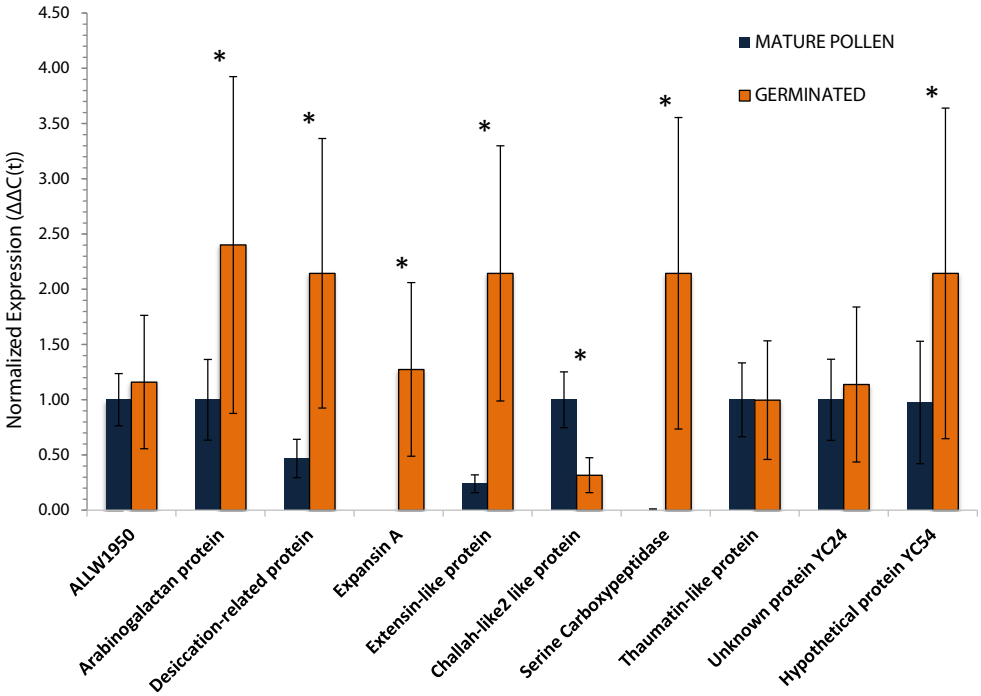

Fig. 2 qReal-Time PCR analysis of the 10 representative genes encoding secretory proteins from germinated pollen of loblolly pine (*Pinus taeda*) (significance determined using two-tailed *t* test, $p \leq 0.05$, $n = 12$, error bar = standard deviation)

tissues were at relatively low levels but nevertheless indicate that they were likely not specific to germinated pollen. On the other hand, expressions of expansin A and serine carboxypeptidase in the stem and needle tissues were at background levels (Fig. 3) and so they are likely specific to germinated pollen.

Sequence characteristics of representative encoded proteins

Computational analysis of the full-length amino acid sequences of the 10 representative cDNAs confirmed our initial predictions of their respective identities and subcellular localizations. It also allowed the prediction of signal peptide lengths and cleavage sites and the presence of conserved (or functional) domains, transmembrane α-helices, and GPI anchors. Conserved domains were predicted for six of the representative secretory proteins including desiccation-related protein, expansin A, extensin-like protein, CHALLAH-like2 line protein, serine carboxypeptidase, and thaumatin-like protein (Table 3). In the full-length aa sequence of desiccation-related protein, a ferritin 2 domain was detected and it spanned from amino acid positions 65 to 231. The expansin A full-length cDNA revealed domain for double-psi beta barrel (DPBB) that spanned from amino acid positions 47 to 264. The full-length sequence of extensin-like protein possessed a pollen Ole e 1 superfamily domain that spanned from the 30 to 110 amino acid positions. CHALLAH-like2 like protein exhibits epidermal patterning factor domain spanning amino

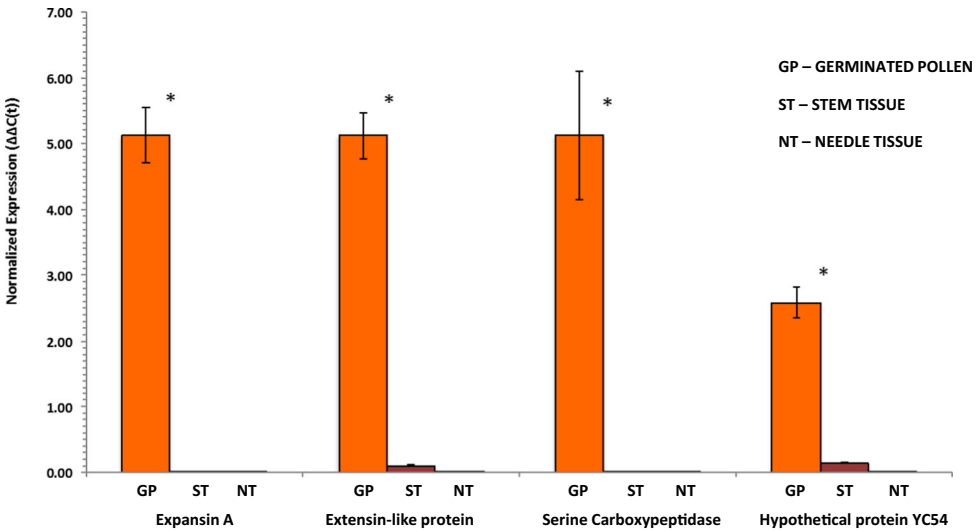

Fig. 3 qReal-Time PCR analysis of representative genes encoding secretory proteins from germinated pollen, stem, and needle tissues of loblolly pine (*Pinus taeda*) (significance determined using two-tailed *t* test, $p \leq 0.05$, $n = 12$, error bar = standard deviation)

