

The post-fertilization archegonium gene network of seedless plants contributed to the origin of seeds

Received: 29 June 2025

Accepted: 9 June 2026

Cite this article as: Plackett, A.R., Catoni, M. The post-fertilization archegonium gene network of seedless plants contributed to the origin of seeds. *Nat Commun* (2026). <https://doi.org/10.1038/s41467-026-74577-w>

Andrew R. G. Plackett & Marco Catoni

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

The post-fertilization archegonium gene network of seedless plants contributed to the origin of seeds

Andrew R.G. Plackett^{1*}, Marco Catoni¹

¹University of Birmingham, School of Biosciences, Edgbaston Campus, Birmingham, B15 2TT, United Kingdom.

*Corresponding author. A.R.G.Plackett@bham.ac.uk

ARTICLE IN PRESS

ABSTRACT

The evolution of seeds was a critical transition in plant evolution, but how ancestral seedless development was modified to form the first seed is only partly-answered through comparative morphology or the fossil record. We investigate seed origins by quantifying gene network conservation between seeds and seedless plant reproductive organs. Characterizing reproduction in the homosporous fern *Ceratopteris richardii* as a proxy for ancestral seedless reproduction, we create a gene expression atlas of fern diploid and haploid reproductive organ development and test for enrichment of these genes in *Arabidopsis thaliana*. Here we show ovule gene networks are enriched in genes regulating post-fertilization development of the fern haploid egg chamber (archegonium), a subset of which also show enrichment in sporophyll primordia, suggesting a mechanistic model to explain the origin of the ovule from a sporangium-bearing axis whereby gene networks in the ancestral post-fertilization archegonium became heterotopically expressed during development of the sporangium-bearing axis.

INTRODUCTION

Plants have not always made seeds, and the evolution of seed-based reproduction had a transformative impact on the Earth's terrestrial biosphere. But how plants first evolved seeds has been a longstanding and intractable question. All living seed plants (gymnosperms and angiosperms) and seeds are inferred to have a single common ancestor^{1,2} arising from within the seedless vascular plants, with monilophytes (horsetails and ferns, hereafter 'ferns') the closest surviving outgroup^{2,3}. The evolution of seed-based reproduction involved complex developmental modifications that resulted in a free-living stage of the plant lifecycle (the haploid gametophyte) becoming entirely dependent on another (the diploid sporophyte), in a reversal of the ancestral bryophyte lifecycle⁴, but how these occurred remains unclear.

Ovule patterning (the pre-fertilization stage of seed development) is ancient and conserved across extant and extinct seed plant lineages, recognisable from Palaeozoic seed ferns onwards: a diploid nucellus apically specialised for pollen reception, containing a multicellular haploid egg-producing structure, with a subtending chalaza from which one or more integuments arise⁵. After fertilization, this ‘body plan’ undergoes further developmental changes to form the seed coat, provide nutrients to the embryo from storage tissues (the nucellus in some gymnosperms, the endosperm in angiosperms)^{6,7,8} and impose dormancy on the developing embryo⁹. Developmental changes involved in the ovule/seed’s origin have been deduced primarily through comparisons to ancestral vascular plant reproduction still found in extant ferns, where separate, free-living diploid and haploid ‘generations’ both form reproductive organs^{10,11}. The diploid sporophyte forms sporangia (generating haplo single-celled haploid spores by meiosis) borne on fern lateral organs (fronds). Evidence increasingly supports the independent evolution of fern fronds and seed plant lateral organs (leaves)^{12,13}, but this does not preclude their homology with ancestral shootlike structures (‘axes’). Notably, fronds are more shoot-like than leaves, developing through a hierarchy of apical cells¹⁴, resulting in the characteristic fern ‘fiddlehead’. Fern spores germinate into thalloid gametophytes that produce gametangia by mitosis: antheridia (containing free-swimming sperm) and archegonia (egg-bearing chambers).

From such comparisons it is inferred that early ovule evolution involved the invention of heterospory, i.e. ‘male’ and ‘female’ sporangia (anther and ovule nucellus) producing a genetically-predetermined male gametophyte (pollen) and female gametophyte (FG), from an ancestral homosporous sporangium¹⁵. Many extant ferns remain homosporous, including the genetic model *Ceratopteris richardii*¹⁶. Although gametophyte morphologies and reproductive strategies are diverse in extant ferns¹⁷, *Ceratopteris* default gametophyte development is

hermaphroditic¹⁸. 'Male' (micro)sporogenesis in seed plants seemingly resembles ancestral sporogenesis whereas 'female' (mega)sporogenesis has apparently undergone significant modification¹⁹, now typically producing a single viable megaspore (monosporous)²². A second crucial early innovation was gametophyte encapsulation inside a sporangium (endosporous)²⁰ and developmental reduction, so extreme in angiosperms that archegonia are considered absent or potentially equivalent to the entire female gametophyte (FG)²¹. However, the ovule is more than the sum of its parts: ovule-specific developmental processes have evolved since its origin that do not have clear parallels, including apical specialisation for pollen capture and the endosperm as a nutrient storage tissue⁵. The origin of seeds thus entailed a complex suite of changes to ancestral seedless vascular plant development, but how development became reprogrammed to achieve this, and how novel developmental processes arose from existing gene networks, remains unclear.

Comparative fossil studies have proposed that the ovule evolved through modification of a sporangium-bearing shoot and progressive sterilization and reduction of branches to form the integument²⁰. However, genetic analyses of extant ovule development, highlighting strong similarity between early ovule and meristematic development together with the modularity and lability of gene networks, challenge this view²³, raising the possibility that integuments arose *de novo* and thus the basis for morphological comparisons may not be robust. Disparities between morphology-based and genetic-based analyses underly similar ambiguity over the evolutionary relationships between gymnosperms and angiosperms²⁴. Even if correct, this model does not explain in detail how underlying gene networks were altered to change developmental patterning and other possibilities remain to be explored: like seeds, fern archegonia demonstrate fertilization-dependent developmental modifications to support embryo development, including post-fertilization growth of the envelope (calyptra) and the transfer of nutrients to the embryo via a placenta²⁵. Fern sporophyte and gametophyte reproductive organs

thus both possess seed-like characteristics, but archegonia were present within early Palaeozoic seeds and remain in most gymnosperm seeds⁵, seemingly precluding their incorporation into modifications leading to the ovule body-plan. The genetic changes underlying these evolutionary transitions remain unclear but could be crucial to independently testing these hypotheses.

A complete molecular explanation for the origin of the ovule has not been proposed previously. A role for the MADS-box gene family, which specifies angiosperm ovule identity^{26,27} has long been suggested from indirect evidence^{28,29}. Similarly, a seed maturation module centred on the gene *LEAFY COTYLEDON1* (*LEC1*) has been proposed to have originated in ferns³⁰. Other studies of individual fern genes to-date suggest at least partial conservation of reproductive gene networks between seedless and seed-bearing vascular plants^{31,32} but to-date the extent of this conservation has not been quantified. While fern genomic information is increasingly available³³⁻³⁷ as is transcriptome data at either whole-lifecycle scale³⁸⁻⁴⁰ or for single reproductive organs^{34,41}, a detailed characterisation of gene networks across fern reproductive development to address this question is still lacking. We thus undertook mRNA-seq analysis to identify genes associated with reproductive processes in both the *Ceratopteris* sporophyte and gametophyte and compared these with *Arabidopsis thaliana* as the most detailed model available of seed development.

In this work we find significant conservation between *Ceratopteris* and *Arabidopsis* reproductive gene networks, identifying a conserved role for the phytohormone gibberellin in shoot reproductive transitions and conserved genes underlying microsporogenesis and fern homosporous sporogenesis. We further identify significant gene conservation between the ovule and the fern post-fertilization archegonium that cannot be attributed to the female gametophyte

within the ovule, suggesting that co-option of this gametophytic organ's gene networks may have contributed to the ovule's evolutionary origin.

RESULTS

Identifying *Ceratopteris* sporophyte reproductive genes

Ceratopteris sporophyte development was first characterised under tissue culture conditions (**Supplementary Fig. 1**). As under soil-grown conditions, successive fronds became increasingly dissected in a heteroblastic sequence until a transition to reproductive development ('sporing') occurred. The *Ceratopteris* shoot was found to spore consistently and reproducibly measured by time (100% by 18 weeks) and development (20.6 ± 0.07 fronds) (**Fig. 1a-c; Supplementary Fig. 2a,b; Supplementary Data 1**). To identify transcriptional changes specifically associated with reproductive development, using mRNA-seq we compared gene expression across frond development between sporophytes in the sporing phase (130 days old) (**Fig. 1d-i**) and the vegetative phase immediately prior to sporing (60 days old) (**Fig. 1j-o; Supplementary Fig. 2c-e**), with 25-day-old apex and primordia samples included as an early vegetative control (**Supplementary Fig. 3a-e**). Frond development was divided into sequential stages common to vegetative fronds and sporophylls: the shoot apical region ('Ap_x') (**Fig. 1e,k**); frond primordium ('Pr_m') (**Fig. 1f,l; Supplementary Fig. 3f-q**); fiddlehead ('Fid') (**Fig. 1g,m**); expanding frond ('Exp') (**Fig. 1h,n**) and mature frond ('Mat') (**Fig. 1i,o**). Sporangia were already visible by the start of fiddlehead unfurling (**Supplementary Fig. 4a-l**), and sporangium development could not be classified into distinct stages based on frond size (**Supplementary Fig. 4m,n**) except that post-meiotic sporangia were only observed in full-expanded (Mat) samples (**Supplementary Fig. 4e-h**).

Sample clustering by gene expression using Principal Component Analysis (PCA) (**Supplementary Fig. 5a**) and correlation of mRNA-seq data (**Supplementary Fig. 6**) confirmed

that clustering reflected frond developmental and reproductive changes. Frond stages separated progressively along PC1, with Apx and Prm samples closely overlapping except for 25d Prm, which clustered more closely with Fid samples. The similarity of Apx and Prm samples is consistent with the common presence of primordia. Within each frond developmental stage, vegetative and reproductive phases separated across PC2. Differential expression (DE) analysis identified 2098 transcripts with increased abundance ($P_{\text{adj}} < 0.05$, $\log_2\text{foldchange} > 1$) between sporophylls and vegetative samples within at least one frond stage (**Fig. 1p**; **Supplementary Data 2**), 38.5% identified in more than one stage (35.3% from consecutive stages). Among these reproductive candidate genes we categorised 295 present (at a threshold of FPKM > 1) only in the sporophyll ('sporophyll-specific'), 1437 present in the sporophyll and at least one vegetative stage ('sporophyll-enriched') and 366 candidates expressed at FPKM < 1 in all sporophyll samples (**Fig. 1q**). These expression categories represent potential direct regulators of reproductive processes, genes either with dual vegetative and reproductive function or frond genes whose expression is modified in sporophylls, and genes with either low or tissue-specific expression, respectively. We compared these candidates to genes with increasing abundance across frond development ($P_{\text{adj}} < 0.05$, $\log_2\text{foldchange} > 1$ between any two frond stages, comparing from earlier to later development), identifying differential regulation specifically in 130d fronds (candidate sporophyll regulators) and co-regulation between 60d and 130d frond stages (frond developmental regulators) (**Supplementary Data 3**). 62.8% of our reproductive candidates were represented in one or both of these classes (**Supplementary Fig. 7**), with sporophyll-specific and sporophyll-FPKM < 1 candidates showing significant enrichment ($P_{\text{adj}} < 0.05$) in sporophyll developmental regulators compared to the full dataset, whereas sporophyll-enriched candidates showed significant enrichment in frond developmental regulators. Our identification of candidate reproductive genes and the three expression categories were thus supported as robust.

Gibberellin promotes the *Ceratopteris* sporing transition

Genes involved in apex function were identified by comparing expression between Apx (shoot apex and primordia) and Prm (primordia only) samples, comparing between different shoot ages to identify genes associated with the shoot sporing transition. DE analysis identified 51 sporing transition candidates with significantly greater abundance ($P_{\text{adj}} < 0.05$) at 130d compared to 60d and 25d (**Fig. 2a**). Genes promoting a reproductive transition similar to angiosperm shoots were further predicted to increase expression immediately prior to the transition, and to be preferentially expressed in the shoot apex vs primordia. Thirteen sporing candidates showed progressive significant increases in abundance ($P_{\text{adj}} < 0.05$, $\log_2\text{fold} > 1$) between 25d, 60d and 130d Apx samples, including the previously-predicted MADS-box reproductive gene *CMADS1*²⁸, but this was present at FPKM > 1 in all ages and not significantly more abundant in Apx vs Prm samples ($P_{\text{adj}} > 0.05$) (**Fig. 2b**) while two other MADS-box homologs identified as sporing candidates showed increased abundance at 130d only (**Supplementary Data 2**).

