Research

New Phytologist

The *Physcomitrium patens* egg cell expresses several distinct epigenetic components and utilizes homologues of *BONOBO* genes for cell specification

Victoria Sanchez-Vera[1]*, Katarina Landberg[1]*, Mauricio Lopez-Obando[1]*, Mattias Thelander[1], Ulf Lagercrantz[2], Rafael Muñoz-Viana[1], Anja Schmidt[3], Ueli Grossniklaus[3] and Eva Sundberg[1]

[1]Department of Plant Biology, The Linnean Centre of Plant Biology in Uppsala, Swedish University of Agricultural Sciences, PO Box 7080, Uppsala SE-75007, Sweden; [2]Department of Ecology and Genetics, Uppsala University, Norbyvägen 18 D, Uppsala SE-752 36, Sweden; [3]Department of Plant and Microbial Biology & Zurich-Basel Plant Science Center, University of Zurich, Zollikerstrasse 107, Zurich CH-8008, Switzerland

Author for correspondence:
Eva Sundberg
Email: Eva.Sundberg@slu.se

Received: *11 August 2021*
Accepted: *14 December 2021*

New Phytologist (2022) **233:** 2614–2628
doi: 10.1111/nph.17938

Key words: *BONOBO*, bryophytes, egg cell, gametes, moss, *Physcomitrium patens*, transcriptome.

Summary

- Although land plant germ cells have received much attention, knowledge about their specification is still limited. We thus identified transcripts enriched in egg cells of the bryophyte model species *Physcomitrium patens*, compared the results with angiosperm egg cells, and selected important candidate genes for functional analysis.
- We used laser-assisted microdissection to perform a cell-type-specific transcriptome analysis on egg cells for comparison with available expression profiles of vegetative tissues and male reproductive organs. We made reporter lines and knockout mutants of the two *BONOBO* (*PbBNB*) genes and studied their role in reproduction.
- We observed an overlap in gene activity between bryophyte and angiosperm egg cells, but also clear differences. Strikingly, several processes that are male-germline specific in Arabidopsis are active in the *P. patens* egg cell. Among those were the moss *PbBNB* genes, which control proliferation and identity of both female and male germlines.
- Pathways shared between male and female germlines were most likely present in the common ancestors of land plants, besides sex-specifying factors. A set of genes may also be involved in the switches between the diploid and haploid moss generations. Nonangiosperm gene networks also contribute to the specification of the *P. patens* egg cell.

Introduction

Whereas animal gametes are directly produced by meiosis, land plant gametes form by mitotic divisions and reprogramming of somatic cells in the haploid gametophyte generation. In bryophytes, a sister phylum to vascular plants, the gametophytic phase dominates the life cycle, and gametes are produced by haploid male (antheridia) and female (archegonia) reproductive organs. In the vascular plant angiosperm lineage, the gametophytic phase has become dramatically reduced and consists of only a few cells, with both female (embryo sac) and male gametophytes (pollen) producing a pair of gametes (see reviews Hater *et al.*, 2020; Hafidh & Honys, 2021). Although land plant germ cells have received much attention, knowledge about the genetic components that specify plant germ cell identity is limited.

In the angiosperm model *Arabidopsis thaliana*, the MYB transcription factor (TF) DUO POLLEN1 (DUO1) is required for cell cycle progression of sperm precursor cells and the promotion of sperm cell differentiation (Brownfield *et al.*, 2009; Borg *et al.*, 2011). A complex network including *DAZ1/2*, *GCS1/HAP2*, and *GEX2* has been suggested to act downstream of DUO1 (see review Hackenberg & Twell, 2019). Interestingly, transcriptome analysis of male antheridia of the bryophyte *Marchantia polymorpha* revealed the enrichment of transcripts homologous to *DUO1*, *DAZ2*, *GCS1/HAP2*, and *GEX2* (Higo *et al.*, 2016), suggesting a conservation of their function in male germ cell development throughout land plant evolution. This is further supported by the fact that Mp*DUO1* is specifically expressed in spermatid mother cells and spermatids, controls Mp*DAZ1*, and is necessary for the production of fertile sperm cells (Higo *et al.*, 2018).

Recently, the first candidates for egg cell specification and fate determination were identified. The RKD TF-clade of the non-metazoan RWP-RK family appears to have a conserved function in germ cell fate regulation throughout land plant evolution. In Arabidopsis, two *RKD* genes are preferentially expressed in the egg cell, and ectopic *AtRKD2* expression induces egg cell-like transcription (Kőszegi *et al.*, 2011). In addition, the single *M. polymorpha RKD* gene is expressed in developing egg and sperm precursors and maintained in the egg until fertilization

*Joint first authors.

(Koi *et al.*, 2016; Rövekamp *et al.*, 2016). The egg cells of lines with reduced Mp*RKD* expression failed to fully differentiate and maintain quiescence, resulting in ectopic cell divisions prior to fertilization, suggesting that Mp*RKD* may also regulate the gametophyte-to-sporophyte transition by suppressing parthenogenesis (Koi *et al.*, 2016; Rövekamp *et al.*, 2016). Mp*RKD* also affects cell division of the sperm precursors in liverwort antheridia (Koi *et al.*, 2016).

To further our understanding of plant egg cell specification, we used laser-assisted microdissection (LAM) for cell-type-specific transcriptome analysis of unfertilized egg cells of another bryophyte model species, the moss *Physcomitrium patens*. Transcription of genes involved in DNA and RNA-processing were overrepresented in the moss egg, as was previously found in Arabidopsis (Wuest *et al.*, 2010). Apart from genes shown to be important in egg cells of other plant species, we also identified expression of genes related to male germ cell differentiation in Arabidopsis. Among those were two homologues of the class VIIIa basic helix–loop–helix (bHLH) TFs BONOBO (BNB) (Yamaoka *et al.*, 2018). Knockout of *PpBNB1* and *PpBNB2* did not affect reproductive organ initiation, as loss of the *M. polymorpha* homologue does (Yamaoka *et al.*, 2018), but disturbs proliferation and identity of both the female and male germlines, suggesting that the moss BNBs play important roles in germ cell specification.

Materials and Methods

Plant material and growth conditions

Plant material, growth conditions, harvesting, and methods for crosses are described in Supporting Information Methods S1.

Sample preparation and RNA extraction for RNA-sequencing

Samples were prepared as described previously (Wuest *et al.*, 2010; Schmidt *et al.*, 2011; Florez-Rueda *et al.*, 2016). Shoots with reproductive organs were fixed in 90% ethanol (EtOH)–10% acetic acid on ice with 2 × 15 min of vacuum followed by an overnight incubation at 4°C. Samples were transferred to 90% EtOH and stored until embedding. Embedding was performed using the automatic system Leica ASP200 (Leica Microsystems, Heerbrugg, Switzerland): 1 h 70% EtOH, 3 × 1 h 90% EtOH, 3 × 1 h 100% EtOH, 2 × 1 h 100% xylol, 1 × 15 min 100% xylol, all at room temperature, followed by 2 × 1 h and 1 × 3 h Paraplast X-tra at 56°C. Samples were mounted on paraffin blocks and stored at 4°C before they were sectioned on a Leica microtome RM2145 (Leica Microsystems) to 6 µm thickness and mounted onto a drop of RNAase-free MilliQ water on RNAase-free polyethylene naphthalate membrane metal frame slides (Molecular Machines & Industries (MMI), Glattbrugg, Switzerland). Slides were dried overnight at 42°C on a heating table before they were deparaffinized in 100% xylol for 2 × 10 min at room temperature and air-dried for 10 min. Then, single cells were dissected using an SL µCut instrument (MMI). Egg cells at stage 8, egg cells at stage 9, cells forming the cavity wall, and cells from the gametophore shoot were harvested separately (on average, 250 cells per cell type; Fig. S1a–e). Microdissected material was stored at −80°C. RNA was extracted using the Picopure RNA extraction kit from Arcturus (Thermo Fisher Scientific, Waltham, MA, USA) and amplified using the SMARTer Ultra Low Input RNA kit for Sequencing v.3 (Clontech Laboratories, Mountain View, CA, USA) using polyT primers and following the manufacturer's protocol applying 16 amplification cycles.

Library preparation, sequencing, and RNA-sequencing analysis

Libraries were created with the ThruPLEX® DNA-seq Kit (Rubicon Genomics, Ann Arbor, MI, USA) and 50 bp single-end reads (high output run SR50) were generated using a HiSeq2500 machine using HiSeq SBS Kit v4 chemistry (Illumina, San Diego, CA, USA). FastQC v.0.10.1 (Andrews, 2010) was used to check read quality. Reads were filtered through fastp (Chen *et al.*, 2018) and bbduk (Bushnell, 2014) to remove low-quality sequences, adapters, runs of polyA, polyT, and sequences matching ribosomal RNA (rRNA; using bbduk and rRNA sequences from www.arb-silva.de). Read counts before and after filtering are shown in Table S1. Filtered reads were pseudoaligned to *P. patens* complementary DNA sequences v.3.3 (https://peatmoss.online.uni-marburg.de/downloads/Annotations/P.patens_v3.3/P.patens_v3.3_cDNA.fasta) and quantified per gene using Kallisto (Bray *et al.*, 2016). The resulting count table is available through Gene Expression Omnibus (Series XX), and the corresponding reads per kilobase of transcript, per million mapped reads (RPKM) are presented in Table S1. We used the zFPKM transformation (Hart *et al.*, 2013; Ammar & Thompson, 2021) and a cutoff of at zFPKM ≥ −3 as the threshold for biologically relevant expression as opposed to noise.

Identification of differentially expressed genes in egg compared with other tissues was conducted using the glmQLFit function in edgeR (Robinson *et al.*, 2010). Processed read count data from additional tissues (Table S5; Perroud *et al.*, 2018) were kindly supplied by Professor Stefan Rensing. Genes with at least 0.8 counts per million in at least two samples were kept for statistical analysis. To illustrate consistency of tissue replicates, we built a multidimensional scaling (MDS) plot for the top 1000 genes with large fold-changes using function plotMDS from the edgeR package (Fig. S1f). Normalization was performed by trimmed mean of *M* values (TMM) (Robinson & Oshlack, 2010), and a model contrasting egg samples (egg stage 8, egg stage 9) to all other samples (samples from green sporophytes excluded) was used to find genes differentially expressed in egg cells. The *P*-values were adjusted for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) method (Benjamini & Hochberg, 1995). The FDR was set to 0.01 if not stated otherwise. A volcano plot was produced by plotting $\log_2$(fold change) against $-\log_{10}$(FDR) (Fig. S1g).

Patterns of gene expression among differentially expressed genes were identified through clustering of genes with the function hclust

in R. RPKM values for antheridia of the Reute ecotype were added from Meyberg *et al.* (2020). All samples, including those of green sporophytes, were normalized using the function scale, Euclidean distance calculated with the function dist, and clustering performed using the method complete in hclust.

For comparison with the egg transcriptome of *A. thaliana*, we downloaded normalized expression values of egg, central, and synergid cells from Song *et al.* (2020).