acids 65 to 118, whereas in serine carboxypeptidase, the peptidase S10 superfamily domain spanned from 88 to 171 amino acid positions. In thaumatin-like protein, the conserved domain for thaumatin spanned from the amino acids 30 to 229. Based on the results from both BUSCA and TMHMM programs, three of the representative secreted proteins are consistently predicted to have transmembrane α-helices (TMHs). Serine carboxypeptidase and thaumatin-like protein are predicted to have one TMH each (amino acid positions 48 to 66 and 3 to 25, respectively), while unknown protein YC24 has two TMHs (amino acid positions 4 to 26 and 38 to 60). The TMH in thaumatin-like protein and the first TMH in unknown protein YC24 are located at their N-terminus and so they strongly overlap in position with their predicted signal peptides (see Table 3). This result indicates that their signal peptide sequences likely also functions as signal anchors. Interestingly, in thaumatin-like protein, the *D* score (0.440) for the probability of cleavage is below the cutoff value of $D = 0.5$, which indicates low confidence that it will be cleaved and thus may represent a typical type II transmembrane protein. In contrast, the other TMH in unknown protein YC24 and that of serine carboxypeptidase are positioned away and differently from their cleavable signal peptides and thus may represent as type I transmembrane proteins. GPI anchors are attached to the C-terminus of polypeptide chains as one of the forms of post-translational protein modifications. Proteins with GPI anchors are found in the outer leaflet of the lipid bilayer and thus, position the protein to face the extracellular environment. Based on PredGPI software, unknown protein YC24 and hypothetical protein YC54 were predicted to have GPI anchors with specificity of 99.6% and 99.2%, respectively. This result collaborates with the prediction of their plasma membrane localizations. Their anchor cleavage sites were predicted at the glutamic acid residue (36st omega-site position) and histidine residue (46th omega-site position), respectively.

Discussion

Pollen germination is characterized by increased in cellular metabolism and the establishment of polarity and cell wall formation. It is a critical developmental process in seed plant reproduction, and our results provide valuable insights on the diversity of populations, subcellular localizations, and functions of genes encoding secreted proteins associated with pine pollen germination under in vitro conditions. The YST approach has been successful in the isolation of larger numbers of genes encoding cell wall–related proteins from germinating pine pollen, particularly in comparison with our previous report (Fernando 2005) and others (Chen et al. 2006, 2012) that used gel-based proteomics on gymnosperms. Of the known genes in the current study, most of the predicted functions include cell wall degradation, biosynthesis, and remodeling (52%), followed by proteins involved in stress and defense responses (19%), signaling (19%), and protein processing (10%). Whereas the known and unknown proteins have either plasma membrane or extracellular space predicted localizations, the hypothetical proteins show more varied predicted localizations that include the chloroplast, mitochondria, and nucleus. Known, unknown, and hypothetical proteins represent 60%, 11%, and 29% of the total proteins predicted in the current study.

To gain further insights on the functions of genes encoding secretory proteins associated with germinated pine pollen, the current study further characterizes 10 predominant cDNAs as representatives in terms of temporal and spatial expression patterns, and sequence characteristics based on full-length sequences. Whereas most of the predicted secreted proteins in germinated pine pollen (54%) appear to be localized to the extracellular space (ECS), 7 of the 10 representatives also have ECS localizations. This includes ALLW1950, arabinogalactan protein, CHALLAH-like2 like protein, desiccation-related protein, expansin A, extensin-like protein, and hypothetical protein YC54. The ECS represents the structure that protects and supports the cell and in the case of germinated pine pollen, it refers to the pollen tube wall. A broader concept of ECS includes the intercellular spaces along with the cell wall and referred to as apoplast which serves to regulate the behavior of the cell. In pine pollen tube growing in situ, the apoplast includes the intercellular spaces between the growing pollen tube and the surrounding nucellar cells. In germinated pollen under in vitro condition, the apoplast also represents the outside of the pollen tube wall including what is secreted into the culture medium. Considering that the current study did not distinguish between these two subcellular locations (i.e., pollen tube wall and intercellular spaces), the predicted proteins that have ECS localizations may actually be localized in the pollen tube wall itself, intercellular space, or both. Thus, these proteins predicted to have ECS localizations may have a role in pollen tube wall formation and/or pollen tube-nucellus interactions, or in the case of the current study, pollen tube and culture medium interactions. It remains to be seen how drastic the difference are between these two growing conditions for pine pollen germination and if the narrow window of the germination stage examined in this study helped reduce this difference. The other three representatives are predicted to be localized in the plasma membrane and include serine carboxypeptidase, thaumatin-like protein, and unknown protein YC24.

Upregulated representative genes encoding secretory proteins

Desiccation-related proteins (DRPs) are usually thought to be involved in protecting spores, seeds, and pollen from the adverse effects of cold stress and dehydration during their

maturation stage, and these DRPs commonly belong to the dehydrin (DHN) and late embryogenesis abundant (LEA) protein groups (Salmi et al. 2005; Yang et al. 2007). However, some DRPs are structurally and functionally different from DHN and LEA proteins since they are characterized by the presence of ferritin or ferritin-like domain and considered to be involved in plant defense against infection by pathogens (Zha et al. 2013). Ferritin as a protein is thought to be involved in iron storage and homeostasis (Andrews 2010), and that plant hosts may use iron to increase local oxidative stress in defense responses against pathogens (Anzar et al. 2015). In the current study, a DRP is predicted to be localized in the ECS and its transcripts are expressed in both mature and germinated pine pollen but significantly highly expressed in the latter, suggesting a greater role in pollen germination. Based on the possession of a ferritin 2 domain, its function appears more likely for defense against pathogens rather than cold or dehydration responses. Pollen grains may catch bacterial and fungal spores during their dispersal, which could grow while the pollen grains germinate in the nucellus and thus defense against pathogens at this stage is essential (Fernando 2005).

