Assessing duplication and loss of *APETALA1/FRUITFULL* homologs in Ranunculales

Natalia Pabón-Mora[1,2]*, Oriane Hidalgo[3], Stefan Gleissberg[4] and Amy Litt[2]

[1] Grupo de Biotecnología, Instituto de Biología, Universidad de Antioquia, Medellín, Colombia
[2] The New York Botanical Garden, Bronx, NY, USA
[3] Laboratori de Botànica, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain
[4] 38 Fairview Avenue, Athens, OH, USA

Gene duplication and loss provide raw material for evolutionary change within organismal lineages as functional diversification of gene copies provide a mechanism for phenotypic variation. Here we focus on the *APETALA1/FRUITFULL* MADS-box gene lineage evolution. *AP1/FUL* genes are angiosperm-specific and have undergone several duplications. By far the most significant one is the core-eudicot duplication resulting in the *euAP1* and *euFUL* clades. Functional characterization of several *euAP1* and *euFUL* genes has shown that both function in proper floral meristem identity, and axillary meristem repression. Independently, *euAP1* genes function in floral meristem and sepal identity, whereas *euFUL* genes control phase transition, cauline leaf growth, compound leaf morphogenesis and fruit development. Significant functional variation has been detected in the function of pre-duplication basal-eudicot *FUL-like* genes, but the underlying mechanisms for change have not been identified. *FUL-like* genes in the Papaveraceae encode all functions reported for *euAP1* and *euFUL* genes, whereas *FUL-like* genes in *Aquilegia* (Ranunculaceae) function in inflorescence development and leaf complexity, but not in flower or fruit development. Here we isolated *FUL-like* genes across the Ranunculales and used phylogenetic approaches to analyze their evolutionary history. We identified an early duplication resulting in the *RanFL1* and *RanFL2* clades. *RanFL1* genes were present in all the families sampled and are mostly under strong negative selection in the MADS, I and K domains. *RanFL2* genes were only identified from Eupteleaceae, Papaveraceae s.l., Menispermaceae and Ranunculaceae and show relaxed purifying selection at the I and K domains. We discuss how asymmetric sequence diversification, new motifs, differences in codon substitutions and likely protein-protein interactions resulting from this Ranunculiid-specific duplication can help explain the functional differences among basal-eudicot *FUL-like* genes.
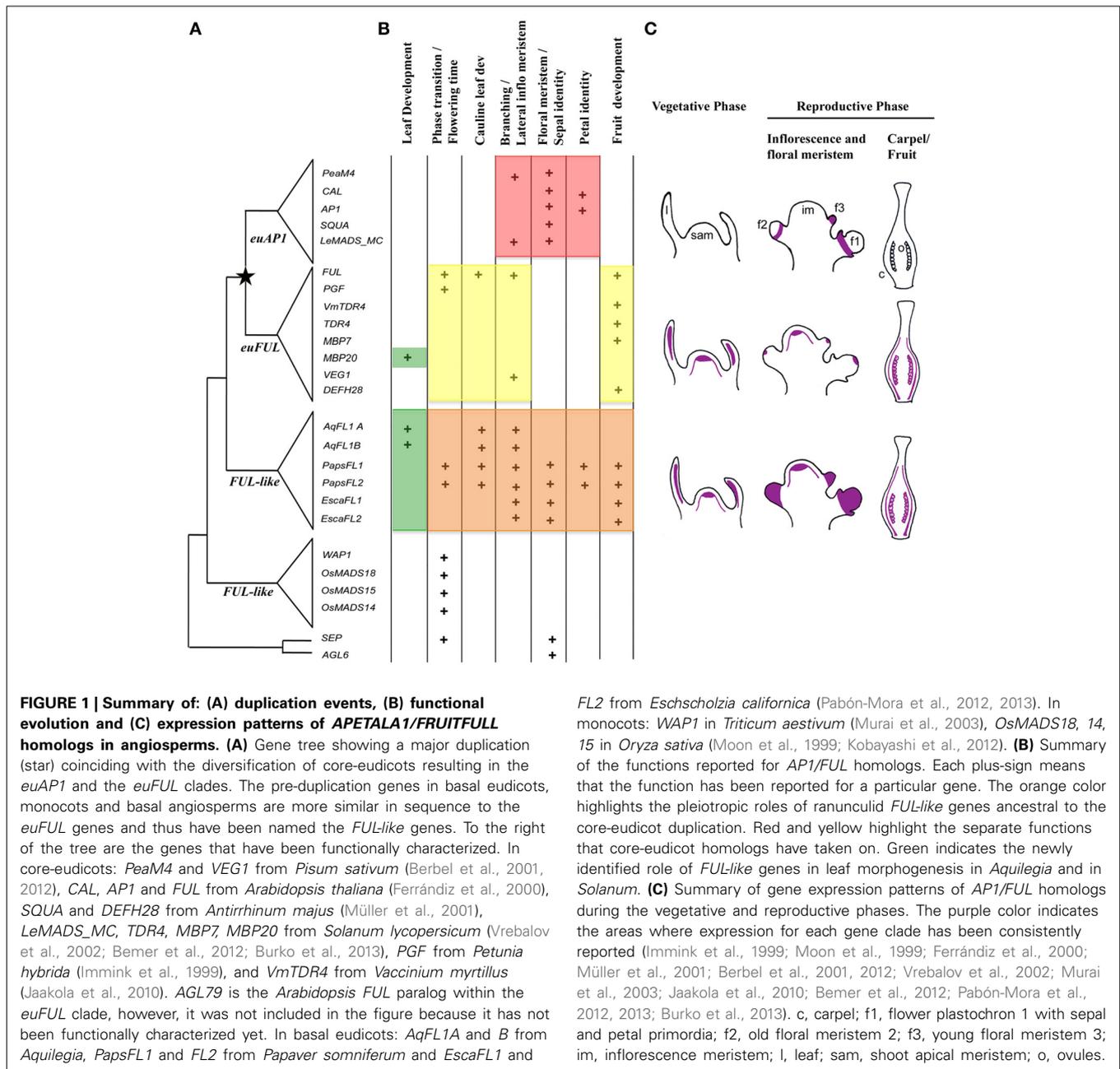
Keywords: gene duplication, *APETALA1*, *FRUITFULL*, basal-eudicots, *FRUITFULL-like*, Ranunculales

INTRODUCTION

The *APETALA1/FRUITFULL* genes are best known for the roles of *APETALA1* (*AP1*), *CAULIFLOWER* (*CAL*) and *FRUITFULL* (*FUL*) paralogs in *Arabidopsis thaliana*. Altogether *AP1*, *CAL* and *FUL* are responsible for proper floral meristem identity (Ferrándiz et al., 2000); in addition, *AP1* plays a key role promoting perianth identity. Because of this, it was included as an A-function gene in the ABC model of flower development (Irish and Sussex, 1990; Coen and Meyerowitz, 1991; Bowman et al., 1993; Gustafson-Brown et al., 1994; Ferrándiz et al., 2000). *CAL* is mostly redundant with *AP1*, however, it has been shown to play an independent role in petal formation (Kempin et al., 1995; Castillejo et al., 2005). *FUL* plays unique roles in proper cauline leaf development and fruit development, and is also a key factor in meristem maintenance and branching (Mandel and Yanofsky, 1995; Gu et al., 1998; Melzer et al., 2008). A fourth, less studied paralog, *AGL79*, is highly divergent in sequence and only expressed in roots, and it has not been functionally characterized

(Parenicová et al., 2003). These paralogous genes are the result of duplications in the *AP1/FUL* gene lineage: whereas the origin of *AP1* and *FUL* is the result of a duplication that resulted in the *euAP1* and *euFUL* gene clades coincident with the origin of the core-eudicots, the close paralogs *AP1* and *CAL* are likely the result of genome duplication events correlated with the diversification of the Brassicaceae (Blanc et al., 2003; Bowers et al., 2003; Alvarez-Buylla et al., 2006; Barker et al., 2009; **Figure 1A**).

The core-eudicot duplication was followed by sequence changes in euAP1 proteins that produced a transcription activation (Cho et al., 1999) and a protein modification motif (Yalovsky et al., 2000). euFUL proteins instead retained the six hydrophobic amino-acid motif that is characteristic of pre-duplication proteins (FUL-like proteins). The function of this motif is unknown (Litt and Irish, 2003; **Figure 1A**). Together *euAP1* and *euFUL* genes promote floral meristem identity (Huijser et al., 1992; Berbel et al., 2001; Vrebalov et al., 2002; Benlloch et al., 2006). Additionally, *euAP1* genes play a unique role in the specification

FIGURE 1 | Summary of: (A) duplication events, (B) functional evolution and (C) expression patterns of *APETALA1/FRUITFULL* homologs in angiosperms. (A) Gene tree showing a major duplication (star) coinciding with the diversification of core-eudicots resulting in the *euAP1* and the *euFUL* clades. The pre-duplication genes in basal eudicots, monocots and basal angiosperms are more similar in sequence to the *euFUL* genes and thus have been named the *FUL-like* genes. To the right of the tree are the genes that have been functionally characterized. In core-eudicots: *PeaM4* and *VEG1* from *Pisum sativum* (Berbel et al., 2001, 2012), *CAL*, *AP1* and *FUL* from *Arabidopsis thaliana* (Ferrándiz et al., 2000), *SQUA* and *DEFH28* from *Antirrhinum majus* (Müller et al., 2001), *LeMADS_MC*, *TDR4*, *MBP7*, *MBP20* from *Solanum lycopersicum* (Vrebalov et al., 2002; Bemer et al., 2012; Burko et al., 2013), *PGF* from *Petunia hybrida* (Immink et al., 1999), and *VmTDR4* from *Vaccinium myrtillus* (Jaakola et al., 2010). *AGL79* is the *Arabidopsis FUL* paralog within the *euFUL* clade, however, it was not included in the figure because it has not been functionally characterized yet. In basal eudicots: *AqFL1A* and *B* from *Aquilegia*, *PapsFL1* and *FL2* from *Papaver somniferum* and *EscaFL1* and

FL2 from *Eschscholzia californica* (Pabón-Mora et al., 2012, 2013). In monocots: *WAP1* in *Triticum aestivum* (Murai et al., 2003), *OsMADS18*, *14*, *15* in *Oryza sativa* (Moon et al., 1999; Kobayashi et al., 2012). **(B)** Summary of the functions reported for *AP1/FUL* homologs. Each plus-sign means that the function has been reported for a particular gene. The orange color highlights the pleiotropic roles of ranunculid *FUL-like* genes ancestral to the core-eudicot duplication. Red and yellow highlight the separate functions that core-eudicot homologs have taken on. Green indicates the newly identified role of *FUL-like* genes in leaf morphogenesis in *Aquilegia* and in *Solanum*. **(C)** Summary of gene expression patterns of *AP1/FUL* homologs during the vegetative and reproductive phases. The purple color indicates the areas where expression for each gene clade has been consistently reported (Immink et al., 1999; Moon et al., 1999; Ferrándiz et al., 2000; Müller et al., 2001; Berbel et al., 2001, 2012; Vrebalov et al., 2002; Murai et al., 2003; Jaakola et al., 2010; Bemer et al., 2012; Pabón-Mora et al., 2012, 2013; Burko et al., 2013). c, carpel; f1, flower plastochron 1 with sepal and petal primordia; f2, old floral meristem 2; f3, young floral meristem 3; im, inflorescence meristem; l, leaf; sam, shoot apical meristem; o, ovules.

