Annals of Botany **121**: 1211–1230, 2018
doi: 10.1093/aob/mcy007, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

Evolution of genes associated with gynoecium patterning and fruit development in Solanaceae

Clara Inés Ortiz-Ramírez[1,2], Sayonara Plata-Arboleda[1] and Natalia Pabón-Mora[1,*]

[1]Instituto de Biología, Universidad de Antioquia, Medellín, Colombia and [2]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas–Universidad Politécnica de Valencia, Valencia, Spain
**For correspondence. E-mail lucia.pabon@udea.edu.co*

Received: 8 November 2017 Returned for revision: 14 December 2017 Editorial decision: 5 January 2018 Accepted: 16 January 2018
Published electronically 17 February 2018

• **Background and Aims** The genetic basis of fruit development has been extensively studied in *Arabidopsis*, where major transcription factors controlling valve identity (i.e. *FRUITFULL*), replum development (i.e. *REPLUMLESS*) and the differentiation of the dehiscence zones (i.e. *SHATTERPROOF*, *INDEHISCENT* and *ALCATRAZ*) have been identified. This gene regulatory network in other flowering plants is influenced by duplication events during angiosperm diversification. Here we aim to characterize candidate fruit development genes in the Solanaceae and compare them with those of Brassicaceae.
• **Methods** *ALC/SPT*, *HEC/IND*, *RPL* and *AG/SHP* homologues were isolated from publicly available databases and from our own transcriptomes of *Brunfelsia australis* and *Streptosolen jamesonii*. Maximum likelihood phylogenetic analyses were performed for each of the gene lineages. Shifts in protein motifs, as well as expression patterns of all identified homologues, are shown in dissected floral organs and fruits in different developmental stages of four Solanaceae species exhibiting different fruit types.
• **Key Results** Each gene lineage has undergone different duplication time-points, resulting in very different genetic complements in the Solanaceae when compared with the Brassicaceae. In general, Solanaceae species have more copies of *HEC1/2* and *RPL* than Brassicaceae, have fewer copies of *SHP* and the same number of copies of *AG*, *ALC* and *SPT*. Solanaceae lack *IND* orthologues, but have pre-duplication *HEC3* homologues. The expression analyses showed opposite expression of *SPT* and *ALC* orthologues between dry- and fleshy-fruited species during fruit maturation. Fleshy-fruited species turn off *RPL* and *SPT* orthologues during maturation.
• **Conclusions** The gynoecium patterning and fruit developmental genetic network in the Brassicaceae cannot be directly extrapolated to the Solanaceae. In Solanaceae *ALC*, *SPT* and *RPL* contribute differently to maturation of dry dehiscent and fleshy fruits, whereas *HEC* genes are not generally expressed in the gynoecium. *RPL* genes have broader expression patterns than expected.

Key words: *ALCATRAZ*, berries, *bHLH*, capsules, fruit development, Solanaceae, *SPATULA*, *REPLUMLESS*.

INTRODUCTION

With close to 3000 species and many major edible crops, as well as ornamental and even extremely toxic species, the Solanaceae is a plant family at the core of human cuisine and addictions, as well as at the centre of historical genetic research breakthroughs (Särkinen *et al.*, 2013; Gebhardt, 2016). Most Solanaceae species possess a bicarpellate syncarpous gynoecium joined at the septum with axile placentation that produces either dry dehiscent or fleshy fruits (Knapp, 2002; Pabón-Mora and Litt, 2011). Other, less frequent fruit types have been reported, including drupes, pyrenes and mericarps (Knapp, 2002; Wang *et al.*, 2015). Optimization of fruit types on recent phylogenetic hypotheses in the family results in the occurrence of dry fruits, both indehiscent and dehiscent, in early-diverging subfamilies (including Goetzeoideae, Schwenckieae, Petunieae, Cestroideae and Nicotianoideae) and a major shift to predominantly fleshy fruits in later-diverging Solanoideae members (Knapp, 2002; Fig. 1). Despite the range of fruit types present in the Solanaceae, extensive attention has been given to early histogenesis and morphogenesis, as well as the hormonal shifts during fruit maturation in tomato, to the point where it has become the most important model system for climacteric fruits (Tanksley *et al.*, 2004; Pesaresi *et al.*, 2014). By comparison, little is known about carpel-to-fruit transformations and the genetic underpinnings of dry dehiscent fruits in other Solanaceae.

The fruit genetic regulatory network was first identified in the model *Arabidopsis thaliana* and has served as reference for comparative studies in fruit crops. *Arabidopsis* possesses a dry dehiscent fruit where the carpel walls form the valves and the septum differentiates into a medial and a lateral zone, of which the outer portion becomes the replum. In between the two, at the valve margins, two layers form the dehiscence zone; the one closer to the valves becomes a lignified layer, and adjacent to it, closer to the replum, there is a separation layer that disintegrates during fruit development to allow dehiscence (Ferrándiz, 2002). Proper valve development is ensured by the MADS-box transcription factor FRUITFULL (FUL), and replum identity is the result of the maintained

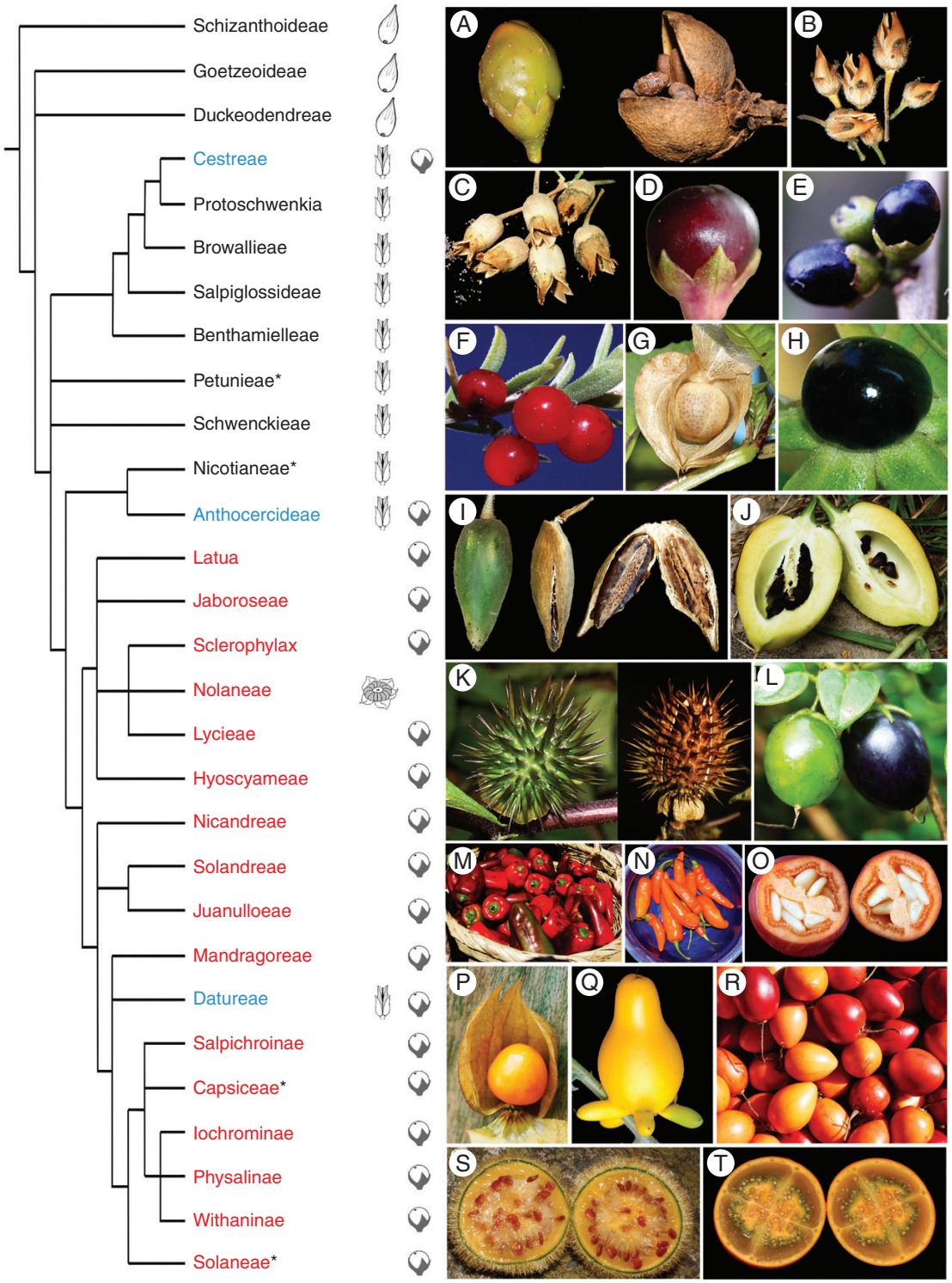

FIG. 1. Fruit diversity in the Solanaceae. (Left) Current phylogenetic circumscription off Solanaceae redrawn after Knapp (2002), Olmstead *et al.* (2008) and Särkinen *et al.* (2013), with drupes, dry dehiscent fruits, berries and mericarps drawn next to the recognized tribes. (Right) (A) *Brunfelsia australis* (Petunieae) early (left) and late (right) fruit developmental stages until dehiscence. (B) Mature dry dehiscent fruits of *Petunia hybrida* 'Mitchell' (Petunieae) (C) Mature dry dehiscent fruits of *Nicotiana sylvestris* (Nicotianeae). (D) Ripe fruit of *Cestrum elegans* (Cestreae). (E) Ripe fruits of *Cestrum cuneifolium* (Cestreae). (F) Ripe fruits of *Lycium chilense* (Lycieae). (G) Ripe fruit of *Nicandra physalodes* (Nicandreae). (H) Ripe fruit of *Atropa belladonna* (Hyoscyameae). (I) *Brugmansia aurea* (Datureae) from early development until fruit dehiscence. (J) *Brugmansia* sp. (Datureae) during early fruit development. (K) *Datura stramonium* (Datureae) during fruit development until dehiscence. (L) Maturing fruits of *Salpichroa tristis* (Salpichroinae). (M–O) Three varieties of *C. annuum* (Capsiceae) after fruit ripening. (P) Ripe fruit of *Physalis peruviana* (Physalinae). (Q) Ripe fruit of *Solanum mamosum* (Solaneae). (R) Ripe fruits of *Solanum betaceum* (Solaneae). (S) Transverse section of *Solanum pseudolulo* (Solaneae) ripen fruits. (T) Transverse section of a *Solanum quitoense* (Solaneae) ripen fruit. Asterisks indicate phylogenetic positions of species selected for expression analyses. Photo credits: (A–C, M–R, T) N. Pabón-Mora; (D, F, G, H) plantsystematics.org; (E, I, J, K, L, S) F. González.

expression of the homeodomain (HD) REPLUMLESS (RPL) protein (Gu *et al.*, 1998; Roeder *et al.*, 2003; Ferrándiz and Fourquin, 2014). FUL and RPL act as repressors of the MADS-box SHATTERPROOF proteins (SHP1 and SHP2) towards the valve margin, which in turn are responsible for the downstream activation of the bHLH genes, *ALCATRAZ* (*ALC*) in the separation layer and *INDEHISCENT* (*IND*) in the lignified layer (Liljegren *et al.*, 2000, 2004; Rajani and Sundaresan, 2001; Kay *et al.*, 2013; Girin *et al.*, 2010, 2011). Tension generated between these two layers during fruit maturation results in fruit dehiscence at the valve margin, leaving the replum intact, and the seeds attached to it eventually disperse out. Finally, regulating the entire genetic network is a member of the *APETALA2/Ethylene Responsive Factor* gene lineage, APETALA2 (AP2), which has been recently identified as an upstream repressor of RPL and SHP (Ripoll *et al.*, 2011). Evo-devo studies in other Brassicaceae members have indicated that putative shifts in upstream major regulators, like AP2, can affect the activation of the valve margin identity genes, likely resulting in shifts from dehiscent to indehiscent fruits (Mühlhausen *et al.*, 2013).

Most of the key regulatory genes of fruit development identified in *Arabidopsis* have also been studied in tomato. At the top of the regulatory network, *SlAP2*, the orthologue of *AP2* in tomato, is known to be a repressor of ripening, as *Slap2* fruits show premature ripening compared with the wild type (Chung *et al.*, 2010). In the second tier of genetic regulation, tomato has two *FRUITFULL* orthologues (of a total of four found), *SlFUL1* (also named *TDR4*) and *SlFUL2* (also named *MBP7*) that are known to promote ripening during fruit development (Bemer *et al.*, 2012). Such regulation is accomplished in part by the interactions between *SlFUL1/2* with *RIPENING INHIBITOR* (*RIN*, the *SEPALLATA4* orthologue) and *RIN* targets (Leseberg *et al.*, 2008; Martel *et al.*, 2011). *RIN* is exclusively expressed in fruits and is known to partially control fruit ripening in climacteric fruits (Vrebalov *et al.*, 2002; Ito *et al.*, 2017). Shifts in *SlAP2* expression in *Slful1/2* mutants suggest that during early fruit development *SlFUL1* and *SlFUL2* act to repress *SlAP2*, as *SlAP2* levels increase in *Slful1/2* mutants (Bemer *et al.*, 2012; Fujisawa *et al.*, 2014). In addition, *Slful1/2* double mutants are quite similar phenotypically to the *tagl1* (*SHP* orthologue) mutant, which also displays ripening defects as well as reduction in the number of pericarp layers, at least in the cultivar 'Alisa Craig' (Itkin *et al.*, 2009; Vrebalov *et al.*, 2009; Pan *et al.*, 2010).

Comparative data in dry dehiscent fruits of Solanaceae are only available for the orthologues of *FUL* and *SHP* in *Nicotiana*. Overexpression of *NtFUL* in *Nicotiana sylvestris* and downregulation of *NbSHP* in *Nicotiana benthamiana* result in indehiscent fruits, both lacking a functional dehiscence zone (Smykal *et al.*, 2007; Fourquin and Ferrándiz, 2012). These observations suggest that FUL-SHP is a genetic switch lying at the core of fruit development and likely evolution (Fourquin and Ferrándiz, 2012). Nevertheless, nothing is known about the *RPL* homologues, or the role of genes downstream of *SHP* homologues, either *NtSHP* or *TAGL1*. These include different clades of bHLH genes, on the one hand orthologues of *ALCATRAZ/SPATULA* and on the other orthologues of *INDEHISCENT/HECATE3*, in the Solanaceae. *ALC/SPT* orthologues are present in petunia, tomato, tobacco and pepper, as well as other Solanaceae members (Pabón-Mora *et al.*, 2014). On the other hand, *IND* orthologues are unique to Brassicaceae, as the *INDEHISCENT/HECATE3* duplication coincides with the Brassicales radiation (Kay *et al.*, 2013; Pabón-Mora *et al.*, 2014). Thus, all other angiosperms only have preduplication genes more similar to *HEC3* than to *IND*, and likely *HEC1/2* as they predate angiosperm diversification (Pfannebecker *et al.*, 2017). The goal of this research was to investigate the evolution and expression patterns of *SHP* transcription factors, the *RPL* transcription factors upstream of *SHP*, and the downstream bHLH genes involved in establishing the putative dehiscence zone to assess: (1) shifts in copy number, as well as shared and exclusive gene duplication events with reference to other core eudicots, in particular the model *A. thaliana*; (2) changes in copy number and functional motifs within Solanaceae that can be correlated to the shifts in fruit type; and (3) variations in expression patterns across different members of Solanaceae exhibiting dry dehiscent versus fleshy fruits.

