

# The *Huperzia selago* Shoot Tip Transcriptome Sheds New Light on the Evolution of Leaves

Anastasiia I. Evkaikina<sup>1,†</sup>, Lidija Berke<sup>2,†</sup>, Marina A. Romanova<sup>3</sup>, Estelle Proux-Wéra<sup>4,5</sup>, Alexandra N. Ivanova<sup>6</sup>, Catarina Rydin<sup>7</sup>, Katharina Pawlowski<sup>7,\*</sup>, and Olga V. Voitsekhovskaja<sup>1</sup>

<sup>1</sup>Laboratory of Molecular and Ecological Physiology, Komarov Botanical Institute, Russian Academy of Sciences, St. Petersburg, Russia

<sup>2</sup>Department of Plant Sciences, Wageningen University, The Netherlands

<sup>3</sup>Department of Botany, St. Petersburg State University, St. Petersburg, Russia

<sup>4</sup>Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Alnarp, Sweden

<sup>5</sup>Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Solna, Sweden

<sup>6</sup>Laboratory of Anatomy and Morphology, Komarov Botanical Institute, Russian Academy of Sciences, St. Petersburg, Russia

<sup>7</sup>Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding author: E-mail: katharina.pawlowski@su.se.

Accepted: August 28, 2017

**Data deposition:** Raw sequence reads used for the *de novo* transcriptome assembly were deposited at NCBI GenBank (accession PRJNA281995). *Huperzia selago* phylogenetic markers: KX761189, KX761188, KX761187 KNOX cDNAs: KX761181 - KX761185 YABBY cDNA: KX761186 YABBY genomic: MF175244.

## Abstract

Lycopodiophyta—consisting of three orders, Lycopodiales, Isoetales and Selaginellales, with different types of shoot apical meristems (SAMs)—form the earliest branch among the extant vascular plants. They represent a sister group to all other vascular plants, from which they differ in that their leaves are microphylls—that is, leaves with a single, unbranched vein, emerging from the protostele without a leaf gap—not megaphylls. All leaves represent determinate organs originating on the flanks of indeterminate SAMs. Thus, leaf formation requires the suppression of indeterminacy, that is, of KNOX transcription factors. In seed plants, this is mediated by different groups of transcription factors including ARP and YABBY.

We generated a shoot tip transcriptome of *Huperzia selago* (Lycopodiales) to examine the genes involved in leaf formation. Our *H. selago* transcriptome does not contain any ARP homolog, although transcriptomes of *Selaginella* spp. do. Surprisingly, we discovered a YABBY homolog, although these transcription factors were assumed to have evolved only in seed plants.