Expansins are produced by plants and bacteria and serve as cell wall–loosening proteins that promote the extension of primary cell walls without the hydrolysis of major structural components. Plant expansins are classified by sequence-based phylogeny into at least two major families, expansin A (EXPA or α-expansin) and expansin B (EXPB or β-expansin), and both exhibit the double-psi beta barrel and pollen allergen 1 domains (Sampedro and Cosgrove 2005). EXPA is associated with the induction of irreversible extension and enhancement of stress relaxation of plant cell walls in a pH-dependent manner (Cosgrove et al. 2002). EXPB includes the group-1 grass pollen allergens, known to facilitate pollen tube penetration into the stigma and style by dissolving the middle lamella in these maternal tissues (Cosgrove et al. 1997). Although expansins from plants and bacteria lack detectable lytic activity, they have been shown to cause paper weakening or cellulose fiber disruption and thus, the mechanism by which they loosen cell walls is likely through the disruption of noncovalent binding between cellulose microfibrils (Cosgrove 2015). An alternative model of cell wall formation indicates that despite the microfibrils being mostly tethered together by xyloglucan, they make direct contacts in a few regions in the cell wall (Park and Cosgrove 2015). This limited contact regions may be the sites targeted by EXPA and bacterial expansins to loosen walls through the mechanism described above (Wang et al. 2013). EXPB, on the other hand, as in the case of ZmEXPB1, has been shown to produce the same loosening effect but through the solubilization of arabinoxylans and homogalacturonan, a mode of action that appears to be specific to grass cell walls (Tabuchi et al. 2011). It seems that the unique grass cell wall composition has evolved selectivity for loosening by EXPB (Sampedro et al. 2015). In the current study, an EXPA is expressed at a significantly higher level in germinated pine pollen, and in contrast, at a background level in the mature pollen, implying that expression of EXPA is germination specific. The expression of EXPA in young stem and needle tissues of loblolly pine are also at background levels, strengthening its status as a germination-specific protein. These results also suggest an important role for EXPA in loblolly pine pollen germination. Our previous report has also identified an EXPA in germinated pollen of *Pinus strobus* (Fernando 2005). Whereas no expansin was reported from most studies that compared mature and germinated pollen in the same study (see Chen et al. 2006; Chen et al. 2012; Zou et al. 2009; Zhu et al. 2011), the type of expansin reported from rice and maize pollen is EXPB (Dai et al. 2007; Kapu and Cosgrove 2010).

Extensin-like protein was the first allergen purified from *Olea europaea* (Lauzurica et al. 1988). Structurally, it is characterized by a pollen Ole e 1 domain that has six conserved cysteine residues which may be involved in the formation of disulfide bonds (Villalba et al. 1993). There is a high diversity of proteins sharing this domain among plant species including proline-rich proteins, extensin-like protein, phosphoglycerate mutase, tyrosine-rich hydroxyproline-rich glycoprotein, and hydroxyproline-rich glycoprotein (Jimenez-Lopez et al. 2011). Various proteins with relevant homology to extensin-like pollen Ole e 1 protein have been described in other members of the Oleaceae family and other families (Villalba et al. 1993; Hanson et al. 1998). Although these pollen proteins are structurally related, their biological function is not yet known. However, they have been considered to be involved in pollen hydration, germination, pollen tube growth, as well as other reproductive and vegetative functions (Alché et al. 2004; Stratford et al. 2001; Hu et al. 2014). The extensin-like protein in germinated pine pollen is predicted to have ECS localization, and its expression level is increased as compared with those in mature pollen and stem and needle tissues. Its predicted function is cell wall remodeling but it is not known how exactly it is involved in this process.

Serine carboxypeptidases (SCPs) of the peptidase S10 superfamily are commonly known to catalyze proteolysis for functional protein maturation (Schaller 2004). They contain a conserved serine-asparagine-histidine catalytic triad, which is dispersed in the primary amino acid sequence and are brought close together in the protein's tertiary structure creating a nucleophilic serine for the hydrolysis of C-terminal peptide bonds in target proteins or peptides (Breddam 1986). SCPs have been isolated from roots, stems, leaves, fruits, and seeds (Breddam 1986; Bienert et al. 2012). In addition to protein maturation, they are also involved in diverse functions including protein turnover, programmed cell death, cell elongation, seed germination, and ABA or GA signaling (Schaller 2004; Bienert et al. 2012). Their mode of action has been mostly attributed to their proteolytic activity, but

some SCPs (i.e., SCP-like) exhibit non-proteolytic activity by catalyzing the transfer of acyl groups (Schaller 2004). It is considered that SCP-like proteins may have been recruited in the acquisition of this novel, non-proteolytic functions as acyltransferases, particularly in the production of plant secondary metabolites involved in herbivory defense and UV protection (Fraser et al. 2005). Schaller (2004) pointed out that the identification of SCPs as peptidases or acyltransferases on the basis of their amino acid sequence or mutant analysis (e.g., analysis of the active site in *Arabidopsis* mutant *BRS1*, Li et al. 2001) is unlikely since both enzyme activities rely on the same triad of amino acids. However, *BRS1* has since been shown to accumulate in the extracellular space and exhibit hydrolytic activity which may process a protein that participates in brassinosteroid signaling (Zhou and Li 2005). In the current study, the SCP in germinated pine pollen reveals a domain for peptidase S10 superfamily including the presence of the conserved catalytic triad. This secretory protein is targeted to the extracellular space and has one of the largest increases in expression levels in germinated pine pollen, suggesting an important role in pollen germination that is yet to be determined. Our functional annotation indicates its involvement in signaling. Interestingly, its expression levels in mature pine pollen and young stem and needle tissues are at background levels indicating germination-specific expression.

YC54 is a hypothetical protein that is expressed in both mature and germinated pine pollen, with significantly higher expression in the later stage. We also detected this protein in pine stem and needle tissues, although at significantly lower levels. It is one of the representative proteins examined in the current study that were predicted to have a GPI anchor, which adds to our understanding of its likely position on the outer layer of the plasma membrane. Although limited sequence information from public databases prevented us from characterizing YC54 and the other hypothetical proteins isolated from this study, this set of hypothetical proteins which are targeted mostly to organelles, may contain some novel secretory proteins involved in pine pollen germination. This also means that a better understanding of the mechanisms involved in pine pollen germination awaits more sequence information, comparative sequence analysis and ideally, functional characterization of these novel proteins. Unfortunately, the long generation time of the gymnosperms prevents their functional gene analysis.

Downregulated representative gene encoding secretory protein

A gene encoding a secretory protein in germinated pine pollen is homologous to CHALLAH-like2 like protein which reveals a conserved domain for epidermal patterning factor. Genes that contain this domain belong to a group of secreted signaling peptides which are cysteine rich which are considered to be required in the formation of disulfide bridges to hold the mature peptide in an active conformation (Pearce et al. 2001). The cysteine-rich peptide families of signaling peptides all have a predicted secretory signal sequence and six C-terminal conserved cysteine residues (Hara et al. 2009; Abrash et al. 2011). Members of this family are involved in regulating stomatal density and positioning (Hara et al. 2009; Abrash and Bergmann 2010; Shimada et al. 2011) or hypocotyl and inflorescence stem elongation (Abrash and Bergmann 2010; Abrash et al. 2011). Because of its varied expression patterns, it was suggested that there is an organ-specific role for these secreted peptides (Shimada et al. 2011; Uchida et al. 2012). In the current study, the gene encoding CHALLAH-like2 like protein is expressed in both mature and germinated pine pollen, and it exhibits the sequence characteristics of the family including the presence of six cysteine residues in the C-terminal region with conserved spacing. It has a significantly higher expression level in the mature pollen as compared with germinated pollen, and although its function in the development of the male gametophytes of loblolly pine is unknown, our results suggest that this type of protein may also be involved in plant development beyond those reported from sporophytic tissues.