MATERIALS AND METHODS

Transcriptome analyses

Each transcriptome was generated from mixed material derived from three biological replicates that included vegetative and reproductive meristems, floral buds, leaves and fruits (when available) in different developmental stages from *Brunfelsia australis* and *Streptosolen jamesonii*. These two species were selected because of their phylogenetic position as members of the early diverging Petunieae and the Browallieae respectively. Since they are ornamentals but lack edible fruits, less transcriptomic and genomic information is available. However, this information is much needed in order to bridge the gaps in the early-diverging members of the family as well as to have a putative reference point in terms of copy number for the Solanaceae. Mixed samples including leaves, floral buds, and fruits for each species were ground using liquid nitrogen and total RNA extraction was carried out using TRizol Reagent (Invitrogen, USA). RNA-seq experiments for each species were conducted using a TruSeq mRNA library construction kit (Illumina) (one library per species) and sequenced in a HiSeq2000 instrument reading 100 bases paired-end reads. The transcriptomes were assembled *de novo*. Read cleaning was performed with PRINSEQ-LITE with a quality threshold of Q35 and contig assembly was computed using the Trinity package, following default settings. For *B. australis*, contig metrics are as follows: total assembled bases, 95 583 446 bp; total number of contigs, 157 563; average contig length, 606 bp; largest contig, 11 983 bp; contig N50, 843 bp; contig GC, 40.20 %. For *S. jamesonii*, contig metrics are as follows: total assembled bases, 107 649 460 bp; total number of contigs, 148 552; average contig length, 724 bp; largest contig, 13 514 bp; contig N50, 1136 bp; contig GC, 40.90 %.

Gene isolation and phylogenetic analyses

For each of the genes, searches were performed using the *Arabidopsis thaliana* sequences (REPLUMLESS, AGAMOUS/SHATTERPROOF, ALCATRAZ/SPATULA and HECATE1/HEC2/HEC3/INDEHISCENT) as a query to

identify homologues in all available Solanaceae species. We used BLAST (Altschul *et al.*,1990) to do searches in all the repositories available for plant genomes (Phytozome, http://www.phytozome.net/; Sol Genomics Network, https://solgenomics.net/) and transcriptomes (OneKP, https://sites.google.com/a/ualberta.ca/onekp/). Whereas in Phytozome and Sol Genomics Network the sequences retrieved are full-length coding sequences with an open reading frame (ORF), those from OneKP are often partial coding sequences. However, we decided to include these incomplete sequences whenever the species belonged to subfamilies in the phylogeny lacking genome sequences and as long as the sequences had the distinctive conserved MADS, bHLH or HD/BELL domains according to the gene lineage. To expand sampling of homologues we isolated sequences using BLAST from our own transcriptomes generated from *S. jamesonii.* and *B. australis*; these sequences can be found under GenBank numbers MG452742–MG452758.

All full-length nucleotide sequences were compiled with Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and manually edited to exclusively keep the ORF for all transcripts. Nucleotide sequences were subsequently 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 (sometimes 4.0), an offset value of 1.0 and all other default settings. The alignment was refined manually using Bioedit taking into account the protein domains and amino acid motifs that have been reported as conserved for each of the gene lineages. The best model of molecular evolution for each gene lineage was calculated using MEGA7.0 (Kumar *et al.*, 2015). Maximum likelihood (ML) phylogenetic analyses using the nucleotide sequences were performed with RaxML-HPC2 BlackBox (Stamatakis *et al.*, 2008) through the CIPRES Science Gateway (Miller *et al.*, 2010). Bootstrapping was performed according to the default criteria in RaxML, where bootstrapping stops after 200–600 replicates. *Amborella trichopoda* genes were used as outgroup as follows: *AmtrAG* for the *AGAMOUS/SHATTERPROOF* analysis; *AmtrSPT* for the *SPATULA/ALCATRAZ* analysis; *AmtrbHLH87* for the *HECATE/INDEHISCENT* analysis; and *AmtrRPL* for the *RPL* analysis. Trees were observed and edited using FigTree v1.4.3. All sequences included in the phylogenetic analyses can be found in Supplementary Data Table S1.

Identification of new protein motifs

To detect both reported and new conserved motifs in REPLUMLESS, 29 sequences including the *Arabidopsis* RPL and the rice orthologue qSH1 were analysed, as expression and/or functional analyses for these genes have been reported. For the SPATULA/ALCATRAZ gene lineage we selected 23 sequences, including *Arabidopsis* SPT and ALC, *Prunus persica* SPT and *Fragaria vesca* SPT. Finally, for HECATE/INDEHISCENT, 34 sequences, including the *Arabidopsis* HEC1/2/3/IND proteins were analysed. Sequences were permanently translated and uploaded as amino acids to the online MEME server (http://meme-suite.org/) and run with all the default options (Bailey *et al.*, 2006). For all motif search analyses we included the same Solanaceae species whenever possible. These included full-length sequences from all taxa with sequenced genomes or transcriptomes that exhibited different fruit types: *Brunfelsia australis*, *Brugmansia sanguinea*, *Capsicum annuum*, *Nicotiana sylvestris*, *Petunia inflata*, *Solanum lycopersicum* and *Solanum tuberosum*.

The motifs retrieved by MEME are reported according to their statistical significance. The MEME suite finds, in the given sequences, the most statistically significant (low E-value) motifs first. The E-value of a motif is based on its log likelihood ratio, width, sites, and the size of the set. We numbered the motifs following the statistical significance given by the analyses. Whenever they coincide with previously reported motifs, labels have been placed accordingly.

Anatomy of fruits and selection of developmental stages for gene expression analyses

Fruits were collected in the field or in the laboratory and immediately fixed in formaldehyde–acetic acid–ethanol (FAA; 3.7 % formaldehyde, 5 % glacial acetic acid, 50 % ethanol). For light microscopy, fixed material was manually dehydrated through an alcohol–histochoice series and embedded in Paraplast X-tra (Fisher Healthcare, Houston, TX, USA). The samples were sectioned at 10–20 µm with an AO Spencer 820 (GMI, MN, USA) rotary microtome. Sections were stained with Johansen's safranin, to identify lignification and presence of cuticle, and 0.5 % Astra Blue and mounted in Permount (Fisher Scientific, Pittsburgh, PA, USA). Sections were viewed and digitally photographed with a Zeiss Axioplan compound microscope equipped with a Nikon DXM1200C digital camera with ACT-1 software for *C. annuum* and *S. lycopersicum* 'MicroTom'. Sections were photographed with an OMAX digital camera with Toplite software for *B. australis* and *Nicotiana obtusifolia*. The two stages F1 and F2 were selected for each species in an attempt to represent an early stage immediately after anthesis and a late stage during fruit maturation. For *B. australis*, collected in the field, F1 corresponds to the 0.5-cm fruit and F2 corresponds to the 1.3-cm fruit, corresponding to the last stage before the fruit turns brown and begins dehiscence. For *N. obtusifolia*, collected in the laboratory, F1 corresponds to the 3-mm diameter fruit at 1 d post-anthesis (1 DPA) and F2 corresponds to the 0.6-mm diameter fruit at 8 DPA. For *C. annuum* 'Black Pearl', collected in the laboratory, F1 corresponds to the 3-mm fruit at 4 DPA with active cell division, and F2 corresponds to the 0.8-cm diameter fruit during breaker stage, close to 30 DPA. For *S. lycopersicum* 'MicroTom', collected in the laboratory, F1 corresponds to the 5-mm fruit at 6 DPA, with active cell division, and F2 corresponds to the 1.5-cm diameter fruit during breaker stage, close to 45 DPA.

Expression analyses by RT–PCR

To examine and compare the expression patterns of *ALC*, *SPT*, *HEC1/2/3* and *RPL* genes in Solanaceae we used dissected sepals, petals, stamens and carpels in preanthetic floral buds, immature and mature fruits and leaves of *B. australis*, *C. annuum*, *N. obtusifolia* and *S. lycopersicum*. These four species represent different subfamilies and exhibit divergent fruit types. For instance, *B. australis* and *N. obtusifolia* have septicidal and

septicidal/loculicidal capsules, respectively, while *C. annuum* has a thin berry and *S. lycopersicum* has a thick berry. Total RNA was prepared from dissected organs, immature and mature fruits and leaves using TRizol Reagent (Invitrogen, Waltham, MA, USA). Samples were treated with DNAseI (Roche, Basel, Switzerland) and quantified with a NanoDrop 2000 (Thermo Scientific, Waltham, MA, USA). Three micrograms of RNA was used as a template for cDNA synthesis (SuperScript III RT, Invitrogen) using OligodT primers. The cDNA was used undiluted for amplification reactions by RT–PCR. The only extraction that was unsuccessful was for *N. obtusifolia* F2, despite several attempts with at least five different kits/protocols. For *RPL* genes, primers were designed flanking both the BELL and the HD whenever possible. For *ALC*/*SPT* and *HEC1/2/3/IND* genes, primers were designed outside of the conserved bHLH domain. All primers used were designed specifically for each paralogue found in *B. australis*, *C. annuum*, *N. obtusifolia* and *S. lycopersicum* (Supplementary Data Table S2). Each amplification reaction incorporated 9 µL of EconoTaq (Lucigen, Middleton, WI, USA), 6 µL of nuclease-free water, 1 µL of BSA (bovine serum albumin) (5 µg/mL), 1 µL of Q solution (betaine 5 µg/µL), 1 µL of forward primer (10 mM), 1 µL of reverse primer (10 mM) and 1 µL of diluted template cDNA, giving a total of 20 µL. Thermal cycling profiles followed an initial denaturation step (94 °C for 30 s), an annealing step (50–62 °C for 30 s) and an extension step with polymerase (72 °C for up to 1 min) repeated for 30–40 amplification cycles. *ACTIN* was used as a control. PCR was repeated at least five times with each primer pair in at least two independent sets of cDNA to check for consistency in the results. PCR products were run on a 1.0 % agarose gel stained with ethidium bromide and digitally photographed using a Whatman Biometra® BioDoc Analyzer.

RESULTS

The REPLUMLESS *(*RPL*) gene lineage*

A total of 108 sequences from angiosperms were included in the phylogenetic analysis (Supplementary Data Fig. S1). The aligned matrix contained 3184 characters, of which 2023 were informative. Using *Amborella trichopoda* single-copy *REPLUMLESS* as outgroup, the ML analysis recovered single-copy *RPL* genes in all angiosperms with the exception of duplicate genes in the Solanaceae, as well as in a few rosid species, including *Brassica rapa* (Brassicaceae), *Glycine max* (Fabaceae), *Gossypium raimondii* and *Theobroma cacao* (Malvaceae), *Malus domestica* (Rosaceae) and *Populus trichocarpa* (Salicaceae) (Fig. 2; Supplementary Data Fig S1). Sampling within Solanaceae included 43 sequences (Supplementary Data Table S1). The Solanaceae-specific duplication (Bootstrap Support (BS) = 100) results in the two clades *SolRPL1* (BS = 100) and *SolRPL2* (BS = 85). By comparison, molecular evolutionary rates have increased in *SolRPL1* more than in *SolRPL2*, as the latter clade exhibits shorter branch lengths (Fig. 2). For the most part, relationships among genes are consistent with the phylogenetic relationships of the sampled taxa (Olmstead *et al.*, 2008; Särkinen *et al.*, 2013). From our screening and based on the genomic data available for *C. annuum*, *S. lycopersicum*, *Solanum pennellii*, *Solanum pimpinellifolium* and *S. tuberosum*, we know that *SolRPL1* orthologues are always found in chromosome 10 while *SolRPL2* copies are found in chromosome 9. Gene losses are harder to determine, but while *C. annuum* 'CM334' possesses two gene copies, one in each clade, the varieties *C. annuum* var. *glabriusculum*, and *C. annuum* var. *Zunla* seem to have lost the *SolRPL2* homologue. A similar case occurs in *S. jamesonii*, where there is only one *RPL* in the *SolRPL2* clade (Fig. 2).

Our MEME analysis resulted in the identification of conserved protein motifs in both Solanaceae clades, the canonical *A. thaliana* RPL, the orthologue ArlyRPL (*Arabidopsis lyrata*, Brassicaceae), the rice qSH1, and ZemaRPL (*Zea mays*, Poaceae). Our MEME analysis resulted in, the HD represented by motifs 1 and 2 preceded upstream by the BELL domain found in motifs 3 and 4. Upstream of BELL we also detected a variation of the SKY motif, as motif 5, which, in all RPL proteins aligned, shifts to SRF and is accompanied downstream by LKPAQxLLEEL (Fig. 3; Supplementary Data Fig. S3A). In addition, our analyses also recovered the ten-amino acid ZIBEL motifs at the N-terminal and C-terminal ends of all proteins as motif 6 (Fig. 3; Supplementary Data Fig. S3B, C). Two new motifs we have detected here include motif 9 at the beginning and motif 7 at the end of the RPL proteins (Supplementary Data Fig. S3D, E). These were screened in other BEL proteins closely related to RPL, including PNF, BLH2 and BLH10, and could not be found, suggesting that all new motifs are indeed exclusive to RPL homologues (Hake *et al.*, 2004; Kanrar *et al.*, 2006; Kumar *et al.*, 2007; Khan *et al.*, 2015).

Finally, motifs 8 and 10 are located between the HD and the C-terminal ZIBEL (Fig. 3; Supplementary Data Fig. S3F). Differences between SolRPL1 and SolRPL2 include the seven amino acids preceding the stop codon, which correspond to motif LLHDFVG in SolRPL1 and FLHDFAG in SolRPL2 and maintain amino acid properties. A comparison of this particular motif with sequences outside the Solanaceae shows that FVG, typical of SolRPL1, is conserved in other angiosperms. In addition, by comparison, the monocot and Brassicaceae RPL homologues that were functionally characterizedin this study, vary in the two amino acids at the beginning of the motif, corresponding to LL in the former and FL in the latter.

In an effort to find putative berry-specific motifs, we identified three regions, one in SolRPL1 and two in SolRPL2 (Supplementary Data Fig. S3G–I). The most divergent protein sequences were found between motifs 10 and 8, characterized by variants of polar uncharged amino acids (Supplementary Data Fig. S3H). The remaining putative berry-specific motif is located between the BELL and HD domains, where sequences from berry-bearing species belonging to SolRPL2 have only 18 amino acids while the remaining proteins have 32–41 amino acids (Fig. 3, Supplementary Data Fig. S3H, I).

