Conserved Molecular Function and Regulatory Subfunctionalization of the LORELEI Gene Family in Brassicaceae

Jennifer A. Noble[1], Ming-Che James Liu[2], Thomas A. DeFalco[3], Martin Stegmann[4,8], Kara McNamara[2], Brooke Sullivan[2], Khanhlinh K. Dinh[2], Nicholas Khuu[2], Sarah Hancock[1], Shin-Han Shiu[5], Cyril Zipfel[3,4], Mark A. Beilstein[1]*, Alice Y. Cheung[2,6,7], and Ravishankar Palanivelu[1]*

[1]School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

[2]Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, MA 01003, USA

[3]Institute of Plant and Microbial Biology, Zurich-Basel Plant Science Center, University of Zurich, Zurich, Switzerland

[4]The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

[5]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

[6]Molecular and Cell Biology Program, University of Massachusetts, Amherst, MA 01003, USA

[7]Plant Biology Graduate Program, University of Massachusetts, Amherst, MA 01003, USA

[8]Department of Phytopathology, Technical University of Munich, School of Life Sciences Weihenstephan, Freising, Germany

*Corresponding author: Email: mbeilstein@email.arizona.edu, rpalaniv@email.arizona.edu

Abstract

A signaling complex comprising members of the LORELEI (LRE)-LIKE GPI-anchored protein (LLG) and *Catharanthus roseus* RECEPTOR-LIKE KINASE 1-LIKE (*Cr*RLK1L) families perceive RAPID ALKALINIZATION FACTOR (RALF) peptides and regulate growth, development, reproduction, and immunity in *Arabidopsis thaliana*. Duplications in each component, which potentially could generate thousands of combinations of this signaling complex, are also evident in other angiosperms. Widespread duplication in angiosperms raises the question what evolutionary mechanisms underlie the expansion and retention of these gene families, as duplicated genes are typically rendered non-functional. As genetic and genomic resources make it a tractable model system, here we investigated this question using *LLG* gene family evolution and function in Brassicaceae. We first established that the *LLG* homologs in the Brassicaceae resulted from duplication events that pre-date the divergence of species in this family. Complementation of vegetative phenotypes in *llg1* by LRE, LLG2, and LLG3 showed that the molecular functions of LLG homologs in *A. thaliana* are conserved. We next tested the possibility that differences in gene expression (regulatory subfunctionalization), rather than functional divergence, played a role in retention of these duplicated genes. For this, we examined the function and expression of *LRE* and *LLG1* in *A. thaliana* and their single copy ortholog in *Cleome violacea* (*Clevi LRE/LLG1*), a representative species outside the Brassicaceae, but from the same order (Brassicales). We showed that expression of *LLG1* and *LRE* did not overlap in *A. thaliana* and that *Clevi-LRE/LLG1* expression in *C. violacea* encompassed all the expression domains of *A. thaliana LRE* + *LLG1*. Still, complementation experiments showed that *LLG1* rescued reproductive phenotypes in *lre* and that *Clevi LRE/LLG1* rescued both vegetative and reproductive phenotypes in *llg1* and *lre.* Additionally, we found that expression of *LLG2* and *LLG3* in *A. thaliana* have also diverged from the expression of their corresponding single copy ortholog (*Clevi LLG2/*LLG3) in *C. violacea*. Our findings demonstrated how regulatory subfunctionalization, rather than functional divergence, underlies the retention of the *LLG* gene family in Brassicaceae. Our findings on the regulatory divergence and

55 functional conservation provide an experimental framework to characterize the
56 combinatorial assembly and function of this critical plant cell signaling complex.

57

58 Introduction

59

60 How cells communicate with each other to ensure coordinated growth and development
61 remains a fundamental question in eukaryotes. The signaling complex comprising
62 members of FERONIA (FER), LORELEI (LRE), and RAPID ALKALINIZATION FACTOR
63 (RALF) families is rapidly emerging as one of the best characterized cell–cell signaling
64 models in Arabidopsis [1]. FER is a receptor kinase (RK) of the *Catharanthus roseus*
65 RECEPTOR-LIKE KINASE 1-LIKE (*Cr*RLK1L) family, LRE is a membrane-associated
66 glycosylphosphatidylinositol (GPI)-anchored protein, and RALF is a small cysteine-rich
67 peptide (CRP). In *Arabidopsis thaliana* reproduction, LRE localizes in the synergid cells
68 of the ovule and interacts with FER and the FER–LRE complex functions at the interface
69 of the synergid cell and pollen tube to mediate pollen tube reception and release of sperm
70 cells to effect double fertilization [2-10]. Consequently, ~80% of ovules remain unfertilized
71 and only ~20% of seeds are produced in *fer* or *lre* mutant pistils [2-10]. FER-based
72 signaling pathway in ovules is also important to prevent late-arriving pollen tubes from
73 entering an ovule that has already engaged with a pollen tube and thus prevent
74 polyspermy [11]. In *A. thaliana* pollen and pollen tubes, the integrity of these cells during
75 reproduction is dependent on a signaling complex involving the *Cr*RLK1Ls ANXUR1
76 (ANX1), ANX2, BUDDHA'S PAPER SEAL1 (BUPS1), and BUPS2, along with the peptide
77 hormones RALF4/19 and LRE homologs LORELEI-LIKE GPI-ANCHORED PROTEIN2
78 (LLG2) and LLG3 [12-15].

79 In addition to functioning in reproduction, variants of the *Cr*RLK1L–LLG–RALF
80 signaling complex mediate diverse processes in vegetative tissues, stress responses,
81 and plant immunity [6, 12, 16-18]. The FER–LLG1–RALF1 complex regulates a Rho-
82 GTPase complex to produce reactive oxygen species (ROS) and is critical for root hair
83 growth and hypocotyl elongation [6, 10, 18]. The FER–LLG1–RALF23 complex regulates
84 pattern recognition receptor complex formation to modulate the perception of the

85 pathogen-associated molecular patterns (PAMPs) flagellin and elongation factor thermo
86 unstable (EF-TU) via the immunogenic epitopes flg22 and elf18, respectively. In this
87 context, RALF23 binds to FER–LLG1 to suppress the scaffolding function of FER, thereby
88 inhibiting ROS production and immunity [16, 17].

89 In *A. thaliana*, the GPI-anchored protein (GPI-AP), receptor kinase, and small CRP
90 components of the trimeric *Cr*RLK1L–LLG–RALF signaling complex are encoded by 17,
91 4, and 37 genes, respectively [19]. These could potentially form 2,516 unique
92 combinations that function in myriad cell types and at different developmental time points.
93 Additionally, this core trimeric complex may interact with other signaling components to
94 form a multimeric signaling complex and mediate a variety of cellular processes [15].
95 Additional components that could interact with this trimeric complex include LEUCINE-
96 RICH REPEAT EXTENSIN (LRX) proteins, which can directly bind RALF peptides [20].
97 LRXs are required for maintaining pollen tube integrity [13] and form a complex with FER
98 that is important for vacuolar expansion [21]. However, a comprehensive evolutionary
99 analysis complemented with functional and expression studies has been done for any of
100 the members of the *Cr*RLK1L–LLG–RALF signaling complex. Consequently, how the
101 expansion of these gene families contributes to the functional diversification of this critical
102 signaling complex and what factors affect the maintenance of duplicate members of these
103 gene families remain unknown.

104 We addressed this question by producing a phylogeny for the *LLG* family members
105 and characterizing their patterns of functional and regulatory evolution. We identified
106 orthologs of the four-member *A. thaliana LLG* gene family and showed that they are
107 conserved throughout the Brassicaceae. We used complementation assays and showed
108 that the molecular functions of GPI-AP proteins encoded by the *LLG* gene family are
109 conserved. Examination of the function and expression of *LRE* and *LLG1* family members
110 in *A. thaliana* and their single copy orthologs in *Cleome violacea* showed that regulatory
111 divergence (i.e. differences in gene expression), rather than functional divergence, likely
112 contributed to the retention of *LRE* and *LLG1* in *A. thaliana* and possibly played a key role
113 in the diversification of this signaling complex.

114

Results

LRE, *LLG1*, *LLG2*, and *LLG3* are maintained in the Brassicaceae

To study the evolution of the *LLG* gene family, we obtained full-length coding sequences (CDS) for orthologs of *LRE*, *LLG1*, *LLG2*, and *LLG3* in eleven species from the Brassicaceae. We also identified orthologs from three species outside the Brassicaceae, but from the same order (Brassicales): *Carica papaya* (Caricaceae), *Tarenaya hassleriana* (Cleomaceae), and *Cleome violacea* (Cleomaceae). Only a single ortholog of the *LLG* family was found in *Amborella trichopoda*, a basal angiosperm, which was used as the outgroup in this analysis (S1 Table). We generated an alignment of all of the full-length CDS, then inferred phylogeny using maximum likelihood methods. The resulting tree was rooted with the single-copy gene from *A. trichopoda*.

Among species of Brassicaceae, *LRE* and its orthologs formed a monophyletic group sister to the clade containing all *LLG1* orthologs (Fig 1). We identified single-copy *LRE/LLG1* orthologs from *C. papaya*, *T. hassleriana*, and *C. violacea* that were sister to the *LRE* + *LLG1* clade, consistent with the possibility that *LRE* and *LLG1* in the Brassicaceae are products of the alpha whole-genome duplication (WGD) [22] (Fig 1).

Our approach identified *LLG2* and *LLG3* orthologs in all the species of Brassicaceae we analyzed with two exceptions: no *LLG2* ortholog was identified in *Brassica rapa* and no LLG3 ortholog was identified in *Aethionema arabicum* (S1 Table). Absence of an ortholog could be due to gene loss, incomplete genome sequencing coverage, or because our approach failed to identify the putative orthologs (see Methods). Our phylogenetic analysis also found that the single-copy *LLG2/LLG3* orthologs in *T. hassleriana*, and *C. violacea* are sister to *LLG3* orthologs in Brassicaceae. This suggested that the duplication occurred prior to the split of *LLG2* and *LLG3* and that these species have an *LLG3* ortholog but likely lack an *LLG2* ortholog (Fig 1). However, this conclusion is based on branches of the phylogenetic tree that are only supported by low bootstrap values. Additional data from species in Brassicaceae and Cleomaceae are required to determine whether the duplication that led to the *LLG2* and *LLG3* clades in Brassicaceae occurred early in the history of Brassicaceae, or whether the duplication predated the split between Brassicaceae and Cleomaceae. Still, our phylogenetic

146 analyses pointed to maintenance of four copies of *LLG* in the genomes of species in the
147 Brassicaceae following at least two duplication events (Fig 1).
148

149 **The molecular functions of LLG1 are likely conserved in LRE, LLG2, and LLG3**

150 Following a whole-genome duplication, the duplicated genes are initially redundant;
151 accumulating mutations quickly render most gene duplicates non-functional, with a half-
152 life of a few million years [23]. For instance, only about 15% of duplicated genes were
153 retained in *A. thaliana* [24, 25]. However, our phylogenetic analysis showed that several
154 duplicates have been retained in the *LLG* family; these genes may have developed non-
155 redundant functions. Therefore, we examined whether divergence in molecular function
156 caused by differences in the transcribed genic regions could explain the maintenance of
157 the *LLG* gene family in the Brassicaceae.

158 To this end, we expressed *A. thaliana LRE*, *LLG2*, or *LLG3* from the *LLG1*
159 promoter in the *llg1-2* mutant and tested if they could complement the vegetative
160 development defects of *llg1-2* mutants [6, 16, 17, 26]. As a positive control in these
161 experiments, we used *llg1-2* mutant carrying a transgenic construct with *LLG1* expressed
162 from its own promoter. Expression of *LLG3* or *LRE* restored root hair phenotypes
163 comparable to the expression of *LLG1* (Fig 2A, S1B Fig) and complemented hypocotyl
164 length and epidermal pavement cell defects in *llg1-2* seedlings (Figs 2B,3A-3E, S1C Fig).
165 Additionally, seedlings expressing either of these two transgenes showed restored
166 RALF1 sensitivity (Fig 3G, S2 Fig). Notably, LRE complemented insensitivity to RALF1 in
167 *llg1-2*, restoring RALF1-induced root growth inhibition (Fig 3G). Expression of *LLG2* also
168 complemented vegetative defects in *llg1-2* mutants, as rosette size was restored to wild
169 type levels (Fig 4A).
170 In addition to vegetative phenotypes, *fer* and *llg1* mutants are defective in immune
171 responses, as they show reduced responsiveness to several PAMPs, including the
172 bacterial elicitors elf18 and flg22 [16, 17, 26]. We found that responses to flg22 and elf18
173 in *llg1-2* plants expressing *LLG2* from the *LLG1* promoter were restored to levels
174 comparable to those expressing *LLG1* from *LLG1* promoter, indicating that LLG2 can

175 substitute for LLG1 and perform its molecular functions (Fig 4B, S3 Fig).
176 Complementation of *llg1-2* with *LLG2* similarly restored responsiveness to exogenous
177 RALF23 peptide to levels detected in *llg1-2* plants expressing *LLG1* (Fig 4C). Taken
178 together, these results demonstrated that the molecular functions of LLG1 are mostly
179 indistinguishable from those of LRE, LLG2, or LLG3. Hence, the retention of *LLG* paralogs
180 in Brassicaceae was unlikely to have been due to divergence in molecular function
181 caused by differences in their transcribed genic regions.
182