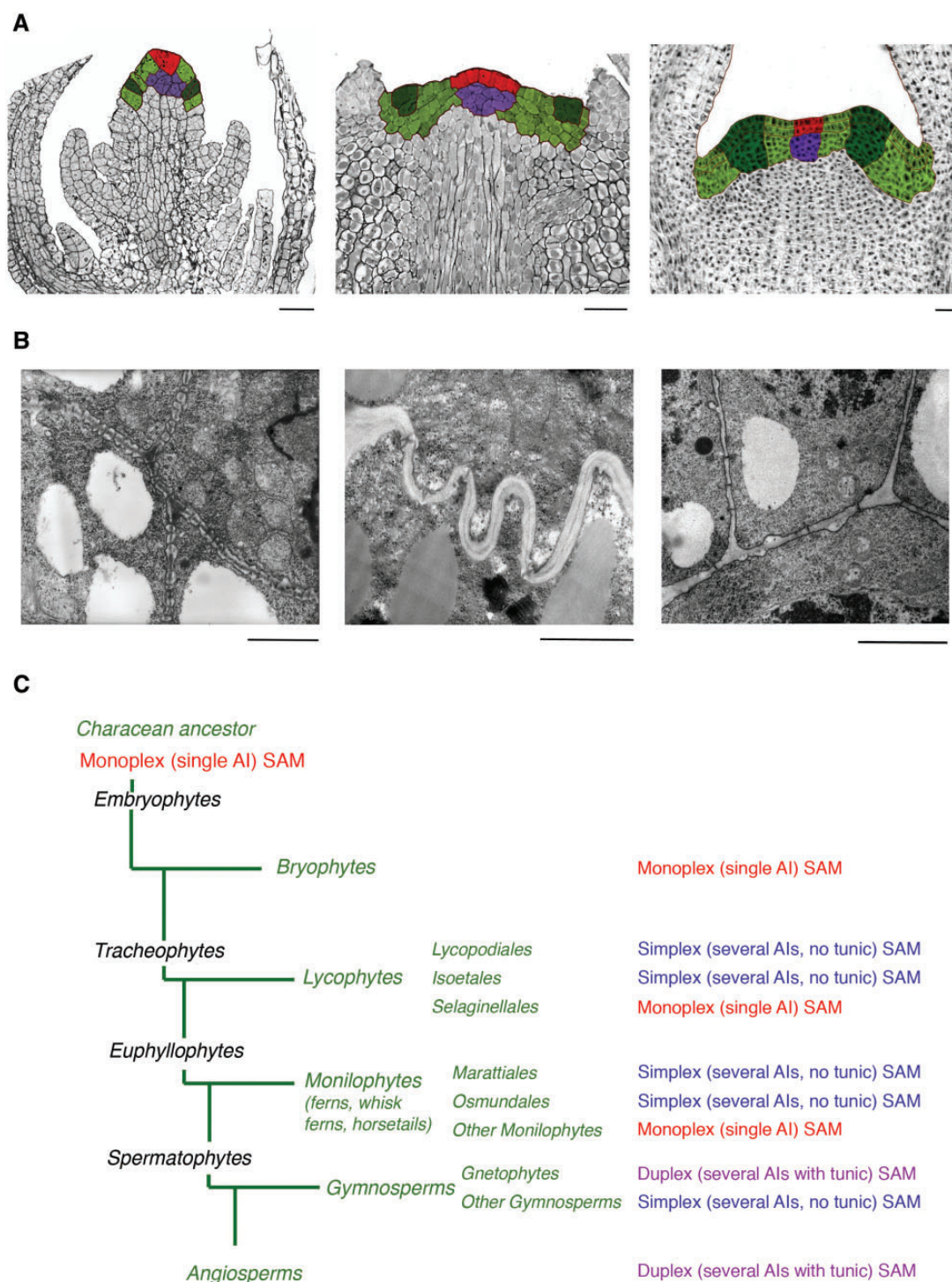
The existence of a YABBY homolog in *H. selago* suggests that YABBY evolved already in the common ancestor of the vascular plants, and subsequently was lost in some lineages like Selaginellales, whereas ARP may have been lost in Lycopodiales. The presence of YABBY in the common ancestor of vascular plants would also support the hypothesis that this common ancestor had a simplex SAM. Furthermore, a comparison of the expression patterns of ARP in shoot tips of *Selaginella kraussiana* (Harrison CJ, et al. 2005. Independent recruitment of a conserved developmental mechanism during leaf evolution. *Nature* 434(7032):509–514.) and YABBY in shoot tips of *H. selago* implies that the development of microphylls, unlike megaphylls, does not seem to depend on the combined activities of ARP and YABBY. Altogether, our data show that Lycopodiophyta are a diverse group; so, in order to understand the role of Lycopodiophyta in evolution, representatives of Lycopodiales, Selaginellales, as well as of Isoetales, have to be examined.

**Key words:** Lycopodiophyta, *Huperzia*, shoot apex, leaf development, KNOX, ARP, YABBY.

## Introduction

All aboveground organs of land plants originate from the shoot apical meristem (SAM). Three main types of SAMs exist

(fig. 1). Angiosperms and Gnetopsida have duplex SAMs whose apical initials (AIs) are located in the central part of the two to three outermost cell layers (the tunica) and divide



**FIG. 1.**—Shoot apical meristems (SAMs) in land plants. (A) Types of the SAM. From left to right: monoplex (single Apical Initial, AI) SAM of *Selaginella kraussiana*, simplex (several AIs in the surface layer, that divide both anticlinally and periclinally) of *Huperzia selago*, duplex (several AIs in two or three outermost layers that divide exclusively anticlinally, also known as tunica-corporis type) in *Syringa vulgaris*. AIs are shown in red, subapical cells in purple, peripheral cells in light green, and cells of leaf primordia in dark green. (B) Morphology and density of plasmodesmata in different SAM types: numerous unbranched (simple) plasmodesmata in the walls between the apical cell and its immediate derivatives in *S. kraussiana*; scarce, mainly branched H-shaped and Y-shaped plasmodesmata in the basal walls of the peripheral AIs in *H. selago*; simple rare plasmodesmata in the walls between cells of layers L1 and L2 of the tunica in the duplex SAM of potato. (C) Distribution of the SAM types throughout the main groups of higher plants (Kenrick and Crane, 1997; Judd et al. 2002): This scheme infers that all SAMs evolved from a plesiomorphic monoplex precursor, with independent origins of simplex SAMs in lycophytes and gymnosperms, respectively. Bar: 50  $\mu\text{m}$  (A); 1  $\mu\text{m}$  (B).