Overrepresentation of biological functions among differentially expressed genes was performed using the TOPGO package (Alexa & Rahnenfuhrer, 2021) in R. Mapping of genes to gene ontology (GO) terms was taken from the annotation file 'P.patens_lookup_table_20181029.xlsx' obtained from https://peatmoss.online.uni-marburg.de/ppatens_db/downloads.php. Tests were performed with TOPGO using Fisher's exact test statistic and the weight01 algorithm. Separate tests were performed for (1) all genes with a fold-change > 2 comparing egg cells with all other samples (6622 genes); (2) the subset of those genes also expressed in *A. thaliana* egg cells (891 genes); (3) the subset not expressed in *A. thaliana* egg cells (891 genes); and (4) those in each of the four clusters identified in the cluster analysis (1745, 630, 340, and 95 genes).

Generation and verification of transgenic lines

The generation and verification of new transgenic lines is described in Methods S2. An overview of reporter constructs is given in Fig. S2, characteristics of crispr RNAs (crRNAs) in guide-RNA-expressing plasmids in Table S2 and of CRISPR-CAS9 gene edited mutants in Table S3. For lists of primers, see Table S4.

Reverse transcription quantitative PCR to determine *PpBNB* transcript abundance

Methods for reverse transcription quantitative PCR (RT-qPCR) to determine *PpBNB* transcript abundance is described in Methods S3. For primer sequences, see Table S4.

Light microscopy and *Ppbnb* mutant phenotyping

Methods for tissue harvesting and light microscopy are described in Methods S4.

Confocal microscopy and *PpBNB* reporter gene analysis

Methods for tissue harvesting and confocal imaging are described in Methods S5.

Results

Transcriptome analysis

Reproductive organs develop at the *P. patens* gametophore shoot apex upon inductive conditions. The egg cell forms by an asymmetric mitotic division of a pre-egg cell (Landberg *et al.*, 2013; see later, Fig. 3c). At archegonium stage 8, the egg is specified, has detached from surrounding cells, but is not yet accessible for fertilization. At stage 9, the archegonium has opened, connecting the egg cavity to the exterior by a canal formed by degradation of a central cell file (Landberg *et al.*, 2013; see later, Fig. 3c). As our strain of the moss ecotype Gransden is male sterile (Landberg *et al.*, 2020; Meyberg *et al.*, 2020), we could harvest egg cells from stages 8 and 9 using single-cell LAM to study the transcriptome of the unfertilized egg (Fig. S1). As internal controls, LAM-captured cells from the shoot apex and the cavity wall were used (Fig. S1a,e). Owing to the detachment of the egg from the surrounding cells, cross-contamination between samples was negligible.

After Illumina short-read sequencing and filtering of the two egg samples and that of the cell wall, between 11 million and 22 million reads remained for mapping to the 32 458 predicted *P. patens* genes (genome v.3.3; https://peatmoss.online.uni-marburg.de/downloads/Annotations/P.patens_v3.3/P.patens_v3.3_cDNA.fasta). Alignment of *c.* 70% was achieved for all samples (Table S1). To identify genes preferentially expressed in egg cells, a statistical analysis was performed that included previously published RNA-sequencing (RNA-seq) data from other *P. patens* tissues (Perroud *et al.*, 2018; Table S5). Preliminary analyses indicated a considerable overlap between egg samples and those from the green sporophyte. Thus, we excluded this early sporophyte stage from the analysis to avoid overlooking important egg genes. To identify genes preferentially expressed in egg cells, we considered stage 8 and stage 9 egg cells as replicates and contrasted those with replicates of protonema, gametophores, leaflets, brown sporophytes, spores, and finally our LAM samples from the cavity wall and shoot apex as replicates of nonegg cell samples. An MDS plot visualizes the relationships among samples (MDSplot; Fig. S1f). The proximity of replicate samples indicated low technical and biological variation and that egg stages 8 and 9 can be used as replicates.

A general linear-model approach in EDGER identified differentially expressed genes. At an FDR of 0.01, 7693 genes were found differentially expressed in the comparison egg cells vs nonegg cells. Of those genes, 3432 were upregulated and 4261 downregulated in the egg cell (Table S6). A volcano plot shows that differentially expressed genes (FDR < 0.01), with a few exceptions, have an absolute fold change > 2 (Fig. S1g).

To visualize expression patterns of egg-expressed genes we performed a cluster analysis of the 6622 genes with a fold change of at least 2. In this analysis, we also included normalized expression data from antheridia, obtained from Meyberg *et al.* (2020), and green sporophytes. Choosing four clusters, one gathered genes highly expressed mainly in egg cells (cluster 1), two showed genes preferentially expressed in egg cells and antheridia (clusters 2 and 3), and one cluster contained genes highly expressed both in egg cells and the green sporophyte (cluster 4; Fig. 1).

Transcripts known to be expressed in egg cells are present in the laser-assisted microdissection egg transcriptome

To validate the egg transcriptome, the literature was surveyed for *P. patens* genes expressed in the egg. Those include the class-1

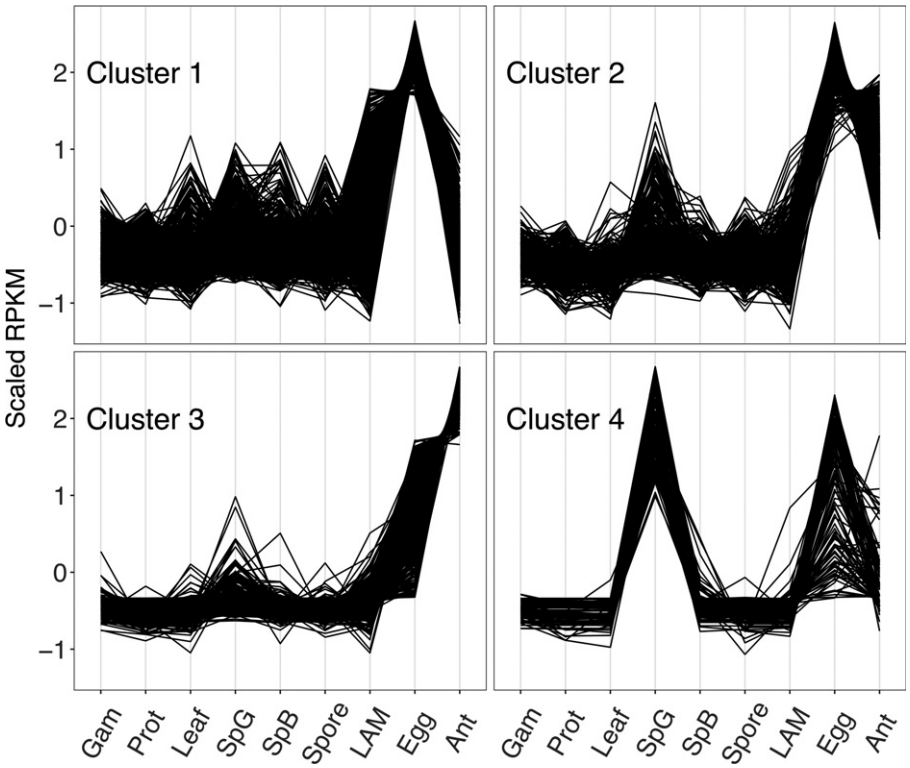

Fig. 1 Cluster analysis of preferentially egg-expressed genes in *Physcomitrium patens*. Expression patterns for genes with a high expression in egg cells (greater than two-fold compared with other tissues) visualized through cluster analysis. Expression in egg cells (Egg) was compared with gametophores (Gam), protonema (Prot), leaf-like structures (Leaf), brown sporophyte (SpB), spores, and the LAM control that were included in statistical analysis plus those of green sporophyte (SpG) and antheridia (Ant). See the Materials and Methods section for details. RPKM, reads per kilobase of transcript per million mapped reads.

KNOX gene *PpMKN2* and the homeobox genes *PpBELL1* and *PpBELL2*. Knock-in translational reporters of these genes are highly active in the egg, although not restricted to it (Sakakibara *et al.*, 2008; Horst *et al.*, 2016). Our data show that the transcripts of these genes are significantly enriched in the egg (Table S7). Egg cell activity of members of five additional TF families (*PpWOX13LA*, *PpCLF*, *PpFIE*, *PpLFY1*, *PpLFY2*, *PpSHI1*, and *PpSHI2*), one auxin efflux carrier (*PpPINA*), and one auxin biosynthesis gene (*PpTARA*), all expressed but not enriched in the egg cell (Table S7), was demonstrated previously through reporters (Tanahashi *et al.*, 2005; Mosquna *et al.*, 2009; Okano *et al.*, 2009; Landberg *et al.*, 2013, 2020; Sakakibara *et al.*, 2014). Loss of function of several of these genes affects reproductive development (Table S7).

In addition, *PpEXO70.3d*, encoding an exocyst subunit important for cell elongation and differentiation, was recently shown to be essential for normal egg cell development (Rawat *et al.*, 2017). Accordingly, we found this gene to be expressed in the egg cell (Table S7). Moreover, we identified egg expression of *PpMSH2* (Table S7), encoding a central component of the mismatch DNA repair pathway essential for reproductive organ development as well as male and female fertility (Trouiller *et al.*, 2006).

We also looked for egg expression of genes belonging to the *RWP-RK* family, which includes members important for egg cell fate determination and differentiation in Arabidopsis and *M. polymorpha* (Kőszegi *et al.*, 2011; Koi *et al.*, 2016; Rövekamp *et al.*, 2016). The *P. patens* homologue of the RKD-clade, *PpRKD*, and two additional *RWP-RK* genes belonging to the RPW-clade were expressed in the egg but not significantly enriched compared with other tissues, whereas the transcripts of three NLP-clade genes were (Table S7).

Finally, four members of the *EGG CELL-SPECIFIC* (*ECS*) gene family encoding eukaryotic A1 aspartic endopeptidases were strongly upregulated in the *P. patens* egg cell (Table S7). Two members of this family (*AtECS1* and *AtECS2*) are exclusively expressed in the Arabidopsis egg from which they are secreted to the extracellular space after fertilization to prevent polytubey (Sprunck *et al.*, 2012; Yu *et al.*, 2021).

Gene ontology analysis suggest similarities in enriched processes in egg cells across kingdoms

We used the R package TOPGO to search for processes and functions enriched in the *P. patens* egg cell. Using a *P*-value cutoff at 0.01, we identified 15 biological process GO terms (Table 1). In accordance with studies of Arabidopsis and human egg transcriptomes Wuest *et al.*, 2010), those include terms related to DNA replication, DNA repair, translation, ribosome biogenesis, chromatin silencing, and histone acetylation (Table 1). We also tested subsets of genes overexpressed in the moss egg for enrichment of GO terms (6622 genes with fold change > 2). For genes having

Table 1 Gene ontology (GO) term enrichment table showing significantly enriched biological processes among significantly overexpressed genes in the egg cell.