183 **_LRE_ and _LLG1_ have distinct expression patterns in _A. thaliana_**

184 The *LLG* gene family members are differentially expressed in *A. thaliana*, with *LLG1*, *LRE*,
185 and *LLG2/LLG3* primarily expressed in vegetative, female, and male reproductive tissues,
186 respectively [9, 17]. Consistent with this, we found that putative transcription factor
187 binding sites in the promoters of *LLG* gene family members are distinct and showed
188 considerable variation (S4 Fig). Based on these findings, we considered the possibility
189 that divergence in expression of the *LLG* gene family (regulatory divergence), rather than
190 divergence in molecular function, underlies the maintenance of the *LLG* gene family in
191 the Brassicaceae. To test this possibility, we performed detailed, cell-specific expression
192 of *LRE* and *LLG1*, as they are closely related paralogs (Fig 1) that are also differentially
193 expressed [9]. Additionally, we chose to use these two genes to investigate this possibility
194 because *lre* and *llg1* single mutants have well-defined, non-overlapping phenotypes [6,
195 9, 17], allowing reciprocal complementation experiments to be performed (see below) to
196 test the functional divergence aspect of the hypothesis. By contrast, *llg2* and *llg3* single
197 mutants do not show phenotypes; phenotypes were detected only in a *llg2 llg3* double
198 mutant, thus making these genes not useful for reciprocal functional tests [6, 9, 17].
199 Previously, RT-PCR experiments indicated that expression of *LRE* is more tightly
200 restricted than *LLG1*, as *LRE* is primarily expressed in reproductive tissues and *LLG1* is
201 expressed throughout plant development [9]. Still, *LRE* and *LLG1* expression overlapped
202 in at least three tissues: 8-day-old seedlings, unfertilized pistils, and pollinated pistils [9,
203 27]. Thus, it is unknown if the domains of *LRE* and *LLG1* expression are indeed distinct
204 within these multicellular tissues. We therefore examined cell-specific expression of *LRE*

205 and *LLG1* using transcriptional fusions of the *LRE* or *LLG1* promoters to β-glucuronidase
206 (GUS). We characterized GUS expression in three *pLRE::GUS* lines (this study), and in
207 a previously reported *pLLG1::GUS* line [6]. GUS expression was examined in the
208 following tissues where either *LRE* or *LLG1* is expressed [9, 27]: 8-day-old seedlings, 21-
209 day-old seedlings, unpollinated pistils 24 hours after emasculation (HAE), and pollinated
210 pistils at 13.5 hours after pollination (HAP) and 18 HAP (Fig 5).

211 In 8-day-old seedlings, *pLLG1::GUS* was expressed in true leaves, the hypocotyl,
212 and to a lesser extent in roots (Fig 5A and 5B). However, we were unable to detect GUS
213 staining at this timepoint in any of the three *pLRE::GUS* lines (Fig 5A and 5C). The
214 *pLRE::GUS* expression results were not consistent with the previous RT-PCR results
215 obtained using 8-day-old seedlings [9, 27]. Perhaps the *LRE* promoter sequence used in
216 this construct did not include all the *cis*-regulatory elements required for expression in 8-
217 day-old seedlings. Alternatively, *pLRE::GUS* may be expressed in these tissues, but at
218 levels below the detection limit of our assay. In 21-day-old seedlings, *pLRE::GUS*
219 expression matched the RT-PCR results [9]. *pLLG1::GUS* was expressed in newly
220 emerged true leaves (Fig 5E) and *pLRE::GUS* was not detected in any cell or tissues of
221 21-day-old seedlings (Fig 5F).

222 In unpollinated pistils, *pLRE::GUS* was strongly expressed in the synergid cells
223 (Fig 5G and 5H). After pollination, *pLRE::GUS* expression was weaker in the zygote at
224 13.5 HAP (Fig 5J) and in proliferating endosperm at 18 HAP (Fig 5I). The cell-specific
225 expression of *pLRE::GUS* at these stages was consistent with the results obtained using
226 *pLRE::GFP* [27]. *pLLG1::GUS* was expressed only in the septum and nectaries in
227 unpollinated and pollinated pistils (Fig 5G, 5I, 5K, and 5M). These results demonstrated
228 that *LRE* and *LLG1* are not expressed in the same cells in seedlings and pistils. These
229 findings are consistent with the regulatory divergence hypothesis and suggested that
230 expression differences between *A. thaliana LRE* and *LLG1* likely contributed to their
231 retention post duplication.

232

233 **LLG1 complements the reproductive functions of LRE in *A. thaliana***

234 Reciprocal complementation experiments are one way to investigate if functional
235 divergence likely contributed to their retention post duplication. We showed that LRE can
236 substitute for LLG1 molecular functions in vegetative development (Fig 2). To investigate
237 if LLG1 can perform the reproductive functions of LRE, we expressed LLG1 from the *LRE*
238 promoter, and transformed it into *lre-7* plants. To allow us to visualize the protein, we also
239 fused LLG1 to citrine Yellow Fluorescent Protein (c*YFP*). In three homozygous single-
240 locus insertion *pLRE::LLG1-cYFP* lines, LLG1-cYFP localized to the filiform apparatus
241 and puncta in the synergid cells (S5 Fig), similar to what we previously reported for LRE
242 using *pLRE::LRE-cYFP* (S5 Fig; [7]. All three homozygous single-locus insertion
243 *pLRE::LLG1-cYFP* lines restored seed set defects in *lre-7* to levels detected in wild type
244 or *pLRE::LRE-cYFP* (Fig 6B). Additionally, when the transgenic plant was used as the
245 female parent in a cross to wild type, there was a significantly increased transmission of
246 *pLRE::LLG1-cYFP* transgene in the progeny of the cross, showing that LLG1-cYFP
247 complemented the defects in the *lre-7* female gametophyte (Table 1); no such increase
248 in transmission was observed in the progeny from a reciprocal cross, when pollen from
249 the transgenic plant was crossed to wild type (Table 1). Based on these results, we
250 concluded that LLG1 expressed under the *LRE* promoter complements the reproductive
251 defects in *lre-7* mutants, and that the molecular functions of LRE and LLG1 are mostly
252 conserved.
253

254 **The single copy LRE/LLG1 ortholog in *C. violacea* can substitute for both LLG1 and**
255 **LRE in *A. thaliana***

256 Based on reciprocal complementation experiments in *A. thaliana*, we concluded that LRE
257 and LLG1 can perform each other's molecular functions. Additionally, phylogenetic
258 analysis (Fig 1) indicated that these paralogs were a product of the alpha WGD that
259 occurred at the base of Brassicaceae [22]. Taken together, these results raise the
260 possibility that the molecular functions performed by the single copy LRE/LLG1 ortholog
261 in *C. violacea*, a member of the Cleomaceae, sister family to Brassicaceae (Fig 1), are
262 conserved in LRE and LLG1. To test this prediction, we fused *Clevi-LRE/LLG1* to cYFP,
263 and expressed it from the *A. thaliana LLG1* or *LRE* promoters (S1A and S5 Figs), then

264 transformed these constructs into *llg1-2* and *lre-7* plants, respectively. We found that
265 Clevi-LRE/LLG1 expressed from the *LLG1* promoter complemented root hair defects (Fig
266 2a and S1B Fig), epidermal cell defects (Fig 3F), and hypocotyl lengths in dark-grown
267 *llg1-2* seedlings (Fig 2B, S1C Fig) to levels seen in *llg-1-2* lines carrying LLG1. Clevi-
268 LRE/LLG1 also complemented *llg1-2* insensitivity to RALF1-induced root growth inhibition
269 comparable to levels seen in *llg1-2* lines carrying LLG1 (Fig 3G and S2 Fig), which also
270 suggests that the *Cr*RLK1L–LLG–RALF signaling complex is conserved outside of the
271 Brassicaceae.

272 In all three *pLRE::Clevi-LRE/LLG1-cYFP* single-locus insertion transgenic lines,
273 we detected partial complementation of reproductive defects in *lre-7*, as the seed set in
274 these lines were significantly higher than that in *lre-7*, yet significantly lower when
275 compared to wild-type (6B Fig). Additionally, when the transgenic plants were used as
276 the female parent in a cross to wild type, there was a significantly increased transmission
277 of *pLRE::Clevi-LRE/LLG1-cYFP* transgene in the progeny of the cross, showing that Clevi
278 LRE/LLG1-cYFP complemented the defects in the *lre-7* female gametophyte (Table 2);
279 no such increase in transmission was observed in the progeny from a reciprocal cross,
280 when pollen from the transgenic plant was crossed to wild type (Table 2). Based on these
281 results, we concluded that the single-copy *LRE/LLG1* in *C. violacea* partially
282 complemented *lre-7* reproductive phenotypes.

283 Partial complementation was perhaps due to relatively lower protein levels of Clevi-
284 LRE/LLG1 compared to LRE as revealed by the cYFP fusion proteins levels in the filiform
285 apparatus of the synergid cells in these lines (S5G Fig). Increased sequence divergence
286 in the single-copy *LRE/LLG1* ortholog in *C. violacea* compared to *LRE* rather than *LLG1*
287 is another possibility for partial complementation, as the branch length leading to the *LRE*
288 clade was longer than the branch length leading to *LLG1*, indicating more substitutions
289 post duplication in LRE compared to LLG1 (Fig 1). Since LRE and LLG1 were able to
290 complement each other's functions (Figs 2–4), a third possibility is that the changes that
291 that affect the ability to complement occurred along the branch leading to *C. violacea*.
292 Nevertheless, our results indicate that the *C. violacea* single-copy LRE/LLG1 ortholog is
293 capable of complementing the functions of both *A. thaliana* LRE and LLG1.

294

295 **The single copy *LRE/LLG1* ortholog in *C. violacea* shows broad expression in both**
296 **vegetative and reproductive tissues**

297 Conserved molecular functions in *LRE* orthologs coupled with non-overlapping
298 expression of *pLRE::GUS* and *pLLG1::GUS* in *A. thaliana* indicate an important role for
299 regulatory divergence in the maintenance of the *LLG* gene family paralogs in
300 Brassicaceae. Such divergence is a hallmark of regulatory sub-functionalization [28, 29].
301 In this case, the ancestral single copy gene would have been expressed in vegetative
302 and reproductive tissues and post duplication, this expression pattern would have been
303 partitioned between the descendant paralogs.

304 To test this prediction, we characterized the expression of the single copy
305 *LRE/LLG1* ortholog in *C. violacea* (*Clevi-LRE/LLG1*; Fig 7). We performed RT-PCR on
306 cDNA isolated from the following three stages of *C. violacea* development in which *LLG1*
307 is expressed, but *LRE* is not expressed, in *A. thaliana*: 1) rosette leaves from 30-day-old
308 plants (equivalent in size to 21-day-old *A. thaliana* rosette leaves; Fig 5), 2) anthers and
309 pollen [9, 17], and 3) mature emasculated pistils without ovules that still included the
310 septum and nectaries (Fig 5). In these RT-PCR experiments, we also included ovules
311 isolated from mature emasculated pistils in which *LRE*, but not *LLG1*, is expressed in *A.*
312 *thaliana* (Fig 5).

313 *Clevi-LRE/LLG1* was expressed in all developmental stages that we tested and
314 *Clevi-LRE/LLG1* expression encompassed all the expression domains of *A. thaliana LRE*
315 + *LLG1* (Fig 7). The RT-PCR products amplified were sequenced to confirm that they
316 were indeed full-length transcripts of *Clevi-LRE/LLG1*. Expression analysis in *A. thaliana*
317 and *C. violacea* showed that the ancestral expression pattern of *Clevi-LRE/LLG1* is
318 divided between the paralogs, leading to non-overlapping expression in *A. thaliana*. This
319 provided additional support to the hypothesis that regulatory sub-functionalization
320 underlies *LLG* paralog retention in Brassicaceae.

321

322 ## Discussion

323

324 ### Regulatory divergence likely led to the retention of *LRE* and *LLG1*

325 Multiple mechanisms have been proposed to explain the retention of duplicated genes,
326 as progressive degeneration of one member of a paralogous set of genes is the default
327 outcome (non-functionalization; [30]). Duplicated genes may acquire new functions (neo-
328 functionalization) and hence may be retained [31]. A third mechanism of retention is sub-
329 functionalization via sequence variations in the protein-coding or regulatory regions of the
330 genes, by which the ancestral functions could be shared between the duplicated copies,
331 resulting in non-overlapping expression of each paralog [29]. The advent of high-
332 throughput sequencing and large-scale transcriptomic studies have allowed evaluation of
333 expression divergence in duplicated genes in *A. thaliana* [28, 32-34], *Glycine max* [35],
334 and *Zea mays* [36]. Although these studies generated critical evidence in support of
335 regulatory sub-functionalization and/or regulatory neo-functionalization, they also relied
336 on comparisons of reconstructed expression of the ancestral expression states rather
337 than direct expression analysis of a species with a single copy ortholog. Similarly,
338 functional complementation studies, particularly those across species, are critical for
339 establishing the mechanisms contributing to duplicate retention; however, very few genes
340 have been subjected to such studies.

341 Here, we demonstrated that a single copy ortholog from *C. violacea*, which
342 diverged prior to the alpha WGD that gave rise to *LRE* and *LLG1* in the Brassicaceae, is
343 able to fully substitute for LLG1 in vegetative tissues and can partially perform the
344 functions of LRE in synergid cells. Expression analyses of *LRE*, *LLG1*, and *Clevi-*
345 *LRE/LLG1* support our hypothesis that after gene duplication, the *LRE* and *LLG1* clades
346 experienced regulatory sub-functionalization yielding non-overlapping expression
347 patterns [33]. The combined expression domains of *LRE* and *LLG1* together with the
348 *Clevi-LRE/LLG1* expression domains suggest that a single-copy *LRE/LLG1* played
349 multiple roles in the common ancestor of the Brassicaceae and Cleomaceae. Based on

these results, we propose that the complementary expression of *LRE* and *LLG1* led to the retention of these paralogs in Brassicaceae, as loss of either copy would cause loss of expression in certain tissues and lower fitness [29].