Non-differentially expressed representative genes encoding secretory proteins

One of the interesting results from the current study is the identification of YST clones with significant homology to a putative human-secreted phosphoprotein, ALLW1950, which was identified through a large-scale effort to identify novel human-secreted proteins (Clark et al. 2003). In addition, ALLW1950 from the transcriptome analysis of semen from a low fertility bull (Lalancette et al. 2008) showed 70% identity to a hypothetical gene from a *Picea glauca* male strobilus cDNA library (GenBank No. DR546867). Thus, our study provides credence to the expression of this putative human-secreted phosphoprotein in the male reproductive structures of conifers. In the current study, ALLW1950 is one of the most abundant secretory proteins in our germinated loblolly pine pollen cDNA library. It has the same expression level in the mature and germinated pollen and so it is likely important in both stages of development. Sequence analysis using ProtFun 2.2, which is a server that produces ab initio predictions of protein function from the primary sequence (Jensen et al. 2002) indicates that the ALLW1950 in germinated pine pollen is involved in signaling, a function that corroborates with its putative identification as a secreted phosphoprotein.

Arabinogalactan proteins (AGPs) constitute a large family of cell wall proteoglycans that are localized on the plasma membrane or extracellular matrix, and the presence of GPI anchor is a characteristic of many AGPs (Costa et al. 2013). AGPs are associated with diverse functions including cell

expansion and division, plant defense, pollen tube growth, and pollen tube guidance into the ovule, but a definitive mode of action has not yet been established (Seifert and Roberts 2007). To this end, Costa et al. (2013) examined two *Arabidopsis* male gametophyte genes (*AGP6* and *AGP11*) and showed that their products are involved in signaling pathways, particularly in sending and recycling proteins to the extracellular matrix in relation to endocytosis-mediated plasma membrane remodeling during pollen tube growth. In the current study, an AGP predicted to be localized in the plasma membrane has appeared to possess a GPI anchor and is expressed in both mature and germinated loblolly pine pollen, but its higher expression level in the latter stage likely indicates where it exerts a greater effect. The exact function of the AGP in pine pollen germination is unknown, although our computational analysis categorizes it under signaling.

Thaumatin-like proteins (TLPs) share sequence similarity with thaumatin, an intensely sweet-tasting protein from the African berry *Thaumatococcus daniellii* (van der Wel and Loeve 1972; Velazhahan et al. 1999). These proteins are characterized by the presence of signal sequences that target them to the vacuole or extracellular matrix. Their inducible expressions under environmental stresses including pathogen/pest attack, drought, and cold have been the basis for their classification as pathogenesis-related protein 5 (PR5). However, TLPs have also been implicated in a wide range of developmental processes including fruit ripening, seed germination, and cell wall breakdown (Singh et al. 2017). TLPs have been identified from mature pollen of several conifers, which they have been considered as allergens (Midoro-Horiuti et al. 2000; Cortegano et al. 2004; Futamura et al. 2005). Their expression levels in pollen have been found to vary according to environmental conditions including air pollutants, UV-B, and oxidative stress. In the current study, a gene encoding a thaumatin-like protein is predicted to have ECS localization and conserved domain for thaumatin, a member of the glycoside hydrolase family 64 and thaumatin-like proteins (GH64-TLP) superfamily. It is expressed in the same level in mature and germinated loblolly pine pollen, which indicates its importance in both stages of development. GH64 are β-1,3-glucanases known to hydrolyze the β-1,3-glucans of fungal cell walls (Singh et al. 2017). Interestingly, the inner wall of mature pine pollen and of the pollen tube wall contain β-1,3-glucans and its hydrolysis by GH64-TLP, perhaps to facilitate pollen germination, needs to be examined.

YC24 is a gene encoding an unknown protein with the highest homology to an EST sequence from *Picea sitchensis* that has no known function. Its expression level is the same in both mature and germinated pine pollen, which suggests an important role in both stages of development. YC24 represents the most common sequence among the genes encoding unknown proteins as thus was chosen as the representative for further analysis. However, limited sequence information in the database prevented its further characterization, as well as the other unknown proteins isolated from this study except for YC9 (homologous to an uncharacterized protein from *Ipomoea nil*) which contains a DUF4409 superfamily conserved domain (domain of unknown function (DUF)). Overall, this set of unknown proteins which are predicted to target the plasma membrane or extracellular space also serves as candidates for the discovery of novel proteins involved in pine pollen germination.

Additional insights from sequence analysis of representative secretory proteins

The plasma membrane is composed of diverse types of proteins, many of which are transmembrane proteins that span the entire lipid bilayer. An example includes signal anchors which are single-pass transmembrane proteins which possess a membrane-spanning domain that targets the polypeptide to the ER membrane. Their amino acid composition is not very different from typical signal peptides, but characteristic features of signal anchors are the presence of positively charged residues on the N-terminal side followed by a hydrophobic segment of about 20 residues (Sakaguchi et al. 1990). Type I membrane proteins are the typically cleavable N-terminal signal peptides and have transmembrane regions close to the C-terminus, whereas type II membrane proteins have non-cleavable hydrophobic transmembrane region close to the N-terminus, which serves as a combined signal and anchor sequence (Goding 1996). Their difference is not only about whether cleavage sites are present or absent, but also on the length of the hydrophobic domain, which is longer in signal anchors since they span the entire lipid bilayer (Neilsen and Krogh 1998). In the current study, thaumatin-like protein and unknown protein YC24 are predicted to have signal peptides and transmembrane α-helixes in their N-terminus. Considering the overlap in the positions of these two elements, the N-terminal sequences from these two proteins seem to fall into the category of combined signal peptide/signal anchor function. However, only that of thaumatin-like protein is predicted to be non-cleavable and so is the most likely to represent a type II transmembrane protein. The presence of cleavage sites and high *D* values indicate that serine carboxypeptidase and unknown protein YC 24 represent type II transmembrane proteins.

