Running Title: *LLG* gene family evolution in angiosperms

Evolutionary analysis of the LORELEI gene family in angiosperms reveals regulatory subfunctionalization

Jennifer A. Noble[1], Nicholas V. Bielski[2], Ming-Che James Liu[3], Thomas A. DeFalco[4,5], Martin Stegmann[5,11], Andrew D.L. Nelson[6], Kara McNamara[3], Brooke Sullivan[3], Khanhlinh K. Dinh[3], Nicholas Khuu[3], Sarah Hancock[1], Shin-Han Shiu[7,8], Cyril Zipfel[4,5], Alice Y. Cheung[3,9,10], Mark A. Beilstein[1]* and Ravishankar Palanivelu[1]*

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

[2]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

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

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

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

[6]Boyce Thompson Institute, Cornell University, Ithaca, NY, 14853, USA

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

[8]Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824, USA

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

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

[11]Phytopathology, TUM School of Life Sciences, Technical University of Munich, Freising, Germany

*Corresponding authors: Email: mbeilstein@email.arizona.edu,

rpalaniv@email.arizona.edu

Preprint server: An earlier version of this manuscript was deposited in the bioRxiv, the preprint server for Biology:

Noble et al 2020, doi: https://doi.org/10.1101/2020.04.27.062893

Keywords: Regulatory subfunctionalization, ancestral state reconstruction, pollen tube growth, angiosperm, closed carpel, reproduction

Word count (Abstract, Introduction, Results, Discussion, and Materials and Methods): 6742

One Sentence Summary

Expression divergence played an important role in maintenance of two sub-groups of *LLG* duplicates in angiosperms

Author Contributions

JAN, NVB, 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, NVB, ADLN, and MAB performed the evolutionary and comparative transcriptome analyses. JAN, MJL, NVB, TAD, MS, ADLN, SS, CZ, AYC, MAB, and RP wrote the manuscript with input from and revisions by all authors. writing. RP and MAB agree to serve as the authors responsible for contact and communication.

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, reproduction, immunity, and stress responses in Arabidopsis. Genes encoding these proteins are members of multi-gene families in most angiosperms and could generate thousands of signaling complex variants. However, the link(s) between expansion of these gene families and the functional diversification of this critical signaling complex as well as the evolutionary factors underlying the maintenance of gene duplicates remain unknown. Here, we investigated *LLG* gene family evolution, function, and expression in angiosperms. We found that *LLGs* in monocots and eudicots are descendants of a duplication early in angiosperm evolution and that both ancient and recent *LLG* duplicates are retained. Complementation and expression analysis showed that expression divergence of *LLGs* (regulatory subfunctionalization), rather than functional divergence, explains the retention of paralogs in Brassicales. All but one extant monocot and eudicot species examined maintained an *LLG* copy with preferential expression in male reproductive tissues, with the other duplicate copies showed highest levels of expression in female or vegetative tissues. Interestingly, the single *LLG* copy in Amborella (sister to all other angiosperms) is expressed vastly higher in male compared to female reproductive or vegetative tissues. Reconstruction of expression evolution showed that the highest inferred expression levels for the single copy ancestral angiosperm *LLG* was in male reproductive tissues. We propose that expression divergence played an important role in maintenance of *LLG* duplicates in angiosperms.

Introduction

How cells communicate with each other is a fundamental question in eukaryotes. In *Arabidopsis thaliana* (hereafter, Arabidopsis), multimeric signaling complexes comprising members of *Catharanthus roseus* RECEPTOR-LIKE KINASE 1-LIKE (*Cr*RLK1L), LORELEI (LRE)-LIKE GPI-anchored protein (LLG), and RAPID ALKALINIZATION FACTOR (RALF) families influence plant fitness by regulating cell-cell communication during growth, reproduction, immunity, and stress responses (Johnson et al., 2019). *Cr*RLK1Ls are receptor kinases (RKs), LLGs are membrane-associated glycosylphosphatidylinositol-anchored proteins (GPI-Aps), and RALFs are secreted cysteine-rich signaling peptides (CRP). In Arabidopsis, the families that make up *Cr*RLK1L–LLG–RALF signaling complexes are represented by 17, 4, and 37 genes, respectively (Campbell and Turner, 2017; Abarca et al., 2021). Thus, thousands of unique combinations of such a complex may function in diverse processes, distinct tissues, and at different developmental time points. Furthermore, the extent to which such functions in Arabidopsis are conserved in other plants remains unknown (Mecchia et al., 2020). Evolutionary analysis-guided functional studies of multimeric signaling complex members offer unique opportunities to answer this question and dissect the impact of signaling at the cellular level on traits underlying fitness.

Arabidopsis *LRE* is primarily expressed in the synergid cells of the female gametophyte located within the ovules, and plays a role in reproduction (Johnson et al., 2019). Although a potential RALF ligand in this context is still unknown, LRE interacts with three C*r*RLK1Ls - FERONIA (FER), HERCULES RECEPTOR KINASE 1 (HERK1), and ANJEA (ANJ) - to form complexes at the interface of the synergid cell and the incoming pollen tube, rupture the pollen tube to release sperm cells, and effect double fertilization (Duan et al., 2014; Li et al., 2015; Liu et al., 2016; Galindo-Trigo et al., 2020). Consequently, failure to induce pollen tube reception in *fer* or *lre* single mutant pistils or in *herk1 anj* double mutant pistils, leads to ~80% reduction in seed set (Escobar-Restrepo et al., 2007; Liu et al., 2016; Galindo-Trigo et al., 2020). Elsewhere in the pistil, a variant of this complex (FER-ANJ-HERK1-LLG1-RALF33/RALF23) functions in the stigma and facilitates pollen hydration (Liu C, 2021).

Downstream effects of *Cr*RLK1L-LRE signaling complex have been characterized. LLG1 (in stigma) or LRE (in synergids) and FER initiate a Rho GTPase-mediated signaling pathway that promotes reactive oxygen species (ROS) production via NADPH oxidases of the RESPIRATORY BURST OXIDASE HOMOLOG (RBOH) family that is critical for pollen hydration in stigma and pollen tube reception in synergids (Duan et al., 2014; Liu C, 2021). After pollen tube arrival in the syngergids, FER mediates three key events in the ovule: changes in the synergid calcium profiles (Ngo et al., 2014), relocalization of NORTIA (NTA) from synergid cytoplasm to the filiform apparatus (Yuan *et al.*, 2020), and accumulation of nitric oxide in the filiform apparatus to prevent polyspermy (Duan et al., 2020).

In Arabidopsis male reproductive tissues, a variant of this multimeric signaling complex involving the *Cr*RLK1Ls ANXUR1 (ANX1), ANX2, BUDDHA'S PAPER SEAL1 (BUPS1), and BUPS2, along with RALF4/19 and LLG2/3 controls pollen tube integrity *in vivo* (Ge et al., 2017; Zhu et al., 2018; Feng et al., 2019; Ge et al., 2019). Consistent with their preferential expression in pollen grains and pollen tubes, loss-of-function mutants in members of this complex result in swollen tips, and/or premature pollen tube burst soon after germination (Feng et al., 2019; Ge et al., 2019). Additionally, RNAi mutant pollen tubes, in which both *LLG2* and *LLG3* are expressed at low levels, exhibited decreased pollen tube length *in vitro* and *in vivo* (Feng et al., 2019; Ge et al., 2019). Slowed pollen tube growth in RNAi *llg2 llg3* pollen tubes likely resulted from insufficient secretion of cell wall components, which is linked to increased accumulation of esterified and de-esterified pectin in the apical and subapical region, respectively, and decreased accumulation of callose (Feng et al., 2019). Upon activation, the ANX-based signaling pathway, coupled with pollen tube-expressed RBOHs, is important for ROS production and for maintenance of a steady tip-focused Ca^{2+} gradient that sustains pollen tube growth (Boisson-Dernier et al., 2013).

Besides reproduction, variants of the *Cr*RLK1L–LLG–RALF signaling complex mediate diverse processes in vegetative tissues, stress responses, and immunity in Arabidopsis (Duan et al., 2010; Li et al., 2015; Stegmann et al., 2017; Feng et al., 2018; Xiao et al., 2019) and the crystal structure of the ternary FER-LLG2-RALF23 complex has been

solved (Xiao et al., 2019). The FER–LLG1–RALF1 complex signals via a Rho GTPase complex to regulate ROS production and is critical for root hair growth and hypocotyl elongation (Duan et al., 2010; Li et al., 2015). Consequently, *fer* and *llg1* mutants show similar vegetative defects, which includes reduced rosette size and many root abnormalities, including insensitivity to RALF1-dependent growth inhibition, decreased sensitivity to auxin-stimulated ROS production, sensitivity to abscisic acid, and abnormal root hair development (Li et al., 2015). Although the malectin-like domains (MLDs) of CrRLK1Ls do not present conserved carbohydrate-binding residues, the extracellular domain of FER binds de-esterified pectin and this interaction is likely important for maintaining Arabidopsis root cell integrity when subjected to salt stress and restoring growth recovery after exposure to high salinity (Feng et al., 2018). A similar FER–LLG1–RALF23 complex regulates immunity by modulating the perception of pathogen-associated molecular patterns (PAMPs) such as flagellin (or the corresponding peptide epitope flg22) and elongation factor thermo unstable (EF-TU, or the corresponding peptide epitope elf18). In this context, RALF23 is recognized by its FER–LLG1 receptor complex to suppress the scaffolding function of FER, thereby inhibiting ROS production and immunity (Stegmann et al., 2017).

Other components that could interact with this *Cr*RLK1L–LLG–RALF module to form higher-order signaling complexes include EARLY NODULIN LIKE-14 (EN14) in the synergid cells and LEUCINE-RICH REPEAT EXTENSIN (LRX) proteins in the pollen tubes and vegetative tissues. EN14, which is a GPI-AP with a plastocyanin-like domain, interacts with FER and LRE and loss-of-function mutants in *EN11-EN15* contained ovules with pollen tube reception defects and ~50% reduction in seed set (Hou et al., 2016). LRX proteins bind RALF peptides and help in maintaining pollen tube integrity (Mecchia et al., 2017; Moussu et al., 2020). LRXs also associate with FER in vegetative tissues, where they regulate immunity (Gronnier et al., 2020), salt stress tolerance (Zhao et al., 2018), as well as root hair growth (Herger et al., 2020) and vacuolar expansion (Dunser et al., 2019).

Phylogenetic analysis of each component of the *Cr*RLK1L–LLG–RALF multimeric complex is essential to elucidate the links between cell-cell communication and the

evolution of traits that influence plant fitness. Phylogenetic analysis of 1050 *CrRLK1L* proteins in 57 plants showed that the C*r*RLK1L subfamily has undergone several duplication events during its evolution and that expression of C*r*RLK1Ls in legumes and Arabidopsis were conserved (Solis-Miranda et al., 2020). In another study, the phylogenetic relationship among *Cr*RLK1L genes identified retention of duplicates in Arabidopsis, moss, rice, medicago, and tomato and that their evolution pointed to functional conservation (Yang et al., 2021). Phylogenetic analysis of 795 RALFs from 51 land plants showed that evolution of the RALF family was distinct in eudicots and monocots and that numerous RALF genes in Arabidopsis resulted from a rapid expansion in the Brassicaceae (Campbell and Turner, 2017). However, a comprehensive evolutionary analysis paired with functional studies and expression profiling has not been performed for any of the *Cr*RLK1L–LLG–RALF signaling complex members. Consequently, links between expansion of these gene families and the functional diversification of this critical signaling complex and evolutionary factors underlying the maintenance of gene duplicates remains unknown.