The AGAMOUS/SHATTERPROOF *(*AG/SHP*) gene lineage*

Sequences recovered by similarity in the transcriptomes generally span the entire coding sequence, although some copies only have a complete MADS domain followed by a premature stop codon. The aligned matrix consists of 1009 characters, of which 595 were informative. Maximum likelihood analysis recovered a core eudicot duplication event resulting

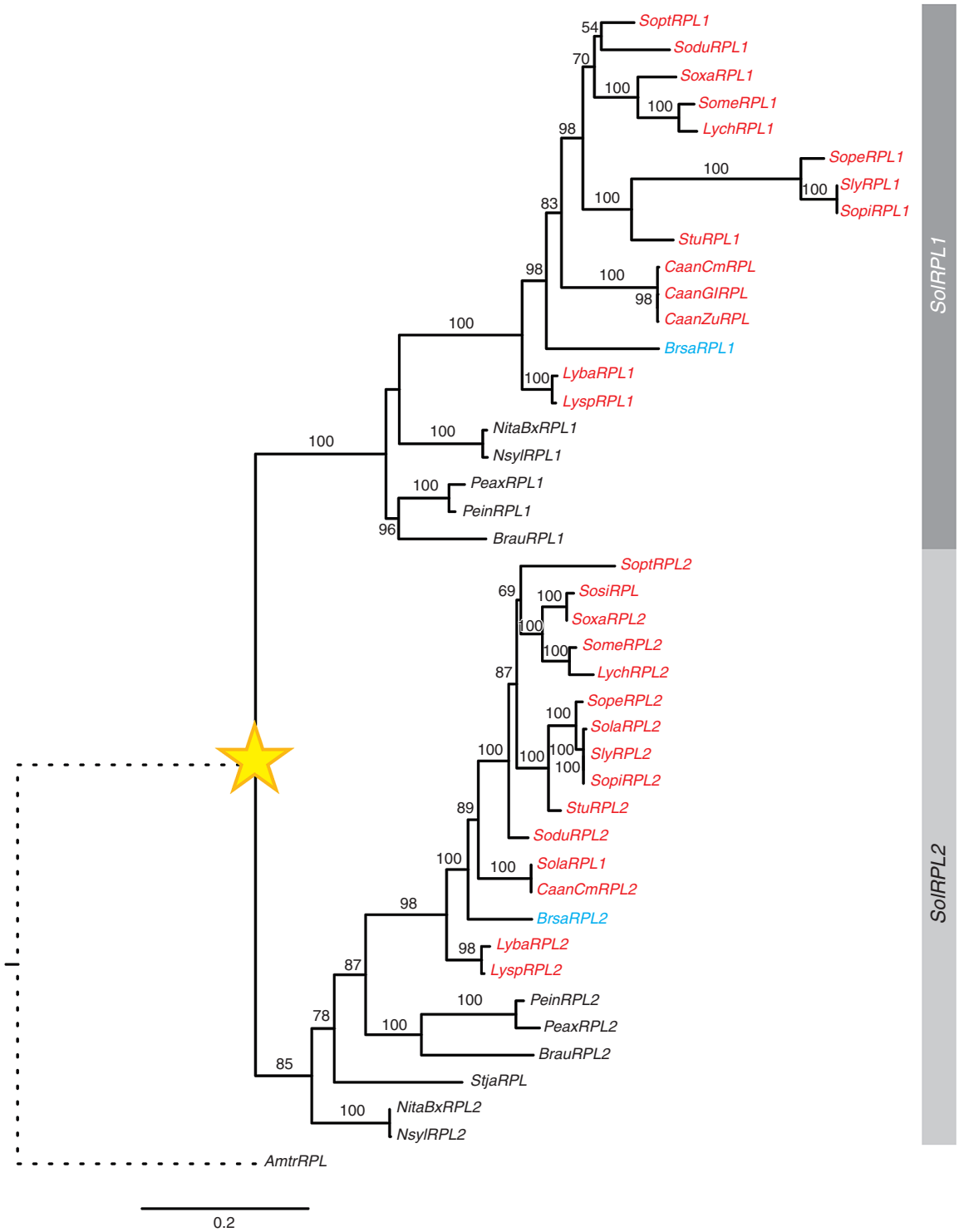

FIG. 2. Maximum likelihood tree of *REPLUMLESS* genes in Solanaceae. *Amborella trichopoda RPL* (*AmtrRPL*) was used as outgroup. The topology recovers two clades, *SolRPL1* and *SolRPL2*, as a result of a specific duplication event in Solanaceae (star). Genes belonging to species with dry dehiscent fruits are labelled in black, those belonging to species with fleshy fruits are labelled in red, and those belonging to species from the Daturae tribe, indicating phylogenetic reversals to dry dehiscent fruits, are labelled in light blue. Branch numbers indicate BS supports and scale bar indicates the number of substitutions divided by the length of the sequence.

in the *AGAMOUS* (BS = 62) and *PLENA/SHATTERPROOF* clades (BS = 69; Fig. 4, Supplementary Data Fig. S4). Thus, all Solanaceae species, similar to other core eudicots, have retained both *AG* and *SHP* orthologues (Fig. 4). From our screening and based on the genomic data available for *C. annuum*, *S. lycopersicum*, *S. pennellii*, *S. pimpinellifolium* and *S. tuberosum*,

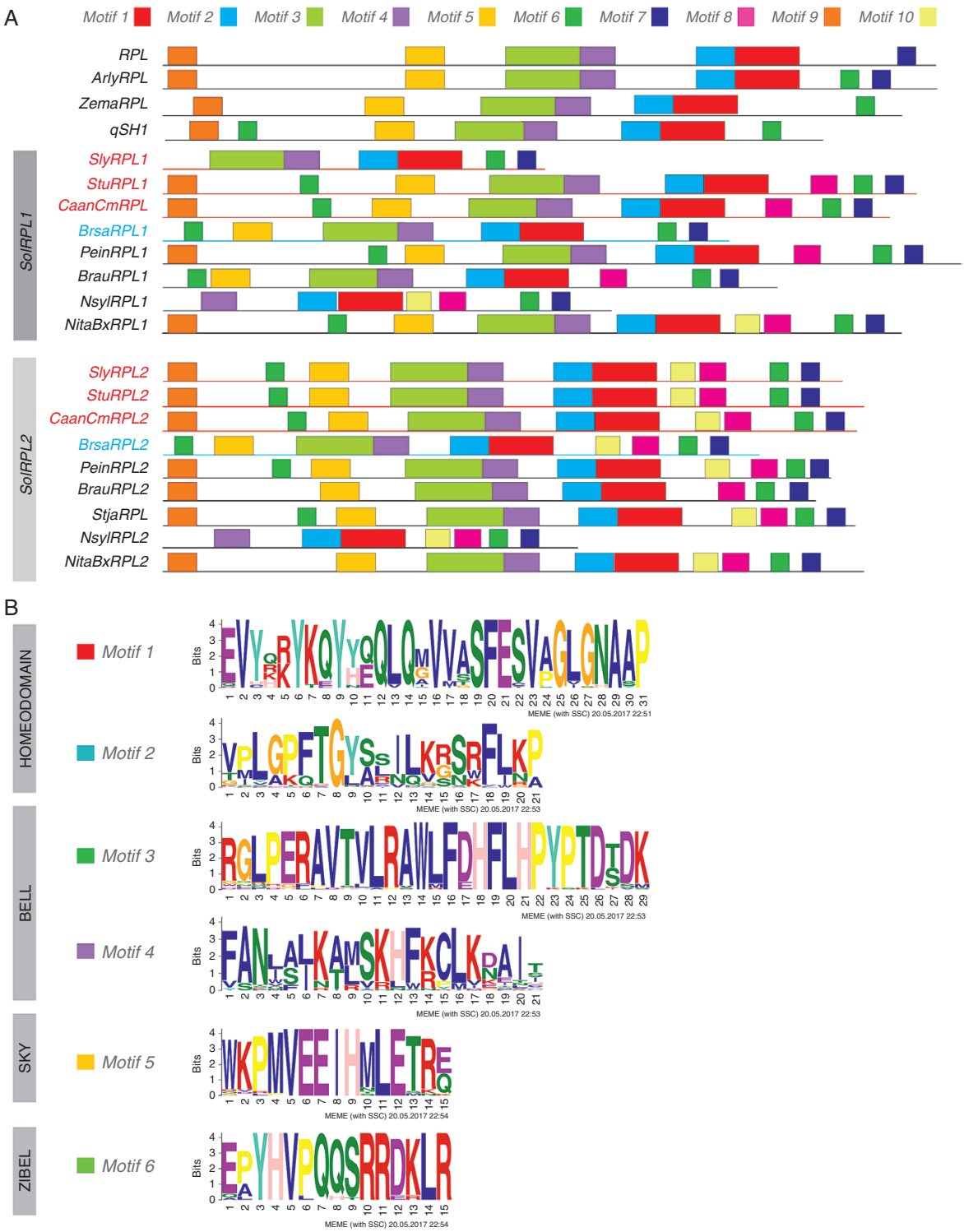

FIG. 3. (A) Conserved motifs mapped on the REPLUMLESS Solanaceae proteins and selected functionally characterized proteins from *Arabidopsis thaliana* (RPL), *Arabidopsis lyrata* (ArlyRPL), *Zea mays* (ZemaRPL) and *Oryza sativa* (qSH1). Colours for protein names follow Fig. 2. (B) Sequences of the conserved motifs previously identified in RPL proteins.

we know that *SolAG* orthologues are always found in chromosome 2 while *SolSHP* copies are found in chromosome 7. Our results show more changes in the coding sequences of *SHP* homologues when compared with *AG* copies, as shown by the branch lengths in the ML analysis (Fig. 4). For the most part, relationships among genes are consistent with the phylogenetic relationships of the sampled taxa (Särkinen *et al.*, 2013). The only exceptions to this are the *AG/SHP* homologues in *Brugmansia*, which have extensively deviant coding sequences that cluster within the *SHP* clade with low support. Alternative

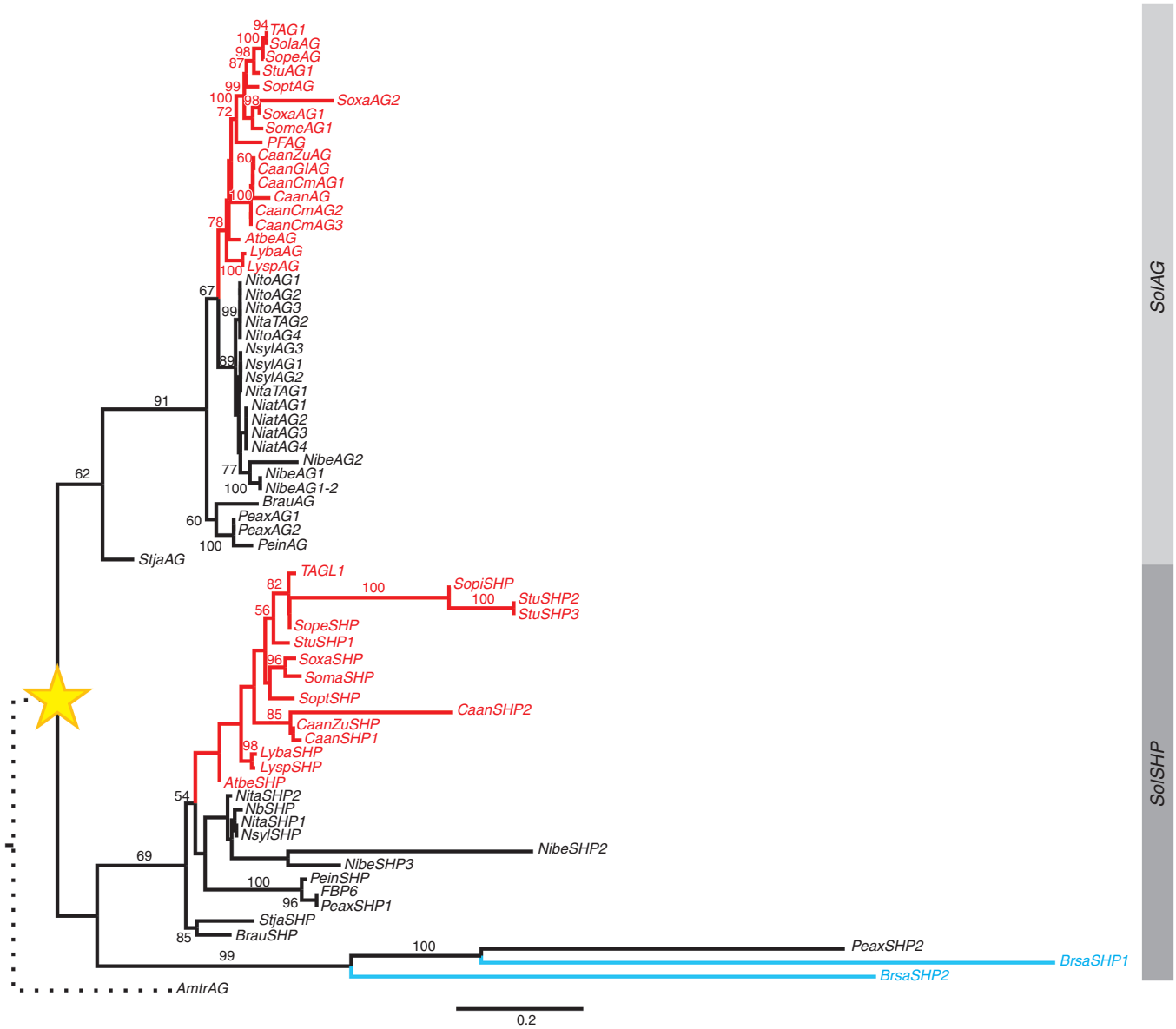

FIG. 4. Maximum likelihood tree of *AGAMOUS/SHATTERPROOF* genes in Solanaceae. *Amborella trichopoda AGAMOUS* (*AmtrAG*) was used as outgroup. The topology recovers *SolAG* and *SolSHP* gene clades as a result of a core eudicot duplication event. The only duplication event found is labelled with a star. Colours for gene names follow Fig. 2. Branch numbers indicate BS supports and scale bar indicates the number of substitutions divided by the length of the sequence.

spliced transcripts are seen in *AG* genes in *Petunia axilaris*, *Nicotiana attenuata, N. benthamiana*, *N. tomentosifolia*, *N. sylvestris* and *C. annuum* but are far less common in *SHP* genes, where they occur only in *N. benthamiana* and *Nicotiana tabacum* (Fig. 4). Putative gene losses may have occurred in *B. sanguinea*, as only the *SHP* homologue was recovered; however, due to the lack of a reference genome for this species, *AG* gene loss remains to be confirmed.

The ALCATRAZ/SPATULA *(*ALC/SPT*) gene lineage*

The combined matrix used here includes all sequences used in previous analyses in addition to the expanded sampling in Solanaceae, resulting in a matrix of 197 sequences using *A. trichopoda* single-copy palaeo-*SPT/ALC* (*AmtrSPT*) as outgroup (Pabón-Mora *et al.*, 2014; Zumajo-Cardona *et al.*, 2017). The aligned matrix consists of 2140 characters, of which 1413 were informative. Our ML analysis recovered a core eudicot duplication (BS = 94) resulting in the two clades *ALC* (BS = 91) and *SPT* (BS = 65), both having representatives among rosids (including *Vitis vinifera*) and asterids (Supplementary Data Fig. S4; Pabón-Mora *et al.*, 2014; Zumajo-Cardona *et al.*, 2017). However, in this new analysis the Brassicaceae *ALC* clade appears as sister to the Brassicaceae *SPT* clade, both nested within the core eudicot *SPT* clade (Supplementary Data Fig. S4).