Expression divergence is also a key player in the evolution of *LLG2* and *LLG3* in Brassicaceae, as we found that the single-copy *LLG2/LLG3* ortholog in *C. violacea* is expressed in all the tissues examined (Fig 7) even though *A. thaliana LLG2/LLG3* are primarily expressed in pollen and pollen tubes [9, 17]. Consistent with these results, highly diverged transcription factor binding sites were found in the promoter regions of *LLG2* and *LLG3* in *A. thaliana* compared to the single copy ortholog *LLG2/LLG3* in *C. violacea* (S4 Fig).

Besides expression divergence, in this study we showed that in *A. thaliana,* members of the *LLG* family share molecular functions, as every member of this family can substitute for LLG1 in vegetative tissues and immune responses and LLG1 can function in ovules in place of LRE. We also showed that *Clevi* LRE/LLG1 can substitute for LLG1 and LRE. Taking the results of expression and molecular function analyses together, we propose that conserved molecular functions and expression divergence are the keys to the expansion and retention of the *LLG* gene family in Brassicaceae.

The *LLG* Gene family may be co-evolving with the *Cr*RLK1L and RALF families

Division of ancestral expression is one mechanism that allows duplicated genes to be retained [29]. Given our results, it is possible that the genes encoding the other members of the trimeric complex (*Cr*RLK1Ls and RALFs) show similar divisions of expression domains. Additional phylogenetic analyses combined with expression analyses will be required to understand the evolution of expression domains in *CrRLK1L* and *RALF* families in the Brassicaceae. *RALFs* and *CrRLK1Ls* are members of large gene families, which poses a challenge in efforts to study all the possible combinations of the co-receptor complex [5, 37]. However, phylogenetic analysis coupled with functional and expression studies, similar to that performed in this study, may offer a viable approach to address this challenge.

379 FER functions with both LRE and LLG1, and correspondingly, *FER* is expressed
380 in both *LRE* and *LLG1* expression domains [6, 9]. Given that specific *Cr*RLK1Ls, RALFs,
381 and LLGs play distinct biological roles, it may be that *Cr*RLK1Ls and RALFs have co-
382 evolved with members of the LLG family to perform their functions in different tissues. *In*
383 *vitro* binding of RALF23 with LLG1, LLG2, and LLG3 but not LRE provide support for this
384 possibility [17]. Methods such as evolutionary rate covariation have been used to link co-
385 evolution with functional associations [38]. Such methods, in combination with the
386 phylogenetics, expression analyses, and molecular genetic assays used in this study will
387 prove invaluable in further characterizing members of this critical signaling complex.

388 Materials and Methods

389 Identifying CDS for the orthologs of the *LLG* gene family

390 CDS of putative orthologs of the *LLG* gene family and single-copy *LRE/LLG1* and
391 *LLG2/LLG3* orthologs were obtained through the Comparative Genomics (CoGe)
392 Platform using CoGeBLAST (tBLASTx) with *A. thaliana LRE* or *LLG1* nucleotide CDS as
393 the search query using standard parameters [39-41]. For each result, if annotations for
394 CDS were available, then they were downloaded directly using "FeatView" on CoGe. In
395 cases of incomplete CDS or when no CDS annotations were available, a 5–7-kb region
396 surrounding the sequence of interest was downloaded and then aligned to the original
397 BLAST query from *A. thaliana LRE* and *LLG1* to identify the entire putative CDS using an
398 exon-by-exon approach with Geneious Alignment (R11.1.2) (https://www.geneious.com).

399 To identify putative orthologs from our tBLASTx results, we performed reciprocal
400 BLASTs to *A. thaliana*, followed by alignments and phylogenetic trees using *A. thaliana*
401 *LRE*, *LLG1*, *LLG2*, and *LLG3* sequences (see below). Putative orthologs were determined
402 based on the most closely related *A. thaliana* paralog. To find *LRE/LLG1* and *LLG2/LLG3*
403 single-copy orthologous genes in *Tarenaya hassleriana*, *Cleome violacea*, and *Carica*
404 *papaya*, we used tBLASTx of *A. thaliana LRE* and *LLG1* nucleotide CDS as the search
405 queries. We built an alignment of these sequences, *LRE*, and its paralogs in *A. thaliana*,
406 and used the alignment to build a phylogenetic tree. Only two loci were identified for each
407 genome: one corresponding to *LRE/LLG1*, the other corresponded to *LLG2/LLG3*;

408 therefore, we named the single-copy genes after their corresponding phylogenetic group.
409 We found in polyploid species that there were typically additional copies of paralogs. We
410 named these numerically, without any particular preference. We identified a single-copy
411 gene in *Amborella trichopoda*, as tBLASTx with any *LLG* gene family member in *A.*
412 *thaliana* only identified the single-copy ortholog in *A. trichopoda*. A list of the genomes
413 and corresponding gene IDs for putative orthologs are presented in Table S1.

414

415 **CDS alignments and phylogenies**

416 Alignments were built using the standard parameters for MUSCLE 3.8.425 in Geneious
417 R11.1.2, followed by manual curation using Geneious [42, 43]. From CDS alignments, we
418 built phylogenetic trees using the RAxML 8.2.11 plugin in Geneious with the following
419 parameters: GTR GAMMA nucleotide model; rapid bootstrapping and search for best-
420 scoring ML tree algorithm, with 100 bootstrap replicates; and starting with a completely
421 random tree [44]. Phylogenetic trees were rooted with *Amborella trichopoda*, which
422 served as the outgroup. The alignment and trees resulting from these analyses were
423 deposited to TreeBASE (Accession URL during review: https://treebase.org/treebase-
424 web/search/study/summary.html?id=25583&x-access-
425 code=2ede7b91dbd6e347d5c22132b139cc70&agreement=ok and Final URL after
426 publication: http://purl.org/phylo/treebase/phylows/study/TB2:S25583).

427

428 **Plant materials and growth conditions**

429 *A. thaliana* and *C. violacea* seeds were liquid sterilized in the following manner: 10–300
430 seeds were placed into a 1.5-mL microcentrifuge tube with 1 mL of 70% EtOH and
431 vortexed for 3 seconds at maximum speed at least 3 times over the course of 3–5 minutes
432 to ensure all seeds were sufficiently exposed to the sterilizing solution. The 70% EtOH
433 solution was discarded and replaced with 1 mL of sterilization solution (50% bleach, 0.2%
434 Tween-20 [Sigma-Aldrich, Catalog # P9416-100ML]), then vortexed for 3 seconds at
435 maximum speed, at least 3 times over 3–5-minutes. The sterilization solution was

436 discarded, and seeds were washed three times with 1 mL of ice-cold autoclaved dH_2O
437 each time. Using a 1-mL pipette, seeds were plated on 1/2X Murashgi and Skoog (MS)
438 plates (Carolina Biological Supply Co., Catalog # 195703), with 1% sucrose for seedling
439 growth assays and with corresponding antibiotics for transmission assays.
440 Seeds on plates were stratified for 3 days in the dark and at 4 °C, then plates were
441 moved to a Percival growth chamber maintained at 21 °C with continuous light (75–100
442 $\mu mol \cdot m^{-2} \cdot s^{-1}$). Ten-to-fourteen-day-old seedlings were transplanted to soil and grown in
443 the following conditions: 16 h light (100–120 $\mu mol \cdot m^{-2} \cdot s^{-1}$) at 21 °C and 8 h dark at 18 °C
444 as described [45].
445 Columbia (Col-0) is the ecotype of all *A. thaliana* seeds used in this study.
446 *pLRE::GUS* was transformed into the Col-0 background. *pLRE::LRE-cYFP*, *pLRE::LLG1-*
447 *cYFP*, and *pLRE::Clevi-LRE/LLG1-cYFP* are all in the homozygous *lre-7* mutant
448 background and were selected on plates containing hygromycin B (20 µg/mL;
449 PhytoTechnology Laboratories, Catalog # H397) and glufosinate ammonium (10 µg/mL;
450 Oakwood Chemical, Catalog # 044851). The *pLLG1::SP-mRFP-LLG1* and *pLLG1::SP-*
451 *mRFP-LLG2* lines were transformed into *llg1-2* plants by floral dip, and transformants
452 were selected on 1x MS-agar with 1% sucrose, supplemented with 25 µg/mL kanamycin
453 or 10 µg/mL glufosinate ammonium, respectively. Seedlings for seedling growth inhibition
454 assays were grown in 12 h light (120 120 $\mu mol \cdot m^{-2} \cdot s^{-1}$) at 19–21 °C. Plants for ROS burst
455 assays were grown in 10 h light (150 $\mu mol \cdot m^{-2} \cdot s^{-1}$) at 20 °C.

456

457 **Cloning transgenic constructs**

458 The *pLLG1::LRE-HA-GPI* and *pLLG1::LLG3* transgenic constructs were prepared using
459 the primers and DNA templates indicated in Supplemental table S4, and contained the
460 *LLG1* promoter region (~2 kb upstream of ATG) as presented in the previously published
461 constructs [6]. For SP-mRFP-LLG1 expression *in planta*, the native *LLG1* promoter was
462 amplified from seedling genomic DNA using primers
463 "ccaagcttgcatgccGTCGTTGTCCCAGATTCGTCG" and
464 "gatctagagtcgaccGGTTCTTTGTTGGTTACAGGAGAAGTCAC" for subsequent cloning
465 into the Gateway destination vector pGWB1 [46] using Infusion cloning (Takara Bio) and

466 the SdaI/SbfI restriction enzyme (Thermo Scientific). The *SP-mRFP-LLG1* construct was
467 synthesized with attached attB1/attB2 sites (Twist Bioscience) and subsequently
468 recombined into pDONR-Zeo using BP II Clonase (Invitrogen). The resulting pDONR-Zeo
469 SP-mRFP-LLG1 construct was recombined with *pGWB1-pLLG1*. The *pLLG1::SP-mRFP-*
470 *LLG2* construct was synthesized (Thermo Scientific) with attB1/attB2 sites added and
471 cloned into *pDONR-Zeo* via BP II Clonase (Invitrogen) and was subsequently recombined
472 with *pGWB601* [47] for *in planta* expression.

473 The *pLRE::LLG1-cYFP, pLRE::GUS*, *pLRE::Clevi-LRE/LLG1-cYFP,* and
474 *pLLG1::Clevi-LRE/LLG1-cYFP* transgenic lines were created by replacing the genomic
475 sequence of *LRE* in the *pLRE::LRE-cYFP* construct with the gene of interest [7]. The
476 *LLG1* promoter region for *pLLG1::Clevi-LRE/LLG1-cYFP* PCR is the same as previously
477 published [6]. PCR was performed with PrimeSTAR GXL DNA Polymerase (TaKaRa Bio
478 Inc.; Catalog # R050A) using primers and DNA templates listed in Table S4. The amplified
479 PCR products were cloned into *pLRE::LRE-cYFP* plasmid linearized with *SpeI-HF* (NEB,
480 Catalog # R3133S) and *AscI* (NEB, Catalog # R0558S) by using the In-Fusion HD Cloning
481 Plus system (Clontech, Catalog # 639645). The recombinant plasmids were transformed
482 into Stellar Competent Cells (Clontech, Catalog # 636763), and positive colonies were
483 selected on LB plates containing spectinomycin (100 µg/mL, Sigma-Aldrich, Catalog #
484 85555).

485 Constructs generated were sequence verified (Eton Bioscience, Inc.) before being
486 transformed into *Agrobacterium tumefaciens* (GV3101 pMP90 strain). The positive colony
487 selected for transforming into *A. thaliana* was also verified by colony PCR for the presence
488 of the transgene.

489

490 **Plant transformation**

491 Transformation solution containing *Agrobacterium tumefaciens* (GV3101 pMP90 strain)
492 harboring the desired transgene was either sprayed onto *A. thaliana* inflorescences or
493 applied by the floral dip method [48]. Hygromycin-resistant T_1 transformants were
494 selected as described [49].

495

496 **Isolation of single-locus insertion lines**

497 For each construct, single-locus insertion lines were isolated as described [7].

498

499 **Scoring cYFP in mature unpollinated pistils**

500 cYFP expression in mature unpollinated pistils were scored as described [7]. Samples
501 were mounted in 5% glycerol with a coverslip, and YFP expression in synergid cells was
502 scored by epifluorescence on a Zeiss Axiophot microscope with a GFP filter (excitation
503 HQ 470/40 and emission HQ 525/50). Images were acquired with Picture Frame
504 (Optronics).
505 Confocal images of the filiform apparatus in synergid cells were taken using a
506 Leica SP5 confocal laser scanning microscope. For cYFP imaging, samples were excited
507 with a 488-nm laser line, and emission spectra between 510 and 550 nm were collected.
508 YFP images were processed with Leica Application Suite X and ImageJ software
509 (http://imagej.nih.gov/ij/).

510

511 **Complementation of root hair and hypocotyl length phenotypes in *Ilg1* seedlings**

512 Root hair analysis was performed as previously described [18]. Root hairs located
513 between 1.5 and 3.5 mm from the primary root tip of four-day-old seedlings were observed
514 with a stereoscope. The number of normal and defective (stunted and collapsed) root
515 hairs was scored. The hypocotyl length assay was performed as described [6]. Three-
516 day-old dark-grown seedlings were imaged using Epson Perfection V370 Photo Scanner
517 at 600 dpi resolution and the length of hypocotyl was measured with Image J.

518

519 **Propidium iodide staining of pavement cells**

520 Six-day-old seedlings were stained with 50 μg/mL propidium iodide (PI) for 20 mins and
521 then excess stain was removed by washing with ddH_2O. Images were acquired using
522 confocal microscopy on a NIKON A1 Spectral System and analyzed by the NIS-Elements
523 AR Analysis Software (V 5.02) with 40x objective. Images were acquired using optimal
524 laser power and gain with excitation wavelength 561 nm and emission wavelength 595
525 nm.