In this study, we addressed maintenance of duplicate members of the *LLG* family by inferring phylogeny and characterizing patterns of functional and regulatory evolution to investigate if ancestral functions are shared between duplicated copies (subfunctionalization) or if the ancestral expression pattern is divided between the duplicates, leading to non-overlapping expression of each paralog (regulatory subfunctionalization) (Ohno, 1970; Force et al., 1999; Des Marais and Rausher, 2008). We identified homologs of the four-member Arabidopsis *LLG* gene family and showed that they are conserved throughout angiosperms. We used complementation assays and found that the molecular functions of GPI-APs encoded by the *LLG* gene family are conserved. In contrast, expression of *LLG* family members in representative angiosperms is highly divergent showing that regulatory divergence (i.e., differences in gene expression), rather than functional divergence, likely contributed to the retention of *LLG* family members in angiosperms. In all but one of the extant monocots and eudicots species, we found that at least one *LLG* copy is preferentially expressed in male reproductive tissues, with the other paralog(s) showing highest levels of expression in

female or vegetative tissues. Interestingly, the single *LLG* copy in the angiosperm *Amborella trichopoda* (hereafter, Amborella) is expressed higher in male reproductive tissues compared to female or vegetative tissues. Additionally, likelihood reconstruction of ancestral expression across plant tissues revealed that elevated *LLG* expression in pollen and pollen tube is ancient. Taken, together, our analyses suggest that the *Cr*RLK1L-LLG-RALF signaling complex likely plays a role in pollen tube growth regulation across angiosperms.

Results

Maintenance of duplicated LLG gene families is widespread in angiosperms

To understand the evolutionary history of the four *LLGs* in Arabidopsis (Capron *et al.*, 2008), we inferred phylogeny for *LLG* homologs recovered from angiosperm genomes. Our motif-driven homology searches (Bailey and Gribskov, 1998) and subsequent filtering returned 114 putative *LLG* family members across 32 focal genomes (Dataset S1). We rooted the tree resulting from our maximum likelihood search with the single *LLG* from Amborella, a putative sister species to all other extant angiosperms (The Angiosperm Phylogeny et al., 2016). The tree indicated that *LLGs* in monocots and eudicots experienced independent duplication events after the two lineages diverged (Figure 1). For this reason, we alphabetized homologs recovered in monocots (*LLGA-H,* Figure 1) to reflect their independent evolution and distinguish them from eudicot *LLGs*, which are numbered. Our tree indicated that *LLGs* in monocots likely underwent several lineage-specific duplication events, including one that resulted in two *LLG* clades in grasses (*LLGA* and *LLGB*) (Figure 1). Importantly, all the monocot genomes included here retained duplicate *LLGs*.

Eudicot *LLG* sequences in our tree descended from a common eudicot ancestor (Figure 1, 88% bootstrap support) that underwent an early duplication event yielding the *LRE/LLG1* and *LLG2/LLG3* copies present in extant eudicots. Except for *Aquilegia coerulea*, which retains only an *LLG2/LLG3* ortholog, sampled eudicots retained both descendants of this early duplication. In Brassicaceae, a subsequent family-wide duplication and retention yielded four *LLGs*. In Brassicaceae, *LRE* and its orthologs

formed a monophyletic group sister to the clade containing all *LLG1* orthologs. Similarly, *LLG2* and its orthologs formed a monophyletic group sister to the clade containing all *LLG3* orthologs (Figure 1). In sum, the evolutionary history of the *LLG* family is characterized by the retention of both ancient and recent duplicates in angiosperms.

LLG molecular functions are conserved in Arabidopsis and members of Cleomaceae

The retention of duplicate genes typically results from evolutionary mechanisms that can produce signatures of selection and divergence in promoter regions, coding sequence, or both (Force et al., 1999; Li et al., 2005; Duarte et al., 2006). To test if retention of the *LRE/LLG1* and *LLG2/LLG3* copies present in extant eudicots resulted from changes in coding sequence thereby yielding distinct functions after an early duplication event in their evolution, we examined *LLG*s in Arabidopsis, where characterized mutants made functional analysis feasible (Li et al., 2015; Liu et al., 2016; Ge et al., 2019). We expressed all Arabidopsis *LLG* coding sequences from the Arabidopsis *LLG1* promoter and tested if they could complement the vegetative development defects of *llg1-2* mutants (Li et al., 2015). Like *LLG1,* expression of Arabidopsis *LRE* or *LLG3* from the *LLG1* promoter complemented root hair defects (Figure 2a-d), hypocotyl length in dark-grown *llg1-2* seedlings (Figure 2e), epidermal pavement cell defects (Figures 3a-3e), and RALF1 sensitivity in *llg1-2* seedlings (Figures 3g and S1).

Expression of Arabidopsis *LLG2* also complemented vegetative defects in *llg1-2* mutants, as rosette size was restored to levels observed with *LLG1* (Figure 4a). In addition to vegetative phenotypes, *llg1* mutants are defective in immune responses, as they show reduced responsiveness to several PAMPs, including the bacterial elicitors elf18 and flg22 (Shen et al., 2017; Stegmann et al., 2017; Xiao et al., 2019). In *llg1-2* plants expressing *LLG2* from the *LLG1* promoter, we found that responses to elf18 and flg22 were restored to levels comparable to those expressing *LLG1* (Figures 4b and S2). Similarly, complementation of *llg1-2* with *LLG2* restored responsiveness to

exogenous RALF23 peptide to levels detected in *llg1-2* plants expressing *LLG1* (Figure 4c). These results demonstrated that LRE, LLG2, or LLG3 can substitute for LLG1 and perform its molecular functions. Results of these complementation experiments are particularly significant, as *LLG2* and *LLG3* are members of a lineage that is distinct from *LLG1* and originated from an ancient duplication event early in the eudicot evolution (Figure1).

We also examined if molecular functions of *LRE* in female reproductive tissues are conserved. For this, we expressed Arabidopsis *LLG1* from the promoter of its closest paralog (Arabidopsis *LRE*) and tested if it could complement the female reproductive defects in the *lre* mutant (Liu et al., 2016). Like LRE fused to citrine YFP (LRE-cYFP) (Liu et al., 2016), LLG1-cYFP localized in the synergid cells (Figure S3) and restored seed set defects in *lre-7* to levels detected in wild type or LRE-cYFP plants based on fertility (Figures 4d and 4e) and transmission assays (Table S1).

To examine conservation of functions in LLGs in other plants besides those in Brassicaceae, we explored *LLGs* in the sister family of Cleomaceae. Unlike Brassicaceae, there are only two single copy orthologs in the Cleomaceae – *LRE/LLG1* and *LLG2/LLG3* (Figure 1). Since LRE and LLG1 can reciprocally complement each of their functions in the vegetative and reproductive tissues of Arabidopsis, respectively, we tested if the molecular functions performed by Arabidopsis LRE and LLG1 are conserved in the LRE/LLG1 single copy ortholog from *Cleome violacea* (Clevi-LRE/LLG1). We expressed *Clevi-LRE/LLG1-cYFP* from the Arabidopsis *LLG1* or *LRE* promoter (Figure 2f) in the *llg1-2* or *lre-7* mutant, respectively, and performed complementation experiments. We found that Clevi-LRE/LLG1 complemented root hair defects (Figures 2b and 2d), hypocotyl lengths in dark-grown *llg1-2* seedlings (Figures 2c and 2e), epidermal cell defects (Figure 3f), and RALF1 sensitivity in *llg1-2* seedlings (Figures 3g and S1) to levels seen in *llg-1-2* lines expressing LLG1. Additionally, Clevi-LRE-cYFP expressed from the *LRE* promoter localized in the synergid cells (Figure S3) and partially restored female reproductive defects in *lre-7* based on fertility (Figure 4f) and transmission assays (Table S2). These results indicated that the *C. violacea* single-copy LRE/LLG1 ortholog complements the functions of both Arabidopsis LRE and

LLG1. Results from complementation experiments led us to conclude that the molecular functions of LLGs are highly conserved and thus changes to the coding regions of these genes (LRE shares 58%, 55%, and 46% amino acid identity with LLG1, LLG3, and LLG2, respectively (Capron et al., 2008), were likely not responsible for the post-duplication retention of the eudicot copies.

LLG gene expression is divergent in Arabidopsis and members of Cleomaceae

We next considered if expression divergence among *LLG* paralogs underlies their retention. Consistent with this possibility, members of the *LLG* gene family in Arabidopsis are differentially expressed (Tsukamoto et al., 2010; Feng et al., 2019; Ge et al., 2019; Xiao et al., 2019). *LLG2* and *LLG3* are primarily expressed in male reproductive tissues (Feng et al., 2019; Ge et al., 2019; Xiao et al., 2019) while *LRE* is primarily expressed in female reproductive tissues and seedlings (Tsukamoto et al., 2010; Wang et al., 2017). Unlike the preferential expression of these three *LLGs* in reproductive tissues, *LLG1* is ubiquitously expressed throughout plant development (Tsukamoto et al., 2010; Feng et al., 2019; Ge et al., 2019; Xiao et al., 2019). However, based on RT-PCR and RNA-seq experiments, there is an overlap of *LRE* and *LLG1* expression in seedlings and unpollinated and pollinated ovaries (Tsukamoto et al., 2010; Ge et al., 2019). To test if expression of these two genes overlapped in cells of these complex tissues, we examined cell-specific β-glucuronidase (GUS) expression in three *pAtLRE::GUS* lines (this study), and one previously reported *pAtLLG1::GUS* line (Li et al., 2015). GUS expression analyses revealed no cell type in seedlings, unpollinated pistils, or pollinated pistils where *LRE* and *LLG1* are co-expressed (Figure 5), demonstrating that *LLG1* and *LRE* indeed have different and non-overlapping expression profiles.

We also explored *LLG* expression divergence in plants other than Arabidopsis by examining published RNA-seq based transcriptome atlases of species in Cleomaceae. Indeed, we found that the single copy *LRE/LLG1* orthologs in *Tarenaya hassleriana* and *Gynandropsis gynandra* were expressed in all developmental stages (Kulahoglu et al.,

2014) and encompassed all the expression domains of Arabidopsis *LRE* + *LLG1* (Table S3). In contrast, the single copy *LLG2/LLG3* ortholog in both plants was expressed nearly exclusively in male reproductive tissues (Table S3). These results showed that *LLG* gene expression is divergent in Arabidopsis and members of Cleomaceae, in striking contrast with the conservation of functions among *LLGs.*

Regulatory subfunctionalization may underlie retention of duplicated LLGs in angiosperms

To examine how widespread expression divergence is among *LLG* family members in other angiosperms, we analyzed published RNA-seq-based transcriptome data (see Methods) and calculated the normalized expression levels of *LLGs* across five expression domains (male and female reproductive tissues, vegetative tissues, and sporophytic reproductive tissues before and after fertilization) (Table S4). We analyzed expression in ten angiosperms for which we recovered *LLG* homologs – four eudicots (Arabidopsis*, T. hassleriana*, *Solanum lycopersicum*, and *Sesamum indicum*) and six monocots (*Oryza sativa, Zea mays, Sorghum bicolor, Brachypodium distachyon, Phoenix* dactylifera, and *Annanas comosus*). We found that expression patterns have diverged post duplication in both eudicots and monocots (Table S4).

Eudicots are characterized by an early duplication that yielded two reciprocally monophyletic clades of *LLGs*: *LRE/LLG1* and *LLG2/LLG3*. Members of the *LRE/LLG1* clade primarily showed the highest level of expression in female and vegetative tissues, while members of the *LLG2/LLG3* clade are expressed at their highest level and almost exclusively restricted to male reproductive tissues (Table S4). In contrast, expression divergence occurred within the *LRE/LLG1* clade. For example, *Solanum lycopersicum* retains two *LRE/LLG1* paralogs; one is expressed in female reproductive tissues and sporophytic tissues before fertilization, while the other is most highly expressed in vegetative tissues (Table S4).