anticlinally to form the epidermis and the outer cortex (Gifford and Foster 1989; Steeves and Sussex 1989). The initials of the layers below the tunica, the corpus, divide in various planes to form the ground and vascular tissues. Therefore, in these plants tunica and corpus are not clonally related (Newman 1965; Philipson 1990; Popham 1951; fig. 1A). Most monilophytes (i.e., ferns s.l.) and some lycophytes (Selaginellales) have monoplex SAMs with a single tetrahedral AI in the superficial cell layer of the apex, which is the ultimate source of all shoot apex cells (Bierhorst 1977; Gifford 1983). Simplex SAMs with few to several AIs in the superficial layer exist in other lycophytes (Lycopodiales and Isoetales), in some monilophytes as well as in most gymnosperms (Ambrose and Vasco 2016; Buvat 1989). Here, the AIs divide both anticlinally and periclinally, thereby producing peripheral and subsurface cells of the SAM, respectively.

The three types of SAMs differ in origin and distribution of the plasmodesmata (PD) connecting their cells. In plants, PD between cells can be primary or secondary with regard to their origin. Primary PD develop during cytokinesis and connect “sister cells” from the same cell lineage. Secondary PD arise de novo, that is, postcytokinetically, between cells, which belong to different cell lineages. All cells within a monoplex SAM are clonally related and connected exclusively by primary PD (Gunning 1978; Imaichi and Hiratsuka 2007); indeed, a hypothesis has been put forward that the presence of a single AI in plants with monoplex SAMs is a consequence of the inability of the cells to form secondary PD (Cooke et al. 1996). This might explain the presence of a gradient of PD density in monoplex SAMs where PD abundance gradually decreases from the AI towards the distant cells in the apex (Gunning 1978; Imaichi and Hiratsuka 2007). In simplex and duplex SAMs, cells can be connected by both primary and secondary PD, and no PD density gradients have been observed (Imaichi and Hiratsuka 2007). Altogether, the organization of the borders between SAM domains is drastically different in the three types of SAMs (Evkaikina et al. 2014; fig. 1B).

Interestingly, two types of SAMs can be found within the earliest branch of vascular plants, the lycophytes (i.e., Lycopodiophyta; fig. 1C), one of them (simplex) with clear seed plant-like traits such as formation of secondary PD between cells and the presence of multiple AIs. The type of SAM in lycophytes shows no correlation with homo-/heterospory; lycophytes with simplex SAMs include homosporous Lycopodiales and heterosporous Isoetales, whereas the heterosporous Selaginellales have monoplex SAMs. Neither does the type of SAM correlate with leaf type since all members of the Lycopodiophyta are inferred to have the same fundamental leaf type, which contrasts with that of other vascular plants. These leaf types have long been assumed to have evolved independently in the two major groups of vascular plants: microphylls in the Lycopodiophyta and megaphylls in euphyllophytes (monilophytes and seed plants). Microphylls

are always simple in shape and contain a single vascular trace, whereas megaphylls develop complex vasculature and variable shape. Although the evolutionary origin of megaphylls as a result of gradual modifications of leafless axes, or telomes of rhyniophytes sensu lato, is widely accepted, the pattern of microphyll origin is still controversial. Three different concepts of their origin are under discussion: concept 1) that microphylls, like megaphylls, evolved by gradual reduction of lateral telomes (Zimmerman 1952); 2) the de novo origin of microphylls as stem tissue outgrowths (enations; e.g., Bower 1908, p. 552–554); and 3) the sterilization theory which sees microphylls as modified sporangia (Kenrick and Crane 1997).

