Sporophyte Directed Gametogenesis via the Ubiquitin Proteasome System

Prakash Sivakumar[1,2,†], Saurabh Pandey[1,†,‡], A Ramesha[1,†,§], Jayeshkumar Narsibhai Davda[1], Aparna Singh[1,¶], Chandan Kumar[1,#], Hardik Gala[1], Veeraputhiran Subbiah[1,††], Harikrishna Adicherla[1], Jyotsna Dhawan[1], L. Aravind[3], Imran Siddiqi[1,2,*]

Affiliations:

[1]CSIR–Centre for Cellular and Molecular Biology; Uppal Road, Hyderabad 500007, India.

[2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad - 201002, India

[3]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health; Bethesda, MD 20894, USA.

*Corresponding author. Email: imran@ccmb.res.in

[†] These authors contributed equally to the work

Present address:

[‡] databaum GmbH; Holsteinischer Kamp 80, 22081 Hamburg, Germany.

[§] Seri-Biotech Research Laboratory, Central Silk Board; Bangalore-560035, India.

[¶] Department of Botany, Banaras Hindu University; Varanasi 221005, India.

[#] University of Texas at Austin; Texas, USA.

[††] Deceased

Abstract

Plants alternate between diploid sporophyte and haploid gametophyte generations. In mosses which retain features of ancestral land plants, the gametophyte is dominant and has an independent existence. However, in flowering plants the gametophyte has undergone evolutionary reduction to just a few cells enclosed within the sporophyte. The gametophyte is thought to retain genetic control of its development even after reduction. Here we demonstrate that male gametophyte development in Arabidopsis, long considered to be autonomous, is also under genetic control of the sporophyte via a repressive mechanism involving large-scale regulation of protein turnover. We identify an Arabidopsis gene *SHUKR* as an inhibitor of male gametogenesis. SHUKR is unrelated to proteins of known function and acts sporophytically in meiosis to control gametophyte development by negatively regulating expression of a large set of ubiquitination genes specific to post-meiotic gametogenesis. This control is late-emerging as *SHUKR* homologs are found only in eudicots. We show that *SHUKR* is rapidly evolving under positive selection suggesting that variation in control of protein turnover during male gametogenesis has played an important role in evolution within eudicots.

Introduction

Alternation between two distinct generations: a diploid sporophyte and a haploid gametophyte is a central feature of the land plant life cycle[1–3]. Land plants likely evolved from an algal ancestor related to extant charophycean algae which comprise the closest relatives to land plants[4,5]. Charophycean algae have a predominantly haploid life cycle, the diploid phase being transient as the zygote directly undergoes meiosis to give haploid spores. The emergence of a multicellular diploid sporophyte contributed to reproductive success on land by increasing the number of spores produced from a single fertilization event, and also increased potential for genetic variability and evolutionary change[1]. From a simple ancestral body plan, land plants comprising bryophytes (mosses, liverworts, and hornworts) and vascular plants evolved progressively specialized cell types, organs, and architectures adapted for life on land. These acquisitions have been based on development and increase in complexity of a multicellular sporophyte[6,7]. In vascular plants the sporophyte evolved to become the dominant phase and the gametophyte has reduced. In flowering plants the gametophyte is reduced to just a few cells and develops enclosed within the sporophyte[8] (Fig. 1a). The differing trajectories taken by the sporophyte and gametophyte in land plant evolution reflects an uncoupling between their body plan evolution consistent with different genetic programs directing their body organization[7,9,10,11,12].

Formation of a germline is an early step in animal development[13,14], however in plants the formation of germ cell precursors takes place late in development[15]. The fact that bryophytes have a multicellular independent gametophyte which produces gametes following extensive mitotic divisions allows overall assignment of the germline to the

gametophyte generation[16]. In flowering plants the assignment of the germline is more difficult because of extreme reduction of the gametophyte and its development enclosed within the sporophyte[15,17,18]. Evidence from comparative evolution showing functional conservation of transcription factors controlling early steps in germ cell development between the liverwort *Marchantia polymorpha* and Arabidopsis argues for the male germline residing in the gametophyte[16]. Asymmetric cell division of the haploid microspore in Arabidopsis at pollen mitosis I forms a generative cell (GC) and vegetative cell (VC)[16,18]. In Arabidopsis, specification of the GC requires bHLH transcription factors, BNB1/BNB2 and DROP1/DROP2 which act in the gametophyte and are orthologous to MpBNB and MpDROP respectively, that also act in the gametophyte to direct gametangia specification and differentiation in Marchantia[19,20]. Sperm differentiation is controlled by gametophytic action of the bHLH transcription factor DUO1 in both Arabidopsis and Marchantia[21,22]. On the female side gametophytically acting transcription, splicing, and signaling related factors are important for cell specification and cellularization in Arabidopsis female gametophytes[23–29]. Therefore, the gametophyte in flowering plants has been considered to largely retain genetic control of its development even though it lacks an independent existence.

The extreme reduction of the gametophyte and its development enclosed within the sporophyte provided a “functionally unitary individual” for natural selection[30], addressing the sensitive requirements for fertilization and reproductive development in a land environment. Formation and function of meiotic cells depends on a range of cellular interactions in the sporophyte[8,31–33]. These interactions involve transcriptional regulators acting nonautonomously[34–36], cell surface receptor signaling[37–39], and small

RNA pathways[36,40,41] together with phytohormone effectors[42]. Small RNAs[43] and the phytohormone auxin[44,45] have also been shown to mediate developmental interactions between cells within gametophytes. The tapetum is a sporophytic nurse tissue surrounding male meiocytes and provides enzymes for release of microspores and components for synthesis of the protective secondary outer coat of the pollen grain, the exine[46,47]. Mutants deficient in tapetal function manifest defects in exine formation and lead to pollen abortion[48–50]. The tapetum also produces small RNAs that direct DNA methylation of transposons and selected genes in male meiosis contributing to genome integrity and paternal epigenetic inheritance[51].

Studies have shown that isolated uninucleate microspores can form functional pollen in vitro suggesting their development into pollen is autonomous[52,53]. It remains unclear whether genetic controls during meiosis broadly regulate postmeiotic gametophyte development. A challenge in uncovering such regulation has been the existence of very many sporophytic mutants that affect core meiosis functions including meiotic recombination and chromosome organization[54,55] or else development of somatic cells within the anther or ovule[56–59]. These mutants have direct or indirect effects on gametophyte development and show spore or gametophyte abortion, making it difficult to distinguish from mutants affecting meiotic regulation of postmeiotic gametogenesis. Here we report the discovery in Arabidopsis that following evolutionary reduction in cell number of the gametophyte, the sporophyte within the eudicot lineage has gone on to evolve a process established in meiosis under control of the *SHUKR* gene, by which it directs male gametophyte development. The *shukr* mutant shows normal meiosis but postmeiotic development is defective. This process controlled by SHUKR involves large-scale control of regulated protein turnover in the developing

gametophyte and is under rapid evolution with positive selection. Our findings suggest that the sporophyte in addition to providing supportive conditions for gametogenesis and fertilization, also exerts genetic control in meiosis of postmeiotic male gametogenesis. This control is part of a process that is capable of generating variation in the gametophyte during evolution, which has adaptive value.

Results

SHUKR is required for male fertility

We identified the *SHUKR* gene in a screen for novel meiotic functions based on coexpression with known meiotic genes in Arabidopsis (see Methods). Three independent insertion lines having insertions in *At2g38690* showed male sterility (Fig. 1b, Extended Data Fig. 1a). Allelism tests indicated the mutants to be allelic. Anthers from plants homozygous for the insertion contained a highly reduced number of viable pollen, resulting in sterility (Fig. 1b, c). Use of wild type pollen as the male donor in crosses to mutant plants resulted in full seed set indicating that female fertility was unaffected (Fig. 1c). The male sterile phenotype was inherited as a single gene recessive trait (Fig. 1c, Extended Data Fig. 1c). We named the *At2g38690* gene *SHUKR* (*SKR*) after the Sanskrit word Shukranu (/śukrāṇu/) for sperm. We confirmed the phenotype was due to loss of function of *SKR* by complementation with a transgenic copy. *skr/+* plants were transformed with wild type *At2g38690* genomic DNA (*gSKR*) by in planta transformation and T1 *skr* progeny were examined for fertility. The *gSKR* transgene rescued male sterility confirming the mutant phenotype was due to disruption of *At2g38690* (Fig. 1b, Extended Data Fig. 1d). Examination of SKR gene expression based on tissue RT-PCR, a proSKR:GFP-GUS reporter, and analysis of publicly available expression profiling data showed that SKR gene

140 expression was specific to reproductive tissues and more specifically to male
141 meiocytes (Fig. 1d-g; Extended Data Fig. 2). Expression of SKR protein was also
142 meiosis-specific (Fig. 2c).
143
144 **Early male gametogenesis defects in the *shukr* mutant**
145 Analysis of meiosis and pollen development was performed to identify the stage at
146 which defects arose in *skr*. Pollen development is divided into two stages:
147 microsporogenesis comprising meiosis together with microspore development, and
148 microgametogenesis or male gametophyte development comprising development of
149 the haploid microspore into pollen[60]. No defects were observed in *skr* during meiosis
150 based on analysis of chromosome spreads (Fig. 2a). Light microscopic examination
151 of whole mount cleared anthers and plastic sections, showed that tetrad formation and
152 release of microspores took place in *skr* (Figs. 1h,2b; Extended Data Fig. 3), however,
153 microspores degenerated shortly after release from tetrads (Fig. 2b; Extended Data
154 Fig. 4). Newly released wild type microspores were angular in shape (Fig. 2b)[61]
155 whereas *skr* microspores were more rounded and variable in cell wall thickness (Fig.
156 2b), suggesting loss of rigidity and possible defects in cell wall development. Hence
157 sterility in *skr* appeared to be associated with defects in male gametogenesis between
158 tetrad formation and early gametophyte development.
159
160 Ultrastructural examination of tetrad and early microgametogenesis stages (Fig. 3a)
161 by transmission electron microscopy revealed that a significant proportion of *skr*
162 tetrads harbored microspores in which the plasma membrane (PM) was retracted from
163 the surrounding callose wall (Fig. 3b-d). PM retraction was associated with presence
164 of material between the PM and callose layer likely reflecting unincorporated cell wall

165 precursors or disrupted material arising from disturbance in cell wall development.
166 Hence onset of the mutant phenotype was detected at the tetrad stage. To test for
167 additional defects in *skr*, we examined expression of the LAT52:GFP reporter that
168 specifically marks the single vegetative cell after pollen mitosis 1[62]. Among *skr*
169 microspores in which expression of LAT52:GFP took place, we observed two signals
170 in 30% of cases (Fig. 3e). Therefore loss of *SKR* led to at least two distinct defects in
171 early gametogenesis: retraction of the PM with associated disruption of cell wall
172 development, and aberrant expression of an early differentiation marker.
173
174 Recessive gene action in the diploid sporophyte would give Mendelian 3:1 segregation
175 of the mutant allele. Gene requirement in the haploid male gametophyte would result
176 in reduced transmission of the mutant allele through the male and a deviation from this
177 ratio. The genetic analysis suggested that *SKR* acts sporophytically as we did not
178 observe a significant deviation from this ratio (Fig. 4a), whereas the mutant phenotype
179 is with respect to the early gametophytic phase. We tested whether *SKR* also has a
180 gametophytic component to its action by reciprocal crosses of *skr/+* to wild type and
181 examining transmission frequency of the mutant *skr* allele through the gametophyte in
182 comparison to the wild type allele. This test measures fitness of mutant relative to wild
183 type gametes from the same flowers. If there is a functional requirement for *SKR* gene
184 expression at the haploid stage in the gametophyte, then mutant pollen would show
185 reduced fitness and the mutant allele would be transmitted to the next generation at
186 lower frequency than the wild type allele. The results indicated no significant difference
187 in transmission frequency of the mutant *skr* allele from that of the wild type (Fig. 4a),
188 thus failing to provide evidence for *SKR* gene action within the gametophyte. Therefore
189 the genetic results together with ultrastructure and marker expression indicated that

190 *SKR* gene action is in the diploid sporophyte and is required for proper execution of
191 early stages of gametogenesis. Further tests confirmed this finding (see below).
192

193 **Meiotic control of male gametophyte gene expression by *SHUKR***

194 The sporophyte is known to play an important role in formation of the pollen wall and
195 aperture[60,63]. Much of the material and machinery for building the outer wall (exine), is
196 synthesized and released by the tapetum, a layer of nurse cells adjacent to the
197 meiocytes[47,64]. Examination of *SKR* gene expression described above indicated that
198 *SKR* is expressed specifically in male meiocytes. Therefore unlike the sporophytic
199 genes which function in the tapetum, *SKR* appears to act in meiosis. We determined
200 the localization of SKR protein in the course of meiosis using a complemented line
201 carrying a *proSKR:GFP-gSKR* gene fusion. SKR protein was detected specifically in
202 male meiocytes from prophase to tetrad stage and extending to early microspores.
203 Subcellular localization of GFP-SKR varied during meiosis, from nuclear enrichment
204 at prophase, changing toward cytoplasmic in late prophase, and back to nuclear at the
205 tetrad stage. Signal intensity was highest in tetrads (Fig. 2c). Localization in early
206 microspores shifted from nuclear towards more of cytoplasmic, before declining to an
207 undetected level in later stage uninucleate microspores (UNM). Expression was not
208 detected in tapetal cells (Fig. 2c). Therefore SKR protein is mainly present during
209 meiosis and the onset of the *skr* mutant phenotype corresponds to the tetrad stage at
210 which SKR signal is highest. High SKR expression in tetrads followed by steep decline
211 in free microspores suggests a possible role in repressing gametogenesis or else in
212 ending the sporophytic program before initiation of gametogenesis.
213

214 To investigate the above possibilities we examined gene expression during meiosis
215 by RNAseq profiling of meiotic anthers (extending upto and not beyond tetrad stage)
216 for *skr* and wild type. 619 genes were differentially expressed in *skr* of which 379 genes
217 were upregulated (Supplementary Table 1). We analyzed upregulated genes based
218 on cell and tissue expression in publicly available datasets using the CoNekT
219 platform[65] and found a dramatic enrichment for male gametogenesis-enriched
220 genes[66,67] (Fig. 4b, Supplementary Table 2). 187/379 (49%) of upregulated genes
221 were male gametogenesis-enriched. Genes showing maximal enrichment in UNMs
222 represented 80%, followed by bicellular and tricellular pollen stages (16%).
223 Upregulated genes included *DML3* which has been suggested to be an early
224 gametogenesis marker[68]. Increased expression of a large set of gametogenesis-
225 enriched genes in meiotic anthers of *skr* suggests that SKR negatively regulates
226 expression of gametogenesis genes and represses gametogenesis.
227
228 **SHUKR acts sporophytically and inhibits gametogenesis**
229 To test for a role in repression of gametogenesis as well as the possibility that SKR
230 protein acts postmeiotically to promote pollen development, we adopted a
231 misexpression strategy by directing expression of SKR in microspores using the well
232 characterized microspore specific MSP1 promoter[69]. *proMSP1:SKR* in a *skr* mutant
233 failed to rescue the mutant phenotype, indicating that SKR expression in the
234 gametophyte is not sufficient to restore function. Furthermore, *proMSP1:SKR* in wild
235 type in fact caused a reduction in pollen viability (Fig. 4c). These results show that
236 expression of SKR during gametogenesis interferes with gametophyte development
237 and argue against the possibility that SKR protein present during postmeiotic stages

contributes to pollen development. Therefore SKR acts sporophytically and is an inhibitor of male gametogenesis.