The *ALC/SPT* gene lineage in Solanaceae was reconstructed based on 63 coding sequences from available databases and our own transcriptomes (Fig. 5). The aligned matrix consists of 1461 characters, of which 830 were informative. The resulting

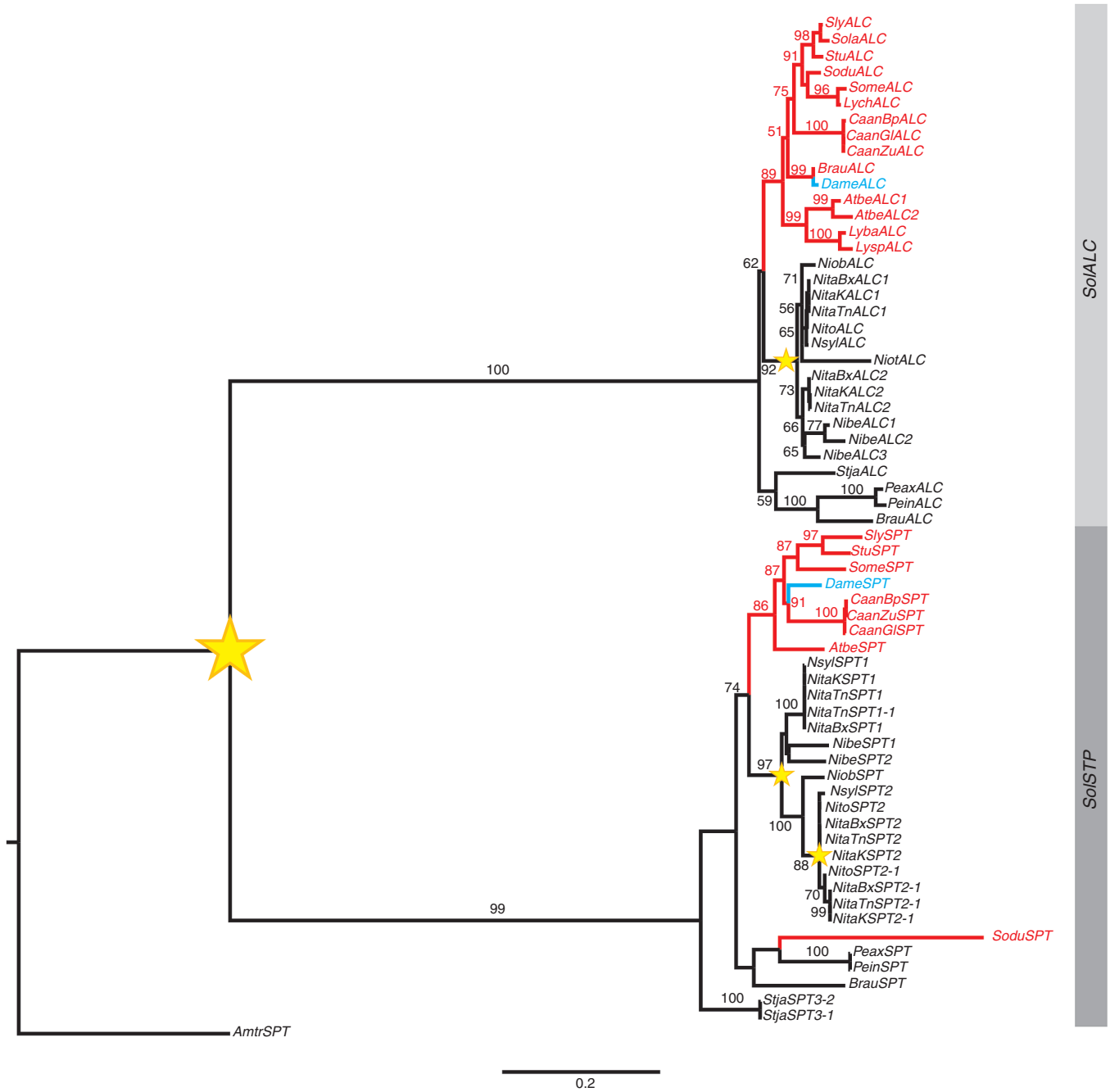

FIG. 5. Maximum likelihood tree of *ALCATRAZ/SPATULA* genes in Solanaceae. *Amborella trichopoda SPATULA* (*AmtrSPT*) was used as outgroup. The topology recovers *SolALC* and *SolSPT* gene clades as a result of a core eudicot duplication event, and additional duplication events (labelled with stars) occurring prior to the diversification of some polyploid *Nicotiana* species. Colours for gene names follow Fig. 2. Branch numbers indicate BS supports and scale bar indicates the number of substitutions divided by the length of the sequence.

ML analysis topology shows a first duplication event separating the two Solanaceae clades coinciding with the core eudicot duplication (Fig. 5; Pabón-Mora *et al.*, 2014; Zumajo-Cardona *et al.*, 2017), named *SolALC* (BS = 100) and *SolSPT* (BS = 99). The additional duplication events, one in *SolALC* and two in *SolSPT*, are specific to *Nicotiana*, but the time-points of these duplications are unclear (Fig. 5). The first *SPT* duplication, resulting in *SolSPT1* and *SolSPT2*, likely predates the diversification of all species in the genus *Nicotiana* as the diploid *N. sylvestris* possesses two copies, one in each clade. The *ALC* duplication resulting in *SolALC1* and *SolALC2* as well as the second *SPT* duplication, resulting in *SolSPT2* and *SolSPT2-1*, have occurred specifically in allotetraploid *Nicotiana* species, like *N. tabacum* 'K326' (Flue-cured), 'TN90' (Burley) and 'Basma Xanthi' (BX, Oriental), and independently in *N. benthamiana* (Fig. 5). In addition, the copies *SPT2* and *SPT2-1* are identical until amino acid 366 and differ only at the N-terminus of the protein. From our screening and based

on the genomic data available for *C. annuum*, *S. lycopersicum*, *S. pennellii*, *S. pimpinellifolium* and *S. tuberosum*, we know that *SolALC* orthologues are always found in chromosome 4 while *SolSPT* copies are found in chromosome 2.

Our MEME analysis resulted in the identification of ten conserved protein motifs, of which motif 1 corresponds to the bHLH domain, motif 4 (immediately upstream of the bHLH) corresponds to the nuclear localization signal (NLS) *sensu* Groszmann *et al.* (2011), motif 2 corresponds to the acidic domain, and motif 7 corresponds to the amphipathic helix (Fig. 6; Pires and Dolan, 2010; Groszmann *et al.*, 2011). The basic region of the bHLH domain is very different in Brassicaceae when compared with other rosids or to Solanaceae. The canonical motif in ALCATRAZ corresponding to NIDAQF is unique to Brassicaceae and it is shifted to SRSAEVH in Solanaceae, and even in *P. persica* (PPERSPT), and *F. vesca* (FaSPT) (Fig. 6; Supplementary Data Fig. S5). This explains the absence of motif 4 in the ALC Brassicaceae orthologues. Comparatively, SPT homologues have fewer changes; while the first eight amino acids of the bHLH domain in Brassicaceae SPT sequences correspond to KR**C**RAAEVH, they shift to KR**S**RAAEV in other species.

The new motifs identified here for SolSPT copies include motifs 5 and 9 at the downstream the bHLH domain towards the end of the protein and motif 10, between the amphipathic helix and the acidic domain (Fig. 6, Supplementary Data Fig. S5). The only exclusive motif for SolALC copies is motif 3 at the 3′ end of the bHLH domain. Motif 8 is rescued in the analysis in all SolALC/SPT sequences, but it varies between SolALC and SolSPT in position. While in SolALC motif 8 (EF**L**EDDKVDNFGFSSE**E**CDGL) is located at the 5′ end of the bHLH domain and is predominantly acidic, in SolSPT motif 8 (RM**L**QQNQLSHQKVGLC**E**GNAF) is located at the 3′ of the bHLH domain and is predominantly polar (Fig. 6; Supplementary Data Fig. S5). In both motifs positions 3 and 17 match a leucine and glutamic acid (**L** and **E** above), respectively. When compared with ALC and SPT in *Arabidopsis*, our data point to the same trends previously identified, where the *Arabidopsis* proteins have reduced conserved motifs compared with other core eudicot ALC/SPT proteins (Fig. 6; Supplementary Data Fig. S5). Changes in the sequences correlated with the occurrence of dry dehiscent and fleshy fruits were identified, but, unlike in *RPL* genes, these changes are often point amino acid substitutions and their biological relevance is yet to be investigated (Supplementary Data Figs S6 and S7).

The HECATE 1/2/3/ INDEHISCENT *(*HEC/IND*) gene lineage*

Our analysis of the *HEC1/2/3/IND* gene lineage was made with 176 sequences from across angiosperms. The aligned matrix consists of 1867 characters, of which 911 were informative. The topology suggests an early duplication event for all angiosperms resulting in the *HEC1/2* and the *HEC3/IND* clades with very low support (Supplementary Data Fig. S8). Within each of these clades, additional duplications have occurred. The *HEC1/2* clade has undergone further independent duplications in Brassicaceae resulting in the *HEC1* and *HEC2* clades, and in Solanaceae, resulting in the *SolHEC1* and *SolHEC2* clades (Fig. 7; Supplementary Data Fig. S8). The *HEC3/IND* clade only underwent additional duplications during the diversification of the Brassicaceae, resulting in the *HEC3* and *IND* clades (Supplementary Data Fig. S8). However, other rosids and most asterids only have *HEC3-like* single copy pre-duplication genes (Pabón-Mora *et al.*, 2014; Pfannebecker *et al.*, 2017).

Within Solanaceae, different duplication trends are observed in this gene lineage. *SolHEC1* (BS = 99) and *SolHEC2* (BS = 99) underwent additional duplications, on the one hand resulting in *SolHEC1* (BS = 81) and *SolHEC1-1* (BS = 58), and on the other in *SolHEC2* (BS = 92) and *SolHEC2-1* (BS = 99; Fig. 7). Thus, most Solanaceae have two *HEC1* copies and two *HEC2* copies (Fig. 7). Additional species-specific copies are only found in *N. benthamiana*, with four *HEC1* copies and three *HEC2* copies, and in *N. tabacum*, having four *HEC2* copies (Fig. 7). This contrasts sharply with the retention of a single-copy *HEC3* in most Solanaceae species (BS = 100), perhaps with the only exception found in the tetraploid *N. benthamiana* and *N. tabacum*, possessing two *SolHEC3* copies (Fig. 7). From our screening and based on the genomic data available for *C. annuum*, *S. lycopersicum*, *S. pennellii*, *S. pimpinellifolium* and *S. tuberosum*, we know that *SolHEC1* orthologues are found in chromosomes 2 and 4, *SolHEC2* orthologues in chromosomes 3 and 12 and *SolHEC3* orthologues in chromosome 11.

For *SolHEC1/2/3* homologues, our MEME analyses resulted in the identification of 11 conserved protein motifs (Fig. 8). We found that the bHLH domain is conserved in all sequences, corresponding to motifs 1, 2 and 3; these are the only motifs conserved with the *Arabidopsis HEC/IND* homologues (Fig. 8). Motif 2 as described here includes 'the HEC exclusive motif (17)' identified in Pires and Dolan, 2010. In Solanacaeae HEC1/2/3, motifs 2 and 3 are different at the start and the end, respectively (Supplementary Data Fig. S9). The beginning of motif 2 varies in the first six amino acids; in SolHEC1 they are LQ_XRNS, in SolHEC2 they are SMNRSN and in SolHEC3 they are E/DEEEEE (Supplementary Data Fig. S7). Likewise, the last five amino acids of motif 3 also vary. In SolHEC1 they correspond to QAAVN/D, in SolHEC2 to RAGAT/N and in SolHEC3 to $QS/L_XNHH/N$ (Supplementary Data Fig. S9). In addition to the bHLH, motif 4 is recovered for all *SolHEC1/2/3* homologues (Fig. 8); in SolHEC1 motif 4 (LMT/NSPPSNFS**F**MGNPIEEPAA) is located upstream of the bHLH domain, while motif 4 for SolHEC2 ($A/_{SX}A_{XX}$GLG**F**PVPMSLSGNY) and SolHEC3 (N/T_XTT**F**VGN_{XX}SD/NPTY) is located downstream of bHLH (Fig. 8). This motif varies extensively except in the phenylalanine at position 11.

Exclusive domains for each clade were also found. Motif 9 is only found in SolHEC1 and is located at the N-terminal end of the protein (Fig. 8). Motifs 8 and 11 are exclusive of SolHEC2 and are located towards the N-terminus (Fig. 8). Motifs 5, 6, 7 and 10 are exclusive to the SolHEC3 clade; while motifs 5, 6 and 10 are located upstream of bHLH, motif 7 is located downstream of the bHLH domain (Fig. 8). Changes in the sequences correlated with the occurrence of dry dehiscent and fleshy fruits were identified, and, as in *SPT/ALC* genes, these changes are often point amino acid substitutions and their biological relevance is yet to be investigated (Supplementary Data Figs S10–S12)

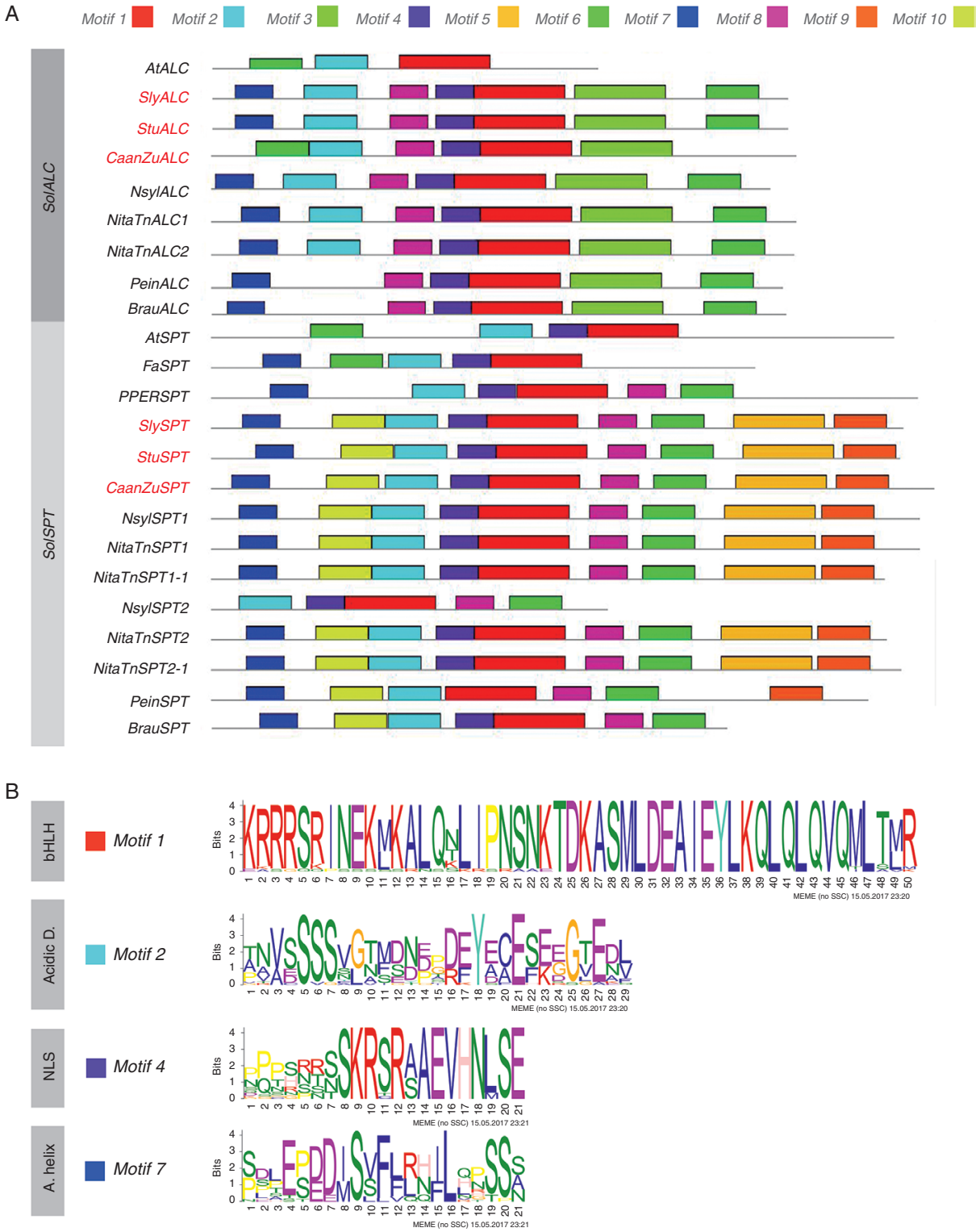

FIG. 6. (A) Conserved motifs of ALCATRAZ/SPATULA proteins in Solanaceae and selected functionally characterized proteins from *Arabidopsis thaliana* (AtALC and AtSPT), *Fragaria vesca* (FaSPT) and *Prunus persica* (PPERSPT). Colours for protein names follow Fig. 2. (B) Sequences of the conserved motifs previously identified in ALC/SPT proteins. Note that AtALC has undergone reduction in terms of conserved motifs when compared with other ALC/SPT proteins.