of sepal (and in *Arabidopsis*, petal) identity (Berbel et al., 2001; Vrebalov et al., 2002; Benlloch et al., 2006) whereas, *euFUL* genes function in the reproductive phase transition, proper cauline leaf development, branching, and fruit development as well as compound leaf development (Immink et al., 1999; Müller et al., 2001; Jaakola et al., 2010; Bemer et al., 2012; Berbel et al., 2012; Torti et al., 2012; Burko et al., 2013; Meyer et al., unpublished data; **Figure 1B**). The functional differences between *euAP1* and *euFUL* genes suggest an evolutionary scenario of either sub- or neofunctionalization after duplication, and studies of the function of *FUL*-like genes in basal eudicot Ranunculales (ranunculids) evaluated these two hypotheses. The *FUL-like* genes of *Papaver somniferum* (opium poppy; Papaveraceae) were shown to play

pleiotropic roles that include essentially all those reported for *euAP1* and *euFUL* genes; thus, sub-functionalization was postulated as the outcome of the core-eudicot *AP1/FUL* duplication (**Figure 1B**; Pabón-Mora et al., 2012). However, functional analyses in *E. californica* (California poppy), also in Papaveraceae, showed that *FUL-like* genes in this species are involved only in a subset of these functions (**Figure 1B**: Pabón-Mora et al., 2012), and studies of *FUL-like* gene function in *Aquilegia coerulea* (columbine; Ranunculaceae) have shown only a role in regulating inflorescence branching and a role in compound leaf morphogenesis (Pabón-Mora et al., 2013; **Figure 1B**).

These studies on the *FUL-like* genes of ranunculids detected significant variation in the function of basal eudicot *FUL-like*

genes. This observed functional diversity is not associated with changes in expression; in general all the ranunculid *FUL-like* genes are turned on in the shoot apical meristem and leaves, and expression is maintained throughout inflorescence and flower development, in all floral organs and fruit (**Figure 1C**). Thus, functional differences may instead be the result of protein sequence changes leading to differences in interactions with other transcription factors or downstream factors. Such sequence changes may hold clues to observed differences in function among genes belonging to different taxa (e.g., Papaveraceae vs. Ranunculaceae) as well as to the selective forces operating on genes of different paralagous lineages. Changes in developmental functions among paralogous genes are often accompanied by changes in rates and patterns of sequence evolution among loci (Purugganan and Suddith, 1998; Lawton-Rauh et al., 1999) and faster rates of evolution are usually associated with the occurrence of genetic redundancy (Lawton-Rauh et al., 1999). To understand functional evolution within ranunculid *FUL-like* genes this study had two main goals: (1) to explore in detail *FUL-like* gene duplications and losses in Ranunculales to establish the relationship among functionally characterized copies, and (2) to investigate differences in protein motifs and rates of evolution and selection across *FUL-like* genes in members of the ranunculids. The results of these analyses were used to understand the variation in *FUL-like* gene function among poppy, California poppy, and columbine and to identify changes in protein evolution that may be linked with differences in protein interaction capabilities across ranunculid FUL-like proteins.

MATERIALS AND METHODS
PLANT MATERIAL
Leaf and floral tissue was obtained from a number of basal eudicots, mostly within Papaveraceae s.l., Berberidaceae and Ranunculaceae, as well as non-eudicots within Aristolochiaceae (Piperales). Fresh material was obtained from living collections at The New York Botanical Garden, Bronx, NY or at the Systematics Garden at Lehman College, Bronx, NY. Voucher information for all species is listed in **Table S1**.

CLONING AND CHARACTERIZATION OF *FUL-like* GENES
Total RNA was extracted from 0.5–1 g of young leaf or floral buds using TRIZOL reagent (Invitrogen) and was DNaseI-treated (Roche) to remove residual genomic DNA. 2 μg were used as template for cDNA synthesis with SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions using the OligodT primer supplied. The resulting cDNA was diluted 1:10 for use in amplification reactions. Initial amplifications using degenerate primers to recover a pool of MADS-box genes were done as in Litt and Irish (2003), with two modifications; (1) the amplification program began with a 5 min activation step at 95°C, and five initial cycles with an incubation step of 30 s at 95°C, a 30 s annealing step at 42°C and a 1 min extension at 72°C, followed by 30 cycles with an incubation step at 95°C for 30 s, a 30 s annealing step at 50°C and a 1 min extension at 72°C. The products of this amplification were diluted 1:20 and used as template in successive reactions. In addition to

the primers used by Litt and Irish (2003) the forward degenerate primer ATGGRDAGAGGWAGGGTWCAG, designed to bind the beginning of the MADS domain, was used in combination with all degenerate reverse primers designed to amplify the full coding sequence towards the 5′ end of the *FUL-like* genes. All PCR products were run on a 1% agarose gel and amplicons between 600 and 900 bp in size were cloned into pCR®2.1-TOPO® (Invitrogen). Clones were grown overnight, plasmid was extracted with the Qiagen miniprep Kit (Invitrogen) and sequenced at the DNA Yale Sequencing Center (CT).

In addition to degenerate PCR, we searched public databases, using BLAST (Altschul et al., 1990) and obtained 16 *FUL-like* genes from the transcriptomes available at the phytometasyn project website (http://www.phytometasyn.ca) and 29 *FUL-like* genes from GenBank (http://www.ncbi.nlm.nih.gov/genbank/). Sequences from 51 species and all families in Ranunculales (Eupteleaceae, Papaveraceae, Lardizabalaceae, Menispermaceae, Berberidaceae and Ranunculaceae) were included except Circaeasteraceae, from which material could not be obtained. Outgroups included representatives of the Magnoliaceae, Lauraceae, Saururaceae and Poaceae (**Table S1**).

PHYLOGENETIC ANALYSES
Between 40 and 60 clones were sequenced per species. If variation was found among clones, the criteria to distinguish allelic variation at a single locus from different loci were the same used by Litt and Irish (2003). *FUL-like* sequences in the transcriptome databases were assembled into contigs and screened for polymorphisms using Sequencher DNA sequencing software (GeneCodes, Ann Arbor, MI): if different hits had less than 5% variation a consensus sequence was generated; if the difference among hits was larger, the two sequences were both kept in the analysis. Only sequences containing at least part of the MADS domain and the FUL-motif were included in the analysis. Sequences were compiled using Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html), and then aligned using the online version of MAFFT (http://mafft.cbrc.jp/alignment/server/) (Katoh et al., 2002), with a gap open penalty of 3.0, an offset value of 0.3, and all other default settings. The alignment was then refined by hand using Bioedit. The nucleotide alignment for 109 full-length sequences from 51 species was used for phylogenetic analyses. The amino acid alignment, also generated in Bioedit, was used to identify conserved motifs as well as single amino acids that were diagnostic of clades; these were optimized and visualized in MacClade4.08a® (Maddison and Maddison, 2005). The Magnoliid sequences (Ma.gr.AP1 and Pe.am.AP1) were used to root the trees, and all non-Ranunculid sequences were used as outgroup.

Maximum Likelihood (ML) phylogenetic analyses were performed in RaxML-HPC2 BlackBox (Stamatakis et al., 2008) on the CIPRES Science Gateway (Miller et al., 2009). The best performing evolutionary model was obtained by the Akaike information criterion (AIC; Akaike, 1974) using the program jModelTest v.0.1.1 (Posada and Crandall, 1998). Bootstrapping was performed according to the default criteria in RAxML where bootstrapping stopped after 200 replicates when the criteria were met.

RELATIVE RATES OF EVOLUTION

To test for evidences of changes in selection constraints in the Ranunculid *FUL-like* gene tree, we performed a series of likelihood ratio tests (LRTs) using the branch-specific model implemented by the CodeML program of PAML package v.4.6 (Yang, 2007). We compared the one ratio model that assumes a constant dN/dS ratio (= ω, per site ratio of nonsynonymous -dN- to synonymous -dS- substitution) along tree branches, against a two-ratio model that assumes a different ratio for a designated ranunculid *FUL-like* subclade (foreground) relative to the remaining sequences (background). For each of the LRTs, twice the difference of log likelihood between the models (2 ΔlnL) was compared to critical values from a χ2 distribution, with degree of freedom equal to the differences in number of estimated parameters between models. The test was conducted for the entire dataset and also for each of the functional domains defined for MADS-box genes. These analyses on the M, IK, and C domains were performed in order to evaluate whether there was a difference in their rates of evolution in different taxa, given their key roles in DNA binding (M), protein dimerization (IK), and multimerization (C).

RESULTS

FUL-like GENE CLONING IN RANUNCULALES

In order to gain a better understanding of the basis of the functional diversity reported for *FUL-like* genes in the basal eudicot order Ranunculales, we looked at patterns of evolution among these genes. We isolated *FUL-like* copies from species representing the phylogenetic breadth of the Ranunculales, an order with nearly 202 genera and 4500 species (APG, 2009; Wang et al., 2009; **Figures 2**, **3**), and reconstructed the evolutionary history of the gene lineage in this clade. Ranunculales includes the early-diverging families Eupteleaceae and Papaveraceae s.l., as well as the core Ranunculales Lardizabalaceae, Circaeasteraceae, Menispermaceae, Berberidaceae and Ranunculaceae. We generated a dataset consisting of 109 *FUL-like* gene sequences (**Table S1**) from Eupteleaceae, Papaveraceae s.l., Lardizabalaceae, Menispermaceae, Berberidaceae and Ranunculaceae, as well as the outgroup basal angiosperm and monocot families Magnoliaceae, Lauraceae, Saururaceae, Aristolochiaceae and the monocot family Poaceae. Sequences from Circeasteraceae were not included due to lack of availability of material.