GO ID	Term	Annotated	Significant	Expected	Weight/Fischer
GO: 0006412	Translation	291	95	52.37	3.7e-09
GO: 0042254	Ribosome biogenesis	229	49	41.21	7.4e-07
GO: 0001510	RNA methylation	92	34	16.56	2.9e-05
GO: 0007018	Microtubule-based movement	101	37	18.18	0.00015
GO: 0080129	Proteasome core complex assembly	67	24	12.06	0.00038
GO: 0051788	Response to misfolded protein	83	27	14.94	0.00095
GO: 0043966	Histone H3 acetylation	4	4	0.72	0.00104
GO: 1903047	Mitotic cell cycle process	53	21	9.54	0.00201
GO: 0000724	Double-strand break repair via homologous recombination	32	12	5.76	0.00336
GO: 0044458	Motile cilium assembly	5	4	0.90	0.00447
GO: 0006094	Gluconeogenesis	118	33	21.24	0.00471
GO: 0006275	Regulation of DNA replication	52	16	9.36	0.00609
GO: 0009826	Unidimensional cell growth	69	17	12.42	0.00634
GO: 0006342	Chromatin silencing	36	13	6.48	0.00746
GO: 0009555	Pollen development	87	25	15.66	0.00915

The subset of genes analyzed is increased in expression at least two-fold in *Physcomitrium patens* egg cells compared with other tissues.

Arabidopsis homologues also expressed in egg cells (1745 genes; data obtained from Song *et al.* (2020)), 24 GO terms, including the aforementioned, were identified (Table S8). For *P. patens* genes with Arabidopsis homologues *not* expressed in the Arabidopsis egg (891 genes), 15 biological GO terms were identified, including 'membrane fusion' and 'microtubule-based movement', suggesting that these processes may be specifically upregulated in the *P. patens* egg (Table S9). Finally, we tested genes from the four clusters in Fig. 1 and found that, for example, the terms 'unidimensional cell growth', 'regulation of DNA replication', and 'ammonium transport' were specific to the egg (cluster 1), whereas terms like 'chromatin silencing', 'floral organ morphogenesis', 'sperm individualization', 'microtubule-based processes', and 'karyogamy' were shared with antheridia (clusters 2 and 3), and 'regulation of telomere maintenance', 'mitotic recombination', 'cell wall modification' and 'sporopollenin biosynthetic process' were shared with the green sporophyte (cluster 4) (Table S10).

Transcripts of genes involved in RNA-based silencing mechanisms are enriched in the *P. patens* egg

The microRNA (miRNA) and RNA-directed DNA methylation (RdDM) pathways are required for the establishment of transcriptionally repressive DNA methylation and include the activity of many genes (Matzke & Mosher, 2014; Erdmann & Picard, 2020). Both pathways are important for the regulation of reproduction in plants at different levels (Vashisht & Nodine, 2014; Martinez & Köhler, 2017; Gehring, 2019; Paro *et al.*, 2021). Three highly enriched transcripts in the *P. patens* egg belong to a family that encodes double-stranded RNA-binding proteins with conserved XH/XS domains and implicated in the regulation of RdDM at chromatin targets. These are similar to the Arabidopsis genes *FACTOR OF DNA METHYLATION1* (*AtFDM1*) and *AtFDM2* (Table S11; Xie *et al.*, 2012; Butt *et al.*, 2014). As transcripts of *AtFDM4* are enriched in the Arabidopsis egg (Wuest *et al.*, 2010; Sprunck *et al.*, 2019; Song *et al.*, 2020), it appears likely that DNA methylation is important to control the egg cell transcriptome in land plants by keeping transposons silent and egg cells quiescent prior to fertilization. Accordingly, transcripts of *DEMETER*, encoding a DNA glycosylase involved in DNA demethylation, were not detected in the Arabidopsis egg (Choi *et al.*, 2002; Morales-Ruiz *et al.*, 2006; Schoft *et al.*, 2011) and, in the *P. patens* egg, a homologous gene was strongly downregulated (Table S11).

Other important players in RdDM are clade III ARGOUNAUTE (AGO) proteins that, in Arabidopsis, include AtAGO4/6/9/8 that interact with Pol V transcripts to target DNA methylation (Mallory & Vaucheret, 2010; Matzke & Mosher, 2014). *AtAGO9* was reported to restrict megasporocyte fate to a single cell in the ovule and to silence transposons in the female gametophyte (Olmedo-Monfil *et al.*, 2010); correspondingly, its transcripts and translational products were found to be enriched in the egg (Wuest *et al.*, 2010; Sprunck *et al.*, 2019; Jullien *et al.*, 2020; Song *et al.*, 2020). Two genes encoding clade III AGOs were expressed in the *P. patens* egg, one (homologous to *AtAGO9*) with 24-fold enrichment and another (homologous to *AtAGO4*) with similar expression levels as in other tissues (Table S11).

Among Arabidopsis clade I AGO proteins (AtAGO1/5/10), at least AGO1 is important for executing miRNA function to regulate developmental programs and repress transposons (Mallory & Vaucheret, 2010; Matzke & Mosher, 2014). Both *AGO1* and *AGO5* are expressed in the Arabidopsis egg (Wuest *et al.*, 2010; Jullien *et al.*, 2020; Song *et al.*, 2020), and a clade I *AGO* homologue also showed significant expression in the *P. patens* stage 8 egg (Table S11). DICER-LIKE (DCL) proteins function upstream of the AGOs in small RNA biogenesis involved in RNA-based silencing (Finnegan *et al.*, 2003; Margis *et al.*, 2006; Nagano *et al.*, 2014; Paro *et al.*, 2021), and one member, *DCL2*, is preferentially expressed in the Arabidopsis egg (Takanashi *et al.*, 2011). Given

this, it is not surprising that the *P. patens DCL* homologue *PpmDCL* (Coruh *et al.*, 2015) also showed five-fold enriched expression in the egg, although the adjusted *P*-value comparing egg cells with other tissues was 0.02 (Table S11).

In addition, the expression of two histone-lysine *N*-methyltransferase (HKMTase) homologues were significantly enriched in the moss egg (Table S11). HKMTases are implicated in the epigenetic regulation of gene expression and catalyze the methylation of histone lysine (K) residues through their SET domains. The egg-enriched moss HKMTases are homologous to *AtSUVH4* and *AtSUVR4*, and a third gene, similar to *AtSUVR3*, was also expressed in the egg. Computational analysis of Arabidopsis HKMTases suggests that several SUVH proteins and SUVR3 catalyze mono or dimethylation of H3K9 (Satish *et al.*, 2018), whereas SUVR4 trimethylates H3K9 to suppress transposon activity in Arabidopsis (Veiseth *et al.*, 2010). In the Arabidopsis gametophyte, the egg showed a stronger H3K9me2 signal compared with neighboring cells, and this signal persisted after fertilization (Pillot *et al.*, 2010), suggesting that the H3K9me2 status plays a specific role in land plant egg cells.

Homologues of genes involved in karyogamy are active in the *P. patens* egg

Our data also revealed the enrichment of transcripts homologous to genes important for female reproductive success in Arabidopsis. The GO term 'karyogamy' came up in cluster 3, mainly consisting of genes highly expressed in both egg and antheridia (Table S12). This includes *GAMETE EXPRESSED1* (*GEX1*), encoding a membrane protein homologous to the yeast karyogamy protein Kar5 (Beh *et al.*, 1997; Alandete-Saez *et al.*, 2011; Ning *et al.*, 2013; Speijer *et al.*, 2015; Nishikawa *et al.*, 2020). GEX1 is localized to the nuclear membrane of the egg and central cell, and, upon double fertilization, it is required for sperm nuclear fusion in both cells. The enriched activity of a *GEX1* homologue in the moss egg suggests that members of this family may represent a conserved karyogamy factor in land plants. Three Arabidopsis genes encoding BINDING IMMUNOGLOBULIN PROTEINs of the heat shock protein 70 family are also involved in the fusion of polar nuclei together with their regulatory partners, ER-resident J-domain-containing proteins (ERDJ3A/B) (Yamamoto *et al.*, 2008; Maruyama *et al.*, 2014a,b, 2015, 2020). Homologues potentially encoding a corresponding complex in moss are also active or even enriched in the moss egg cell (Table S12).

Ammonium transporters are expressed in the *P. patens* egg

One of the most strongly upregulated genes in the *P. patens* egg encodes a conserved membrane-bound ammonium transporter homologous to Arabidopsis AMT2;1 (Hao *et al.*, 2020), and the transcripts of two additional homologues are also significantly enriched (Table S13). Ammonium transport also came up as an enriched GO term in cluster 1, containing genes mainly expressed in the egg (Table S13). Although *AtAMT2;1* expression is not elevated in the Arabidopsis female gametophyte (Wuest *et al.*, 2010; Julca *et al.*, 2021), our data suggest a specific role of ammonium transport in early land plant reproduction.

Unexpectedly, homologues of several genes important to male reproduction in angiosperms are active in the *P. patens* egg

Interestingly, the GO term 'pollen development' was overrepresented among egg-enriched genes (Table 1), and we identified enriched expression of homologues of several genes implicated in male reproductive development in Arabidopsis, primarily in the process of pollen and pollen wall development. These include homologues of the Arabidopsis gene *MS1*, encoding a PHD-finger TF expressed in the sporophytic tapetum layer surrounding developing pollen, where it activates the production of pollen coat proteins required for pollen wall development and pollen viability (Wilson *et al.*, 2001; Ito & Shinozaki, 2002; Ito *et al.*, 2007; Yang *et al.*, 2007; Lu *et al.*, 2020). Though *AtMS1* is not active in the egg (Julca *et al.*, 2021), transcripts of the two *P. patens MS1* homologues (*PpMS1A* and *PpMS1B*) are significantly enriched in the egg. As these genes are also expressed in antheridia of the male fertile ecotype Reute, they may be involved in processes common to both male and female germlines (Tables 2, S14; Meyberg *et al.*, 2020). The single *M. polymorpha MS1* homologue, Mp*MS1*, is also expressed in male spermatogenous tissues (Higo *et al.*, 2016), but its activity in female reproductive tissues has not yet been studied.

In an attempt to determine if the unexpected expression of the two moss *MS1* homologues in the egg cell is indeed due to sense transcripts, we analyzed how RNA-seq reads from egg cells, cavity wall, antheridia, and gametophores mapped to these genes, as well as to all other genes listed in Table S14 (Fig. S3). We also included a reference gene expressed in all four tissue types (Fig. S3a). For this reference gene, the strong overrepresentation of RNA-seq reads mapping only in the sense direction in the gametophore library shows that this library, unlike the others, was prepared to retain strand information. However, the LAM-captured samples display a strong 3′ bias in read mapping, most likely as a combinatorial effect of partial RNA degradation and polyA-based messenger RNA amplification prior to library preparation. This bias, which is also present in the antheridia library, suggests that sense transcripts are formed from the reference gene in all samples. Correct split mapping of reads across introns in many instances further supports sense rather than antisense transcription. To further test our interpretation, we also analyzed the RNA-seq reads from egg cells of genes known to express sense transcripts in the egg cell (i.e. the first 15 genes listed in Table S7) and found the expected 3′ bias in read mapping (Fig. S4).