Expression analyses of bHLH and RPL genes in Solanaceae

In order to identify how these genes were expressed in different Solanaceae species we studied their expression patterns in dissected floral organs of four different species and two fruit developmental stages. The species selected include *B. australis* (Petunieae), *C. annuum* (Capsiceae), *N. obtusifolia* (Nicotianeae) and *S. lycopersicum* (Solaneae). The four

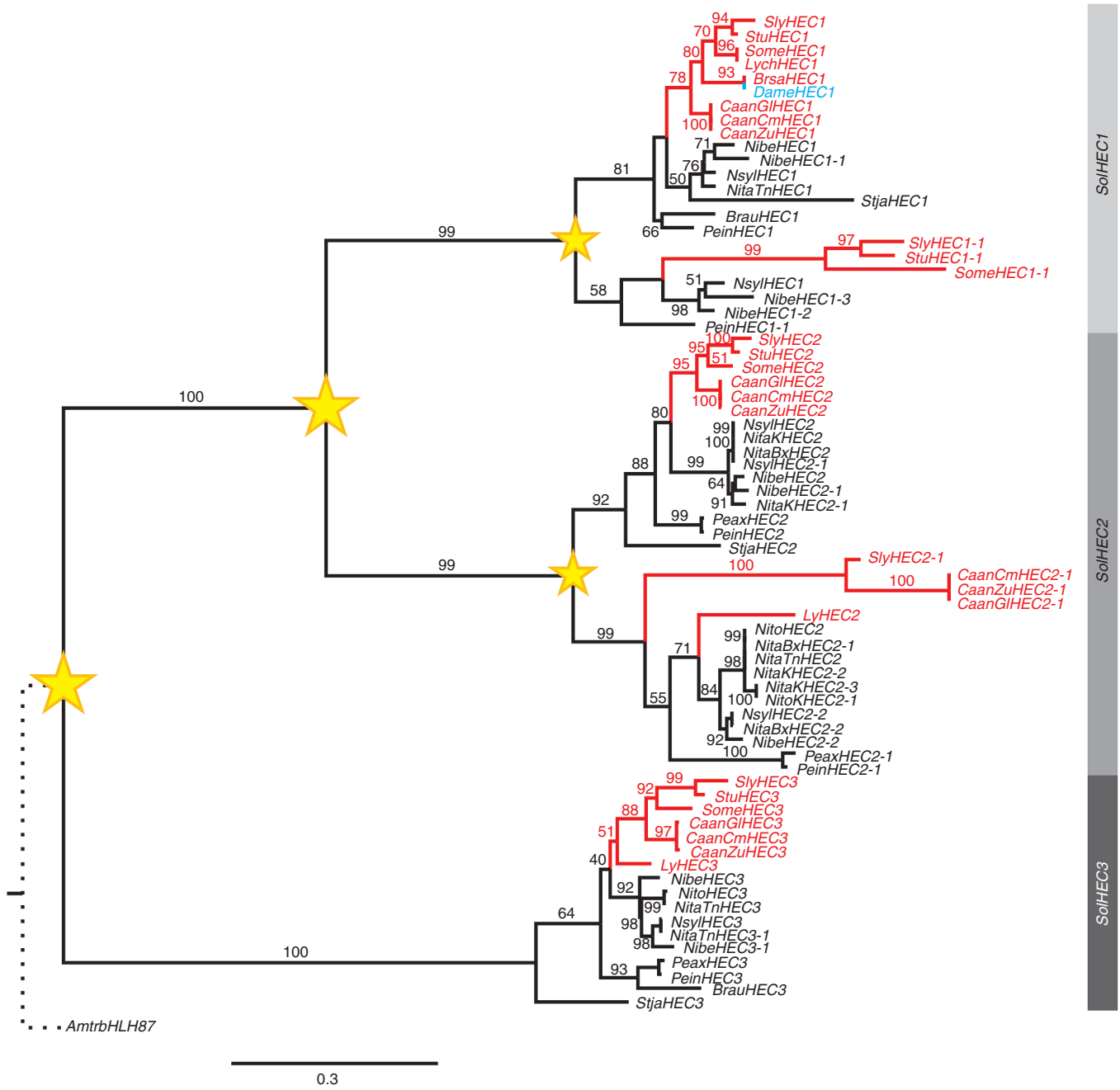

FIG. 7. Maximum likelihood tree of *HECATE/INDEHISCENT* genes in Solanaceae. *Amborella trichopoda bHLH87* (*AmtrbHLH87*) was used as outgroup. The topology recovers a split between *HEC3* and *HEC1/2* that predates angiosperm diversification, followed by additional duplication events (labelled with stars) within *HEC1* and *HEC2* specific to Solanaceae. Since *INDEHISCENT* genes are unique to Brassicaceae, the Solanaceae only possess the closely related *HECATE3* genes. Colours for gene names follow Fig. 2. Branch numbers indicate BS supports and scale bar indicates the number of substitutions divided by the length of the sequence.

species for comparison were selected as they represent four tribes in the phylogeny diverging at different time-points and exhibiting a unique fruit type (see below). These stages represent an early stage with active cell division in all fruits (F1), and a late developmental stage with cessation of cell division at the beginning of maturation (F2). Transverse sections were made to help visualize the developmental stages evaluated during gene expression analyses (Fig. 9). Descriptions of the pericarp follow Pabón-Mora and Litt (2011). Early fruit development (F1) in *B. australis* is characterized by both anticlinal and periclinal cell division in the 21–24 cell layers of the pericarp. At this stage the endocarp, the mesocarp and the exocarp are all parenchymatous and small intercellular spaces can be observed (Fig. 9A). The exocarp is covered by a thick cuticle (Fig. 9A). Late fruit development in *B. australis* (F2) is characterized by the lignification of the inner endocarp going into the septum and the continuation of both cell expansion and anticlinal cell division in the outer

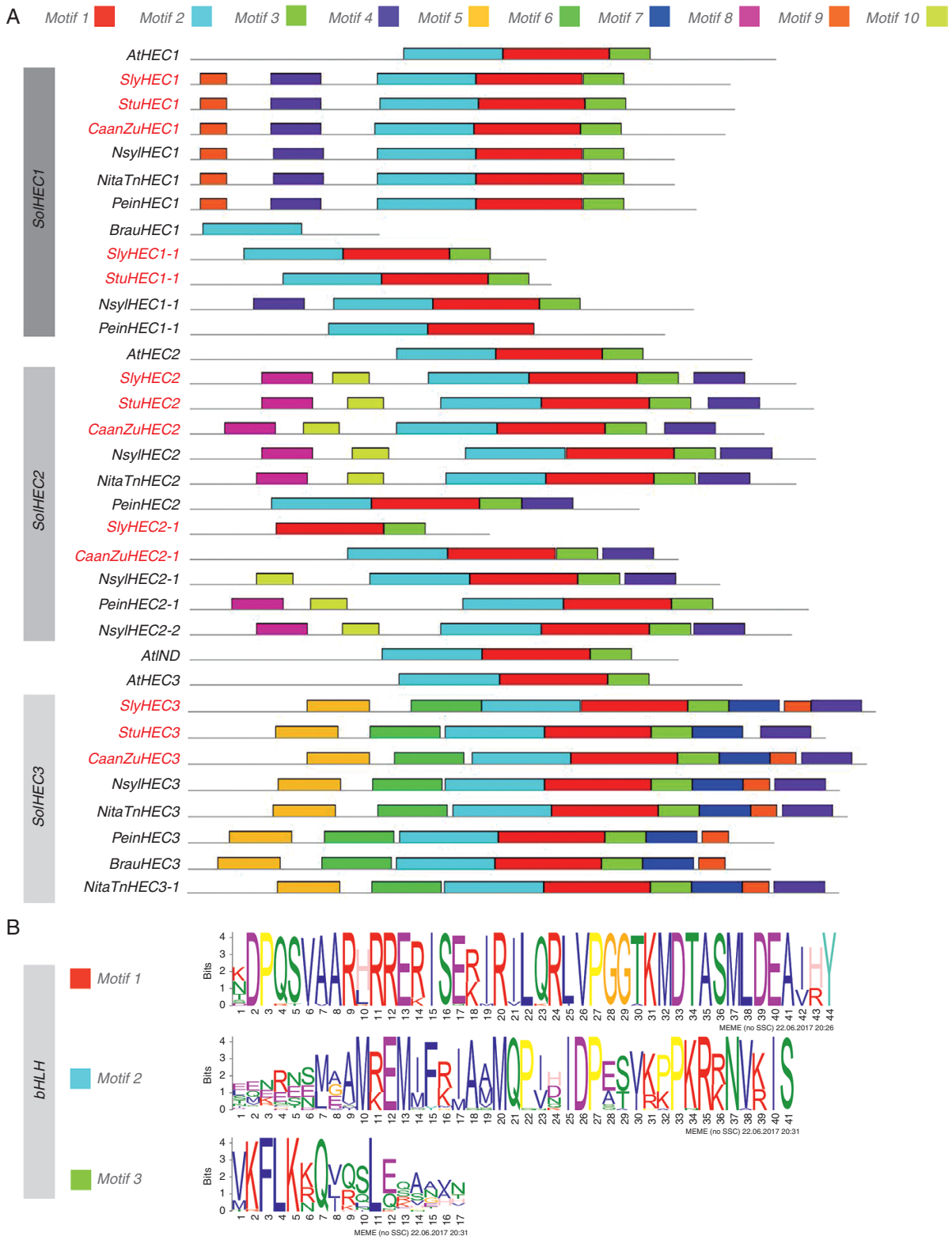

FIG. 8. (A) Conserved motifs of HECATE/INDEHISCENT proteins in Solanaceae and selected functionally characterized proteins from *Arabidopsis thaliana* (AtHEC1, AtHEC2, AtHEC3 and AtIND). Colours for protein names follow Fig. 2. (B) Sequences of the three motifs that form the bHLH domain.

endocarp and the mesocarp. However, periclinal cell division does not occur, as the number of cell layers remains the same (Fig. 9B). At this stage isodiametrical smaller cells are found at the septum marking the future dehiscence zone (data not shown). In *N. obtusifolia* early fruit development (F1) exhibits an eight-layered pericarp homogeneously parenchymatous

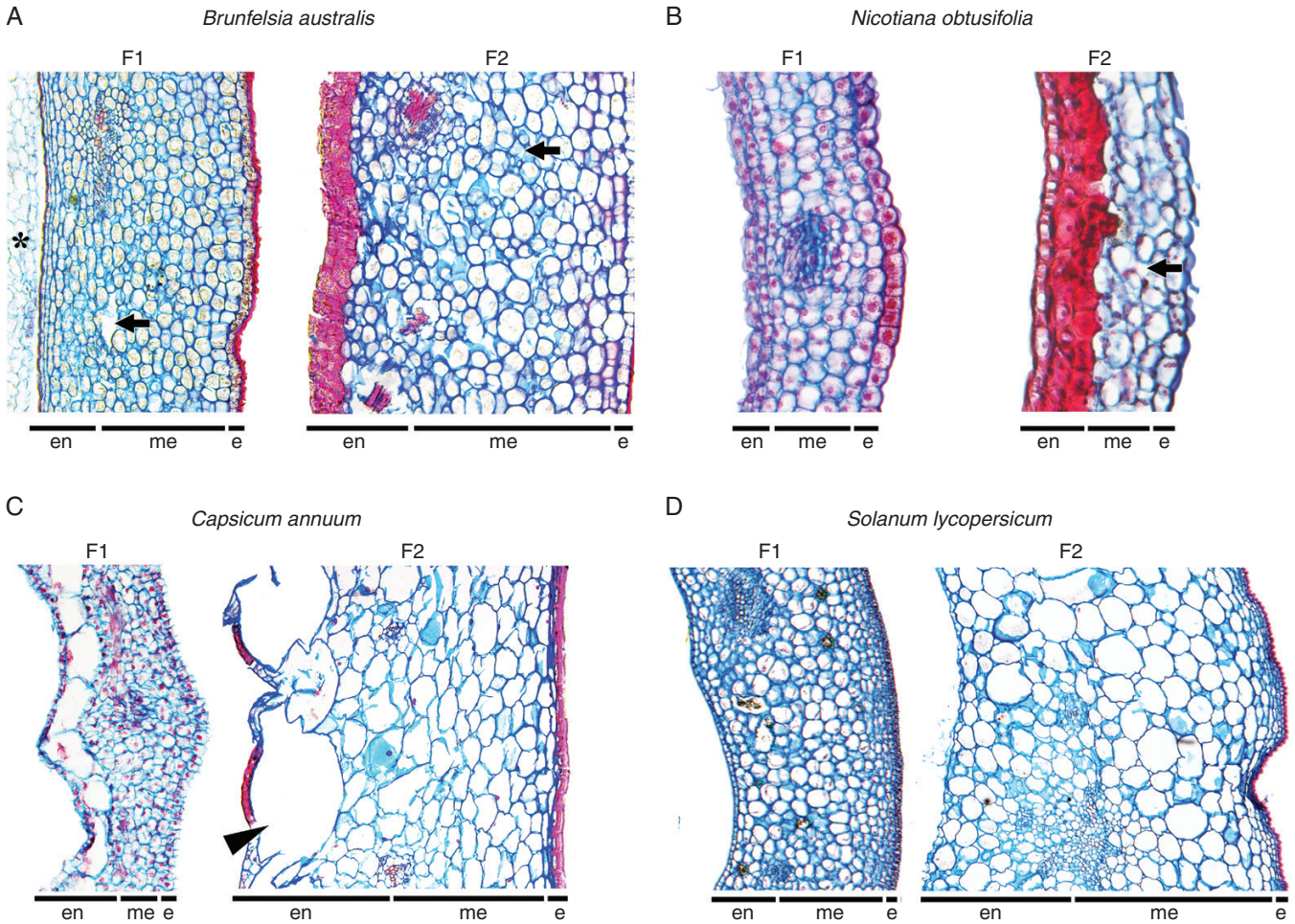

FIG. 9. Cross-sections of selected Solanaceae with dry dehiscent (A, B) and fleshy (C, D) fruits. (A) *Brunfelsia australis* prior to endocarp lignification in early development (F1) and after endocarp lignification at late (F2) developmental stages. Asterisk indicates the placenta. (B) *Nicotiana obtusifolia* in the same two developmental stages (F1 and F2). Arrows in (A) and (B) point to the typical lacunae (intercellular spaces) during dry dehiscent fruit development. (C) *Capsicum annuum* 'Black Pearl' during early development with cell expansion restricted to the endocarp in early development (F1) and in late development with cell expansion in endocarp, mesocarp and exocarp, with a fully developed cuticle (F2). Arrowhead indicates the inner layer of the endocarp characterized by giant cells that accumulate capsaicin. (D) *Solanum lycopersicum* 'MicroTom' during early development, with abundant cell division and limited cell expansion (F1), and during late development, with increased cell expansion and a fully developed cuticle (F2). Abbreviations: e, exocarp; en, endocarp; me, mesocarp.

with three layers of smaller cells marking the inner endocarp (Fig. 9B). No cuticle is formed covering the exocarp (Fig. 9B). During maturation in the *N. obtusifolia* fruit (F2) the inner endocarp becomes lignified and the mesocarp and exocarp continue to divide anticlinally, leaving extensive intercellular spaces in the pericarp (Fig. 9B).

On the other hand, the fleshy fruits undergo completely different processes during maturation. In *C. annuum* 'Black Pearl' early fruit developmental stages (F1) already show a clear differentiation of the two inner layers of the endocarp with respect to the rest of the 14-layered pericarp. The innermost layer of the endocarp is characterized by having small parenchymatous cells, and the one adjacent to it exhibits gigantic cells (almost 5–6 times the normal cell size in the rest of the pericarp) with larger nuclei (Fig. 9C). The rest of the endocarp, the mesocarp and the exocarp have parenchymatous cells with active cell division, both anticlinal and periclinal (Fig. 9C). During maturation (F2) the fruits of *C. annuum* are the result of extensive cell expansion in all layers except the inner endocarp, whose cells undergo lignification in parallel to the expansion of the adjacent layer of gigantic cells (Fig. 9C). The outer mesocarp and the exocarp exhibit flattened cells with a thickened cell wall. In addition, a cuticle develops over the exocarp at late developmental stages (Fig. 9C). In *S. lycopersicum* 'MicroTom' early fruit development stages (F1) are characterized by a parenchymatous 16-layered pericarp with extensive anticlinal and periclinal cell division, where only the inner endocarp layer, the outer mesocarp layers and the exocarp present smaller, almost square cells (Fig. 9D). In *S. lycopersicum*, maturation (F2) is accompanied by cell expansion and both anticlinal and periclinal cell division, resulting in a 27 layered pericarp (Fig. 9D). As in *C. annuum*, the outer mesocarp and the exocarp exhibit thickened cells walls and a cuticle is present over the exocarp (Fig. 9D).