Clones that were recovered with degenerate primers either span the entire coding sequence or are missing 10–20 amino acids (AA) from the start of the 60 AA MADS domain. The alignment includes 60 AA in the MADS domain, 35–40 in the I domain, 70–75 in the K domain, and 90 in the C-terminal domain. Among Ranunculales, paralogous gene sequence similarity ranges from 52 to 95%, and the variation in sequence similarity between outgroup and ingroup ranges from 50 to 75%. In the C-terminal portion, all protein sequences show the previously described FUL-like motif (Litt and Irish, 2003; Preston and Kellogg, 2006; Shan et al., 2007). Alignment of the predicted amino acid sequences of the entire dataset reveals a high degree of conservation in the M, I, and K regions until position 184. In most plant MADS proteins, the structurally conserved Keratin-like domain (K), forms three amphipathic helices (K1,

K2, K3) that are important for strength and specificity of protein dimerization (Yang et al., 2003). Usually the three putative amphipathic α-helices of the K domain have heptad repeats (**a**bc**d**efg)$_n$, in which **a** and **d** positions are occupied by hydrophobic amino-acids. The putative amphipathic α-helices of ranunculid FUL-like proteins, K1 (AA 97–110), K2 (AA 121–143) and K3 (AA 152–258), conform to this expected pattern. (**Figure S1**). Within K1, positions 99 (E), 102 (K), 104 (K), 106 (K), 108 (E), and 111 (Q), and within K2 positions 119 (G), 128 (K), 129 (E), 134 (E), 136 (Q), are conserved in all Ranunculales and outgroup FUL-like predicted protein sequences, with a few exceptions (**Figure S1**). The C-terminal domain, beginning after the hydrophobic amino acid located in position 184, is more variable, but three regions of high similarity can be identified: (1) a region rich in tandem repeats of polar uncharged amino acids (QNQ), particularly glutamine (Q), between positions 190–230 in the alignment; (2) a highly conserved, predominantly hydrophobic motif unique to ranunculids at positions 226–256, with the sequence QNS-P/LS/TFLLSQSE/LP-SLN/TI, and (3) a negatively charged region rich in glutamic acid (E) before the conserved FUL-motif LMPPWML (**Figure 2**).

GENE DUPLICATION AND LOSS OF *FUL-like* GENES IN RANUNCULALES

A total of 910 characters were included in the matrix, of which 645 (71%) were informative. Maximum likelihood analysis recovered a single duplication event early in the diversification of the Ranunculales resulting in two clades of *FUL-like* genes, here named *RanFL1* and *RanFL2* (**Figure 3**). Bootstrap support for the *RanFL1* and *RanFL2* clades is low (<50), however, within each clade, gene copies from the same family are grouped together with strong support, and the relationships among gene clades are mostly consistent with the phylogenetic relationships of the sampled taxa (Wang et al., 2009). An exception is the position of the Menispermaceae sequences as sister to the Papaveraceae s.l. sequences—although with long branches and low support—in both gene clades; phylogenetic analyses have shown Menispermaceae as the sister group to [Ranunculaceae + Berberidaceae] (Wang et al., 2009). Other inconsistent positioning is the placement of Lardizabalaceae as sister to [Papaveraceae + Menispermaceae], while it was sister to [Menispermaceae (Ranunculaceae + Berberidaceae)] in Wang et al. (2009).

Additional duplications and putative losses can also be detected. The *RanFL1* clade contains two paralogous Lardizabalaceae clades, *LarFL1a* and *LarFL1b*, but the *RanFL2* clade lacks sequences from this family. This suggests that *LarFL1* genes underwent an independent duplication, and that *LarFL2* members have been lost or are yet to be found. *RanFL2* sequences were also not recovered from Berberidaceae. Additional taxon-specific duplications were found in *Pseudofumaria lutea, E. californica* (Papaveraceae sl.), *Berberis gilgiana* and *Nandina domestica* (Berberidaceae), *A. coerulea, Eranthis hyemalis* and *Ranunculus sceleratus* (Ranunculaceae) within the *RanFL1* clade. Similarly, duplications were found in *Bocconia frutescens* (Papaveraceae) within the *RanFL2* clade. Finally, duplications in both clades (*RanFL1* and *RanFL2*) were evident in *Argemone*
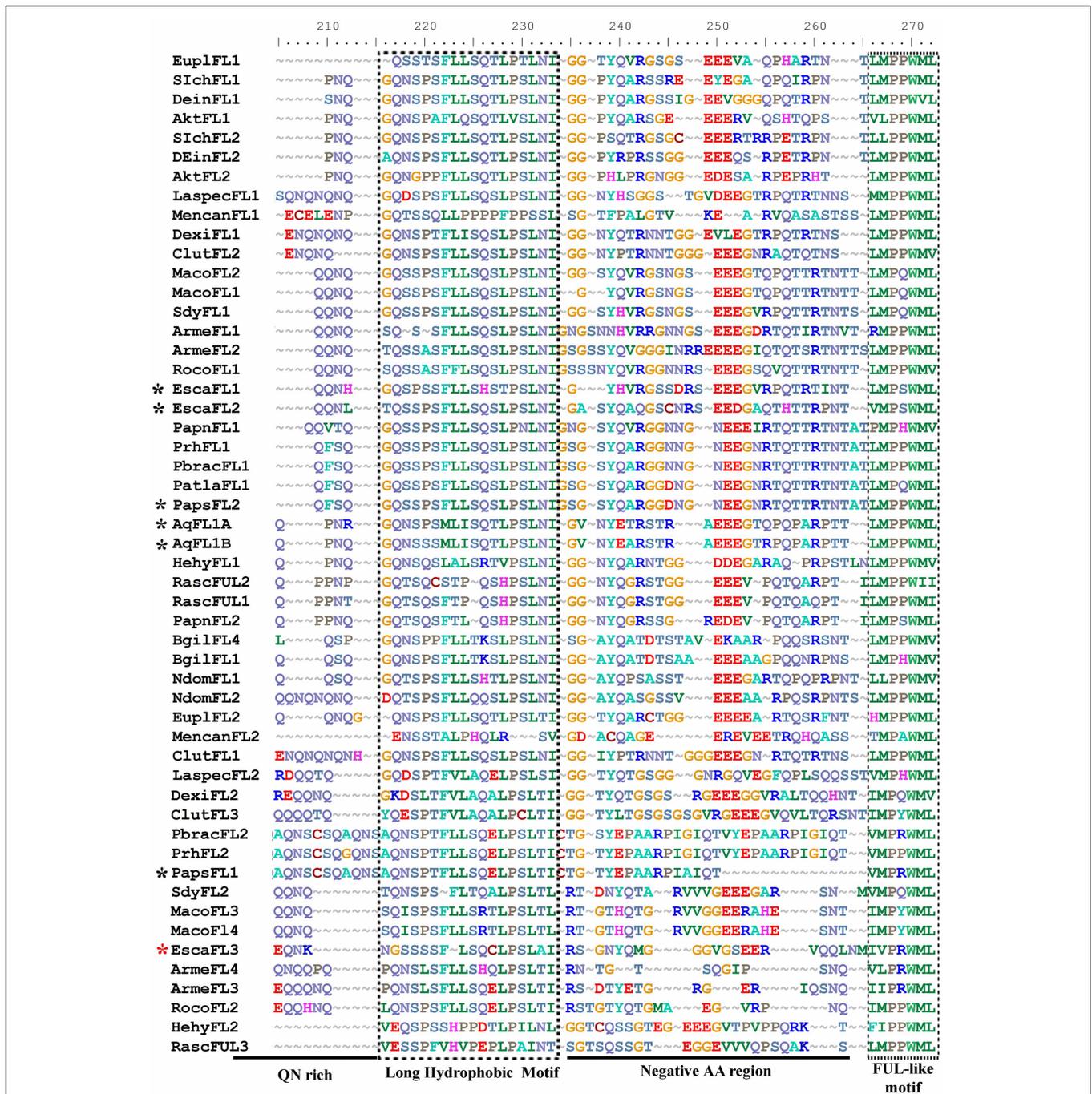
FIGURE 2 | Sequence alignment including the end of the K domain (K) and the complete C-terminal domain of ranunculid FUL-like proteins. The alignment shows a region rich in glutamine (Q), asparagine (N) and serine (S), labeled as the QN rich zone, followed by the conserved hydrophobic motif newly identified (boxed), a region negatively charged and rich in glutamic acid (E), labeled the Negative AA region, and the FUL-like motif (boxed), typical of FUL-like and euFUL proteins. CmFL1 was excluded from the alignment because is the only sequence that has an additional insertion in the "hydrophobic motif" with 8 additional AA in between positions 229–236. Black asterisks show proteins that have been functionally characterized, red asterisk points to EscaFL3 that was not previously identified and has not been functionally characterized.

mexicana, Macleaya cordata (Papaveraceae), and *Ranunculus bulbosus* (Ranunculaceae). Since most of these species are thought to be polyploid (Index to Plant Chromosome Numbers; Missouri Botanical Garden, http://www.tropicos.org/Project/IPCN), additional duplicates are likely derived from whole genome duplications. If so, these transcription factors, that are thought to function as tetramers with other MADS box proteins at least in flower development (Smaczniak et al., 2012),
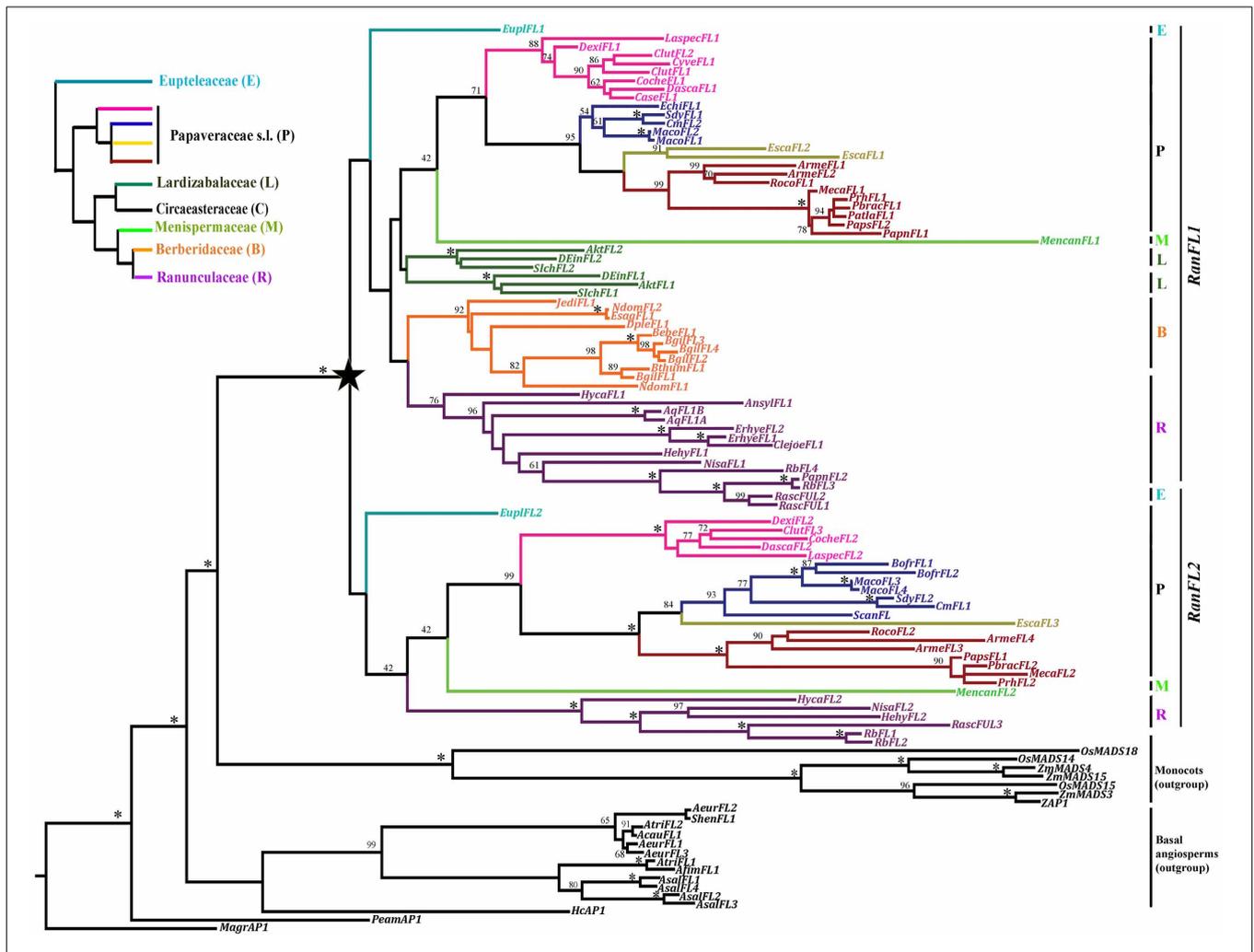
FIGURE 3 | Best Maximum Likelihood tree of *FUL-like* genes in Ranunculales. Bootstrap values (above 40%) are placed at nodes. Asterisks indicate bootstrap values of 100%. The star indicates the duplication event that resulted in the Ran*FUL-like1* (*RanFL1*) and Ran*FUL-like2* (*RanFL2*) clades. Branch colors and vertical lines on the right denote different plant families as indicated in the organismal tree in the inset at the left (Wang et al., 2009). Papaveraceae s.l. is here shown with four different colors belonging to specific clades: bright pink shows the subfamily Fumarioideae; subfamily Papaveroideae is subdivided into the tribes Chelidonieae (blue), Eschscholtzieae (yellow)