Of two sporing candidates whose expression profile closely matched that predicted for sporing regulators, one was annotated as a *GA 20-OXIDASE*⁴² (**Fig. 2b**), supported by its phylogenetic position in a fern outgroup to seed plant *GA20ox* orthologs (**Supplementary Fig. 8**). Sporing candidates further included a putative homolog of *RGA TARGET 1* (*RGAT1*), a direct target of GA signalling in *Arabidopsis* pollen development⁴³, whereas a putative *GA 3-OXIDASE* biosynthesis gene significantly decreases in abundance ($P_{\text{adj}} < 0.05$) at 130d (**Supplementary Data 2**), a known homeostatic response to GA signalling⁴⁴. These results suggested a role for GA in promoting *Ceratopteris* sporing. We tested this experimentally through exogenous application of GA_3 , which significantly accelerated the sporing transition ($p < 0.05$) (**Fig. 2c**). These did not phenocopy the frond dissection of natural sporophylls (**Fig. 2d-g**), suggesting that the sporing transition network is separable from frond heteroblasty. In a reciprocal test we applied the canonical GA biosynthesis inhibitor paclobutrazol (PBZ)⁴⁵ to sporophytes from 4

weeks onwards. In contrast to mock-treated controls, frond morphology reverted to simple spade-shaped fronds (**Fig. 2h,i**) but a morphological phase transition was still observed whereby new fronds were elongated, reminiscent of sporophyll pinnae (**Fig. 2j**) and bore apparent immature sporangia in 7/20 sporophytes examined. This phase transition was significantly delayed compared to sporangia under mock treatment both chronologically and developmentally (**Fig. 2k,l**). These results are consistent with a conserved function for GA promoting the shoot reproductive transition between ferns and angiosperms, separably from a potential role in frond morphogenesis.

Sporophyll reproductive genes include sporogenesis orthologs

54.6% of candidate genes possessed GO term annotations (**Fig. 3a**) and were significantly enriched in these ($P_{\text{adj}} < 0.05$) compared to the whole genome (**Supplementary Data 5**). GO term analysis identified significant enrichment ($P < 0.05$) predominantly of processes including metabolism, nutrition and stress response (**Fig. 3b**). Reproductive GO terms were not enriched in all candidates but were so in the sporophyll-specific expression category, relating to meiosis (**Supplementary Fig. 9**). Hierarchical clustering analysis of reproductive candidates by normalised expression identified seven expression clusters exhibiting maximum abundance in distinct stages across sporophyll development (**Fig. 3c**; **Supplementary Data 4**). Meiotic GO terms were significantly enriched ($p < 0.05$) in mature sporophylls (cluster 3) (**Fig. 3d**), consistent with our developmental characterization (**Supplementary Fig. 4**) and the only cluster significantly enriched in sporophyll-specific candidates ($P_{\text{adj}} < 0.05$; **Supplementary Data 4**). Meiotic GO terms resolved to a single gene annotated as an ortholog of *DISRUPTED MEIOTIC CDNA1 (DMC1)* (**Supplementary Data 5**), a eukaryotic-conserved regulator of meiosis⁴⁶. However, reproductive candidates included further annotated orthologs of regulators of meiotic recombination (*MSH4*, *ZIP4*)^{47,48}, sporogenous cell fate (*EXS/EMS*, *TPD1*) and post-meiotic

pollen development (*AMS1*, *MS1* and *MS2*)⁴⁹. These suggest that our candidate dataset has successfully captured the sporogenesis pathway in *Ceratopteris*.

Beyond sporogenesis, GO term enrichment analysis identified potential mechanisms regulating sporophyll development. GO terms relating to hormone response were significantly enriched ($P < 0.05$) in sporophyll candidates (**Fig. 3a**), auxin-related terms significantly enriched in mature sporophylls (cluster 5) and jasmonate (JA)-related terms enriched in sporophyll primordia (cluster 7) amongst others (cluster 6) (**Fig. 3d**). Sporophyll primordia were specifically enriched in GO terms relating to protein phosphorylation (cluster 1) and gene regulation (clusters 1 and 7), including two MADS-box homologs (Ceric.05G029400, Ceric.34G033600; **Supplementary Data 5**). Comparing GO term enrichment between sporophyll differentially-regulated genes and frond co-regulated genes found that hormone GO terms were enriched specifically in sporophyll development whereas phosphorylation was enriched in both (**Supplementary Fig. 10**).

Remarkably, clusters with maximum expression during the fiddlehead (cluster 2) and expanding frond stages (cluster 4) were not enriched in recognisable regulatory pathways, enriched GO terms instead relating to metabolism (**Supplementary Data 5**). This was not due to poor annotation because these clusters demonstrated a similar proportion of annotated genes as the whole genome ($P_{\text{adj}} > 0.05$) as did clusters 5, 6 and 7 (**Supplementary Data 5**). Collectively, these findings suggest that sporophyll-specific development is regulated at least in part through activating reproduction-specific regulatory mechanisms to modify frond regulatory pathways, primarily during early organ development and at the onset of sporogenesis.

Identifying *Ceratopteris* gametophyte reproductive genes

The *Ceratopteris* hermaphrodite gametophyte (from hereon ‘gametophyte’) develops multiple archegonia from a ‘notch’ meristem while antheridia develop at the thallus margin (**Fig. 4a**). Archegonia are fertilized by free-swimming sperm²⁵ released in the presence of water. Green

when unfertilized (**Fig. 4b**), all archegonia turn brown within one day of being flooded (**Fig. 4c,d**) but typically only a single embryo develops (**Fig. 4e,f**). Fertilization and embryo development were experimentally manipulated by restricting water availability: by 14 days after sowing (0 DAF) $77.5 \pm 12.5\%$ of well-watered controls bore embryos at various developmental stages (**Supplementary Fig. 11**) compared to $1.0 \pm 1.0\%$ under restricted water (**Fig. 4g**). Flooding of restricted-water populations successfully triggered fertilization, with $83.0 \pm 5.74\%$ bearing visible embryos by 7 DAF compared to $5.5 \pm 2.2\%$ without flooding. Under this system most embryos became visible 3-4 DAF (**Fig. 4h**) and showed closer developmental synchronization (**Supplementary Fig. 11**), which was further improved by restricting water after initial flooding (**Fig. 4i**). The notch meristem also exhibited fertilization-dependent shut down of archegonium production by 2 DAF (**Supplementary Fig. 12; Supplementary Data 6**).

To identify genes relating to archegonium formation we dissected out the notch region ('Ntc') from the gametophyte (**Fig. 4j-m**) and used the surrounding thallus ('Thl') as a vegetative control (**Fig. 4n-q**). To further capture genes involved in fertilization and post-fertilization development these samples were compared between unflooded gametophytes at 14 DAS (**Fig. 4j,n**) and at 1 (**Fig. 4k,o**), 2 (**Fig. 4l,p**) and 4 DAF (**Fig. 4m,q**). The 4 DAF notch was subdivided into embryo-bearing archegonia (Aeb) and embryo-less archegonia (Arc) (**Fig. 4m**). To discriminate embryo from archegonium gene networks, we isolated 4 DAF embryos as an additional control ('Dem') (**Supplementary Fig. 3r-w**). PCA analysis (**Supplementary Fig. 5b**) and sample correlation (**Supplementary Fig. 13**) found that biological replicates clustered more closely within each sample than between different developmental stages except for 1 DAF and 2 DAF (which overlapped within the Ntc and Thl tissue types). PC1 separated Ntc from Thl and DAF separated progressively along PC2 (except for Dem). A total of 7817 reproductive candidate genes were identified with increased abundance in Ntc compared to Thl within at least one developmental stage (**Fig. 4r; Supplementary Data 7**). Similar to the sporophyll,

candidate genes were categorised as 'notch-specific' (1435), 'notch-enriched' (5057) and 'notch FPKM<1' (1325) (**Fig. 4s**), representing potential direct reproductive regulators, downstream developmental gene targets and low or tissue-specific expression respectively. Comparing these to genes demonstrating differential upregulation specific to the notch during development (**Supplementary Fig. 14**), 35.95% of notch reproductive candidates were represented within this dataset, with the notch-specific and notch-FPKM1 expression categories significantly enriched ($P_{\text{adj}} < 0.05$) compared to all candidates whereas the notch-enriched category was significantly enriched in genes showing co-regulation between the notch and thallus (**Supplementary Data 3**). The successful detection of robust candidate notch reproductive genes was thus supported.

Notch reproductive genes include seed development orthologs

52.1% of notch candidates possessed GO term annotation (**Fig. 5a**) significantly enriched ($P_{\text{adj}} < 0.05$) compared to the whole genome (**Supplementary Data 5**). 24.2% of significantly-enriched GO terms ($p < 0.05$) related to the cell cycle or cell division (**Fig. 5b**), in contrast to sporophyll candidates (3.0%). 17.9% of notch candidate enriched GO terms related to gene regulation, including post-transcriptional and post-translational categories. Our candidates thus reflect the notch's status as an actively-dividing tissue and suggest multiple regulatory mechanisms are involved in its development.

Hierarchical clustering of notch candidates by normalised expression successfully identified six clusters with greatest transcript abundance in distinct notch developmental stages (**Fig. 5c**). Cluster 4, with expression greatest in the pre-fertilization notch, was significantly enriched in notch-specific candidates ($P_{\text{adj}} < 0.05$; **Supplementary Data 4**) and GO terms relating to the regulation of gene expression and negative regulation of reproduction (**Fig. 5d**). This revealed putative orthologs of the *TFL1-like* PEBP subfamily (Ceric.19G00340) and the *SIN3* family of

transcription factors (Ceric.02G103600, Ceric.02G103700; **Supplementary Data 5**) which regulate the timing of *Arabidopsis* flowering^{50,51}. Cluster 6, where expression peaks immediately after fertilization and during notch meristem shutdown, was also significantly enriched for GO terms regulating gene expression, revealing a putative ortholog of the HB21/40/53 transcription factors (Ceric.07G084400) which in *Arabidopsis* trigger inflorescence meristem arrest⁵². These similarities suggest that fern notch meristem gene networks could be conserved with seed plant reproductive meristem gene networks.

GO terms for the regulation of cell division and gene expression via post-transcriptional and post-translational regulatory mechanisms (noncoding RNA, protein phosphorylation, histone and DNA methylation) were enriched specifically in clusters with expression greatest in the Aeb (clusters 2 and 3) (**Fig 5d**), with cluster 3 also significantly enriched in notch-specific candidates and genes differentially-upregulated specifically during notch development (**Supplementary Data 4**). The genes associated with these GO terms include multiple putative orthologs of *Arabidopsis* genes with functions across ovule and seed development (**Supplementary Table 1**), including integument development, megaspore mother cell (MMC) specification, FG development, embryogenesis and endosperm development. Surprisingly, many were annotated as orthologs of genes with functions in *Arabidopsis* meiosis (*MSH2*, *MSH7*, *SAE2*, *RAD51C*, *TOP43A*, *RFC1*)⁵³⁻⁵⁸. A proportion of these enriched GO terms could be ascribed to transcripts with significantly greater abundance ($P_{\text{adj}} < 0.05$, $\log_2\text{fold} > 1$) in Dem vs Aeb samples (**Fig. 5e**). Two further candidate genes in cluster 3 (Ceric.31G036600, Ceric.33G031700) were resolved by sequence similarity as previously-published *Ceratopteris* embryo genes *CrANT*⁵⁹ and *CrLFY1*⁶⁰ (**Supplementary Fig. 15**), supporting the accuracy of this cluster. Our findings thus suggest significant conservation between gene networks of the ovule/seed and post-fertilization archegonium development.

In further support of this hypothesis, our RNA-seq analysis successfully predicted two regulatory mechanisms in *Ceratopteris* archegonium development similar to seeds. During angiosperm seed development sugar signalling via trehalose 6-phosphate (T6P) promotes seed filling and inhibits post-fertilization abortion⁶¹⁻⁶³. In *Ceratopteris* we found significant enrichment ($P < 0.05$) of carbohydrate metabolism during notch development immediately after fertilization (cluster 6) (**Fig. 6a**). Disaccharide biosynthesis and metabolism GO terms were also significantly enriched ($P < 0.05$) in notch-specific candidates and genes showing differential up-regulation during notch development (**Supplementary Fig. 16**), with the associated genes annotated as T6P synthase and T6P phosphatase orthologs (**Supplementary Dataset 5**). In contrast, cell wall-related carbohydrate metabolism showed more general enrichment. To test the effect of carbohydrate status on *Ceratopteris* post-fertilization development, gametophytes were flooded with a 2% sucrose solution and the effect on fertilization success compared to mock treatment (**Fig. 6b,c**). The proportion of the gametophyte population successfully bearing embryos was not significantly affected by sucrose ($p = 0.5393$; **Supplementary Data 6**) but the mean number of embryos per fertilized gametophyte was significantly increased ($p < 0.0001$; **Fig. 6d**). Sucrose has previously been shown to promote apogamy in *Ceratopteris* gametophytes⁶⁴ but the additional embryos that developed were confirmed as developing inside archegonia (**Fig. 6c**). Our RNA-seq analysis thus successfully predicted a role for carbohydrate status in *Ceratopteris* archegonium fertilization, which may be conserved with sugar signalling mechanisms during seed development.