In order to hypothesize functional roles of *RPL*, *ALC*, *SPT* and *HEC1/2/3* orthologues in *B. australis*, *N. obtusifolia*, *C. annuum* and *S. lycopersicum*, respectively, the expression

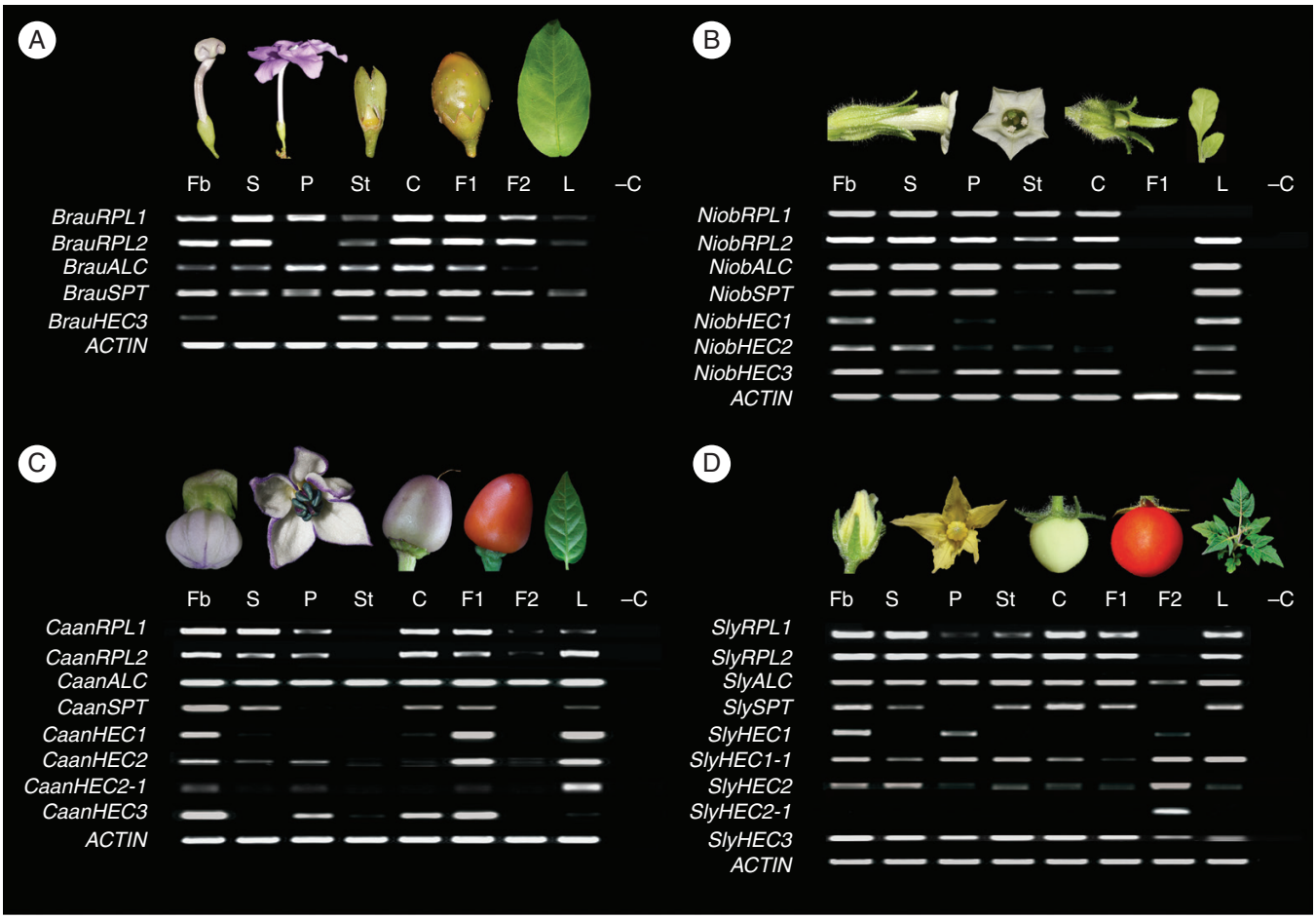

FIG. 10. Expression analyses of *ALCATRAZ*, *HECATE*, *REPLUMLESS* and *SPATULA* homologues in four species of Solanaceae with dry dehiscent and fleshy fruits. *ACTIN* was used as a loading control. Expression of all homologues is shown in dissected floral organs, fruits (F1 and F2) and leaves of (A) *Brunfelsia australis*, (B) *Nicotiana obtusifolia*, (C) *Capsicum annuum* 'Black Pearl' and (D) *Solanum lycopersicum* 'MicroTom'. C, carpels; Fb, floral bud; F1, early stages of fruit development; F2, late stages of fruit development; L, leaves; P, petals; S, sepals; St, stamens. -C indicates the amplification reaction loaded without cDNA.

patterns were assessed in sepals, petals, stamens and carpels of pre-anthetic floral buds, immature (F1) and mature (F2) fruits (as defined above), and leaves of each species (Figs 9 and 10).

Genes belonging to the *SolRPL1* and *SolRPL2* clades have very similar expression patterns. In *B. australis* the *RPL* copies *BrauRPL1* and *BrauRPL2* are expressed in floral buds, sepals, stamens, carpels and fruits in F1 and F2. Additionally, *BrauRPL1* is expressed in petals (Fig. 10A). *NiobRPL1* and *NiobRPL2* are expressed in floral buds and all dissected floral organs; moreover, only *NiobRPL2* is expressed in leaves (Fig. 10B). *CaanRPL1* and *CaanRPL2* have the same expression patterns; they are present in floral buds, sepals, petals, carpels and F1 fruits. Of the two, only *CaanRPL2* is strongly expressed in leaves (Fig. 10C). Finally, *SlyRPL1* and *SlyRPL2* are expressed in all organs evaluated except in F2 fruits; however, *SlyRPL1* is expressed at a low level in petals and stamens in comparison with *SlyRPL2* (Fig. 10D).

All the *SolALC* homologues are expressed ubiquitously in floral buds, sepals, petals, stamens and carpels, F1 and F2 fruits and leaves in all four species, except in F2 fruits and leaves of *B. australis* and F1 fruits of *N. obtusifolia* (Fig. 10). On the other hand, *SolSPT* homologues are expressed in floral buds and leaves of *B. australis*, *N. obtusifolia*, *C. annuum* and *S. lycopersicum*, and have different expression patterns among floral organs and fruits in each species. While *BrauSPT* is broadly expressed in all the floral organs, F1 and F2 fruits, *NiobSPT* is only expressed in sepals and petals; finally, *CaanSPT* and *SlySPT* are expressed in sepals, carpels and F1 fruits, with only *SlySPT* present in stamens (Fig. 10).

The genes belonging to the *SolHEC1/2/3* clades have more restricted expression patterns. In *B. australis*, of the two copies *BrauHEC1* and *BrauHEC3*, the former is not detected and the latter is expressed in stamens, carpels and F1 fruits (Fig. 10A). In *N. obtusifolia*, *NiobHEC1* is expressed in floral buds and leaves, *NiobHEC2* is expressed in floral buds, sepals and leaves and *NiobHEC3* is expressed in floral buds, in all floral organs and leaves (Fig. 10B). In *C. annum*, *CaanHEC1* and *CaanHEC2* genes have similar expression patterns in floral buds, F1 fruits and leaves, while *CaanHEC2-1* is strongly expressed only in leaves. Finally, *CaanHEC3* is expressed in floral buds, petals, carpel and F1 fruits (Fig. 10C). In *S. lycopersicum*, *SlyHEC1* is expressed in floral buds and petals; *SlyHEC1-1* is broadly expressed in all samples except F1 fruits. *SlyHEC2* is expressed in floral buds, sepals and F2 fruits, while *SlyHEC2-1* is expressed only in leaves. Finally, *SlyHEC3* is expressed in all floral organs, fruits and leaves (Fig. 10D).

DISCUSSION

Previous studies have shown that both *FUL* and *SHP* genes have maintained key roles in fruit development across eudicots and thus in Solanaceae (reviewed in Ferrandiz and Fourquin, 2014). However, very little is known about the rest of the fruit developmental genetic network outside Brassicaceae. Here we discuss how the genetic complement has changed for each of the gene lineages involved in fruit development in the Solanaceae compared with the canonical Brassicaceae transcription factors.

REPLUMLESS *homologues*

The *RPL* gene lineage reconstruction showed a duplication unique for Solanaceae, which contrasts greatly with the prevalence of single-copy *RPL* genes in most other angiosperms (Pabón-Mora *et al.*, 2014). Previous studies in potato had identified two copies of *RPL*, but a larger sampling across Solanaceae was lacking (Supplementary Data Figs S1 and S2; Sharma *et al.*, 2014). The fact that gene copies are found consistently in chromosomes 9 and 10 further supports the idea that the duplication coincides with an ancient whole-genome duplication (WGD) event prior to Solanaceae diversification (The Tomato Genome Consortium, 2012). Paralogues outside Solanaceae are only present in recent polyploids, like *B. rapa*, *G. max*, *G. raimondii*, *M. domestica*, *P. trichocarpa* and *T. cacao* (Walling *et al.*, 2006; Tang *et al.*, 2008; Velasco *et al.*, 2010; The Brassica rapa Genome Sequencing Project Consortium, 2011; Paterson *et al.*, 2012; Chen *et al.*, 2013).

Our MEME analysis was able to identify the SKY and the EAR/ZIBEL motifs, two conserved motifs likely important in protein function and interaction. The SKY region, together with the BELL domain, forms the MID (i.e. the MEINOX interacting domain), which is responsible for the phylogenetically conserved BELL–KNOX interaction (Hake *et al.*, 2004; Luo *et al.*, 2005; Lee *et al.*, 2008; Hamant and Pautot, 2010; Hay and Tsiantis, 2010; Yoon *et al.*, 2017). The SolRPL proteins show a consistent shift from SKY to SRF; however, as amino acids are in both cases hydrophobic and charged, it is likely that functions and interactions are largely unaffected (Supplementary Data Fig. 3A). In addition, we have identified two EAR-like motifs in RPL Solanaceae proteins. These largely correspond to the ZIBEL motifs identified by Mukherjee *et al.* (2009) at the N-terminal and C-terminal ends of the BELL proteins. However, in Solanaceae RPL sequences the N-terminal motif is GLSLSLSS and the C-terminal motif is VSLTLGL (Supplementary Data Fig. S3B, C). EAR motifs (LxLxLx) have been associated with transcriptional repression of target transcription factors by chromatin modifications, mainly through the recruitment of co-repressors, like TOPLESS, as well as an HDAC and AtHDA19 (Hiratsu *et al.*, 2004; Kieffer *et al.*, 2006; Kagale and Rozwadowski, 2011; Fujiwara *et al.*, 2014). Thus, it is likely that RPL proteins in general, but also specifically in Solanaceae, are driving repression by indirectly regulating epigenetic modifications in their targets during plant development.

Our results show that *SolRPL* genes are broadly expressed in floral buds, sepals, carpels, fruits in F1 and F2, and leaves, with some species-specific differences. Interestingly while *RPL* paralogues are expressed in early and late stages of dry dehiscent fruit development in *B. australis*, they are turned off during late developmental stages of fruit maturation in the fleshy-fruited species *C. annuum* and *S. lycopersicum* (Fig. 10). As the sequence between the BELL and the HD domains is in general shorter for fleshy-fruited species when compared with species having dry dehiscent fruits, regulation may also be changing in the downstream targets (Bencivenga *et al.*, 2016). The expression patterns observed here overlap with those reported in *Arabidopsis* RPL in the fruits, style, stem and pedicels, sepal vasculature and inflorescence meristem (Roeder *et al.*, 2003). The data suggest that it is possible that Solanaceae *RPL* genes also function in carpel and fruit patterning, as has been shown to occur in *A. thaliana* and in *Orchis italica* (Dust *et al.*, 2014; Yoon *et al.*, 2017). During gynoecium pattering and fruit development in *Arabidopsis*, *RPL* itself is negatively regulated by *APETALA2* (Ripoll *et al.*, 2011) and the role of *RPL* in fruit formation is fulfilled by the negative regulation of *SHATTERPROOF* homologues to the replum boundary (Roeder *et al.*, 2003; Hake *et al.*, 2004; Dinneny *et al.*, 2005; Kanrar *et al.*, 2006; Kumar *et al.*, 2007; Østergaard *et al.*, 2009; Etchells *et al.*, 2012; Chung *et al.*, 2013; Reyes-Olalde *et al.*, 2013; Marsch-Martínez and de Folter, 2016). Expression patterns of *AP2* in *C. annuum* show complete opposite patterns in carpel and fruit development to *RPL*, suggesting that negative regulation is conserved in fleshy fruits of Solanaceae (Zumajo-Cardona and Pabón-Mora, 2016). In addition, both expression and functional studies show that *SHP* genes are expressed during carpel development as well as early and late fruit developmental stages of tomato (Hileman *et al.*, 2006; Vrebalov *et al.*, 2009). Although more detailed spatio-temporal expression analyses are needed, the data suggest that the AP2-RPL-SHP negative regulation could be present in fleshy fruit development in the Solanaceae.

When expression patterns of *RPL* Solanaceae homologues are compared with eFP Browser data available from other core eudicots, it is clear that *RPL* genes also have broad and variable expression patterns in *Arabidopsis*, potato and soybean, with consistent expression in the SAM (shoot apical meristem) and early fruits. Previous reports describe broad expression patterns of *RPL* genes in floral organs (Yu *et al.*, 2008). Such patterns are likely to be indicative of pleiotropic redundant roles, usual for recent gene duplicates (Panchy *et al.*, 2016). However, based on the expression patterns observed, other roles, including those reported for *RPL* in *Arabidopsis* and rice, SAM initiation and boundary maintenance, stem elongation, flowering transition, internode patterning in inflorescences and formation of the abscission zone in the floral peduncle may also be part of the functions of these genes in Solanaceae (Byrne *et al.*, 2003; Roeder *et al.*, 2003; Bao *et al.*, 2004; Dinneny *et al.*, 2005; Kanrar *et al.*, 2006; Yu *et al.*, 2008; Østergaard *et al.*, 2009; Hamant and Pautot, 2010; Avino *et al.*, 2012; Etchells *et al.*, 2012; Khan *et al.*, 2012, 2015; Chung *et al.*, 2013; Arnaud and Pautot, 2014; Andrés *et al.*, 2015; Chávez-Montes *et al.*, 2015; Bencivenga *et al.*, 2016).

bHLH ALCATRAZ/SPATULA *homologues*

Previous analyses had shown different results for duplication time-points in *ALCATRAZ/SPATULA* gene evolution.

Maximum likelihood analyses revealed that *ALC* and *SPT* were the result of a core eudicot duplication, thus rescuing paralogous clades for all rosids and asterids, including the Solanaceae (Fig. 5; Pabón-Mora *et al.*, 2014; Zumajo-Cardona *et al.*, 2017). On the other hand, Bayesian analyses recover a clade of *SPT* and a grade of *ALC* genes pointing to unclear duplication events (Pfannebecker *et al.*, 2017). Here we have recovered the core eudicot duplication, with each Solanaceae clade nested in each of the paralogous clades. Evidence of ancestral WGD is further supported by the position of the two paralogues in different chromosomes. Nevertheless, this analysis also shows ALC Brassicaceae genes nested within the core eudicot SPT clade as sister to the Brassicaceae SPT clade (Supplementary Data Fig. S4). There are two alternative interpretations for such topology. The first is that the Brassicaceae had undergone a taxon-specific duplication and a loss of core eudicot ALC orthologues, which is in agreement with a more recent duplication unique to Brassicales pointed out previously by Groszmann *et al.* (2011). The second, and the one we favour, is that long-branch attraction can be occurring for the Brassicaceae ALC homologues, as they have experienced size reduction due to loss of several domains compared with ALC in other core eudicots. Additional duplications of both *SolALC* and *SolSPT* have occurred only in polyploid *Nicotiana* species, like *N. benthamiana* and *N. tabacum* (Leitch *et al.*, 2008) but not in known polyploid species within *Solanum*, like *S. tuberosum*, which shows instead alternative splicing for both *ALC* and *SPT* (The Potato Genome Sequencing Consortium, 2011). Thus, *ALC* and *SPT* genes have mostly been maintained as single copies across Solanaceae species, suggesting strong selective pressure and likely retention of function (Fig. 5).