and Papavereae (red). Note that both the *RanFL1* and *RanFL2* clades have representative members from Eupteleaceae, Papaveraceae, Menispermaceae and Ranunculaceae, whereas, only *RanFL1* genes were amplified from Lardizabalaceae and Berberidaceae, suggesting that *RanFL2* genes from these families have been lost. In addition Lardizabalaceae *FL1* genes have undergone an independent duplication resulting in the Lardizabalaceae *FL1a* and *b* clades. B, Berberidaceae; E, Eupteleaceae; L, Lardizabalaceae; M, Menispermaceae; P, Papaveraceae; R, Ranunculaceae. Outgroup includes Basal angiosperms and Monocots in black.

are likely to maintain their functions and partners, given that during polyploidization events their partners also duplicate (Otto and Whitton, 2000; Blanc and Wolfe, 2004). Duplicates in *E. californica* are likely tandem-repeats or transcripts inserted by retro-transposition, as this is thought to be a diploid species with a chromosome number of $2n = 14$ (Hidalgo et al., in prep). Similar local *FUL-like* gene duplications may have occurred in *E. hyemalis* and *R. bulbosus*, which are also thought to be diploids ($2n = 16$; Index to Plant Chromosome Numbers; Missouri Botanical Garden, http://www.tropicos.org/Project/IPCN).

Taxon-specific losses are harder to confirm, since is possible that some copies were not recovered through our cloning strategy. Nevertheless, our results suggest that *RanFL1* copies were lost in

Sanguinaria canadensis and *B. frutescens* (Papaveraceae s.str.), and that *RanFL2* copies were lost in *Cysticapnos vesicaria*, *Capnoides sempervirens* and *Eomecon chionanta* (Papaveraceae s.l.) as well as in *Anemone sylvestris*, *E. hyemalis*, *Clematis sp* and *A. coerulea* (Ranunculaceae). The loss can only be confirmed in the case of *A. coerulea* as in this case the genome has been sequenced (Joint Genome Institute, 2010).

Finally we identified amino acid synapomorphies for subclades within the *RanFL1* and RanFL2 subclades, but no synapomorphies for those two clades themselves, consistent with the low support values in the deeper branches of the tree (**Figures 3**, **4**). Nearly all the terminal subclades have at least one synapomorphy or as many as nine, however, the number of synapomorphies

for each paralogous subclade differs greatly according to the family. For instance whereas Papaveraceae s. str. FL1 and FL2 have a single synapomorphy supporting each clade, Ranunculaceae FL1 and FL2 have one and nine synapomorphies respectively, suggesting that conserved aminoacids may have been fixed at different rates in the coding sequences of different paralogous clades.

SHIFTS IN SELECTION CONSTRAINTS IN THE HISTORY OF RANUNCULALES *FUL-like* GENES

Likelihood ratio tests, carried out to determine whether there were differences in selection acting on the ranunculid FUL-like sequences, show all tested ranunculid lineages to have $\omega < 1$, indicating purifying selection (**Table 1**). This purifying pressure, however, exhibits significant variation (strengthening and release) across FUL-like subclades and in different protein domains (**Figure 5A**; **Table 1**). Indeed, while Ranunculales do not show a significant difference in the selective pressure acting on FUL proteins with respect to background taxa (basal angiosperms and grasses) at the level of the whole sequence, purifying pressure is significantly reinforced in the MADS domain and released in the IK region. In addition the analyses revealed that although both gene clades are under purifying selection, the degree of purifying selection is stronger in RanFL1 ($\omega_f = 0.18$ vs. $\omega_b = 0.25$) and significantly relaxed in RanFL2 ($\omega_b = 0.29$ vs. $\omega_0 = 0.19$) (**Figure 5A**, ω_f vs. ω_b values and statistical significance are listed in **Table 1**).

An expanded analysis using the two-ratio test in each gene subclade showed a more complicated pattern of molecular evolution with each plant family showing unique selection constraints. Strengthening of purifying selection is detected in Papaveraceae s.l. FL1 and Berberidaceae FL1 ($\omega_f = 0.13$ vs. $\omega_b = 0.23$ and $\omega_f = 0.15$ vs. $\omega_b = 0.22$ respectively), whereas purifying selection is relaxed in Lardizabalaceae FL1a ($\omega_f = 0.46$ vs. $\omega_b = 0.21$) and FL1b ($\omega_f = 0.33$ vs. $\omega_b = 0.21$), Papaveraceae FL2

($\omega_f = 0.30$ vs. $\omega_b = 0.19$) and Ranunculaceae FL2 ($\omega_f = 0.21$ vs. $\omega_b = 0.27$). In addition, these analyses also detected strong purifying selection in Menispermaceae FL1 and FL2 ($\omega_f = 0.16$ vs. $\omega_b = 0.21$ and $\omega_f = 0.16$ vs. $\omega_b = 0.21$ respectively) as well as relaxed purifying selection in Eupteleaceae FL1 and FL2 ($\omega_f = 0.30$ vs. $\omega_b = 0.21$ and $\omega_f = 0.36$ vs. $\omega_b = 0.21$ respectively), however, significant statistical support is lacking in these cases (**Figure 5A**; **Table 1**).

In order to test whether specific regions of the proteins were experiencing different selective pressures, we repeated the tests on the three distinct protein regions: the MADS (1–180 nt), the I + K (181–541 nt) and the C-terminal (542–910 nt) domains. The results showed that the MADS domain was under strong purifying selection in the Papaveraceae s.l. FL1 ($\omega_f = 0.01$ vs. $\omega_b = 0.05$) and under relaxed purifying selection in Lardizabalaceae FL1a and FL1b ($\omega_f = 0.26$ vs. $\omega_b = 0.04$ and $\omega_f = 0.22$ vs. $\omega_b = 0.04$ respectively) and in the Eupteleaceae FL2 ($\omega_f = 0.19$ vs. $\omega_b = 0.04$). Changes in selection were also evident in the I + K domains, showing strong purifying selection in Papaveraceae s.l. FL1 ($\omega_f = 0.08$ vs. $\omega_b = 0.19$) and Berberidaceae FL1 ($\omega_f = 0.08$ vs. $\omega_b = 0.18$) and a relaxed purifying selection in Eupteleaceae FL1 and FL2 ($\omega_f = 0.47$ vs. $\omega_b = 0.16$ and $\omega_f = 0.43$ vs. $\omega_b = 0.17$), Lardizabalaceae FL1a ($\omega_f = 0.67$ vs. $\omega_b = 0.16$), Papaveraceae FL2 ($\omega_f = 0.28$ vs. $\omega_b = 0.15$) and Ranunculaceae FL2 ($\omega_f = 0.33$ vs. $\omega_b = 0.16$). Significative changes in selection at the C terminus were only detected in Papaveraceae s.l. ($\omega_f = 0.62$ vs. $\omega_b = 0.39$) (**Figure 5A**; **Table 1**).

DISCUSSION

FUL-like GENES UNDERWENT DUPLICATION EARLY IN THE DIVERSIFICATION OF THE RANUNCULALES

The ML analysis showed a single major duplication in the ranunculid FUL-like genes which gave rise to the *RanFL1* and *RanFL2* gene clades early in the diversification of the order

FIGURE 4 | Diagnostic amino acid characters of the Ranunculales FUL-like proteins, indicating the position in our alignments, mapped on a summary gene tree. The star denotes the duplication event. Colors and names of the gene clades follow **Figure 3** and are here abbreviated.
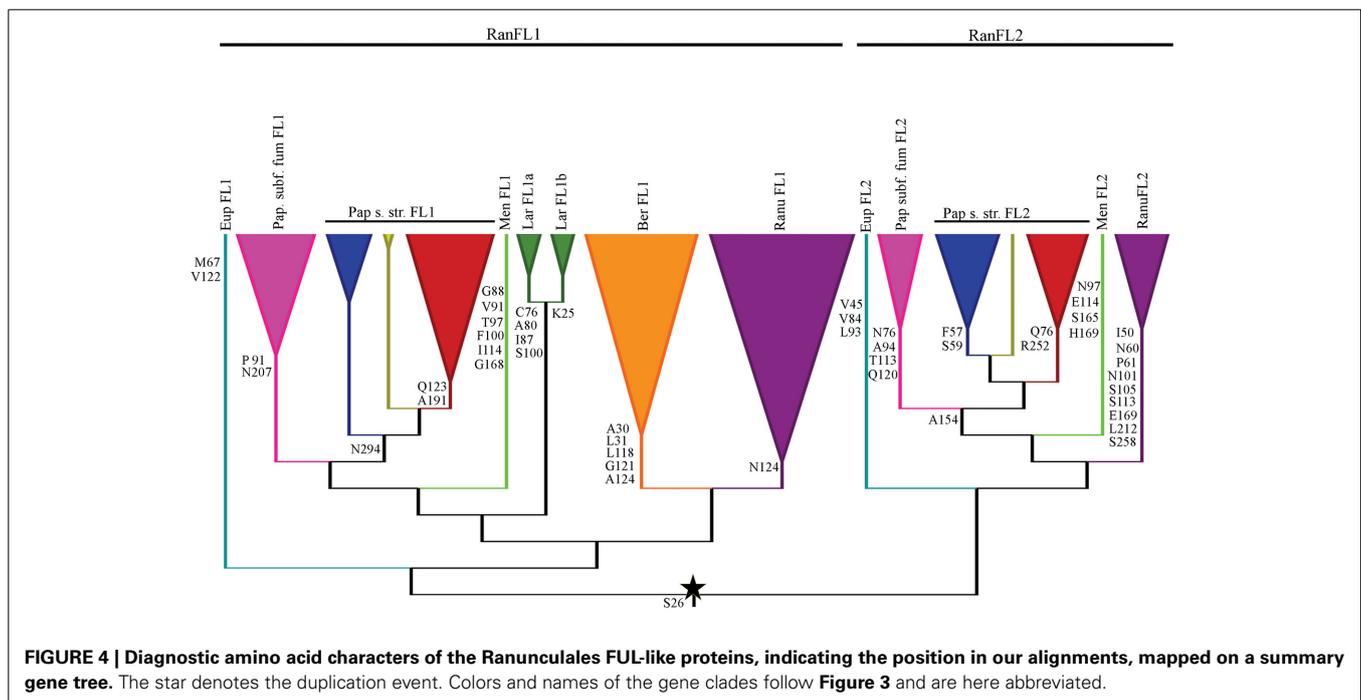
Table 1 | Comparison of the one ratio model that assumes a constant dN/dS ratio (= ω, per site ratio of nonsynonymous -dN- to synonymous -dS- substitution) along tree branches, against a two-ratio model that assumes a different ratio for a designated ranunculid FUL-like subclade (foreground -ω_f) relative to the remaining sequences (background -ω_b).