526

527 **RALF1-induced root growth inhibition assays**

528 The root sensitivity assays with RALF1 treatment were performed as described [6, 50] :
529 three-day-old light-grown seedlings were treated with RALF1 for 2 days at concentrations
530 indicated in Figs 3, S2. Primary root length was measured using Image J at the beginning
531 and end of treatments to obtain growth during treatment.

532

533 **RALF23-induced seedling growth inhibition assays**

534 Surface-sterilized T2 *pLLG1::SP-mRFP-LLG1* and *pLLG1::SP-mRFP-LLG2* seeds were
535 selected on 1x MS-agar with 1% sucrose supplemented with kanamycin or glufosinate
536 ammonium, respectively, alongside Col-0 and *llg1-2* on 1x MS-agar with 1% sucrose
537 without selection. Five-day-old seedlings were then transferred to liquid 1x MS with 1%
538 sucrose in sterile 48-well plates, with or without 1 μM RALF23 peptide (Scilight) [17].
539 Seedling fresh weight was measured after 7 days of growth in liquid medium.

540

541 **ROS burst measurements**

542 ROS burst measurements were performed as described previously [17]. Briefly, twelve 4-
543 mm leaf discs from 4.5-week-old plants were harvested and equilibrated overnight in 96-
544 well plates in sterile, deionized water. The next day, water was replaced with PAMP

545 solution (100 nM flg22 or elf18 peptide plus 10 µg/mL horse radish peroxidase and 100
546 µM luminol). ROS was immediately measured using a charge-coupled device camera
547 (Photek). Total ROS production was calculated as the sum of Random Light Units value
548 over 40 minutes of PAMP treatment.

549

550 **Seed set scoring**

551 Unfertilized, viable (enlarged after fertilization), and aborted ovules in siliques (10 days
552 after pollination) were scored as described [9]. Three to five self-pollinated siliques
553 located between 5th and 15th siliques from the bottom of the main branch of an *A. thaliana*
554 plant were scored.

555

556 **GUS staining of *pLRE::GUS* and *pLLG1::GUS* transgenic plants**

557 Tissues were stained for GUS activity as described [51]. Stained samples were mounted
558 in 50% glycerol and observed for epifluorescence on a Zeiss Axiovert 100 microscope
559 and images were taken using Metamorph (Version 7; Molecular Devices) using a red-
560 green-blue filter with autoexposure settings. For each tissue, at least five samples were
561 observed for each genotype.

562

563 **Transcription factor binding site analysis in the promoters of *LLG* gene family**
564 **members in *A. thaliana* and *C. Violacea***

565 DAP-seq data were obtained from the Plant Cistrome Database
566 (http://neomorph.salk.edu/dev/pages/shhuang/dap_web/pages/index.php) in the form of
567 MEME motif file format. These 838 motifs were scanned against the putative promoter
568 sequences (1 kb upstream of the transcriptional start, not overlapping with neighboring
569 genes) of the *LLG* gene family members in *A. thaliana* (*LRE*, *LLG1*, *LLG2*, and *LLG3)*
570 and the single-copy orthologs *LRE/LLG1* and *LLG2/LLG3* in *Cleome violacea* with MAST
571 [52] using default parameters. The scanned results were used to determine the presence

572 of motif sites. A motif site was considered present in a putative promoter sequence if its
573 mapping p-value was <1e-3. The motif site information was combined for all motifs to
574 generate a motif presence (1)/absence (0) matrix for hierarchical clustering using the
575 heatmap.2 function in R.

576

577 **RNA isolation and RT-PCR of *Cleome violacea***

578 The following tissues were collected from *C. violacea* for RNA isolation: rosette leaves
579 from 30-day-old plants, mature pistils 24 HAE with ovules removed, and mature ovules
580 24 HAE removed from the pistils. For each tissue type, three biological replicates were
581 collected. For rosette tissues, 3 rosette leaves were collected for each biological replicate.
582 Unlike *A. thaliana*, where every pistil has the potential to mature in every flower, we
583 observed that not all flowers in *C. violacea* develop a mature pistil. We determined mature
584 pistils alternate every 2–4 flowers; therefore, to collect pistils and ovules we first
585 monitored the maturation pattern and then emasculated 2–3 pistils which may mature.
586 Twenty-four hours later, mature pistils were dissected, ovules collected separately from
587 the other pistil tissues (septum, transmitting tract, stigma, style, carpel walls, nectaries).
588 100 ovules were collected for each biological replicate. For pistil without ovule tissues, 4
589 pistils were harvested for each biological replicate. All tissues were flash-frozen in liquid
590 nitrogen and stored at -80 °C until RNA extraction.

591 RNA was isolated using RNeasy Plant Mini Kit (Qiagen, Catalog # 74904)
592 according to the manufacturer's instructions and treated with RNase-free DNase I (Life
593 Technologies, Catalog # AM2222) to remove residual genomic DNA. Samples were
594 cleaned up using RNeasy MinElute Cleanup Kit (QIAGEN, Catalog #74204) and tested
595 for RNA integrity by Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was reverse
596 transcribed from 550 ng of total RNA using SuperScript IV First-Strand Synthesis System
597 (ThermoFisher Scientific, Catalog # 18091050).

598 PCR was performed with 20 ng of cDNA for each reaction, using the following PCR
599 cycle conditions: 1. 98 °C for 2 minutes; 2. 95 °C for 30 seconds, 3. 56 °C for 20 seconds,
600 4. 72 °C for 1 minute 10 seconds, 5. Repeat cycles 2–5 for 34 cycles, 6. 72 °C for 10
601 minutes, 7. Hold at 4 °C. PCR products were analyzed by gel electrophoresis on a 1%

602 agarose gel with ethidium bromide in 1X TAE buffer. Gel images were processed using
603 ImageJ. *Clevi-LRE/LLG1* and the control gene *Clevi-ACTIN2* were amplified using
604 primers listed in Table S4. Two technical replicates of each biological replicate were
605 performed.

606

607 **Image processing**

608 ImageJ was used to assemble image panels, insert scale bars, and prepare figures.

609

610 **Genomes and accession numbers**

611 *A. thaliana LRE* (*At4g26466*), *LLG1* (*At5g56170*), *LLG2* (*At2g20700*), and *LLG3*
612 (*At4g28280*) were obtained from The Arabidopsis Information Resource (Version:
613 TAIR10). *LRE, LLG1, LLG2, LLG3, LRE/LLG1,* and *LLG2/LLG3* orthologs were identified
614 using the Comparative Genomics platform (CoGe) [40, 41]. Genome references and
615 accession numbers can be found in Table S1.

616

617 ## Acknowledgements

618 We acknowledge Dr. Patrick Edger from Michigan State University for providing us
619 access to the *Cleome violacea* genome on CoGe and Dr. Jocelyn Hall from the University
620 of Alberta for providing *Cleome violacea* seeds. We thank Ramin Yadegari (University of
621 Arizona) for the Zeiss Axiophot microscope. We thank the past and present members of
622 the Palanivelu Lab for discussions. J.A.N. was supported by the following: IGERT
623 Comparative Genomics Program at the University of Arizona (Award ID: 0654435); NSF
624 Graduate Research Fellowship: Grant DGE-1143953; the Boynton Graduate Fellowship;
625 the American Society of Plant Biologists; and the University of Arizona Graduate College
626 Office of Diversity and Inclusion. Additional support for this work was provided by an NSF
627 grant to R.P. (IOS-1146090) and University of Arizona Undergraduate Biology Research
628 Program fellowship to S.H. (private donors). This work was also supported by the

University of Zürich (C.Z.), the European Research Council under the Grant Agreement 773153 (grant IMMUNO-PEPTALK to C.Z.), as well as fellowships from the European Molecular Biology Organization, the Natural Sciences and Engineering Research Council of Canada and the Deutsche Forschungsgemeinschaft (fellowships EMBO-LTF 100-2017 and NSERC PDF-532561-2019 to T.A.D.; DFG STE 2448/1 to M.S.T.). Work in A.Y.C. lab was supported by Natural Science Foundation (IOS-1645854 and MCB-1715764) to A.Y.C. and Hen-Ming Wu, and the National Institute of Food and Agriculture, U.S. Department of Agriculture, the Center for Agriculture, Food and the Environment under project number MAS00525. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the USDA or NIFA. We also thank H.M.W. for contribution of LLG1 and LLG3 constructs. K.M. was supported by the Torrey Summer Research Scholarship from UMass Plant Biology Program and the Linda Slakey Summer Research Scholarship from UMass BMB department. This study was also supported by funding from NSF pollen RCN grant for sponsoring activities and meetings that forged collaborations between A.Y.C. and R.P. labs (MCB0955910).

Author contributions

JAN, MJL, TAD, MS, KM, BS, KD, NK, and SH planned and designed the research in consultation with and guidance of their respective principal investigators, performed experiments, collected and/or analyzed data. SS performed the transcription factor binding site analysis in the promoters of *LLG* family members. JAN and MAB performed all the evolutionary analysis reported in this study. JAN, RP, MJL, MS, TAD, SS, CZ, MAB, and AYC wrote the manuscript with input from and revisions by all authors.

653 ## References

654

655 1. Johnson MA, Harper JF, Palanivelu R. A Fruitful Journey: Pollen Tube
656 Navigation from Germination to Fertilization. Annual Review of Plant Biology. 2019. doi:
657 10.1146/annurev-arplant-050718-100133.

658 2. Capron A, Gourgues M, Neiva LS, Faure JE, Berger F, Pagnussat G, et al.
659 Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-
660 anchored protein encoded by the *LORELEI* gene. Plant Cell. 2008;20(11):3038-49.
661 PubMed PMID: 19028964.

662 3. Escobar-Restrepo J-M, Huck N, Kessler S, Gagliardini V, Gheyselinck J, Yang
663 W-C, et al. The FERONIA Receptor-like Kinase Mediates Male-Female Interactions
664 During Pollen Tube Reception. Science. 2007;317(5838):656-60. doi:
665 10.1126/science.1143562.

666 4. Huck N, Moore JM, Federer M, Grossniklaus U. The Arabidopsis mutant feronia
667 disrupts the female gametophytic control of pollen tube reception. Development.
668 2003;130(10):2149-59. doi: 10.1242/dev.00458.

669 5. Kessler SA, Lindner H, Jones DS, Grossniklaus U. Functional analysis of related
670 CrRLK1L receptor-like kinases in pollen tube reception. EMBO reports. 2015;16(1):107-
671 15. doi: 10.15252/embr.201438801.

672 6. Li C, Yeh F-L, Cheung AY, Duan Q, Kita D, Liu M-C, et al.
673 Glycosylphosphatidylinositol-anchored proteins as chaperones and co-receptors for
674 FERONIA receptor kinase signaling in Arabidopsis. eLife. 2015;4:e06587. doi:
675 10.7554/eLife.06587.

676 7. Liu X, Castro C, Wang Y, Noble J, Ponvert N, Bundy M, et al. The Role of
677 LORELEI in Pollen Tube Reception at the Interface of the Synergid Cell and Pollen

678 Tube Requires the Modified Eight-Cysteine Motif and the Receptor-Like Kinase
679 FERONIA. The Plant Cell. 2016;28(5):1035-52. doi: 10.1105/tpc.15.00703.

680 8. Rotman N, Rozier F, Boavida L, Dumas C, Berger F, Faure JE. Female control of
681 male gamete delivery during fertilization in *Arabidopsis thaliana*. Curr Biol.
682 2003;13(5):432-6. PubMed PMID: 12620194.

683 9. Tsukamoto T, Qin Y, Huang Y, Dunatunga D, Palanivelu R. A role for LORELEI,
684 a putative glycosylphosphatidylinositol-anchored protein, in Arabidopsis thaliana double
685 fertilization and early seed development. Plant J. 2010;62(4):571-88. PubMed PMID:
686 20163554.

687 10. Duan Q, Kita D, Johnson EA, Aggarwal M, Gates L, Wu HM, et al. Reactive
688 oxygen species mediate pollen tube rupture to release sperm for fertilization in
689 Arabidopsis. Nature communications. 2014;5:3129. doi: 10.1038/ncomms4129.
690 PubMed PMID: 24451849.

691 11. Duan Q, Liu M-CJ, Kita D, Jordan SS, Yeh F-LJ, Yvon R, et al. FERONIA
692 controls pectin- and nitric oxide-mediated male–female interaction. Nature. 2020. doi:
693 10.1038/s41586-020-2106-2.

694 12. Ge Z, Bergonci T, Zhao Y, Zou Y, Du S, Liu M-C, et al. Arabidopsis pollen tube
695 integrity and sperm release are regulated by RALF-mediated signaling. Science. 2017.
696 doi: 10.1126/science.aao3642.

697 13. Mecchia MA, Santos-Fernandez G, Duss NN, Somoza SC, Boisson-Dernier A,
698 Gagliardini V, et al. RALF4/19 peptides interact with LRX proteins to control pollen tube
699 growth in Arabidopsis. Science. 2017. doi: 10.1126/science.aao5467.

700 14. Feng H, Liu C, Fu R, Zhang M, Li H, Shen L, et al. LORELEI-LIKE GPI-
701 ANCHORED PROTEINS 2/3 regulate pollen tube growth as chaperones and

702 coreceptors for ANXUR/BUPS receptor kinases in Arabidopsis. Molecular plant. 2019.
703 doi: https://doi.org/10.1016/j.molp.2019.09.004.

704 15. Ge Z, Dresselhaus T, Qu L-J. How CrRLK1L Receptor Complexes Perceive
705 RALF Signals. Trends in plant science. 2019;24(11):978-81. doi:
706 10.1016/j.tplants.2019.09.002.

707 16. Stegmann M, Monaghan J, Smakowska-Luzan E, Rovenich H, Lehner A, Holton
708 N, et al. The receptor kinase FER is a RALF-regulated scaffold controlling plant immune
709 signaling. Science. 2017;355(6322):287-9. doi: 10.1126/science.aal2541.