Secondly, hormone response GO terms were significantly enriched prior to fertilization (cluster 4), during fertilization (cluster 6), in the 4 DAF embryoless notch (cluster 5) (**Fig. 6e**) and in transcripts with greater expression in Aeb tissues than the dissected embryo (**Fig. 6f**). JA-related GO terms were enriched in the first three of these. To test the function of JA in *Ceratopteris* reproduction, we applied exogenous methyl jasmonate (MeJA) solution at 0 DAF

but found no significant effect of MeJA ($P > 0.05$) on fertilization success or embryo number (**Supplementary Fig. 17**). JA GO term enrichment in clusters 4 and 6 was also consistent with a role in notch meristem activity. To test if this could explain the observed enrichment we applied MeJA to 7 day-old gametophytes (at which age the notch meristem initiates) and found that by 14 days old MeJA-treated gametophytes had significantly fewer archegonia ($P < 0.05$; **Supplementary Fig. 17**) and visibly-reduced thallus size, suggesting a role for JA signalling in more general notch meristem activity. In contrast to JA, an auxin-related GO term was enriched in 4 DAF transcripts with greater expression in Aeb tissues than the dissected embryo (**Fig. 6f**). Auxin is a critical regulator of fertilization-dependent endosperm and seed coat development^{65,66}, sufficiently so to drive parthenocarpic *Arabidopsis* seed development in the absence of fertilization⁶⁷. To test for a role of auxin in *Ceratopteris* archegonium post-fertilization development, exogenous auxin solution was applied to gametophytes at fertilization. At 4 DAF under mock treatment growth of the embryo and archegonium calyptra (envelope) was tightly co-ordinated (**Fig. 6g**) whereas embryo-less archegonia showed no further development compared to pre-fertilization (**Fig. 6h**). Under auxin treatment embryo and calyptra growth was visibly decoupled (**Fig. 6i**), and embryo-less archegonia exhibited apparent ectopic cell divisions resembling post-fertilization calyptra development (**Fig. 6j**). Our results thus support a function for auxin in calyptra development which resembles and may be conserved with auxin seed mechanisms.

The ovule is enriched in *Ceratopteris* notch reproductive orthologs

To quantify the extent of conservation between *Ceratopteris* and *Arabidopsis* reproductive gene networks we undertook a gene orthology-based comparison. *Ceratopteris* reproductive candidates partially overlapped between the sporophyll and notch (44.5% and 11.9%, respectively) (**Fig. 7a**). All *Arabidopsis* gene orthologs of these candidates were identified (**Fig. 7b,c**; **Supplementary Fig. 18**; **Supplementary Data 8**) and the populations of sporophyll

orthologs, notch orthologs and common orthologs were each tested for enrichment within published *Arabidopsis* reproductive datasets⁶⁸⁻⁸⁰ compared to vegetative-specific genes⁷⁶. Enrichment between *Ceratopteris* and *Arabidopsis* datasets occurred within reproduction-enriched candidate orthologs in the sporophyll and notch, accounting for all enrichment detected when testing all orthologs together (with the exception of notch vs embryo sac) with multiple additional significant enrichments detected (**Fig. 7d-i**; **Supplementary Fig. 19**) and the notch-enriched expression category comprising a significantly enriched proportion of candidates with *Arabidopsis* orthologs compared to the whole dataset (**Supplementary Data 9**). No significant enrichment was detected between *Arabidopsis* reproductive datasets and reproduction-specific or reproduction-FPKM<1 expression categories or orthologs of candidates common between notch and sporophyll datasets (**Supplementary Data 9**). As such, downstream analysis focussed on the reproduction-enriched expression category.

Orthologs of sporophyll candidates were significantly enriched ($P_{\text{adj}} < 0.05$) in stamen datasets (**Fig. 7d-e**), but not in pollen, ovule or seed datasets ($P_{\text{adj}} > 0.05$; **Fig. 7f-i**). We also tested for enrichment in sporophyll and frond differentially-regulated genes (**Supplementary Fig. 20**) which further supported conservation between the sporophyll and stamen ($P_{\text{adj}} < 0.05$) but not the ovule ($P_{\text{adj}} > 0.05$). In contrast, notch orthologs were significantly enriched in carpel samples containing ovules (**Fig. 7d**) and multiple independently-derived ovule datasets: whole early (functional megaspores to 4 celled FG) and late ovules (8-celled to mature FG)⁷¹, nucellus and MMC genes from the premeiotic ovule⁷⁸, and the egg, synergid and central cell of the mature female gametophyte⁷⁷ (**Fig. 7g,h**). Notch orthologs also showed significant enrichment in the stamen (**Fig. 7e**), with both sporophyll and notch orthologs enriched in whole intermediate stamen and *spl/nzz*-downregulated datasets, representing genes relating to microsporogenesis^{81,82}. Notch orthologs showed further significant enrichment in pollen,

including *ms1*-downregulated stamen genes in which post-meiotic microspore development fails⁸³ (**Fig. 7e,f**).

Ovule genes are conserved in the post-fertilization archegonium

To determine the developmental origin of these putative conserved networks in *Ceratopteris* reproductive organs, orthologs within each notch and sporophyll pHeatmap cluster were tested for enrichment in *Arabidopsis* reproductive genes. Notch clusters 2 and 3, both showing peak expression in the 4 DAF Aeb, were significantly enriched ($P_{\text{adj}} < 0.05$) in ovule datasets (**Fig. 8a**), whereas clusters 1 (4 DAF Aeb), 3 (4 DAF Arc) and 5 (0 DAF Thl) were significantly enriched in stamen datasets. Within the sporophyll, stamen orthologs were enriched in clusters 3 and 5 (mature sporophyll) and clusters 2 (fiddlehead) (**Fig. 8b**). We tested the robustness of these associations by reciprocal enrichment: the *Ceratopteris* orthologs of *Arabidopsis* genes from the stamen, pollen ovule and FG significantly enriched in the notch, sporophyll and frond co-upregulated tests were identified (**Supplementary Data 10**) and their enrichment in *Ceratopteris* pHeatmap clusters tested against the overall candidate datasets. Notch-conserved *Arabidopsis* stamen genes were confirmed as reciprocally significantly enriched ($P_{\text{adj}} < 0.05$) in clusters 1 and 5 (**Fig. 8c**). Conserved pollen genes were enriched in the same clusters. Notch-conserved ovule genes were similarly confirmed as reciprocally enriched in notch clusters 2 and 3, with genes conserved with only the FG enriched specifically in cluster 3 (**Fig. 8d**). Surprisingly, notch-conserved ovule genes also demonstrated significant cross-enrichment in sporophyll cluster 1, with expression greatest in primordia (**Fig. 8e**), distinct from sporophyll-conserved stamen genes enriched in sporophyll cluster 3 (**Supplementary Fig. 21**). These cross-enriched orthologs include regulators of flowering (FWA)⁸⁴, reproductive meristem development (*MRLK*)⁸⁵, ovule specification (BAM2, BAM3)⁸⁶ and ovule development (BHLH11, AGL80)^{87,88}.

Our analyses thus suggest that gametophyte post-fertilization archegonium gene networks are conserved with the ovule (including premeiotic development), with some of these genes also conserved in sporophyll primordia. Sporophyll and notch reproductive candidates also both demonstrate significant and separate conservation with stamen sporogenous gene networks, respectively localised to late sporophyll development and gametophyte tissues distinct from the embryo-bearing archegonium.

Conserved archegonium genes include ovule sporophytic genes

To resolve the contribution of embryo-bearing archegonium gene networks to ovule development we tested for reproductive enrichment in gene sets specific to or common between ovule tissue and cell types (**Supplementary Fig. 22**). Notch orthologs significantly enriched within ovule tissues showed higher correlation within premeiotic tissues (nucellus and MMC) and mature FG tissues than between them, but still with up to 59% cross-overlap (**Fig. 9a**). Notch orthologs were significantly enriched ($P_{\text{adj}} < 0.05$) in nucellus-specific and MMC-specific premeiotic genes (**Fig. 9b**), FG-specific and FG-common cell types (**Fig. 9c**) and in genes specific to premeiotic tissues compared to the FG (**Fig. 9d**). We also determined that the enrichment in these tissues was distinct from that in whole early and late ovule samples, whose genes partially overlapped (**Fig. 9e**): notch ortholog enrichment within whole ovule samples was found to relate only to early-specific or common genes (**Fig. 9f**), but significant enrichment remained in gene populations absent from the whole ovule data (**Fig. 9g**). 627 *Arabidopsis* putative orthologs of notch reproductive candidates were significantly enriched specifically in the nucellus, of which 32 were also conserved with the sporophyll primordium (including *FWA*, *MRLK*, *BAM2*, *BAM3*; **Supplementary Data 11**). Within genes uniquely conserved between the nucellus and notch we identified *HUA1* and *BEL1*, regulators of carpel and early ovule identity respectively^{89,90}. 136 FG-specific orthologs included *FERTILIZATION INDEPENDENT SEED2* (*FIS2*) which represses precocious endosperm development in the FG⁹¹ and 141 notch

orthologs significantly enriched in all four datasets included *AtMYB65*, highly expressed in the FG central cell to also prevent endosperm division⁹² (**Supplementary Data 11**). Other enriched orthologs included regulators of MMC specification (*EXS/EMS1*, *DRM1*)^{93,94} and cell divisions including female meiosis (*MSH2*, *SAE2*, *RAD51C*)^{53,55,56}. Significant co-enrichment ($P_{\text{adj}} < 0.05$) was also detected between notch and sporophyll orthologs enriched in *Arabidopsis* stamen sporogenous tissue (**Supplementary Data 11**) including the GAMYBs *MYB65* and *MYB101*⁹⁵ and multiple auxin-responsive genes. Post-fertilization archegonium gene networks thus show significant separate and overlapping conservation with sporophytic and gametophytic tissues of the ovule, distinct from that found in the sporophyll primordium.

DISCUSSION

Our RNA-seq analysis identified significant genetic conservation across *Ceratopteris* and *Arabidopsis* reproductive development, consistent with the hypothesis that seed plant reproductive gene networks arose from reproductive gene networks in the last common ancestor with ancestral seedless vascular plants. Multiple predictions arising from this analysis were experimentally validated within *Ceratopteris*, and we independently identified candidate reproductive genes reported in previous individual studies (*CMADS1*²⁸, *CrANT*⁵⁹, *CrLFY*⁶⁰). This conservation allows us to identify molecular changes between reproductive development in seedless and seed-bearing vascular plants that were potentially important to the evolution of the ovule.

GA signalling was identified as promoting the shoot sporangium transition in *Ceratopteris* similar to the shoot floral transition in angiosperms, with partially-separable roles regulating frond complexity and sporogenesis. Canonical GA signalling is inferred to have evolved in the vascular plant lineage prior to the divergence of ferns^{96,97}. In lycophytes GA promotes microspore wall development⁹⁸ but not growth responses⁹⁷. A role for GA in lycophyte shoot

reproductive transitions has not been reported but our data suggest this function was present at least in the last common ancestor of ferns and seed plants.

The extent of conservation between the sporing and floral transition remains to be determined. In this study the sensitivity of our analysis was limited by the unavoidable presence of primordia within our Apx samples. However, evidence suggests not all components of the angiosperm floral transition are conserved, with *CrLFY1* and -2 playing no role in *Ceratopteris*⁹⁹. Interestingly, these instead have gametophytic reproductive functions and the capacity to regulate notch meristem activity. Here we detected further genes in the pre-fertilization notch associated with flowering and the inflorescence meristem (*TFL1-LIKE*, *SIN3-LIKE*)^{50,51}. In the gymnosperm *Picea abies* *TFL1-like* genes are thought to lack a flowering function and instead act to promote growth cessation¹⁰⁰, consistent here with the shutdown of the notch meristem. To what extent gametophytic gene networks contributed to the seed plant reproductive shoot network remains to be elucidated.

The putative *Ceratopteris* sporophyll gene network contains multiple orthologs of the angiosperm microsporogenesis programme (*EXS/EMS*, *TPD1*, *GAMYB*, *AMS1*, *MS1* and *MS2*)⁴⁹ that results in a highly-derived three-celled male gametophyte enclosed in a sporopollenin coat¹⁰¹. Components of this network regulating sporogenesis, gametogenesis and spore/pollen wall formation are also conserved in the moss *Physcomitrium patens* (*PpMS1*, *PpMS2* and *PpGAMYB*)^{98,102,103}. Our results suggest that this sporogenesis network is conserved across land plants and that the gene network in *Ceratopteris* strongly resembles that regulating microsporogenesis in seed plants, consistent with the developmental similarities reported between them¹⁹.