A large proportion of membrane proteins also carry a C-terminal glycosylphosphatidylinositol (GPI) that mediates their association with the plasma membrane (Brown and Waneck 1992). GPI is a glycolipid that replaces the hydrophobic amino acids at the C-terminus of a protein during post-translational modification in the ER lumen to enable targeting to the outer leaflet of the plasma membrane (Seifert and Roberts 2007). In plants, GPI-anchored proteins have diverse functions including directional cell expansion, cellulose

deposition, cell wall attachment and remodeling, and immunity (Yadeta et al. 2013). Two of the representative secretory proteins in germinated pine pollen, which are initially predicted to be localized in the plasma membrane, i.e., an arabinogalactan protein and hypothetical protein YC54, are also predicted to have GPI anchors with corresponding anchor cleavage sites. Unlike other peripheral membrane proteins, GPI-anchored proteins are located almost exclusively on the extracellular side of the plasma membrane (Brown and Waneck 1992; Seifert and Roberts 2007). This result corroborates with the predicted plasma membrane localizations of the two genes encoding secreted proteins and provides insights on their possible position on the outer layer of the plasma membrane. This particular localization implies that these proteins can establish bonds with the cell wall components, and thus, may also serve as promising candidates to analyze plasma membrane and cell wall interactions (Liu et al. 2015). The presence of cleavage sites indicates that their GPI anchors can be cleaved and released into the extracellular space or external environment. The release of GPI anchor from the plasma membrane may trigger second messengers for signal transduction (Brown and Waneck 1992).

Genes encoding non-conventional secretory proteins

The conventional ER-Golgi pathway, which involves N-terminal signal peptides, accounts for how most of the secretory proteins with signal peptides enter the secretory pathway and become transported to the various subcellular compartments of the cell, except for those that are retained in the ER or Golgi apparatus. Another type of protein trafficking involves secretory proteins that do not have the canonical signal peptides and thus, they are exported to the apoplast through routes that are independent of the ER-Golgi pathway and referred to as non-conventional secretion (Rose and Lee 2010). Although not further analyzed in the current study, four YST clones from germinated pine pollen were found to be devoid of signal peptides and so they may represent non-conventional secreted proteins. Alternatively, they may represent false positives since some truncated proteins are considered to exhibit abnormally exposed N-terminal hydrophobic or highly basic regions that can cause secretion (Lee et al. 2014). Another form of non-conventional protein trafficking involves post-translational transport where proteins are translated in the cytosol and then translocated to specific organelles by means of N-terminal transit peptides in the case of chloroplast and mitochondria or nuclear localization signals in the case of nuclei (Savojardo et al. 2015). In the current study, three genes encoding hypothetical proteins with *D* values below the confidence level for presence of signal peptides were predicted to be localized in the chloroplast, mitochondria, and nuclei and thus may represent non-conventional secretion. Many secretory proteins without signal peptides have been isolated from semi–in vitro germinated tobacco pollen and they have been considered to be involved in guiding pollen tubes towards the ovule (Hafidh et al. 2016).

Conclusions

This study presents known, unknown, and hypothetical genes that encode secretory proteins associated with pine pollen germination based on sequence homology, subcellular localizations, and other sequence characteristics derived from computational analysis. The encoded proteins are related to cell wall degradation, biosynthesis and remodeling, stress and defense responses, signaling, and protein processing, and thus, expand our understanding of the molecular players involved in pollen germination. The study also provides information on the spatial and temporal expression patterns of representative genes, which lend support to the distinct expression profiles of mature (ungerminated) versus germinated pollen. Considering that the study was conducted under in vitro conditions, it is possible that the expression profile and pattern reported here would differ from what occurs under in vivo conditions and so a comparative analysis needs to be done. Another step for the study is functional gene characterization, but it is currently not feasible in pines and other gymnosperms because of their long generation time and lack of efficient genetic transformation system. Therefore, the development of a pollen transformation system to facilitate characterization of pollen expressed genes is necessary to help advance our understanding of the mechanism underlying pollen germination in pines and other gymnosperms.

Acknowledgments The technical and invaluable contributions of Drs. Sang-Jik Lee and Jocelyn Rose (Cornell University), particularly on the use and implementation of the YST approach, are gratefully acknowledged.

Data archiving Full-length cDNA sequences of the 10 representative proteins have been submitted to GenBank and their accession numbers are presented in Table 3.

References

Abrash EB, Bergmann DC (2010) Regional specification of stomatal production by the putative ligand CHALLAH. Development 137(3):447–455. https://doi.org/10.1242/dev.040931

Abrash EB, Davies KA, Bergmann DC (2011) Generation of signaling specificity in Arabidopsis by spatially restricted buffering of ligand–receptor interactions. Plant Cell 23(8):2864–2879. https://doi.org/10.1105/tpc.111.086637

Alché JD, M'rani-Alaoui M, Castro AJ, Rodríguez-García MI (2004) Ole e 1, the major allergen from olive (*Olea europaea* L) pollen, increases its expression and is released to the culture medium during

in vitro germination. Plant Cell Physiol 45:1149–1157. https://doi.org/10.1093/pcp/pch127

Andrews SC (2010) The ferritin-like superfamily: evolution of the biological iron storeman from a rubrerythrin-like ancestor. Biochim Biophys Acta 1800:691–705. https://doi.org/10.1016/j.bbagen.2010.05.010

Anzar A, Chen NWG, Thomine S, Dellagi A (2015) Immunity to plant pathogens and iron homeostasis. Plant Sci 240:90–97. https://doi.org/10.1016/j.plantsci.2015.08.022

Bienert MD, Delannoy SE, Navarre M, Boutry M (2012) NtSCP1 from tobacco is an extracellular serine carboxypeptidase III that has an impact on cell elongation. Plant Physiol 158:1220–1122. https://doi.org/10.1104/pp.111.192088

Breddam K (1986) Serine carboxypeptidases: a review. Carlsberg Res Commun 51:83–128. https://doi.org/10.1007/BF02907561

Brown D, Waneck GL (1992) Glycosyl-phosphatidylinositol-anchored membrane proteins. J Am Soc Nephrol 3:895–906 https://jasn.asnjournals.org/content/3/4/895

Chen Y, Chen T, Shen S, Zheng M, Guo Y, Lin J, Baluska F, Samaj J (2006) Differential display proteomic analysis of *Picea meyeri* pollen germination and pollen tube growth after inhibition of actin polymerization by latrunculin B. Plant J 47:174–195. https://doi.org/10.1111/j.1365-313X.2006.02783.x