In terms of conserved functional motifs, we have been able to detect important shifts in the flanking regions of the HLH domain, both in the basic upstream region (motif 4) and in the 5′ flanking region (motif 3). Since these regions are important for DNA binding, it is possible that ALC orthologues in the Brassicaceae are changing interactions when compared with ALC genes outside Brassicaceae, given the absence of motifs 3 and 4 in the Brassicaceae ALC proteins (Fairman *et al.*, 1993; Ferre-D'Amare *et al.*, 1994; Murre *et al.*, 1989; Nair and Burley, 2000; Toledo-Ortiz *et al.*, 2003).

As for all other functionally characterized motifs, there are no changes between the Solanaceae proteins when compared with *Arabidopsis*. bHLH (motif 1), known to control homodimerization of SPT and in general protein–protein interactions, is present in all ALC and SPT sequences across core eudicots (including Solanaceae) as well as in palaeo-SPT/ALC proteins (Groszmann *et al.*, 2008; Zumajo-Cardona *et al.*, 2017). Similarly, the acidic domain (motif 2) and the amphipathic helix (motif 7) are found across eudicots and in Solanaceae (Daingwall and Laskey, 1991; Groszmann *et al.*, 2008; Zumajo Cardona *et al.*, 2017). Interestingly, the specific amino acids present in such motifs exhibit some variation between the Solanaceae ALC and SPT, but interaction data would be needed in order to better understand the functional effect of such shifts.

Our results show that *SolALC* and *SolSPT* have broad expression patterns in flowers of all Solanaceae species sampled (Fig. 10), which largely overlap with the *Arabidopsis* expression patterns of both copies in petal margins, stamens and the developing gynoecia (Alvarez and Smyth, 1999, 2002; Rajani and Sundaresan, 2001; Groszmann *et al.*, 2010, 2011). Expression in the carpels of all Solanaceae species suggests potentially conserved roles in controlling carpel fusion, particularly in the distalmost regions, possibly regulating the medial regions, transmitting tract development and regulating style patterning (Alvarez and Smyth, 1999, 2002). Differences can be seen during fruit development, when *SolALC* and *SolSPT* genes show opposite expression patterns. In dry dehiscent fruits of *B. australis*, *ALC* is turned off during fruit maturation in F2 and *SPT* is maintained in F1 and F2, which would suggest redundant roles early in fruit development prior to the lignification of the endocarp and a more prevalent role of *SPT* during endocarp lignification and fruit maturation. Conversely, in fleshy fruits of *C. annuum* and *S. lycopersicum SPT* is turned off during fruit maturation in F2 and the expression of *ALC* remains largely unaffected (Fig. 10). Interestingly, in *N. obtusifolia* neither *ALC* nor *SPT* seems to be playing a key role in early fruit development (Fig. 10). This suggests shifts in the regulation of the two genes during fruit maturation in dry and fleshy fruits, deviating from *Arabidopsis* fruit patterning, where both *ALC* and *SPT* are expressed during fruit development in the separation layer (Groszmann *et al.*, 2010). Other reports in peach have shown expression of *PPERSPT* in the lignified endocarp during fruit maturation (Tani *et al.*, 2011), similar to our observations in *B. australis*. On the other hand, there are reports, also in peach, of largely invariable steady expression of *ALC* genes during fruit development in different tissues, more similar to our observations of fleshy fruit development in *C. annuum* and *S. lycopersicum*. Also, expression levels of *ALC* homologues in peach are independent of the levels of upstream regulators like *SHP* (Dardick *et al.*, 2010). The data available suggest that *SPT* and *ALC* genes may have specialized during fruit maturation, with a more prevalent role of *SPT* in the formation and maintenance of lignified layers, as in peach (Tani *et al.*, 2011; Dardick *et al.*, 2010), and *ALC* in the formation and maintenance of parenchymatous unlignified layers, as in *Arabidopsis* (Rajani and Sundaresan, 2001). Thus, it is possible that significant changes have likely occurred between the networks controlling dry dehiscent fruits when compared with drupes or berries.

bHLH INDEHISCENT/HECATE *homologues*

Few phylogenetic analyses are available for *HEC/IND* genes, but all analyses available coincide in that *HEC1/2* and *IND/HEC3* form two separate clades that diverged prior to the diversification of flowering plants (Supplementary Data Fig. S8; Pabón-Mora *et al.*, 2014; Pfannebecker *et al.*, 2017). In addition, *IND* is a Brassicaceae-specific paralogue and all other flowering plants only have pre-duplication homologues more similar to *HEC3* than to *IND* (Pabón-Mora *et al.*, 2014). Thus, all members of Solanaceae exhibit homologues of *HEC1*, *HEC2* and *HEC3* and lack orthologues of *IND* (Fig. 7). Moreover, all Solanaceae species investigated have additional copies of *HEC1* and *HEC2* as a result of family-specific duplications, while *HEC3* remains as a single copy in all species. There are no functional analyses of protein motifs for *HECATE* genes, which complicates the comparison between Solanaceae and Brassicaceae homologues, but our data show that Solanaceae proteins possess unique additional motifs flanking the bHLH

motif that are not found in the canonical *Arabidopsis* orthologues (Supplementary Data Figs S10–S12). Thus, protein–protein interaction studies are necessary in order to establish whether such changes are important in the activation of downstream targets during gynoecium or fruit development or both.

In *Arabidopsis HECATE1*, *HECATE2* and *HECATE3* are expressed in the developing septum, the transmitting tract, the developing ovules and the stigma, with *HEC3* always being expressed more strongly and longer during gynoecium development than *HEC1* and *HEC2* (Gremski *et al*., 2007) and *HEC1*, controlling *PINOID* (*PIN*) expression and promoting auxin transport (Schuster *et al*., 2015). In addition, as they show some degree of redundancy, only *hec1*, *hec2* and *hec3* mutants in *Arabidopsis* show absence of stigmatic tissue accompanied by complete infertility, a phenotype that becomes enhanced due to defects in the formation of a reproductive tract when *spt* is also mutated, as all *HECATE* paralogues interact with *SPT* (Gremski *et al*., 2007). On the other hand *ind* mutants have defects in the formation of the lignified layer of the dehiscence zone (Liljegren *et al*., 2004). In Solanaceae, only *HECATE3* genes seem to be involved in gynoecium patterning as they show consistent expression during carpel development in all Solanaceae species sampled (Fig. 10). All other *HEC* genes are expressed in sepals, as in *N. obtusifolia*, and during fruit development in early stages, as in *C. annuum*, or in late developmental stages, as in *S. lycopersicum*, perhaps compensating, together with *HEC3*, for the absence of *IND* orthologues during fruit maturation. It is unclear what the role of *HEC* genes during fruit maturation is, but it has been suggested that in peach *IND* (i.e. the *HEC3-like* homologue) does not show tissue specificity or substantial expression changes during maturation, and only declining expression is seen in late developmental stages of peach fruit development (Dardick *et al*., 2010). Our data also show that recent duplicates exclusive to Solanaceae, like *HEC2-1*, have acquired restriction of expression to leaves and are likely no longer involved in gynoecium patterning of fruit development.

Conclusions

Based on our analyses, the Solanaceae would have a lot more genetic redundancy when compared with Brassicaceae in all gene lineages involved in gynoecium patterning and fruit development, with the sole exception of *SHP* genes, which are duplicated in Brassicaceae and are single-copy in Solanaceae. However, only *RPL*, *SPT*, *ALC* and *HEC3* are consistently expressed during gynoecium development in all four species evaluated, suggesting that *HEC1* and 2 are likely not redundant in carpel patterning with *HEC3*, as they are in Brassicaceae. In addition, our data show opposite expression patterns of *RPL*, *ALC* and *SPT* during fleshy fruit development versus dry dehiscent fruit development. Finally, our data suggest that it is downstream of *FUL–SHP* regulation where major shifts occur that are likely to result in fruit histogenesis changes, at least in Solanaceae.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Fig. S1: maximum likelihood tree of *REPLUMLESS* genes in angiosperms. Fig. S2: sequences of the conserved motifs detected by the MEME analysis in the REPLUMLESS homologues in Solanaceae and selected functionally characterized proteins from *Arabidopsis thaliana* (RPL), *Arabidopsis lyrata* (ArlyRPL), *Zea mays* (ZemaRPL) and *Oryza sativa* (qSH1). Fig. S3: specific regions of the REPLUMLESS alignments pointing to the known motifs for RPL homologues (A–C), the new motifs exclusive to Solanaceae (D–F) and the differential motifs within Solanaceae that are divergent between the species having dry dehiscent and fleshy fruits (G–I). Fig. S4: maximum likelihood tree of *ALCATRAZ/SPATULA* genes in angiosperms. Outgroup used corresponds to *Amborella trichopoda SPATULA* (*AmtrSPT*). Fig. S5: sequences of the conserved motifs detected by the MEME analysis in *ALCATRAZ/SPATULA* homologues in Solanaceae and selected functionally characterized proteins from *Arabidopsis thaliana* (AtALC and AtSPT), *Fragaria vesca* (FaSPT) and *Prunus persica* (PPERSPT). Fig. S6: specific regions of the ALCATRAZ alignments pointing to the differential motifs within Solanaceae that are divergent between species having dry dehiscent and fleshy fruits. Fig. S7: specific regions of the SPATULA alignments pointing to the differential motifs within Solanaceae that are divergent between species having dry dehiscent and fleshy fruits. Fig. S8: maximum likelihood tree of *HECATE/INDEHISCENT* genes in angiosperms. Fig. S9: sequences of the conserved motifs detected by the MEME analysis of the HECATE/INDEHISCENT proteins in Solanaceae and selected functionally characterized proteins from *Arabidopsis thaliana* (AtHEC1, AtHEC2, AtHEC3 and ATIND). Fig. S10: specific regions of the HECATE1 alignments pointing to the differential motifs within Solanaceae that are divergent between species having dry dehiscent and fleshy fruits. Fig. S11: specific regions of the HECATE2 alignments pointing to the differential motifs within Solanaceae that are divergent between species having dry dehiscent and fleshy fruits. Fig. S12: specific regions of the HECATE3 alignments pointing to the differential motifs within Solanaceae that are divergent between species having dry dehiscent and fleshy fruits. Table S1: accession numbers for all sequences used in the main figures of this study. Table S2: 5′–3′ sequence for all primers used in Fig. 10.

ACKNOWLEDGEMENTS

This work was funded by COLCIENCIAS (111565842812), the Committee for Research development (CODI), Convocatoria de Internacionalización 2015 at the Universidad de Antioquia and iCOOP+2016 grant COOPB20250 from Centro Superior de Investigación Científica, CSIC. We thank Juan Fernando Alzate (Centro Nacional de Secuenciación de Genómica, SIU, Universidad de Antioquia, Medellín, Antioquia) for the assembly and storage of our own generated transcriptomes. We thank the OneKP project for allowing us to access the database. We thank Amy Litt (UC Riverside) and Barbara Ambrose (The New York Botanical Garden) for providing additional reagents for RT–PCR. We thank Cecilia Zumajo-Cardona, Harold Suárez-Baron, Favio González, Ricardo Callejas, and Cristina Ferrandiz for their help and suggestions during different parts of this research project.

LITERATURE CITED

Altschul SF, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ. 1990**. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Alvarez J, **Smyth DR. 1999**. CRABS CLAW and SPATULA, two *Arabidopsis* genes that control carpel development in parallel with AGAMOUS. *Development* **126**: 2377–2386.

Alvarez J, **Smyth DR. 2002**. CRABS CLAW and SPATULA genes regulate growth and pattern formation during gynoecium development in *Arabidopsis thaliana*. *International Journal of Plant Science* **163**: 17–41.

Andrés F, **Romera Branchat M**, **Martinez-Gallegos R**, ***et al.*** **2015**. Floral induction in *Arabidopsis* by FLOWERING LOCUS T requires direct repression of BLADE-ON-PETIOLE genes by the homeodomain protein PENNYWISE. *Plant Physiology* **169**: 2187–2199.

Arnaud N, **Pautot V. 2014**. Ring the BELL and tie the KNOX: roles for TALEs in gynoecium development. *Frontiers in Plant Science* **5**: 1–7.

Avino M, **Kramer EM**, **Donohue K**, **Hammel AJ**, **Hall JC. 2012**. Understanding the basis of a novel fruit type in Brassicaceae: conservation and deviation in expression patterns of six genes. *EvoDevo* **3**: 20.

Bailey TL, **Williams N**, **Misleh C**, **Li WW. 2006**. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* **34**: 369–373.

Bao X, **Franks RG**, **Levin JZ**, **Liu Z. 2004**. Repression of AGAMOUS by BELLRINGER in floral and inflorescence meristems. *Plant Cell* **16**: 1478–1489.

Bemer M, **Karlova R**, **Ballester AR**, ***et al.*** **2012**. The tomato FRUITFULL homologs TDR4/FUL1 and MBP/FUL2 regulate ethylene independent aspects of fruit ripening. *Plant Cell* **24**: 4437–4451.

Bencivenga S, **Serrano-Mislata A**, **Bush M**, **Fox S**, **Sablowski R. 2016**. Control of oriented tissue growth through repression of organ boundary genes promotes stem morphogenesis. *Developmental Cell* **39**: 198–208.

Byrne M, **Groover AT**, **Fontana JR**, **Martienssen RA. 2003**. Phyllotactic pattern and stem cell fate are determined by the *Arabidopsis* homeobox gene BELLRINGER. *Development* **130**: 3941–3950.

Chávez Montes RA, **Herrera-Ubaldo H**, **Serwatowska J**, **de Folter S. 2015**. Towards a comprehensive and dynamic gynoecium gene regulatory network. *Current Plant Biology* 3–4: 3–12. http://dx.doi.org/10.1016/j.cpb.2015.08.002.

Chen ECH, **Abad Najar CFB**, **Zheng C**, ***et al.*** **2013**. The dynamics of functional classes of plant genes in rediploidized ancient polyploids. *BMC Bioinformatics* **14**: S15–S19.

Chung KS, **Lee JH**, **Lee JS**, **Ahn JH. 2013**. Fruit indehiscence caused by enhanced expression of NO TRANSMITTING TRACT in *Arabidopsis thaliana*. *Molecular Cell* **35**: 519–525.

Chung MY, **Vrebalov J**, **Alba R**, ***et al.*** **2010**. A tomato (*Solanum lycopersicum*) APETALA2/ERF gene, SlAP2a, is a negative regulator of fruit ripening. *Plant Journal* **64**: 936–947.

Daingwall C, **Laskey RA. 1991**. Nuclear targeting sequence a consensus? *Trends Biochem Sci* **16**: 478–481.

Dardick CD, **Callahan AM**, **Chiozzotto R**, **Schaffer RJ**, **Piagnani MC**, **Scorza R. 2010**. Stone formation in peach fruit exhibits spatial coordination of the lignin and flavonoid pathways and similarity to *Arabidopsis* dehiscence. *BMC Biology* **8**: 13.

Dinneny JR, **Weigel D**, **Yanofsky MF. 2005**. A genetic framework for fruit patterning in *Arabidopsis thaliana*. *Development* **132**: 4687–4696.

Dust AN, **Mauro-Herrera M**, **Francis AD**, **Shand L. 2014**. Morphological diversity and genetic regulation of inflorescence abscission zones in grasses. *American Journal of Botany* **101**: 1759–1769.

Etchells JP, **Moore L**, **Jiang WZ**, ***et al.*** **2012**. A role for *BELLRINGER* in cell wall development is supported by loss-of-function phenotypes. *BMC Plant Biology* **12**: 212.