Surprisingly, the same genes relating to sporogenesis were also significantly enriched in the post-meiotic gametophyte notch. A number of these have dual roles in angiosperms:

EXS1/EMS1 and *DRM1* regulate the specification of sporogenous mother cells in both anthers and ovules^{93,94,104}; *MSH2*, *SAE2*, *RAD51C* function in *Arabidopsis* male and female meiosis and also in mitosis^{53,55,56}; the *GAMYB* family is essential for microsporogenesis by promoting tapetum programmed cell death (PCD)¹⁰⁵ but also promotes seed germination through aleurone PCD⁹⁵. Our findings suggest that some of these ‘cross-generational’ genes already had functions beyond sporogenesis in the gametophyte of the last common ancestor of ferns and seed plants, and we speculate that this could have prefigured the evolution of female meiosis.

Beyond sporogenesis, *Ceratopteris* sporophyll reproductive candidate genes showed conservation with *Arabidopsis* stamen gene networks, consistent with RNA-seq analysis of *Adiantum capillus-veneris* homosporangia³⁴. Gene expression changes specific to sporophyll rather than frond development were enriched in hormone-related genes, including auxin which regulates both *Arabidopsis* early pollen development and stamen maturation^{106,107}. Similarly, *Ceratopteris* auxin-related genes were associated with late sporophyll development and coincident with meiosis. A second hormone identified in sporophyll development (JA) also regulates angiosperm stamen development and, more variedly, broader aspects of floral development¹⁰⁸. As such, our results support the evolution of the stamen gene network from those of ancestral sporangium-bearing sporophytic organs.

We found multiple strands of evidence that gene networks regulating post-fertilization development of the fern archegonium contributed to the origin of ovule patterning. Although conservation of FG networks between ferns and seed plants is to be expected as homologous organisms, importantly this enrichment included genes specific to premeiotic ovules and thus cannot be explained simply by the presence of the FG in the ovule, whilst genes conversed with

the angiosperm FG alone do not account for the full enrichment between ovule and post-fertilization archegonium. This putative conserved network did not include *LEC1*, previously predicted as the origin of a seed maturation network from the fern gametophyte³⁰, but it is possible that further ancestral modules were co-opted later in seed evolution.

Although *Ceratopteris* sporophyll reproductive gene networks were not directly conserved with *Arabidopsis* ovule gene networks, archegonium genes conserved with the ovule were also enriched in the sporophyll primordium. This suggests that at least some part of the ovule gene network arose from an ancestral sporangium-bearing sporophytic axis, consistent with hypotheses based on morphological and fossil data²⁰. This limited conservation must be considered in the context of by differences observed in shoot gene networks between ferns and lycophytes¹⁰⁹, raising the possibility that extant fern sporophylls do not closely resemble the ancestral spore-bearing shoot gene network from which ovules arose. Notably however, those conserved genes identified in the sporophyll did not include all notch orthologs conserved with ovule sporophytic tissues, suggesting that part of the sporophytic network in extant ovules was gametophytic in evolutionary origin.

If current evolutionary theories are correct, can the underlying changes in development and patterning necessary to create the ovule from a sporophytic sporangiate lateral organ be explained by changes in ancestral reproductive gene networks? Our results support a model whereby that ancestral organ acquired 'ovule-like' characteristics from the post-fertilization archegonium: gene networks regulating post-fertilization archegonium development were expressed in that developing organ prior to sporangium initiation, resulting in developmental modifications that began to resemble that fertilized archegonium. Consistent with this, our findings suggest this novel expression may have occurred in the primordium stage of that organ's development. This model reconciles the morphology and fossil-based evidence of

ovules evolving from modified sporangia-bearing axes²⁰, the developmental genetic evidence of extant ovules as lateral organs²³ and the observed but so-far unexplained seed-like characteristics of the gametophyte archegonium²⁵.

There is precedent for large-scale ontogenetic rearrangement of gene networks that supports this hypothesised transfer from the gametophyte archegonium to a sporophytic organ. In *Physcomitrium*, gametophytic or sporophytic gene networks can be mis-expressed in the opposing generation through the mutation of regulatory protein complexes like PRC2¹¹⁰ or KNOX/BELL dimers¹¹¹. Importantly, conservation between *Ceratopteris* and *Arabidopsis* reproductive gene networks was detected specifically in 'reproduction-enriched' gene populations, suggesting that the ovule gene network arose through the repurposing of downstream targets of seedless reproductive gene networks. This model predicts that only the modifications to archegonium development associated with fertilization were expressed in the developing sporophytic axis, not the whole archegonium programme, thus allowing gametophytic reproductive development to continue without disturbance whilst modifying sporophyte organ development. Therefore, this model is also consistent with the persistence of archegonia in ancestral seed plant and extant gymnosperm ovules⁵.

METHODS

Plant material and growth conditions

Ceratopteris richardii lab strain Hn-n¹¹² was used for all experiments. Spores were sterilised through 15 minutes incubation in dilute sodium hypochlorite solution (2-3% active chlorine) containing 0.1% Tween20 (v/v), rinsed 5 times in sterile water and incubated in sterile water at room temperature under darkness for 3-4 days to imbibe¹¹³. Spores were sown onto 90 mm-diameter petri dishes (hereafter 'plates') containing 25 ml C-fern 1% agar medium (pH6.0) with a sterile cellophane disc between spores and media¹¹³. Plates were sown with a density of 200-

300 spores per plate, wetted with 1 ml sterile water and sealed with micropore tape. Plates were kept humid during subsequent incubation by adding 1.5ml sterile water at 3 and 7 days after sowing (DAS), with no water added beyond this time to prevent premature sperm release and fertilization. Fertilization was triggered at 14 DAS by flooding plates with 7.5 ml sterile water, with embryo development assumed to have begun on the day of flooding. Individual sporophytes transferred individually to fresh C-fern agar media once visible (9-11 days after flooding [DAF]). Individual sporophytes were transferred to sterile GA-7 Magenta boxes containing 100 ml C-fern 1% agar medium (pH6.0) when 2-3 fronds and elongating roots were visible (14-16 days) and wetted with 5 ml sterile water. Sporophytes were subsequently maintained in Magenta boxes until 130 days old, with sterile water re-applied whenever no surface water was visible (typically every 2-3 weeks). At 100 and 120 days 5 ml of sterile C-fern liquid media (pH6.0) was applied instead of water to maintain healthy growth. Sporophytes grown for harvesting 25d samples were grown on 90 mm plates, transferred at 14 DAF at a density of 30-40 per plate. All plant material was grown in an MLR-352H-PE controlled environment cabinet (PHC Europe, Breda, Netherlands) at 28°C, constant 50% relative humidity and a 16hr light/ 8hr dark photoperiod at a fluence of 150 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

Embryo development synchronisation

To confirm the efficacy of restricted watering at synchronising embryo development, populations of gametophytes were grown with parallel growth regimes applied: either restricted watering (as described above) or 'well-watered' conditions with sterile water applied at 9 and 11 DAS to maintain a constant layer of surface water, at ratio of 2 restricted water:1 well-watered plates. At 14 DAS, all well-watered plates and half of the restricted water plates were flooded with sterile water as described above ('+Flooding'), with the remaining restricted water plates left dry ('-Flooding'). One plate per treatment was sacrificed for analysis at 14 DAS (restricted-water, well-watered), 4 DAF and 7 DAF (restricted-water $\pm$ Flooding, well-watered +Flooding). From

each plate 50 hermaphrodite gametophytes were selected at random and scored for the presence of an embryo using an ultraZOOM-2 stereomicroscope (GT Vision Ltd, Newmarket, UK) at 20x magnification. The experiment was replicated on four different occasions.

To characterise the timing of post-fertilization embryo development and archegonium production in response to flooding, populations of gametophytes were grown at a density of 200-300 spores per plate under restricted water conditions (as described above) and flooded at 14 DAS. At 0 (unflooded), 1, 2, 3, 4, and 7 DAF single plates were sacrificed for analysis. From each plate 50 hermaphrodite gametophytes were selected at random and scored for the presence of an embryo (as described above). At 1-7 DAF gametophyte selection was restricted to those with evidence of fertilization in response to flooding (brown archegonia). The experiment was replicated on five different occasions. During three of these replicates the number of archegonia per gametophyte were counted alongside scoring for the presence of an embryo. Numbers of embryo-bearing and embryo-less flooded gametophytes recorded per replicate varied due to availability within the population.

Sporing transition characterisation

Sporophytes were grown in Magenta boxes (as described above) in populations of 25. Eleven replicate populations were grown, with populations initiated at intervals over 13 months. From 4 weeks old, the number of sporophytes with visible sporophylls within each population was recorded once per week ± 1 days until 19 weeks old, by which time 100% of sporophytes had begun to spore. For nine of these replicates the number of fronds per sporophyte was also recorded at the same time.

Sporangium development characterisation

Single sporophylls were collected from 130 day-old (130d) sporophytes, secured dorsal-side down to microscope slides using double-sided tape and pinnae dissected open under a stereomicroscope using hypodermic needles as micro-scalpels to expose developing sporangia on the ventral surface. Sections of pinnae were recorded photographically to capture sporangium planar area. Sporangia were recorded at two locations per pinna (base and tip) and from two pinnae per frond, positioned basally and distally. Four sporophytes were sampled per frond developmental stage. Sporangia areas were measured from scaled micrographs using the FIJI image analysis software package¹¹⁴.

Sporophyte hormone treatments

Sporophytes were treated with sterile 100 μM GA₃ (0.1% EtOH, 0.01% Tween20 [v/v]) or 100 μM PBZ (0.1% EtOH, 0.01% Tween20 [v/v]) vs mock solution. For GA experiments 14d sporophytes were transferred to Magenta boxes (as above) containing C-fern 1% agar media containing 100 μM GA₃ or 0.1% EtOH. Sporophytes were then treated with corresponding GA/PBZ or mock solution every 2-3 days thereafter, with 0.5ml solution applied to the shoot tip and emerging fronds. For PBZ experiments 14d sporophytes were transferred to Magenta boxes (as above) containing C-fern 1% agar media and treatment with PBZ or mock solution applied (as above) at 4 weeks old and once per week thereafter. Sporophytes were grown in populations of 25 with 12/13 individuals per treatment.

Gametophyte experimental treatments

For analysis of post-fertilization development, plates of gametophytes were grown to 14 DAS as described above. For sucrose experiments plates were flooded with 7.5 ml sterile 2% sucrose (w/v) or water and incubated under normal growing conditions for 7 days. For JA experiments plates were flooded with 7.5 ml sterile 100 μM MeJA or mock solution (0.00002% EtOH [V/V]) and then incubated under normal growing conditions for 7 days. At 7 DAF 30 (sucrose) or 50

(JA) hermaphrodite gametophytes from each plate were randomly selected and scored for the presence of a visible embryo at 2x magnification using a dissecting microscope and the number of embryos counted. If <30/<50 embryo-bearing gametophytes were collected, further embryo-bearing gametophytes were subsequently screened and the number of embryos counted. For auxin experiments plates were flooded with 7.5 ml sterile 10 μ M NAA or mock solution (0.1 μ M NaOH) then incubated under normal growing conditions for 4 days. Gametophytes were imaged under brightfield microscopy.

For analysis of notch meristem activity, plates of gametophytes were grown to 7 DAS and then transferred to fresh C-fern 1% agar media containing 100 μ M MeJA or mock treatment (0.00002% EtOH [v/v]), and then wetted with 1.5 ml of the corresponding sterile JA or mock solution (as above). Plates were then incubated under normal growing conditions. At 15 DAS 20 hermaphrodite gametophytes were randomly selected from each plate and the number of visible archegonia scored at 2x magnification using a dissecting microscope.

mRNA-seq sample preparation

Tissue samples for mRNA-seq were harvested by dissection and immediately frozen in liquid nitrogen. Sequential stages of sporophyte frond development (shoot apex [Apx], frond primordium [Prm], frond fiddlehead [Fid], expanding frond [Exp] and mature frond [Mat], see **Fig.1**) were harvested from 60d sporophytes (vegetative phase) and 130d sporophytes (reproductive phase) within 2 hours per population grown, with replicate populations harvested between 5 and 10 hours after dawn (ZT5-ZT10). Frond stages were determined by eye and populations were harvested on separate occasions to prevent confusion of similar-looking stages between 60d and 130d. Dissection of Apx and Prm samples was performed under a stereomicroscope using watchmaker's forceps. Tissues were frozen immediately after dissection by collection into tubes floating on liquid nitrogen. All biological replicates comprised

tissue from at least 4 independent sporophytes (**Supplementary Data 12**), harvested from one (all Mat, Exp, Fid, 60d Apx and Prm), 2-3 (130d Apx and Prm) or 3-4 independently-harvested populations (25d Apx and Prm), respectively. Exp samples were designed to encompass 'early' (unfurling), 'mid' (unfurled to 50% final size) and 'late' frond expansion (50-75% final size) within each biological replicate (**Supplementary Fig. 4**). Samples across frond development at 60d and 130d comprised tissue from the same harvested population(s) per biological replicate.