***SHUKR* represses the ubiquitin proteasome system**

Gene ontology (GO)[70] together with further analysis of individual gene families revealed a striking signature of regulated protein turnover among the 379 genes upregulated in *skr* RNAseq. 23% of upregulated genes were associated with ubiquitin-mediated degradation of proteins[71] (Fig. 5a, Supplementary Table 3; $p < 2.0e-33$). At the level of molecular function, there was enrichment for ubiquitin protein transferase activity covering the Skp1-Cullin-Fbox (SCF) complex of E3 ubiquitin ligase as well as other components of the ubiquitin protease system (UPS) including *RING/U-box* and *ARIADNE* genes which encode E3 ubiquitin ligases[72,73]. F-box genes encode the substrate specificity component of SCF and comprised the largest family among upregulated genes in *skr*. F-box genes in Arabidopsis make up a family of approximately 740 genes[74,75] (Supplementary Table 3), several of which show expression during male gametogenesis[76]. 69 F-box genes representing 9% of all F-box genes in Arabidopsis were upregulated in *skr* (R.F. = 6.6, $p < 1.74e-36$; Fig. 5a; Supplementary Table 3). Hence a significant fraction of F-box genes are implicated in pollen development. 75% of upregulated F-box genes were maximally enriched at the UNM stage and 9% at later stages comprising bicellular and tricellular pollen. Among downregulated genes, there was enrichment for stress and hypoxia related genes (Supplementary Table 3). Upregulation of a large number of UPS genes suggested involvement of SKR in general control of protein homeostasis and further experiments were performed to test this hypothesis (see below). We also examined stage-wise expression of a subset of upregulated UPS genes at the cellular level, by generating

263 lines carrying a GFP gene fusion for three different F-box genes (*At3g28330*,
264 *At5g02980* and *At5g62510*) and segregating for *skr*. In each case, we observed F-
265 box:GFP expression in wild type only during postmeiotic development of microspores
266 into pollen, whereas *skr* showed precocious F-box:GFP expression in meiosis (Fig.
267 5b, Extended Data Fig. 5). Hence SKR can be seen to act in meiosis to bring about
268 repression of these F-box genes. Taken together, these observations led us to
269 hypothesize that SKR represses onset of male gametogenesis during meiosis through
270 timing of expression of a large set of male gametophyte-preferred F-box and other
271 UPS genes.
272
273 **Control of protein homeostasis by *SHUKR***
274 We directly investigated the functional basis for the *skr* phenotype by performing a
275 suppressor screen on *skr* using EMS mutagenesis of seeds collected from a large
276 number of *skr* plants followed by screening of M2 plants for improved seed set. We
277 isolated a suppressor of *skr* which showed elongated siliques and increased pollen
278 viability, whereas *skr* shows low pollen viability. The suppressor line was backcrossed
279 to the parent *skr* strain and F1 plants were sterile. F2 seed were collected from multiple
280 F1 plants and grown to give F2 plants which showed recessive Mendelian inheritance
281 for fertility (Extended Data Fig. 6a). Whole genome sequencing of bulk segregant
282 fertile and sterile plants from the backcrossed F2 population led to identification of an
283 SNP at the *SKI2* locus which created a stop codon after 517 amino acids and showed
284 100% linkage to the suppressor phenotype (Fig. 5c, Extended Data Fig. 6b). *SKI2*
285 encodes the RNA helicase subunit of the SKI complex and prevents secondary siRNA
286 production in Arabidopsis[77]. *ski2* has also been isolated as a suppressor of
287 overexpression phenotypes in a different context[78]. We tested an independent allele

288 of *ski2* by crossing to *skr* and observed suppression of the *skr* mutant phenotype,
289 confirming *ski2* as a suppressor of *skr* (Extended Data Fig. 6c). We then performed
290 RNAseq on meiotic anthers of the *skr ski2* backcrossed suppressor line. Comparison
291 of differentially expressed genes with the *skr* parent revealed a prominent signature of
292 the proteasome and translation among downregulated genes: 56% of genes encoding
293 subunits of the proteasome and 29% of translation related genes were down-regulated
294 in *skr ski2* (Fig. 5d, Supplementary Table 4). Notably, F-box genes as a class were
295 not downregulated in *skr ski2* (R.F.= 0.3; $p < 7.7e-19$). The striking reduction in
296 transcript levels of the majority of proteasomal genes in *skr ski2* strongly supports the
297 hypothesis that the *skr* phenotype arises from uncoordinated degradation by the
298 proteasome, of regulator(s) of gametogenesis due to inappropriate expression of
299 gametogenesis specific F-box and possibly other ubiquitination genes. Reduction of
300 proteasome activity by downregulation of a major subset of proteasome genes
301 rescues the mutant phenotype.
302
303 We confirmed involvement of the proteasome in the *skr* phenotype using two different
304 approaches. First, we performed knockdown of PBE1, a core component of the
305 proteasome[79] in the *skr* mutant background using *pro*SKR to direct meiosis-specific
306 expression of PBE1 hairpin RNA and scored for pollen viability in RNAi plants.
307 Knockdown of PBE1 by RNAi has been used in a different context[80] and we used the
308 same region for hairpin RNA construction in our experiments. Second, we treated
309 inflorescences of *skr* mutant plants with two physiological inhibitors of the proteasome
310 (MG132 and MG115). In both regimes, we observed an increase in formation of viable
311 pollen that was significant for PBE1:RNAi and MG115 (Fig. 5e,f; $p < 0.05$). Taken
312 together, these results clearly establish deregulation of proteasome activity as a major

313 basis for the *skr* mutant phenotype. The large number of F-box genes showing
314 deregulation during meiosis in *skr*, most of which normally are preferentially expressed
315 only after meiosis during gametogenesis, highlights the scope of this control. Hence,
316 SKR plays a crucial role in meiosis to repress a proteasomal program that promotes
317 gametogenesis and which comes into play in free microspores following meiosis. SKR
318 therefore sporophytically regulates postmeiotic development of the gametophyte
319 through timing of regulated protein turnover by the proteasome (Fig. 5g). Notably, this
320 programme is distinct from the canonical meiotic processes of chromosome reduction
321 and spore formation, which remain unaffected in *skr*.
322

323 **_SHUKR_ is eudicot specific and rapidly evolving**

324 A cDNA of the *SKR* coding region was amplified by RT-PCR, sequenced, and found
325 to encode a predicted protein of 191 amino acids in agreement with the annotated
326 CDS. Homologs were found only in plants and bore no similarity to any known protein
327 family. SHUKR is therefore the founding member of a new protein family (hereinafter
328 SHUKR). The only alignable region conserved across all SHUKR homologs is always
329 at the C-terminus and this is usually the only globular domain in the protein from
330 Shannon entropy-based compositional analysis. The domain has a central
331 compositionally biased, poorly structured variable region enriched in negatively
332 charged residues. Thus, the SHUKR domain varies in length from 67 to 119 residues.
333 It has a hallmark CxC motif and a dyad of aromatic amino acids (usually Y or F) close
334 to the end of the domain (Fig. 6a, Supplementary Figure 1). The SHUKR family is
335 present in both great divisions of eudicots: the Superrosids and Superasterids. It is
336 additionally present in Proteales (Nelumbo) and Ranunculales (Macleaya, Papaver),

337 which are the two basal-most branches of eudicots. It is absent in monocots and basal
338 angiosperms. Hence the origin of SHUKR is at the base of eudicots.
339
340 To understand evolution of the SHUKR family we used the SHUKR domain of 119
341 representatives from 85 species for tree construction. Branch length analysis of the
342 tree revealed two lineages displaying elevated evolutionary rates:
343 Brassicaceae+Cleomaceae and Asteraceae (Fig. 6b; Extended Data Fig. 7). This was
344 further tested using the maximum-likelihood distance matrix computed from the input
345 alignment for the tree. The mean pairwise distance was calculated between each
346 operational taxonomic unit (OTU) and rest of the OTUs, and centered using overall
347 mean distance for the matrix. These mean-centered distances showed a bimodal
348 distribution with the main mode between -1.25 to 0.5, a second mode from 0.5 to 1.5,
349 and a few outliers beyond that (Fig. 6c). The second mode corresponds to sequences
350 with higher values of mean-centered distance, indicating their more extensive
351 divergence. Brassicaceae+Cleomaceae ($p = 1.1472e-04$) and Asteraceae ($p = 1.6e-$
352 16) were significantly enriched among sequences with mean-centered distance
353 greater than 0.5 indicating they represent atypically fast-evolving groups. We note that
354 the high divergence of the SHUKR family outside the SHUKR domain posed
355 challenges in alignment across the entire coding sequence. Therefore the analysis
356 likely represents a conservative assessment of evolutionary rates within the *SHUKR*
357 gene family.
358
359 ***SHUKR* is under positive selection**
360 We assessed rapid evolution in particular plant clades by computing column-wise
361 Shannon entropy from an alignment of the SHUKR domain for four plant families with

362 good representation in the Genbank nr database: Brassicaceae, Asteraceae,
363 Fabaceae and Rosaceae. The column-wise entropy was computed using both the 20
364 amino acid alphabet and a reduced 7 amino acid alphabet that groups together
365 residues as per their shared sidechain chemistry (see Methods). Elevation of entropy
366 in the reduced alphabet indicates changes in sidechain properties along a column and
367 is a proxy for diversifying selection[81]. Entropy plots in both alphabets indicated
368 significantly elevated entropy in Brassicaceae and Asteraceae relative to other
369 families, suggesting action of diversifying selection in these lineages (Fig. 6d). At least
370 30% of species examined had more than one paralog per genome most of which
371 appeared to represent recent lineage-specific duplications as they tended to group
372 together in the tree. However, Asteraceae appears to have experienced a duplication
373 early in their history. Likewise the duplication in Brassicaceae and Cleomaceae
374 predated their divergence. We further assessed diversifying selection at *SHUKR* within
375 Brassicaceae based on an alignment spanning the length of the gene, using branch-
376 site models[82] that consider proteins as a mosaic of conserved and divergent domains,.
377 The branch-site test results provided further support for existence of positive selection
378 ($p = 0.0031$) (Extended Data Table 1; Extended Data Fig. 8; Supplementary Figure 2),
379 demonstrating that gene duplication has been followed by diversifying selection at
380 *SHUKR*.
381

382 **Adaptive value for variation in protein turnover during male gametogenesis**

383 Plant F-box genes have undergone independent lineage-specific expansions as well
384 as reductions during land plant evolution[83,84]. The Brassicaceae family has an
385 expanded set of F-box genes, a majority of which are of short taxonomic scale
386 indicative of more recent emergence[84]. These recent members were significantly

387 enriched among upregulated F-box genes in *skr* ($p < 1.5e-05$) (Fig. 7a,c;
388 Supplementary Table 5). Orthology analysis of the total set of F-box genes from 10
389 eudicot species[84] indicated that 58% (37/64) of F-box genes upregulated in *skr*
390 belonged to orthogroups that included *Arabidopsis lyrata* but not 8 other eudicot
391 species examined spanning 7 families including *Carica papaya* (family *Caricaceae*,
392 order *Brassicales*), suggesting that these were likely specific to the Brassicaceae
393 lineage (Fig. 7b; Supplementary Table 5). 8% (5/64) of genes belonged to orthogroups
394 having only *A. thaliana* and not *A. lyrata* suggesting a more recent origin, and 23%
395 (15/64) had at least one other species but not *A. lyrata* indicative of recent loss (Fig.
396 7b). We examined the possibility of diversifying selection at the set of *skr* upregulated
397 genes by testing whether the ratio K_a/K_s of non-synonymous to synonymous
398 substitutions per site was higher relative to the overall set of F-box genes. Plots of
399 K_a/K_s gave significantly higher values for the set of F-box genes upregulated in *skr*
400 relative to the overall set of F-box genes ($p = 0.0075$; Fig. 7d). Hence the set of *skr*
401 upregulated F-box is itself undergoing diversifying selection. Interestingly, male
402 gametes have been proposed as a source for emergence of new genes in both
403 animals and plants[85–87], and pollen-preferred F-box genes also appear to belong to
404 this set. Variation in protein turnover during male gametophyte development can
405 therefore be seen to operate at two levels in evolution. One is through rapid evolution
406 and positive selection in a regulator of UPS genes involved in male gametophyte
407 development. The second is in changes and diversifying selection in components of
408 the UPS system itself, the set of F-box genes that direct substrates for proteasomal
409 degradation in gametogenesis. The large number of F-box genes involved and their
410 distribution across the F-box phylogenetic tree (Fig. 7c) suggests the likelihood of
411 levels of multiple proteins being subject to regulation by SKR. Variation in this control

412 provides a plausible route for the sporophyte to influence development and evolution
413 of the gametophyte. These findings lead us to hypothesize that processes that provide
414 for variation in levels of large sets of proteins rather than just a single target during
415 male gametophyte development have adaptive value and have contributed
416 significantly to evolution in Brassicaceae.
417

418 ## Discussion

419 Extensive functional and gene expression studies of gametophyte development[15,18,88]
420 have supported the view that the male gametophyte retains control of its development
421 even after evolutionary reduction to just a few cells enclosed within the sporophyte.
422 Experiments in angiosperm species have also shown that isolated microspores can
423 form pollen in culture[52,53,89], supporting this view. Here we have demonstrated in
424 Arabidopsis that the sporophyte in fact plays an important role in control of
425 gametogenesis. This direction is by a repressive mechanism involving large scale
426 control of regulated protein turnover in male gametophyte development by a novel
427 gene, *SHUKR*. *SHUKR* acts during meiosis and is an inhibitor of gametogenesis (Fig.
428 5g). Our findings in Arabidopsis suggest that significant aspects of the male
429 gametogenesis programme are established in the diploid sporophyte during meiosis
430 but held in a repressed condition until microspores are released. We suggest the term
431 sporophyte directed gametogenesis (SDG) for this phenomenon.
432

433 Recent emergence of *SHUKR* at the base of eudicots suggests that SDG arose late
434 in evolution following reduction of the gametophyte. Gametophyte reduction and its
435 close proximity to the sporophyte may have provided conditions necessary for the
436 emergence of SDG. Rapid evolution and diversifying selection at *SHUKR* and the set

437 of F-box genes under its control points to an adaptive value for variation in protein
438 turnover in male gametophyte evolution. This variation provides a way for the
439 sporophyte to generate diversity in the male gametophytes which face intense
440 selective pressure in the course of transmission and intraspecific competition[90]. The
441 location of this control in the sporophyte is also of interest because it raises the
442 possibility that gametophyte development can be modulated by the sporophyte as a
443 function of its own status and environment.

444 ## References

445 1. Graham, L. E. The Origin of the Life Cycle of Land Plants: A simple modification
446 in the life cycle of an extinct green alga is the likely origin of the first land plants.
447 *American Scientist* **73**, 178–186 (1985).
448 2. Kenrick, P. Alternation of Generations in Land Plants: New Phylogenetic and
449 Palaeobotanical Evidence. *Biological Reviews* **69**, 293–330 (1994).
450 3. Niklas, K. J. & Kutschera, U. The evolution of the land plant life cycle. *New*
451 *Phytologist* **185**, 27–41 (2010).
452 4. Graham, L. E. Origin of Land Plants. *Wiley.com* https://www.wiley.com/en-
453 us/Origin+of+Land+Plants-p-9780471615279 (1993).
454 5. Delwiche, C. F. & Cooper, E. D. The Evolutionary Origin of a Terrestrial Flora.
455 *Curr Biol* **25**, R899-910 (2015).
456 6. Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land.
457 *Nature* **389**, 33–39 (1997).
458 7. Graham, L. E., Cook, M. E. & Busse, J. S. The origin of plants: Body plan
459 changes contributing to a major evolutionary radiation. *Proceedings of the*
460 *National Academy of Sciences* **97**, 4535–4540 (2000).
461 8. Ma, H. & Sundaresan, V. Chapter Thirteen - Development of Flowering Plant
462 Gametophytes. in *Current Topics in Developmental Biology* (ed. Timmermans,
463 M. C. P.) vol. 91 379–412 (Academic Press, 2010).
464 9. Harrison, C. J. & Morris, J. L. The origin and early evolution of vascular plant
465 shoots and leaves. *Philos Trans R Soc Lond B Biol Sci* **373**, 20160496 (2018).
466 10. Sakakibara, K. *et al.* KNOX2 genes regulate the haploid-to-diploid morphological
467 transition in land plants. *Science* **339**, 1067–1070 (2013).

468 11. Mosquna, A. *et al.* Regulation of stem cell maintenance by the Polycomb protein
469 FIE has been conserved during land plant evolution. *Development* **136**, 2433–
470 2444 (2009).
471 12. Okano, Y. *et al.* A polycomb repressive complex 2 gene regulates apogamy and
472 gives evolutionary insights into early land plant evolution. *Proc Natl Acad Sci U S*
473 *A* **106**, 16321–16326 (2009).
474 13. Strome, S. & Lehmann, R. Germ versus soma decisions: lessons from flies and
475 worms. *Science* **316**, 392–393 (2007).
476 14. Hayashi, K. & Surani, M. A. Self-renewing epiblast stem cells exhibit continual
477 delineation of germ cells with epigenetic reprogramming in vitro. *Development*
478 **136**, 3549–3556 (2009).
479 15. Berger, F. & Twell, D. Germline Specification and Function in Plants. *Annual*
480 *Review of Plant Biology* **62**, 461–484 (2011).
481 16. Hisanaga, T. *et al.* Building new insights in plant gametogenesis from an
482 evolutionary perspective. *Nat Plants* **5**, 663–669 (2019).
483 17. Grossniklaus, U. Plant germline development: a tale of cross-talk, signaling, and
484 cellular interactions. *Sex Plant Reprod* **24**, 91–95 (2011).
485 18. Twell, D. Male gametogenesis and germline specification in flowering plants. *Sex*
486 *Plant Reprod* **24**, 149–160 (2011).
487 19. Yamaoka, S. *et al.* Generative Cell Specification Requires Transcription Factors
488 Evolutionarily Conserved in Land Plants. *Curr Biol* **28**, 479-486.e5 (2018).
489 20. Saito, M. *et al.* A bHLH heterodimer regulates germ cell differentiation in land
490 plant gametophytes. *Curr Biol* **33**, 4980-4987.e6 (2023).