Model	Whole FUL sequence					MADS region					IK region					C region				
	p	Results	LnL	2ΔlnL	(LRT)	p	Results	LnL	2ΔlnL	(LRT)	p	Results	LnL	2ΔlnL	(LRT)	p	Results	LnL	2ΔlnL	(LRT)
One ratio	213	$w_0 = 0.2166$	−29.104,26			213	$w_0 = 0.0478$	−4.533,45			213	$w_0 = 0.1714$	−12.248,26			213	$w_0 = 0.4560$	−11.194,84		
Two ratios																				
Ranunculales-FUL	214	$w_B = 0.1986$; $w_F = 0.2216$	−29.103,01	2,50		214	$w_B = 0.0799$; $w_F = 0.0421$	−4.528,98	8,93	**	214	$w_B = 0.1134$; $w_F = 0.1916$	−12.237,11	22,31	***	214	$w_B = 0.3741$; $w_F = 0.4759$	−11.192,97	3,74	
Ranunculales-FUL1	214	$w_B = 0.2476$; $w_F = 0.1846$	−29.090,34	27,84	***	214	$w_B = 0.0539$; $w_F = 0.0416$	−4.532,35	2,20		214	$w_B = 0.2016$; $w_F = 0.1385$	−12.239,44	17,65	***	214	$w_B = 0.5104$; $w_F = 0.4086$	−11.192,02	5,62	*
Eupteleaceae-EuplFL1	214	$w_B = 0.2159$; $w_F = 0.3013$	−29.103,61	1,30		214	$w_B = 0.0483$; $w_F = 0.0001$	−4.532,01	2,89		214	$w_B = 0.1694$; $w_F = 0.4769$	−12.245,50	5,53	*	214	$w_B = 0.4557$; $w_F = 0.4817$	−11.194,83	0,01	
Pa paveraceae-FUL1	214	$w_B = 0.2350$; $w_F = 0.1386$	−29.080,80	46,91	***	214	$w_B = 0.0563$; $w_F = 0.0122$	−4.521,08	24,74	***	214	$w_B = 0.1925$; $w_F = 0.0809$	−12.226,45	43,64	***	214	$w_B = 0.4759$; $w_F = 0.3760$	−11.193,00	3,67	
Fumarioideae-FUL1	214	$w_B = 0.2223$; $w_F = 0.1260$	−29.095,32	17,87	***	214	$w_B = 0.0504$; $w_F = 0.0053$	−4.528,08	10,75	***	214	$w_B = 0.1791$; $w_F = 0.0580$	−12.236,65	23,22	***	214	$w_B = 0.4584$; $w_F = 0.4159$	−11.194,73	0,21	
Papaveroideae-FUL1	214	$w_B = 0.2272$; $w_F = 0.1431$	−29.091,30	25,92	***	214	$w_B = 0.0527$; $w_F = 0.0155$	−4.527,17	12,57	***	214	$w_B = 0.1820$; $w_F = 0.0919$	−12.238,80	18,93	***	214	$w_B = 0.4713$; $w_F = 0.3621$	−11.193,11	3,44	
Menispermaceae-FUL1	214	$w_B = 0.2177$; $w_F = 0.1671$	−29.103,48	1,56		214	$w_B = 0.0486$; $w_F = 0.0179$	−4.532,44	2,02		214	$w_B = 0.1735$; $w_F = 0.0992$	−12.246,81	2,91		214	$w_B = 0.4580$; $w_F = 0.3009$	−11.194,47	0,74	
Lardizabalaceae-FUL1	214	$w_B = 0.2087$; $w_F = 0.3919$	−29.089,81	28,89	***	214	$w_B = 0.0415$; $w_F = 0.2277$	−4.517,81	31,27	***	214	$w_B = 0.1610$; $w_F = 0.4418$	−12.232,87	30,79	***	214	$w_B = 0.4533$; $w_F = 0.4965$	−11.194,72	0,23	
Lardizabalaceae-FUL1a	214	$w_B = 0.2114$; $w_F = 0.4631$	−29.092,38	23,75	***	214	$w_B = 0.0448$; $w_F = 0.2641$	−4.525,63	15,64	***	214	$w_B = 0.1634$; $w_F = 0.6704$	−12.231,83	32,87	***	214	$w_B = 0.4570$; $w_F = 0.4288$	−11.194,81	0,06	
Lardizabalaceae-FUL1b	214	$w_B = 0.2140$; $w_F = 0.3295$	−29.101,01	6,49	*	214	$w_B = 0.0445$; $w_F = 0.2206$	−4.525,91	15,08	***	214	$w_B = 0.1690$; $w_F = 0.2832$	−12.246,45	3,63		214	$w_B = 0.4530$; $w_F = 0.5652$	−11.194,52	0,64	
Berberidaceae-FUL1	214	$w_B = 0.2224$; $w_F = 0.1507$	−29.098,01	12,49	***	214	$w_B = 0.0462$; $w_F = 0.0653$	−4.532,75	1,40		214	$w_B = 0.1791$; $w_F = 0.0815$	−12.240,17	16,18	***	214	$w_B = 0.4689$; $w_F = 0.3356$	−11.192,92	3,84	
Ranunculaceae-FUL1	214	$w_B = 0.2197$; $w_F = 0.1968$	−29.103,36	1,79		214	$w_B = 0.0499$; $w_F = 0.0343$	−4.532,52	1,86		214	$w_B = 0.1744$; $w_F = 0.1520$	−12.247,71	1,11		214	$w_B = 0.4529$; $w_F = 0.4749$	−11.194,78	0,12	
Eupteleaceae-EuplFL2	214	$w_B = 0.2155$; $w_F = 0.3605$	−29.102,73	3,05		214	$w_B = 0.0468$; $w_F = 0.1902$	−4.531,34	4,22	*	214	$w_B = 0.1693$; $w_F = 0.4318$	−12.245,56	5,40	*	214	$w_B = 0.4542$; $w_F = 0.7561$	−11.194,46	0,75	
Ranunculales-FUL2	214	$w_B = 0.1893$; $w_F = 0.2865$	−29.079,86	48,79	***	214	$w_B = 0.0492$; $w_F = 0.0449$	−4.533,33	0,25		214	$w_B = 0.1317$; $w_F = 0.2825$	−12.214,86	66,81	***	214	$w_B = 0.3977$; $w_F = 0.6255$	−11.185,07	19,53	***
Pa paveraceae-FUL2	214	$w_B = 0.1990$; $w_F = 0.3002$	−29.086,45	35,62	***	214	$w_B = 0.0479$; $w_F = 0.0475$	−4.533,45	0,00		214	$w_B = 0.1496$; $w_F = 0.2833$	−12.230,50	35,52	***	214	$w_B = 0.4132$; $w_F = 0.6871$	−11.186,05	17,58	***
Fumarioideae-FUL2	214	$w_B = 0.2116$; $w_F = 0.3331$	−29.098,05	12,41	***	214	$w_B = 0.0477$; $w_F = 0.0512$	−4.533,44	0,02		214	$w_B = 0.1691$; $w_F = 0.2169$	−12.247,44	1,66		214	$w_B = 0.4391$; $w_F = 0.9332$	−11.189,22	11,24	***
Papaveroideae-FUL2	214	$w_B = 0.2065$; $w_F = 0.2851$	−29.095,76	17,00	***	214	$w_B = 0.0480$; $w_F = 0.0467$	−4.533,44	0,01		214	$w_B = 0.1552$; $w_F = 0.3002$	−12.233,85	28,84	***	214	$w_B = 0.4343$; $w_F = 0.6190$	−11.191,52	6,63	*

(Continued)

Table 1 | Continued

Model	Whole FUL sequence				MADS region				IK region				C region			
	p	Results	LnL	2ΔlnL (LRT)	p	Results	LnL	2ΔlnL (LRT)	p	Results	LnL	2ΔlnL (LRT)	p	Results	LnL	2ΔlnL (LRT)
Menispermaceae-FUL2	214	$w_B = 0.2178$ $w_F = 0.1628$	−29.103,34	1,84	214	$w_B = 0.0473$ $w_F = 0.0663$	−4.533,23	0,45	214	$w_B = 0.1731$ $w_F = 0.1092$	−12.247,26	2,01	214	$w_B = 0.4574$ $w_F = 0.3487$	−11.194,68	0,31
Ranunculaceae-FUL2	214	$w_B = 0.2119$ $w_F = 0.2785$	−29.100,74	7,03 **	214	$w_B = 0.0500$ $w_F = 0.0252$	−4.531,65	3,60	214	$w_B = 0.1604$ $w_F = 0.3307$	−12.237,24	22,04 ***	214	$w_B = 0.4519$ $w_F = 0.5084$	−11.194,62	0,43

*For each of the LRTs, twice the difference of log likelihood between the models (2 ΔlnL) was compared to critical values from a χ2 distribution, with degree of freedom equal to the differences in number of estimated parameters between models. The test was conducted for the entire dataset and also for each of the functional domains defined for MADS-box genes. These analyses were repeated on the M, IK, and C domains in order to evaluate whether there was a difference in their rates of evolution in different taxa, given their key roles in DNA binding (M), protein dimerization (IK) and multimerization (C). The color of the asterisks indicates whether the proteins show an increase in the degree of purifying selection (red), or a relaxed degree of purifying selection (black). Significance: *P < 0.05, **P < 0.01, ***P < 0.001.*

Ranunculales. This duplication was not recovered in previous analyses of the AP1/FUL gene lineage (Litt and Irish, 2003; Shan et al., 2007). Although these analyses suggested major duplications occurred in the *FUL-like* genes in Ranunculales, it was not clear when they occurred. Our analyses on an expanded sample of Ranunculales clearly established that there was a single major event very early in the diversification of the order, however, is still unclear whether this duplication occurred before or after the divergence of Eupteleaceae. In fact, low support within each major clade and high similarity between EUplFL1 and EUplFL2 suggest that an alternative topology to **Figure 3** tree would be possible, in which two independent duplications occurred, one within the Eupteleaceae and another after the divergence of the Eupteleaceae but before the diversification of all other Ranunculiids. This would be similar to the scenario found in the reconstruction of the evolutionary history of the *APETALA3* (*AP3*) genes in the Ranunculales, in which three duplications occurred: one in the Eupteleaceae and two in the remaining Ranunculales (Sharma et al., 2011). This indication that *FUL-like* and *AP3* genes underwent duplication events early in the diversification of most Ranunculales, before or right after the split of Eupteleaceae, suggests a possible ancestral genome-wide polyploidization event (Cui et al., 2006) in the Ranunculales, independent to the already well-established gamma-duplication in the core eudicots (Jiao et al., 2011; Vekemans et al., 2012).