Gametophyte stages were necessarily harvested on separate days, individual populations harvested within 2-4 hours with replicate populations harvested between 5 and 10 hours after dawn (ZT5-ZT10). Dissection of the notch was performed under a stereomicroscope using hypodermic needles as micro-scalpels, the developmental status of each individual confirmed by eye. Each biological replicate comprised notch or thallus tissue from 150-225 individual gametophytes (**Supplementary Data 12**) pooled from 1 (0 DAF, 1 DAF, 2DAF), 2-3 (4 DAF) or 3-4 independently-harvested populations (Dem, 761-788 embryos per sample) respectively. Notch and thallus from the same individuals remained paired when preparing biological replicates for sequencing.

RNA was extracted using the Qiagen RNeasy plant mini kit with on-column DNase treatment (QIAGEN, Hilden, Germany). RNA quality was determined using a 2100 Bioanalyser RNA 6000 nano assay (Agilent Technologies, Santa Clara, USA). Individual sample RIN values are supplied in **Supplementary Data 12**. Library preparation, 150bp paired-end Illumina sequencing was performed separately for sporophyte and gametophyte samples by Novogene (Novogene UK, Cambridge UK) to a depth of ≥ 20 million reads per sample. QC data is provided in **Supplementary Data 12**.

Imaging

Whole sporophytes and gametophyte plates were photographed using a Cybershot DSC-HXTV digital camera (Sony Corporation, Tokyo, Japan). Individual gametophytes and dissected tissue samples were photographed using an ultraZOOM-2 stereomicroscope with an attached GXCAM-U3PRO-X digital camera and GXCam software (GT Vision Ltd, Newmarket, UK). Photographs were minimally processed by adjusting brightness and contrast and sharpness for whole images.

Statistics and reproducibility

No statistical method was used to predetermine sample size. Randomization was used during the growing of plant material for each experiment. Sporophyte populations under tissue culture were grown in a 5x5 grid pattern, with the positions of Magenta boxes randomised once per week. Similarly, the positions of gametophyte plates growing within an experiment was randomised once per week. For gametophyte characterisation experiments, individuals were selected for scoring at random within each plate, with the exception of when additional embryo-bearing gametophytes were required. For gametophyte sucrose, auxin and JA response experiments, the investigators were blinded to treatment allocation until after data collection was complete. For sporophyte GA and PBZ response experiments, the requirement to apply treatments repeatedly and the clear growth phenotypes exhibited before completion of the experiment made blinding an impracticable approach.

Statistical analyses of characterization experiments were conducted in R version 4.2.2 using the car package¹¹⁵. Pairwise comparisons were performed using two-tailed students T-tests or Mann-Whitney nonparametric tests (as shown within each figure), following analysis of the data structure using Shapiro-Wilk for Normal distribution¹¹⁶ and Levene's Test for equal variance¹¹⁷ to determine what test was appropriate. Where necessary data transformation was applied (indicated within each figure legend). Datasets were assessed for significant outlying datapoints

but it was not necessary to exclude any data from the analyses performed. A post-hoc correction for multiple comparisons was applied to p-values using the Benjamini-Hochberg method¹¹⁸.

mRNA-seq data analysis

Library preparation and sequencing were conducted by Novogene using NovaSeq 6000 PE150. We obtained an average of over 56 million pair reads per samples, with each sample covered by at least by 30 million paired reads. We sequenced three biological samples (replicates) for each developmental stage analysed. FAST QC (v.0.11.5) was used to assess the quality of the sequenced libraires (Andrew et al. 2010). Reads were aligned to the *Ceratopteris richardii* genome v2.1³⁶ using HISAT2 (v.2.2.1)¹¹⁹. SAMtools (v.1.10)¹²⁰ was used to sort and index Sam files into Bam format. Read counts were generated using HTSeq (v.0.13.5)¹²¹. Read counts per gene were analysed in R using DEseq2 (v.1.34.0)¹²². Differentially expressed genes (DEGs) were computed in pairwise comparisons across all developmental stages (each composed by three biological samples). We considered DEG any gene with an adjusted p-value (P_{adj}) below 0.05 and a log2 foldchange >1. For each gene the transcripts per million reads (TPM) was computed in R starting from the raw read count.

Orthology of *Ceratopteris* genes to *Arabidopsis* was identified using Orthofinder¹²³ with default settings against the following proteomes: *Adiantum capillus-veneris*³⁴, *Alsophila spinulosa* (v3.1)³⁵, *Arabidopsis thaliana* (TAIR10)¹²⁴, *Amborella trichopoda* (v1.0)¹²⁵, *Cycas panzhihuaensis*¹²⁶, *Gnetum montanum*¹²⁷, *Glycine max* (Wm82.a4.v1)¹²⁸, *Lycopodium clavatum*¹²⁹, *Manihot esculenta* (v8.1) (https://phytozome-next.jgi.doe.gov/info/Mesculenta_v8_1), *Marsilea vestita* (v3.0)¹³⁰, *Micromonas pusilla* CCMP1545 (v3.0)¹³¹, *Ostreococcus lucimarinus* (v2.0)¹³², *Oryza sativa* (v7.0)¹³³, *Physcomitrium patens* (v3.3)¹³⁴, *Quercus robur* (PM1N)¹³⁵, *Solanum lycopersicum* (ITAG4.0)¹³⁶, *Selaginella*

moellendorffii (v1.0)¹³⁷, *Spirodela polyrhiza* (v2)¹³⁸, *Thuja plicata* (v3.1)¹³⁹, and *Volvox carteri* (v2.1)¹⁴⁰. Orthology data is provided in **Supplementary Data 8**. Additional GO term annotations were assigned from gene orthologs in the model species *Arabidopsis*, *Oryza*, *Selaginella*, *Physcomitrium* and *Marchantia*. Higher-order Venn comparisons were performed in Interactivenn¹⁴¹. Further analysis of transcriptome data was performed in R, with cluster analysis using the pHeatmap package and GO term enrichment testing using the TopGO package employing Fisher's Exact Test (classic algorithm). *Arabidopsis* reproductive gene datasets were obtained from past publications and filtered against published vegetative-specific genes (**Supplementary Data 9**). Enrichment testing of *Arabidopsis* orthologs of *Ceratopteris* genes against these datasets was performed in R using Fisher's Exact Test (one-sided, 'greater than') and corrected for multiple comparisons using the Benjamini-Hochberg method¹¹⁸. All other plots were prepared in R using the ggplot2¹⁴² and the ggvenn and ggcorrplot packages.

DATA AVAILABILITY

The raw sequencing read data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO) database under accession code GSE291236 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE291236>]. The processed gene expression data are provided in the Source Data file. The statistical analyses generated in this study are provided in the Supplementary Information. The published *Arabidopsis* microarray expression data used in this study are available at the NCBI GEO under accession codes GSE1265 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1265>]⁷⁰, GSE1266 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1266>]⁷⁰, GSE1267 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1267>]⁷⁰, GSE1268 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1268>]⁷⁰, GSE1269 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1269>]⁷⁰, GSE1270 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1270>]⁷⁰, GSE1271 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1271>]⁷⁰, GSE1272 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1272>]⁷⁰, GSE1273 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1273>]⁷⁰, GSE1274

[<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1274>]⁷⁰, GSE1275
 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1275>]⁷⁰, GSE8864
 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8864>]⁷², GSE680
 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE680>]⁷⁶, GSE12404
 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12404>]⁸⁰; the EMBL-EBI Array
 Express database under accession codes E-MEXP-1246
 [<https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MEXP-1246>]⁷³, E-MEXP-1920
 [<https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MEXP-1920>]⁷⁵, E-MEXP-3137 MMC
 [<https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MEXP-3137>]⁷⁸, E-MEXP-3138
 [<https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MEXP-3138>]⁷⁸; or online
 supplementary material^{68,69,71,74,77,79}. The processed *Arabidopsis* gene datasets used in this
 study are provided in the Supplementary Information. Source data are provided with this paper.

REFERENCES

1. Hilton, J. & Bateman R.M. Pteridosperms are the backbone of the seed-plant phylogeny. *Journal of the Torrey Botanical Society* **133**, 119-168 (2006).
2. Wickett, N.J. *et al.* Phylotranscriptomic analysis of the origin and early diversification of land plants. *PNAS* **111**, E4859-E4868 (2014).
3. Pryer, K.M. *et al.* Horestails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**, 618-622 (2001).
4. Harrison, C.J. Development and genetics in the evolution of land plant body plans. *Philos Trans R Soc Lond B Biol Sci* **372**, 20150490 (2017).
5. Rudall, P.J. Evolution and patterning of the ovule in seed plants. *Biological Reviews* **96**, 943-960 (2021).
6. Singh, H. & Johri, B.M. *Seed Biology: Importance, development and germination Volume 1* Ch. 2. (Elsevier Academic Press, Oxford, 1972).
7. Linkies, A., Graeber, K., Knight, C. & Leubner-Metzger, G. 2010. The evolution of seeds. *New Phytol.* **186**, 817-831 (2010).

8. Matilla, A.J. Seed coat formation: its evolution and regulation. *Seed Science Research* **29**, 215-226 (2019).
9. Willis, C.G. *et al.* The evolution of seed dormancy: environmental cues, evolutionary hubs, and diversification of the seed plants. *New Phytol.* **203**, 300-309 (2014).
10. Qiu, Y.L., Taylor, A.B. & McManus, H.A. Evolution of the life cycle in land plants. *Journal of Systematics and Evolution* **50**, 179-194 (2012).
11. Conway, S.J. & Di Stilio, V.S. An ontogenetic framework for functional studies in the model fern *Ceratopteris richardii*. *Developmental Biology* **457**, 20-29 (2020).
12. Harrison, C.J., Morris, J.L. The origin and early evolution of vascular plant shoots and leaves. *Phil Trans Roy Soc B* **373**, 20160496 (2017).
13. Tsuda, K. Evolution of the sporophyte shoot axis and functions of TALE HD transcription factors in stem development. *Curr. Op. Plant Biol.* **81**, 102594 (2024).
14. Plackett, A.R.G., Di Stilio, V.S. & Langdale, J.A. Ferns: the missing link in shoot evolution and development. *Frontiers in Plant Science* **6**, 972 (2015).
15. Petersen, K.B. & Burd, M. Why did heterospory evolve? *Biological Reviews* **92**, 1739-1754 (2017).
16. Hickok, L.G., Warne, T.R. & Fribourg, R.S. The biology of the fern *Ceratopteris* and its use as a model system. *International Journal of Plant Sciences* **156**, 332–345 (1995).
17. Haufler, C.H. *et al.* Sex and the Single Gametophyte: Revising the Homosporous Vascular Plant Life Cycle in Light of Contemporary Research. *Bioscience* **66**, 928-937 (2016).
18. Banks, J. Gametophyte development in ferns. *Annual Review of Plant Biology* **50**, 163-186 (1999).
19. Ambrose, B.A. & Stevenson, D.W. The evolution and development of sporangia- the fundamental reproductive organ of land plant sporophytes. *Curr. Op. Plant Biol.* **81**, 102563 (2024).

20. Taylor, E.L., Taylor, T.N. & Krings, M. *Palaeobotany: The biology and evolution of fossil plants*. (Academic Press, Boston, 2009).
21. Sokoloff, D.D., Remizowa, M.W. Diversity, development and evolution of archegonia in land plants. *Bot. J. Linn. Soc.* **195**, 380-419 (2021).
22. Ranganath, R.M. Female gametophyte development in higher plants- meiosis and mitosis break the cellular barrier. *Plant Biology* **5**, 42-49 (2003).
23. Mathews, S., Kramer, E.M. The evolution of reproductive structures in seed plants: a re-examination based on insights from developmental genetics. *New Phytol.* **194**, 910-923 (2012).
24. Ran, J.H., Shen, T.T., Wang, M.M. & Wang, X.Q. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proc. R. Soc. B.* **285**, 20181012 (2018).
25. Johnson, G.P. & Renzaglia, K.S. Embryology of *Ceratopteris richardii* (Pteridaceae, tribe Ceratopterideae), with emphasis on placental development. *Journal of Plant Research* **121**, 581-592 (2008).
26. Favaro, R. *et al.* MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* **15**, 2603-2611 (2003).
27. Pinyopich A, *et al.* Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**, 85-88 (2003).
28. Hasebe, M., Wen, C.K., Kato, M. & Banks, J.A. Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. *PNAS* **95**, 6222-6227 (1998)
29. Theissen, G., Melzer, R., Rümpler, F. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* **143**, 3259-3271 (2016).
30. Bai, S.N., Rao G.Y. & Yang J. Origins of the seed: the 'golden-trio' hypothesis. *Frontiers in Plant Science* **13**, 965000 (2022).