491 21. Borg, M. *et al.* The R2R3 MYB Transcription Factor DUO1 Activates a Male
492 Germline-Specific Regulon Essential for Sperm Cell Differentiation in
493 Arabidopsis. *The Plant Cell* **23**, 534–549 (2011).
494 22. Higo, A. *et al.* Transcription factor DUO1 generated by neo-functionalization is
495 associated with evolution of sperm differentiation in plants. *Nat Commun* **9**, 5283
496 (2018).
497 23. Kasahara, R. D., Portereiko, M. F., Sandaklie-Nikolova, L., Rabiger, D. S. &
498 Drews, G. N. MYB98 is required for pollen tube guidance and synergid cell
499 differentiation in Arabidopsis. *Plant Cell* **17**, 2981–2992 (2005).
500 24. Gross-Hardt, R. *et al.* LACHESIS restricts gametic cell fate in the female
501 gametophyte of Arabidopsis. *PLoS Biol* **5**, e47 (2007).
502 25. Moll, C. *et al.* CLO/GFA1 and ATO are novel regulators of gametic cell fate in
503 plants. *Plant J* **56**, 913–921 (2008).
504 26. Rabiger, D. S. & Drews, G. N. MYB64 and MYB119 Are Required for
505 Cellularization and Differentiation during Female Gametogenesis in Arabidopsis
506 thaliana. *PLOS Genetics* **9**, e1003783 (2013).
507 27. Yuan, L. *et al.* The CKI1 Histidine Kinase Specifies the Female Gametic
508 Precursor of the Endosperm. *Developmental Cell* **37**, 34–46 (2016).
509 28. Portereiko, M. F. *et al.* AGL80 Is Required for Central Cell and Endosperm
510 Development in Arabidopsis. *The Plant Cell* **18**, 1862–1872 (2006).
511 29. Bemer, M., Wolters-Arts, M., Grossniklaus, U. & Angenent, G. C. The MADS
512 Domain Protein DIANA Acts Together with AGAMOUS-LIKE80 to Specify the
513 Central Cell in Arabidopsis Ovules. *The Plant Cell* **20**, 2088–2101 (2008).

514 30. Bateman, R. M. & DiMICHELE, W. A. Heterospory: The Most Iterative Key
515 Innovation in the Evolutionary History of the Plant Kingdom. *Biological Reviews*
516 **69**, 345–417 (1994).
517 31. Feng, X., Zilberman, D. & Dickinson, H. A conversation across generations:
518 soma-germ cell crosstalk in plants. *Dev Cell* **24**, 215–225 (2013).
519 32. Schmidt, A., Schmid, M. W. & Grossniklaus, U. Plant germline formation:
520 common concepts and developmental flexibility in sexual and asexual
521 reproduction. *Development* **142**, 229–241 (2015).
522 33. Petrella, R., Cucinotta, M., Mendes, M. A., Underwood, C. J. & Colombo, L. The
523 emerging role of small RNAs in ovule development, a kind of magic. *Plant*
524 *Reprod* **34**, 335–351 (2021).
525 34. Yang, W. C., Ye, D., Xu, J. & Sundaresan, V. The SPOROCYTELESS gene of
526 Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear
527 protein. *Genes Dev* **13**, 2108–2117 (1999).
528 35. Schiefthaler, U. *et al.* Molecular analysis of NOZZLE, a gene involved in pattern
529 formation and early sporogenesis during sex organ development in Arabidopsis
530 thaliana. *Proc Natl Acad Sci U S A* **96**, 11664–11669 (1999).
531 36. Mendes, M. A. *et al.* The RNA-dependent DNA methylation pathway is required
532 to restrict SPOROCYTELESS/NOZZLE expression to specify a single female
533 germ cell precursor in Arabidopsis. *Development* **147**, dev194274 (2020).
534 37. Zhao, D.-Z., Wang, G.-F., Speal, B. & Ma, H. The excess microsporocytes1 gene
535 encodes a putative leucine-rich repeat receptor protein kinase that controls
536 somatic and reproductive cell fates in the Arabidopsis anther. *Genes Dev* **16**,
537 2021–2031 (2002).

538 38. Canales, C., Bhatt, A. M., Scott, R. & Dickinson, H. EXS, a putative LRR receptor
539 kinase, regulates male germline cell number and tapetal identity and promotes
540 seed development in Arabidopsis. *Curr Biol* **12**, 1718–1727 (2002).
541 39. Jia, G., Liu, X., Owen, H. A. & Zhao, D. Signaling of cell fate determination by the
542 TPD1 small protein and EMS1 receptor kinase. *Proc Natl Acad Sci U S A* **105**,
543 2220–2225 (2008).
544 40. Nonomura, K.-I. *et al.* A germ cell specific gene of the ARGONAUTE family is
545 essential for the progression of premeiotic mitosis and meiosis during
546 sporogenesis in rice. *Plant Cell* **19**, 2583–2594 (2007).
547 41. Singh, M. *et al.* Production of viable gametes without meiosis in maize deficient
548 for an ARGONAUTE protein. *Plant Cell* **23**, 443–458 (2011).
549 42. Su, Z. *et al.* The THO Complex Non-Cell-Autonomously Represses Female
550 Germline Specification through the TAS3-ARF3 Module. *Curr Biol* **27**, 1597-
551 1609.e2 (2017).
552 43. Martínez, G., Panda, K., Köhler, C. & Slotkin, R. K. Silencing in sperm cells is
553 directed by RNA movement from the surrounding nurse cell. *Nat Plants* **2**, 16030
554 (2016).
555 44. Pagnussat, G. C., Alandete-Saez, M., Bowman, J. L. & Sundaresan, V. Auxin-
556 dependent patterning and gamete specification in the Arabidopsis female
557 gametophyte. *Science* **324**, 1684–1689 (2009).
558 45. Panoli, A. *et al.* Auxin Import and Local Auxin Biosynthesis Are Required for
559 Mitotic Divisions, Cell Expansion and Cell Specification during Female
560 Gametophyte Development in Arabidopsis thaliana. *PLoS One* **10**, e0126164
561 (2015).

562 46. Gómez, J. F., Talle, B. & Wilson, Z. A. Anther and pollen development: A
563 conserved developmental pathway. *Journal of Integrative Plant Biology* **57**, 876–
564 891 (2015).
565 47. Ariizumi, T. & Toriyama, K. Genetic regulation of sporopollenin synthesis and
566 pollen exine development. *Annu Rev Plant Biol* **62**, 437–460 (2011).
567 48. Wilson, Z. A., Morroll, S. M., Dawson, J., Swarup, R. & Tighe, P. J. The
568 Arabidopsis MALE STERILITY1 (MS1) gene is a transcriptional regulator of male
569 gametogenesis, with homology to the PHD-finger family of transcription factors.
570 *Plant J* **28**, 27–39 (2001).
571 49. Sorensen, A.-M. *et al.* The Arabidopsis ABORTED MICROSPORES (AMS) gene
572 encodes a MYC class transcription factor. *Plant J* **33**, 413–423 (2003).
573 50. Zhang, W. *et al.* Regulation of Arabidopsis tapetum development and function by
574 DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription
575 factor. *Development* **133**, 3085–3095 (2006).
576 51. Long, J. *et al.* Nurse cell--derived small RNAs define paternal epigenetic
577 inheritance in Arabidopsis. *Science* **373**, eabh0556 (2021).
578 52. Tanaka, I. & Ito, M. Induction of typical cell division in isolated microspores of
579 Lilium longiflorum and Tulipa gesneriana. *Plant Science Letters* **17**, 279–285
580 (1980).
581 53. Moreno, R. M., Macke, F., Alwen, A. & Heberle-Bors, E. In-situ seed production
582 after pollination with in-vitro-matured, isolated pollen. *Planta* **176**, 145–148
583 (1988).
584 54. Chaudhury, A. M. Nuclear Genes Controlling Male Fertility. *Plant Cell* **5**, 1277–
585 1283 (1993).

586 55. Mercier, R. & Grelon, M. Meiosis in plants: ten years of gene discovery.
587 *Cytogenet Genome Res* **120**, 281–290 (2008).
588 56. Blackmore, S., Wortley, A. H., Skvarla, J. J. & Rowley, J. R. Pollen wall
589 development in flowering plants. *New Phytol* **174**, 483–498 (2007).
590 57. Scott, R. J., Spielman, M. & Dickinson, H. G. Stamen structure and function.
591 *Plant Cell* **16 Suppl**, S46-60 (2004).
592 58. Bencivenga, S., Colombo, L. & Masiero, S. Cross talk between the sporophyte
593 and the megagametophyte during ovule development. *Sex Plant Reprod* **24**,
594 113–121 (2011).
595 59. Figueiredo, D. D. & Köhler, C. Bridging the generation gap: communication
596 between maternal sporophyte, female gametophyte and fertilization products.
597 *Curr Opin Plant Biol* **29**, 16–20 (2016).
598 60. McCormick, S. Male Gametophyte Development. *Plant Cell* **5**, 1265–1275
599 (1993).
600 61. Owen, H. A. & Makaroff, C. A. Ultrastructure of microsporogenesis and
601 microgametogenesis inArabidopsis thaliana (L.) Heynh. ecotype Wassilewskija
602 (Brassicaceae). *Protoplasma* **185**, 7–21 (1995).
603 62. Twell, D. Use of a nuclear-targeted β-glucuronidase fusion protein to
604 demonstrate vegetative cell-specific gene expression in developing pollen. *The*
605 *Plant Journal* https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-
606 313X.1992.00887.x (1992).
607 63. Dobritsa, A. A. & Coerper, D. The novel plant protein INAPERTURATE
608 POLLEN1 marks distinct cellular domains and controls formation of apertures in
609 the Arabidopsis pollen exine. *Plant Cell* **24**, 4452–4464 (2012).

610 64. Ma, H. Molecular genetic analyses of microsporogenesis and
611 microgametogenesis in flowering plants. *Annu Rev Plant Biol* **56**, 393–434
612 (2005).
613 65. Proost, S. & Mutwil, M. CoNekT: an open-source framework for comparative
614 genomic and transcriptomic network analyses. *Nucleic Acids Res* **46**, W133–
615 W140 (2018).
616 66. Honys, D. & Twell, D. Comparative analysis of the Arabidopsis pollen
617 transcriptome. *Plant Physiol* **132**, 640–652 (2003).
618 67. Pina, C., Pinto, F., Feijó, J. A. & Becker, J. D. Gene Family Analysis of the
619 Arabidopsis Pollen Transcriptome Reveals Biological Implications for Cell
620 Growth, Division Control, and Gene Expression Regulation. *Plant Physiol* **138**,
621 744–756 (2005).
622 68. Ichino, L. *et al.* Single-nucleus RNA-seq reveals that MBD5, MBD6, and
623 SILENZIO maintain silencing in the vegetative cell of developing pollen. *Cell Rep*
624 **41**, 111699 (2022).
625 69. Honys, D. *et al.* Identification of microspore-active promoters that allow targeted
626 manipulation of gene expression at early stages of microgametogenesis in
627 Arabidopsis. *BMC Plant Biol* **6**, 31 (2006).
628 70. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and
629 conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191–W198
630 (2019).
631 71. Moon, J., Parry, G. & Estelle, M. The ubiquitin-proteasome pathway and plant
632 development. *Plant Cell* **16**, 3181–3195 (2004).
633 72. Smalle, J. & Vierstra, R. D. The ubiquitin 26S proteasome proteolytic pathway.
634 *Annu Rev Plant Biol* **55**, 555–590 (2004).

635 73. Mladek, C., Guger, K. & Hauser, M.-T. Identification and characterization of the
636 ARIADNE gene family in Arabidopsis. A group of putative E3 ligases. *Plant*
637 *Physiol* **131**, 27–40 (2003).
638 74. Gagne, J. M., Downes, B. P., Shiu, S.-H., Durski, A. M. & Vierstra, R. D. The F-
639 box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes
640 in Arabidopsis. *Proc Natl Acad Sci U S A* **99**, 11519–11524 (2002).
641 75. Kuroda, H. *et al.* Classification and expression analysis of Arabidopsis F-box-
642 containing protein genes. *Plant Cell Physiol* **43**, 1073–1085 (2002).
643 76. Ikram, S. *et al.* Functional redundancy and/or ongoing pseudogenization among
644 F-box protein genes expressed in Arabidopsis male gametophyte. *Plant Reprod*
645 **27**, 95–107 (2014).
646 77. Branscheid, A. *et al.* SKI2 mediates degradation of RISC 5'-cleavage fragments
647 and prevents secondary siRNA production from miRNA targets in Arabidopsis.
648 *Nucleic Acids Res* **43**, 10975–10988 (2015).
649 78. Zhang, X. *et al.* Suppression of endogenous gene silencing by bidirectional
650 cytoplasmic RNA decay in Arabidopsis. *Science* **348**, 120–123 (2015).
651 79. Fu, H., Doelling, J. H., Arendt, C. S., Hochstrasser, M. & Vierstra, R. D.
652 Molecular organization of the 20S proteasome gene family from Arabidopsis
653 thaliana. *Genetics* **149**, 677–692 (1998).
654 80. Hatsugai, N. *et al.* A novel membrane fusion-mediated plant immunity against
655 bacterial pathogens. *Genes Dev* **23**, 2496–2506 (2009).
656 81. Krishnan, A., Iyer, L. M., Holland, S. J., Boehm, T. & Aravind, L. Diversification of
657 AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread
658 roles in immunity. *Proc Natl Acad Sci U S A* **115**, E3201–E3210 (2018).

659 82. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*
660 **24**, 1586–1591 (2007).
661 83. Xu, G., Ma, H., Nei, M. & Kong, H. Evolution of F-box genes in plants: Different
662 modes of sequence divergence and their relationships with functional
663 diversification. *Proc Natl Acad Sci U S A* **106**, 835–840 (2009).
664 84. Hua, Z., Zou, C., Shiu, S.-H. & Vierstra, R. D. Phylogenetic Comparison of F-Box
665 (FBX) Gene Superfamily within the Plant Kingdom Reveals Divergent
666 Evolutionary Histories Indicative of Genomic Drift. *PLOS ONE* **6**, e16219 (2011).
667 85. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes.
668 *Genome Res* **20**, 1313–1326 (2010).
669 86. Wu, D.-D. *et al.* “Out of Pollen” Hypothesis for Origin of New Genes in Flowering
670 Plants: Study from Arabidopsis thaliana. *Genome Biology and Evolution* **6**, 2822–
671 2829 (2014).
672 87. Cui, X. *et al.* Young Genes out of the Male: An Insight from Evolutionary Age
673 Analysis of the Pollen Transcriptome. *Mol Plant* **8**, 935–945 (2015).
674 88. Nelms, B. & Walbot, V. Gametophyte genome activation occurs at pollen mitosis
675 I in maize. *Science* **375**, 424–429 (2022).
676 89. Barinova, I. *et al.* Antirrhinum majus microspore maturation and transient
677 transformation in vitro. *J Exp Bot* **53**, 1119–1129 (2002).
678 90. Delph, L. F. Pollen competition is the mechanism underlying a variety of
679 evolutionary phenomena in dioecious plants. *New Phytol* **224**, 1075–1079
680 (2019).
681 91. Zhao, L. & Kunst, L. SUPERKILLER Complex Components Are Required for the
682 RNA Exosome-Mediated Control of Cuticular Wax Biosynthesis in Arabidopsis
683 Inflorescence Stems. *Plant Physiol* **171**, 960–973 (2016).

684 92. Toufighi, K., Brady, S. M., Austin, R., Ly, E. & Provart, N. J. The Botany Array
685 Resource: e-Northerns, Expression Angling, and promoter analyses. *Plant J* **43**,
686 153–163 (2005).
687 93. Schmid, M. *et al.* A gene expression map of Arabidopsis thaliana development.
688 *Nat Genet* **37**, 501–506 (2005).
689 94. Alexander, M. P. Differential staining of aborted and nonaborted pollen. *Stain*
690 *Technol* **44**, 117–122 (1969).
691 95. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-
692 mediated transformation of Arabidopsis thaliana. *Plant J* **16**, 735–743 (1998).
693 96. Karimi, M., Depicker, A. & Hilson, P. Recombinational cloning with plant gateway
694 vectors. *Plant Physiol* **145**, 1144–1154 (2007).
695 97. Reddy, T. V., Kaur, J., Agashe, B., Sundaresan, V. & Siddiqi, I. The DUET gene
696 is necessary for chromosome organization and progression during male meiosis
697 in Arabidopsis and encodes a PHD finger protein. *Development* **130**, 5975–5987
698 (2003).
699 98. Sanders, P. M. *et al.* Anther developmental defects in Arabidopsis thaliana male-
700 sterile mutants. *Sex Plant Reprod* **11**, 297–322 (1999).
701 99. Schneitz, K., Hülskamp, M. & Pruitt, R. E. Wild-type ovule development in
702 Arabidopsis thaliana: a light microscope study of cleared whole-mount tissue.
703 *The Plant Journal* **7**, 731–749 (1995).
704 100. Andreuzza, S., Nishal, B., Singh, A. & Siddiqi, I. The Chromatin Protein
705 DUET/MMD1 Controls Expression of the Meiotic Gene TDM1 during Male
706 Meiosis in Arabidopsis. *PLoS Genet* **11**, e1005396 (2015).

707 101. Singh, D. K., Spillane, C. & Siddiqi, I. PATRONUS1 is expressed in meiotic
708 prophase I to regulate centromeric cohesion in Arabidopsis and shows synthetic
709 lethality with OSD1. *BMC Plant Biol* **15**, 201 (2015).
710 102. Siddiqi, I., Ganesh, G., Grossniklaus, U. & Subbiah, V. The dyad gene is
711 required for progression through female meiosis in Arabidopsis. *Development*
712 **127**, 197–207 (2000).
713 103. Schwartz, J. *et al.* Removing Stripes, Scratches, and Curtaining with
714 Nonrecoverable Compressed Sensing. *Microscopy and Microanalysis* **25**, 705–
715 710 (2019).
716 104. Jiao, W.-B. & Schneeberger, K. Chromosome-level assemblies of multiple
717 Arabidopsis genomes reveal hotspots of rearrangements with altered
718 evolutionary dynamics. *Nat Commun* **11**, 989 (2020).
719 105. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature
720 Analysis. *Curr Protoc Bioinformatics* **47**, 11.12.1-34 (2014).
721 106. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read
722 sequencing. Preprint at https://doi.org/10.48550/arXiv.1207.3907 (2012).
723 107. Mansfeld, B. N. & Grumet, R. QTLseqr: An R Package for Bulk Segregant
724 Analysis with Next-Generation Sequencing. *Plant Genome* **11**, (2018).
725 108. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
726 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
727 109. Walker, J. *et al.* Lineage-specific DNA methylation regulates Arabidopsis
728 sexual development. *Nat Genet* **50**, 130–137 (2018).
729 110. Xu, D. *et al.* BBX21, an Arabidopsis B-box protein, directly activates HY5 and
730 is targeted by COP1 for 26S proteasome-mediated degradation. *Proc Natl Acad*
731 *Sci U S A* **113**, 7655–7660 (2016).