Like our findings in eudicots, duplication in the sampled grass monocots yielded one copy with the highest expression in male reproductive tissues and another copy that

was expressed more broadly. For example, the *LLGB* clade members exhibited male reproductive tissue expression. However, members of the sister clade, the grass specific *LLGA*, underwent additional duplication events and expansion of expression across domains. In *O. sativa*, transcription of *LLGA-1* was at its highest levels in the female domain and reproductive tissues before fertilization, while the domain of highest expression for *LLGA-2* was in vegetative tissues (Table S4). Thus, the pollen- and pollen tube-preferential expression patterns observed for Arabidopsis *LLG2* and *LLG3* appeared to mirror that of *LLG2/3* orthologs in other eudicots and *LLGB* copies in grass monocots (Table S4).

Copies of *LLGs* from non-grass monocots (*P. dactylifera* and *A. comosus*) were derived from a combination of older and more recent duplication events (Figure 1). For *P. dactylifera* the pattern of expression divergence observed in other species holds; *LLGH* is restricted to male reproductive tissues and *LLGE-1 and 2* are more broadly expressed, including in female reproductive and vegetative tissues (Table S4). Thus, our analyses revealed that preferential expression of at least one *LLG* copy in male reproductive tissues has independently evolved in eudicots and in monocots (Table S4). Still, in contrast to all other species, *A. comosus* appears to lack a male reproductive tissue specific *LLG;* however, *A. comosus LLGG* is expressed at much higher levels than other *A. comosus LLG* paralogs in both male and female domains (Table S4). Hence, it is possible that *LLGG* in *A. comosus* is sufficient to perform the functions required in the male domain.

To obtain additional evidence in support of regulatory subfunctionalization of *LLG* expression in angiosperms, we examined *LLG* expression in Amborella, which is sister to all other extant angiosperms and contained only one *LLG.* In Amborella, the expression pattern of the sole *LLG* present (Figure 1) encompasses the expression domains of the other sampled *LLGs*. Interestingly, this single *LLG* showed vastly higher levels of expression in the male reproductive tissues compared to expression in the female or vegetative tissues (Table S4). Taken together, the expression analyses in 11 angiosperms showed that expansion of *LLG* gene family in angiosperms was

accompanied by divergence in their expression and that regulatory subfunctionalization likely underlies the retention of duplicated *LLGs* in angiosperms.

The inferred ancestral expression pattern of single copy angiosperm ancestor LLG is primarily in pollen and pollen tubes

Expression divergence among *LLGs* yielding copies that are primarily expressed in male reproductive tissues, while other gene copies are expressed in female reproductive tissues or more broadly raises the question: what was the ancestral expression pattern of *LLGs* in the earliest angiosperms? This question is also relevant given that the single copy *LLG* in Amborella showed vastly higher levels of expression in the male reproductive tissues. To better understand the evolution of divergent expression in *LLG* paralogs, we calculated normalized expression values (z-scores; Table S4) across the five expression domains for each *LLG* and used these values to reconstruct likelihood-based ancestral character states (Figures 6a and S4, Datasets S2 and S3). This expression reconstruction showed that a single copy *LLG* ancestor, with highest expression in male reproductive tissues, underwent independent expansions in monocots and eudicots.

Male reproductive tissue comprises pollen grain, pollen tube, and sperm cells. To pinpoint relative expression level of *LLG*s within the male reproductive domain, we divided the male domain into pollen, pollen tubes, and sperm and then recalculated expression z-scores for each *LLG* present in five representative angiosperms from which expression data were available – two eudicots (Arabidopsis and *S. lycopersicum*), two monocots (*O. sativa* and *Z. mays*), and Amborella (Figure 6b, Dataset S4). These analyses revealed that expression is most pronounced in mature pollen and pollen tubes rather than sperm (Figure 6b, Dataset S4). Ancestral reconstruction of the expression in these tissues indicated that the single copy angiosperm *LLG* ancestor was likely highly expressed in pollen and pollen tube and expressed at low levels in sperm (Figures 6c–e, Datasets S2 and S3), pointing to critical functions for the ancestral *LLG*, and the signaling complex it was a part of, in male reproductive tissues.

Discussion

Regulatory Divergence likely led to the Retention of LLGs in Angiosperms

Multiple mechanisms have been proposed to explain the retention of duplicated genes, as progressive degeneration of one member of a paralogous set of genes is the default outcome (non-functionalization; (Panchy et al., 2016). Duplicated genes may acquire new functions (neo-functionalization) and hence may be retained (Ohno, 1970). Another mechanism of retention is sub-functionalization via sequence variations in the protein-coding or regulatory regions of the genes, by which the ancestral functions could be shared between the duplicated copies, resulting in non-overlapping expression of each paralog (Force et al., 1999). The advent of high-throughput sequencing and large-scale transcriptomic studies have allowed evaluation of expression divergence in duplicated genes in *A. thaliana* (Duarte et al., 2006; Zou et al., 2009; Liu et al., 2011; De Smet et al., 2017), *Glycine max* (Roulin et al., 2013), and *Zea mays* (Hughes et al., 2014). These studies used expression analysis to generate critical evidence in support of regulatory sub-functionalization and/or regulatory neo-functionalization. Here, we performed the first phylogenetic analysis of *LLGs* in angiosperms coupled with expression and functional studies of *LLG* gene family to show that regulatory subfunctionalization is likely the basis of retention of *LLG* duplicates.

In many organisms, there is a positive correlation between expression divergence and *cis*-regulatory motif divergence among duplicate genes (Li et al., 2005), including Arabidopsis (Arsovski et al., 2015). Consistent with this, we identified considerable variation among putative transcription factor binding sites in the putative promoter regions of *LLGs* (Figure S5). Additionally, in support of their shared expression in male reproductive tissues, specific transcription factor binding sites are conserved in the promoter regions of *LLG2* or *LLG3* of Arabidopsis and the single copy ortholog *LLG2/LLG3* in *C. violacea* (Figure S5). Similarly, a different set of transcription factor binding sites are conserved in the promoter regions of *LLG1* or *LRE* of Arabidopsis and

the single copy ortholog *LRE/LLG1* in *C. violacea*, which show expression in female reproductive and vegetative tissues (Figure S5). Therefore, a conserved set of *cis*-regulatory elements is likely responsible for these divergent expression patterns in both Brassicaceae and Cleomaceae and ultimately contributed to the retention of duplicates in eudicots.

Preferential expression of at least one *LLG* copy in male reproductive tissues and highest expression of the single copy *LLG* in Amborella in the male domain raises the possibility of transcription factor-binding motifs that are conserved among the putative promoter regions of these genes. Using a set of *in vitro* binding data (see Methods), there was a clear pattern of presence of motifs in the *LLG2/3* clade in eudicots*,* as they shared AP2 (APETALA2) and EREBP (ethylene-responsive element binding proteins) transcription factor binding sites (Chandler and Werr, 2020) with the single copy Amborella *LLG* (Dataset S5). These differences are consistent with the distinct evolutionary history and divergence between the *LLGs* in eudicots and monocots. It is likely that the monocots have lost these motifs since their divergence from eudicots. In addition, it is likely that Arabidopsis *LLG2* and *Sesame indicum LRE/LLG1-2* have also lost these motifs compared to other orthologs (Dataset S5). Still, there is not one motif that is completely conserved among these genes (Dataset S5) and perhaps uncharacterized novel motifs are shared by these male tissue-expressed genes.

The LLG Gene Family may be Co-evolving with the CrRLK1L and RALF Gene Families

Like *LLGs,* genes encoding the other two members of the trimeric complex (*Cr*RLK1Ls and RALFs), also show expression divergence in Arabidopsis (Campbell and Turner, 2017; Ge et al., 2017) and given our results, it is possible that they show similar divisions of expression domains in other angiosperms. The large size of *RALF* and *CrRLK1L* gene families pose a challenge in efforts to study all the possible combinations of the signaling complex (Kessler et al., 2015; Sharma et al., 2016). However, phylogenetic analysis coupled with functional and expression studies, like that

performed in this study, may offer a viable approach to understand the evolution of expression domains in *CrRLK1L* and *RALF* families in angiosperms.

FER functions with both LRE and LLG1, and correspondingly, *FER* is expressed in both *LRE* and *LLG1* expression domains (Tsukamoto et al., 2010; Li et al., 2015). Given that specific *Cr*RLK1Ls, RALFs, and LLGs play distinct biological roles, *CrRLK1Ls* and *RALFs* may have co-evolved with members of the *LLG* family to perform their functions in different tissues. *In vitro* interactions of RALF23 with LLG1, LLG2, and LLG3 but not LRE provide support for this possibility (Xiao et al., 2019). Methods such as evolutionary rate covariation have been used to link co-evolution with functional associations (Findlay et al., 2014). Such methods, in combination with the phylogenetics, expression analyses, and molecular genetic assays used in this study will prove invaluable in further characterizing members of this critical signaling complex.

The inferred ancestral Expression Pattern in Angiosperms Suggests an Important Role for LLG in the Evolution of Rapid Pollen Tube Growth During Angiosperm Reproduction

Results of our ancestral *LLG* expression reconstruction suggest that a signaling complex containing the single copy *LLG* may have mediated pollen tube growth in the earliest angiosperms. Moreover, the evolution of a faster reproductive cycle is often posited as a key innovation of early angiosperms (Stebbins, 1974). Several traits are believed to influence this increase in reproductive rate, including reduced male and female gametophytes and closed carpels. Moreover, evolution of faster pollen tube growth has been identified as a key link between the reduction in gametophyte size and closed carpels (Williams, 2008; Williams, 2012). Faster pollen tube growth in angiosperms is hypothesized to be facilitated by the accumulation of esterified pectins in pollen tube tips and deesterified pectins and callose in their walls (Williams, 2008). Consistent with this hypothesis, Arabidopsis *llg2 and llg3* mutant pollen tubes showed decreased length *in vivo* and *in vitro,* increased accumulation of esterified and deesterified pectins, and decreased levels of callose (Feng et al., 2019). Interestingly, RNA-seq-based transcriptome analyses of early land plants (Julca et al., 2021) revealed

male-preferential expression of one of the three *LLG* homologs from the moss *Physcomitrella patens*, while maximal expression of both *LLG* homologs in the liverwort *Marchantia polymorpha* is in antheridia (Figure S6). Thus, it is possible that proto angiosperms co-opted an existing male expressed molecular signaling module, present across land plants, to regulate pollen tube growth through the carpel, thereby allowing relatively rapid delivery of sperm to the female gametophyte and setting the stage for the evolution of the suite of traits that unite all angiosperms.

Materials and Methods

Homolog Identification

Putative orthologs and paralogs from species representing major lineages of angiosperms were retrieved by first identifying conserved motifs in the four coding sequences (CDS) of *LLG* family members in the Arabidopsis genome. Homolgos were searched using CDS obtained from Phytozome v. 13.0 (Goodstein et al., 2012) or from the Database Resources of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) in MEME Suite v. 5.1.1 (Bailey, 1994) (Dataset S1). CDS of putative orthologs for *Cleome violacea* and *Tarenaya hassleriana* were obtained through the Comparative Genomics (CoGe) Platform using CoGe BLAST (tBLASTx) with *A. thaliana LRE* or *LLG1* nucleotide CDS as query (CoGe version (Altschul et al., 1990; Lyons and Freeling, 2008; Lyons et al., 2008)). Motif results and CDS were input into MAST (Bailey and Gribskov, 1998) under default parameters and only those with an E-value ≤ 1e-05 were retained as putative *LLG*s. Putative *LLGs* were edited to the longest possible ORF.

Alignment and Phylogenetic Analyses

Based on preliminary alignments made using MUSCLE v. 3.8.425 (Edgar, 2004) in Geneious 2019.2.1 (https://www.geneious.com) under default parameters, we inferred maximum likelihood trees with RAxML-HPC MPI AVX2 v. 8.2.11 (Stamatakis, 2014). We used these preliminary trees to further assess the homology of sequences returned via motif searching and eliminated sequences whose homology could not be confidently

assigned due to phylogenetic position, branch length, and/or increase to the length of the alignment.