Fairman R, **Beran-Steed RK**, **Anthony-Cahill SJ**, ***et al.*** **1993**. Multiple oligomeric states regulate the DNA binding of helix-loop-helix peptides. *Proceedings of the National Academy of Sciences of the USA* **90**: 10429–10433.

Ferrándiz C. 2002. Regulation of fruit dehiscence in *Arabidopsis*. *Journal of Experimental Botany* **53**: 2031–2038.

Ferrándiz C, **Fourquin C. 2014**. Role of the FUL-SHP network in the evolution of fruit morphology and evolution. *Journal of Experimental Botany* **65**: 4505–4513.

Ferre-D'Amare AR, **Pognonec P**, **Roeder RG**, **Burley SK. 1994**. Structure and function of the b/HLH/Z domain of USF. *EMBO Journal* **13**: 180–189.

Fourquin C, **Ferrandiz C. 2012**. Functional analyses of AGAMOUS family members in *Nicotiana benthamiana* clarify the evolution of early and late roles of C-function genes in eudicots. *Plant Journal* **71**: 990–1001.

Fujisawa M, **Shima Y**, **Nakagawa H**, ***et al.*** **2014**. Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS-box proteins. *Plant Cell* **26**: 89–101.

Fujiwara S, **Sakamoto S**, **Kigoshi K**, **Suzuki K**, **Ohme-Takagi M. 2014**. VP16 fusion induces the multiple-knockout phenotype of redundant transcriptional repressors partly by Med25-independent mechanisms in *Arabidopsis*. *FEBS Letters* **588**: 3665–3672.

Gebhardt C. 2016. The historical role of species from the Solanaceae plant family in genetic research. *Theoretical and Applied Genetics* **129**: 2281–2294.

Girin T, **Stephenson P**, **Goldsack CM**, **Kempin SA**, **Perez A**, **Pires N. 2010**. Brassicaceae INDEHISCENT genes specify valve margin cell fate and repress replum formation. *Plant Journal* **63**: 329–338.

Girin T, **Paicu T**, **Fuentes S**, **O'Brien M**, ***et al.*** **2011**. INDEHISCENT and SPATULA interact to specify carpel and valve margin tissue and thus promote seed dispersal in *Arabidopsis*. *Plant Cell* **23**: 3641–3653.

Gremski K, **Ditta G**, **Yanofsky MF. 2007**. The HECATE genes regulate female reproductive tract development in Arabidopsis thaliana. *Development* **134**: 3593–3601.

Groszmann M, **Paicu T**, **Smyth DR. 2008**. Functional domains of SPATULA, a bHLH transcription factor involved in carpel and fruit development in *Arabidopsis*. *Plant Journal* **55**: 40–52.

Groszmann M, **Bylstra Y**, **Lampugnani ER**, **Smyth DR. 2010**. Regulation of tissue-specific expression of SPATULA, a bHLH gene involved in carpel development, seedling germination, and lateral organ growth in *Arabidopsis*. *Journal of Experimental Botany* **61**: 1495–1508.

Groszmann M, **Paicu T**, **Alvarez JP**, **Swain SM**, **Smyth DR. 2011**. SPATULA and ALCATRAZ are partially redundant, functionally divergent bHLH genes required for *Arabidopsis* gynoecium and fruit development. *Plant Journal* **68**: 816–829.

Gu Q, **Ferrándiz C**, **Yanofsky MF**, **Martienssen R. 1998**. The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* **125**: 1509–1517.

Hake S, **Smith HMS**, **Holtan H**, **Magnani E**, **Mele G**, **Ramirez J. 2004**. The role of KNOX genes in plant development. *Annual Review of Cell and Developmental Biology* **20**: 125–151.

Hamant O, **Pautot V. 2010**. Plant development: a TALE story. *Comptes Rendus Biologies* **333**: 371–381. https://doi.org/10.1016/j.crvi.2010.01.015.

Hay A, **Tsiantis M. 2010**. KNOX genes: versatile regulators of plant development and diversity. *Development* **137**: 3153–3165.

Hileman LC, **Sundstrom JF**, **Litt A**, **Chen M**, **Shumba T**, **Irish VF. 2006**. Molecular and phylogenetic analyses of the MADS-box gene family in tomato. *Molecular Biology and Evolution* **23**: 2245–2258.

Hiratsu K, **Mitsuda N**, **Matsui K**, **Ohme-Takagi M. 2004**. Identification of the minimal repression domain of SUPERMAN shows that the DLELRL hexapeptide is both necessary and sufficient for repression of transcription in *Arabidopsis*. *Biochemical and Biophysical Research Communications* **321**: 172–178.

Itkin M, **Seybold H**, **Breitel D**, **Rogachev I**, **Meir S**, **Aharoni A. 2009**. TOMATO AGAMOUS-LIKE 1 is a component of the fruit ripening regulatory network. *Plant Journal* **60**: 1081–1095.

Ito Y, **Nishizawa-Yokoi A**, **Endo M**, ***et al.*** **2017**. Re-evaluation of the rin mutation and the role of RIN in the induction of tomato ripening. *Nature Plants* **3**: 866–874.

Katoh K, **Misawa K**, **Kuma K**, **Miyata T. 2002**. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.

Kanrar S, **Onguka O**, **Smith HMS. 2006**. *Arabidopsis* inflorescence architecture requires the activities of KNOX-BELL homoeodomain heterodimers. *Planta* **224**: 1163–1173.

Kay P, **Groszmann M**, **Ross JJ**, **Parish RW**, **Swain SM. 2013**. Modifications of a conserved regulatory network involving INDEHISCENT controls multiple aspects of reproductive tissue development in *Arabidopsis*. *New Phytologist* **197**: 73–87.

Kagale S, **Rozwadowski K. 2011**. EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* **6**: 2.

Khan M, **Tabb P**, **Hepworth SR. 2012**. BLADE-ON-PETIOLE1 and 2 regulate *Arabidopsis* inflorescence architecture in conjunction with homeobox genes KNAT6 and ATH1. *Plant Signaling and Behavior* **7**: 7.

Khan M, **Ragni L**, **Tabb P**, **Salasini BC**, ***et al.*** **2015**. Repression of lateral organ boundary genes by PENNYWISE and POUND-FOOLISH is

essential for meristem maintenance and flowering in *Arabidopsis*. *Plant Physiology* **169**: 2166–2186.
Kieffer M, **Stern Y**, **Cook H**, ***et al.*** **2006**. Analysis of the transcription factor WUSCHEL and its functional homologue in *Antirrhinum* reveals a potential mechanism for their roles in meristem maintenance. *Plant Cell* **18**: 560–573.
Knapp S. 2002. Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *Journal of Experimental Botany* **53**: 2001–2022.
Kumar R, **Kushalappa K**, **Godt D**, ***et al.*** **2007**. The *Arabidopsis* BEL1-LIKE HOMEODOMAIN proteins SAW1 and SAW2 act redundantly to regulate KNOX expression spatially in leaf margins. *Plant Cell* **19**: 2719–2735..
Kumar S, **Stecher G**, **Tamura K. 2015**. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Mol Biol Evol*. **33**: 1870–1874.
Lee JH, **Lin H**, **Joo S**, **Goodenough U. 2008**. Early Sexual origins of homeoprotein heterodimerization and evolution of the plant KNOX/BELL family. *Cell* **133**: 829–840.
Leitch IJ, **Hanson L**, **Lim KY**, ***et al.*** **2008**. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Annals of Botany* **101**: 805–814.
Leseberg CH, **Eissler CL**, **Wang X**, **Johns MA**, **Duvall MR**, **Mao L. 2008**. Interaction study of MADS-domain proteins in tomato. *Journal of Experimental Botany* **59**: 2253–2265.
Liljegren SJ, **Ditta GS**, **Eshed Y**, **Savidge B**, **Bowman JL**, **Yanofsky MF. 2000**. SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**: 766–770.
Liljegren SJ, **Roeder AHK**, **Kempin SA**, **Gremski K**, **Ostergaard L. 2004**. Control of fruit patterning in *Arabidopsis* by INDEHISCENT. *Cell* **116**: 843–853.
Luo H, **Song F**, **Goodman RM**, **Zheng Z. 2005**. Up-regulation of OsBIHD1, a rice gene encoding BELL homeo-domain transcriptional factor, in disease resistance responses. *Plant Biology* **7**: 459–468.
Martel C, **Vrebalov J**, **Tafelmeyer P**, **Giovannoni JJ. 2011**. The tomato MADS-box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening process in a COLORLESS NONRIPENING dependent manner. *Plant Physiology* **157**: 1568–1579.
Marsch-Martínez N, **de Folter S. 2016**. Hormonal control of the development of the gynoecium. *Current Opinion in Plant Biology* **29**: 104–114..
Miller MA, **Pfeiffer W**, **Schwartz T. 2010**. Creating the CIPRES Science Gateway for interference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop* (GCE). doi:10.1109/GCE.2010.5676129. http://ieeexplore.ieee.org/document/5676129/.
Mühlhausen A, **Lenser T**, **Mummenhoff K**, **Theißen G. 2013**. Evidence that an evolutionary transition from dehiscent to indehiscent fruits in *Lepidium* (Brassicaceae) was caused by a change in the control of valve margin identity genes. *Plant Journal* **73**: 824–835.
Mukherjee K, **Brocchieri L**, **Bürglin TR. 2009**. A comprehensive classification and evolutionary analysis of plant homeobox genes. *Molecular Biology and Evolution* **26**: 2775–2794.
Murre C, **McCaw PS**, **Baltimore D. 1989**. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD and myc proteins. *Cell* **56**: 777–783.
Nair SK, **Burley SK. 2000**. Recognizing DNA in the library. *Nature* **404**: 715–717.
Olmstead RG, **Bohs L**, **Migi HA**, **Valentin ES**, **Garcia VF**, **Collier SM. 2008**. A molecular phylogeny of the Solanaceae. *Molecular phylogeny* **57**: 1159–1181.
Østergaard L. 2009. Don't leaf now. The making of a fruit. *Current Opinion in Plant Biology* **12**: 36–41.
Pabón-Mora N, **Litt A. 2011**. Comparative anatomical and developmental analysis of dry and fleshy fruits of Solanaceae. *American Journal of Botany* **98**: 1415–1436.
Pabón-Mora N, **Wong KG**, **Ambrose BA. 2014**. Evolution of fruit development genes in flowering plants. *Frontiers in Plant Science* **5**: 300.
Pan IL, **McQuinn R**, **Giovannoni JJ**, **Irish VF. 2010**. Functional diversification of AGAMOUS lineage genes in regulating tomato flower and fruit development. *Journal of Experimental Botany* **61**: 1795–1806.
Panchy N, **Lehti-Shiu M**, **Shiu S-H. 2016**. Evolution of gene duplication in plants. *Plant Physiology* **171**: 2294–2316.
Paterson AH, **Wendel JF**, **Gundlach H**, ***et al.*** **2012**. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**: 423–428.
Pesaresi P, **Mizzotti C**, **Colombo M**, **Masiero S. 2014**. Genetic regulation and structural changes during tomato fruit development and ripening. *Frontiers in Plant Science* **5**: 124.
Pfannebecker KC, **Lange M**, **Rupp O**, **Becker A. 2017**. Seed plant-specific gene lineages involved in carpel development. *Molecular Biology and Evolution* **34**: 925–942.
Pires N, **Dolan L. 2010**. Early evolution of bHLH proteins in plants. *Plant Signal Behav* **7**: 911–912.
Rajani S, **Sundaresan V. 2001**. The *Arabidopsis* myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Current Biology* **11**: 1914–1922.
Reyes-Olalde JI, **Zuñiga-Mayo VM**, **Chavez Montes RA**, **Marsch-Martinez N**, **de Folter S. 2013**. Inside the gynoecium: at the carpel margin. *Trends in Plant Science* **18**: 644–655.
Ripoll JJ, **Roeder AH**, **Ditta GS**, **Yanofsky MF. 2011**. A novel role for the floral homeotic gene APETALA2 during *Arabidopsis* fruit development. *Development* **138**: 5167–5176.
Roeder AHK, **Ferrándiz C**, **Yanofsky MF. 2003**. The role of the REPLUMLESS homeodomain protein in patterning the *Arabidopsis* fruit. *Current Biology* **13**: 1630–1635.
Särkinen T, **Bohs L**, **Olmstead RG**, **Knapp S. 2013**. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology* **13**: 214.
Schuster C, **Gaillochet C**, **Lohmann JU. 2015**. Arabidopsis HECATE genes function in phytohormone control during gynoecium development. *Development* **142**: 3343–3350. doi: 10.1242/dev.120444.
Sharma P, **Lin T**, **Grandellis C**, **Yu M**, **Hannapel DJ. 2014**. The BEL1-like family of transcription factors in potato. *Journal of Experimental Botany* **65**: 709–723.
Smykal P, **Gennen J**, **De Bodt S**, **Ranganath V**, **Melzer S. 2007**. Flowering of strict photoperiodic *Nicotiana* varieties in non-inductive conditions by transgenic approaches. *Plant Molecular Biology* **65**: 233–242.
Stamatakis A, **Hoover P**, **Rougemont J. 2008**. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**: 758–771.
Tanksley SD. 2004. The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* **16**: 181–189.
Tang H, **Bowers J**, **Wang XX**, **Ming R**, **Alam M**, **Paterson AH. 2008**. Synteny and collinearity in plant genomes. *Science* **320**: 486–488. https://www.ncbi.nlm.nih.gov/pubmed/18436778.
Tani E, **Tsaballa A**, **Stedel C**, ***et al.*** **2011**. The study of a SPATULA-like bHLH transcription factor expressed during peach (*Prunus persica*) fruit development. *Plant Physiology and Biochemistry* **49**: 654–663.
The Brassica rapa Genome Sequencing Project Consortium. **2011**. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43**: 1035–1039.
The Potato Genome Sequencing Consortium. **2011**. Genome sequence and analysis of the tuber crop potato. *Nature* **475**: 189–195.
The Tomato Genome Consortium. **2012**. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**: 635–641.
Toledo-Ortiz G, **Huq E**, **Quail PH. 2003**. The *Arabidopsis* basic/helix-loop-helix transcription factor family. *Plant Cell* **15**: 1749–1770.
Velasco R, **Zharkikh A**, **Affourtit J**, ***et al.*** **2010**. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* **42**: 833–839.
Vrebalov J, **Ruezinsky D**, **Padmanabhan V**, ***et al.*** **2002**. A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science* **296**: 343–346.
Vrebalov J, **Pan IL**, **Arroyo AJ**, ***et al.*** **2009**. Fleshy fruit expansion and ripening are regulated by the tomato SHATTERPROOF gene TAGL1. *Plant Cell* **21**: 3041–3062.
Walling JG, **Shoemaker R**, **Young N**, **Mudge J**, **Jackson S. 2006**. Chromosome-level homeology in paleopolyploid soybean (*Glycine max*) revealed through integration of genetic and chromosome maps. *Genetics* **172**: 1893–1900.
Wang L, **Li J**, **Zhao J**, **He C. 2015**. Evolutionary developmental genetics of fruit morphological variation within the Solanaceae. *Frontiers in Plant Science* **6**: 248.
Yoon J, **Cho L-H**, **Antt HW**, **Koh H-J**, **An G. 2017**. KNOX protein OSH15 induces grain shattering by repressing lignin biosynthesis genes. *Plant Physiology* **174**: 312–325.
Yu L, **Patibanda V**, **Smith HMS. 2008**. A novel role of BELL1-like homeobox genes PENNYWISE and POUND-FOOLISH in floral patterning. *Planta* **229**: 693–707.
Zumajo-Cardona C, **Pabón-Mora N. 2016**. Evolution of the APETALA2 gene lineage in seed plants. *Molecular Biology and Evolution* **33**: 1818–1832.
Zumajo-Cardona C, **Ambrose B**, **Pabón Mora N. 2017**. Evolution of the *SPATULA/ALCATRAZ* gene lineage and expression analyses in the basal eudicot, *Bocconia frutescens* L. (Papaveraceae). *EvoDevo* **8**: 5.