732 111. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein
733 database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).
734 112. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res* **46**,
735 W200–W204 (2018).
736 113. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics
737 Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**, 2237–2243 (2018).
738 114. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids*
739 *Res* **49**, D412–D419 (2021).
740 115. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242
741 (2000).
742 116. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein
743 secondary structure prediction server. *Nucleic Acids Res* **43**, W389-394 (2015).
744 117. Baek, M. *et al.* Accurate prediction of protein structures and interactions using
745 a three-track neural network. *Science* **373**, 871–876 (2021).
746 118. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant
747 genomics. *Nucleic Acids Res* **40**, D1178-1186 (2012).
748 119. Lyons, E. *et al.* Finding and comparing syntenic regions among Arabidopsis
749 and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol*
750 **148**, 1772–1781 (2008).
751 120. Numa, H. & Itoh, T. MEGANTE: A Web-Based System for Integrated Plant
752 Genome Annotation. *Plant Cell Physiol* **55**, e2 (2014).
753 121. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software
754 version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780
755 (2013).

756 122. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-
757 analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
758 123. Gao, F. *et al.* EasyCodeML: A visual tool for analysis of selection using
759 CodeML. *Ecol Evol* **9**, 3891–3898 (2019).
760 124. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for
761 Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).
762 125. Guindon, S. *et al.* New algorithms and methods to estimate maximum-
763 likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**,
764 307–321 (2010).
765 126. Lechner, M. *et al.* Proteinortho: Detection of (Co-)orthologs in large-scale
766 analysis. *BMC Bioinformatics* **12**, 124 (2011).
767

768 ## Figure Legends

769 **Fig. 1: *SHUKR* is required for pollen development.** (a) Schematic showing
770 reduction of the gametophyte in flowering plants. (b) Sterility (short siliques) and
771 reduced pollen viability (green staining) in *skr*, and complementation by a genomic
772 *SKR* transgene. (c) Tables showing reduced seed set, male-specific sterility, and
773 recessive Mendelian segregation for *skr*; numbers in brackets are those expected for
774 a 3:1 ratio and the χ^2 test p value. (d) RT-PCR analysis of SKR expression in
775 vegetative and reproductive tissues. (e) GUS staining for *proSKR:GFP-GUS*
776 transgene expression in successive unopened buds. (f) GUS staining for
777 *proSKR:GFP-GUS* expression in meiotic stage anthers. (g) anti-GFP antibody staining
778 for *proSKR:GFP-GUS* expression in male meiosis and microspores. (h) Normal tetrad
779 formation in *skr*. Scale bar: (e) 0.5 mm, (f) 20 μm.
780

781 **Fig. 2: Staging of *skr* phenotype and SKR protein expression**. (a) Meiotic
782 chromosome development and segregation in wild type (a-j) and corresponding stages
783 in *skr* (a'-j'). a-d,a'-d' prophase I; e,e' metaphase I; f,f' telophase I; g,g' metaphase II;
784 h,h' anaphase II; i,i' telophase II; j,j' tetrad. 5846 meiocytes were scored for *skr* and
785 2281 for WT. (b) Whole mount DIC images of cleared anthers from successive stage
786 buds; arrowheads indicate degenerating microspores starting in anthers containing
787 early microspores. (c) Anti-GFP staining of male meiosis stages in a *proSKR:GFP-*
788 *gSKR* complementing line; asterisks: somatic/tapetal cells; arrowheads: older
789 uninucleate microspores with low GFP signal. Scale bar: (a) 5 µm, (b,c) 20 µm.
790
791 **Fig. 3: Ultrastructure and marker expression reveal early gametogenesis defects**
792 **in *skr*.** (a-c) Ultrastructure of tetrad and early uninucleate microspores showing onset
793 of mutant phenotype at tetrad stage in *skr*. (a) WT: Normal tetrad and microspore; *skr*:
794 normal tetrad, degenerating microspore with retracted PM. (b) *skr* tetrad in which one
795 spore shows PM retraction from callose wall. (c) Magnified image of boxed region from
796 (b) showing PM retraction from callose wall (Cal) and disrupted cell wall material
797 (arrowhead). (d) Quantitation of normal spores and spores showing PM retraction and
798 broken cell wall material (R&B) in tetrads (Fisher's exact test $p < 0.001$). (e) Presence
799 of two LAT52:GFP signals in a proportion of *skr* microspores (Fisher's exact test $p <$
800 0.001). Scale bar (a): Tetrad images 5 µm; microspore images 2 µm; (b) 5 µm; (C) 1
801 µm.
802
803 **Fig. 4: *SHUKR* is a gametogenesis inhibitor**. (a) Reciprocal crosses comparing
804 transmission of *skr* and wild type alleles; numbers in brackets are those expected for
805 equal transmission and χ^2 test p-value. (b) Tissue/cell enrichment categories for

806 upregulated genes in *skr*. (c) Pollen viability in WT and *skr* plants transgenic for
807 *proMSP1:SKRcds* or control. Data are pools of 4-10 anthers from 5 independent
808 transgenic lines in each case. Scale bar (c) 100 µm.
809
810 **Fig. 5: *SHUKR* regulates the proteasome in pollen development.** (a) Gene family
811 categories for upregulated genes in *skr*. (b) Stagewise expression of an
812 *At3g28330:2xnlsGFP* transgene during meiosis and in microspores for *skr* and
813 wildtype. Frames: a,f meiocytes; b,g tetrad; c-e,h microspores; upper half: GFP, lower
814 half: DIC. (c) Suppression of *skr* male sterility by *ski2*. Bar graph shows viable pollen
815 counts in single anthers from 2-4 plants. (d) Downregulation of proteasome genes in
816 *skr ski2*. (e) Mean viable pollen counts/anther in independent *skr; proSKR:PBE1RNAi*
817 and control *T1* transgenic plants. (f) Pollen viability increase in *skr* following
818 proteasome inhibitor treatment in 3 replicate experiments. (g) Model for sporophytic
819 control of pollen development by *SHUKR*. During meiosis *SHUKR* establishes a
820 programme that determines timing of expression of F-box and other UPS genes
821 following meiosis. Scale bar: (b) 5 µm; (c) plant 1 cm, anthers 100 µm. Abbreviations:
822 UNM, uninucleate microspore; BCP, bicellular pollen; TCP, tricellular pollen; MP,
823 mature pollen; SGT, sporophyte-gametophyte transition.
824
825 **Fig. 6: Sequence and evolutionary analysis of the SHUKR domain.** (a) Multiple
826 sequence alignment of the SHUKR domain. Lines mark CxC and aromatic dyad. (b)
827 Phylogenetic tree of the SHUKR domain from 119 species. Dark green,
828 Brassicaceae+Cleomaceae; Light green, Asteraceae. (c) Distribution of mean-
829 centered Maximum Likelihood (ML) distance differences. (d) Columnwise Shannon

830 entropy distributions for four plant families. P-value calculated using Kruskal-Wallis
831 test for Brassicaceae and Asteraceae relative to the other two families.
832
833 **Fig. 7: Diversifying selection in SHUKR regulated F-box genes.** (a) Enrichment for
834 short taxonomic scale (STS) F-box genes (R.F. = 1.2) and underenrichment for long
835 taxonomic scale (LTS) genes (R.F. = 0.1) among *skr* upregulated set ($p < 1.49e-05$).
836 (b) 37 genes were unique to the Brassicaceae lineage of which 5 were found only in
837 *Arabidopsis thaliana*. 15 genes were present in *Arabidopsis thaliana* and at least one
838 other species but absent in *Arabidopsis lyrata*. (c) Phylogenetic tree of Arabidopsis F-
839 box genes showing *skr* upregulated, LTS, and STS sets. (d) Violin plot of K_a/K_s for *skr*
840 upregulated F-box and total F-box genes. Other eudicots: *Cucumis sativus*, *Glycine*
841 *max*, *Manihot esculenta*, *Mimulus guttatus*, *Populus trichocarpa*, *Ricinus communis*,
842 *Vitis vinifera*. Abbreviations: Al *Arabidopsis lyrata*, At *Arabidopsis thaliana*, Cp *Carica*
843 *papaya*.
844

845 ## Extended Data Figure Legends

846 **Extended Data Fig. 1:**
847 (a) Gene diagram of *SKR* showing location of insertions: *skr-1* (GT_5_101780), *skr-2*
848 (CSHL_GT5061), *skr-3* (SALKSeq_034497.1). (b) Toluidine blue stained tetrads. (c)
849 Inflorescence stem and Alexander staining showing recessive male sterility of *skr.* (d)
850 Complementation of *skr-3* by a *gSKR* transgene. Scale bar: 1 cm for inflorescence,
851 100 μm for anthers.
852

853 **Extended Data Fig. 2:**
854 Profiling of SKR expression in male germline development.

855 TPM values for SKR were collected from the CoNekt database, except for meiocytes.
856 Meiocyte TPM value were taken from (Walker 2018)[21]. UNM – Uninucleate
857 microspore; BCP – Bicellular pollen; TCP – Tricellular pollen; MP – Mature pollen.
858
859 **Extended Data Fig. 3:**
860 Sections of anther development stages.
861 Corresponding stages are shown for wild type (A-D, I-L) and *skr* (E-H, M-P). Stages
862 are numbered according to (Sanders 1999)[9]. Abbreviations: E epidermis, En
863 endothecium, ML middle layer, T tapetum, Td tetrad, Msp microspore, DM
864 degenerating microspore, PG pollen grain, St stomium, MD microspore debris. Scale
865 bar 20 µm.
866
867 **Extended Data Fig. 4:**
868 Staging of microspore degeneration in *skr*.
869 Microspore stage and phenotypes are shown together with stage of ovules from the
870 same bud (Schneitz 1995)[10] as a reference for comparison to wild type
871
872 **Extended Data Fig. 5:**
873 Precocious expression of F-box gene:nlsGFP fusions in meiosis in *skr*.
874 (a) At5g02980. A-C, A'-C': wild type; D-F, D'-F': skr; A,A',D,D' meiocytes; B,B',E,E':
875 tetrad; C,C',F,F': uninucleate microspores. (b) At5g62510. A-E, A'-E': wild type; F-H,
876 F'-H': skr; A,A',F,F' meiocytes; B,B',G,G': tetrad; C-E, C'-E',H,H': uninucleate
877 microspores. A-H: GFP fluorescence; A'-H': corresponding DIC. GFP expression in
878 *skr* is first detected in meiocytes (At5g02980) or tetrad stage At5g62510) whereas in
879 wild type earliest GFP signal was in released microspores. Scale bar: 5 µm

880

881 **Extended Data Fig. 6:**

882 *ski2* is a suppressor of *skr*.

883 (a) Mendelian segregation of the suppressor mutation in the 08-10 *skr* suppressor line
884 (b) A C>T transition in the *ski2* allele (named *ski2-7*) present in the 08-10 *skr*
885 suppressor line results in a TGA termination codon at position 518. (c) Restoration of
886 fertility and pollen viability in *skr-3* by the independent allele *ski2-6*. Scale bar:
887 inflorescence 1 cm; anthers 100 μm.

888

889 **Extended Data Fig. 7:**

890 Phylogenetic tree of the SHUKR family from eudicots.

891 119 representatives from 85 species were used for tree construction using only the
892 SHUKR domain that is shared by all representatives of the family. Genbank ID and
893 species are indicated. Brassicaceae+Cleomaceae (dark green) and Asteraceae (light
894 green)

895

896 **Extended Data Fig. 8:**

897 Phylogenetic tree of Brassicales *SHUKR* and *SHUKR-like* (*SKL*) genes used for
898 branch-site test.

899 The *SHUKR* gene family is represented by a single gene in *Carica papaya*, whereas
900 Brassicaceae and Cleomaceae experienced a gene duplication event prior to their
901 divergence. CDS were aligned using MAFFT, and RAxML was used to build the
902 phylogenetic tree with 1000 bootstraps. The red branch was selected as foreground
903 for *SKR*. Branch-site test for positive selection was performed using the CODEML
904 package implemented in EASYCoDEML. Likelihood Ratio Test (LRT) results of Model

905 A vs Model A null (Extended Data Table 1), gave a p-value of 0.0031 for *SKR*. The
906 tree was unrooted.
907
908 **Extended Data Table 1:**
909 Branch-site test results for positive selection in the SKR branch in the Brassicaceae.
910 Branch site models postulate three classes of sites: those under purifying selection
911 ($0< \omega<1$), neutral ($\omega=1$), and those under positive selection ($\omega>1$) in the foreground
912 branch being tested (Yang 2007)[53]. Model A is compared to Model A null in which the
913 third class is set to $\omega=1$. Positive sites are indicated together with posterior
914 probabilities. P-value for likelihood ratio test (LRT) results are calculated from the
915 χ^2 distribution taking as critical value $\chi^2= 2(LnL_{MA}-LnL_{MAnull})$ and dividing by 2 (Yang
916 2007)[53].
917

918 ## Methods

919 **Plant Material and Growth Conditions**
920 Arabidopsis mutant strains used in the study were *skr-1* (GT_5_101780), *skr-2*
921 (CSHL_GT5061) and *skr-3* (SALKSeq_034497.1), and *ski2-6* (SALK_122393)[91]. *skr-*
922 *1* and *skr-2* alleles are in Ler ecotype and have a Ds insertion in AT2G38690; *skr-3* is
923 in Col-0 ecotype. Plants were grown at 22 ^{o}C under 16 hours of light and 8 hours of
924 dark in a 1:1:1 proportion of peat, perlite, and vermiculite soaked with 0.5x Murashige
925 Skoog (MS) solution initially followed by irrigation during growth with water. For growth
926 on plates and selection of transformants, seeds were surface sterilized and plated on
927 MS plates supplemented with 2% sucrose and appropriate antibiotics.
928

929 **Coexpression screen for meiotic genes**

930 We selected 23 known meiotic genes (*DUET/MMD1, MS5/TDM1, DMC1, TAM, SDS,*
931 *AML5, RAD51, RAD50, ASY1, ZYP1, REC8, SPO11-1*, *SPO11-2*, *PHS1*, *MRE11*,
932 *MND1*, *OSD1*, *NBS1*, *MSH4*, *MSH5*, *HOP2*, *MER3*, *PRD1*), and each of these genes
933 was used as a query in the Expression Angler program[92] to identify co-expressed
934 genes. For each query, the top 100 hits were selected from each of three data sets,
935 NASCArrays, Bio-analytic resource, and AtGenExpress Plus-Extended tissue
936 compendium. We selected those genes common to three data sets and which
937 appeared at least twice in independent queries. Single copy genes were selected
938 using the Arabidopsis single copy genes list from the University of California, Davis
939 (http://www.cgpdb.ucdavis.edu/COS_Arabidopsis/). We manually selected 29 genes
940 that showed floral stage 9 peak/specific expression corresponding to meiosis stages
941 from AtGenExpress Visualization Tool[93]. We tested one to two insertion alleles for
942 each of the selected genes and for each of these alleles about 20 plants were
943 genotyped by PCR to identify homozygous mutant plants and these were phenotyped
944 by examining silique length and pollen viability by Alexander stain[94].
945
946 **Plasmid constructs and transgenics**
947 For complementation, a genomic fragment of 2220 bp (gSKR) comprising the *SKR*
948 gene including 3'UTR was amplified from Ler ecotype and cloned into a modified
949 binary vector pZP222 by Gateway cloning (Invitrogen). The binary construct was
950 introduced into Agrobacterium strain AGL1 and transformed into *skr-1/*+ and *skr-3/*+
951 plants by in planta transformation[95]. *skr* mutant T1 plants were genotyped by PCR and
952 tested for complementation. Plant vectors were ordered from
953 (https://gatewayvectors.vib.be/index.php/gateway-vectors). The pro*SKR*:*GFP-gSKR*
954 gene fusion construct was made by assembling the SKR promoter, GFP, and SKR

955 genomic coding region plus 3'UTR using three-way Gateway technology in the binary
956 destination vector pB7m34GW,0[96]. The pro*SKR:nlsGFP-GUS* construct comprised
957 1170 bp genomic DNA from 690 bp upstream of ATG upto 30 bp of the second exon
958 fused in frame with GFP-GUS in the pBGWFS7 vector. F-box gene nlsGFP fusions
959 were assembled in the pBnRGW_Redseed vector and introduced into plants. Single
960 locus insert transgenic skr/+ lines were made homozygous for the transgene and
961 examined in T4 for GFP fluorescence. The hairpin RNAi construct for PBE1[80] was
962 expressed under control of SKR regulatory elements comprising promoter and 3'UTR
963 and contained the PDK intron from pBm42GWIWG8_1. The construct and a noRNAi
964 control were introduced into *skr/*+ plants and transgenics that were *skr* mutant were
965 analysed for pollen viability. The *skr-1* mutant allele was used for all studies unless
966 indicated otherwise. proMSP1:SKRcds constructs were made via overlap extension
967 PCR, cloned into pENTR D TOPO vector using NotI and AscI restriction mediated
968 cloning, and further cloned into pZP222 via Gateway cloning.
969