Chen Y, Liu P, Hoehenwarter W, Lin J (2012) Proteomic and phosphoproteomic analysis of *Picea wilsonii* pollen development under nutrient limitation. J Proteome Res 11:4180–4190. https://doi.org/10.1021/pr300295m

Clark HF, Gurney AL, Abaya E, Baker K, Baldwin D, Brush J, Chen J, Chow B, Chui C, Crowley C (2003) The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. Genome Res 13:2265–2270. https://doi.org/10.1101/gr.1293003

Cortegano I, Civantos E, Aceituno E, del Moral A, Lopez E, Lombardero M, del Pozo V, Lahoz C (2004) Cloning and expression of a major allergen from Cupressus arizonica pollen, Cup a 3, a PR-5 protein expressed under polluted environment. Allergy 59:485–490. https://doi.org/10.1046/j.1398-9995.2003.00363.x

Cosgrove DJ (2015) Plant expansins: diversity and interactions with plant cell walls. Curr Op Plant Biol 25:162–172. https://doi.org/10.1016/j.pbi.2015.05.014

Cosgrove DJ, Bedinger PA, Durachko DM (1997) Group I allergens of grass pollen as cell wall loosening agents. Proc Natl Acad Sci U S A 94:6559–6564. https://doi.org/10.1073/pnas.94.12.6559

Cosgrove DJ, Li LC, Cho HT, Hoffmann-Benning S, Moore RC, Blecker D (2002) The growing world of expansins. Plant Cell Physiol 43:1236–1244. https://doi.org/10.1093/pcp/pcf180

Costa M, Nobre MS, Becker JD, Masiero S, Amorin MI, Pereira LG, Coimbra S (2013) Expression-based and co-localization detection of arabinogalactan protein 6 and arabinogalactan protein 11 interactors in Arabidopsis pollen and pollen tubes. BMC Plant Biol 13:7. https://doi.org/10.1186/1471-2229-13-7

Crane PR, Leslie AB (2014) Major events in the evolution of land plants. In: Losos JB, Baum DA, Futuyma DJ, Hoekstra HE, Lenski RE, Moore AJ, Peichel CL, Schluter D, Whitlock MJ (eds) The Princeton guide to evolution. Princeton University Press, Princeton, pp 143–151. http://www.jstor.org/stable/j.ctt4cgc5m.22

Dai S, Chen T, Chong K, Xue Y, Liu S, Wang T (2007) Proteomics identification of differentially expressed proteins associated with pollen germination and tube growth reveals characteristics of germinated *Oryza sativa* pollen. Mol Cell Proteomics 6:207–230. https://doi.org/10.1074/mcp.M600146-MCP200

Daystar J, Gonzalez R, Reeb CW, Venditti RA, Treasure T, Abt R, Kelley S (2014) Economics, environmental impacts, and supply chain analysis of cellulosic biomass for biofuels in southern US: pine, eucalyptus, unmanaged hardwoods, forest residues, switchgrass and sweet sorghum. BioResouces 9:393–444. https://doi.org/10.15376/biores.9.1.393-444

Fernando DD (2005) Characterization of pollen tube development in *Pinus strobus* (eastern white pine) through proteomic analysis of differentially expressed proteins. Proteomics 5:4917–4926. https://doi.org/10.1002/pmic.200500009

Fernando DD (2014) The pine reproductive process in temperate and tropical regions. New For 45:333–352. https://doi.org/10.1007/s11056-013-9403-7

Fernando DD, Owens JN, von Aderkas P, Takaso T (1997) In vitro pollen tube growth and penetration of female gametophyte in Douglas fir (*Pseudotsuga menziesii*). Sexual Plant Reprod 10:209–216. https://doi.org/10.1007/s004970050089

Fernando DD, Quinn CR, Brenner ED, Owens JN (2010) Male gametophyte development and evolution in extant gymnosperms. Int J Plant Dev Biol 4:47–63. http://www.academia.edu/25338811/Male_Gametophyte_Development_and_Evolution_in_Extant_Gymnosperms

Fraser CM, Rider LW, Chapple C (2005) An expression and bioinformatics analysis of the Arabidopsis serine carboxypeptidase-like gene family. Plant Physiol 138:1136–1148. http://www.plantphysiol.org/content/138/2/1136

Futamura N, Tani N, Tsumura Y, Nakajima N, Sakaguchi M, Shinohara K (2005) Characterization of genes for novel thaumatin-like proteins in Cryptomeria japonica. Tree Physiol 26:51–62. https://doi.org/10.1093/treephys/26.1.51

Ge W, Song Y, Zhang C, Zhang Y, Burlingame AL, Guo Y (2011) Proteomic analyses of apoplastic proteins from germinating *Arabidopsis thaliana* pollen. Biochim Biophys Acta 1814:1964–1973. https://doi.org/10.1016/j.bbapap.2011.07.013

Goding JW (1996) Analysis of antigens recognized by monoclonal antibodies. In: Monoclonal antibodies (3rd Ed): principles and practice. Elsevier Ltd, pp 234–326. https://doi.org/10.1016/B978-012287023-1/50058-4

Hafidh S, Potesil D, Fila J, Capkova Z, Zdrahal Z, Honys D (2016) Quantitative proteomics of the tobacco pollen tube secretome identifies novel pollen tube guidance proteins important for fertilization. Genome Biol 17:81. https://doi.org/10.1186/s13059-016-0928-x

Han B, Chen SX, Dai SJ, Yang N, Want T (2010) Isobaric tags for relative and absolute quantification- based comparative proteomics reveals the features of plasma membrane-associated proteomes of pollen grains and pollen tubes from *Lilium davidii*. J Integr Plant Biol 52:1043–1058. https://doi.org/10.1111/j.1744-7909.2010.00996.x

Hanson DD, Hamilton DS, Travis JL, Bashe DM, Mascarenhas JP (1998) Characterization of a pollen- specific cDNA clone from *Zea mays* and its expression. Plant Cell 1:173–179. https://doi.org/10.1105/tpc.1.2.173

Hara K, Yokoo T, Kajita R, Onishi T, Yahata S, Peterson KM, Torii KU, Kakimoto T (2009) Epidermal cell density is autoregulated via a secretory peptide, EPIDERMAL PATTERNING FACTOR 2 in Arabidopsis leaves. Plant Cell Physiol 50(6):1019–1031. https://doi.org/10.1093/pcp/pcp068