31. Rodríguez-Pelayo, C., Ambrose, B., Vasco, A., Alzate, J.F. & Pabón-Mora, N. Evolution and expression of LEAFY genes in ferns and lycophytes. *EvoDevo* **13**, 2 (2022).
32. Withers, K.A. *et al.* A *Ceratopteris* EXCESS MICROSPOROCTES1 suppresses reproductive transition in the fern vegetative leaves. *Plant Science* **353**, 111812 (2023).
33. Li, F.W. *et al.* Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* **4**, 460-472 (2018).
34. Fang, Y. *et al.* The genome of homosporous maidenhair fern sheds light on the euphylllophyte evolution and defences. *Nature Plants* **8**, 1024-1037 (2022).
35. Huang, X. *et al.* The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nature Plants* **8**, 500-512 (2022).
36. Marchant, D.B. *et al.* Dynamic genome evolution in a model fern. *Nature Plants* **8**, 1038-1051 (2022).
37. Fernández, P. *et al.* A 160 Gbp fork fern genome shatters size record for eukaryotes. *iScience* **27**, 109889 (2024).
38. Aya, K. *et al.* De novo transcriptome assembly of a fern, *Lygodium japonicum*, and a web resource database, LjtransDB. *Plant and Cell Physiology* **56**, e5 (2015).
39. Geng, Y. *et al.* A de novo transcriptome assembly of *Ceratopteris richardii* provides insights into the evolutionary dynamics of complex gene families in land plants. *Genome Biology and Evolution* **13**, evab042 (2021).
40. Ali, Z. *et al.* Comparative transcriptomics in ferns reveals key innovations and divergent evolution of the secondary cell walls. *Nat. Plants* **11**, 1028-1048 (2025).
41. Fu, Q. & Chen, L.Q. Comparative transcriptome analysis of two reproductive modes in *Adiantum reniforme* var. *sinense* targeted to explore possible mechanism of apogamy. *BMC Genetics* **20**, 55 (2019).
42. Huang, S. *et al.* Overexpression of 20-oxidase confers a gibberellin-overproduction phenotype in *Arabidopsis*. *Plant Physiol.* **118**, 773-781 (1998).

43. Qian Q *et al.* A novel Arabidopsis gene RGAT1 is required for GA-mediated tapetum and pollen development. *New Phytol.* **231**, 137-151 (2021).
44. Hedden, P. & Thomas, S.G. Gibberellin biosynthesis and its regulation. *Biochemical Journal* **444**, 11-25 (2012).
45. Lee, E.H., Byun, J.K., Wilding, S.J. A new gibberellin biosynthesis inhibitor, paclobutrazol (PP₃₃₃), confers increased SO₂ tolerance on snap bean plants. *Environmental and Experimental Botany* **25**, 265-275 (1985).
46. Couteau, F. *et al.* Random chromosome segregation without meiotic arrest in both male and female meiocytes of a *dmc1* mutant of Arabidopsis. *The Plant Cell* **11**, 1623-1634 (1999).
47. Higgins, J.D., Armstrong, S.J., Franklin, F.C.H. & Jones, G.H. The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes Dev.* **18**, 2557-2570 (2004).
48. Chelysheva, L. *et al.* Zip4/Spo22 is required for Class I CO formation but not for synapsis completion in Arabidopsis thaliana. *PLoS Genetics* **3**, e83 (2007).
49. Wiese, A.J., Torotaeva, E. & Honys, D. The transcription factors and pathways underpinning male reproductive development in Arabidopsis. *Front. Plant Sci.* **15**, 1354418 (2024).
50. Yang, K.Z., Wei, X.X. & Wang, X.Q. Revisiting the phosphatidylethanolamine-binding protein (PEBP) gene family reveals cryptic *FLOWERING LOCUS T* gene homologs in gymnosperms and sheds new light on functional evolution. *New Phytol.* **212**, 730-744 (2016).
51. Huang, F., Yuan, W., Tian, S., Zheng, Q. & He, Y. *SIN3 LIKE* genes mediate long-day induction of flowering but inhibit the floral transition in short days through histone deacetylation in Arabidopsis. *Plant. J.* **100**, 101-113 (2019).
52. Sánchez-Gerschon, V. *et al.* Transcription factors HB21/40/53 trigger inflorescence arrest through abscisic acid accumulation at the end of flowering. *Plant Physiol.* **195**, 2743-2756 (2024).

53. Emmanuel, E., Yehuda, E., Melamed-Bessudo, C., Avivi-Ragolsky, N. & Levy, A.A. The role of *AtMSH2* in homologous recombination in *Arabidopsis thaliana*. *EMBO reports* **7**, 101-105 (2006).
54. Lario, L.D., Botta, P., Casati, P. & Spampinato, C.P. Role of *AtMSH7* in UV-B-induced DNA damage recognition and recombination. *JXB* **66**, 3019-3026 (2014).
55. Uanschou C *et al.* A novel plant gene essential for meiosis is related to the human *CtIP* and the yeast *COM1/SAE2* gene. *EMBO J.* **26**, 5061-5070 (2007).
56. Abe, K. *et al.* *Arabidopsis RAD51C* gene is important for homologous recombination in meiosis and mitosis. *Plant Physiol.* **139**, 896-908 (2005).
57. Hartung, F., Suer, S., Knoll, A., Wurz-Wildersinn, R. & Puchta, H. Topoisomerase 3 α and RMI1 suppress somatic crossovers and are essential for resolution of meiotic recombination intermediates in *Arabidopsis thaliana*. *PLoS Genet.* **4**, e1000285 (2008).
58. Liu, Y., Deng, Y., Li, G. & Zhao, J. Replication factor C1 (*RFC1*) is required for double-strand break repair during meiotic homologous recombination in *Arabidopsis*. *Plant J.* **73**, 154-165 (2013).
59. Bui, L.T. *et al.* A fern *AINTEGUMENTA* gene mirrors BABY BOOM in promoting apogamy in *Ceratopteris richardii*. *Plant J.* **90**, 122-132 (2017).
60. Plackett, A.R.G. *et al.* *LEAFY* maintains apical stem cell activity during shoot development in the fern *Ceratopteris richardii*. *eLife* **7**, e39625 (2018).
61. Yadav, U.P. *et al.* The sucrose–trehalose 6-phosphate (Tre6P) nexus: specificity and mechanisms of sucrose signalling by Tre6P. *Journal of Experimental Botany* **65**, 1051–1068 (2014).
62. Griffiths, C.A. *et al.* Chemical intervention in plant sugar signalling increases yield and resilience. *Nature* **540**, 574-578 (2016).
63. Meitzel, T. *et al.* Trehalose 6-phosphate promotes seed filling by activating auxin biosynthesis. *New Phytol.* **229**, 1553–1565 (2021).

64. Cordle, A.R., Irish, E.E. & Cheng, C.L. Apogamy induction in *Ceratopteris richardii*. *International Journal of Plant Sciences* **168**, 361-369 (2007).
65. Figueiredo, D.D., Batista, R.A., Roszak, P.J. & Köhler, C. Auxin production couples endosperm development to fertilization. *Nature Plants* **1**: 15184 (2015).
66. Figueiredo, D.D., Batista, R.A., Roszak, P.J., Hennig, L. & Köhler, C. Auxin production in the endosperm drives seed coat development in *Arabidopsis*. *eLife* **5**: e20542 (2016).
67. Goetz, M. *et al.* Expression of aberrant forms of *AUXIN RESPONSE FACTOR8* stimulates parthenocarpy in *Arabidopsis* and tomato. *Plant Physiol.* **145**, 351-366 (2007).
68. Honys, D. & Twell, D. Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* **132**, 640-652 (2003).
69. Honys, D. & Twell, D. Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biology* **5**, R85 (2004).
70. Wellmer, F., Reichmann, J.L., Alves-Ferreira, M. & Meyerowitz, E.M. Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers. *The Plant Cell* **16**, 1314-1326 (2004).
71. Yu, H.J., Hogan, P. & Sundaresan, V. Analysis of the female gametophyte transcriptome of *Arabidopsis* by comparative expression profiling. *Plant Physiology* **139**, 1853-1869 (2005).
72. Alves-Ferreira, M. *et al.* Global expression profiling applied to the analysis of *Arabidopsis* stamen development. *Plant Physiol.* **145**, 747-762 (2007).
73. Johnston, A.J. *et al.* Genetic subtraction profiling identifies genes essential for *Arabidopsis* reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. *Genome Biology* **8**,R204 (2007).
74. Peiffer, J.A. *et al.* A spatial dissection of the *Arabidopsis* floral transcriptome by MPSS. *BMC Plant Biology* **8**, 43 (2008).
75. Skinner, D.J. & Gasser, C.S. Expression-based discovery of candidate ovule development regulators through transcriptional profiling of ovule mutants. *BMC Plant Biology* **9**, 29 (2009).

76. Le, B.H. *et al.* Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *PNAS* **107**, 8063-8070 (2010).
77. Wuest, S.E. *et al.* Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Current Biology* **20**, 506-512 (2010).
78. Schmidt, A. *et al.* Transcriptome analysis of the Arabidopsis Megaspore Mother Cell uncovers the importance of RNA helicases for plant germline development. *PLoS Biology* **9**, e1001155 (2011).
79. Xiang, D. *et al.* Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in Arabidopsis. *Plant Physiol.* **156**, 346-356 (2011).
80. Belmonte, M.F. *et al.* Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *PNAS* **110**: E435-E444 (2013).
81. Schiefthaler, U. *et al.* Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. *PNAS* **96**, 11664-11669 (1999).
82. Yang, Y.C., Ye, D., Xu, J. & Sundaresan, V. The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.* **13**, 2108-2117 (1999).
83. Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R. & Tighe, P.J. 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 (2001).
84. Soppe, W.J.J. *et al.* The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791-802 (2000).
85. Fujita, H. *et al.* An Arabidopsis MADS-box protein, AGL24, is specifically bound to and phosphorylated by Meristematic Receptor-Like Kinase (MRLK). *Plant and Cell Physiology* **44**, 735-742 (2003).

86. DeYoung, B.J. *et al.* The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. *Plant J.* **45**, 1-16 (2006).
87. Wynn, A.N., Rueschhoff, E.E. & Franks, R.G. Transcriptomic characterization of a synergistic genetic interaction during carpel margin meristem development in *Arabidopsis thaliana*. *PLoS One* **6**, e26231 (2011).
88. Cucinotta, M., Colombo, L. & Roig-Villanova, I. Ovule development, a new model for lateral organ formation. *Frontiers Plant Science* **5**, 117 (2014).
89. Ray, A. *et al.* Arabidopsis floral homeotic gene BELL (*BEL1*) controls ovule development through negative regulation of AGAMOUS gene (*AG*). *PNAS* **91**, 5761-5765 (1994)
90. Li, J., Jia, D. & Chen, X. *HUA1*, a Regulator of Stamen and Carpel Identities in Arabidopsis, Codes for a Nuclear RNA Binding Protein. *Plant Cell* **13**, 2269-2281 (2001).
91. Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J. & Chaudhury, A. Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. *PNAS* **97**, 10637–10642 (2000).
92. Zhao, Y. *et al.* Clearance of maternal barriers by paternal miR159 to initiate endosperm nuclear division in *Arabidopsis*. *Nature Commun.* **9**, 5011 (2018).
93. Nonomura, K.I. *et al.* 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 (2003).
94. Mendes, M.A. *et al.* The RNA-dependent DNA methylation pathway is required to restrict *SPOROCTELESS/NOZZLE* expression to specify a single female germ cell precursor in *Arabidopsis*. *Development* **147**, dev194274 (2020).
95. Alonso-Peral, M.M. *et al.* The microRNA159-regulated *GAMYB-like* genes inhibit growth and promote programmed cell death in Arabidopsis. *Plant Physiol.* **154**, 757-771 (2010).