970 **Cytology and imaging**

971 For Alexander staining, anthers from unopened buds were dissected out on a slide,
972 placed under a coverslip, and Alexander stain was added from the side to fill the entire
973 coverslip followed by sealing using nail-polish. The slides were incubated at room
974 temperature for 3-4 days and visualized under 10x /20 x objective.Whole anther and
975 ovule clearing was performed as in[97], anther staging of plastic sections was performed
976 as in[98] and ovule staging as in[99]. Immunostaining for GFP:SKR in meiosis was
977 performed as in[100]. Acid chromosomal spreads were prepared as described in[101]. GFP
978 / DAPI fluorescence was visualized after mounting of meiotic anthers under the
979 coverslip with 30-50 % glycerol / Antifade. ImageJ and Photoshop were used to edit

980 and assemble images. Electron microscopy sample preparation and imaging was
981 done as in[102] and stripe artifacts in the EM images were removed using the method
982 mentioned in[103].
983
984 **Suppressor screening and analysis**
985 *skr-1* mutant seeds (M0) were bulked from multiple plants and 0.3 % EMS solution
986 was used to mutagenise the M0 seeds for 5 hours followed by extensive washing with
987 water. Seeds from 934 M1 plants were collected individually and M2 families were
988 screened for fertility restoration. One suppressor line 08-10 was identified and
989 backcrossed to *skr* mutants to get BC1 F1 seeds. Seeds from BC1 F1 plants were
990 collected and bulk segregant analysis on BC1 F2 plants was performed on fertile and
991 sterile pools. Genomic DNA was isolated from the fertile and sterile pools (50 plants
992 each) and used for whole genome sequencing.
993
994 To map the suppressor, the parental un-mutagenised *skr* genome was sequenced and
995 *skr* specific reference genome was made by replacing the SNPs of *skr* to Ler-0
996 genome[104] sequence using the *bedtools – consensus*[105]. Variants were called for the
997 fertile and sterile bulk reads from the *skr* reference genome using *freebayes*[106] and
998 indels were removed. SNPindex was calculated using the R package QTLSeqr[107] for
999 the fertile bulk and sterile bulk. SNPs with a deltaSNP of fertile to sterile score greater
1000 than 0.5 were analysed for genic perturbations. The 08-10 line had an in frame stop
1001 codon in the SKI2 gene, at amino acid position 518. The suppressor phenotype was
1002 confirmed by generating the double mutant using the independent alleles *ski2-6* and
1003 *skr-3.* The *ski2* allele in line 08-10 was named *ski2-7.*
1004

1005 **Expression profiling**

1006 RNA from meiotic stage anthers was isolated using Qiagen kits for *skr-1,* Ler and *skr-*
1007 *1 ski2-7*. RNA sequencing was performed on Illumina platforms following the
1008 manufacturer's instructions. Differentially expressed genes were found using DESeq2
1009 keeping a FDR adjusted *P* value < 0.05[108]. Validation for eight upregulated F-box
1010 genes was performed by RT-qPCR (Supplementary Figure 3) which confirmed the
1011 RNAseq results. The tissue-specific expression analysis of deregulated genes was
1012 performed by downloading TPM values for each gene from CoNekT
1013 (https://evorepro.sbs.ntu.edu.sg) database[65]. Meiocyte RNAseq TPM value[109] was
1014 combined with the CoNekT dataset. The tissue enrichment score was calculated by
1015 row normalization using the largest TPM score. Enrichment score of 1, in a tissue/cell
1016 type was taken to represent maximal expression in that type. The number of non-
1017 overlapping genes with enrichment score 1, was calculated and plotted for meiocytes,
1018 UNM, BCP, TCP/MP and others.

1019

1020 **Proteasome inhibitor treatment**

1021 MG132 and MG115 stocks at 20 mM concentration were made in DMSO. Inhibitor
1022 solutions were used at 100 µM concentration[110] by dissolving in half strength MS
1023 where 0.5 % DMSO was present. Control solution was 0.5 % DMSO in half strength
1024 MS. Sections of freshly cut *skr* inflorescence stems were placed in 200 µl inhibitor or
1025 control solution in Eppendorf tubes for 3 hrs, followed by 60 hrs of recovery in half
1026 strength MS solution. Anthers were dissected and viable pollens were identified by
1027 Alexander staining. At least 100 anthers were analysed per sample (treatment and
1028 control) per experiment. Number of anthers with 10 or more viable pollen were counted
1029 per 100 anthers for treatment and control and the ratio was calculated. The experiment

1030 was conducted blind and repeated thrice. The ratio paired t test (two-tailed) on the
1031 logarithm of ratios was used to calculate p values (GraphPad).
1032

1033 **Sequence analysis**

1034 Sequence searches to detect homologs of the SHUKR domain were conducted using
1035 the PSI-BLAST (RRID:SCR_001010)[111] and JACKHMMER (RRID:SCR_005305)[112]
1036 programs with profile-inclusion threshold of expect (e)-value at 0.01 against the non-
1037 redundant database of National Center for Biotechnology Information (NCBI) clustered
1038 down to 50%. Clustering of proteins based on bit score density and length of aligned
1039 sequence was performed using the MMseqs program. Remote homology searches
1040 were performed using profile-profile comparisons with HHpred program
1041 (RRID:SCR_010276)[113] against profile libraries from the PFAM (RRID:
1042 SCR_004726)[114] and PDB (RRID:SCR_012820)[115] databases. Multiple sequence
1043 alignments were built using Mafft program with maxiterate parameter set to 3000
1044 followed by manual adjustments based on secondary structure prediction. Secondary
1045 structures were predicted using the JPred (RRID:SCR_016504)[116] and
1046 RoseTTaFold[117] programs. The globular domains in the protein were objectively
1047 identified using Shannon entropy-based compositional analysis with the SEG program
1048 with a window size of 45, low-cut of 3.4, high-cut of 3.75, and minimal length of the
1049 high-complexity segment of 15 residues.
1050

1051 The analysis of the Shannon entropy (H_j) for a given multiple sequence alignment was
1052 performed using the equation:

1053 $$H_j = -\sum_{i=1}^{M} P_i(log_2(P_i))$$

1054 P is the fraction of residues of amino acid type i and M is the number of amino acid
1055 types. The Shannon entropy for the jth position in the alignment can range from 0 (only
1056 one residue at that position) to 4.32 (all 20 residues equally represented at that
1057 position). The reduced alphabet was created by the following grouping: h= L, I, M, V,
1058 A; a= F, Y, W; b=K, R, H; n=E, D; p=Q, N; o=T, C, S; t=P, G, d=- (gap).
1059

1060 **Phylogenetic analysis**

1061 SKR and SKL coding sequences of different species were extracted from Genbank,
1062 Phytozome, and CoGeBLAST[118,119] (Supplementary File 1). Genomic fragments were
1063 annotated to CDS by the online tool MEGANTE[120]. For the branch site analysis,
1064 MAFFT was used to build the alignment and RAxML was used to build a phylogenetic
1065 tree with 1000 bootstraps[121,122]. Branch site model tests of the CODEML package[82]
1066 were performed as implemented in EasyCoDEML[123] by selecting the branch of choice.
1067 The tree was unrooted. Protein phylogenetic analysis was performed using the
1068 maximum likelihood method implemented in the IQtree program (RRID:
1069 SCR_017254)[124] under multiple parameter regimes using: 1) the Q.plant substitution
1070 matrix derived from alignments specific to plant nuclear genes and 1 invariant site
1071 category with 8 gamma distributed sites; 2) the LG substitution matrix with 1 invariant
1072 site category with 8 gamma distributed sites or 3) with a 20-profile mixture model.
1073 Bootstrap values were calculated using the Shimodaira-Hasegawa-like approximate
1074 likelihood ratio (SH-aLRT) [125].
1075

1076 F-box gene sequences of 710 Arabidopsis genes used for ortholog analysis were
1077 obtained from Ref. 84 of which 64/69 were common to *skr* upregulated. Classification

1078 of LTS and STS genes was according to Ref. 84. Identification of orthologs was done
1079 using Proteinortho[126].
1080

1081 **Acknowledgments**

1082 We thank David Twell for kindly providing the LAT52:GFP reporter line, ABRC and
1083 CSHL for seeds. Kathakali Bannerjee and Manoj Kamble for technical support, Mekala
1084 Sai Kiran for maintenance of plants, the NGS and Advanced Microscopy facilities at
1085 CCMB, and Priyanka Pant for help in illustrations. We are grateful to Elliot Meyerowitz,
1086 Utpal Nath, and Manjula Reddy for valuable comments on the manuscript.
1087

1088 **Funding**

1089 This work was supported by a CSIR Network Project Grant MLP0120 and a
1090 Department of Biotechnology Centre of Excellence Grant
1091 BT/PR21305/COE/34/44/2016 to IS. SP and PS were supported by CSIR Fellowships.
1092 AR was supported by a Department of Biotechnology Research Associate Fellowship.
1093 IS acknowledges a JC Bose fellowship from the Department of Science and
1094 Technology and a Senior Scientist fellowship from the Indian National Science
1095 Academy.
1096

1097 **Author contributions**

1098 PS, SP, JND, JD and IS conceived and planned experiments. SP, PS, AR, JND, VS,
1099 HA, AS, and IS performed experiments. SP, PS, AR, AS, CK, HG, JD, and IS provided
1100 analysis of the experimental results. LA contributed the evolutionary analysis of Figure
1101 6. IS and PS wrote the manuscript with contributions from LA and JD.
1102

1103 **Competing interests**

1104 Authors declare no competing interests.

1105

1106 **Additional information**

1107 Supplementary information is available for this paper.

1108

1109 **Materials and correspondence**

1110 Correspondence and requests for materials to be addressed to IS. Sequencing data

1111 will be deposited at the European Nucleotide Archive.

1112

1113 Supplementary information

1114 **Supplementary Table 1:** Differentially expressed gene lists.

1115 **Supplementary Table 2:** Male gametogenesis enrichment analysis of *skr* upregulated

1116 genes.

1117 **Supplementary Table 3:** Analysis of genes differentially expressed in *skr.*

1118 **Supplementary Table 4:** Differentially expressed translation-related genes in *skr*

1119 *ski2*.

1120 **Supplementary Table 5:** Enrichment for short taxonomic scale (STS) F-box genes

1121 upregulated in *skr*.

1122 **Supplementary Table 6:** List of primers

1123 **Supplementary Figure 1:** Multiple sequence alignment of the SHUKR domain.

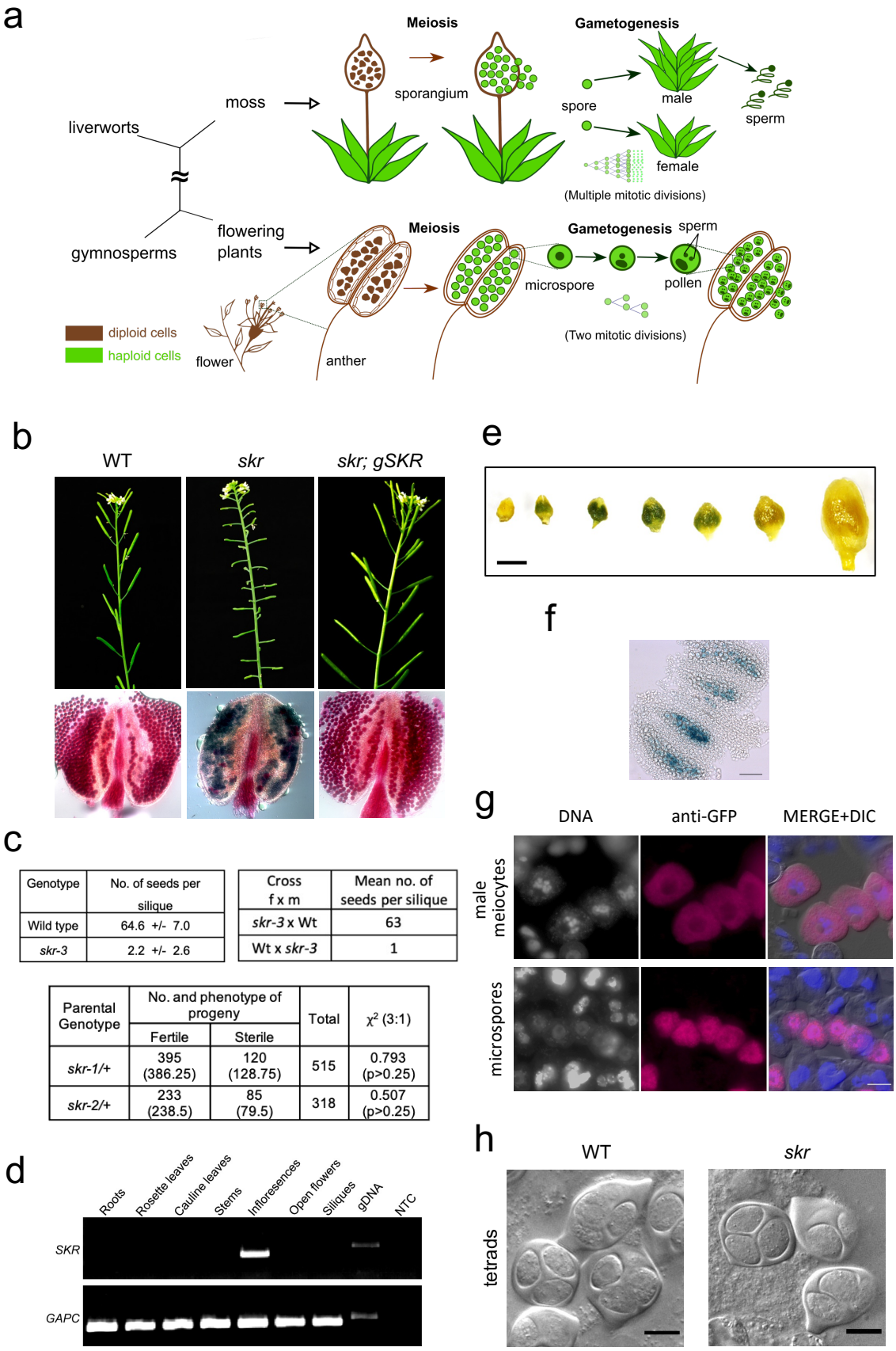
1124 **Supplementary Figure 2:** Multiple sequence alignment of SKR protein sequences

1125 used for branch-site analysis.

1126 **Supplementary Figure 3:** Validation for eight *skr* upregulated F-box genes by RT-

1127 qPCR.

1128 **Supplementary File 1:** Sequences of *SHUKR* family genes used in branch-site
1129 analysis.



Genotype	No. of seeds per silique
Wild type	64.6 +/- 7.0
skr-3	2.2 +/- 2.6

Cross f x m	Mean no. of seeds per silique
skr-3 x Wt	63
Wt x *skr-3*	1

Parental Genotype	No. and phenotype of progeny		Total	χ^2 (3:1)
	Fertile	Sterile		
skr-1/+	395 (386.25)	120 (128.75)	515	0.793 (p>0.25)
skr-2/+	233 (238.5)	85 (79.5)	318	0.507 (p>0.25)

Fig. 1: *SHUKR* is required for pollen development.