The full amino acid sequences of putative *LLG* family members were aligned in MUSCLE v. 3.8.425 (Edgar, 2004). Using the Arabidopsis *LLG* gene structures as guides (Liu et al., 2016), the initial alignment was divided into signal peptide, main chain, and GPI anchor region. The alignment for each region was corrected by eye before being concatenated into a full-length amino acid alignment and submitted to TranslatorX (Abascal et al., 2010) along with the FASTA nucleotide sequences to obtain an amino acid corrected nucleotide alignment. The full alignment was partitioned and subjected to maximum likelihood phylogenetic analyses in RAxML-HPC MPI AVX2 v. 8.2.11 (Stamatakis, 2014) under the GTR+GAMMA model of sequence evolution. Support for branches of the tree was assessed from 1000 bootstrap replicates of the rapid bootstrapping algorithm in RaxML.

Plant Growth

Arabidopsis (Columbia ecotype, Col-0) and *C. violacea* seeds were sterilized with 50% bleach and plated on 1/2X Murashigi and Skoog (MS) plates (Carolina Biological Supply Co.), with 1% sucrose for seedling growth assays and with corresponding antibiotics for transmission assays. After stratification (3 days in dark and 4 °C), plated seeds were incubated at 21 °C with continuous light (75–100 $\mu mol \cdot m^{-2} \cdot s^{-1}$). Ten- to fourteen-day-old seedlings were transplanted to soil and grown in 16 h light (100–120 $\mu mol \cdot m^{-2} \cdot s^{-1}$) at 21 °C and 8 h dark at 18 °C.

Cloning Transgenic Constructs

The *pLLG1::LRE-HA-GPI* and *pLLG1::LLG3* transgenic constructs were prepared using the primers and DNA templates (Table S4), and with the Arabidopsis *LLG1* promoter region (~2 kb upstream of ATG) (Li et al., 2015). For SP-mRFP-LLG1 expression *in planta*, the Arabidopsis *LLG1* promoter was amplified from seedling genomic DNA using primers "ccaagcttgcatgccGTCGTTGTCCCAGATTCGTCG" and "gatctagagtcgaccGGTTCTTTGTTGGTTACAGGAGAAGTCAC" for subsequent cloning into the Gateway destination vector pGWB1 (Nakagawa *et al.*, 2007a) using Infusion

cloning (Takara Bio) and the SdaI/SbfI restriction enzyme (Thermo Scientific). The *SP-mRFP-LLG1* construct was synthesized with attached attB1/attB2 sites (Twist Bioscience) and subsequently recombined into pDONR-Zeo using BP II Clonase (Invitrogen). The resulting pDONR-Zeo SP-mRFP-LLG1 construct was recombined with *pGWB1-pLLG1*. The *pAtLLG1::SP-mRFP-LLG2* construct was synthesized (Thermo Scientific) with attB1/attB2 sites added and cloned into *pDONR-Zeo* via BP II Clonase (Invitrogen) and was subsequently recombined with *pGWB601* (Nakagawa et al., 2007) for *in planta* expression.

The *pAtLRE::LLG1-cYFP, pAtLRE::GUS*, *pAtLRE::Clevi-LRE/LLG1-cYFP,* and *pAtLLG1::Clevi-LRE/LLG1-cYFP* transgenic lines were created by replacing the genomic sequence of *LRE* in the *pAtLRE::LRE-cYFP* construct with the desired gene and/or promoter (Liu et al., 2016). PCR amplified products (PrimeSTAR GXL DNA Polymerase, TaKaRa Bio Inc.) using primers and DNA templates listed in Table S4 were cloned into *pLRE::LRE-cYFP* plasmid linearized with *SpeI-HF* and *AscI* by using the In-Fusion HD Cloning Plus system (Clontech). Transformants in Col-0 background (*pLRE::GUS*) or *lre-7* mutant (*pLRE::LRE-cYFP*, *pLRE::LLG1-cYFP*, and *pLRE::Clevi-LRE/LLG1-cYFP*) were generated and single-locus insertion lines were isolated (Liu et al., 2016).

The *pLLG1::SP-mRFP-LLG1* and *pLLG1::SP-mRFP-LLG2* lines were transformed into *llg1-2* plants by floral dip (Clough and Bent, 1998), and transformants were selected on 1x MS-agar with 1% sucrose, supplemented with 25 µg/mL kanamycin or 10 µg/mL glufosinate ammonium, respectively. Seedlings for seedling growth inhibition 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 assays were grown in 10 h light (150 $\mu mol \cdot m^{-2} \cdot s^{-1}$) at 20 °C.

cYFP and Seed Set Scoring

cYFP expression in mature unpollinated pistils were scored and imaged (Liu *et al.*, 2016). Unfertilized, viable (enlarged after fertilization), and aborted ovules in siliques (10 days after pollination) were scored (Noble and Palanivelu, 2020). Three to five self-

pollinated siliques located between 5th and 15th siliques from the bottom of the main branch of an Arabidopsis plant were scored.

Root Hair and Hypocotyl Length Phenotypes in llg1 Seedlings

Root hair analysis was performed (Duan et al., 2010). Root hairs located between 1.5 and 3.5 mm from the primary root tip of four-day-old seedlings were observed with a stereoscope. The number of normal and defective (stunted and collapsed) root hairs was scored. The hypocotyl length assay was performed (Li et al., 2015). Three-day-old dark-grown seedlings were imaged using Epson Perfection V370 Photo Scanner at 600 dpi resolution and the length of hypocotyl was measured with Image J.

Propidium Iodide Staining of Pavement Cells

Six-day-old seedlings were stained with 50 µg/mL propidium iodide for 20 min and washed with ddH_2O. Images were acquired using confocal microscopy on a NIKON A1 Spectral System (excitation wavelength 561 nm and emission wavelength 595 nm) and analyzed by the NIS-Elements AR Analysis Software (V 5.02).

RALF1-induced Root Growth Inhibition Assays

The root sensitivity assays with RALF1 treatment were performed (Haruta et al., 2014; Li et al., 2015): three-day-old light-grown seedlings were treated with RALF1 for 2 days at concentrations indicated in Figures 3 and S2. Growth during treatment was determined by measuring primary root length using Image J at the beginning and end of treatments.

RALF23-induced Seedling Growth Inhibition Assays

Five-day-old T2 *pLLG1::SP-mRFP-LLG1* and *pLLG1::SP-mRFP-LLG2* T2 seedlings were transferred to liquid 1x MS with 1% sucrose in sterile 48-well plates, with or without 1 µM RALF23 peptide (Scilight) (Xiao et al., 2019). Seedling fresh weight was measured after 7 days of growth in liquid medium.

ROS Burst Measurements

ROS burst measurements were performed (Xiao et al., 2019). Twelve 4-mm leaf discs from 4.5-week-old plants were harvested and equilibrated overnight in 96-well plates in sterile, deionized water. The next day, water was replaced with PAMP solution (100 nM flg22 or elf18 peptide plus 10 µg/mL horseradish peroxidase and 100 µM luminol). ROS was immediately measured using a charge-coupled device camera (Photek). Total ROS production was calculated as the sum of Random Light Units value over 40 minutes of PAMP treatment.

GUS Staining

Tissues were stained for GUS activity (Johnson and Kost, 2010), mounted in 50% glycerol, observed on a Zeiss Axiovert 100 epifluorescence microscope, and imaged using Metamorph (Version 7; Molecular Devices). For each tissue and genotype, at least five samples were observed.

Transcription Factor Binding Site Analysis

DAP-seq data were obtained from the Plant Cistrome Database (http://neomorph.salk.edu/dev/pages/shhuang/dap_web/pages/index.php) in the MEME motif file format. These 838 motifs were scanned against the putative promoter sequences (1 kb upstream of the transcriptional start, not overlapping with neighboring genes) of the *LLG* paralogs with MAST (Bailey and Gribskov, 1998) using default parameters. The scanned results were used to determine the presence of motif sites (mapping p-value was <1e-3). The motif site information was combined for all motifs to generate a motif presence (1)/absence (0) matrix for hierarchical clustering using the heatmap.2 function in R.

Expression Analyses

TPM (transcript per kilobase million) values from CoNekT (https://evorepro.sbs.ntu.edu.sg/) for Arabidopsis, *S. lycopersicum*, *Z. mays*, *O. sativa*, and Amborella (Mutwil et al., 2011; Ruprecht et al., 2017; Julca et al., 2021). For *LLGs*

(Dataset S1) in *S. bicolor* and *B. distachyon* (Davidson *et al.*, 2012), TPM values were obtained from the EMBL-EBI expression database (https://www.ebi.ac.uk/gxa/home). FPKM (fragments per kilobase million) expression values reported for *T. hassleriana* (Kulahoglu et al., 2014) and *A. comosus* (Wang et al., 2020) were converted to TPM by dividing the FPKM value for each gene by the sum of FPKM values for all genes, then multiplying that number by one million. For *S. indicum* and *P. dactylifera* raw RNA-seq data were retrieved from the NCBI Sequence Read Archive (SRA - https://www.ncbi.nlm.nih.gov/sra; Bioprojects PRJNA74261 and PRJNA472694, respectively). Both retrieval and mapping were performed in the CyVerse Discovery Environment (Merchant et al., 2016) using the RMTA pipeline v. 2.6.3 (Peri et al., 2019) under default parameters. Read count data for focal genes were then extracted and converted to TPM. We defined five expression domains based on available RNA-seq data: male, female, reproductive before and after fertilization, and vegetative tissues. We binned sampled tissues with TPM values into one of these five categories, calculated the mean TPM of each expression domain, and then z-transformed these values in R v. 4.0.2 (R Core Team, 2020). Expression within the male domain was obtained by calculating expression z-scores for pollen, pollen tube, and sperm for species in which these tissues were available in RNA-seq experiments.

Ancestral Character State Reconstruction

We used the z-score of expression for each *LRE/LLG* sequence along with our phylogeny, pruned to only those family members from which expression data were available, to infer the ancestral level of relative expression at nodes in the tree. Ancestral state reconstruction was performed in R v. 4.0.2 (R Core Team, 2020). We used both Brownian Motion (BM) and Ornstein-Uhlenbeck (OU) models in phytools v0.7-47 (50) and then compared their log-likelihoods via Akaike information criterion (AIC) (Akaike, 1974) and AICc (Mazerolle, M. J. (2019), R package version 2.1-1. Retrieved from https://CRAN.R-project.org/package=AICcmodavg) (Dataset S2). The latter employs a correction and may be more accurate for data sets of the size we analyzed.

Image Processing

ImageJ was used to assemble image panels and insert scale bars.

Acknowledgements

We acknowledge Dr. Patrick Edger from Michigan State University for providing us access to the *Cleome violacea* genome on CoGe and Dr. Jocelyn Hall from the University of Alberta for providing *Cleome violacea* seeds. We thank Dr. Ramin Yadegari (University of Arizona) for the Zeiss Axiophot microscope. We thank Dr. Mark Johnson, Brown University and members of the Palanivelu Lab for insightful discussions. JAN was supported by the following: IGERT Comparative Genomics Program at the University of Arizona (Award ID: 0654435); NSF Graduate Research Fellowship: Grant DGE-1143953; the Boynton Graduate Fellowship; and the University of Arizona Graduate College Office of Diversity and Inclusion. NVB was supported by a University of Arizona Graduate College University Fellowship and an NSF grant to MAB (IOS-1546825). Additional support for this work was provided by an NSF grant to RP (IOS-1146090) and University of Arizona Undergraduate Biology Research Program fellowship to SH (private donors). This work was also supported by the Gatsby Charitable Foundation (C.Z.), the University of Zürich (C.Z.), the European Research Council under the European Union (EU)'s Horizon 2020 research and innovation programme (grant agreement No 773153, project 'IMMUNO-PEPTALK'), 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 AYC lab was supported by Natural Science Foundation (IOS-1645854 and MCB-1715764) to AYC 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 HMW for the contribution of LLG1 and LLG3 constructs. KM was supported by the Torrey Summer

Research Scholarship from UMass Plant Biology Program and the Linda Slakey Summer Research Scholarship from UMass BMB department. We also acknowledge funding from NSF pollen RCN grant (MCB0955910) for sponsoring activities and meetings that forged collaborations between AYC and RP labs. Work in the ADLN lab was supported by the National Science Foundation (PGRP-IOS 2023310 and PGRP-IOS 1758532). Work in the Shiu lab was partly supported by the National Science Foundation (IOS-1546617, DEB-1655386) and U.S. Department of Energy (Great Lakes Bioenergy Research Center BER DE-SC0018409).