Based on the aforementioned reasoning, biased read mapping of antheridia and egg cell samples towards the annotated 3′ end of *PpMS1A* and *PpMS1B* suggests that they represent sense transcripts (Fig. S3b,c). Additional smaller clusters of egg-cell-derived reads also mapped closer to the 5′ end of *PpMS1B*. As several alternative splice variants have been identified with exons other than the three in the primary transcript (Fig. S3c), it is difficult

Table 2 Expression levels (transcripts per million normalized) of *PpMS1*, *PpBNB*, and homologs to *TPD* and *DAZ* genes in vegetative tissue, egg cells, and antheridia of *Physcomitrium patens*.

Tissue	*PpMS1A*	*PpMS1B*	*PpBNB1*	*PpBNB2*	*TPD1* hom. Pp3c19_9830	*TPD1* hom. Pp3c5_9010	*TPD1* hom. Pp3c22_22416	*DAZ* hom. Pp3c18_22010
Egg	15.1	9.1	4.9	11.7	133	40.8	22.5	17.7
Antheridia	10.9	13.7	8.2	13.5	5.8	9.3	25	8.9
Cavity	0.1	0	0	0	3.5	0.5	0	1.3
Shoot tip	0	0	0	0	0.1	0	2.5	0
Gametophore	0	0	0	0	1.2	2.5	3.5	5.3

to determine if these reads represent sense or low level of additional antisense transcripts.

Furthermore, predicted sense transcripts of three *P. patens* genes encoding TAPETUM DETERMINANT1 (TPD1)-like cysteine-rich ligands (Huang *et al.*, 2016) are significantly enriched in the egg and also expressed in antheridia (Tables 2, S14; Fig. S3d–f; Meyberg *et al.*, 2020). In Arabidopsis, TPD1 is produced in, and secreted from, the microsporocytes, which are diploid precursors of the male gametophytes. Differentiation of the surrounding tapetum layer depends on the perception of the TPD1 signal peptide via the EXCESS MICROSPOROCYTES1 (EMS1) receptor kinase (Canales *et al.*, 2002; Zhao *et al.*, 2002; Huang *et al.*, 2016). Although mutations in *TPD1* and *EMS1* have no effect on female reproductive development in Arabidopsis, homologues in rice (*Oryza sativa*) and maize (*Zea mays*) are required to limit the number of megasporocytes in the ovules, besides their conserved roles in anther development (Sheridan *et al.*, 1996; Nonomura *et al.*, 2003; Zhao *et al.*, 2008; Hong *et al.*, 2012; Wang *et al.*, 2012). Though two *EMS1*-like genes produce predicted sense transcripts in the *P. patens* egg, none of them appear significantly enriched, opening the possibility that the egg-produced TPD1 may be exported to activate receptors expressed elsewhere (Fig. S3g,h; Table S14).

DUO1 was identified as a male-germline-specific TF in Arabidopsis and *M. polymorpha* (Durbarry *et al.*, 2005; Rotman *et al.*, 2005; Borg *et al.*, 2011; Higo *et al.*, 2018). As expected, both *P. patens DUO1* homologues are expressed in antheridia, but to our surprise we also found a strong enrichment of transcripts of one of them, *PpDUO1B* (Higo *et al.*, 2018), in the egg (Table S14; Meyberg *et al.*, 2020). As the previous annotation of *PpDUO1B* to Pp3c24_11770 disappeared during a recent genome curation, we used the RNA-seq reads of antheridia to map the *PpDUO1B* gene in the genomic region around the current Pp3c24_11770V.3 gene (Fig. S3i). This revealed a gene structure for *PpDUO1B* with four exons, the last of which coincides with the single exon of the predicted Pp3c24_11770V.3 gene (Fig. S3i), encoding a protein with an almost identical sequence to the uncharacterized protein XP_024363644 at the National Center for Biotechnology Information, and high similarity to that of Pp3c8_16720, designated as PpDUO1A (Higo *et al.*, 2018) (Fig. S3j). The read mapping suggests that *PpDUO1B* is producing sense transcripts in antheridia, whereas predicted antisense transcripts are enriched in the egg cell (Fig. S3i). The egg cell reads thus accumulate in the 5′ end of the *PpDUO1B* gene and also over the last intron. This suggests that PpDUO1B protein production is likely silenced in the egg cell. Similarly, *PpDUO1A* is likely silenced in the gametophore, as the retained read strand orientation reveals antisense transcription (Fig. S3k). Both AtDUO1 and MpDUO1 directly control the expression of *DUO1-ACTIVATED ZINC FINGER* (*DAZ*) genes in the male germline (Borg *et al.*, 2011, 2014; Higo *et al.*, 2018), and although PpDUO1 protein production appears repressed in the moss egg, three *P. patens DAZ* homologues produce predicted sense transcripts there, one of which is significantly enriched (Tables 2, S14; Fig. S3l–n), suggesting a role of these components in the egg.

In contrast to *PpDUO1B*, sense transcripts homologous to the *BNB* genes, which are important for male generative cell specification in Arabidopsis, accumulated in the *P. patens* egg cell (Fig. S3o,p; Table S14). The *BONOBO* genes encode class VIIIa bHLH TFs (Yamaoka *et al.*, 2018; Zhang *et al.*, 2020). In *M. polymorpha*, the single *BNB* gene Mp*BNB* is expressed in the initial cells of both male and female reproductive organs, as well as in the egg cell of immature archegonia and in sperm progenitors (Yamaoka *et al.*, 2018). Because Mp*bnb* knockout mutants fail to develop reproductive organs, the role of Mp*BNB* in *M. polymorpha* gamete development could not be elucidated. However, Mp*BNB* could, at least partially, replace the *AtBNB1* and *AtBNB2* function in Arabidopsis male generative cell specification (Yamaoka *et al.*, 2018). We found that the expression of the two *BNB* homologues in *P. patens* (Yamaoka *et al.*, 2018), here called *PpBNB1* (*PpbHLH081*) and *PpBNB2* (*PpbHLH074*), are significantly enriched in the egg and that both are also expressed in antheridia (Tables 2, S14; Meyberg *et al.*, 2020). To elucidate the potential role of *BNB* in the generative cells of bryophytes, we analyzed their expression profiles and the consequence of their loss of function.

*PpBNB*s are expressed in reproductive organ apical stem cells and in the immature female and male germ cells

We could confirm that both *PpBNB* genes are expressed in archegonia and antheridia of the *P. patens* Reute ecotype using RT-qPCR, and we made reporter lines in the same ecotype to determine their expression in more detail (Figs 2, S4). The activity of the translational knock-in *PpBNB1* reporter and the transcriptional *PpBNB2* reporter lines strongly overlapped during reproductive organ development, indicating that the PpBNB proteins may have redundant functions (Figs 3a,b, 4a,b). Expression initiated as soon as the respective reproductive organ

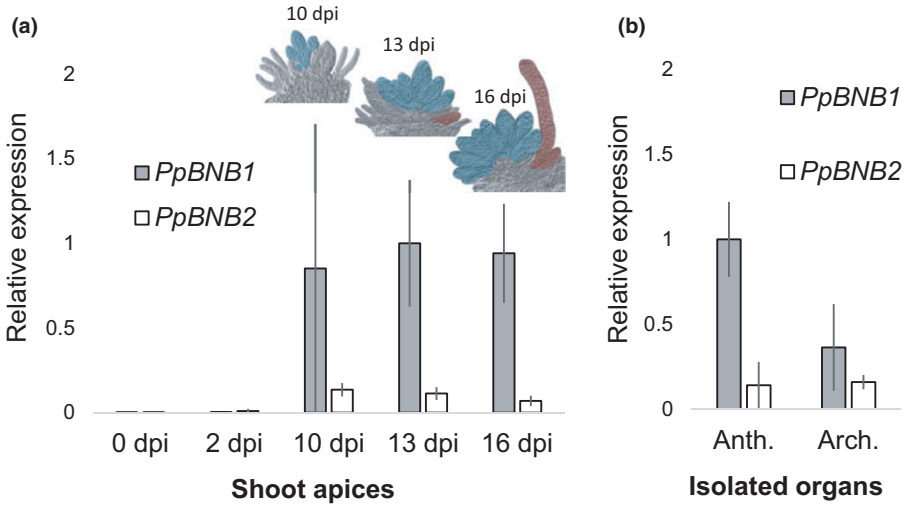

Fig. 2 Relative *PpBNB1* and *PpBNB2* transcript abundance in *Physcomitrium patens* assayed by quantitative PCR. (a) Expression in detached shoot apices harvested at different time points after transfer to inductive conditions. Image insets demonstrate that apices possess young antheridia (Anth.) at 10 days post-induction (dpi), apices possess mid-stage antheridia and young archegonia (Arch.) at 13 dpi, and apices possess mature antheridia and mid-stage archegonia (antheridia and archegonia are false colored in blue and red, respectively) at 16 dpi. (b) Expression in isolated antheridia and archegonia bundles. Each data point represents an average of three independent biological replicates, and error bars indicate SD.

primordia were defined in the initial apical stem cells (defined in Landberg *et al.* (2020)) and remained in all daughter cells formed in two cell files during the first developmental stages, with the strongest signal in the apicalmost cells. When inner cells formed and further divided during stages 3–6 (stages as in Landberg *et al.* (2013)), *PpBNB1/2* expression became restricted to these cells and the apical tip cells. In archegonia, a total of six inner cells forms in a single row. The basalmost of these cells matures into the egg cell at stages 7–9, whereas the remaining inner cells degrade to leave an empty canal connecting the egg to the outside as the organ tip opens. *PpBNB1/2* remained active in the canal cells until they degraded and, in accordance with the egg transcriptome data, in the pre-egg up to stage 8, but the reporter signals faded in the mature egg cell at stage 9. The antheridial inner cells constitute the spermatogenous cell initials that undergo spermatogenesis during stages 7–9; at the initiation of this process, *PpBNB1/2* expression declined.

Loss of *PpBNB* function results in aberrant inner cell division and gamete differentiation

Using the highly self-fertile ecotype Reute (Hiss *et al.*, 2017), we produced several independent single and double loss-of-function *Ppbnb* mutants by CRISPR-Cas9 genome editing, using crRNAs specifically targeting the two genes (Table S2). By sequencing the two loci in the resulting lines, we could identify three *Ppbnb1*, two *Ppbnb2*, and two *Ppbnb1Ppbnb2* mutants, in which the targeted gene(s) were destroyed (Table S3). As putative off-targets of all crRNAs used had at least four mismatches, it is extremely unlikely that off-target editing events occur (Table S2; Modrzejewski *et al.*, 2020), and sequencing of the putative off-targets in the double mutant lines revealed no deviations from the wild-type (WT).