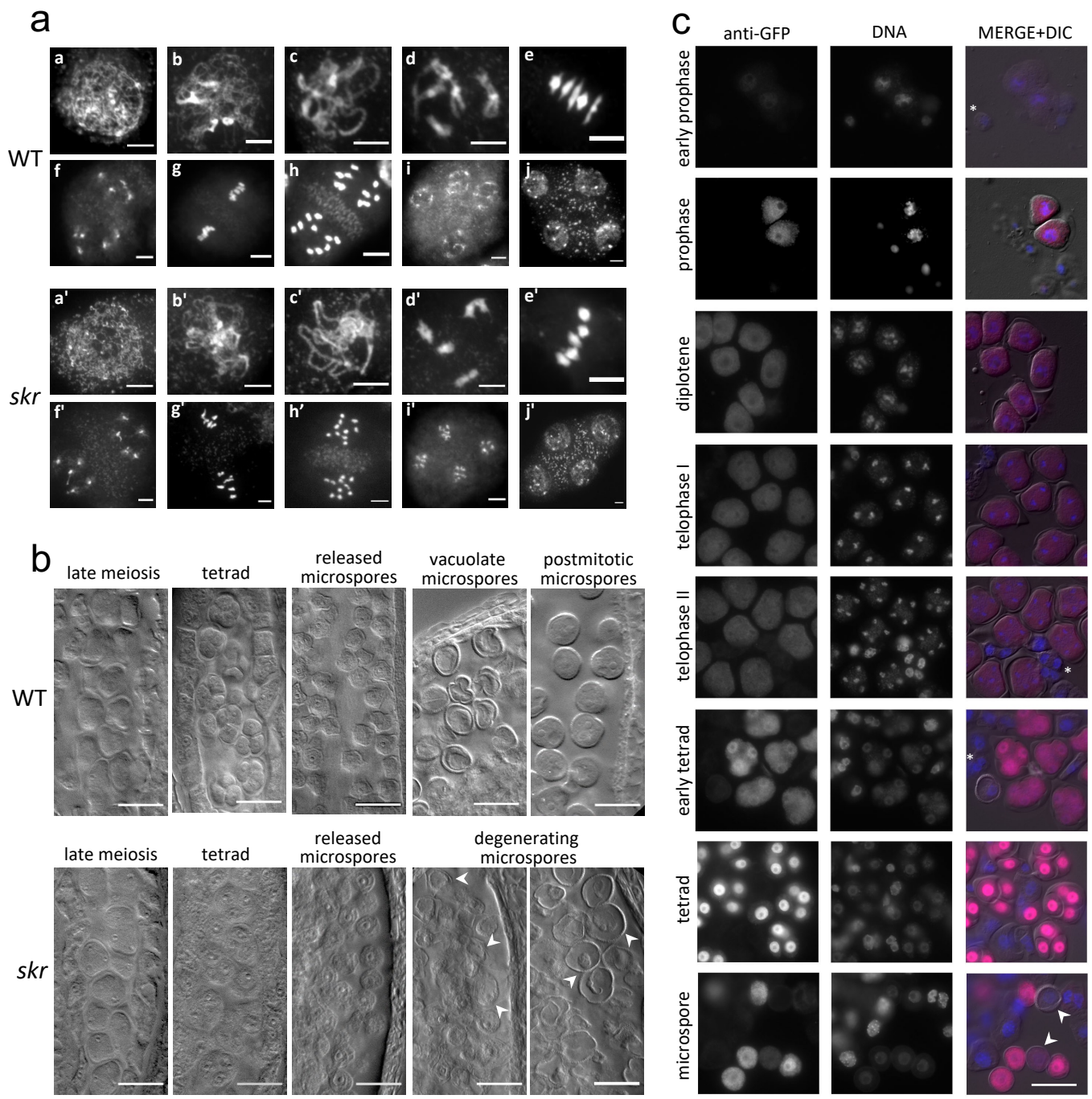


Fig. 2: Staging of *skr* phenotype and SKR protein expression

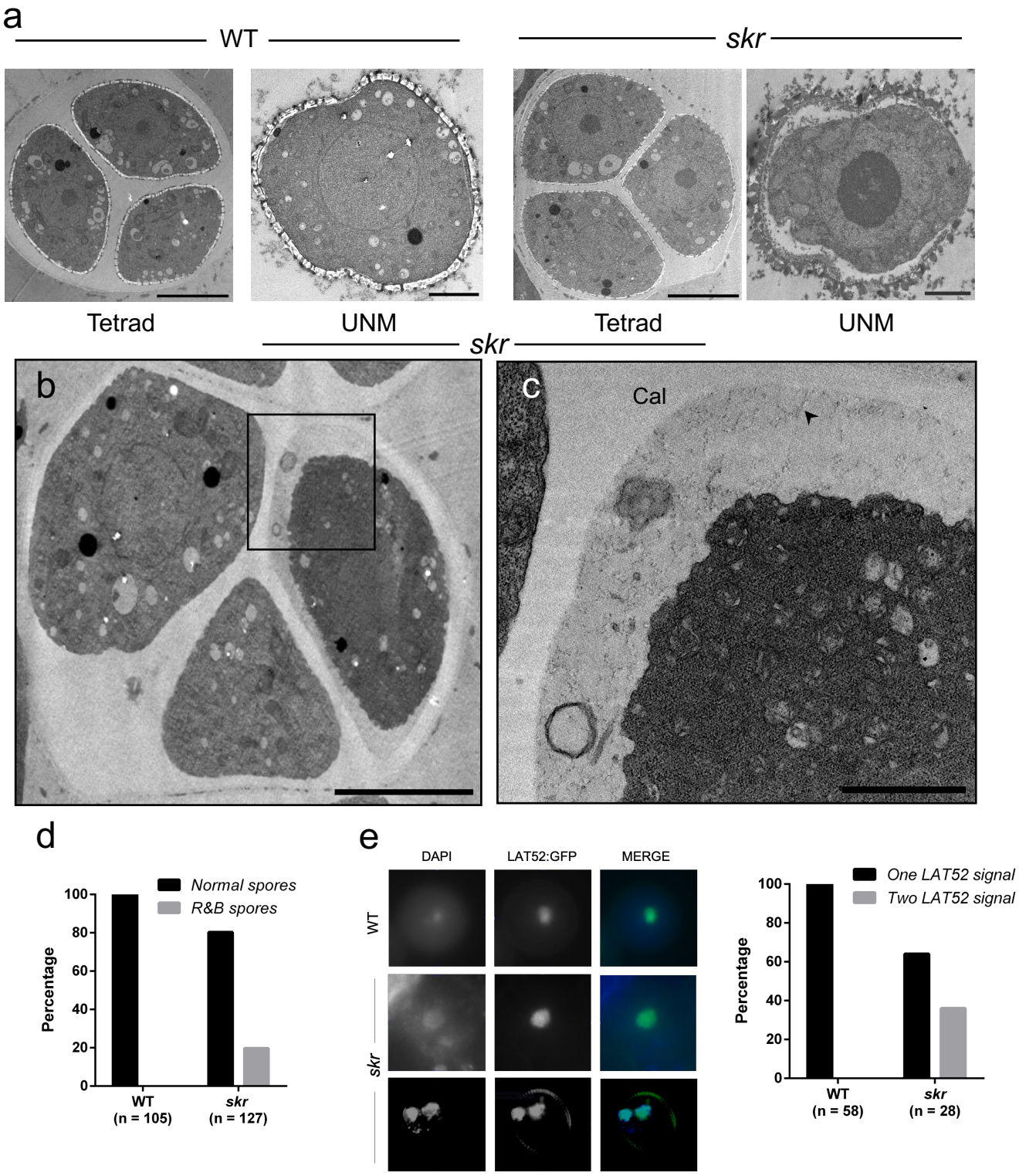
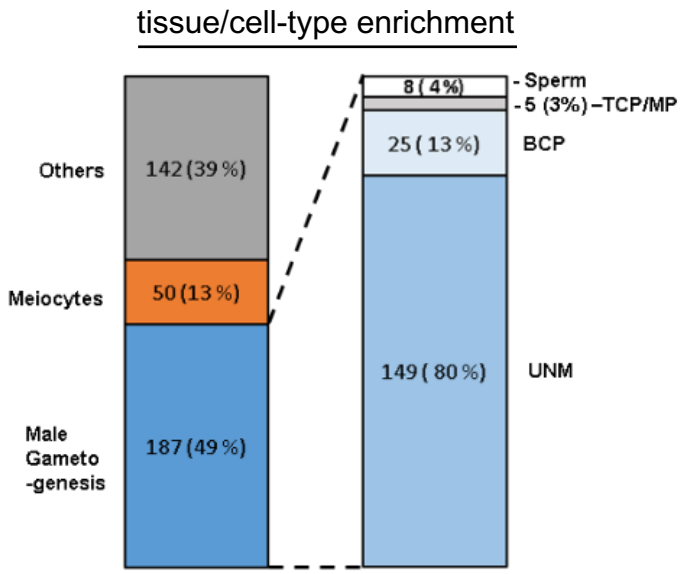


Fig. 3: Ultrastructure and marker expression reveal early gametogenesis defects in *skr*

a

Parents F x M	Total seeds	KanR	KanS	KanR:KanS	χ2
Wt x *skr-1/+*	187	87 (93.5)	100 (93.5)	1:1.15 (1:1)	0.39 {p>0.5}
skr-1/+ x Wt	222	116 (111)	106 (111)	1:0.91 (1:1)	0.23 {p>0.5}
skr-1/+ (selfed)	969	741 (727)	228 (242)	3:0.92 (3:1)	1.09 {p>0.3}