References

Abarca A, Franck CM, Zipfel C (2021) Family-wide evaluation of RAPID ALKALINIZATION FACTOR peptides. Plant Physiology

Abascal F, Zardoya R, Telford MJ (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res **38:** W7-13

Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control **19:** 716-723

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403-410

Arsovski AA, Pradinuk J, Guo XQ, Wang S, Adams KL (2015) Evolution of Cis-Regulatory Elements and Regulatory Networks in Duplicated Genes of Arabidopsis. Plant Physiol **169:** 2982-2991

Bailey TL, and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *In* Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. , Vol 2, pp 28–36

Bailey TL, Gribskov M (1998) Methods and statistics for combining motif match scores. J Comput Biol **5:** 211-221

Boisson-Dernier A, Lituiev DS, Nestorova A, Franck CM, Thirugnanarajah S, Grossniklaus U (2013) ANXUR receptor-like kinases coordinate cell wall integrity with growth at the pollen tube tip via NADPH oxidases. PLoS Biol **11:** e1001719

Campbell L, Turner SR (2017) A Comprehensive Analysis of RALF Proteins in Green Plants Suggests There Are Two Distinct Functional Groups. Front Plant Sci **8:** 37

Capron A, Gourgues M, Neiva LS, Faure JE, Berger F, Pagnussat G, Krishnan A, Alvarez-Mejia C, Vielle-Calzada JP, Lee YR, Liu B, Sundaresan V (2008) Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene. Plant Cell **20:** 3038-3049

Chandler JW, Werr W (2020) A phylogenetically conserved APETALA2/ETHYLENE RESPONSE FACTOR, ERF12, regulates Arabidopsis floral development. Plant Mol Biol **102:** 39-54

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant Journal **16:** 735-743

De Smet R, Sabaghian E, Li Z, Saeys Y, Van de Peer Y (2017) Coordinated Functional Divergence of Genes after Genome Duplication in Arabidopsis thaliana. The Plant Cell **29:** 2786

Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature **454:** 762-765

Duan Q, Kita D, Johnson EA, Aggarwal M, Gates L, Wu HM, Cheung AY (2014) Reactive oxygen species mediate pollen tube rupture to release sperm for fertilization in Arabidopsis. Nat Commun **5:** 3129

Duan Q, Kita D, Li C, Cheung AY, Wu HM (2010) FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development. Proc Natl Acad Sci U S A **107:** 17821-17826

Duan Q, Liu M-CJ, Kita D, Jordan SS, Yeh F-LJ, Yvon R, Carpenter H, Federico AN, Garcia-Valencia LE, Eyles SJ, Wang C-S, Wu H-M, Cheung AY (2020) FERONIA controls pectin- and nitric oxide-mediated male–female interaction. Nature

Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol **23:** 469-478

Dunser K, Gupta S, Herger A, Feraru MI, Ringli C, Kleine-Vehn J (2019) Extracellular matrix sensing by FERONIA and Leucine-Rich Repeat Extensins controls vacuolar expansion during cellular elongation in Arabidopsis thaliana. EMBO J **38**

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:** 113

Escobar-Restrepo JM, Huck N, Kessler S, Gagliardini V, Gheyselinck J, Yang WC, Grossniklaus U (2007) The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. Science **317:** 656-660

Feng H, Liu C, Fu R, Zhang M, Li H, Shen L, Wei Q, Sun X, Xu L, Ni B, Li C (2019) LORELEI-LIKE GPI-ANCHORED PROTEINS 2/3 Regulate Pollen Tube Growth as Chaperones and Coreceptors for ANXUR/BUPS Receptor Kinases in Arabidopsis. Mol Plant **12:** 1612-1623

Feng W, Kita D, Peaucelle A, Cartwright HN, Doan V, Duan Q, Liu MC, Maman J, Steinhorst L, Schmitz-Thom I, Yvon R, Kudla J, Wu HM, Cheung AY, Dinneny JR (2018) The FERONIA Receptor Kinase Maintains Cell-Wall Integrity during Salt Stress through Ca(2+) Signaling. Curr Biol **28:** 666-675 e665

Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF (2014) Evolutionary Rate Covariation Identifies New Members of a Protein Network Required for Drosophila melanogaster Female Post-Mating Responses. PLoS Genet **10:** e1004108

Force A, Lynch M, Pickett FB, Amores A, Yan Y-l, Postlethwait J (1999) Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics **151:** 1531-1545

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531-1545

Galindo-Trigo S, Blanco-Tourinan N, DeFalco TA, Wells ES, Gray JE, Zipfel C, Smith LM (2020) CrRLK1L receptor-like kinases HERK1 and ANJEA are female determinants of pollen tube reception. EMBO Rep **21:** e48466

Ge Z, Bergonci T, Zhao Y, Zou Y, Du S, Liu MC, Luo X, Ruan H, Garcia-Valencia LE, Zhong S, Hou S, Huang Q, Lai L, Moura DS, Gu H, Dong J, Wu HM, Dresselhaus T, Xiao J, Cheung AY, Qu LJ (2017) Arabidopsis pollen tube integrity and sperm release are regulated by RALF-mediated signaling. Science **358:** 1596-1600

Ge Z, Zhao Y, Liu MC, Zhou LZ, Wang L, Zhong S, Hou S, Jiang J, Liu T, Huang Q, Xiao J, Gu H, Wu HM, Dong J, Dresselhaus T, Cheung AY, Qu LJ (2019) LLG2/3 Are Co-receptors in BUPS/ANX-RALF Signaling to Regulate Arabidopsis Pollen Tube Integrity. Curr Biol **29:** 3256-3265 e3255

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40:** D1178-1186

Gronnier J, Franck CM, Stegmann M, DeFalco TA, Cifuentes AA, Dünser K, Lin W, Yang Z, Kleine-Vehn J, Ringli C, Zipfel C (2020) FERONIA regulates FLS2 plasma membrane nanoscale dynamics to modulate plant immune signaling. bioRxiv**:** 2020.2007.2020.212233

Haruta M, Sabat G, Stecker K, Minkoff BB, Sussman MR (2014) A peptide hormone and its receptor protein kinase regulate plant cell expansion. Science **343:** 408-411

Herger A, Gupta S, Kadler G, Franck CM, Boisson-Dernier A, Ringli C (2020) Overlapping functions and protein-protein interactions of LRR-extensins in Arabidopsis. PLoS Genet **16:** e1008847

Hou Y, Guo X, Cyprys P, Zhang Y, Bleckmann A, Cai L, Huang Q, Luo Y, Gu H, Dresselhaus T, Dong J, Qu LJ (2016) Maternal ENODLs Are Required for Pollen Tube Reception in Arabidopsis. Curr Biol **26:** 2343-2350

Hughes TE, Langdale JA, Kelly S (2014) The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. Genome Research **24:** 1348-1355

Johnson MA, Harper JF, Palanivelu R (2019) A Fruitful Journey: Pollen Tube Navigation from Germination to Fertilization. Annu Rev Plant Biol **70:** 809-837

Johnson MA, Kost B (2010) Pollen tube development. Methods Mol Biol **655:** 155-176

Julca I, Ferrari C, Flores-Tornero M, Proost S, Lindner AC, Hackenberg D, Steinbachova L, Michaelidis C, Gomes Pereira S, Misra CS, Kawashima T, Borg M, Berger F, Goldberg J, Johnson M, Honys D, Twell D, Sprunck S, Dresselhaus T, Becker JD, Mutwil M (2021) Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. Nat Plants

Kessler SA, Lindner H, Jones DS, Grossniklaus U (2015) Functional analysis of related CrRLK1L receptor-like kinases in pollen tube reception. EMBO Rep **16:** 107-115

Kulahoglu C, Denton AK, Sommer M, Mass J, Schliesky S, Wrobel TJ, Berckmans B, Gongora-Castillo E, Buell CR, Simon R, De Veylder L, Brautigam A, Weber AP (2014) Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C3 and C4 plant species. Plant Cell **26:** 3243-3260

Li C, Yeh F-L, Cheung AY, Duan Q, Kita D, Liu M-C, Maman J, Luu EJ, Wu BW, Gates L, Jalal M, Kwong A, Carpenter H, Wu H-M (2015) Glycosylphosphatidylinositol-anchored proteins as chaperones and co-receptors for FERONIA receptor kinase signaling in Arabidopsis. eLife **4:** e06587

Li C, Yeh FL, Cheung AY, Duan Q, Kita D, Liu MC, Maman J, Luu EJ, Wu BW, Gates L, Jalal M, Kwong A, Carpenter H, Wu HM (2015) Glycosylphosphatidylinositol-anchored proteins as chaperones and co-receptors for FERONIA receptor kinase signaling in Arabidopsis. Elife **4**

Li WH, Yang J, Gu X (2005) Expression divergence between duplicate genes. Trends Genet **21:** 602-607

Liu C SL, Yu, Xiao Y, Vyshedsky D*, Peng C, Sun X, Liu Z, Cheng L, Zhang H, Han Z, Chai J, Wu H-M*, Cheung AY, Li C. (2021) Pollen PCP-B peptides unlock a stigma peptide-receptor kinase gated mechanism for pollination. Science **Accepted (March 1, 2021)**

Liu S-L, Baute GJ, Adams KL (2011) Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in Arabidopsis thaliana. Genome biology and evolution **3:** 1419-1436

Liu X, Castro C, Wang Y, Noble J, Ponvert N, Bundy M, Hoel C, Shpak E, Palanivelu R (2016) The Role of LORELEI in Pollen Tube Reception at the Interface of the Synergid Cell and Pollen Tube Requires the Modified Eight-Cysteine Motif and the Receptor-Like Kinase FERONIA. Plant Cell **28:** 1035-1052

Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J **53:** 661-673

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D, Freeling M (2008) Finding and comparing syntenic

regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol **148:** 1772-1781

Mecchia MA, Rövekamp M, Giraldo-Fonseca A, Meier D, Gadient P, Bowman JL, Grossniklaus U (2020) Characterization of the single FERONIA homolog in Marchantia polymorpha reveals an ancestral role of CrRLK1L receptor kinases in regulating cell expansion and morphological integrity. bioRxiv**:** 2020.2012.2023.424085

Mecchia MA, Santos-Fernandez G, Duss NN, Somoza SC, Boisson-Dernier A, Gagliardini V, Martinez-Bernardini A, Fabrice TN, Ringli C, Muschietti JP, Grossniklaus U (2017) RALF4/19 peptides interact with LRX proteins to control pollen tube growth in Arabidopsis. Science **358:** 1600-1603

Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, Antin P (2016) The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. PLoS Biol **14:** e1002342

Moussu S, Broyart C, Santos-Fernandez G, Augustin S, Wehrle S, Grossniklaus U, Santiago J (2020) Structural basis for recognition of RALF peptides by LRX proteins during pollen tube growth. Proc Natl Acad Sci U S A **117:** 7494-7503

Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, Fernie AR, Usadel B, Nikoloski Z, Persson S (2011) PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. Plant Cell **23:** 895-910

Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, Watanabe Y, Nakamura K, Kimura T, Ishiguro S (2007) Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci Biotechnol Biochem **71:** 2095-2100

Ngo QA, Vogler H, Lituiev DS, Nestorova A, Grossniklaus U (2014) A calcium dialog mediated by the FERONIA signal transduction pathway controls plant sperm delivery. Dev Cell **29:** 491-500

Noble JA, Palanivelu R (2020) Workflow to Characterize Mutants with Reproductive Defects. Methods Mol Biol **2160:** 109-128

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag

Panchy N, Lehti-Shiu M, Shiu S-H (2016) Evolution of Gene Duplication in Plants. Plant Physiology **171:** 2294

Peri S, Roberts S, Kreko IR, McHan LB, Naron A, Ram A, Murphy RL, Lyons E, Gregory BD, Devisetty UK, Nelson ADL (2019) Read Mapping and Transcript Assembly: A Scalable and High-Throughput Workflow for the Processing and Analysis of Ribonucleic Acid Sequencing Data. Front Genet **10:** 1361

Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA (2013) The fate of duplicated genes in a polyploid plant genome. The Plant Journal **73:** 143-153

Ruprecht C, Proost S, Hernandez-Coronado M, Ortiz-Ramirez C, Lang D, Rensing SA, Becker JD, Vandepoele K, Mutwil M (2017) Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. Plant J **90:** 447-465

Sharma A, Hussain A, Mun BG, Imran QM, Falak N, Lee SU, Kim JY, Hong JK, Loake GJ, Ali A, Yun BW (2016) Comprehensive analysis of plant rapid alkalization factor (RALF) genes. Plant Physiol Biochem **106:** 82-90

Shen Q, Bourdais G, Pan H, Robatzek S, Tang D (2017) Arabidopsis glycosylphosphatidylinositol-anchored protein LLG1 associates with and modulates FLS2 to regulate innate immunity. Proc Natl Acad Sci U S A **114:** 5749-5754

Solis-Miranda J, Fonseca-García C, Nava N, Pacheco R, Quinto C (2020) Genome-Wide Identification of the CrRLK1L Subfamily and Comparative Analysis of Its Role in the Legume-Rhizobia Symbiosis. Genes **11:** 793

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30:** 1312-1313

Stebbins GL (1974) Flowering plants: evolution above the species level. Harvard University Press, Cambridge, Massachusetts

Stegmann M, Monaghan J, Smakowska-Luzan E, Rovenich H, Lehner A, Holton N, Belkhadir Y, Zipfel C (2017) The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. Science **355:** 287-289

The Angiosperm Phylogeny G, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, Stevens PF (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society **181:** 1-20

Tsukamoto T, Qin Y, Huang Y, Dunatunga D, Palanivelu R (2010) A role for LORELEI, a putative glycosylphosphatidylinositol-anchored protein, in Arabidopsis thaliana double fertilization and early seed development. Plant J **62:** 571-588

Wang L, Li Y, Jin X, Liu L, Dai X, Liu Y, Zhao L, Zheng P, Wang X, Liu Y, Lin D, Qin Y (2020) Floral transcriptomes reveal gene networks in pineapple floral growth and fruit development. Commun Biol **3:** 500

Wang Y, Tsukamoto T, Noble JA, Liu X, Mosher RA, Palanivelu R (2017) Arabidopsis LORELEI, a Maternally Expressed Imprinted Gene, Promotes Early Seed Development. Plant Physiol **175:** 758-773

Williams J (2012) Pollen tube growth rates and the diversification of flowering plant reproductive cycles. INternational Journal of Plant Sciences **173:** 649-661

Williams JH (2008) Novelties of the flowering plant pollen tube underlie diversification of a key life history stage. Proc Natl Acad Sci U S A **105:** 11259-11263

Xiao Y, Stegmann M, Han Z, DeFalco TA, Parys K, Xu L, Belkhadir Y, Zipfel C, Chai J (2019) Mechanisms of RALF peptide perception by a heterotypic receptor complex. Nature **572:** 270-274

Yang H, Wang D, Guo L, Pan H, Yvon R, Garman S, Wu H-M, Cheung AY (2021) Malectin/Malectin-like domain-containing proteins: A repertoire of cell surface molecules with broad functional potential. The Cell Surface **7:** 100056

Zhao C, Zayed O, Yu Z, Jiang W, Zhu P, Hsu CC, Zhang L, Tao WA, Lozano-Duran R, Zhu JK (2018) Leucine-rich repeat extensin proteins regulate plant salt tolerance in Arabidopsis. Proc Natl Acad Sci U S A **115:** 13123-13128

Zhu L, Chu LC, Liang Y, Zhang XQ, Chen LQ, Ye (2018) The Arabidopsis CrRLK1L protein kinases BUPS1 and BUPS2 are required for normal growth of pollen tubes in the pistil. Plant J **95:** 474-486

Zou C, Lehti-Shiu MD, Thomashow M, Shiu S-H (2009) Evolution of Stress-Regulated Gene Expression in Duplicate Genes of Arabidopsis thaliana. PLOS Genetics **5:** e1000581

Figure Legends

Figure 1. Full LLG family phylogeny across angiosperms. Maximum likelihood tree (ln(L) = -33601.508911) inferred using RAxML (RAxMLHPC MPI AVX2 v. 8.2.11) on a partitioned nucleotide alignment. Support for nodes was assessed using the rapid bootstrap algorithm (1000 replicates). The tree is rooted with the single copy Amborella *LLG*. Grey stars indicate gene duplication events. White stars indicate duplications that may be impacted by phylogenetic uncertainty. Monocots and eudicots experienced independent duplication events. LLGs duplicated early in eudicot evolution resulting in reciprocally monophyletic *LRE/LLG1* and *LLG2/LLG3* clades. In Brassicaceae, a subsequent set of duplication events yielded the four copies present in *Arabidopsis* (indicated in green).

Figure 2. LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complemented vegetative defects in *llg1-2* seedlings.
(a) Diagrams of the *pLLG1::LRE-HA, pLLG1::LLG3,* and *pLLG1::Clevi-LRE/LLG1-cYFP* constructs. (b) Root hair length defects were restored when LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP were expressed in *llg1-2* seedlings to wild-type (Col-0) levels. Images were taken from T2 seedlings of single insertion lines. (c) LRE-HA, LLG3, and Clevi-LRE/LLG1-cYFP complementation of short hypocotyl length phenotype in *llg1-2* seedlings. Representative images of wild-type (Col-0) or T2 seedlings of single-locus insertion lines of *pLLG1::LRE-HA, pLLG1::LLG3,* or *pLLG1::Clevi-LRE/LLG1-cYFP* are shown. (d) Quantification of complementation of root hair defects in *llg1-2* seedlings shown in B are reported here. 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. (e) Quantification of complementation of de-etiolated hypocotyl in dark grown *llg1-2* mutant seedlings shown in C are reported here. 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.

Figure 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* seedling cotyledons 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 Figure S2 Root length was measured two days after RALF1 treatment, and three trials were performed. Error bars represent ±SD.

Figure 4. mRFP-LLG1 and mRFP-LLG2 complemented defects in *llg1-2* and LLG1-cYFP and Clevi-LRE/LLG1-cYFP complemented reproductive defects in *lre.* (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) Growth sensitivity to RALF23 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. (d) Images of opened siliques of indicated genotypes in *Arabidopsis.* A representative unfertilized ovule (*) and viable (V) or aborted (A) seed is marked in the *lre* silique. (e) LLG1-cYFP complemented *lre* mutant seed set defects in self-pollinated pistils of indicated three independent transformants (ANOVA, $p = 0.18$). (f) 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, $p > 0.05$). (e, f) Number in the middle of each column refers to the number of ovules/seeds scored. Groups sharing the same lowercase letters are like seach other in statistical tests.

Figure 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 hours after emasculation (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 the area marked in red rectangles are shown below. (h-i) *pLRE::GUS* was expressed in synergid cells at 24 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 hours after pollination (HAP) (j-k) or 18 HAP (l-m) 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.

Figure 6. Phylogeny and expression analysis of *LLG* orthologs in angiosperms.

(a) Maximum likelihood phylogenetic tree from putative ortholog CDS. The heatmap (right) displays the z-score expression of the orthologs in male, female, vegetative, reproductive tissues before fertilization (RBF), and reproductive tissues after fertilization (RAF). Ancestral state reconstruction was performed in phytools under either a Brownian motion or Ornstein-Uhlenbeck model. Reconstruction results of select nodes are displayed on the adjacent branches of the tree. Nodes with missing expression values are the result of missing expression data at the tips, precluding robust inference of the ancestral state at that node. (b) Heatmap of core male tissues for *LLGs* from a subset of species. The tissues shown are mature pollen, pollen tube (PT), and sperm respectively. (c-e) Ancestral state reconstruction trees for pollen, pollen tube, and sperm

respectively. Tip expression and node expression are colored according to the z-score scales presented.

Arabidopsis thaliana LRE
Arabidopsis lyrata LRE
Capsella rubella LRE
Capsella grandiflora LRE
Camelina sativa LRE-1
Camelina sativa LRE-2
Camelina sativa LRE-3
Cardamine hirsuta LRE
Eutrema salsugineum LRE-2
Eutrema salsugineum LRE-1
Brassica rapa LRE
Arabidopsis lyrata LLG1
Arabidopsis thaliana LLG1
Cardamine hirsuta LLG1-1
Cardamine hirsuta LLG1-2
Brassica rapa LLG1-1
Brassica rapa LLG1-2
Eutrema salsugineum LLG1
Camelina sativa LLG1-3
Camelina sativa LLG1-2
Camelina sativa LLG1-1
Capsella grandiflora LLG1
Capsella rubella LLG1
Tarenaya hassleriana LRE/LLG1
Cleome violacea LRE/LLG1
Carica papaya LRE/LLG1
Citrus clementina LRE/LLG1-1
Citrus clementina LRE/LLG1-2
Gossypium raimondii LRE/LLG1-1
Theobroma cacao LRE/LLG1
Gossypium raimondii LRE/LLG1-2
Populus trichocarpa LRE/LLG1-1
Populus trichocarpa LRE/LLG1-2
Fragaria vesca LRE/LLG1
Glycine max LRE/LLG1-3
Glycine max LRE/LLG1-4
Glycine max LRE/LLG1-2
Glycine max LRE/LLG1-1
Sesame indicum LRE/LLG1-1
Sesame indicum LRE/LLG1-2
Helianthus annuus LRE/LLG1-4
Helianthus annuus LRE/LLG1-1
Helianthus annuus LRE/LLG1-3
Helianthus annuus LRE/LLG1-2
Daucus carota LRE/LLG1-2
Daucus carota LRE/LLG1-1
Solanum lycopersicum LRE/LLG1-2
Solanum lycopersicum LRE/LLG1-1
Eudicot Lorelei / LLG1
Arabidopsis lyrata LLG2
Arabidopsis thaliana LLG2
Eutrema salsugineum LLG2
Camelina sativa LLG2-1
Camelina sativa LLG2-2
Capsella grandiflora LLG2
Capsella rubella LLG2
Cardamine hirsuta LLG2
Cardamine hirsuta LLG3
Eutrema salsugineum LLG3
Brassica rapa LLG3-1
Brassica rapa LLG3-3
Brassica rapa LLG3-2
Arabidopsis lyrata LLG3
Arabidopsis thaliana LLG3
Camelina sativa LLG3-1
Camelina sativa LLG3-2
Camelina sativa LLG3-3
Capsella rubella LLG3
Capsella grandiflora LLG3
Cleome violacea LLG2/LLG3
Tarenaya hassleriana LLG2/LLG3
Carica papaya LLG2/3
Sesame indicum LLG2/LLG3
Helianthus annuus LLG/LLG3
Solanum lycopersicum LLG2/LLG3
Daucus carota LLG2/LLG3-2
Daucus carota LLG2/LLG3-1
Citrus clementina LLG2/LLG3
Gossypium raimondii LLG2/LLG3
Theobroma cacao LLG2/LLG3
Fragaria vesca LLG2/LLG3
Glycine max LLG3/LLG3-1
Glycine max LLG2/LLG3-2
Aquilegia coerulea LLG2/LLG3
Populus trichocarpa LLG2/LLG3-3
Populus trichocarpa LLG2/LLG3-2
Populus trichocarpa LLG2/LLG3-1
Eudicot LLG2 / LLG3
Zea mays LLGA-1
Zea mays LLGA-2
Sorghum bicolor LLGA
Setaria italica LLGA-1
Setaria italica LLGA-2
Oryza sativa LLGA-1
Brachypodium distachyon LLGA
Oryza sativa LLGA-2
Zea mays LLGB
Sorghum bicolor LLGB
Setaria italica LLGB
Brachypodium distachyon LLGB
Oryza sativa LLGB
Ananas comosus LLGC-1
Ananas comosus LLGC-2
Ananas comosus LLGC-3
Musa acuminata LLGD-4
Musa acuminata LLGD-3
Musa acuminata LLGD-1
Musa acuminata LLGD-2
Phoenix dactylifera LLGE-1
Phoenix dactylifera LLGE-2
Asparagus officinalis LLGF-2
Asparagus officinalis LLGF-3
Asparagus officinalis LLGF-1
Ananas comosus LLGG
Phoenix dactylifera LLGH
Amborella trichopoda LLG
Monocot LLGA – H
0.3