Archegonia initiated at approximately the same time in the mutants and the WT. Whereas the *Ppbnb2* single mutant archegonia were indistinguishable from the WT, some *Ppbnb1* inner cells (Fig. 3c,d), most commonly the pre-egg, the upper basal cell (UBC), and the two basal canal cells, underwent one or two extra transverse cell divisions (Fig. 3e). Already at stage 5, the pre-egg divided into two similarly sized cells, which either arrested or divided once more, either asymmetrically to form a smaller UBC and a pre-egg as in WT, or into two equally sized cells resembling immature canal cells. In rare cases, the pre-egg or UBC divided longitudinally. The extra divisions blocked further development of the inner cells; the pre-egg never matured and the extra UBC and canal cells did not degrade properly. With a few exceptions, the apical part of the neck developed more normally, the apical canal cells degraded to a large extent, and the apex opened. The double mutant archegonia were even further affected (Fig. 3f). Instead of the strict transversal inner cell division of the WT, the *Ppbnb1bnb2* inner cells divided in all directions, forming an almost globular organ without signs of a specified egg cell. In addition, neck development arrested in the majority of archegonia (76%, $n = 33$), and although some showed partial neck elongation, they never fully developed and always lacked a proper single-celled canal that degraded and opened at the tip.

Though antheridia developed normally in the single mutants (Fig. 4c,d), the number of antheridia was reduced in *Ppbnb2* and even further so in *Ppbnb1bnb2* double mutant lines (Table S15). In addition, double mutant antheridia were severely affected (Fig. 4e). Although the first cell divisions appeared normal, the number of inner cells formed was dramatically reduced (on average, 1.3 cells compared with six in WT) (Landberg *et al.*, 2020). The majority arrested and appeared morphologically similar to the outer vegetative cells. Occasionally, the inner cell(s) divided once more and, rarely, one or two of them produced a few small and round spermatogenous cells. In contrast to the inner cell defects, the development of the organ's tip and base appeared normal, indicating that the regulation of these tissues is independent of *PpBNB*s and of the progression of inner cell development.

As expected, the *Ppbnb1* and *Ppbnb1bnb2* mutants did not develop sporophytes, and even though the reproductive organs of *Ppbnb2* single mutants appeared normal, the frequency of sporophyte formation was clearly reduced, suggesting that gametes of one or both sexes were affected (Table 3). When the self-sterile *Ppbnb1* mutant was used as a sperm donor in crossings with the

New Phytologist

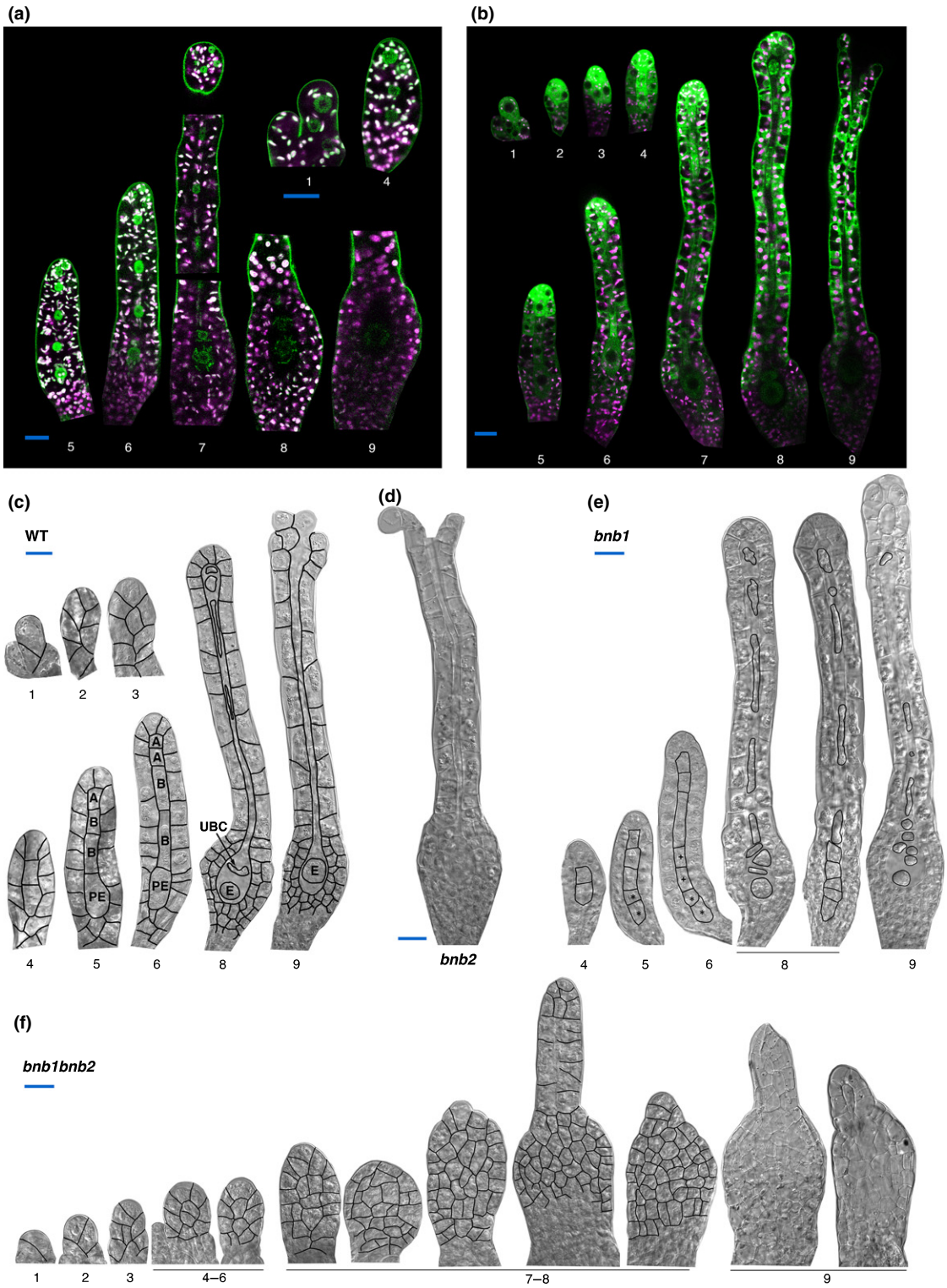

Fig. 3 *PpBNB* expression pattern and *Ppbnb* mutant phenotype during *Physcomitrium patens* archegonial development. Expression of the (a) translational knock-in reporter *PpBNB1pro::PpBNB1-GFPGUS-1* and (b) transcriptional reporter *PpBNB2pro::GFPGUS* integrated in locus P108 at the stages of organ development indicated. For each archegonium stage in (a) and (b) a merge of confocal channels detecting green fluorescent protein (green) and chloroplast autofluorescence (magenta) is shown. (c–f) Differential interference contrast images of representative stages of (c) wild-type (WT), (d) *Ppbnb2*, (e) *Ppbnb1*, and (f) *Ppbnb1bnb2* archegonia. Organs at stages marked with a thin black line show examples of the phenotypic variation. It is not possible to distinguish individual stages 4–6 and 7–8 in the double mutant. Numbers 1–9 indicate developmental stages as defined in Landberg *et al*. (2013, 2020). Cell types are indicated in (c) as follows: A, apical canal cell; B, basal canal cell; PE, pre-egg; E, egg; UBC, upper basal cell. * and + in (e) indicate extra cell division of the pre-egg and the basalmost canal cell, respectively. Inner cells and their borders have been traced in black for clarity at some stages. Bar, 20 μm.

Table 3 Analysis of number of sporophytes formed per shoot after selfing or after crossing between indicated *PpBNB* mutant line and the wild-type (WT) strains Gransden (Gd) or Reute (R) of *Physcomitrium patens*.

Female genotype	Male genotype	Frequency of shoots with initiated sporophyte development (%)	No. shoots analyzed	No. shoots with at least one sporophyte formed
bnb1bnb2-1	*bnb1bnb2-1*	0	793	0
bnb1bnb2-2	*bnb1bnb2-2*	0	786	0
bnb1-1	*bnb1-1*	0	290	0
bnb1-2	*bnb1-2*	0	259	0
bnb2-1	*bnb2-1*	65	455	295
bnb2-2	*bnb2-2*	71	506	358
R WT	R WT	98	300	294
Gd WT	Gd WT	1	448	4
Gd WT	R WT	71	48	34
Gd WT	*bnb1-1*	42	146	62
Gd WT	*bnb1-2*	33	75	25
bnb1-1	R WT	0	253	0

largely male sterile WT ecotype Gransden, the frequency of sporophytes on Gransden shoots clearly increased, although not as much as when the WT ecotype Reute was used as sperm donor, revealing some *Ppbnb1* male fertility. However, when adding sperm cells from the highly fertile WT ecotype Reute to *Ppbnb1* mutants, no sporophytes formed, indicating that *Ppbnb1* mutant egg cells are nonfunctional (Table 3).

Discussion

Comparisons of egg cell expression from different plant lineages help to highlight commonly expressed genes for the specification of this key reproductive cell type. Processes shared between bryophyte and angiosperm egg cells include epigenetic modifications, DNA repair, and the activity of genes important for egg cell specification and function, such as the RWP-RK TFs of the RKD, RWP, and NLP clades and ECS-like endopeptidases (Wuest *et al.*, 2010; Kőszegi *et al.*, 2011; Sprunck *et al.*, 2012; Koi *et al.*, 2016; Rövekamp *et al.*, 2016; Yu *et al.*, 2021). In *M. polymorpha*, Schmid *et al.* (2018) showed that epigenetic reprogramming occurs by similar but (compared with the sporophytic tissues) distinct processes in male and female reproductive organs. They also revealed that the level of DNA methylation increased in reproductive organs compared with the vegetative tissues from which they were derived. Whether there is a similar increase in the *P. patens* egg is not possible to elucidate from our work, but it appears that the genetic program driving RdDM is largely distinct from vegetative tissues.

Despite a clear overlap in gene activity between bryophyte and angiosperm egg cells, there are also several differences revealed by our transcriptome analysis. The upregulation of genes involved in, for example, microtubule-based movement and membrane fusion appears specific to the *P. patens* egg. This may correlate with the extensive number of small vesicles with electron-dense material that form in the moss egg. These eventually fuse with the plasma membrane to release the contents to the surrounding cavity matrix to form a proper environment for the egg and the entering sperm cells (Sanchez-Vera *et al.*, 2017).

Another process that appears specific to the moss egg cell is ammonium transport. Besides the nutritious role of ammonium, which is the most obvious potential function, it can also act as a signaling molecule to trigger physiological and morphological changes in plants (Liu & von Wirén, 2017). In the green unicellular flagellated algae *Chlamydomonas reinhardtii*, ammonium is used as a chemoattractant (Ermilova *et al.*, 2007). It would be interesting to know if ammonium also acts as a chemoattractant for sperm cells. Autophagy is highly active in the *P. patens* egg (Sanchez-Vera *et al.*, 2017) and may contribute to ammonium accumulation from, for example, protein and amino acid degradation. If so, the PpAMT2;1 proteins could potentially act as exporters of ammonium out of the egg cell instead of into it.