710 17. Xiao Y, Stegmann M, Han Z, DeFalco TA, Parys K, Xu L, et al. Mechanisms of
711 RALF peptide perception by a heterotypic receptor complex. Nature.
712 2019;572(7768):270-4. doi: 10.1038/s41586-019-1409-7.

713 18. Duan Q, Kita D, Li C, Cheung AY, Wu H-M. FERONIA receptor-like kinase
714 regulates RHO GTPase signaling of root hair development. Proceedings of the National
715 Academy of Sciences. 2010;107(41):17821-6. doi: 10.1073/pnas.1005366107.

716 19. Campbell L, Turner SR. A Comprehensive Analysis of RALF Proteins in Green
717 Plants Suggests There Are Two Distinct Functional Groups. Frontiers in plant science.
718 2017;8:37. Epub 2017/02/09. doi: 10.3389/fpls.2017.00037. PubMed PMID: 28174582;
719 PubMed Central PMCID: PMCPMC5258720.

720 20. Moussu S, Broyart C, Santos-Fernandez G, Augustin S, Wehrle S, Grossniklaus
721 U, et al. Structural basis for recognition of RALF peptides by LRX proteins during pollen
722 tube growth. Preprint at https://wwwbiorxivorg/content/101101/695874v1 2019. doi:
723 10.1101/695874.

724 21. Dünser K, Gupta S, Herger A, Feraru MI, Ringli C, Kleine-Vehn J. Extracellular
725 matrix sensing by FERONIA and Leucine-Rich Repeat Extensins controls vacuolar

726 expansion during cellular elongation in Arabidopsis thaliana. The EMBO Journal.
727 2019;38(7):e100353. doi: 10.15252/embj.2018100353.

728 22. Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm
729 genome evolution by phylogenetic analysis of chromosomal duplication events. Nature.
730 2003;422(6930):433-8. Epub 2003/03/28. doi: 10.1038/nature01521. PubMed PMID:
731 12660784.

732 23. Lynch M, Conery JS. The evolutionary demography of duplicate genes. Journal
733 of Structural and Functional Genomics. 2003;3(1):35-44. doi:
734 10.1023/A:1022696612931.

735 24. Blanc G, Hokamp K, Wolfe KH. A Recent Polyploidy Superimposed on Older
736 Large-Scale Duplications in the Arabidopsis Genome. Genome Research.
737 2003;13(2):137-44.

738 25. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, et al.
739 Modeling gene and genome duplications in eukaryotes. Proceedings of the National
740 Academy of Sciences of the United States of America. 2005;102(15):5454. doi:
741 10.1073/pnas.0501102102.

742 26. Shen Q, Bourdais G, Pan H, Robatzek S, Tang D. Arabidopsis
743 glycosylphosphatidylinositol-anchored protein LLG1 associates with and modulates
744 FLS2 to regulate innate immunity. Proceedings of the National Academy of Sciences.
745 2017. doi: 10.1073/pnas.1614468114.

746 27. Wang Y, Tsukamoto T, Noble JA, Liu X, Mosher RA, Palanivelu R. Arabidopsis
747 LORELEI, a Maternally Expressed Imprinted Gene, Promotes Early Seed Development.
748 Plant physiology. 2017;175(2):758-73. doi: 10.1104/pp.17.00427.

749 28. Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, et al.
750 Expression pattern shifts following duplication indicative of subfunctionalization and

751 neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol. 2006;23(2):469-
752 78. Epub 2005/11/11. doi: 10.1093/molbev/msj051. PubMed PMID: 16280546.

753 29. Force A, Lynch M, Pickett FB, Amores A, Yan Y-l, Postlethwait J. Preservation of
754 Duplicate Genes by Complementary, Degenerative Mutations. Genetics.
755 1999;151(4):1531-45.

756 30. Panchy N, Lehti-Shiu M, Shiu S-H. Evolution of Gene Duplication in Plants. Plant
757 physiology. 2016;171(4):2294. doi: 10.1104/pp.16.00523.

758 31. Ohno S. Evolution by Gene Duplication: Springer-Verlag; 1970.

759 32. De Smet R, Sabaghian E, Li Z, Saeys Y, Van de Peer Y. Coordinated Functional
760 Divergence of Genes after Genome Duplication in Arabidopsis thaliana. The Plant Cell.
761 2017;29(11):2786. doi: 10.1105/tpc.17.00531.

762 33. Liu S-L, Baute GJ, Adams KL. Organ and cell type-specific complementary
763 expression patterns and regulatory neofunctionalization between duplicated genes in
764 Arabidopsis thaliana. Genome biology and evolution. 2011;3:1419-36. Epub
765 2011/11/04. doi: 10.1093/gbe/evr114. PubMed PMID: 22058183.

766 34. Zou C, Lehti-Shiu MD, Thomashow M, Shiu S-H. Evolution of Stress-Regulated
767 Gene Expression in Duplicate Genes of Arabidopsis thaliana. PLOS Genetics.
768 2009;5(7):e1000581. doi: 10.1371/journal.pgen.1000581.

769 35. Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, et al. The fate of
770 duplicated genes in a polyploid plant genome. The Plant Journal. 2013;73(1):143-53.
771 doi: 10.1111/tpj.12026.

772 36. Hughes TE, Langdale JA, Kelly S. The impact of widespread regulatory
773 neofunctionalization on homeolog gene evolution following whole-genome duplication in
774 maize. Genome Research. 2014;24(8):1348-55. doi: 10.1101/gr.172684.114.

775 37. Sharma A, Hussain A, Mun BG, Imran QM, Falak N, Lee SU, et al.
776 Comprehensive analysis of plant rapid alkalization factor (RALF) genes. Plant
777 physiology and biochemistry : PPB / Societe francaise de physiologie vegetale.
778 2016;106:82-90. Epub 2016/05/08. doi: 10.1016/j.plaphy.2016.03.037. PubMed PMID:
779 27155375.

780 38. Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF. Evolutionary
781 Rate Covariation Identifies New Members of a Protein Network Required for Drosophila
782 melanogaster Female Post-Mating Responses. PLoS Genet. 2014;10(1):e1004108. doi:
783 10.1371/journal.pgen.1004108.

784 39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment
785 search tool. J Mol Biol. 1990;215(3):403-10. Epub 1990/10/05. doi: 10.1016/s0022-
786 2836(05)80360-2. PubMed PMID: 2231712.

787 40. Lyons E, Freeling M. How to usefully compare homologous plant genes and
788 chromosomes as DNA sequences. The Plant journal : for cell and molecular biology.
789 2008;53(4):661-73. Epub 2008/02/14. doi: 10.1111/j.1365-313X.2007.03326.x. PubMed
790 PMID: 18269575.

791 41. Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, et al. Finding and
792 Comparing Syntenic Regions among Arabidopsis and the Outgroups Papaya, Poplar,
793 and Grape: CoGe with Rosids. Plant physiology. 2008;148(4):1772. doi:
794 10.1104/pp.108.124867.

795 42. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
796 throughput. Nucleic Acids Res. 2004;32(5):1792-7. Epub 2004/03/23. doi:
797 10.1093/nar/gkh340. PubMed PMID: 15034147; PubMed Central PMCID:
798 PMCPMC390337.

799 43. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time
800 and space complexity. BMC Bioinformatics. 2004;5(1):113. doi: 10.1186/1471-2105-5-
801 113.

802 44. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-
803 analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. doi:
804 10.1093/bioinformatics/btu033.

805 45. Kessler SA, Shimosato-Asano H, Keinath NF, Wuest SE, Ingram G, Panstruga
806 R, et al. Conserved molecular components for pollen tube reception and fungal
807 invasion. Science. 2010;330(6006):968-71. PubMed PMID: 21071669.

808 46. Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, et al.
809 Development of series of gateway binary vectors, pGWBs, for realizing efficient
810 construction of fusion genes for plant transformation. Journal of Bioscience and
811 Bioengineering. 2007;104(1):34-41. doi: https://doi.org/10.1263/jbb.104.34.

812 47. Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, et al. Improved
813 Gateway Binary Vectors: High-Performance Vectors for Creation of Fusion Constructs
814 in Transgenic Analysis of Plants. Bioscience, Biotechnology, and Biochemistry.
815 2007;71(8):2095-100. doi: 10.1271/bbb.70216.

816 48. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated
817 transformation of Arabidopsis thaliana. The Plant journal : for cell and molecular biology.
818 1998;16(6):735-43. Epub 1999/03/09. doi: 10.1046/j.1365-313x.1998.00343.x. PubMed
819 PMID: 10069079.

820 49. Harrison SJ, Mott EK, Parsley K, Aspinall S, Gray JC, Cottage A. A rapid and
821 robust method of identifying transformed Arabidopsis thaliana seedlings following floral
822 dip transformation. Plant Methods. 2006;2(1):19. doi: 10.1186/1746-4811-2-19.

823 50. Haruta M, Sabat G, Stecker K, Minkoff BB, Sussman MR. A Peptide Hormone
824 and Its Receptor Protein Kinase Regulate Plant Cell Expansion. Science.
825 2014;343(6169):408-11. doi: 10.1126/science.1244454.

826 51. Johnson MA, Kost B. Pollen tube development. Methods Mol Biol. 2010;655:155-
827 76. PubMed PMID: 20734260.

828 52. Bailey TL, Gribskov M. Combining evidence using p-values: application to
829 sequence homology searches. Bioinformatics. 1998;14(1):48-54. Epub 1998/04/01. doi:
830 10.1093/bioinformatics/14.1.48. PubMed PMID: 9520501.

831

832 ## Figure Legends

833 **Fig. 1. The *LLG* gene family is maintained in the Brassicaceae family.**

834 A maximum likelihood phylogenetic tree of full-length CDS of single copy orthologs of
835 *LRE* and *LLG1* (*LRE/LLG1*) and *LLG2* and *LLG3* (*LLG2/LLG3)* in *Carica papaya* (pink
836 and grey boxes) and two species in the Cleomaceae (*Cleome violacea* and *Tarenaya*
837 *hassleriana*) (purple and red boxes, respectively) and 11 species in the Brassicaceae.
838 The *LRE* and its orthologs formed a distinct clade (blue box) from *LLG1* and its orthologs
839 (orange box). *LLG2* and its orthologs (green box) formed a separate clade than *LLG3* and
840 its orthologs (yellow box). The phylogenetic tree was generated with 100 bootstrap
841 replicates and was rooted using a single copy ortholog identified in *Amborella trichopoda*.
842 Only bootstrap values ≥ 50 are represented along each branch.

843

844 **Fig 2. LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complemented vegetative defects**
845 **in *llg1*.**

846 (A) Complementation of root hair defects in *llg1-2* seedlings. Root hairs were scored as
847 normal or defective in at least 10 seedlings in each indicated single-insertion line carrying
848 LRE-HA, LLG3, or Clevi-LRE/LLG1-cYFP fusion protein. (B) Complementation of short
849 hypocotyl lengths in dark grown *llg1-2* mutant seedlings. Hypocotyl length was assayed
850 in T2 seedlings from selfed seeds of single insertion lines carrying LRE-HA, LLG3, or
851 Clevi-LRE/LLG1-cYFP fusion protein. Quantification of hypocotyl lengths was done in at
852 least three trials of 20-25 seedlings for each line. Error bars represent ±SD.

853

854 **Fig 3. LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complemented epidermal defects**
855 **and RALF1 insensitivity in *llg1-2* seedlings.**

856 (A-F) Epidermal pavement cells of 6-day-old *llg1-2* seedlings expressing LRE-HA, LLG3,
857 or Clevi-LRE/LLG1-cYFP showed restored normal pavement cell morphology like that
858 seen in wild-type (Col-0) and *llg1-2* seedlings expressing LLG1. In each genotype, 5-10
859 seedlings were stained with Propidium Iodide (PI) and visualized with confocal

860 microscopy. (G) Percentage of root growth after RALF1 treatment of three-day-old
861 seedlings as described in S2 Fig. Root length was measured two days after RALF1
862 treatment, and three trials were performed. Error bars represent ±SD.

863

864 **Fig 4. mRFP-LLG1 and mRFP-LLG2 complemented defects in *llg1-2*.**

865 (A) Expression of *pLLG1::SP-mRFP-LLG*1 or *pLLG1::SP-mRFP-LLG2* restores rosette
866 size in 4.5-week-old *llg1-2* plants. (B) ROS production in response to flg22 (left) or elf18
867 (right) is restored in SP-mRFP-LLG1 or SP-mRFP-LLG2 lines driven by the *LLG1* native
868 promoter. Total Relative Light Unit (RLU) over 40 minutes of exposure to 100 nM flg22 or
869 elf18 treatment is displayed. Letters indicate significantly different values (n=12 leaf discs,
870 two-way ANOVA with Tukey test, flg22 $p<0.0001$; elf18 $p=0.0009$). Error bars show ±SD.
871 (C) RALF23 sensitivity is restored in seedlings expressing SP-mRFP-LLG1 or SP-mRFP-
872 LLG2. Letters indicate significantly different values (n=16 seedlings, two-way ANOVA
873 with Tukey test, $p<0.0001$). Error bars show ±SD.