96. Hirano, K. *et al.* The GID1-mediated gibberellin perception mechanism is conserved in the lycophyte *Selaginella moellendorffii* but not in the bryophyte *Physcomitrella patens*. *Plant Cell* **19**, 3058-3079 (2007).
97. Yasamura, Y., Crumpton-Taylor, M., Fuentes, S. & Harberd, N.P. Step-by-step acquisition of the gibberellin-DELLA growth-regulatory mechanism during land-plant evolution. *Current Biology* **17**, 1225-1230 (2007).
98. Aya, K., *et al.* The Gibberellin perception system evolved to regulate a pre-existing GAMYB-mediated system during land plant evolution. *Nature Commun.* **2**, 544 (2011).
99. McConnell, H. *et al.* *LEAFY* demonstrates functions in reproductive development of the gametophyte but not the sporophyte of the fern *Ceratopteris richardii*. *Development* **153**, dev204808 (2026).
100. Karlgren, A., Gyllenstrand, N., Clapham, D. & Lagercrantz, U. *FLOWERING LOCUS T / TERMINAL FLOWER1*-Like Genes Affect Growth Rhythm and Bud Set in Norway Spruce. *Plant Physiol.* **163**, 792-803 (2013).
101. Hackenberg, D. & Twell, D. The evolution and patterning of male gametophyte development. *Current Topics in Developmental Biology* **131**, 257-298 (2019).
102. Wallace, S. *et al.* Conservation of *Male Sterility 2* function during spore and pollen wall development supports an evolutionarily early recruitment of a core component in the sporopollenin biosynthetic pathway. *New Phytol.* **205**, 390-401 (2015).
103. Landberg, K., Lopez-Obando, M., Sanchez Vera, V., Sundberg, E. & Thelander, M. *MS1/MMD1* homologues in the moss *Physcomitrium patens* are required for male and female gametogenesis. *New Phytol.* **236**, 521-524 (2022).
104. Canales, C., Bhatt, A.M., Scott, R. & Dickinson, H. *EXS*, a putative LRR Receptor Kinase, regulates male germline cell number and tapetal identity and promotes seed development in *Arabidopsis*. *Current Biology* **12**, 1718-1721 (2002).

105. Millar, A.A. & Gubler, F. The Arabidopsis *GAMYB-Like* genes, *MYB33* and *MYB65*, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* **17**, 705-721 (2005).
106. Cardarelli, M. & Cecchetti, V. Auxin polar transport in stamen formation and development: how many actors? *Front. Plant Sci.* **5**, 333 (2014).
107. Yao, X. *et al.* Auxin production in diploid microsporocytes is necessary and sufficient for early stages of pollen development. *PLoS Genet.* **14**, e1007397 (2018).
108. Huang, H., Chen, Y., Wang, S., Qi, T., Song, S. Jasmonate action and crosstalk in flower development and fertility. *JXB* **74**, 1186-1197 (2022).
109. Frank, M.H. *et al.* Dissecting the molecular signatures of apical cell-type shoot meristems from two ancient land plant lineages. *New Phytol.* **207**, 893-904 (2015).
110. Okano, Y. *et al.* A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. *PNAS* **106**, 16321-16326 (2019).
111. Sakakibara, K. *et al.* KNOX2 genes regulate the haploid-to-diploid morphological transition in land plants. *Science* **339**, 1067-1070 (2013).
112. Warne, T.R. & Hickok L.G. (2-Chloroethyl)phosphonic acid promotes germination of immature spores of *Ceratopteris richardii* Brongn. *Plant Physiol.* **83**, 723-725 (1987).
113. Plackett, A.R.G., Rabbionwitsch, E.H. & Langdale J.A. Protocol: genetic transformation of the fern *Ceratopteris richardii* through microparticle bombardment. *Plant Meth.* **11**, 37 (2015).
114. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676-682 (2012).
115. Fox, J. & Weisberg, S. *An R Companion to Applied Regression*, 3rd Ed. (Sage, Thousand Oaks, 2019).
116. Shapiro, S.S. & Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (1965).

- 117.** Levene, H. Robust tests for equality of variances. *In Contributions to Probability and Statistics*, I. Olkin, ed. pp. 278–292. (Stanford University Press, Redwood City, 1960)
- 118.** Benjamini, Y. & Hochberg, Y. 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 (1995).
- 119.** Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907-915 (2019).
- 120.** Li, H., *et al.* The sequence alignment/map format and SAMTOOLS. *Bioinformatics* **25**, 2078–2079 (2009).
- 121.** Anders, S., Pyl, P.T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
- 122.** Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESEQ2. *Genome Biology* **15**, 550 (2014).
- 123.** Emms, D.M. & Kelly, S. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).
- 124.** Bernardini, T.Z. *et al.* The Arabidopsis Information Resource: making and mining the “gold standard” annotated reference plant genome. *Genesis* **53**, 474-485 (2015).
- 125.** Chamala, S. *et al.* Assembly and validation of the genome of the nonmodel basal angiosperm *Amborella*. *Science* **342**, 1516-1517 (2013).
- 126.** Liu, Y. *et al.* The *Cycas* genome and the early evolution of seed plants. *Nat. Plants* **8**, 389-401 (2022).
- 127.** Wan, T., *et al.* A genome for gnetophytes and early evolution of seed plants. *Nat. Plants* **4**, 82-89 (2018).
- 128.** Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178-183 (2010).

- 129.** Yu, J.G. *et al.* The first homosporous lycophyte genome revealed the association between the recent dynamic accumulation of LTR-RTs and genome size variation. *Plant Molecular Biology* **112**, 325-340 (2023).
- 130.** Rahmatpour, N. *et al.* Analyses of *Marsilea vestita* genome and transcriptomes do not support widespread intron retention during spermatogenesis. *New Phytol.* **237**, 1490-1494 (2023).
- 131.** Worden, A.Z. *et al.* Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**, 268-272 (2009).
- 132.** Palenik, B. *et al.* The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *PNAS* **104**, 7705-7710 (2007).
- 133.** Ouyang, S. *et al.* The TIGR Rice Genome Annotation Resource: improvements and new features., *Nucleic Acids Res.* **35**, D883-D887 (2007).
- 134.** Lang, D. *et al.* The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515-533 (2018).
- 135.** Plomion, C. *et al.* Oak genome reveals facets of long lifespan. *Nature Plants* **4**, 440-452 (2018).
- 136.** Hosmani, P.S. *et al.* An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. Preprint at <https://www.biorxiv.org/content/10.1101/767764v1> (2019).
- 137.** Banks, J.A. *et al.* The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* **332**, 960-963 (2011).
- 138.** Wang, W. *et al.* The *Spirodela polyrhiza* genome reveals insights into its neotenus reduction fast growth and aquatic lifestyle. *Nat. Commun.* **5**, 3311 (2014).
- 139.** Shaley, T.J. *et al.* The western redcedar genome reveals low genetic diversity in a self-compatible conifer. *Genome Res.* **32**, 1952-1964 (2022).

140. Prochnik, S.E. *et al.* Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* **329**, 223-226 (2010).
141. Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169 (2015).
142. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016)
143. Golden, T.A. *et al.* *Short Integuments1/suspensor1/carpel Factory*, a dicer homolog, is a maternal effect gene required for embryo development in *Arabidopsis*. *Plant Physiol.* **130**, 808-822 (2002).
144. Holding, D.R., Springer, P.S. The *Arabidopsis* gene *PROLIFERA* is required for proper cytokinesis during seed development. *Planta* **214**, 373-382 (2002).
145. Griffith, M.E. *et al.* The *TORMOZ* gene encodes a nuclear protein required for regulated division planes and embryo development in *Arabidopsis*. *The Plant Cell* **19**, 2246-2263 (2007).
146. Van Damme, D. *et al.* *Arabidopsis* α Aurora kinases function in formative cell division plane orientation. *Plant Cell* **23**, 4013-4024 (2011).
147. Jenik, P.D, Jurkata R.E.J. & Barton M.K. Interactions between the cell cycle and embryonic patterning in *Arabidopsis* uncovered by a mutation in DNA Polymerase ϵ . *Plant Cell* **17**, 3362-3377 (2005).
148. Autran, D. *et al.* Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* **145**, 707-719 (2011).
149. Chen, B. *et al.* BABY BOOM regulates early embryo and endosperm development. *PNAS* **119**, e2201761119 (2022).
150. Tzafrir, I. *et al.* Diversity of TITAN functions in *Arabidopsis* seed development. *Plant Physiol.* **128**, 38-51 (2002).
151. Andreuzza, S. *et al.* DNA LIGASE I exerts a maternal effect on seed development in *Arabidopsis thaliana*. *Development* **137**, 73-81 (2010).

152. Molitor, A.M., Bu, Z., Yu, Y., Sen, W.H. *Arabidopsis* AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes. *PLoS Genet.* **10**, e1004091 (2014).

ACKNOWLEDGEMENTS

We thank Dr Juliet Coates and Prof Daniel Gibbs for helpful discussions about mRNA-seq experimental design, Dr Megan McDonald, Dr Peter Keane and Prof Dr Annette Becker for helpful discussions about data analysis and Dr Eugenio Sanchez-Moran, Prof Daniel Gibbs and Prof Christine Foyer for helpful discussions about the manuscript.

Funding Statement

This research was funded by Royal Society University Research Fellowship URF\R1\191326 (AP).

AUTHOR CONTRIBUTIONS

A.R.G.P conceived and planned experiments, performed developmental characterisation, tissue harvesting, RNA extraction, GO term analysis and enrichment testing. M.C. performed gene orthology assignment. A.R.G.P. and M.C. wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Fig. 1. Reproduction-associated genes identified during *Ceratopteris* sporophyll development.

a-c, The timing and consistency of sporing of the *Ceratopteris* sporophyte. Plots show the

mean sporing time by chronological age (**a**) and shoot position (**b**) and the proportion of sporophytes sporing over time per population (**c**). Black arrowheads denote informative shoot ages used in mRNA-seq experiments. Boxplots denote the data median and interquartile range (box) ± 1.5 the interquartile range (whiskers). $n = 271$ (**a,b**) and 11 (**c**).

d-o, The reproductive-phase *Ceratopteris* shoot (**d**) was dissected for sequential samples across frond development: the shoot apical region (Apx) (**e**) comprising the shoot apex (white arrowhead) and early primordia (P_1); the next-most recent primordium (P_{rm}) (**f**) undergoing morphogenesis; the frond fiddlehead (Fid) (**g**) emerged from the shoot apical region; expanding fronds (Exp) (**h**) where the fiddlehead had terminated, from unfurling to approximately 75% of maturity; and mature non-senescent fronds (Mat) (**i**) where growth had ceased compared to prior fronds. The vegetative phase immediately preceding the sporing transition (**j**) was dissected for corresponding frond developmental stages (**k-o**). Candidate reproductive genes were identified through significantly increased expression ($P_{adj} < 0.05$, log2fold expression change > 1) in the sporophyll vs vegetative frond within each frond developmental stage. Scale bars = 10 mm (grey), 200 μ m (white) and 1 mm (black).

p,q, Comparing reproductive candidate reproductive genes between frond developmental stages identified 2098 candidates (**p**). These were categorised based on transcript abundance (**q**): candidates at FPKM > 1 at 130d but not in vegetative phases 'sporophyll-specific', at FPKM > 1 in both 130d and at least one vegetative phase 'sporophyll-enriched', and at FPKM < 1 at 130d as 'sporophyll-FPKM < 1 '. Candidate gene identities and corresponding expression categories are provided in **Supplementary Data 2**.

Source data are provided as a **Source Data** file.

Fig. 2. Gibberellin promotes the *Ceratopteris* sporing transition independently of frond heteroblasty.

a,b, Differential gene expression analysis identified 51 candidates with significantly greater expression in the 130d Apx vs 60d and 25d, with 13 showing a progressive increase immediately prior to the sporing transition (**a**). These were ranked according to the predicted expression characteristics of a shoot apex sporing regulator (**b**). GA-related annotations are highlighted in bold.

c, Comparison of the number of fronds at sporing under mock and exogenous GA treatment (100 μ M GA₃). Pairwise comparisons were performed on square root-transformed data using a two-tailed Mann-Whitney test. Asterisk denotes a significant difference ($P = 0.0000928$). Boxplots denote the data median and interquartile range (box) ± 1.5 x the interquartile range (whiskers). $n = 11$.

d-g, Under exogenous GA treatment sporangia typically first became visible on frond 9 (**d,e**) with margin rolling and sporangium production compared to mock-treated fronds at the same position (**f**) but otherwise are similar without increased dissection seen in mock-treated sporophylls at later frond positions (**g**). White arrowheads denote visible sporangia within the rolled frond margin.

h-j, Under exogenous PBZ treatment (100 μ M) frond morphology became strongly simplified compared to mock-treated sporophytes (**h,i**). A phase-change nevertheless emerged as shoot development progressed (red arrow) (**i**, inset shows prolonged shoot development post phase-change), with fronds demonstrating increased hyponasty and pinna elongation (**j**). White arrowheads denote apparent immature sporangia on the abaxial surface.

k,l, Comparison of phase-change timing under PBZ treatment compared to sporing under mock treatment, measured chronologically (**k**) and in number of fronds produced per shoot (**l**).

Pairwise comparisons were performed using a two-tailed Mann-Whitney test (chronological age, untransformed data; number of fronds, log₁₀-transformed data). Asterisk denotes a significant difference ($P = 0.00000004382$ and 0.000000000000448 , respectively). Boxplots denote the data median and interquartile range (box) ± 1.5 x the interquartile range (whiskers). $n = 36$

(mock), 33 (PBZ). Scale bars = 10 mm (black, grey), 500 μ m (white). Statistical analysis are provided in **Supplementary Data 1**. Source data are provided as a **Source Data** file.