Strikingly, processes that are male germline specific in Arabidopsis are active not only in antheridia but also in the *P. patens* egg. This may suggest that pathways shared between the two germlines were present in the common ancestors of all land plants besides the sex-specifying factors. Work on volvocine algae suggests that the on–off shift of a single sex-determining factor controls the gender of a gamete precursor (Geng *et al.*, 2014), revealing that a separation of gamete and sex determination is also theoretically possible in land plants. This assumption is supported by the fact that the germ-cell-specifying *BNB* gene is also active in both germlines in *M. polymorpha*. However, unlike the Mp*BNB*, the two moss *PpBNB*s are dispensable for reproductive organ initiation from apical stem cells but are needed for the continued formation of antheridia. The main function of the *PpBNB*s is inner cell specification, an apparent requirement for proper cell divisions and gamete differentiation. The changes in inner cell identity, and thus division, resemble that of auxin biosynthesis mutants: reduced division combined with developmental arrest in antheridia and ectopic cell division combined with loss of cell identity in archegonia (Landberg *et al.*, 2020). Although this may not necessarily imply that auxin sensing is affected in *Ppbnb* mutant reproductive organs, it clearly demonstrates the importance of strictly controlled cell division patterns for proper gamete differentiation. Albeit Mp*BNB* function at these stages could not be analyzed, the fact that Mp*BNB* is also expressed in immature egg and sperm cell progenitors (Yamaoka *et al.*, 2018) indicates that the BNBs have a conserved function in both male and female germ cell specification in bryophytes, whereas their function in flowering plants is restricted to the male germ cells.

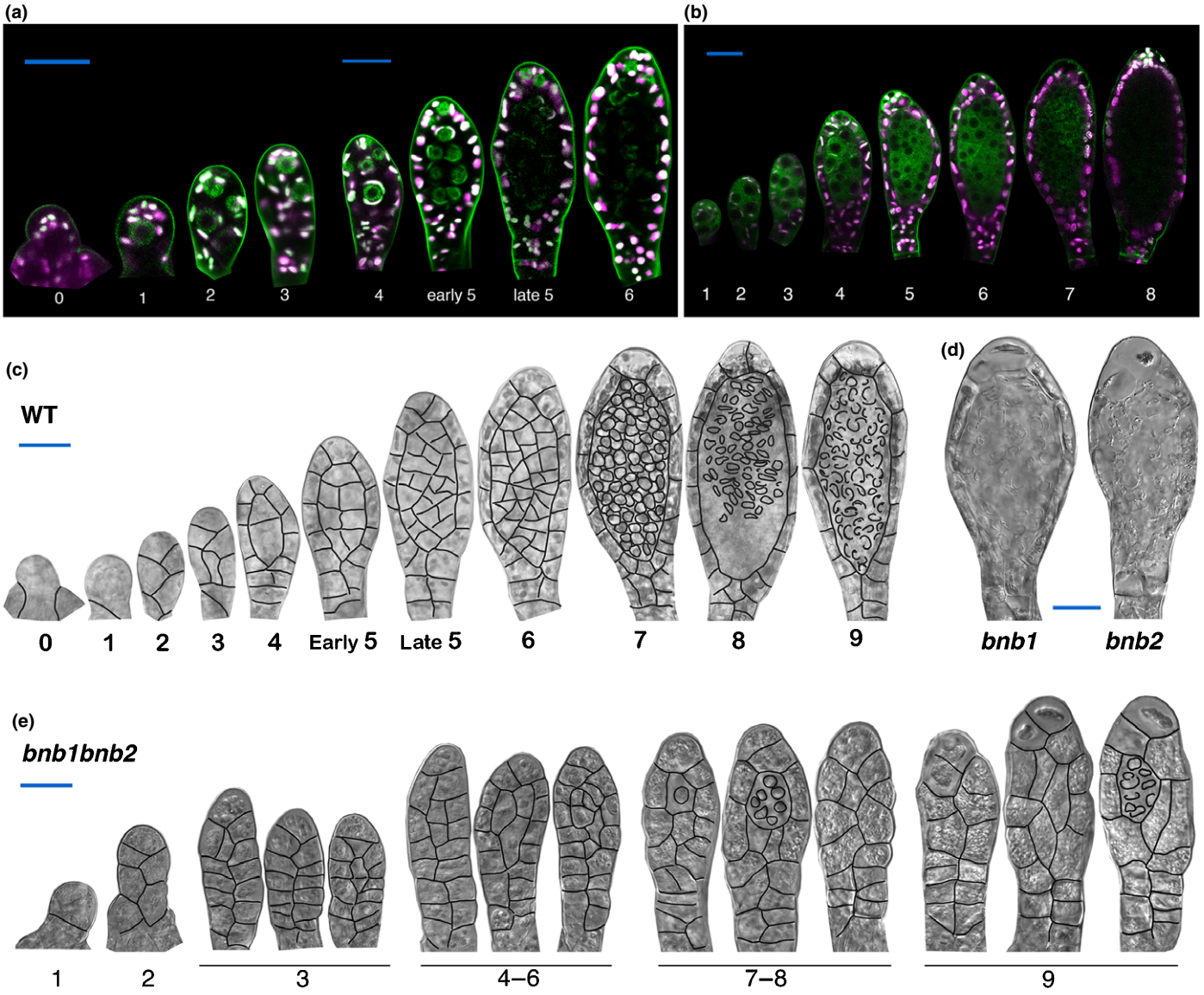

Fig. 4 *PpBNB* expression and mutant phenotype relating to antheridia development in *Physcomitrium patens*. Expression of the (a) translational knock-in reporter *PpBNB1pro::PpBNB1-GFPGUS-1* and (b) transcriptional reporter *PpBNB2pro::GFPGUS* integrated in locus P108 at indicated stages of antheridia development. The images in (a) and (b) are a merge of confocal channels detecting green fluorescent protein (green) and chloroplast autofluorescence (magenta). (c) Differential interference contrast (DIC) images of representative stages from wild-type (WT) antheridial development. (d) DIC images of stage 9 antheridia of the single mutants *Ppbnb1* and *Ppbnb2*. (e) DIC images of representative organs from *Ppbnb1bnb2* double mutant. For stages 3–9, three organs are shown exemplifying the phenotypic variation. It is not possible to distinguish individual stages 4–6 and 7–8 in the double mutant. Numbers 0–9 indicate antheridia stages according to Landberg *et al*. (2013, 2020). In (c, e) the inner cells have been traced in black for clarity. Bar, 20 μm. In (a) the left scale bar represents stages 0–3 and the right scale bar represents stages 4–6.

A subset of the genes upregulated in the *P. patens* egg is also highly active in the green sporophyte stage. We know little about cell-specific expression patterns in the embryo/sporophyte, but it is tempting to speculate that several of these genes are involved in the preparation of upcoming phase changes: the transitions between the generations of the plant life cycle from diploid to haploid in the sporophyte and from haploid to diploid in the egg.

It is important to note that the proportion of largely egg-specific genes (the 524 genes with a fold change > 256) without clear homology to Arabidopsis genes reached more than 72%, whereas the corresponding figure for all genes was only 31%. Therefore, the transcriptome data suggest that nonangiosperm gene networks widely contribute to the specification and differentiation of the *P. patens* egg. This is not surprising given the distinct environment in which the egg cell develops and the distinct mechanisms of fertilization (Sharma *et al.*, 2021).

Acknowledgements

This work was supported by grants from the Knut and Alice Wallenberg Foundation (KAW; 2012.0087 to ES), the Swedish Research Council (VR; 621-2014-4941; 2018-04068

to ES), the University of Zurich, and a grant from the 'Staatssekretariat für Bildung und Forschung' in the framework of COST action FA0903 (to UG and AS). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, which is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The platform is also supported by VR and KAW. We thank Dr German Martinez Arias and two anonymous reviewers for helpful comments on the manuscript and Prof. Stefan Rensing for kindly sharing RNA-seq data.

Author contributions

VS-V, KL, ML-O and MT conducted the experiments and analyzed the data together with RM-V, UL and ES; VS-V, AS, UG, ML-O, KL, MT and ES designed the experiments; ES and UG provided materials and resources; VS-V, ML-O, KL, MT, UL and ES interpreted the results and wrote the manuscript. All authors reviewed and commented on the manuscript. VS-V, KL and ML-O have shared first authorship.

ORCID

Ueli Grossniklaus https://orcid.org/0000-0002-0522-8974
Ulf Lagercrantz https://orcid.org/0000-0003-2440-0677
Katarina Landberg https://orcid.org/0000-0002-2945-8571
Mauricio Lopez-Obando https://orcid.org/0000-0002-1380-0643
Rafael Muñoz-Viana https://orcid.org/0000-0002-1363-6978
Victoria Sanchez-Vera https://orcid.org/0000-0001-8615-5270
Anja Schmidt https://orcid.org/0000-0002-3276-3243
Eva Sundberg https://orcid.org/0000-0003-4228-434X
Mattias Thelander https://orcid.org/0000-0002-6663-7405

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The RNA-seq data are openly available in the Gene Expression Omnibus at https://www.ncbi.nlm.nih.gov/geo, reference number GSE182112.

References

Alandete-Saez M, Ron M, Leiboff S, McCormick S. 2011. *Arabidopsis thaliana GEX1* has dual functions in gametophyte development and early embryogenesis. *The Plant Journal* **68**: 620–632.

Alexa A, Rahnenfuhrer J. 2021. *topGO: enrichment analysis for gene ontology*. R package v.2.44.0. [WWW document] URL https://bioconductor.org/packages/release/bioc/html/topGO.html [accessed 28 April 2021].

Ammar R, Thompson J. 2021. *zFPKM: a suite of functions to facilitate zFPKM transformations*. R package v.1.14.0. [WWW document] URL https://github.com/ronammar/zFPKM/ [accessed 28 April 2021].

Andrews S. 2010. *FastQC: a quality control tool for high throughput sequence data*. [WWW document] URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc [accessed 29 April 2021].

Beh CT, Brizzio V, Rose MD. 1997. *KAR5* encodes a novel pheromone-inducible protein required for homotypic nuclear fusion. *Journal of Cell Biology* **139**: 1063–1076.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B: Methodological* **57**: 289–300.

Borg M, Brownfield L, Khatab H, Sidorova A, Lingaya M, Twell D. 2011. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. *Plant Cell* **23**: 534–549.

Borg M, Rutley N, Kagale S, Hamamura Y, Gherghinoiu M, Kumar S, Sari U, Esparza-Franco MA, Sakamoto W, Rozwadowski K *et al.* 2014. An EAR-dependent regulatory module promotes male germ cell division and sperm fertility in *Arabidopsis*. *Plant Cell* **26**: 2098–2113.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**: 525–527.

Brownfield L, Hafidh S, Borg M, Sidorova A, Mori T, Twell D. 2009. A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genetics* **5**: e1000430.

Bushnell B. 2014. *BBMap: a fast, accurate, splice-aware aligner*. https://sourceforge.net/projects/bbmap/ [accessed 3 February 2014].