874

875 **Fig 5. *pLRE::GUS* and *pLLG1::GUS* showed non-overlapping expression in**
876 **vegetative and reproductive tissues.**

877 (A-F) *pLLG1::GUS* was expressed, while *pLRE::GUS* was not, in vegetative tissues. In 8-
878 day-old seedlings, *pLLG1::GUS* was expressed in true leaves, hypocotyls, and roots (A-
879 C). In 21-day-old seedlings, *pLLG1::GUS* was expressed in the epicotyl, the hypocotyl,
880 and weakly expressed in roots (D-F). (G) At 24 HAE, *pLRE::GUS* and *pLLG1::GUS* were
881 both expressed in pistils but in different cell-types. *pLRE::GUS* was expressed in synergid
882 cells, while *pLLG1::GUS* was expressed in septum. Close up of area marked in red
883 rectangles are shown below. (H-I) *pLRE::GUS* was expressed in synergid cells at 24
884 hours after emasculation (HAE) (H) but *pLLG1::GUS* is not expressed in the ovule (I). (J-
885 M) *pLRE::GUS* and *pLLG1::GUS* showed non-overlapping expression after pollination.
886 Mature unpollinated pistils were pollinated with Col-0 pollen and collected at 13.5 HAP
887 (J-K) or 18 HAP (l-M) and stained for GUS activity. (J-K) At 13.5 HAP, *pLRE::GUS* was
888 expressed in the micropylar end of the female gametophyte (J), while *pLLG1::GUS*

889 continues to be expressed in the septum (K). (L-M) At 18 HAP, *pLRE::GUS* and
890 *pLLG1::GUS* were both expressed in pollinated pistils but in different cell-types.
891 *pLRE::GUS* was expressed in the zygote and developing endosperm nuclei, while
892 *pLLG1::GUS* was expressed in septum.

893

894 **Fig 6. LLG1-cYFP and Clevi-LRE/LLG1-cYFP complemented reproductive defects**
895 **in *lre*.**

896 (A) Images of opened siliques of indicated genotypes in *A. thaliana.* A representative
897 unfertilized ovule (*) and viable (V) or aborted (A) seed is marked in the *lre* silique. (B)
898 LLG1-cYFP complemented lre mutant seed set defects in self-pollinated pistils of
899 indicated three independent transformants (ANOVA, $p = 0.18$). (C) Clevi-LRE/LLG1-cYFP
900 partially complemented lre mutant seed set defects in self-pollinated pistils of indicated
901 three independent transformants (pairwise two-tailed t-tests, $p > 0.05$). (B,C) Number in
902 the middle of each column refers to the number of ovules/seeds scored. Groups sharing
903 same lowercase letters are similar to each other in statistical tests.

904

905 **Fig 7. *Clevi-LRE/LLG1* and *Clevi-LLG2/LLG3* are expressed in vegetative and**
906 **reproductive tissues of *Cleome violacea*.**

907 RT-PCR of full-length *Clevi-LRE/LLG1* or *Clevi-LLG2/LLG3* in cDNAs isolated from 30-
908 day-old rosette leaves, anther and pollen, emasculated pistils without ovules, and ovules
909 from emasculated pistils of *Cleome violacea*. A homolog of *A. thaliana ACTIN2* (*Clevi-*
910 *ACTIN2*) was used as a control in these experiments. gDNA, genomic DNA isolated from
911 of Cleome leaves was used as a positive control in PCR portion of the RT-PCR
912 experiment. RT-PCR was repeated with two additional biological replicates with similar
913 results and the amplified bands were sequenced to confirm the identity of amplified
914 cDNAs.

915

916

917 ## Supplemental Information

918 **Fig S1. LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complemented vegetative**
919 **defects in *llg1-2* seedlings.**

920 (A) Diagrams of the *pLLG1::LRE-HA, pLLG1::LLG3,* and *pLLG1::Clevi-LRE/LLG1-cYFP*
921 constructs. (B) Root hair length defects were restored when LRE-HA, LLG3, and Clevi-
922 LRE/LLG1-cYFP were expressed in *llg1-2* seedlings to wild-type (Col-0) levels. Images
923 were taken from T2 seedlings of single insertion lines. Quantification of data shown here
924 are reported in Fig 2A. (C) LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complementation
925 of short hypocotyl length phenotype in *llg1-2* seedlings. Representative images of wild-
926 type (Col-0) or T2 seedlings of single-locus insertion lines of *pLLG1::LRE-HA,*
927 *pLLG1::LLG3,* or *pLLG1::Clevi-LRE/LLG1-cYFP* are shown. Quantification of data shown
928 here are reported in Fig 2B.

929

930 **Fig S2. LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complemented root hair lengths**
931 **in *llg1-2* seedlings prior to RALF treatment and increased RALF1 sensitivity after**
932 **treatment.**

933 Root lengths before and after RALF1 treatment were measured in single representative
934 lines from *LRE-HA, LLG3,* and *Clevi-LRE/LLG1-cYFP*. Root lengths in three-day-old
935 seedlings were measured, then treated with 0 μM RALF1 (untreated), 0.5 μM RALF1, or
936 1 μM RALF1. Roots were measured two days after RALF1 treatments with three trial
937 replicates. Error bars represent ±SD.

938

939 **Fig S3 ROS production kinetics in response to flg22 (top) or elf18 (bottom) is**
940 **restored in *pLLG1::SP-mRFP-LLG1* or *pLLG1::SP-mRFP-LLG2* lines.**

941 ROS burst levels in response to 100 nM flg22 (top) or elf18 (bottom) over time is indicated
942 in Relative Light Unit (RLU). In each trace of indicated genotypes, data shown is average
943 of n=12 leaf discs, ±SE

944

945 **Fig S4 Distribution of putative transcription factor (TF) binding sites in the putative**
946 **promoter regions of LRE homologs in *Arabidopsis thaliana* and *Cleome violacea*.**

947 (A) TFs with putative binding sites as determined using DNA affinity purification-
948 sequencing (DAP-seq) are shown as present (yellow boxes, $p < 1e\text{-}4$) or absent (blue
949 boxes). The TF family and gene name information were based on the Plant Cistrome
950 Database. *: amplified DAP-seq where secondary DNA modifications were removed. (B)
951 Table showing the frequency of putative TF binding site occurrence in the promoters of
952 *LRE*, *LLG1*, and the single copy ortholog *LRE/LLG1* in *Cleome violacea.* (C) Table
953 showing the frequency of putative TF binding site occurrence in the promoters of *LLG2*,
954 *LLG3*, and the single copy ortholog *LLG2/LLG3* in *Cleome violacea.*

955

956 **Fig S5 LLG1-cYFP and Clevi-LRE/LLG1-cYFP were expressed in synergid cells and**
957 **localized to the FA.**

958 (A-B) Diagrams of the *pLRE::LLG1-cYFP* and *pLRE::Clevi-LRE/LLG1-cYFP* constructs.
959 (C) A diagram of a mature ovule with a 7-celled female gametophyte. Synergid cells are
960 located in the micropylar end of the ovule, adjacent to the egg cell. The finger-like
961 projections of the FA are shown in yellow. A red arrow points to the FA. (D) In mature
962 unpollinated pistils, LRE-cYFP is expressed in the synergid cells, with localization in the
963 puncta in the synergid cell cytoplasm and in the FA. The ovule is outlined in light gray
964 dashed line, while the female gametophyte is outlined in dark gray dashed line. The red
965 rectangle marks the synergid cells. (E) Close-up image of the LRE-cYFP in the synergid
966 cells marked by the red rectangle in Fig C, outlined in dark gray dashed line. YFP localized
967 in the FA and puncta in the synergid cells. (F) Close-up image of the LLG1-cYFP in the
968 synergid cells with YFP localization in the puncta and the FA of the synergid cells. (g)
969 Close-up image of the Clevi-LRE/LLG1-cYFP in the synergid cells. YFP is weakly
970 expressed in the FA, but was not present elsewhere in the synergid cells, including the
971 puncta

972 **Table S1.** List of genes and genomes.

973 **Table S2.** List of primers used in this study.

974

975

Table 1. Enhanced transmission of the *pLRE::LLG1-cYFP* transgene through the *lre-7* female gametophyte.

Parents		Observed No. of progeny		Transmission Efficiency (TE) Analysis		
Female parent+	Male parent+	Hygr*	Hygs*	TE (R/S)	χ2†	P-value
WT	*LRE-cYFP-23*	129	116	1.11	0.34	0.56#
LRE-cYFP-23	*WT*	156	31	5.03	47.13	6.64E-12
WT	*LLG1-cYFP-A2*	213	171	1.25	2.30	0.13#
LLG1-cYFP-A2	*WT*	163	24	6.79	60.05	9.25E-15
WT	*LLG1-cYFP-10*	60	71	0.85	0.46	0.50#
LLG1-cYFP-10	*WT*	179	30	5.97	60.95	5.87E-15
WT	*LLG1-cYFP-11*	147	79	1.86	10.47	0.001#
LLG1-cYFP-11	*WT*	138	16	8.63	57.32	3.71E-14

+ Line Numbers refer to three independent transformants in the *lre-7/lre-7* background containing single insertion of the *pLRE::LLG1-cYFP* transgene. Genotype of each transgenic line is heterozygous for the transgene (*pLRE::LLG1-cYFP/+*) and homozygous for the *lre-7* mutation (*lre-7/lre-7*).

***** Hygromycin resistant (Hygr) and susceptible (Hygs) progeny. Hygromycin resistance gene is linked to the construct carrying the *pLRE::LLG1-cYFP* transgene.

TE, Transmission efficiency was calculated as the ratio of hygromycin resistance (R) to susceptibility (S) in the progeny of the indicated cross

† χ2 is calculated based on an expected segregation ratio of hygromycin resistant to susceptibility of 1:1

No significant deviation from 1:1 segregation through the male gametophyte indicates that pollen parent contains a single insertion of the *pLRE::LLG1-cYFP* transgene. Additional details on our protocol to isolate single insertion lines can be found in the methods.

976

977

Table 2. Enhanced transmission of the *pLRE::Clevi-LRE/LLG1- cYFP* transgene through the *lre-7* female gametophyte.

Parents		Observed No. of progeny		Transmission Efficiency (TE) Analysis		
Female parent+	Male parent+	Hygr*	Hygs*	TE (R/S)	χ2†	P-value
WT	*LRE-cYFP-23*	212	182	1.16	1.14	0.28#
LRE-cYFP-23	*WT*	335	43	7.79	132.56	1.13E-30
WT	*Clevi-LRE/LLG1-cYFP-10*	127	130	0.98	0.02	0.89#
Clevi-LRE/LLG1-cYFP-10	*WT*	87	52	1.67	4.49	0.03
WT	*Clevi-LRE/LLG1-cYFP-11*	150	113	1.33	2.62	0.10#
Clevi-LRE/LLG1-cYFP-11	*WT*	109	52	2.10	10.44	0.001
WT	*Clevi-LRE/LLG1-cYFP-12*	90	88	1.02	0.01	0.92#
Clevi-LRE/LLG1-cYFP-12	*WT*	84	28	3	14.93	0.0001

+ Line Numbers refer to three independent transformants in the *lre-7/lre-7* background containing single insertion of the *pLRE::Clevi-LRE/LLG1-cYFP* transgene. Genotype of each transgenic line is heterozygous for the transgene (*pLRE::Clevi-LRE/LLG1-cYFP/*+) and homozygous for the *lre-7* mutation (*lre-7/lre-*7).

***** Hygromycin resistant (Hygr) and susceptible (Hygs) progeny. Hygromycin resistance gene is linked to the construct carrying the *pLRE::Clevi-LRE/LLG1-cYFP* transgene.

TE, Transmission efficiency was calculated as the ratio of hygromycin resistance (R) to susceptibility (S) in the progeny of the indicated cross

† χ2 is calculated based on an expected segregation ratio of hygromycin resistant to susceptibility of 1:1

978

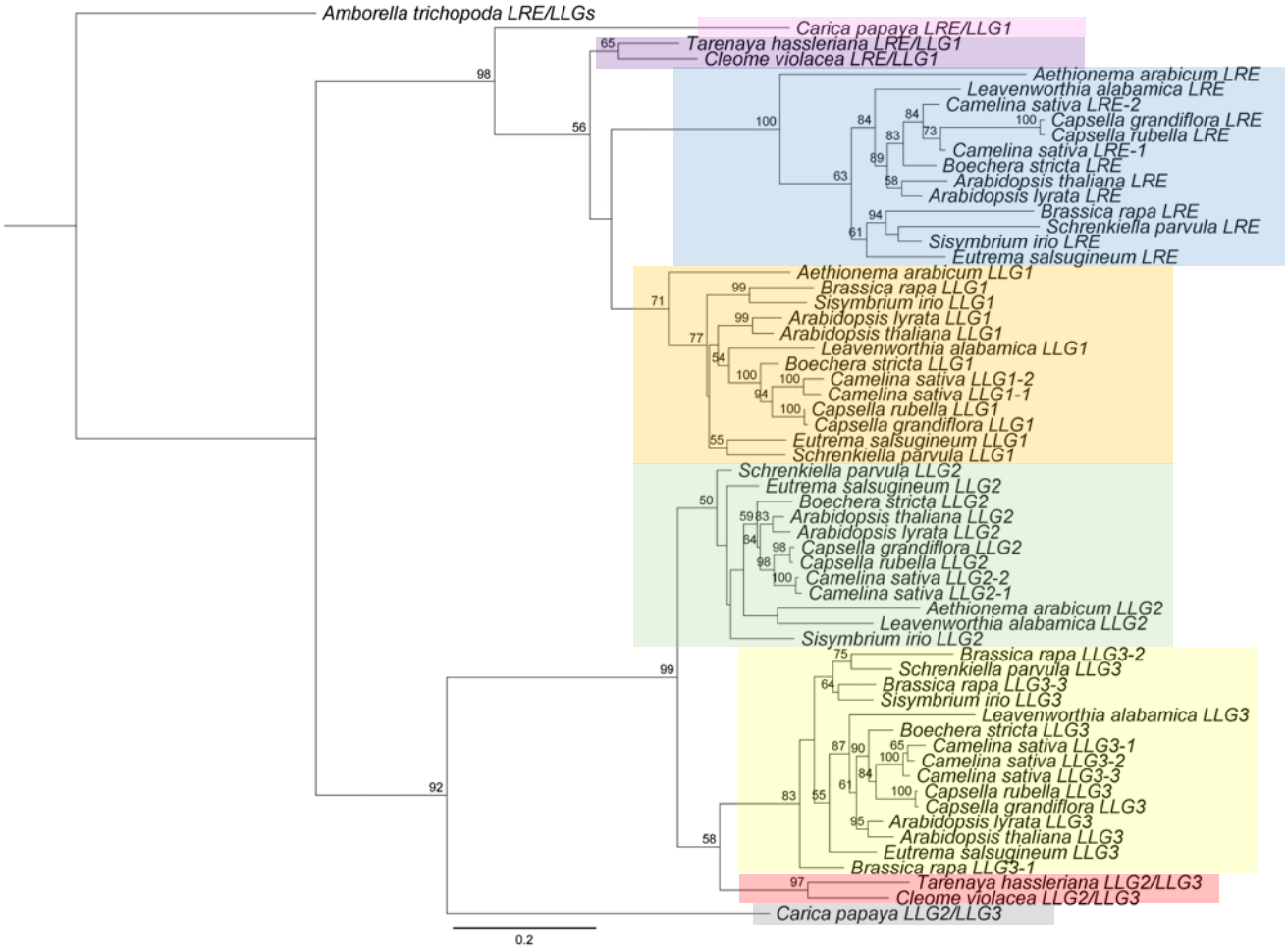

Fig 1 The *LLG* gene family is maintained in the Brassicaceae family.