Fig. 3. *Ceratopteris* sporophyll candidate reproductive genes include conserved meiosis and stamen developmental regulators.

a, The proportion of sporophyll reproductive candidate genes with GO terms annotated from orthologous genes. Numbers in brackets correspond to the frequency of genes within each category.

b, The proportion of GO term categories significantly enriched ($p < 0.05$) within annotated sporophyll reproductive candidates. Numbers in brackets correspond to the frequency of GO terms within each category.

c, Clustering of sporophyll reproductive candidate genes based on normalised expression identified seven discrete clusters of genes with maximum expression in specific stages of sporophyll development. The distribution of genes annotated with GO terms and of genes in sporophyll-specific, sporophyll-enriched and sporophyll-FPKM<1 expression categories are shown within each cluster.

d, The frequency of sporophyll reproductive candidate genes with enriched reproductive GO terms within each cluster identified in (**c**) against the frequency expected based on the whole genome. Clusters are shown in order of their peak in abundance during sporophyll development. Obs., observed frequency, Exp., expected frequency. Enrichment testing employed Fishers Exact Test (one-sided, greater than). Asterisks denote a significant increase ($P < 0.05$) in the frequency of genes with that GO term compared to the expected frequency. P-values for GO term enrichment testing are given in **Supplementary Data 5**.

Source data are provided as a **Source Data** file.

Fig. 4. Identification of reproduction-associated genes in the *Ceratopteris* gametophyte.

a-f, *Ceratopteris* gametophyte and archegonium morphology before flooding (**a,b**), 1 day after flooding (DAF) (**c,d**) and 4 DAF after successful fertilization, as shown by a visible embryo (**e,f**).

Ag, archegonium; asterisk, antheridia; Cp, calyptra; Em, embryo; Nt, notch meristem.

g, The proportion of embryo-bearing gametophytes before and after flooding when grown to sexual maturity (14 days) under well-watered or restricted water conditions. Lineplot shows the mean $\pm$ SEM. $n = 4$.

h,i, The proportion of flooded gametophytes with visible embryos per DAF under repeated flooding (0 +3 DAF) (**h**) or when flooded once (0 DAF only) (**i**). Boxplots show the median and interquartile range (box) $\pm 1.5 \times$ interquartile range (whiskers). $n = 5$. Different letters denote a significant difference ($P_{\text{adj}} < 0.05$) between groups. Pairwise comparisons were performed using two-tailed Mann-Whitney tests on logit-transformed data (**h**) and corrected for multiple comparisons. P-values are given in **Supplementary Data 6**.

j-q, Gametophytes were dissected into the notch region (Ntc) containing the notch meristem and developing archegonia (**j-m**) and surrounding thallus tissue (Thl) (**n-q**) before fertilization (0 DAF; **j,n**); recently after flooding (1 DAF and 2 DAF; **k,l,o,p**) and once an embryo was visible (4 DAF; **m,q**). 4 DAF Ntc samples were subdivided into the archegonium bearing an embryo (Aeb) and embryo-less archegonia (Arc). Candidate reproductive genes were identified through significantly increased expression ($P_{\text{adj}} < 0.05$, log2fold expression change > 1) in the notch vs thallus within each stage.

r,s, Comparing reproductive candidate reproductive genes between gametophyte stages identified 7187 candidates (**r**). These were categorised based on transcript abundance (**s**): candidates expressed at FPKM > 1 in at least one Ntc sample but not in any Thl sample were categorised as 'notch-specific', those expressed at FPKM > 1 in both Ntc and Thl as 'notch-enriched', and those with FPKM < 1 in all Ntc samples as 'notch-FPKM < 1 '. Candidate gene identities and corresponding expression categories are provided in **Supplementary Data 7**.

Scale bars = 500 μm (black), 50 μm (white). Source data are provided as a **Source Data** file.

Fig. 5. The *Ceratopteris* embryo-bearing archegonium is enriched in seed-related genes.

a,b, The proportion of *Ceratopteris* notch candidate genes with GO terms inherited from orthologous genes (**a**) and of different categories of significantly-enriched GO terms ($p < 0.05$) within all notch reproductive candidates (**b**). Numbers in brackets within the key correspond to the numbers of genes (**a**) or GO terms (**b**) within each category.

c, Clustering of notch reproductive candidate genes based on normalised expression identified six discrete clusters of genes with maximum expression in specific tissues during gametophyte reproductive development.

d,e The frequency of observed notch reproductive candidate genes with individual reproductive GO terms against the frequency expected based on the whole genome, analysed within each cluster identified in (**c**) ordered by notch developmental stage (**d**), and within genes differentially-expressed ($P_{\text{adj}} < 0.05$, $\log_2\text{foldchange} > 1$) between Aeb and Dem samples at 4DAF (**e**). Obs., observed frequency, Exp., expected frequency. Enrichment testing employed Fishers Exact Test (one-sided, greater than). Asterisks denote a significant increase ($P < 0.05$) in the frequency of genes with that GO term compared to the expected frequency. P-values for GO term enrichment testing are given in **Supplementary Data 5**. Source data are provided as a **Source Data** file.

Fig. 6. Mechanisms regulating seed-set also regulate *Ceratopteris* fertilization-dependent archegonium development.

a, Enrichment of GO terms relating to nutrients (carbohydrate metabolism and amino acid transport) vs cell wall polysaccharides within each notch expression cluster following the stages of notch development. showing the observed ('Obs.') vs expected ('Exp.') gene frequencies per GO term compared to the entire genome. Enrichment was tested using Fishers Exact Test (one-

sided, greater than). Asterisks denote a significant increase ($p < 0.05$) in genes with that GO term compared to that expected.

b,c, Gametophyte phenotypes 6 days after flooding with either mock treatment (**b**) or 2% sucrose (**c**). Mock-treated gametophytes typically carry one visible embryo but those flooded with 2% frequently carried two.

d, The mean number of visible embryos per replicate population (30 embryo-bearing gametophytes) was significantly different ($p = 0.01$) between mock and sucrose treatment. Boxplot denotes the data median and interquartile range (box) ± 1.5 x the interquartile range (whiskers), $n = 5$. Pairwise comparison was performed using a two-tailed Mann-Whitney test on untransformed data (**Supplementary Data 6**).

e,f, Enrichment of GO terms relating to development or hormone response within each notch expression cluster following the stages of notch development (**e**) and within genes differentially-expressed ($P_{\text{adj}} < 0.05$, $\log_2\text{foldchange} > 1$) between Aeb and Dem samples at 4DAF (**f**), showing the observed ('Obs.') vs expected ('Exp.') gene frequencies per GO term compared to the entire genome. Enrichment testing employed Fishers Exact Test (one-sided, greater than). Asterisks denote a significant increase ($P < 0.05$) in genes with that GO term compared to that expected. P-values for GO term enrichment testing are given in **Supplementary Data 5**.

g-j, Brightfield microscopy of archegonia from embryo-bearing gametophytes after flooding with mock solution (**g,h**) or 10 μM NAA (**i,j**), comparing embryo-containing (**g,i**) and empty archegonia phenotypes (**h,j**). At 4 DAF under NAA treatment calyptra and embryo growth was visibly decoupled (**i**) while embryo-less archegonia of the same gametophytes exhibited ectopic cell division (**j**). Experiment repeated 3 times.

Scale bars = 1 mm (black), 100 μm (white). Ag, archegonium, Cp, calyptra, Em, embryo, Nc, neck canal. Source data are provided as a **Source Data** file.

Fig 7. *Ceratopteris* reproductive organ gene networks are enriched within the *Arabidopsis* stamen, pollen and ovule.

a-c, Sporophyll and notch reproductive candidates are partially overlapping (**a**). All *Arabidopsis* orthologs were identified separately for those candidate genes with direct orthology (**b**), of which 87.0% and 92.0% respectively possessed 5 *Arabidopsis* orthologs or less. Populations of nonredundant *Arabidopsis* orthologs for sporophyll, notch and common candidate genes were thus established (**c**) to enable enrichment testing against *Arabidopsis* reproductive gene datasets.

d-i, Enrichment testing of *Arabidopsis* gene orthologs for the *Ceratopteris* sporophyll-enriched, notch-enriched and common-enriched orthologs within published datasets of *Arabidopsis* reproductive genes expressed in floral organs^{70,74} (**d**), stamen development⁷² (**e**), pollen development^{68,69} (**f**), whole ovule/seed development^{71,76,80} (**g**), ovule tissues^{73,75,77,78} (**h**) and seed tissues^{79,80} (**i**) versus published vegetative-specific genes⁷⁶. Barplots show the overlap of candidate orthologs against each target as a proportion of that dataset. Enrichment tested employed Fishers Exact Test (one-sided, greater than), adjusted for multiple comparisons (Benjamini-Hochberg method). Asterisks denote a significant increase ($P_{adj} < 0.05$) in the proportion of overlapping genes between candidate orthologs and the target reproductive organ compared to the overlap with vegetative-specific genes. Dotted horizontal lines track the baseline overlap of each candidate ortholog dataset with the vegetative control. P-values for individual comparisons are given in **Supplementary Data 9**. Source data are provided as a **Source Data** file.

Fig. 8. Reciprocal enrichment analysis identifies the developmental origin of conserved *Arabidopsis* reproductive gene networks in *Ceratopteris* organs.

a,b Enrichment testing of *Arabidopsis* gene orthologs of *Ceratopteris* notch-enriched (**a**) and sporophyll-enriched (**b**) candidate genes by pHeatmap cluster (see **Fig.3, 5**) within published

datasets of *Arabidopsis* reproductive genes expressed in the ovule and stamen. Barplots show the overlap of each *Ceratopteris* pHeatmap cluster against *Arabidopsis* reproductive datasets as a proportion of that dataset. Enrichment testing employed Fishers Exact Test (one-sided, greater than), adjusted for multiple comparisons (Benjamini-Hochberg method). Asterisks denote a significant increase ($P_{\text{adj}} < 0.05$) in the proportion of overlapping genes between candidate orthologs and the target cluster compared to the vegetative control (dotted vertical lines). Individual P-values are given in **Supplementary Data 9**.

c-e, Reciprocal enrichment testing of *Ceratopteris* gene orthologs of *Arabidopsis* stamen (**c**) and ovule genes conserved within the notch (**d,e**) against pHeatmap clusters within notch (**c,d**) and sporophyll reproductive candidates (**e**). Barplots shows the overlap of candidate orthologs against each pHeatmap cluster as a proportion of that dataset. Enrichment testing employed Fishers Exact Test (one-sided, greater than), adjusted for multiple comparisons (Benjamini-Hochberg method). Asterisks denote a significant increase ($P_{\text{adj}} < 0.05$) in the proportion of overlapping genes between candidate orthologs and the target cluster compared to the overlap with all sporophyll/notch reproductive candidates (dotted horizontal lines). Individual P-values are given in **Supplementary Data 10**.

Source data are provided as a **Source Data** file.

Fig. 9. Notch candidates are enriched in ovule sporophyte and gametophyte tissue-specific datasets.

a, Pairwise correlation matrix comparing the overlap in significantly-enriched notch orthologs between the different ovule tissues tested.

b-d, Enrichment testing of all notch-enriched orthologs within genes specific to or common between tissues of the premeiotic ovule (**b**), the female gametophyte ('FG') (**c**) and both gene networks in total (**d**). 'FG' represents a nonredundant summation of all the genes expressed in egg, synergid and central cell tissues⁷⁷ where significant enrichment was identified, and

'Nuc/MMC' sums the nucellus and MMC⁷⁸ respectively. The assignment of genes to specific or shared ovule/seed tissues is shown in **Supplementary Fig. 22**. Barplots show the overlap of notch-enriched orthologs against *Arabidopsis* reproductive genes distributed by tissue/sample specificity as a proportion of that gene population. Dotted horizontal lines track the baseline overlap with the vegetative control, bar colours denote the tissue/sample of origin. Enrichment testing employed Fishers Exact Test (one-sided, greater than), adjusted for multiple comparisons (Benjamini-Hochberg method). Asterisks denote a significant increase ($P_{\text{adj}} < 0.05$) in the proportion of overlapping genes between candidate orthologs and the target cluster compared to the vegetative control.

e, Overlap in *Arabidopsis* genes expressed in whole ovule stages^{71,78} and Nuc/MMC/FG samples combined.

f, Enrichment testing of all notch-enriched orthologs within genes specific to or common between whole ovule samples (**f**) and within genes specific to these or overlapping with ovule tissue datasets (**g**). Barplots show the overlap of notch-enriched orthologs against *Arabidopsis* reproductive genes distributed by tissue/sample specificity as a proportion of that gene population. Dotted horizontal lines track the baseline overlap with the vegetative control, bar colours denote the tissue/sample of origin. Enrichment testing employed Fishers Exact Test (one-sided, greater than), adjusted for multiple comparisons (Benjamini-Hochberg method). Asterisks denote significant enrichment ($P_{\text{adj}} < 0.05$) compared to the vegetative control.

Individual P-values are given in **Supplementary Data 11**. Source data are provided as a **Source Data** file.

