At any rate, however, both microphylls and megaphylls are determinate organs originating on the flanks of indeterminate SAMs (Gifford and Foster 1989). Thus, in all vascular plants leaf formation involves the addition of a determinate growth program to the indeterminate growth program of the SAM. Class I KNOTTED1-like homeobox (*KNOX*) genes encode the key transcription factors that maintain the cells in an undifferentiated state (Byrne 2012; Kessler and Sinha 2004). They act in a noncell-autonomous manner, moving between cells via PD (Carles and Fletcher 2003; Gaillochet et al. 2015; Holt et al. 2014; Perales and Reddy 2012). The functions of *KNOX* I proteins are conserved among Selaginellales, ferns and angiosperms (Harrison et al. 2005). Chen et al. (2013, 2014) have suggested that the ability of *KNOX* I proteins to efficiently traffic through PD was acquired early in the evolution of land plants, that is, already in Bryophyta, which only contain primary PD. Simplex SAMs in Lycopodiales/Isoetales as well as duplex SAMs in angiosperms require secondary PD. There is evidence for differences in trafficking through primary and secondary PD (Ding et al. 1992; Hake and Freeling 1986).

Determinate leaf growth thus requires the repression of *KNOX* I activity in the progeny of stem cells destined to generate lateral organs. A complex genetic network is responsible for leaf lamina outgrowth and specification of adaxial–abaxial polarity. The most important members of this network are genes encoding MYB homologs (*ASYMMETRIC LEAVES1*, *ROUGH SHEATH2* and *PHANTASTICA*, collectively called *ARP*), the adaxial determinants, and the *YABBY* genes, which were first proposed to represent abaxial determinants, are now thought to play a key role in regulating laminar outgrowth in response to adaxial–abaxial juxtaposition (reviewed by Yamaguchi et al. 2012).

Most studies have been conducted on angiosperms, but the genome sequence of a heterosporous lycophyte, *Selaginella moellendorffii* (Selaginellales), was published in 2011 (Banks et al. 2011) and has since then played an important role in comparative genomics and improved our understanding of land plant evolution. However, to progress in the field and obtain a better understanding of evolution of SAMs in land plants, more data and better taxonomic coverage are needed, not least from the Lycopodiophyta. We therefore

sequenced the shoot tip transcriptome of a homosporous lycophyte, *Huperzia selago* (Lycopodiales).

Both the telome theory of leaf evolution and the sterilization hypothesis of microphyll evolution can be interpreted to mean a common origin of microphylls and megaphylls and are supported by the data of Harrison et al. (2005) and the gymnosperm transcriptome data of Finet et al. (2016). However, the enation hypothesis of microphyll evolution is supported by the data of Floyd and Bowman (2006) on *HD-Zip III* genes in *Selaginella kraussiana*. Similarly, the evolution of the different types of SAMs has not been resolved. Did the duplex SAM of angiosperms evolve from the simplex SAM of the common ancestor of vascular plants, with the monoplex SAMs in Selaginellaceae and monilophytes represent convergent evolution after independent losses? Or do the simplex SAMs of Isoetales/Lycopodiales and gymnosperms, which all contain secondary PD, represent cases of independent convergent evolution from monoplex SAMs, whereas the duplex SAMs of angiosperms evolved later from the simplex SAMs of gymnosperms? In this context, it is interesting that the evolution of angiosperms (with duplex SAMs) involved a large expansion of the *KNOX* gene family (Gao et al. 2015).

In order to see whether the evolution of simplex meristems in Lycopodiales led to a similar expansion, we analyzed the phylogeny of *KNOX* proteins in the *H. selago* transcriptome. As the down-regulation of *KNOX I* genes represents a key event in differentiation of leaf primordia in all plants, we set about to search for transcripts of the two major *KNOX I* repressors, *ARP* and *YABBY*, in the *H. selago* transcriptome.

## Materials and Methods

### Plant Material

All plants of *Huperzia selago* were collected near Kuznechnoe village at the shore cliffs of Ladoga Lake (Leningrad Region, Russian Federation). If necessary, *H. selago* plants were cultivated in a greenhouse at 19 °C and a light regimen of 8 h light/16 h dark with ca. 200  $\mu\text{Mol photons m}^{-2}\text{sec}^{-1}$  in pots using Kuznechnoye forest soil; otherwise, they were shock-frozen in a dry ice/EtOH mixture and stored at  $-80$  °C. A voucher specimen prepared from cultivated material that had been harvested on June 16, 2012 has been deposited in the herbarium of the Swedish Museum of Natural History (S), leg. K. Pawlowski s.n. (S; Reg. No. S17-11258). Based on the study by Case (1943) describing the rhythmic pace of *H. selago* shoot growth and development and on our own observations, plant material for molecular studies was collected on June 11, 2011 (transcriptome), and on May 30, 2016 (RT-PCR), when the plants had completed a zone with sterile leaves and began to form fertile leaves with sporangia. Material for DNA isolations was harvested on July 24, 2016, when the fertile zone had developed further. Plants for shoot tips for in situ hybridization were harvested in September

2016; their uppermost region was sterile, followed by a completed fertile zone. Shoot tips were fixed after these plants had been cultivated in the greenhouse for a week.