A maximum likelihood phylogenetic tree of full-length CDS of single copy orthologs of *LRE* and *LLG1* (*LRE/LLG1*) and *LLG2* and *LLG3* (*LLG2/LLG3)* in *Carica papaya* (pink and grey boxes) and two species in the Cleomaceae (*Cleome violacea* and *Tarenaya hassleriana*) (purple and red boxes, respectively) and 11 species in the Brassicaceae. The *LRE* and its orthologs formed a distinct clade (blue box) from *LLG1* and its orthologs (orange box). *LLG2* and its orthologs (green box) formed a separate clade than *LLG3* and its orthologs (yellow box). The phylogenetic tree was generated with 100 bootstrap replicates and was rooted using a single copy ortholog identified in *Amborella trichopoda*. Only bootstrap values ≥ 50 are represented along each branch.

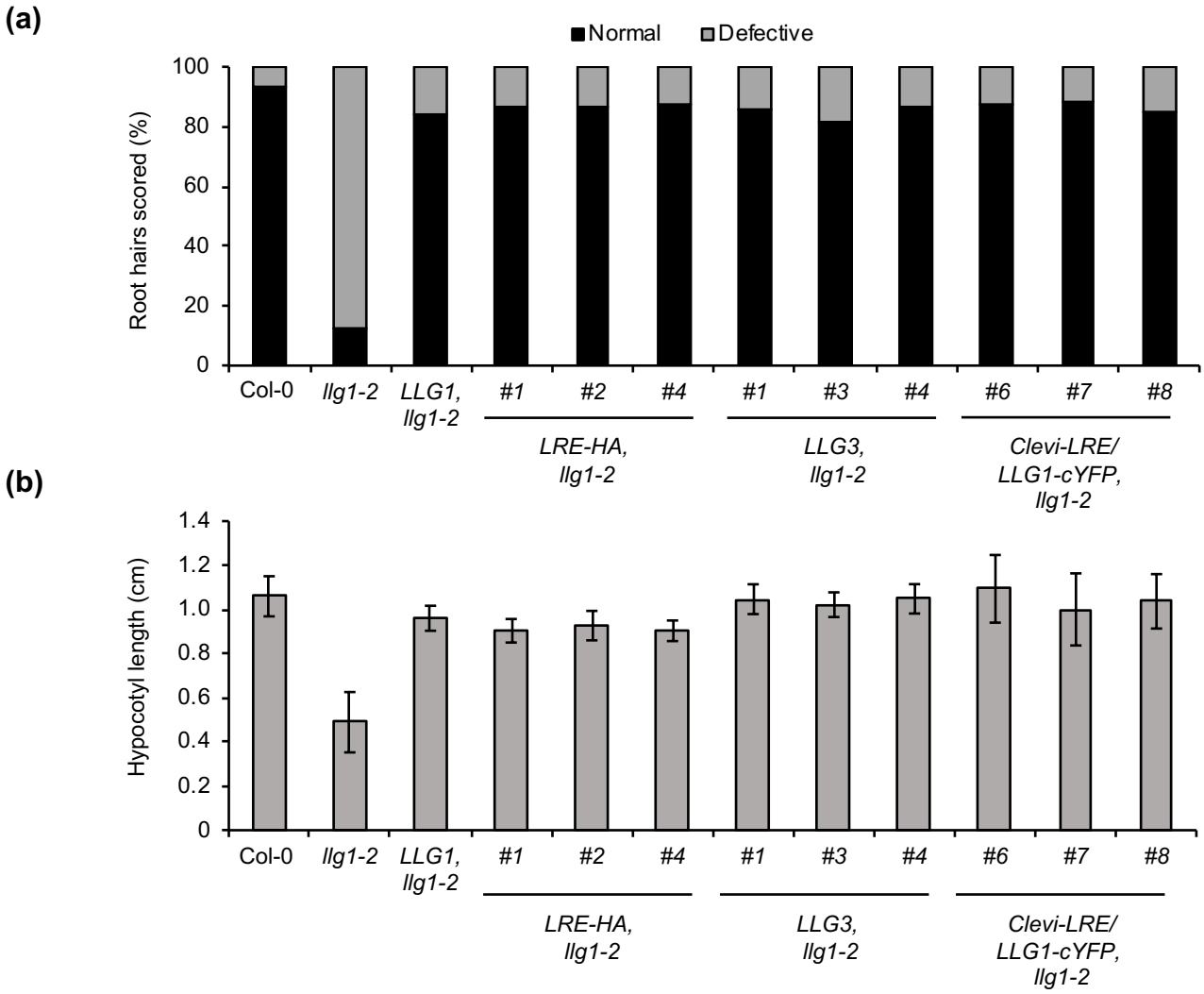

Fig 2 LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complemented vegetative defects in *llg1*.

(a) Complementation of root hair defects in *llg1-2* seedlings. Root hairs were scored as normal or defective in at least 10 seedlings in each indicated single-insertion line carrying LRE-HA, LLG3, or Clevi-LRE/LLG1-cYFP fusion protein.

(b) Complementation of short hypocotyl lengths in dark grown *llg1-2* mutant seedlings. Hypocotyl length was assayed in T2 seedlings from selfed seeds of single insertion lines carrying LRE-HA, LLG3, or Clevi-LRE/LLG1-cYFP fusion protein. Quantification of hypocotyl lengths was done in at least three trials of 20-25 seedlings for each line. Error bars represent ±SD.

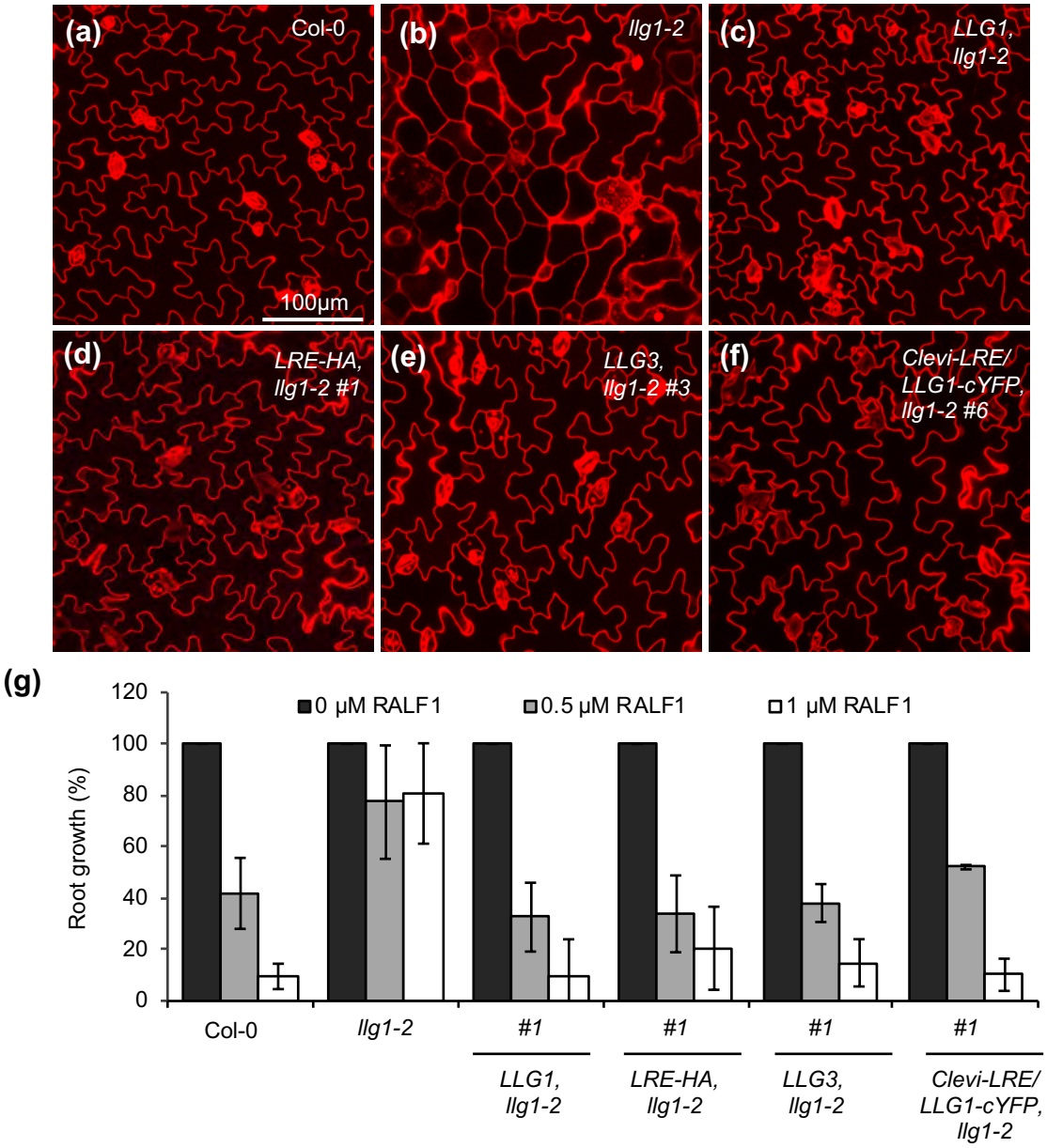

Fig 3 LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complemented epidermal defects and RALF1 insensitivity in *llg1-2* seedlings.

(a-f) Epidermal pavement cells of 6-day-old *llg1-2* seedlings expressing LRE-HA, LLG3, or Clevi-LRE/LLG1-cYFP showed restored normal pavement cell morphology like that seen in wild-type (Col-0) and *llg1-2* seedlings expressing LLG1. In each genotype, 5-10 seedlings were stained with Propidium Iodide (PI) and visualized with confocal microscopy.

(g) Percentage of root growth after RALF1 treatment of three-day-old seedlings as described in Supplemental Figure 2. Root length was measured two days after RALF1 treatment, and three trials were performed. Error bars represent ±SD.

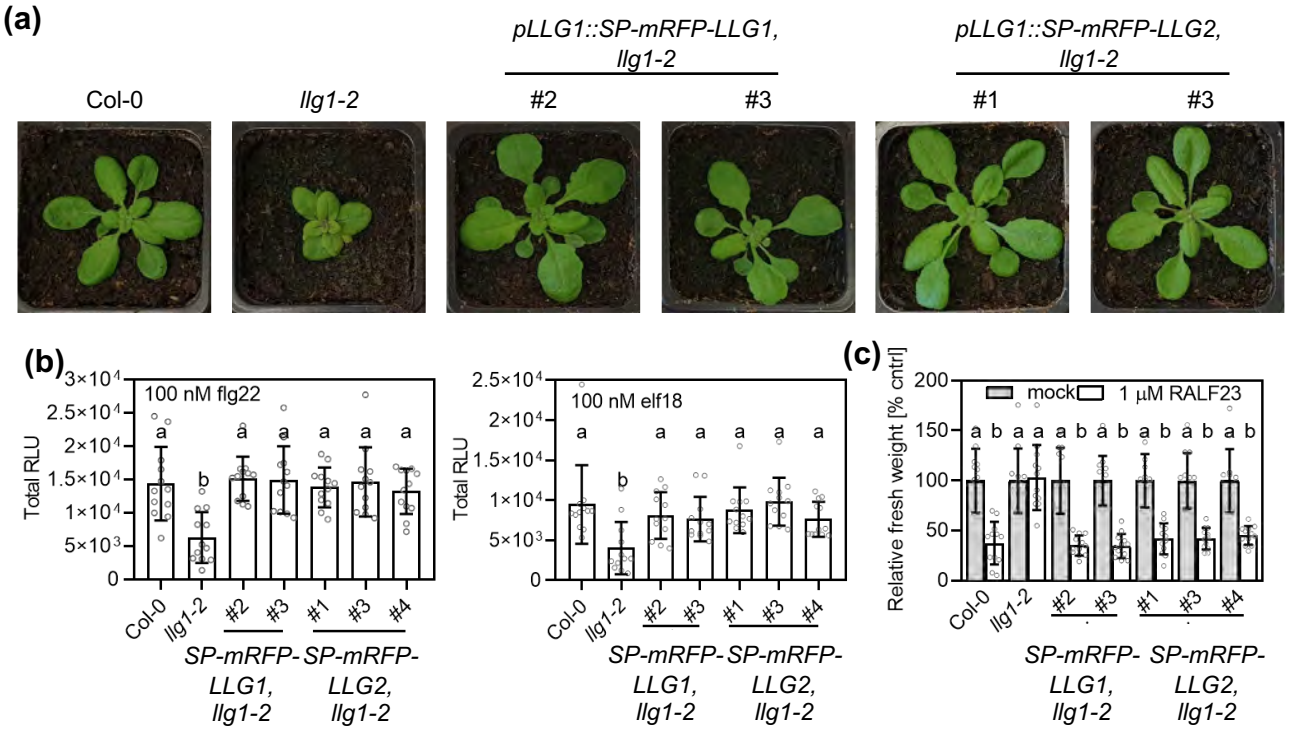

Fig 4 mRFP-LLG1 and mRFP-LLG2 complemented defects in *llg1-2*.