Hu B, Liu B, Liu L, Liu C, Xu L, Ruan Y (2014) Epigenetic control of pollen Ole e 1 allergen and extensin family gene expression in Arabidopsis thaliana. Acta Physiol Plant 36:2203–2209. https://doi.org/10.1007/s11738-014-1597-6

Jensen L, Gupta R, Blom N, Devos D, Tamames J, Kesmir C, Nielsen H, Stærfeldt H, Rapacki K, Workman C, Andersen C, Knudsen S, Krogh A, Valencia A, Brunak S (2002) Prediction of human protein function from post-translational modifications and localization features. J Mol Biol 319:1257–1265. https://doi.org/10.1016/S0022-2836(02)00379-0

Jimenez-Lopez JC, Rodriguez-Garcia MI, Alche JD (2011) Systematic and phylogenetic analysis of the Ole e 1 pollen protein family members in plants. In: Yang N-S (ed) Systems and computational biology

– bioinformatics and computational modeling. InTech, pp 245–260. https://doi.org/10.5772/19175

Kapu NUS, Cosgrove DJ (2010) Changes in growth and cell wall extensibility of maize silks following pollination. J Expt Bot 61:4097–4107. https://doi.org/10.1093/jxb/erq225

Klee EW, Ellis LBM (2005) Evaluating eukaryotic secreted protein prediction. BMC Bioinformatics 6:256. https://doi.org/10.1186/1471-2105-6-256

Lalancette C, Thibault C, Bachand I, Caron N, Bissonnette N (2008) Transcriptome analysis of bull semen with extreme nonreturn rate: use of suppression-subtractive hybridization to identify functional markers for fertility. Biol Reprod 78:618–635. https://doi.org/10.1095/biolreprod.106.059030

Lauzurica P, Gurbindo C, Maruri N, Galocha B, Diaz R, Gonzalez J, Garcia R, Lahoz C (1988) Olive (*Olea europea*) pollen allergens. 1. Immunochemical characterization by immunoblotting, crie and immunodetection by a monoclonal-antibody. Mol Immunol 25:329–335. https://doi.org/10.1016/0161-5890(88)90027-2

Lee SJ, Kelley BS, Damasceno CM, St John B, Kim BS, Kim BD, Rose JKC (2006) A functional screen to characterize the secretomes of eukaryotic pathogens and their hosts in planta. Mol Plant-Microbe Interac 19:1368–1377. https://doi.org/10.1094/MPMI-19-1368

Lee JM, Lee SJ, Rose JKC, Yeam I, Kim BD (2014) Mining secreted proteins that function in pepper fruit development and ripening using a yeast secretion trap (YST). Biochem Biophys Res Commun 446:882–888. https://doi.org/10.1016/j.bbrc.2014.03.022

Li J, Lease KA, Tax FE, Walker JC (2001) BRS1, a serine carboxypeptidase, regulates BRI1 signaling in Arabidopsis thaliana. PNAS 98:5916–5921. https://doi.org/10.1073/pnas.091065998

Link TI, Voegele RT (2008) Secreted proteins of *Uromyces fabae*: similarities and stage specificity. Mol Plant Pathol 9:59–66. https://doi.org/10.1111/j.1364-3703.2007.00448.x

Liu Z, Persson S, Sanchez-Rodriguez C (2015) At the border: the plasma membrane-cell wall continuum. J Exp Bot 66:1553–1563. https://doi.org/10.1093/jxb/erv019

Lora J, Hormaza JI, Herrero M (2015) The diversity of the pollen tube pathway in plants: toward an increasing control by the sporophyte. Front Plant Sci 7:107. https://doi.org/10.3389/fpls.2016.00107

Martelli PL, Fariselli P, Casadio R (2003) An ENSEMBLE machine learning approach for the prediction of all-alpha membrane proteins. Bioinformatics 19:205–211. https://doi.org/10.1093/bioinformatics/btg1027

Midoro-Horiuti T, Goldblum RM, Kurosky A, Wood TG, Brooks EG (2000) Variable expression of pathogenesis-related protein allergen in mountain cedar (*Juniperus ashei*) pollen. J Immunol 164:2188–2192. https://doi.org/10.4049/jimmunol.164.4.2188

Neilsen H, Krogh A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. In J Glasgow et al. (eds), Proc Sixth Int Conf Intell Syst Mol Bio 122–130. http://www.aaai.org/Library/ISMB/1998/ismb98-015.php

Owens JN, Takaso T, Runions CJ (1998) Pollination in conifers. Trends Plant Sci 3:479–485. https://doi.org/10.1016/S1360-1385(98)01337-5

Park YB, Cosgrove DJ (2015) Xyloglucan and its interactions with other components of the growing cell wall. Plant Cell Physiol 56:180–194. https://doi.org/10.1093/pcp/pcu204

Pearce G, Moura DS, Stratmann J, Ryan CA (2001) Production of multiple plant hormones from a single polyprotein precursor. Nature 411(6839):817–820. https://doi.org/10.1038/35081107

Pierleoni A, Martelli P, Casadio R (2008) PredGPI: a GPI-anchor predictor. BMC Bioinformatics 9:392. https://doi.org/10.1186/1471-2105-9-392

Rose JKC, Lee SK (2010) Straying off the highway: trafficking of secreted plant proteins and complexity in the plant cell wall proteome. Plant Physiol 153:433–436. https://doi.org/10.1104/pp.110.154872

Rutley N, Twell D (2015) A decade of pollen transcriptomics. Plant Reprod 28:73–89. https://doi.org/10.1007/s00497-015-0261-7

Sakaguchi M, Tomiyoshi R, Kuroiwa T, Mihara K, Omra T (1990) Functions of signal and signal-anchor sequences are determined by the balance between the hydrophobic segment and the N-terminal charge. Proc Natl Acad Sci 89:16–19. https://doi.org/10.1073/pnas.89.1.16

Salmi ML, Bushart TJ, Stout SC, Roux SJ (2005) Profile and analysis of gene expression changes during early development in germinating spores of *Ceratopteris richardii*. Plant Physiol 138:1734–1745. https://doi.org/10.1104/pp.105.062851

Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6:242. https://doi.org/10.1186/gb-2005-6-12-242

Sampedro J, Guttman M, Li LC, Cosgrove DJ (2015) Evolutionary divergence of beta-expansin structure and function in grasses parallels emergence of distinctive primary cell wall traits. Plant J 81:108–120. https://doi.org/10.1111/tpj.12715