In addition, whereas *RanFL1* genes are found in all the families of the order sampled so far, *RanFL2* genes were not found in Lardizabalaceae and Berberidaceae. This may be because in those two families our primers did not pick up *RanFL2* genes, or those genes are not expressed in leaf or floral tissue, or they were lost. None of these hypotheses can be rejected at this time, but after numerous amplification attempts with multiple degenerate primers specifically targeted to *RanFL2* genes, as well as extensive database searches, we favor the second and the third.

The clarification of orthology and paralogy of previously functionally characterized *FUL-like* genes sheds light on why these *FUL-like* genes might have both overlapping and unique functions (**Figure 1**). Our results show that *PapsFL2* and *EscaFL1* and *EscaFL2* are orthologs belonging to the *RanFL1* clade (**Figure 3**). On the other hand, *PapsFL1* is orthologous to *EscaFL3*, which was not discovered in previous studies in *E. californica* (**Figure 3**). These latter two genes belong to the *RanFL2* clade. These results suggest that the reason *escafl1-fl2* double mutants in *E. californica* did not show defects in cauline leaf development, flowering time and petal identity as did *papsfl1-fl2* mutants may be because *EscaFL3* is redundant for these functions (Pabón-Mora et al., 2012). Our results also confirm that the two *A. coerulea* FUL-like copies are the result of an independent duplication, as *AqcFL1A* and *AqcFL1B* are recent paralogs belonging to the *RanFL1* clade. *RanFL2* copies are not present in the *Aquilegia* genome. This gene loss may explain why results from functional analyses in poppies could not be extrapolated to *Aquilegia* (Pabón-Mora et al., 2012, 2013), and indeed probably suggests results from *Aquilegia* cannot even be applied to other members of Ranunculaceae. Gene loss in *Aquilegia* might have resulted in
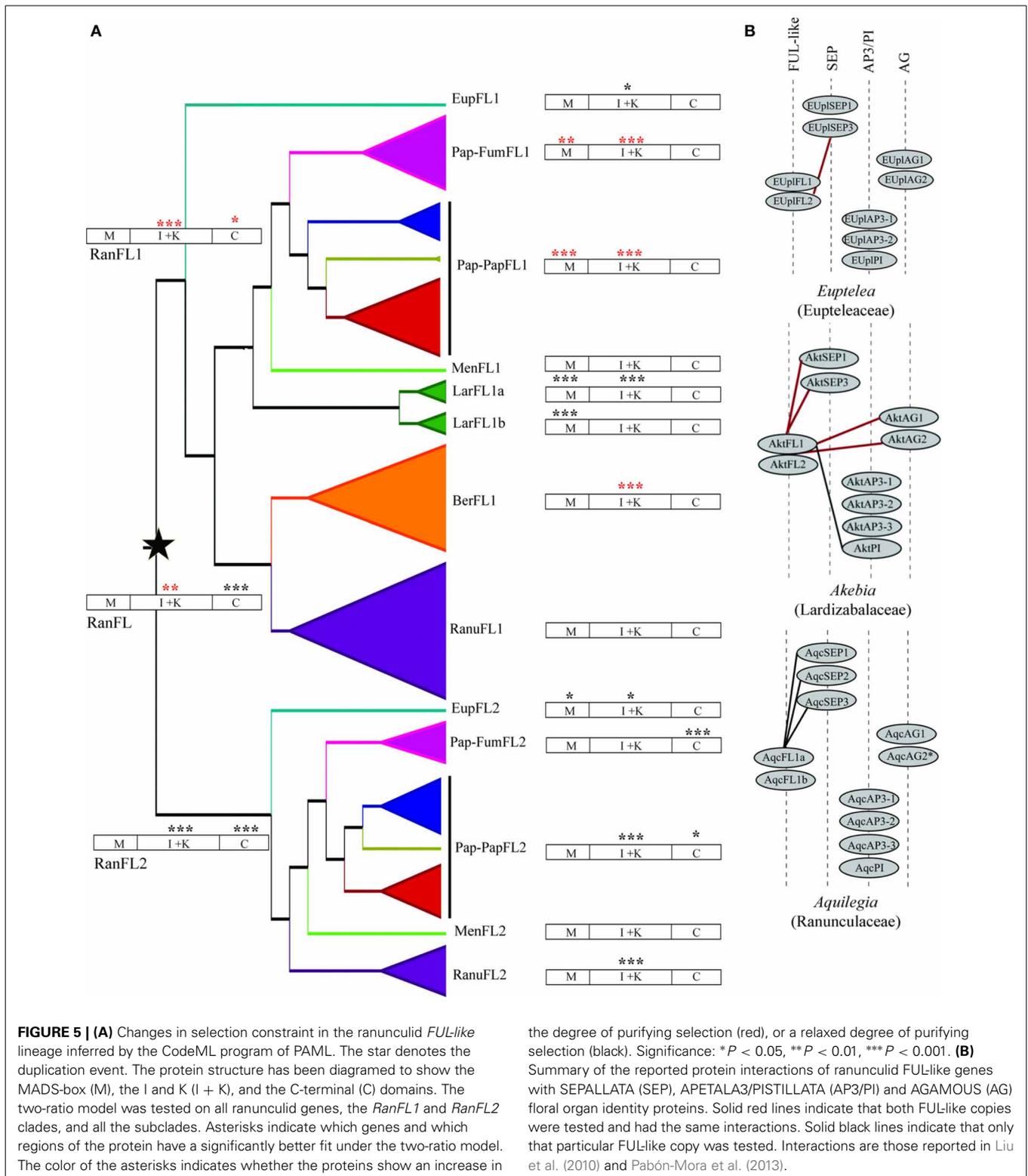
FIGURE 5 | (A) Changes in selection constraint in the ranunculid *FUL-like* lineage inferred by the CodeML program of PAML. The star denotes the duplication event. The protein structure has been diagramed to show the MADS-box (M), the I and K (I + K), and the C-terminal (C) domains. The two-ratio model was tested on all ranunculid genes, the *RanFL1* and *RanFL2* clades, and all the subclades. Asterisks indicate which genes and which regions of the protein have a significantly better fit under the two-ratio model. The color of the asterisks indicates whether the proteins show an increase in

the degree of purifying selection (red), or a relaxed degree of purifying selection (black). Significance: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. **(B)** Summary of the reported protein interactions of ranunculid FUL-like genes with SEPALLATA (SEP), APETALA3/PISTILLATA (AP3/PI) and AGAMOUS (AG) floral organ identity proteins. Solid red lines indicate that both FUL-like copies were tested and had the same interactions. Solid black lines indicate that only that particular FUL-like copy was tested. Interactions are those reported in Liu et al. (2010) and Pabón-Mora et al. (2013).

the rewiring of flower and fruit developmental networks such that *FUL-like* genes are excluded from roles in floral meristem identity, floral organ identity, or fruit development, and instead have been co-opted into leaf development. Nevertheless, it is

also possible that *AqcFL1* residual transcript, or redundancy with other transcription factors masked the roles of *AqcFL1* genes in flower and fruit development in previous experiments (Pabón-Mora et al., 2013).

SEQUENCE CHANGES IN THE C-TERMINAL DOMAIN RESULTED IN NEW MOTIFS THAT MIGHT PLAY ROLES IN ACTIVATION AND PROTEIN MULTIMERIZATION CAPABILITIES

We have shown that ranunculid FUL-like proteins have, at the beginning of the C terminal domain, glutamine-rich segments carrying from 3 to 9 consecutive glutamines (Q) and 3–4 non-consecutive glutamines. Glutamine-rich motifs are also found in grass FUL-like proteins (Preston and Kellogg, 2006), and glutamine-rich domains in plants, carrying from 4 to 20 repeats, have been known to behave as transcription activation domains (Gerber et al., 1994; Schwechheimer et al., 1998; Xiao and Jeang, 1998; Wilkins and Lis, 1999; Immink et al., 2009); this suggests that FUL-like proteins may have transcription activation capability similar to euAP1 proteins (Cho et al., 1999). However, AqFL1A and AqFL1B (with 2 consecutive and 2 non-consecutive Q), as well as PapsFL1 and PapsFL2 (both with 4 consecutive Q) have not been shown to auto-activate in yeast systems (Pabón-Mora et al., 2012, 2013). Other ranunculid FL proteins, like those of *Eschscholzia*, have a larger number of glutamines but have not yet been tested for transcription activation capability. Glutamine repeats in eukaryotes have also been hypothesized to behave as "polar zippers" in protein-protein interactions (Perutz et al., 1994; Michleitsch and Weissman, 2000), thus these regions might mediate strength and specificity of FUL-like protein interactions.

This study identified two additional protein regions conserved in ranunculid FUL-like proteins including the sequence QNS-P/LS/TFLLSQSE/LP-SLN/TI, and a negatively charged region rich in glutamic acid (E) before the conserved FUL-motif LMPPWML (**Figure 2**). There are no functional studies specific for these regions, however, it has been shown that the N/SS at positions 227–228 are consistently found in AP1/FUL proteins and shared with SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) and some SEPALLATA proteins, and that mutations in these amino acids influence interaction specificity and can result in changes in protein partners (Van Dijk et al., 2010).

RELEASE OF PURIFYING SELECTION IN THE I+K PROTEIN DOMAINS MIGHT HAVE INFLUENCED FUNCTIONAL DIVERSIFICATION

Variation in the rates of evolution of different FUL-like protein regions may also explain the functional differences among characterized proteins in different species. This is based on the premise that the rate of amino acid substitution is limited by functional or structural constraints on proteins (Liu et al., 2008). Previous studies have shown that differences in the rates and patterns of molecular evolution appear to be associated with divergence of developmental function between paralogous MADS-box loci (Lawton-Rauh et al., 1999). A common way to measure changes in protein sequence evolution is the dN/dS ratio, which calculates the ratio of non-synonymous to synonymous changes in protein sequences and provides an estimate of selective pressure. A dN/dS < 1 suggests that strong purifying selection has not allowed for fixation of most amino acid substitutions, dN/dS > 1 suggests that constraints are reduced and new amino acids have been able to become fixed due to positive selection, and dN/dS = 1 suggests neutral evolution, in which synonymous changes occur at the same rate as non-synonymous changes and fixation of new amino acids occurs at a neutral rate (Li, 1997; Hurst, 2002).