Figure 1

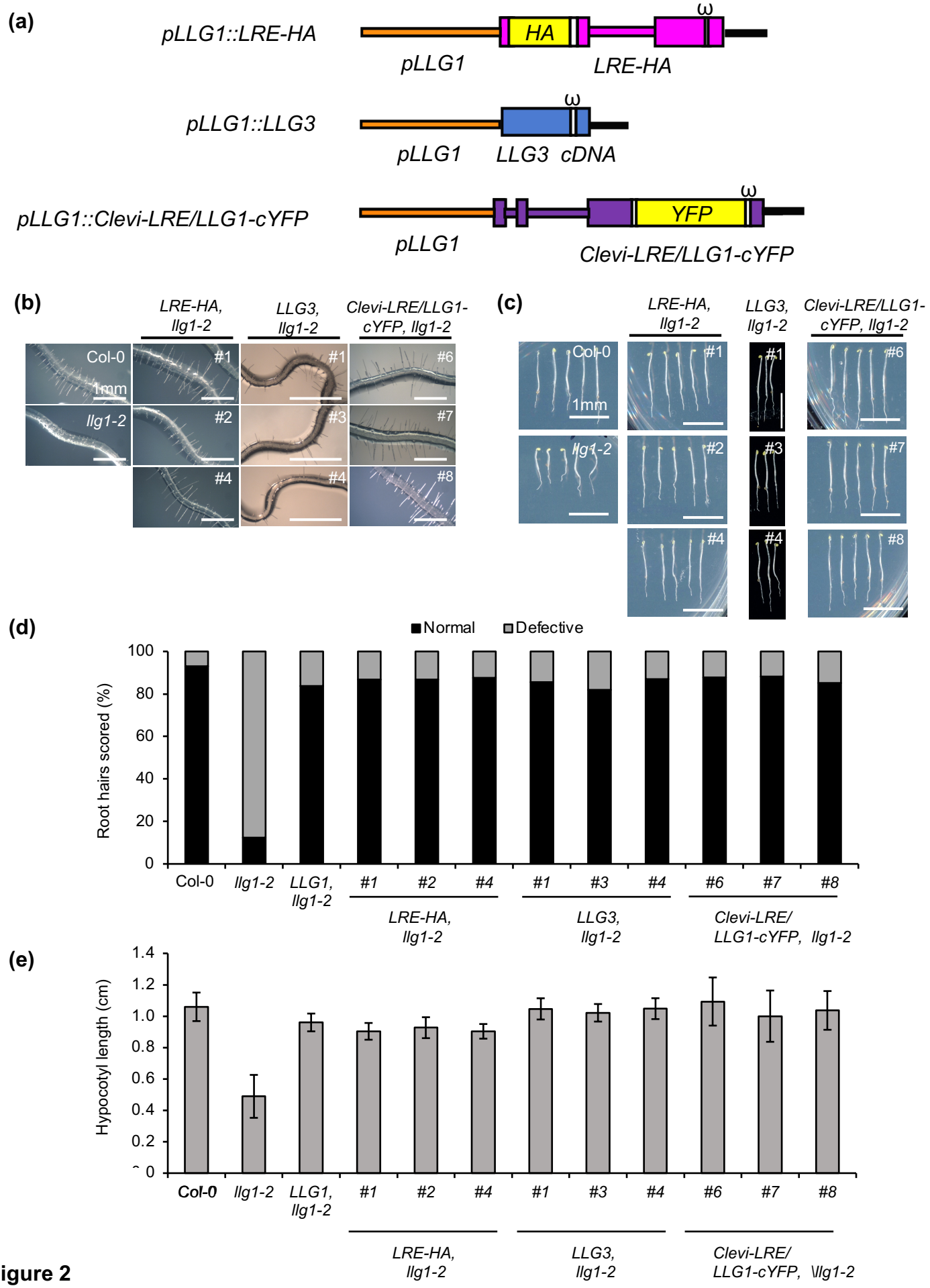
(a)
pLLG1::LRE-HA
HA
ω
pLLG1
LRE-HA
pLLG1::LLG3
ω
pLLG1
LLG3 cDNA
pLLG1::Clevi-LRE/LLG1-cYFP
YFP
ω
pLLG1
Clevi-LRE/LLG1-cYFP
(b)
LRE-HA, llg1-2
LLG3, llg1-2
Clevi-LRE/LLG1-cYFP, llg1-2
Col-0
1mm
llg1-2
#1
#2
#3
#4
#6
#7
#8
(c)
LRE-HA, llg1-2
LLG3, llg1-2
Clevi-LRE/LLG1-cYFP, llg1-2
Col-0
1mm
llg1-2
#1
#2
#3
#4
#6
#7
#8
(d)
Normal
Defective
Root hairs scored (%)
100
80
60
40
20
0
Col-0
llg1-2
LLG1, llg1-2
#1
#2
#4
#1
#3
#4
#6
#7
#8
LRE-HA, llg1-2
LLG3, llg1-2
Clevi-LRE/ LLG1-cYFP, llg1-2
(e)
Hypocotyl length (cm)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
Col-0
llg1-2
LLG1, llg1-2
#1
#2
#4
#1
#3
#4
#6
#7
#8
LRE-HA, llg1-2
LLG3, llg1-2
Clevi-LRE/ LLG1-cYFP, llg1-2

Figure 2

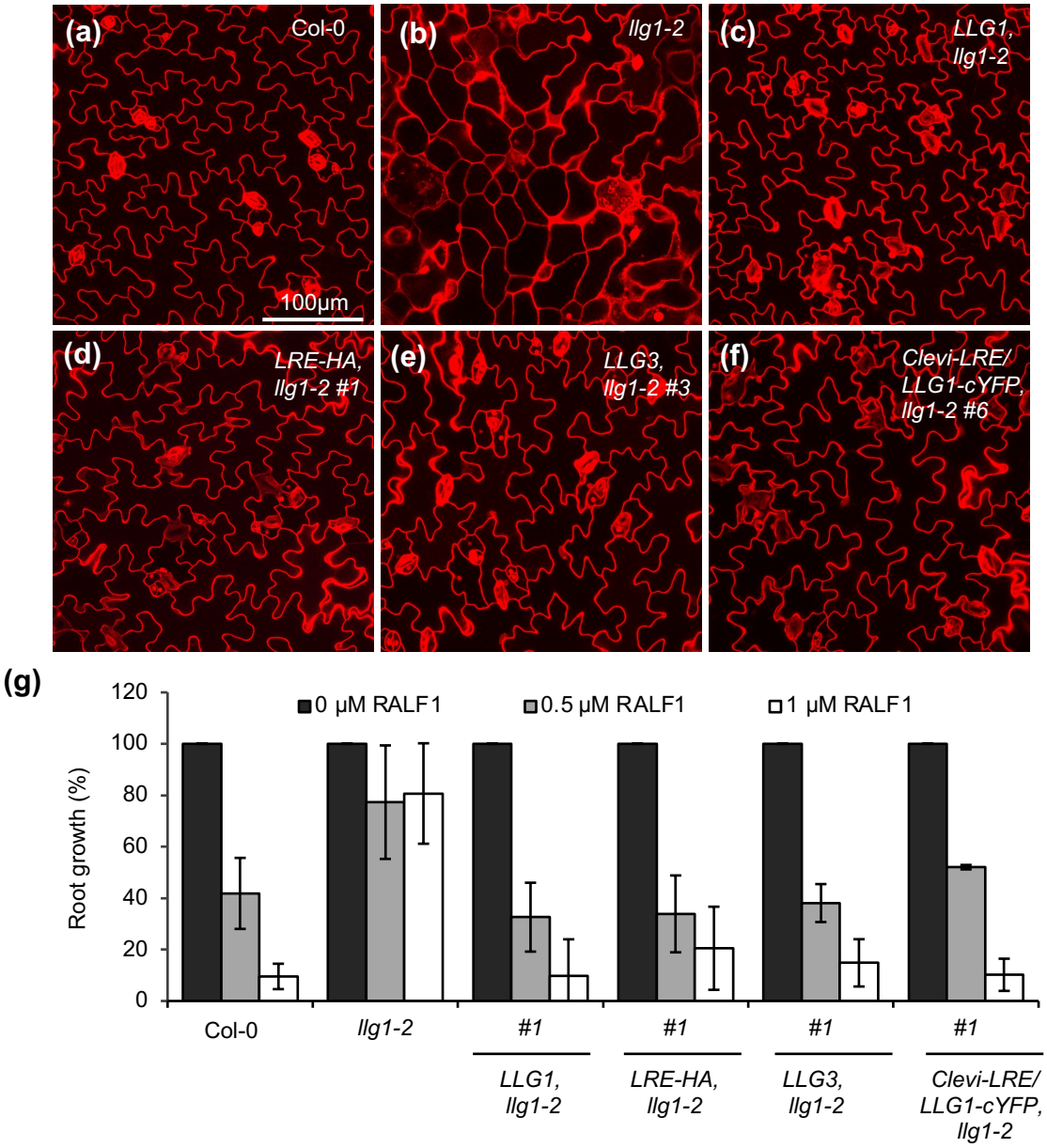
(a)
Col-0
100µm
(b)
llg1-2
(c)
LLG1,
llg1-2
(d)
LRE-HA,
llg1-2 #1
(e)
LLG3,
llg1-2 #3
(f)
Clevi-LRE/
LLG1-cYFP,
llg1-2 #6
(g)
0 µM RALF1
0.5 µM RALF1
1 µM RALF1
Root growth (%)
120
100
80
60
40
20
0
Col-0
llg1-2
#1
#1
#1
#1
LLG1,
llg1-2
LRE-HA,
llg1-2
LLG3,
llg1-2
Clevi-LRE/
LLG1-cYFP,
llg1-2