b



c

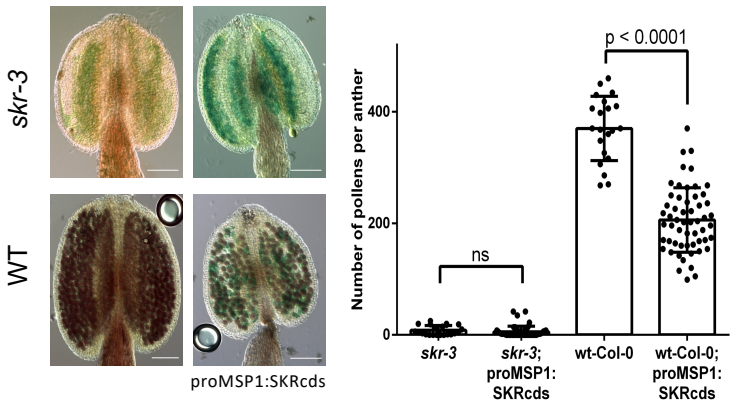


Fig. 4: *SHUKR* is a gametogenesis inhibitor

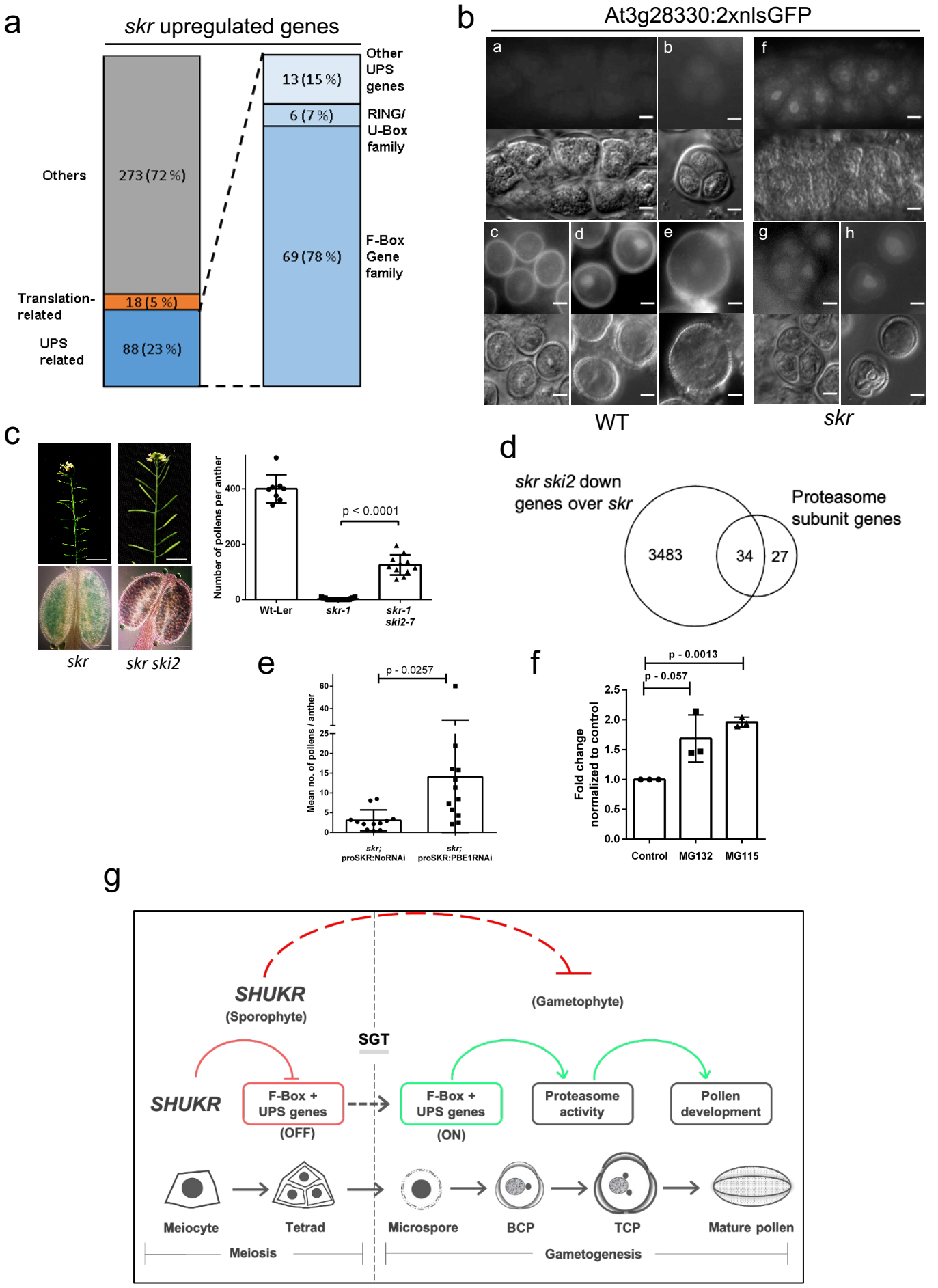


Fig. 5: *SHUKR* regulates the proteasome in pollen development

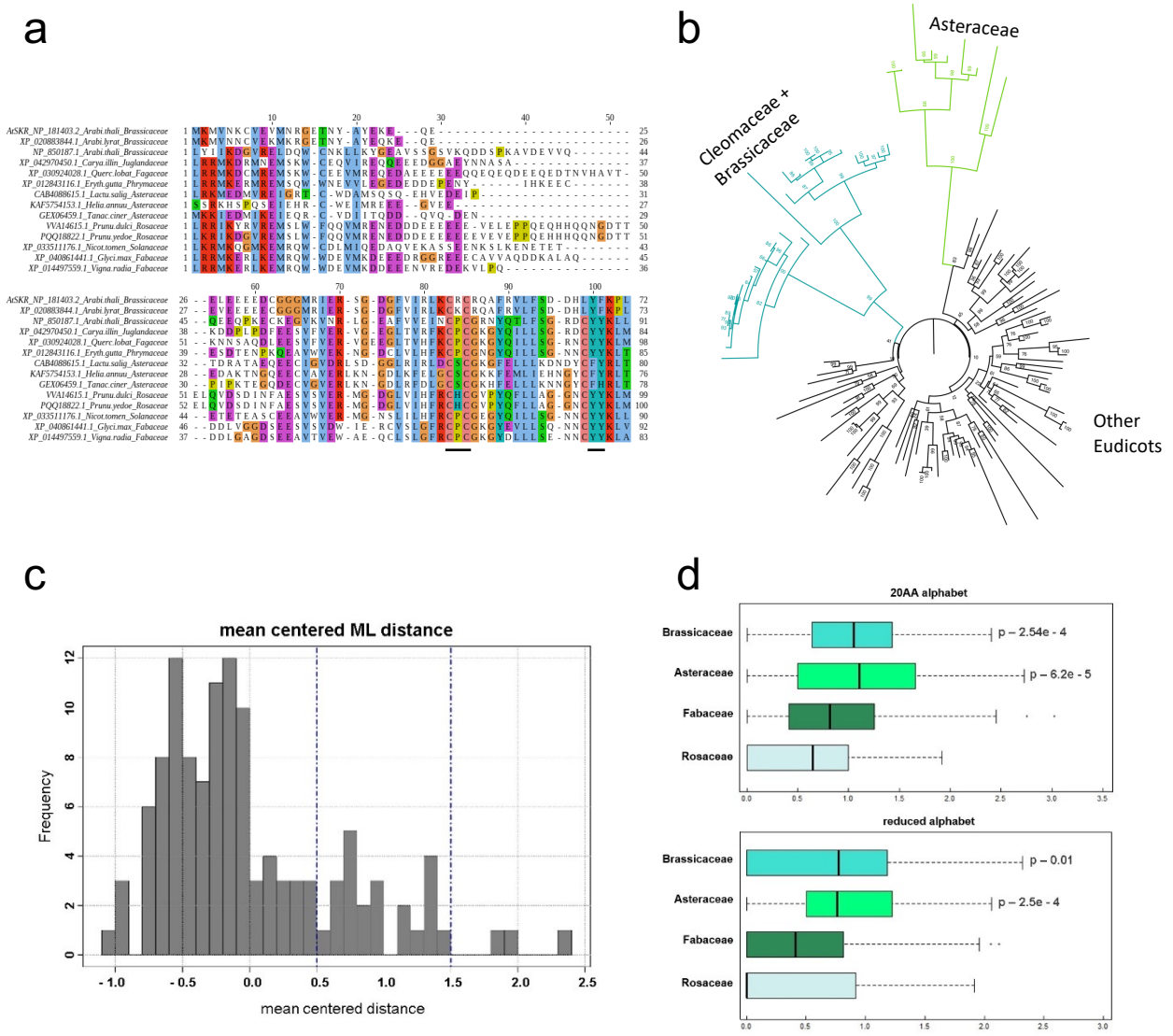


Fig. 6: Sequence and evolutionary analysis of the SHUKR domain

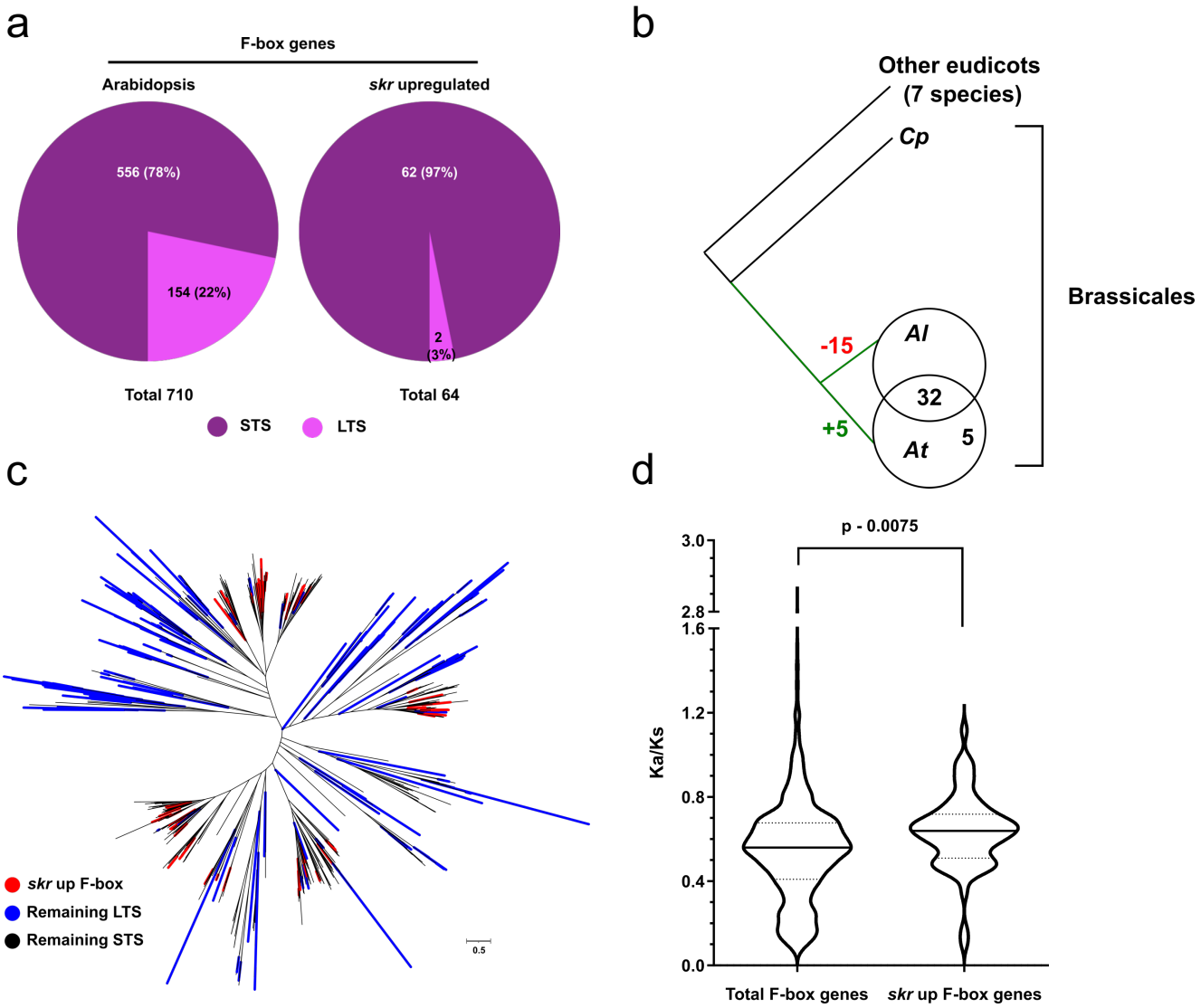
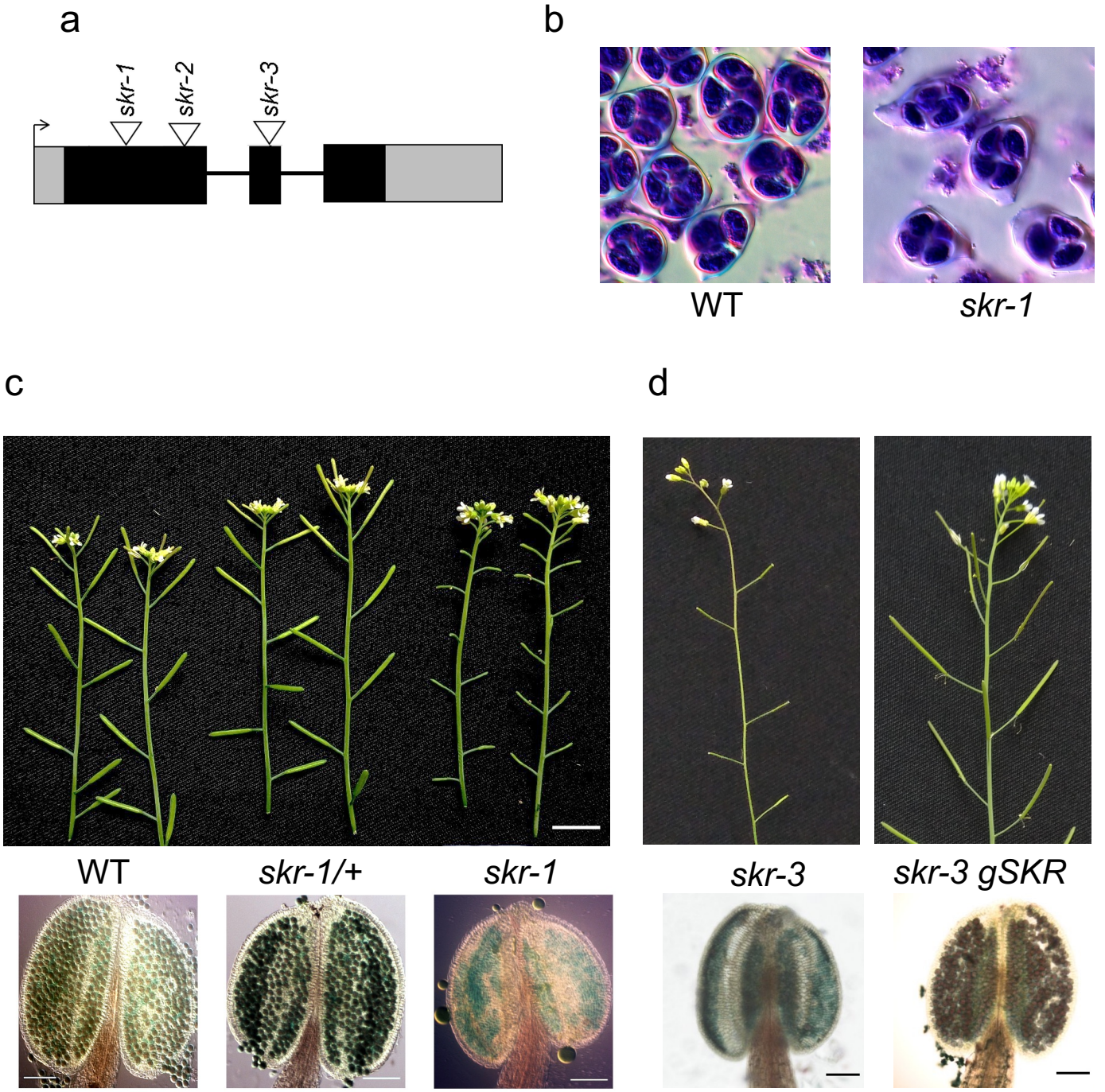
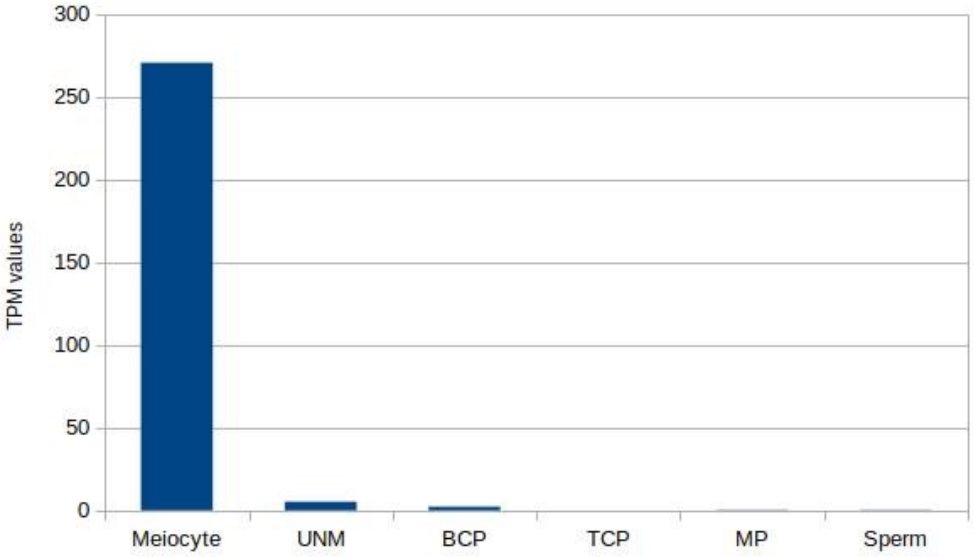


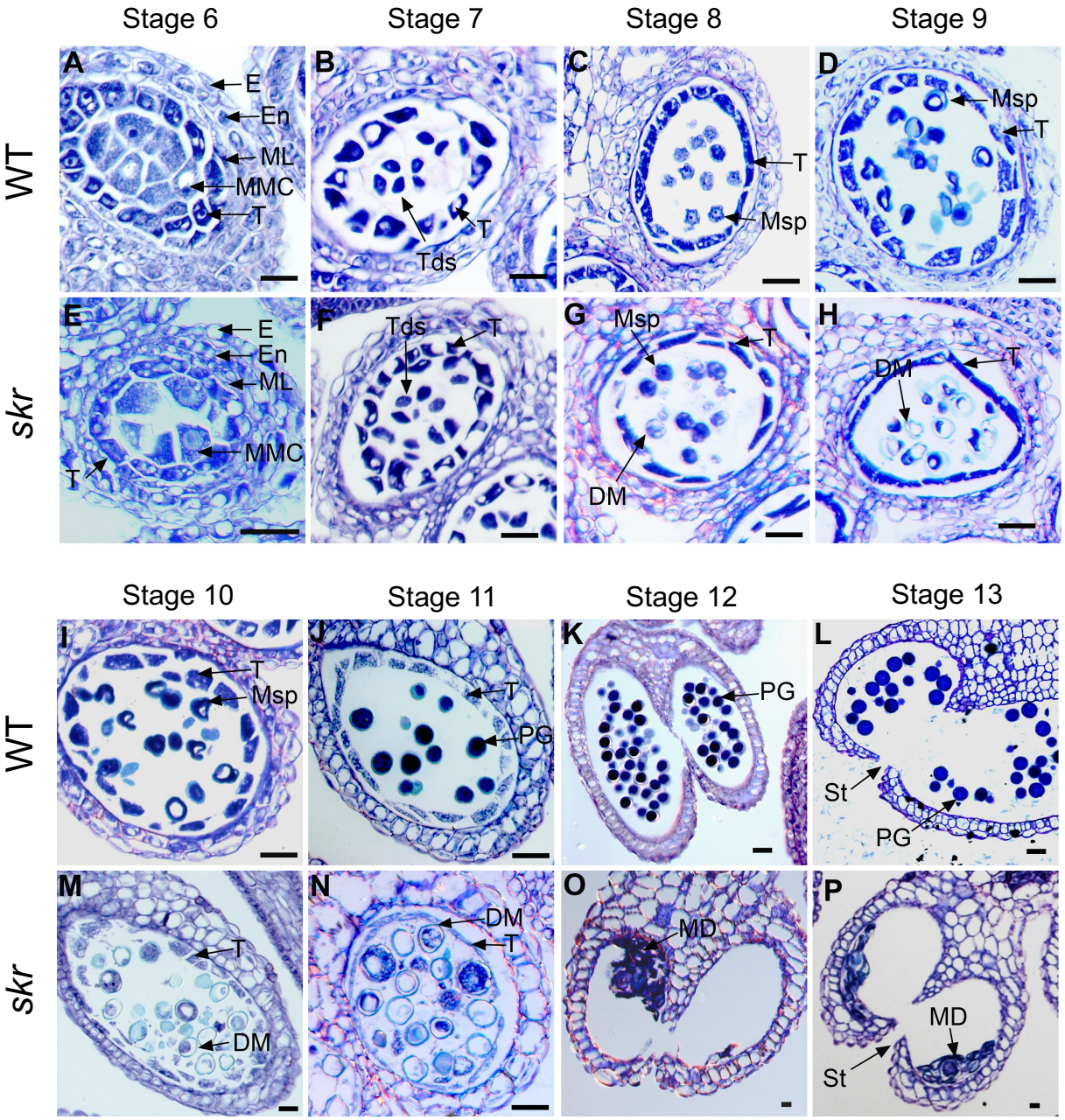
Fig. 7: Diversifying selection in SHUKR regulated F-box genes



Extended Data Fig. 1:
(a) Gene diagram of *SKR* showing location of insertions: *skr-1* (GT_5_101780), *skr-2* (CSHL_GT5061), *skr-3* (SALKSeq_034497.1). (b) Toluidine blue stained tetrads. (c) Inflorescence stem and Alexander staining showing recessive male sterility of *skr.* (d) Complementation of *skr-3* by a *gSKR* transgene. Scale bar: 1 cm for inflorescence, 100 μm for anthers.

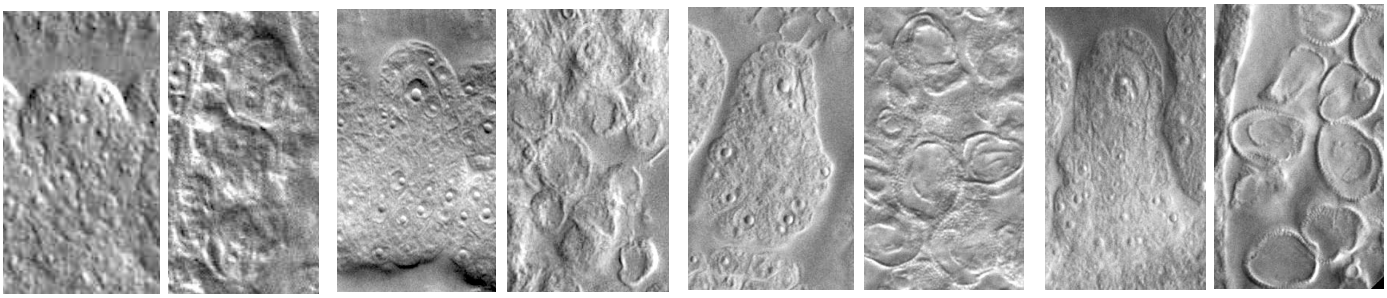


Extended Data Fig. 2:
Profiling of SKR expression in male germline development.
TPM values for SKR were collected from the CoNekt database, except for meiocytes. Meiocyte TPM value were taken from (Walker 2018)[21]. UNM – Uninucleate microspore; BCP – Bicellular pollen; TCP – Tricellular pollen; MP – Mature pollen.



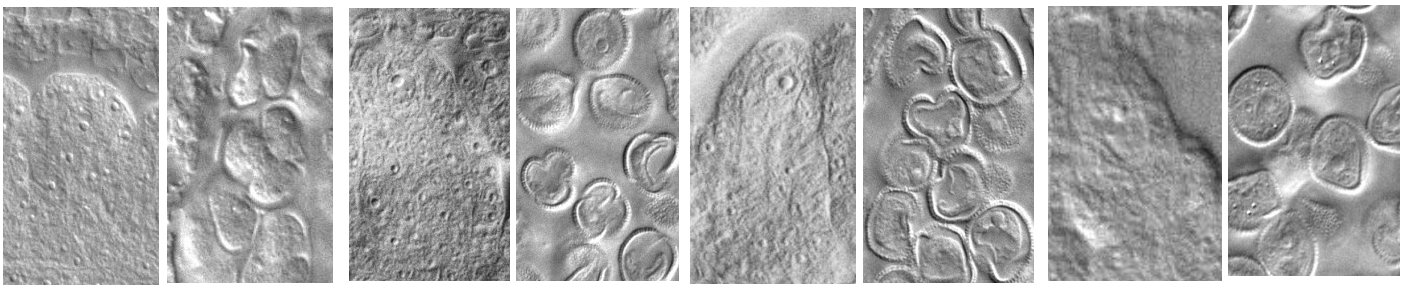
Extended Data Fig. 3:
Sections of anther development stages.
Corresponding stages are shown for wild type (A-D, I-L) and *skr* (E-H, M-P). Stages are numbered according to (Sanders 1999)[9]. Abbreviations: E epidermis, En endothecium, ML middle layer, T tapetum, Td tetrad, Msp microspore, DM degenerating microspore, PG pollen grain, St stomium, MD microspore debris. Scale bar 20 µm.

shukr



Ovule stg 1-2 | tetrad | Ovule stg 2-1 | degenerating microspore | Ovule stg 2-2 | degenerating microspore | Ovule stg 2-3 | degenerating microspore