Butt H, Graner S, Luschnig C. 2014. Expression analysis of *Arabidopsis* XH/XS domain proteins indicates overlapping and distinct functions for members of this gene family. *Journal of Experimental Botany* **65**: 1217–1227.

Canales C, Bhatt AM, Scott R, Dickinson H. 2002. EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in Arabidopsis. *Current Biology* **12**: 1718–27.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.

Choi YH, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fisher RL. 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* **110**: 33–42.

Coruh C, Cho SH, Shahid S, Liu Q, Wierzbicki A, Axtell MJ. 2015. Comprehensive annotation of *Physcomitrella patens* small RNA loci reveals that the heterochromatic short interfering RNA pathway is largely conserved in land plants. *Plant Cell* **27**: 2148–2162.

Durbarry A, Vizir I, Twell D. 2005. Male germ line development in Arabidopsis. *duo pollen* mutants reveal gametophytic regulators of generative cell cycle progression. *Plant Physiology* **137**: 297–307.

Erdmann RM, Picard CL. 2020. RNA-directed DNA methylation. *PLoS Genetics* **16**: e1009034.

Ermilova EV, Nikitin MM, Fernández E. 2007. Chemotaxis to ammonium/methylammonium in *Chlamydomonas reinhardtii*: the role of transport systems for ammonium/methylammonium. *Planta* **226**: 1323–1332.

Finnegan J, Margis R, Waterhouse PM. 2003. Posttranscriptional gene silencing is not compromised in the *Arabidopsis CARPEL FACTORY* (*DICER-LIKE1*) mutant, a homolog of Dicer-1 from *Drosophila*. *Current Biology* **13**: 236–240.

Florez Rueda AM, Grossniklaus U, Schmidt A. 2016. Laser-assisted microdissection (LAM) as a tool for transcriptional profiling of individual cell types. *Journal of Visualized Experiments* **10**: e53916.

Gehring M. 2019. Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming? *New Phytologist* **224**: 91–96.

Geng S, De Hoff P, Umen JG. 2014. Evolution of sexes from an ancestral mating-type specification pathway. *PLoS Biology* **12**: e1001904.

Hackenberg D, Twell D. 2019. The evolution and patterning of male gametophyte development. *Current Topics in Developmental Biology* **131**: 257–298.

Hafidh S, Honys D. 2021. Reproduction multitasking: the male gametophyte. *Annual Review of Plant Biology* **72**: 581–614.

Hao DL, Zhou JY, Yang SY, Qi W, Yang KJ, Su YH. 2020. Function and regulation of ammonium transporters in plants. *International Journal of Molecular Science* **18**: e3557.

Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. 2013. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics* **14**: e778.

Hater F, Nakel T, Groß-Hardt R. 2020. Reproductive multitasking: the female gametophyte. *Annual Review of Plant Biology* **71**: 517–546.

Higo A, Kawashima T, Borg M, Zhao M, López-Vidriero I, Sakayama H, Montgomery SA, Sekimoto H, Hackenberg D, Shimamura M *et al.* 2018. Transcription factor DUO1 generated by neo-functionalization is associated with evolution of sperm differentiation in plants. *Nature Communications* **9**: e5283.

Higo A, Niwa M, Yamato KT, Yamada L, Sawada H, Sakamoto T, Kurata T, Shirakawa M, Endo M, Shigenobu S *et al.* 2016. Transcriptional framework of male gametogenesis in the liverwort *Marchantia polymorpha* L. *Plant and Cell Physiology* **57**: 325–338.

Hiss M, Meyberg R, Westermann J, Haas FB, Schneider L, Schallenberg-Rüdinger M, Ullrich KK, Rensing SA. 2017. Sexual reproduction, sporophyte development and molecular variation in the model moss *Physcomitrella patens*: introducing the ecotype Reute. *The Plant Journal* **606**: 20–110.

Hong L, Tang D, Shen Y, Hu Q, Wang K, Li M, Lu T, Cheng Z. 2012. *MIL2* (*MICROSPORELESS2*) regulates early cell differentiation in the rice anther. *New Phytologist* **196**: 402–413.

Horst NA, Katz A, Pereman I, Decker EL, Ohad N, Reski R. 2016. A single homeobox gene triggers phase transition, embryogenesis and sexual reproduction. *Nature Plants* **2**: e15209.

Huang J, Zhang T, Linstroth L, Tillman Z, Otegui MS, Owen HA, Zhao D. 2016. Control of anther cell differentiation by the small protein ligand TPD1 and its receptor EMS1 in *Arabidopsis*. *PLoS Genetics* **12**: e1006147.

Ito T, Nagata N, Yoshiba Y, Ohme-Takagi M, Ma H, Shinozaki K. 2007. *Arabidopsis MALE STERILITY1* encodes a PHD-type transcription factor and regulates pollen and tapetum development. *Plant Cell* **19**: 3549–3562.

Ito T, Shinozaki K. 2002. The *MALE STERILITY1* gene of *Arabidopsis*, encoding a nuclear protein with a PHD-finger motif, is expressed in tapetal cells and is required for pollen maturation. *Plant and Cell Physiology* **43**: 1285–1292.

Julca I, Ferrari C, Flores-Tornero M, Proost S, Lindner AC, Hackenberg D, Steinbachová L, Michaelidis C, Gomes Pereira S, Misra CS *et al.* 2021. Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. *Nature Plants* **7**: 1143–1159.

Jullien PE, Bonnet DMV, Pumplin N, Schröder JA, Voinnet O. 2020. Asymmetric expression of ARGONAUTES in *Arabidopsis* reproductive tissues. *bioRXiv* doi: 10.1101/2020.05.18.102863.

Koi S, Hisanaga T, Sato K, Shimamura M, Yamato KT, Ishizaki K, Kohchi T, Nakajima K. 2016. An evolutionarily conserved plant RKD factor controls germ cell differentiation. *Current Biology* **11**: 1775–1781.

Kőszegi D, Johnston AJ, Rutten T, Czihal A, Altschmied L, Kumlehn J, Wüst SEJ, Kirioukhova O, Gheyselinck J, Grossniklaus U *et al.* 2011. Members of the RKD transcription factor family induce an egg cell-like gene expression program. *The Plant Journal* **67**: 280–291.

Landberg K, Pederson ERA, Viaene T, Bozorg B, Friml J, Jönsson H, Thelander M, Sundberg E. 2013. The moss *Physcomitrella patens* reproductive organ development is highly organized, affected by the two *SHI/STY* genes and by the level of active auxin in the *SHI/STY* expression domain. *Plant Physiology* **162**: 1406–1419.

Landberg K, Šimura J, Ljung K, Sundberg E, Thelander M. 2020. Studies of moss reproductive development indicate that auxin biosynthesis in apical stem cells may constitute an ancestral function for focal growth control. *New Phytologist* **229**: 845–860.

Liu Y, von Wirén N. 2017. Ammonium as a signal for physiological and morphological responses in plants. *Journal of Experimental Botany* **68**: 2581–2592.

Lu JY, Xiong SX, Yin W, Teng XD, Lou Y, Zhu J, Zhang C, Gu JN, Wilson ZA, Yang ZN. 2020. *MS1*, a direct target of MS188, regulates the expression of key sporophytic pollen coat protein genes in *Arabidopsis*. *Journal of Experimental Botany* **71**: 4877–4889.

Mallory A, Vaucheret H. 2010. Form, function, and regulation of ARGONAUTE proteins. *Plant Cell* **22**: 3879–3889.

Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM. 2006. The evolution and diversification of Dicers in plants. *FEBS Letters* **580**: 2442–2450.

Martinez G, Köhler C. 2017. Role of small RNAs in epigenetic reprogramming during plant sexual reproduction. *Current Opinion in Plant Biology* **36**: 22–28.

Maruyama D, Endo T, Nishikawa S. 2015. BiP3 supports the early stages of female gametogenesis in the absence of BiP1 and BiP2 in *Arabidopsis thaliana*. *Plant Signaling and Behavior* **10**: e1035853.

Maruyama D, Higashiyama T, Endo T, Nishikawa S. 2020. Fertilization-coupled sperm nuclear fusion is required for normal endosperm nuclear proliferation. *Plant and Cell Physiology* **61**: 29–40.

Maruyama D, Sugiyama T, Endo T, Nishikawa S. 2014a. Multiple BiP genes of *Arabidopsis thaliana* are required for male gametogenesis and pollen competitiveness. *Plant and Cell Physiology* **255**: 801–810.

Maruyama D, Yamamoto M, Endo T, Nishikawa S. 2014b. Different sets of ER-resident J-proteins regulate distinct polar nuclear-membrane fusion events in *Arabidopsis thaliana*. *Plant and Cell Physiology* **255**: 937–1944.

Matzke AM, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Review Genetics* **15**: 394–408.

Meyberg R, Perroud PF, Haas FB, Schneider L, Heimerl T, Renzaglia KS, Rensing SA. 2020. Characterisation of evolutionary conserved key players affecting eukaryotic flagellar motility and fertility using a moss model. *New Phytologist* **227**: 440–454.

Modrzejewski D, Hartung F, Lehnert H, Sprink T, Kohl C, Keilwagen J, Wilhelm R. 2020. Which factors affect the occurrence of off-target effects caused by the use of CRISPR/Cas: a systematic review in plants. *Frontiers in Plant Science* **11**: e574959.

Morales-Ruiz T, Ortega-Galisteo AP, Ponfarrada-Marín MI, Martínez-Macías MI, Ariza RR, Roldán-Arjona R. 2006. *DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. *Proceedings of the National Academy of Sciences, USA* **103**: 6853–6858.

Mosquna A, Katz A, Decker EL, Rensing SA, Reski R, Ohad N. 2009. Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. *Development* **136**: 2433–2444.

Nagano H, Fukudome A, Hiraguri A, Moriyama H, Fukuhara T. 2014. Distinct substrate specificities of *Arabidopsis* DCL3 and DCL4. *Nucleic Acids Research* **42**: 1845–1856.

Ning J, Otto TD, Pfander C, Schwach F, Brochet M, Bushell E, Goulding D, Sanders M, Lefebvre PA, Pei J *et al.* 2013. Comparative genomics in *Chlamydomonas* and *Plasmodium* identifies an ancient nuclear envelope protein family essential for sexual reproduction in protists, fungi, plants, and vertebrates. *Genes and Development* **27**: 1198–1215.

Nishikawa SI, Yamaguchi Y, Suzuki C, Yabe A, Sato Y, Kurihara D, Sato Y, Susaki D, Higashiyama T, Maruyama D. 2020. *Arabidopsis* GEX1 is a nuclear membrane protein of gametes required for nuclear fusion during reproduction. *Frontiers in Plant Science* **11**. doi: 10.3389/fpls.2020.548032.

Nonomura KI, Miyoshi K, Eiguchi M, Suzuki T, Miyao A, Hirochika H, Kurata N. 2003. The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *Plant Cell* **15**: 1728–1739.

Okano Y, Aono N, Hiwatashi Y, Murata T, Nishiyama T, Ishikawa T, Kubo M, Hasebe M. 2009. A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. *Proceedings of the National Academy of Sciences, USA* **106**: 16321–16326.

Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada F-P. 2010. Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* **464**: 628–632.

Paro R, Grossniklaus U, Santoro R, Wutz A. 2021. Epigenetics and metabolism. In: *Introduction to epigenetics. Learning materials in biosciences*. Cham, Switzerland: Springer, 179–201.

Perroud P-F, Haas FB, Hiss M, Ullrich KK, Alboresi A, Amirebrahimi M, Barry K, Bassi R, Bonhomme S, Chen H *et al.* 2018. The *Physcomitrella patens* gene atlas project: large-scale RNA-seq based expression data. *The Plant Journal* **95**: 168–182.

Pillot M, Baroux C, Arteaga Vazquez M, Autran D, Leblanc O, Vielle-Calzada JP, Grossniklaus U, Grimanelli D. 2010. Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in *Arabidopsis*. *Plant Cell* **22**: 307–320.

Rawat A, Brejšková L, Hála M, Cvrčková F, Žárský V. 2017. The *Physcomitrella patens* exocyst subunit EXO70.3d has distinct roles in growth and development, and is essential for completion of the moss life cycle. *New Phytologist* **216**: 438–454.

Robinson MD, McCarthy DJ, Smyth GK. 2010. EDGER: a BIOCONDUCTOR package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25.

Rotman N, Durbarry A, Wardle A, Yang WC, Chaboud A, Faure JE, Berger F, Twell D. 2005. A novel class of MYB factors controls sperm-cell formation in plants. *Current Biology* **15**: 244–248.

Rövekamp M, Bowman JL, Grossniklaus U. 2016. *Marchantia* Mp*RKD* regulates the gametophyte–sporophyte transition by keeping the egg cells quiescent in the absence of fertilization. *Current Biology* **26**: 1782–1789.

Sakakibara K, Nishiyama T, Deguchi H, Hasebe M. 2008. Class 1 *KNOX* genes are not involved in shoot development in the moss *Physcomitrella patens* but do function in sporophyte development. *Evolution and Development* **10**: 555–566.

Sakakibara K, Reisewitz P, Aoyama T, Friedrich T, Ando S, Sato Y, Tamada Y, Nishiyama T, Hiwatashi Y, Kurata T *et al.* 2014. *WOX13-like* genes are required for reprogramming of leaf and protoplast cells into stem cells in the moss *Physcomitrella patens*. *Development* **141**: 1660–1670.

Sanchez VV, Shekar Kenchappa C, Landberg K, Bressendorff S, Schwarzbach S, Martin T, Mundy J, Petersen M, Thelander M, Sundberg E. 2017. Autophagy is required for gamete differentiation in the moss *Physcomitrella patens*. *Autophagy* **13**: 1939–1951.

Satish M, Nivya MA, Abhishek S, Nakarakanti NK, Shivani D, Vani MV, Rajakumara E. 2018. Computational characterization of substrate and product specificities, and functionality of *S*-adenosylmethionine binding pocket in histone lysine methyltransferases from Arabidopsis, rice and maize. *Proteins* **86**: 21–34.

Schmid MW, Giraldo-Fonseca A, Rövekamp M, Smetanin D, Bowman JL, Grossniklaus U. 2018. Extensive epigenetic reprogramming during the life cycle of *Marchantia polymorpha*. *Genome Biology* **19**: 9–25.

Schmidt A, Wuest SE, Vijverberg K, Baroux C, Kleen D, Grossniklaus U. 2011. Transcriptome analysis of the Arabidopsis megaspore mother cell uncovers the importance of RNA helicases for plant germline development. *PLoS Biology* **9**: e1001155.

Schoft VK, Chumak N, Choi Y, Hannon M, Garcia-Aguilar M, Machlicova A, Slusarx L, Moisiolek M, Park J-S, Park GT *et al.* 2011. Function of the DEMETER DNA glycosylase in the *Arabidopsis thaliana* male gametophyte. *Proceedings of the National Academy of Sciences, USA* **108**: 8042–8047.

Sharma V, Clark AJ, Kawashima T. 2021. Insights into the molecular evolution of fertilization mechanism in land plants. *Plant Reproduction* **34**: 353–364.

Sheridan WF, Avalkina NA, Shamrov II, Batygina TB, Golubovskaya IN. 1996. The *mac1* gene: controlling the commitment to the meiotic pathway in maize. *Genetics* **142**: 1009–1020.

Song Q, Ando A, Jiang N, Ikeda Y, Chen ZJ. 2020. Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in *Arabidopsis* female gametophytes. *Genome Biology* **21**: e178.

Speijer D, Luke J, Eliá M. 2015. Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. *Proceedings of the National Academy of Sciences, USA* **112**: 8827–8834.

Sprunck S, Rademacher S, Vogler F, Gheyselinck J, Grossniklaus U, Dresselhaus T. 2012. Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* **338**: 1093–1097.

Sprunck S, Urban M, Strieder N, Lindemeier M, Bleckmann A, Evers M, Hackenberg T, Möhle C, Dresselhaus T, Engelmann JC. 2019. Elucidating small RNA pathways in *Arabidopsis thaliana* egg cells. *bioRxiv* doi: 10.1101/525956.

Takanashi H, Ohnishi T, Mogi M, Hirata Y, Tsutsumi N. 2011. *DCL2* is highly expressed in the egg cell in both rice and *Arabidopsis*. *Plant Signaling and Behavior* **6**: 604–606.

Tanahashi T, Sumikawa N, Kato M, Hasebe M. 2005. Diversification of gene function: homologs of the floral regulator *FLO/LFY* control the first zygotic cell division in the moss *Physcomitrella patens*. *Development* **132**: 1727–1736.

Trouiller B, Schaefer DG, Charlot F, Nogué F. 2006. *MSH2* is essential for the preservation of genome integrity and prevents homeologous recombination in the moss *Physcomitrella patens*. *Nucleic Acid Research* **34**: 232–242.

Vashisht D, Nodine MD. 2014. MicroRNA functions in plant embryos. *Biochemical Society Transactions* **42**: 352–357.

Veiseth SV, Rahman MA, Yap KL, Fischer A, Egge-Jacobsen W, Reuter G, Zhou MM, Aalen RB, Thorstensen T. 2010. The SUVR4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in *Arabidopsis*. *PLoS Genetics* **7**: e1001325.

Wang CJ, Nan GL, Kelliher T, Timofejeva L, Vernoud V, Golubovskaya IN, Harper L, Egger R, Walbot V, Cande WZ. 2012. Maize *multiple archesporial cells1* (*mac1*), an ortholog of rice *TDL1A*, modulates cell proliferation and identity in early anther development. *Development* **139**: 2594–2603.

Wilson ZA, Morroll SM, Dawson J, Swarup R, Tighe PJ. 2001. The *Arabidopsis MALE STERILITY1* (*MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *The Plant Journal* **28**: 27–39.

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenführer J, von Mering C, Grossniklaus U. 2010. *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Current Biology* **20**: 506–512.

Xie M, Ren G, Costa-Nunes P, Pontes O, Yu B. 2012. A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. *Nucleic Acids Research* **10**: 4422–4431.

Yamamoto M, Maruyama D, Endo T, Nishikawa S. 2008. *Arabidopsis thaliana* has a set of J proteins in the endoplasmic reticulum that are conserved from yeast to animals and plants. *Plant and Cell Physiology* **49**: 1547–1562.

Yamaoka S, Nishihama R, Yoshitake Y, Ishida S, Inoue K, Saito M, Okahashi K, Bao H, Nishida H, Yamaguchi K *et al.* 2018. Generative cell specification requires transcription factors evolutionarily conserved in land plants. *Current Biology* **28**: 479–486.

Yang C, Vizcay-Barrena G, Conner K, Wilson ZA. 2007. *MALE STERILITY1* is required for tapetal development and pollen wall biosynthesis. *Plant Cell* **19**: 3530–3548.

Yu X, Zhang X, Zhao P, Peng X, Chan H, Bleckmann A, Bazhenova A, Shi C, Dresselhaus T, Sun M-X. 2021. Fertilized egg cells secrete endopeptidases to avoid polytubey. *Nature* **592**: 433–437.

Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH, Zwaenepoel A, Ma H, Goffinet B, Guan YL *et al.* 2020. The hornwort genome and early land plant evolution. *Nature Plants* **6**: 107–118.

Zhao X, de Palma J, Oane R, Gamuyao R, Luo M, Chaudhury A, Hervé P, Xue Q, Bennett J. 2008. OsTDL1A binds to the LRR domain of rice receptor kinase MSP1, and is required to limit sporocyte numbers. *The Plant Journal* **54**: 375–387.

Zhao DZ, Wang GF, Speal B, Ma H. 2002. The *EXCESS MICROSPOROCYTES1* gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther. *Genes & Development* **16**: 2021–2031.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Sampling of egg cells, archegonium and shoot tissues and data quality.

Fig. S2 Overview of reporter constructs.

Fig. S3 Illustration of RNA-seq read mapping from selected tissues obtained using different extraction and library preparation methods.

Fig. S4 Illustration of RNA-seq read mapping from egg cells for known egg expressed genes.

Methods S1 Plant material and growth conditions.

Methods S2 Generation and verification of transgenic lines.

Methods S3 RT-qPCR to determine *PpBNB* transcript abundance.

Methods S4 Light microscopy and *Ppbnb* mutant phenotyping.

Methods S5 Confocal microscopy and *PpBNB* reporter gene analysis.

Table S1 Summary of RNA-seq mapping.

Table S2 Characteristics of crRNAs in gRNA-expressing plasmids.

Table S3 Characteristics of knock-out and knock-in lines obtained by CRISPR-CAS9 gene editing.

Table S4 Primers used in this study.

Table S5 RNA-seq libraries (Perroud *et al.*, 2018) included in statistical analyses.

Table S6 Expression data and statistical evaluation of gene differentially expressed genes in egg cells.

Table S7 Literature validation of the *P. patens* egg cell transcriptome.

Table S8 Gene ontology (GO) term enrichment table showing the significantly enriched biological processes based on significantly overexpressed genes in the *P. patens* egg cell that are also expressed in the Arabidopsis egg cell.

Table S9 Gene ontology (GO) term enrichment table showing the significantly enriched biological processes based on significantly overexpressed genes in the *P. patens* egg cell that are not expressed in the Arabidopsis egg cell.

Table S10 Gene ontology (GO) term enrichment table showing the significantly enriched biological processes based on significantly overexpressed genes in the *P. patens* egg in each of the four clusters presented in Fig. 1.

Table S11 Homologs of genes that are involved in RNA-based silencing are active in the *P. patens* egg cell.

Table S12 Homologs of genes that are important for karyogamy in Arabidopsis are active in the *P. patens* egg cell.

Table S13 Homologs of genes that are important in ammonium transport in Arabidopsis are active in the *P. patens* egg cell.

Table S14 Homologs of genes that are male specific in Arabidopsis are active in the *P. patens* egg cell.

Table S15 Total number of antheridia formed per shoot.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.