For RNA isolation, ca. 4 mm long shoot tips were cut off with a razor; then the leaves were cut off under a stereomicroscope with razor blade, leaving shoot apices with shoot apical meristems and primordia of young leaves and sporangia. The material was divided into several samples of 100 mg fresh weight each, which were ground with mortar and pestle in liquid nitrogen upon addition of 100 mg Polyclar AT (Serva, Mannheim, Germany). The ground samples were transferred into Eppendorf tubes precooled in liquid nitrogen and stored at  $-80$  °C until being used for RNA isolation. For DNA isolation, plant material was shock-frozen in liquid nitrogen and stored at  $-80$  °C. For in situ hybridization, plants with soil were transferred to the laboratory and observed for 1 week to make sure they were still growing; then, shoot tips were fixed.

### RNA Isolation and Transcriptome Sequencing

RNA was isolated using TRIzol reagent according to Chomczynski and Mackey (1995). The absence of contaminating DNA was confirmed by direct PCR for ubiquitin (Heidstra et al. 1994). After poly(A) selection, a strand-specific library was prepared using the TruSeq Stranded mRNA Sample Prep Kit (Illumina RS-122-2103; Illumina, San Diego, CA, USA). A bead based RoboCut size selection was performed using the 430A and 280B buffers, resulting in fragments of ca. 280–430 bp with an average fragment size of ca. 350 bp. The library was sequenced in an Illumina MiSeq v2 flow cell to produce 21.29 million paired end reads of 250 bp (SciLifeLab, Stockholm, Sweden). The raw data were submitted to GenBank (BioProject PRJNA281995).

### Transcriptome Assembly

We ran Trimmomatic v0.32 (Bolger et al. 2014) to trim the RNA-Seq reads, with a quality cut-off of 20. With these trimmed reads, we assembled a de novo transcriptome with Trinity r2013-08-14 (Grabherr et al. 2011; Haas et al. 2013), using default parameters. We discarded all the sequences whose length was  $<300$  nucleotides. We then mapped back the same set of trimmed reads onto the Trinity assembly with bowtie (Langmead et al. 2009), and performed abundance estimation using RSEM (Li and Dewey 2011). For each Trinity component, we selected the most abundant transcript as representative sequence. We ended up with a data set of 63,908 transcript sequences.

### Transcriptome Analysis

We ran a blastx search of our representative transcriptome against all the plant protein sequences in the nr (nonredundant) database using blast 2.2.29+ (Camacho et al. 2009) and kept the 20 best hits for each transcript. We then used

Blast2GO V.2.8.0 (Conesa et al. 2005) to retrieve the corresponding GO IDs associated with these hits. Following the Gene Ontology Consortium guidelines (Gene Ontology Consortium 2001), we removed the GO IDs that should not be used for direct annotations. Five transcripts of *KNOX* genes were found in the transcriptome (GenBank accession KX761181 - KX761185), as well as one transcript of *YABBY* (GenBank accession KX761186).

### Molecular Cloning

Molecular cloning of *YABBY* cDNAs from fresh shoot tips of *H. selago* was performed by PCR amplification. Total RNA was isolated via the method of Chomczynski and Mackey (1995) using TRIzol reagent. First strand cDNA was synthesized using Maxima First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA). Full length *H. selago* *YABBY* cDNA was amplified using the gene-specific primers 5'-GAGCGTACTTGCTACAAATCATGTCATCC-3' and 5'-GGATGAACTTCTGGTGAAGGAGTTACA-3' and 1.25 U of Taq DNA polymerase (Syntol, Moscow, Russia) according to the manufacturer's instructions. The absence of contaminating DNA was confirmed by direct PCR for ubiquitin (Heidstra et al. 1994). PCR conditions were: 40 cycles of 95 °C for 30 s, 63 °C for 30 s, and 72 °C for 1 min. PCR products were cloned in pJET1.2 (Thermo Scientific) and sequenced. The genomic sequence of *YABBY* was amplified using *H. selago* DNA as a template with the same gene-specific primers and 0.4 U of Phusion High-Fidelity DNA Polymerase (Thermo Scientific). DNA was extracted according to Doyle and Doyle (1990). The PCR protocol was 98 °C for 5 min, followed by 35 cycles of 98 °C for 10 s, 72 °C for 2.5 min. PCR products were cloned in pJET1.2 (Thermo Scientific) and sequenced (GenBank accession MF175244).