Figure 3

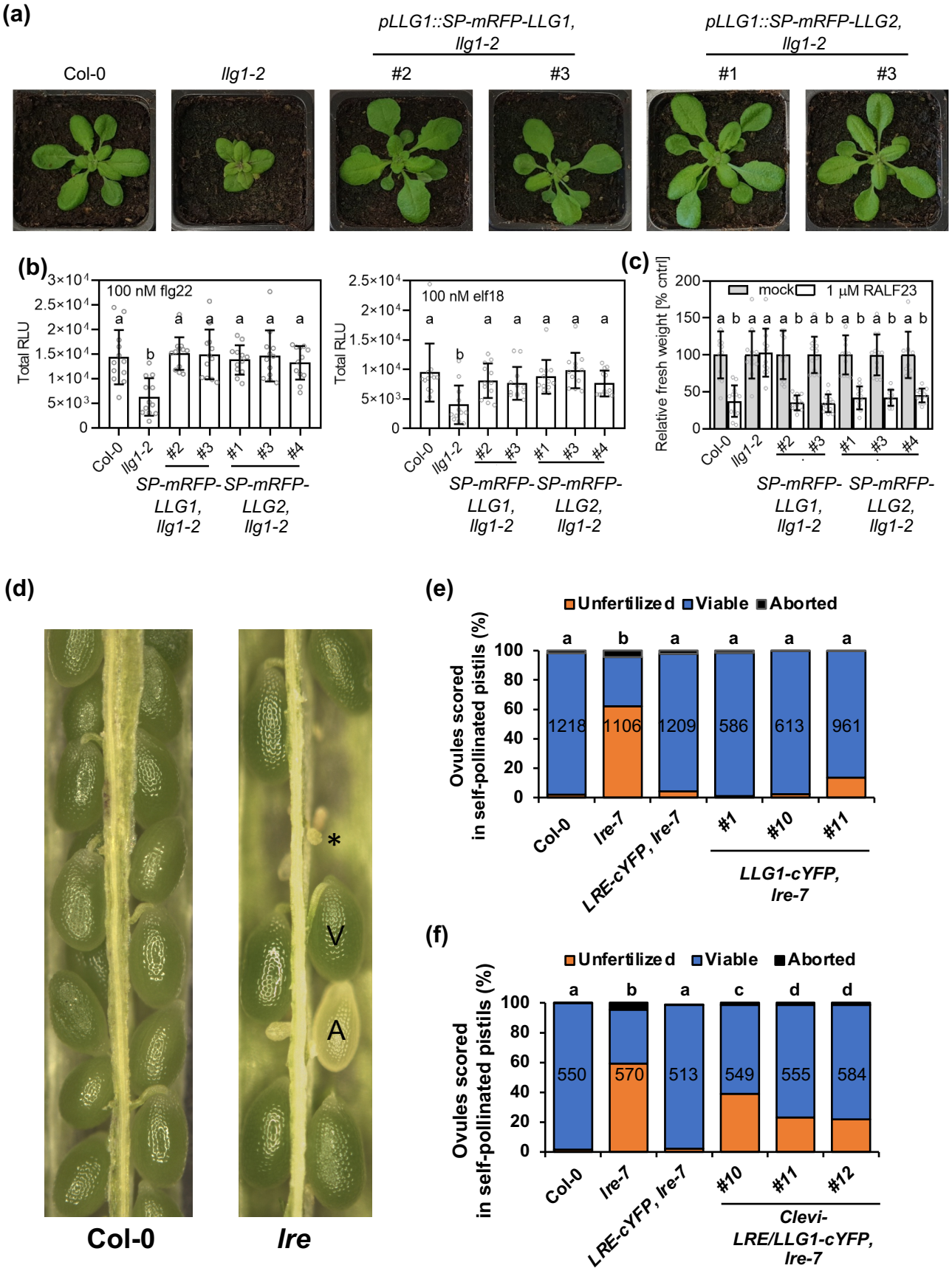
(a)
pLLG1::SP-mRFP-LLG1, llg1-2
pLLG1::SP-mRFP-LLG2, llg1-2
Col-0
llg1-2
#2
#3
#1
#3
(b)
100 nM flg22
Total RLU
Col-0
llg1-2
#2
#3
#1
#3
#4
SP-mRFP-LLG1, llg1-2
SP-mRFP-LLG2, llg1-2
100 nM elf18
(c)
mock
1 µM RALF23
Relative fresh weight [% cntrl]
(d)
Col-0
lre
V
A
*
(e)
Unfertilized
Viable
Aborted
Ovules scored in self-pollinated pistils (%)
1218
1106
1209
586
613
961
lre-7
LRE-cYFP, lre-7
#1
#10
#11
LLG1-cYFP, lre-7
(f)
550
570
513
549
555
584
#12
Clevi-LRE/LLG1-cYFP, lre-7

Figure 4

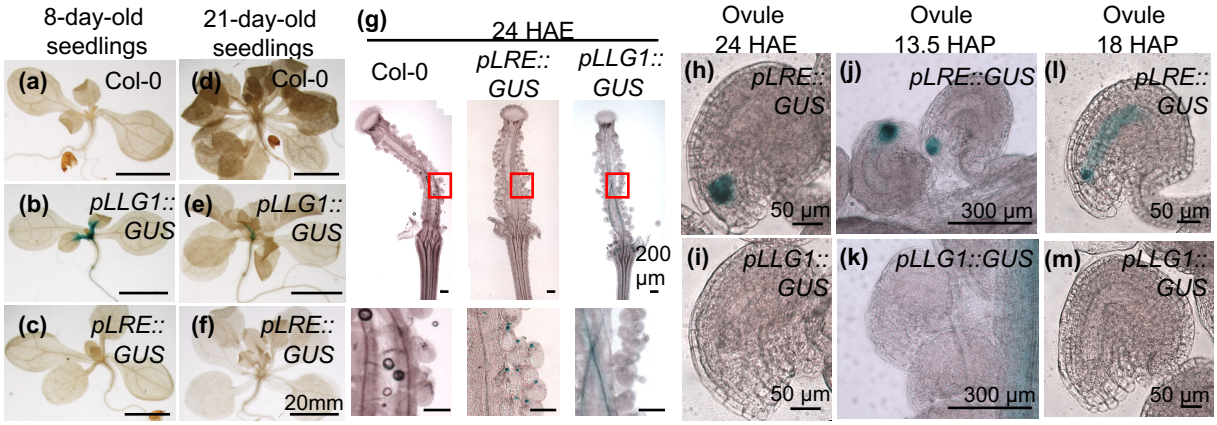
8-day-old seedlings
21-day-old seedlings
(g)
24 HAE
Ovule 24 HAE
Ovule 13.5 HAP
Ovule 18 HAP
(a) Col-0
(d) Col-0
Col-0
pLRE:: GUS
pLLG1:: GUS
(h) pLRE:: GUS
(j) pLRE::GUS
(l) pLRE:: GUS
(b) pLLG1:: GUS
(e) pLLG1:: GUS
200 µm
(i) pLLG1:: GUS
(k) pLLG1::GUS
(m) pLLG1:: GUS
(c) pLRE:: GUS
(f) pLRE:: GUS
20mm
50 µm
300 µm
50 µm
50 µm
300 µm
50 µm

Figure 5

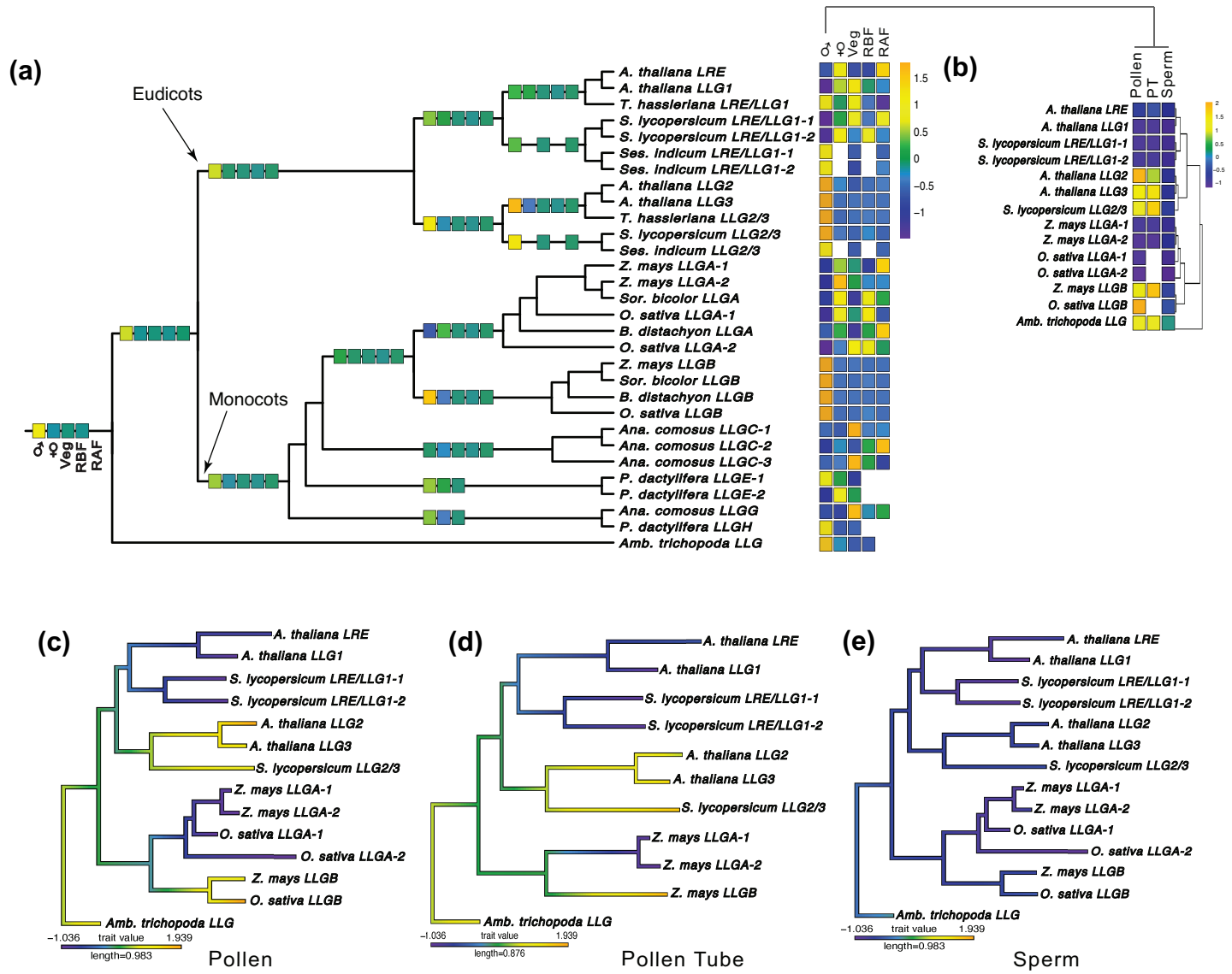
(a)
Eudicots
Monocots
O2
+O
Veg
RBF
RAF
A. thaliana LRE
A. thaliana LLG1
T. hassleriana LRE/LLG1
S. lycopersicum LRE/LLG1-1
S. lycopersicum LRE/LLG1-2
Ses. indicum LRE/LLG1-1
Ses. indicum LRE/LLG1-2
A. thaliana LLG2
A. thaliana LLG3
T. hassleriana LLG2/3
S. lycopersicum LLG2/3
Ses. indicum LLG2/3
Z. mays LLGA-1
Z. mays LLGA-2
Sor. bicolor LLGA
O. sativa LLGA-1
B. distachyon LLGA
O. sativa LLGA-2
Z. mays LLGB
Sor. bicolor LLGB
B. distachyon LLGB
O. sativa LLGB
Ana. comosus LLGC-1
Ana. comosus LLGC-2
Ana. comosus LLGC-3
P. dactylifera LLGE-1
P. dactylifera LLGE-2
Ana. comosus LLGG
P. dactylifera LLGH
Amb. trichopoda LLG
1.5
1
0.5
0
−0.5
−1
(b)
Pollen
PT
Sperm
(c)
−1.036
trait value
1.939
length=0.983
Pollen
(d)
−1.036
trait value
1.939
length=0.876
Pollen Tube
(e)
−1.036
trait value
1.939
length=0.983
Sperm

Figure 6