Savojardo C, Martelli PL, Fariselli P, Casadio R (2015) TPpred3 detects and discriminates mitochondrial and chloroplastic targeting peptides in eukaryotic proteins. Bioinformatics 31(20):3269–3275. https://doi.org/10.1093/bioinformatics/btv367

Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R (2018) BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic Acid Res 46:W459–W466. https://doi.org/10.1093/nar/gky320

Schaller A (2004) A cut above the rest: the regulatory function of plant proteases. Planta 220:183–197. https://doi.org/10.1007/s00425-004-1407-2

Seifert GJ, Roberts K (2007) The biology of arabinogalactan proteins. Annu Rev Plant Biol 58:137–161. https://doi.org/10.1146/annurev.arplant.58.032806.103801

Shimada T, Sugano SS, Hara-Nishimura I (2011) Positive and negative peptide signals control stomatal density. Cell Mol Life Sci 68(12):2081–2088. https://doi.org/10.1007/s00018-011-0685-7

Singh S, Tripathi RK, Lemaux PG, Buchanan BB, Singh J (2017) Redox-dependent interaction between thaumatin-like protein and B-glucan influences malting quality of barley. Proc Natl Acad Sci 114:7725–7730. https://doi.org/10.1073/pnas.1701824114

Stratford S, Barne W, Hohorst DL, Sagert JG, Cotter R, Golubiewski A (2001) A leucine-rich repeat region is conserved in pollen extensin-like (Pex) proteins in monocots and dicots. Plant Mol Biol 46:43–56. https://doi.org/10.1023/A:1010659425399

Tabuchi A, Li LC, Costgrove DJ (2011) Matrix solubilization and cell wall weakening by beta-expansin (group-1 allergen) from maize pollen. Plant J 68:546–559. https://doi.org/10.1111/j.1365-313X.2011.04705.x

Uchida N, Lee JS, Horst RJ, Lai HH, Kajita R, Kakimoto T, Tasaka M, Torii KU (2012) Regulation of inflorescence architecture by inter tissue layer ligand–receptor communication between endodermis and phloem. Proc Natl Acad Sci 109(16):6337–6342. https://doi.org/10.1073/pnas.1117537109

van der Wel H, Loeve K (1972) Isolation and characterization of thaumatin I and II, the sweet-tasting proteins from *Thaumatococcus daniellii* Benth. Eur J Biochem 31:221–225. https://doi.org/10.1111/j.1432-1033.1972.tb02522.x

Velazhahan R, Datta SK, Muthukrishnan S (1999) The PR-5 family: thaumatin-like proteins in plants. In: Datta SK, Muthukrishnan S (eds) Pathogenesis related proteins in plants. CRC Press, Boca Raton, pp 107–129

Villalba M, Batanero E, Lopez-Otin C, Sanchez LM, Monsalve RI, Gonzalez de la Pena MA, Lahoz C, Rodriguez R (1993) The amino acid sequence of Ole e I, the major allergen from olive tree (*Olea europaea*) pollen. Eur J Biochem 216:863–869. https://doi.org/10.1111/j.1432-1033.1993.tb18208.x

Wang T, Park YB, Caporini MA, Rosay M, Zhong L, Cosgrove DJ, Homg M (2013) Sensitivity- enhanced solid-state NMR detection

of expansin's target in plant cell walls. Proc Natl Acad Sci U S A 110:16444–16449. https://doi.org/10.1073/pnas.1316290110

Williams JH (2012) Pollen tube growth rates and the diversification of flowering plant reproductive cycles. Int J Plant Sci 173:649–661. https://doi.org/10.1086/665822

Yadeta KA, Elmore JM, Coaker G (2013) Advancements in the analysis of the Arabidopsis plasma membrane proteome. Front Plant Sci 4(86). https://doi.org/10.3389/fpls.2013.00086

Yamane H, Lee SJ, Kim BD, Tao R, Rose JKC (2005) A coupled yeast signal sequence trap and transient plant expression strategy to identify genes encoding secreted proteins from peach pistils. J Expt Bot 56:2229–2238. https://doi.org/10.1093/jxb/eri222

Yang N, Wang T (2017) Comparative proteomic analysis reveals a dynamic plasma membrane protein map and the membrane landscape of receptor-like kinases and transporters important for pollen tube growth and interaction with pistils in rice. BMC Plant Biol 17(2):2. https://doi.org/10.1186/s12870-016-0961-7

Yang P, Li X, Wang X, Chen H, Chen F, Shen S (2007) Proteomic analysis of rice (*Oryza sativa*) seeds during germination. Proteomics 7:3358–3368. https://doi.org/10.1002/pmic.200700207

Yeom SI, Baek HK, Oh SK, Kang WH, Lee SJ, Seo E, Rose JKC, Kim BD, Choi D (2011) Use of a secretion trap screen in pepper following *Phytophthora capsici* infection reveals novel functions of secreted plant proteins in modulating cell death. Mol Plant Microbe Interac 24:671–684. https://doi.org/10.1094/MPMI-08-10-0183

Zha HG, Liu T, Zhou JJ, Sun H (2013) MS-desi, a desiccation-related protein in the floral nectar of the evergreen velvet bean (*Mucuna sempervirens*): molecular identification and characterization. Planta 238:77–89. https://doi.org/10.1007/s00425-013-1876-2

Zhang Z, Hu M, Feng X, Gong A, Cheng L, Yuan H (2017) Proteomes and phosphoproteomes of anther and pollen: availability and progress. Proteomics 17:1600458. https://doi.org/10.1002/pmic.201600458

Zhou A, Li J (2005) Arabidopsis BRS1 is a secreted and active serine carboxypeptidase. J Biol Chem 280:35554–35561. https://doi.org/10.1074/jbc.M503299200

Zhu Y, Yang J, Showalter AM (2011) AtAGP18, a lysine-rich arabinogalactan protein in *Arabidopsis thaliana*, functions in plant growth and development as a putative co-receptor for signal transduction. Plant Signal Behav 6:855–857. https://doi.org/10.4161/psb.6.6.15204

Zou J, Song L, Zhang W, Wang Y, Ruan S, Wu WÄ (2009) Comparative proteomic analysis of Arabidopsis mature pollen and germinated pollen. J Integr Plant Biol 51:438–455. https://doi.org/10.1111/j.1744-7909.2009.00823