Our results show that strong purifying selection can be detected in the RanFL1 clade compared to more relaxed purifying selection in the RanFL2 proteins ($p < 0.001$). This would suggest that RanFL2 proteins are evolving at a faster rate, having been released from strong purifying selection after the duplication, and suggests a scenario of long-term maintenance of ancestral functions in one clade (RanFL1) and sub or neo-functionalization in the other clade (RanFL2), (Aagaard et al., 2006). When the same analyses are applied to the subclades within *RanFL1* and *RanFL2*, this pattern can also be seen for the duplicates in Papaveraceae s.l. and Ranunculaceae, but not in other families. For instance a contradictory pattern is found in Lardizabalaceae, in which both FL1a and FL1b proteins (paralogous clades within RanFL1) show relaxed purifying selection, suggesting that within this family, ancestral *FUL*-like gene functions may have been redistributed among the paralogs or lost, with the potential for new functions to appear in the evolutionary process (Force et al., 1999; Conant and Wagner, 2002).

Our analyses also showed that relaxation in purifying selection occurred preferentially in the I + K domains (in Eupteleaceae FL1, FL2, Lardizabalaceae FL1a, FL1b, Papaveraceae s str. FL2 and Ranunculaceae FL2), where dimerization functions have been localized, and less frequently in the MADS domain (in Lardizabalaceae FL1 a and FL1b), important for DNA binding, and the C terminus (in Papaveraceae s str. FL2), the function of which is not known. Most protein motifs maintained in MADS box duplicates and involved in dimerization occur at a hot-spot at the junction between the MADS and the I domain and is clear that non-synonymous changes in this region can dramatically change protein interactions (Van Dijk et al., 2010). For instance, three spots between the MADS and the I domain are maintained in most MADS box proteins and are thought to control DNA binding, these include Alanine A57, Asparagine N60 and Methionine M61 (Van Dijk et al., 2010). In FUL-like proteins the A57 is replaced by another hydrophobic amino-acid more often Tyrosine Y or Phenylalanine F, the M61 appears in position M63 and is conserved in all sequences, and finally the hydrophobic N60 is maintained in Ranunculaceae FL2, but changed in the rest of RanFL2 and RanFL1 proteins for Aspartic Acid D. The importance of the IK domains in protein-protein interactions has been long recognized. For instance, the end of the I domain and the entire K domain have been identified as the most important regions for the interactions between FUL-like and SEPALLATA proteins in rice (Moon et al., 1999). Likewise, residues in position 148–158 in APETALA1 seem to be crucial for recovery of floral meristem identity (Alvarez-Buylla et al., 2006) and a point mutation in Y148N is known to cause the loss of interaction between AP1 and SEPALLATA4, AGAMOUS-Like6 and AGAMOUS-Like15 (Van Dijk et al., 2010). Altogether the data suggests that changes in the IK regions might be key in explaining the different functions reported in ranunculid FUL-like proteins *via* changes in protein interactions. This is in agreement with observations in paralogous regulatory genes in which relaxed purifying selection is associated with the partitioning or even the acquisition of new interacting protein partners compared to the ancestral (pre-duplication) protein interactions (Dermitzakis and Clark, 2001; see also He and Zhang, 2006; Wagner and Zhang, 2011).

A comparison of protein-protein interaction data gathered from ranunculid FUL-like proteins and the outgroup Poaceae proteins partially supports this hypothesis. Protein interactions in grasses show that *Oryza sativa* FUL-like proteins OsMADS14, OsMADS15 and OsMADS18 can only interact with a narrow set of floral organ identity proteins, the SEPALLATA proteins (Moon et al., 1999). Similarly, the *Euptelea* FUL-like proteins (EuplFL1 and EuplFL2) only interact with SEPALLATA proteins (Liu et al., 2010). The same interactions with floral organ identity proteins have been recorded for *Aquilegia* (AqFL1a) FUL-like proteins (Pabón-Mora et al., 2013), under strong purifying selection. In contrast, *Akebia* (Lardizabalaceae) FUL-like proteins, under relaxed purifying selection, appear to have been able to expand the repertoire of protein partners and can interact with SEPALLATA, PISTILLATA and AGAMOUS orthologs (Liu et al., 2010). Clearly more data are required to test the hypothesis that Ranunculales FUL-like protein interactions are maintained under strong purifying selection but diverge under relaxed selection, with resulting diversification of functional outcomes (**Figure 5B**).

The data presented here and in previous publications (Pabón-Mora et al., 2012, 2013) allow us to hypothesize that: (1) *FUL-like* genes across ranunculids perform overlapping and unique roles in a manner that cannot be predicted by their expression patterns. (2) Variation in function is possibly due to key amino acid changes in the I and K domains, important in dimerization, as well as unique protein motifs in the C-domain likely important for multimerization. In combination, these might have provided FUL-like homologs in the Ranunculales with different biochemical capabilities and protein interactions. (3) Understanding the evolution of gene pleiotropy in terms of protein regions that might be important for different functions in pre-duplication *FUL-like* genes across basal eudicots, provides clues on how *FUL-like* genes might have taken on different roles. Future

directions include expression analyses and functional characterization of *FUL-like* genes in other Ranunculales, tests on the protein interactions between FUL-like proteins and other floral organ identity proteins in different ranunculid taxa, and functional characterization of the conserved motifs, particularly at the IK domains and the C-terminus.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Plant_Evolution_and_Development/10.3389/fpls.2013.00358/abstract

Figure S1 | K-domain sequence alignment of ranunculid FUL-like proteins.
Hydrophobic amino-acids in the a and d positions in the heptad repeats (abcdefg)$_n$ are in bold. The predicted protein sequence at this domain contains three amphipathic α-helices: K1, K2, and K3. Within K1, positions 99 (E), 102 (K), 104 (K) are conserved in all ranunculid sequences and the outgroup, except for Mencan1 y Mencan2. Similarly, positions 106 (K), 108 (E) are also conserved, except in RocoFL2, ArmeFL4. Finally 111 (Q) is also conserved except in MacoFL3, MacoFL4. Within K2 positions 119 (G), 128 (K), 129 (E), 134 (E), 136 (Q) are conserved except in ArmeFL3. Conserved hydrophobic amino-acids outside of the predicted helices are highlighted and labeled with h.

Table S1 | Accession numbers of *FUL-like* sequences used in this study.

REFERENCES

Aagaard, J. E., Willis, J. H., and Phillips, P. C. (2006). Relaxed selection among duplicate floral regulatory genes in Lamiales. *J. Mol. Evol.* 63, 493–503. doi: 10.1007/s00239-005-0306-x

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1006/jmbi.1990.9999

Alvarez-Buylla, E. R., García-Ponce, B., and Garay-Arroyo, A. (2006). Unique and redundant functional domains of APETALA1 and CAULIFLOWER, two recently duplicated Arabidopsis thaliana floral MADS-box genes. *J. Exp. Bot.* 57, 3099–3107. doi: 10.1093/jxb/erl081

APG. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1111/j.1095-8339.2009.00996.x

Barker, M. S., Vogel, H., and Schranz, M. E. (2009). Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biol. Evol.* 1, 391–399. doi: 10.1093/gbe/evp040

Bemer, M., Karlova, R., Ballester, A. R., Tikunov, Y. M., Bovy, A. G., Wolters-Arts, M., et al. (2012). The tomato FRUITFULL homologs TDR4/FUL1 and MBP7/FUL2 regulate ethylene-independent aspects of fruit ripening. *Plant Cell* 24, 4437–4451. doi: 10.1105/tpc.112.103283

Benlloch, R., d'Erfurth, I., Ferrandiz, C., Cosson, V., Beltrán, J. P., Cañas, L. A., et al. (2006). Isolation of *mtpim* proves *Tnt1* a useful reverse genetics tool in *Medicago truncatula*

and uncovers new aspects of *AP1*-like functions in legumes. *Plant Physiol.* 142, 972–983. doi: 10.1104/pp.106.083543

Berbel, A., Ferrandiz, C., Hecht, V., Dalmais, M., Lund, O. S., Sussmilch, F. C., et al. (2012). *VEGETATIVE1* is essential for development of the compound inflorescence in pea. *Nat. Commun.* 3 797. doi: 10.1038/ncomms1801

Berbel, A., Navarro, C., Ferrandiz, C., Cañas, L. A., Madueño, F., and Beltrán, J. (2001). Analysis of PEAM4, the pea AP1 functional homologue, supports a model for AP1-like genes controlling both floral meristem and floral organ identity in different plant species. *Plant J.* 25, 441–451. doi: 10.1046/j.1365-313x.2001.00974.x

Blanc, G., Hokamp, K., and Wolfe, K. H. (2003). A recent polyploidy superimposed on older large scale duplications in the *Arabidopsis* genome. *Genome Res.* 13, 137–144. doi: 10.1101/gr.751803

Blanc, G., and Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell* 16, 1679–1691. doi: 10.1105/tpc.021410

Bowers, J. L., Chapman, B. A., Rong, J., and Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438. doi: 10.1038/nature01521

Bowman, J., Alvarez, J., Weigel, D., Meyerowitz, E. M., and Smyth, D. R. (1993). Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* 119, 721–743.

Burko, Y., Shleizer-Burko, S., Yanai, O., Shwartz, I., Zelnik, I. D., Jacob-Hirsch, J., et al. (2013). A role for APETALA1/FRUITFULL transcription factors in tomato leaf development. *Plant Cell* 25, 2070–2083. doi: 10.1105/tpc.113.113035

Castillejo, C., Romera-Branchat, M., and Pelaz, S. (2005). A new role of

the Arabidopsis SEPALLATA3 gene revealed by its constitutive expression. *Plant J.* 43, 586–596. doi: 10.1111/j.1365-313X.2005.02476.x

Cho, S., Jang, S., Chae, S., Chung, K. M., Moon, Y., An, G., et al. (1999). Analysis of the C-terminal region of *Arabidopsis thaliana APETALA1* as a transcription activation domain. *Plant Mol. Biol.* 40, 419–429. doi: 10.1023/A:1006273127067

Coen, E., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0

Conant, G., and Wagner, A. (2002). Asymmetric sequence divergence of duplicate genes. *Genome Res.* 13, 2052–2058. doi: 10.1101/gr.1252603

Cui, L., Wall, P. K., Leebens-Mack, J., Lindsay, B. G., Soltis, D. E., and Doyle, J. J. (2006). Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749. doi: 10.1101/gr.4825606

Dermitzakis, E., and Clark, A. (2001). Differential selection after duplication in mammalian developmental genes. *Mol. Biol. Evol.* 18, 557–562. doi: 10.1093/oxfordjournals.molbev.a003835

Ferrándiz, C., Gu, Q., Martienssen, R., and Yanofsky, M. F. (2000). Redundant regulation of meristem identity and plant architecture by *FRUITFULL, APETALA1* and *CAULIFLOWER. Development* 127, 725–734.