(a) Expression of *pLLG1::SP-mRFP-LLG*1 or *pLLG1::SP-mRFP-LLG2* restores rosette size in 4.5-week-old *llg1-2* plants.

(b) ROS production in response to flg22 (left) or elf18 (right) is restored in SP-mRFP-LLG1 or SP-mRFP-LLG2 lines driven by the *LLG1* native promoter. Total Relative Light Unit (RLU) over 40 minutes of exposure to 100 nM flg22 or elf18 treatment is displayed. Letters indicate significantly different values (n=12 leaf discs, two-way ANOVA with Tukey test, flg22 $p<0.0001$; elf18 $p=0.0009$). Error bars show ±SD.

(c) RALF23 sensitivity is restored in seedlings expressing SP-mRFP-LLG1 or SP-mRFP-LLG2. Letters indicate significantly different values (n=16 seedlings, two-way ANOVA with Tukey test, $p<0.0001$). Error bars show ±SD.

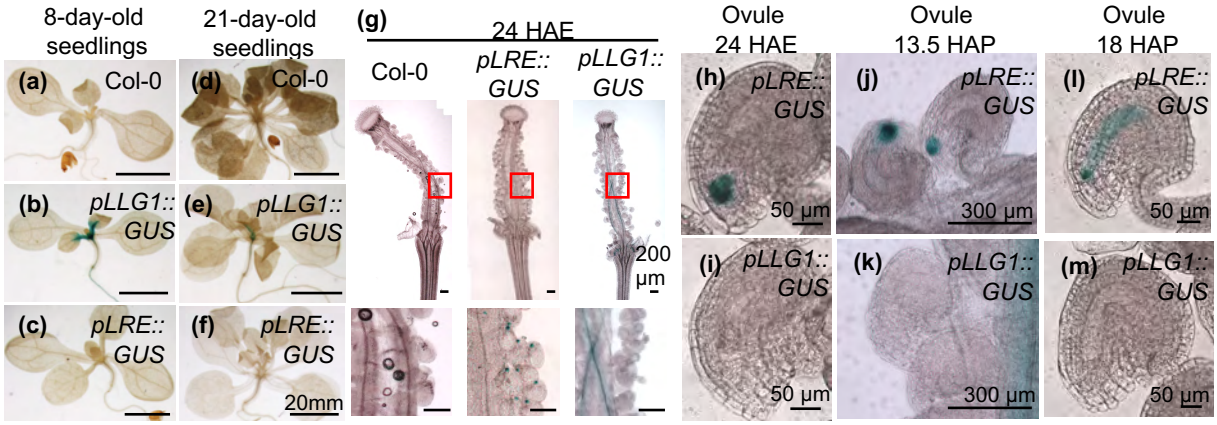

Fig 5 *pLRE::GUS* and *pLLG1::GUS* showed non-overlapping expression in vegetative and reproductive tissues.

(a-f) *pLLG1::GUS* was expressed, while *pLRE::GUS* was not, in vegetative tissues. In 8-day-old seedlings, *pLLG1::GUS* was expressed in true leaves, hypocotyls, and roots (a-c). In 21-day-old seedlings, *pLLG1::GUS* was expressed in the epicotyl, the hypocotyl, and weakly expressed in roots (d-f).

(g) At 24 HAE, *pLRE::GUS* and *pLLG1::GUS* were both expressed in pistils but in different cell-types. *pLRE::GUS* was expressed in synergid cells, while *pLLG1::GUS* was expressed in septum. Close up of area marked in red rectangles are shown below.

(h-i) *pLRE::GUS* was expressed in synergid cells at 24 hours after emasculation (HAE) (H) but *pLLG1::GUS* is not expressed in the ovule (i).

(j-m) *pLRE::GUS* and *pLLG1::GUS* showed non-overlapping expression after pollination. Mature unpollinated pistils were pollinated with Col-0 pollen and collected at 13.5 HAP (j-k) or 18 HAP (l-Mm and stained for GUS activity.

(j-k) At 13.5 HAP, *pLRE::GUS* was expressed in the micropylar end of the female gametophyte (j), while *pLLG1::GUS* continues to be expressed in the septum (k).

(l-m) At 18 HAP, *pLRE::GUS* and *pLLG1::GUS* were both expressed in pollinated pistils but in different cell-types. *pLRE::GUS* was expressed in the zygote and developing endosperm nuclei, while *pLLG1::GUS* was expressed in septum.

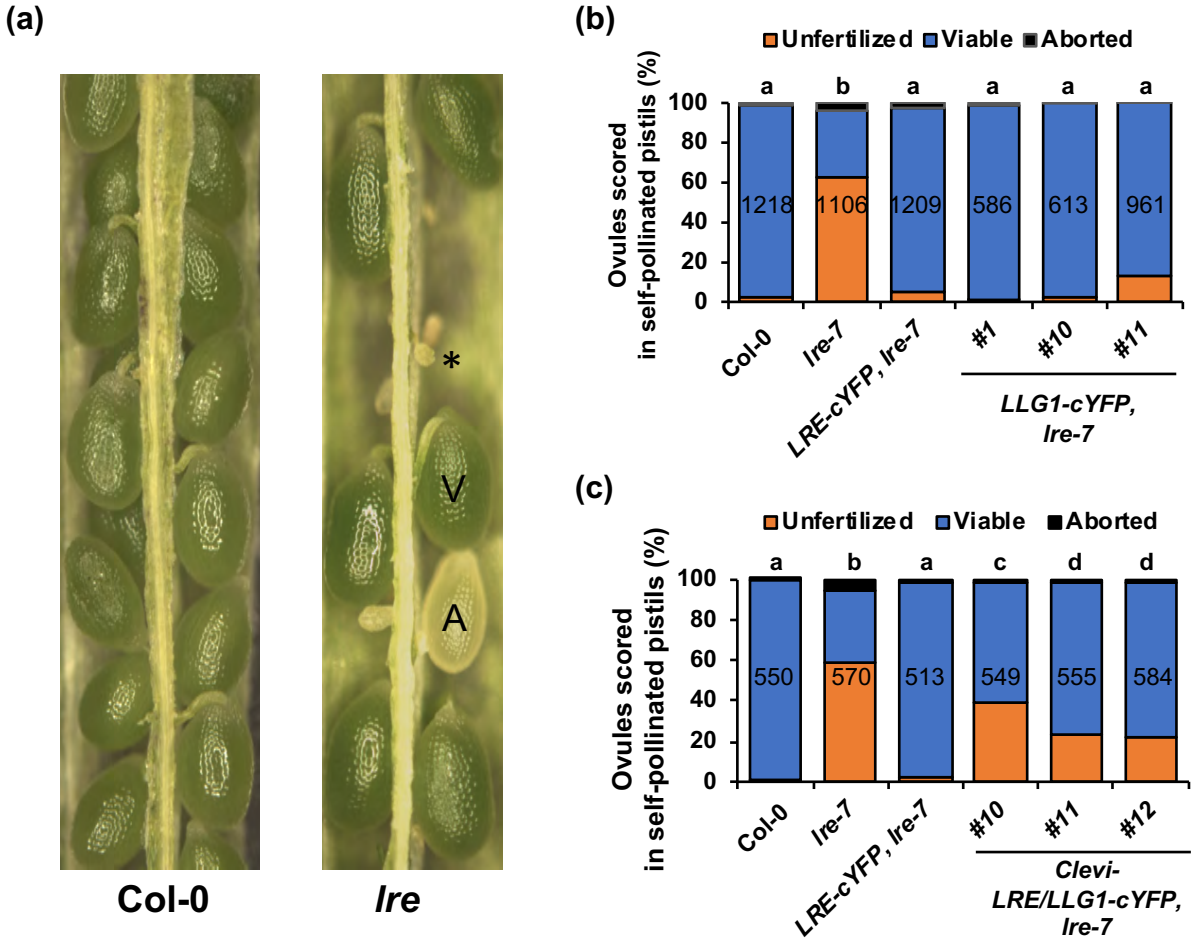

Fig 6 LLG1-cYFP and Clevi-LRE/LLG1-cYFP complemented reproductive defects in *lre*.

(a) Images of opened siliques of indicated genotypes in *A. thaliana.* A representative unfertilized ovule (*) and viable (V) or aborted (A) seed is marked in the *lre* silique.

(b) LLG1-cYFP complemented lre mutant seed set defects in self-pollinated pistils of indicated three independent transformants (ANOVA, $p = 0.18$).

(c) Clevi-LRE/LLG1-cYFP partially complemented lre mutant seed set defects in self-pollinated pistils of indicated three independent transformants (pairwise two-tailed t-tests, $p > 0.05$).

(b,c) Number in the middle of each column refers to the number of ovules/seeds scored. Groups sharing same lowercase letters are similar to each other in statistical tests.

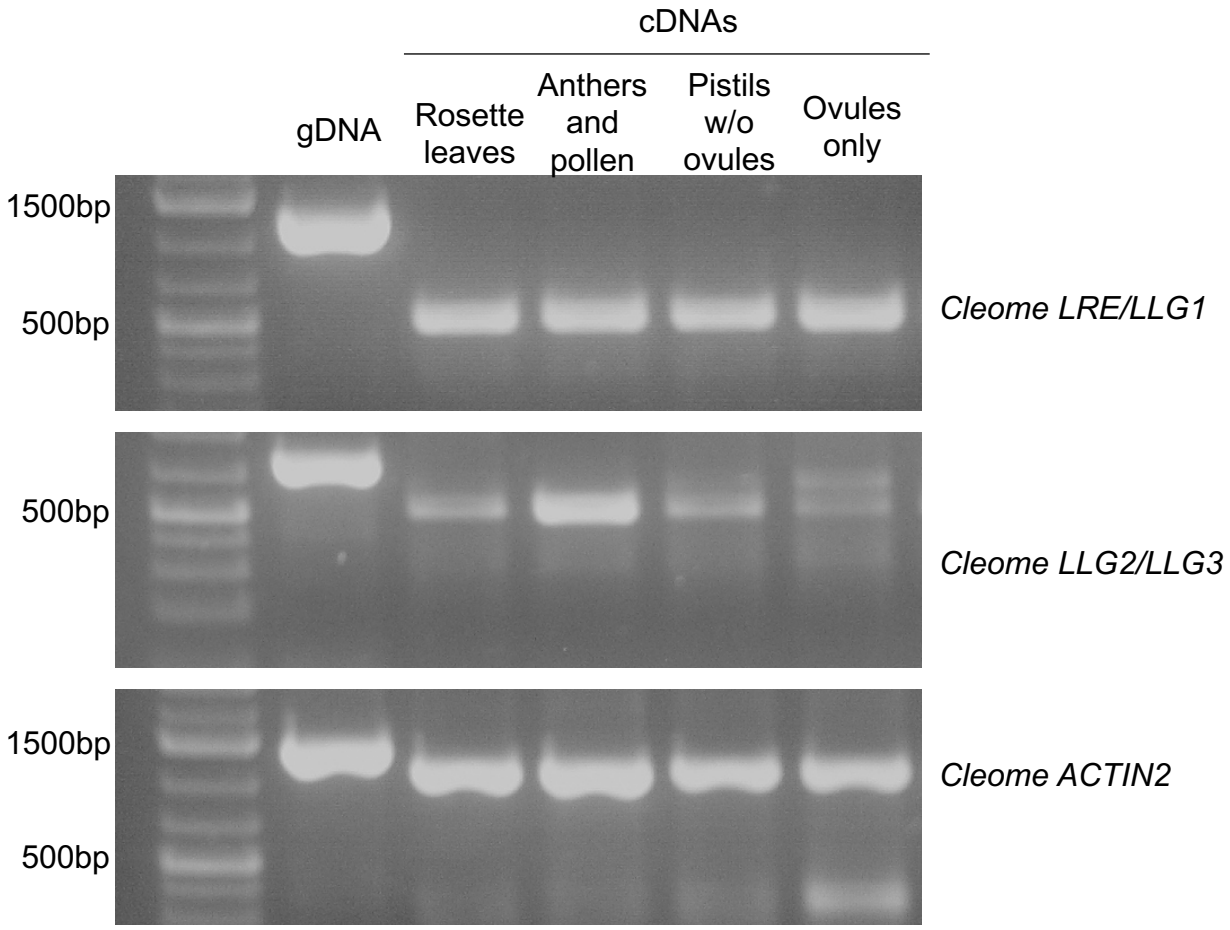

Fig 7 *Clevi-LRE/LLG1* and *Clevi-LLG2/LLG3* are expressed in vegetative and reproductive tissues of *Cleome violacea*.

RT-PCR of full-length *Clevi-LRE/LLG1* or *Clevi-LLG2/LLG3* in cDNAs isolated from 30-day-old rosette leaves, anther and pollen, emasculated pistils without ovules, and ovules from emasculated pistils of *Cleome violacea*. A homolog of *A. thaliana ACTIN2* (*Clevi-ACTIN2*) was used as a control in these experiments. gDNA, genomic DNA isolated from of Cleome leaves was used as a positive control in PCR portion of the RT-PCR experiment. RT-PCR was repeated with two additional biological replicates with similar results and the amplified bands were sequenced to confirm the identity of amplified cDNAs.