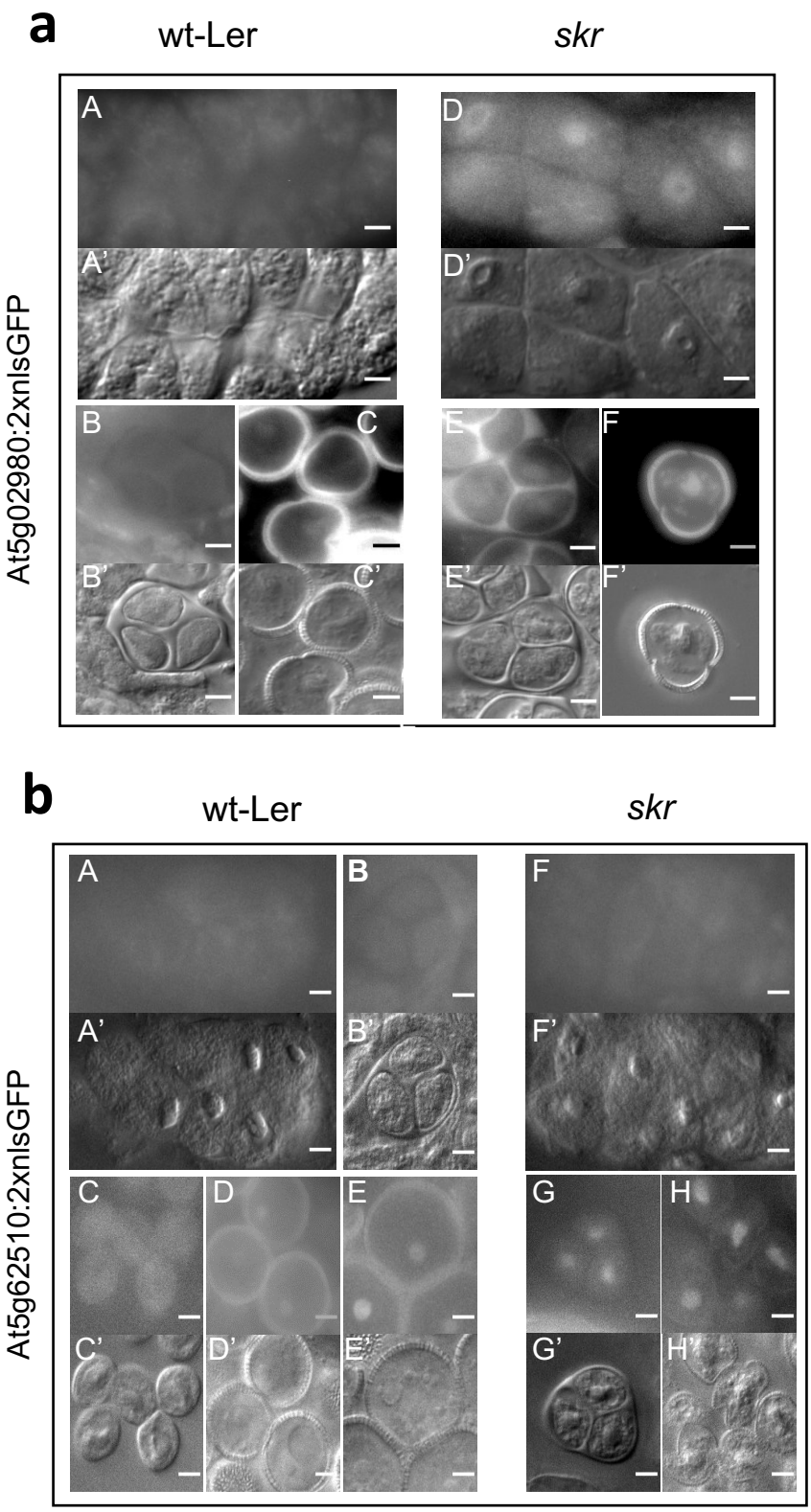
Wild type



Ovule stg 1-2 | tetrad | Ovule stg 2-1 | vacuolated microspore | Ovule stg 2-2 | late vacuolated microspore | Ovule stg 2-3 | post-UNM

Ovule stage	Pollen development stage	
	Wild type	***shukr***
1-2	Tetrad	Tetrad
2-1	Vacuolated microspores	Degenerating microspores
2-2	Late vacuolated microspores	Degenerating microspores
2-3	Post-mitotic	Degenerating microspores

Extended Data Fig. 4:
Staging of microspore degeneration in *skr*.
Microspore stage and phenotypes are shown together with stage of ovules from the same bud (Schneitz 1995)[10] as a reference for comparison to wild type.

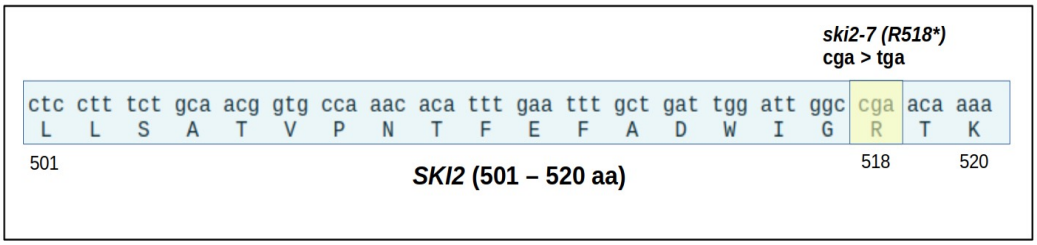


Extended Data Fig. 5:
Precocious expression of F-box gene:nlsGFP fusions in meiosis in *skr*.
(a) At5g02980. A-C, A'-C': wild type; D-F, D'-F': skr; A,A',D,D' meiocytes; B,B',E,E': tetrad; C,C',F,F': uninucleate microspores. (b) At5g62510. A-E, A'-E': wild type; F-H, F'-H': skr; A,A',F,F' meiocytes; B,B',G,G': tetrad; C-E, C'-E',H,H': uninucleate microspores. A-H: GFP fluorescence; A'-H': corresponding DIC. GFP expression in *skr* is first detected in meiocytes (At5g02980) or tetrad stage At5g62510) whereas in wild type earliest GFP signal was in released microspores. Scale bar: 5 µm

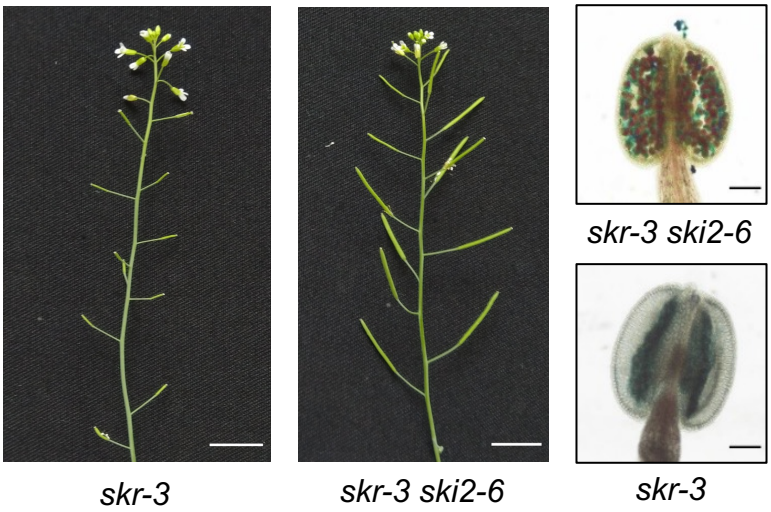
a

Cross	Tested Generation	Plants tested	Sterile	Fertile	Sterile : Fertile
skr X 08-10 suppressor line	Backcrossed F1	20	20	0	1:0
skr X 08-10 suppressor line	Backcrossed F2	180	129	51	2.53:1

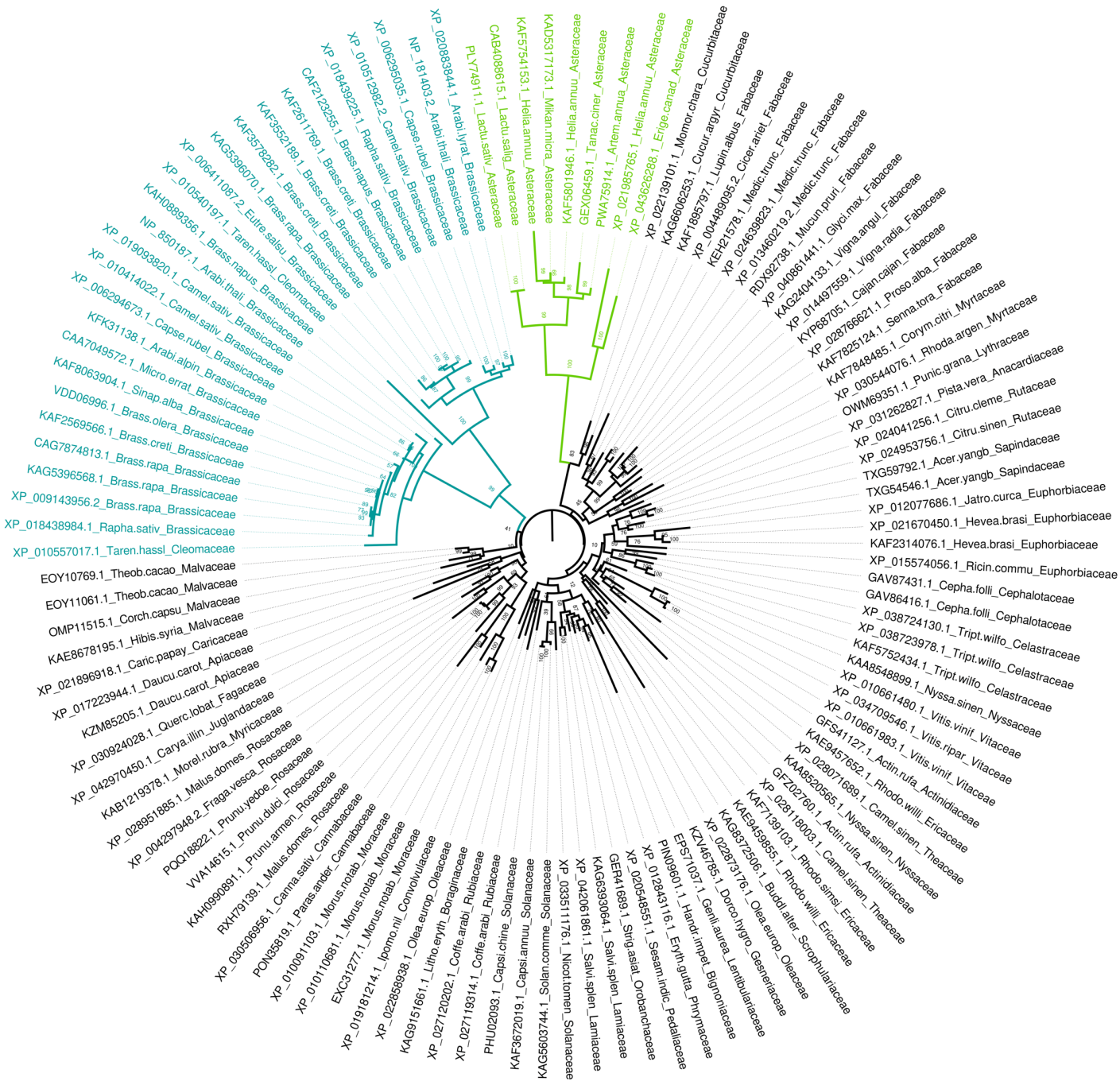
b



c



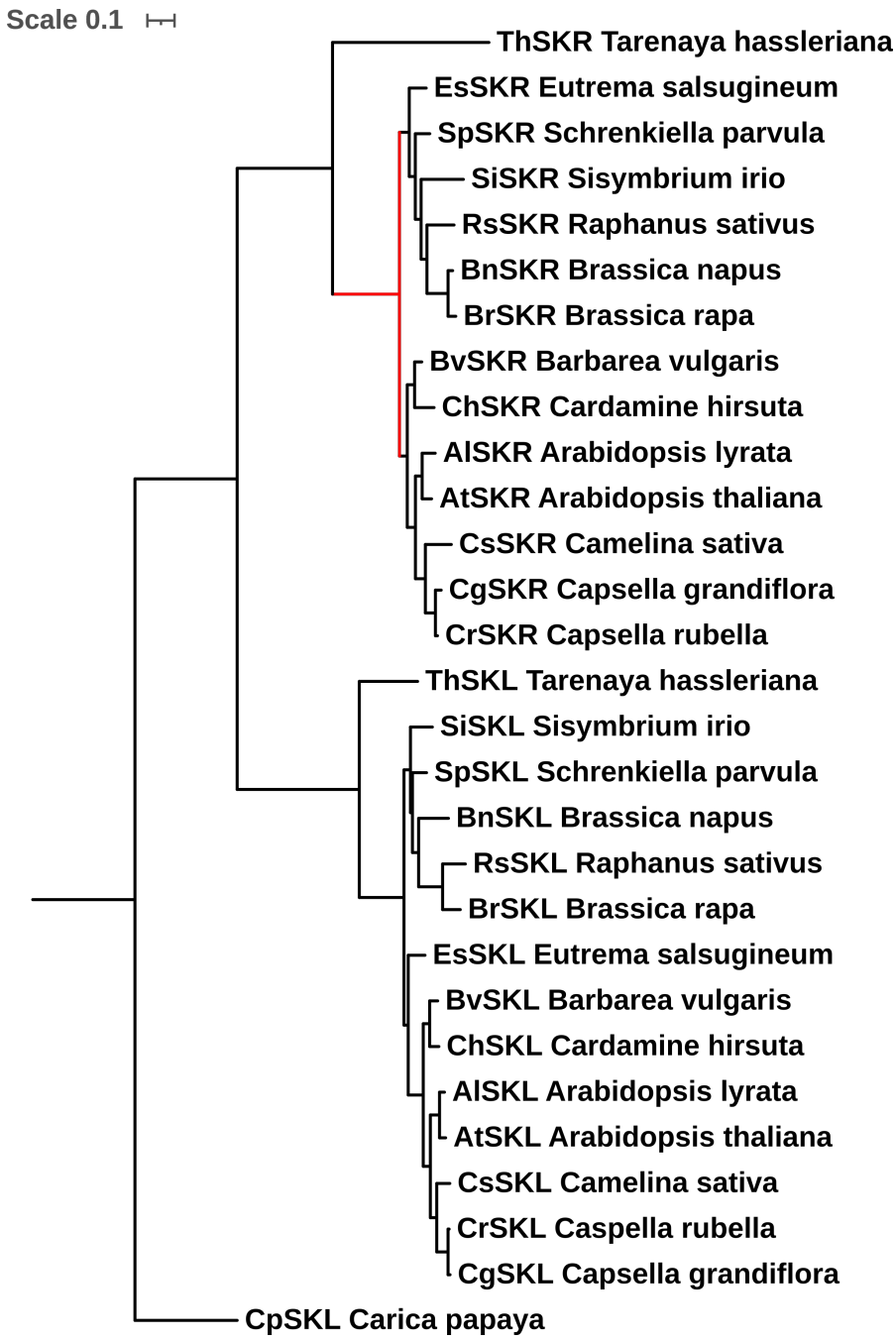
Extended Data Fig. 6:
ski2 is a suppressor of *skr*.
(a) Mendelian segregation of the suppressor mutation in the 08-10 *skr* suppressor line (b) A C>T transition in the *ski2* allele (named *ski2-7*) present in the 08-10 *skr* suppressor line results in a TGA termination codon at position 518. (c) Restoration of fertility and pollen viability in *skr-3* by the independent allele *ski2-6*. Scale bar: inflorescence 1 cm; anthers 100 μm.



Extended Data Fig. 7:

Phylogenetic tree of the SHUKR family from eudicots.

119 representatives from 85 species were used for tree construction using only the SHUKR domain that is shared by all representatives of the family. Genbank ID and species are indicated. Brassicaceae+Cleomaceae (dark green) and Asteraceae (light green)



Extended Data Fig. 8:
Phylogenetic tree of Brassicales *SHUKR* and *SHUKR-like* (*SKL*) genes used for branch-site test.
The *SHUKR* gene family is represented by a single gene in *Carica papaya*, whereas Brassicaceae and Cleomaceae experienced a gene duplication event prior to their divergence. CDS were aligned using MAFFT, and RAxML was used to build the phylogenetic tree with 1000 bootstraps. The red branch was selected as foreground for *SKR*. Branch-site test for positive selection was performed using the CODEML package implemented in EASYCoDEML. Likelihood Ratio Test (LRT) results of Model A vs Model A null (Extended Data Table 1), gave a p-value of 0.0031 for *SKR*. The tree was unrooted.

Model	LnL	Estimates of parameters					Model compared	LRT P-value	Positive sites
Model A	-4394.302739	Site class	0	1	2a	2b	Model A vs.Model A null	0.003120857	7 E 0.932,22 F 0.672,25 V 0.529,31 E 0.613,43 F 0.553,57 E 0.962*,63 L 0.974*,82 K 0.594,100 P 0.747,112 R 0.668,114 C 0.643
		f	0.66753	0.21979	0.08477	0.02791			
		ω0	0.20115	1.00000	0.20115	1.00000			
		ω1	0.20115	1.00000	22.04175	22.04175			
Model A null	-4398.670454	1							Not Allowed

Extended Data Table 1: Branch-site test results for positive selection in the SKR branch in the Brassicaceae. Branch site models postulate three classes of sites: those under purifying selection (0< ω<1), neutral (ω=1), and those under positive selection (ω>1) in the foreground branch being tested (Yang, 2007)[82]. Model A is compared to Model A null in which the third class is set to ω=1. Positive sites are indicated together with posterior probabilities. P-value for likelihood ratio test (LRT) results are calculated from the χ^2 distribution taking as critical value $\chi^2 = 2(LnL_{MA} - LnL_{MAnull})$ and dividing by 2 (Yang, 2007)[82].