Force, A., Lynch, M., Pickett, F., Amores, A., Yan, Y., and Postlehwait, J. (1999). Preservation of duplicate genes by complementary degenerative mutations. *Genetics* 151, 1531–1545.

Gerber, H. P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S., et al. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 263, 808–811. doi: 10.1126/science.8303297

Gu, Q., Ferrandiz, C., Yanofsky, M. F., and Martienssen, R. (1998). The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125, 1509–1517.

Gustafson-Brown, C., Savidge, B., and Yanofsky, M. F. (1994). Regulation of the Arabidopsis floral homeotic gene *APETALA1. Cell* 76, 131–143. doi: 10.1016/0092-8674(94)90178-3

He, X., and Zhang, J. (2006). Toward a molecular understanding of pleiotropy. *Genetics* 173, 1885–1891. doi: 10.1534/genetics.106.060269

Huijser, P., Klein, J., Lonnig, W-E., Meijer, H., Saedler, H., and Sommer, H. (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene squamosa in *Antirrhinum majus. EMBO J.* 11, 1239–1249.

Hurst, L. D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genetics* 18, 486. doi: 10.1016/S0168-9525(02)02722-1

Immink, R. G. H., Hannapel, D. J., Ferrario, S., Busscher, M., Franken, J., Lookeren-Campagne, M. M., et al. (1999). A petunia MADS box gene involved in the transition from vegetative to reproductive development. *Development* 126, 5117–5126.

Immink, R. G. H., Tonaco, I. A. N., deFolter, S., Shchennikova, A., vanDijk, A. D. J., Busscher-Lange, J., et al. (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biol.* 10:R24. doi: 10.1186/gb-2009-10-2-r24

Irish, V. F., and Sussex, I. M. (1990). Function of the *apetala-1* gene during *Arabidopsis* floral development. *Plant Cell* 2, 741–753. doi: 10.2307/3869173

Jaakola, L., Poole, M., Jones, M. O., Kamarainen-Karppinen, T., Koskimaki, J. J., Hohtola, A., et al. (2010). A SQUAMOSA MADS box gene is involved in the regulation of anthocyanin accumulation in bilberry fruits. *Plant Physiol.* 153, 1619–1629. doi: 10.1104/pp.110.158279

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916

Joint Genome Institute. (2010). *Phytozome v6.0.* Avilable online at: http://www.phytozome.net/

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on a fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kempin, S., Savidge, B., and Yanofsky, M. F. (1995). Molecular basis of the cauliflower phenotype in Arabidopsis. *Science* 27, 522–525. doi: 10.1126/science.7824951

Kobayashi, K., Yasuno, N., Sato, Y., Yoda, M., Yamazaki R., Kimizu, M., et al. (2012). Inflorescence meristem identity in rice is specified by overlapping functions of three AP1/FUL-like MADS box genes and PAP2, a SEPALLATA MADS box

gene. *Plant Cell* 24, 1848–1859. doi: 10.1105/tpc.112.097105

Lawton-Rauh, A. L., BucklerIV, ES, and Purugganan, M. D. (1999). Patterns of molecular evolution among paralogous floral homeotic genes. *Mol. Biol. Evol.* 16, 1037–1045. doi: 10.1093/oxfordjournals.molbev.a026193

Li, W.-H. (1997). *Molecular Evolution.* Sunderland, MA: Sinauer Associates.

Litt, A., and Irish, V. F. (2003). Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165, 821–833.

Liu, C., Zhang, J., Zhang, N., Shan, H., Su, K., Zhang, J., et al. (2010). Interactions among proteins of floral MADS-box genes in basal eudicots: implications for evolution of the regulatory network for flower development. *Mol. Biol. Evol.* 27, 1598–1611. doi: 10.1093/molbev/msq044

Liu, J., Zhang, Y., Leu, X., and Zhang, Z. (2008). Natural selection of protein structural and functional properties: a single nucleotide polymorfism perspective. *Genome Biol.* 9, R69. doi:10.1186/gb-2008-9-4-r69

Maddison, D. R., and Maddison, W. P. (2005). MacClade 4: analysis of phylogeny and character evolution. *Version* 4.08a. Avilable online at: http://macclade.org

Mandel, M. A., and Yanofsky, M. F. (1995). The Arabidopsis AGL8 MADS box gene is expressed in inflorescence meristems and is negatively regulated by *APETALA1. Plant Cell* 7, 1763–1771.

Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., and Beeckman, T. (2008). Flowering-time genes modulate meristem determinacy and growth form in Arabidopsis thaliana. *Nat. Genetics* 40, 1489–1492. doi: 10.1038/ng.253

Michleitsch, M., and Weissman, J. (2000). A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11910–11915. doi: 10.1073/pnas.97.22.11910

Miller, M. A., Holder, H. T., Vos, R., Midford, P. E., Liebowitz, T., Chan, L., et al. (2009). *The CIPRES portals.* Avilable online at: http://www.phylo.org

Moon, Y.-H., Hang, H.-G., Jung, J.-Y., Jeon, J.-S., Sung, S.-K., and An, G. (1999). Determination of the motif responsible for interaction between the rice *APETALA1/*

AGAMOUS-LIKE9 family proteins using a yeast two-hybrid System. *Plant Physiol.* 120, 1193–1203. doi: 10.1104/pp.120.4.1193

Müller, B. M., Saedler, H., and Zachgo, S. (2001). The MADS-box gene DEFH28 from Antirrhinum is involved in the regulation of floral meristem identity and fruit development. *Plant J.* 28, 169–179. doi: 10.1046/j.1365-313X.2001.01139.x

Murai, K., Miyamae, M., Kato, H., Takumi, S., and Ogihara, Y. (2003). *WAP1*, a wheat *APETALA1* homolog, plays a central role in the phase transition from vegetative to reproductive growth. *Plant Cell Physiol.* 44, 1255–1265. doi: 10.1093/pcp/pcg171

Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437. doi: 10.1146/annurev.genet.34.1.401

Pabón-Mora, N., Ambrose, B. A., and Litt, A. (2012). Poppy APETALA1/FRUITFULL orthologs control flowering time, branching, perianth identity, and fruit development. *Plant Physiol.* 158, 1685–1704. doi: 10.1104/pp.111.192104

Pabón-Mora, N., Sharma, B., Holappa, L., Kramer, E., and Litt, A. (2013). The *Aquilegia* FRUITFULL-like genes play key roles in leaf morphogenesis and inflorescence development. *Plant J.* 74, 197–212. doi: 10.1111/tpj.12113

Parenicová, L., deFolter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., et al. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15, 1538–1551. doi: 10.1105/tpc.011544

Perutz, M., Johnson, T., Suzuki, M., and Finch, J. (1994). Glutamine repeats as polar zippers: their possible role in inherited neurodegerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* 91, 5355–5358. doi: 10.1073/pnas.91.12.5355

Posada, D., and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817

Preston, J. C., and Kellogg, E. A. (2006). Reconstructing the evolutionary history of paralogous *APETALA1/FRUITFULL*-like genes in grasses (Poaceae). *Genetics* 174, 421–437. doi: 10.1534/genetics.106.057125

Purugganan, M. D., and Suddith, J. I. (1998). Molecular population

genetics of the Arabidopsis CAULIFLOWER regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8130–8134. doi: 10.1073/pnas.95.14.8130

Schwechheimer, C., Smith, C., and Bevan, M. W. (1998). The activities of acidic and glutamine-rich transcriptional activation domains in plant cells: design of modular transcription factors for high level expression *Plant Mol. Biol.* 36, 195–204. doi: 10.1023/A:1005990321918

Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z., et al. (2007). Patterns of gene duplication and functional diversification during the evolution of the AP1/SQUA subfamily of plant MADS-box genes. *Mol. Phylogenet. Evol.* 44, 26–41. doi: 10.1016/j.ympev.2007.02.016

Sharma, B., Guo, C., Kong, H., and Kramer, E. M. (2011). Petal-specific subfunctionalization of an APETALA3 paralog in the Ranunculales and its implications for petal evolution. *New Phytol.* 191, 870–883. doi: 10.1111/j.1469-8137.2011.03744.x

Smaczniak, C., Immink, R. G. H., Muiño, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., et al. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc. Natl. Acad.*

Sci. U.S.A. 109, 1560–1565. doi: 10.1073/pnas.1112871109

Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A fast bootstrapping algorithm for the RAxML web-servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642

Torti, S., Fornara, F., Vincent, C., Andrés, F., Nordström, K., Göbel, U., et al. (2012). Analysis of the *Arabidopsis* shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. *Plant Cell* 24, 444–462. doi: 10.1105/tpc.111.092791

Van Dijk, A. D. J., Morabito, G., Fiers, M., van Ham, RCHJ., Angenent, G. C., and Immink, R. G. H. (2010). Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. *PLOS Comput. Biol.* 6:e1001017. doi:10.1371/journal.pcbi.1001017

Vekemans, D., Proost, S., Vanneste, K., Coenen, H., Viaene, T., Ruelens, P., et al. (2012). Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. *Mol. Biol. Evol.* 29, 3793–3806. doi: 10.1093/molbev/mss183

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., et al. (2002). A MADS-box gene necessary for fruit ripening at the tomato

ripening-inhibitor (rin) locus. *Science* 296, 343–346. doi: 10.1126/science.1068181

Wagner, G. P., and Zhang, J. (2011). The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* 12, 204–213. doi: 10.1038/nrg2949

Wang, W., Lu, A.-M., Ren, Y., Endress, M. E., and Chen, Z.-D. (2009). Phylogeny and classification of Ranunculales: evidence from four molecular loci and morphological data. *Perspect. Plant Ecol. Evol. Syst.* 11, 81–110. doi: 10.1016/j.ppees.2009.01.001

Wilkins, R., and Lis, J. (1999). DNA distortion and multimerization: novel functions of the glutamine-rich domain of gaga factor. *J. Mol. Biol.* 285, 515–525. doi: 10.1006/jmbi.1998.2356

Xiao, H., and Jeang, K. (1998). Glutamine-rich domains activate transcription in yeast saccharomyces cerevisiae. *J. Biol. Chem.* 273, 22873–22876. doi: 10.1074/jbc.273.36.22873

Yalovsky, S., Rodríguez-Concepción, M., Bracha, K., Toledo-Ortiz, G., and Gruissem, W. (2000). Prenylation of the floral transcription factor *APETALA1* modulates its function. *Plant Cell* 12, 1257–1266. doi: 10.2307/3871127

Yang, Y. Z., Fanning, L., and Jack, T. (2003). The K domain mediates heterodimerization of the Arabidopsis floral organ identity porteins, APETALA3 and

PISTILLATA. *Plant J.* 33, 47–59. doi: 10.1046/j.0960-7412.2003.01473.x

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.