### In situ RNA–RNA Hybridization and Cytology

Full length *H. selago* *YABBY* cDNA in pJET1.2 was used for both sense and antisense probe synthesis. Plasmids containing the cDNA in different orientations were linearized with *Nco*I (Thermo Scientific) according to the manufacturer's instructions, followed by extraction with phenol/chloroform 1:1 (v/v). 0.5 µg of each linearized plasmid were used for in vitro transcription with T7 RNA polymerase in the presence of DIG-11-UTP according to the manufacturer's instructions (Roche, Mannheim, Germany) in the presence of RNase inhibitor (Syntol). After in vitro transcription, the reaction mix was treated with 1 Unit of DNase (Thermo Scientific) for 15 min at 37 °C, then 2 µl of 200 mM EDTA (pH 8.0) were added, followed by 10 min incubation at 65 °C. Probes were hydrolyzed in 60 mM Na<sub>2</sub>CO<sub>3</sub>/40 mM NaHCO<sub>3</sub>, pH 10.2 for 19 min at 60 °C to obtain fragments of 200–300 nucleotides (Roche). Hydrolysis was stopped using 3 M sodium acetate (pH 6.0), followed by EtOH precipitation. Unincorporated nucleotides

were removed using an RNA probe purification kit (E.Z.N.A., Omega bio-tek, Norcross, GA, USA).

*Huperzia selago* shoot apices were fixed with 4% paraformaldehyde in microtubule-stabilizing buffer (MTSB; 80 mM PIPES pH 6.8, 1 mM MgCl<sub>2</sub>, 5 mM EGTA plus 0.2% Tween-20 and 0.2% Triton X-100) or in 3% paraformaldehyde in 1/3 MTSB (containing 0.2% Tween-20, 0.2% Triton X-100 and additionally 10% DMSO) overnight at 4 °C after vacuum infiltration. Both treatments yielded the same results; the results depicted in this manuscript are from the fixation in 4% PFA. Afterwards, the tissue was dehydrated in a graded ethanol series and embedded in paraffin with 60 °C melting temperature following Karlgren et al. (2009; Arabidopsis embedding version).

In situ hybridizations were performed according to Coen et al. (1990) and Eisel et al. (2008) with some modifications. Results were documented using an Olympus BX51 light microscope (Olympus, Hamburg, Germany) equipped with a Colorview II camera (Olympus).

For the analysis of the leaf structure of *H. selago*, plant material was fixed in FAA (Formalin-Acetic Acid-Alcohol), dehydrated with a graded ethanol series, embedded in paraffin, cross-sectioned with Accu-Cut SRM 200 Rotary Microtome (Sakura), deparaffinized in xylenes, rehydrated and stained with 0.1% Safranin O solution according to standard plant anatomy protocols (Ruzin 1999).

### Plant Evolution: Data, Sequence Alignment, and Phylogenetic Analyses Using Parsimony

Data from the newly produced transcriptome of *Huperzia* was analyzed along with data from GenBank (accessions are given in supplementary table S1; Supplementary Material online). Sequences of three chloroplast markers were utilized, *rbcl*, *atpB* and *rps4*. Alignment was performed using ClustalX (Chenna et al. 2003) with minor subsequent manual editing. The aligned data matrix comprised 105 taxa representing the major clades of land plants, and a total of 3,359 characters (1,344 from *rbcl*, 1,398 from *atpB*, and 613 from *rps4*). Parsimony analyses were performed in Paup\* (Swofford 2002). A starting tree for divergence time analyses (see below) was produced by using the heuristic search option, 100 random sequence additions, TBR branch swapping. Bootstrap analysis was conducted, performing 1,000 bootstrap replicates with ten random sequence additions in each. Trees were rooted using liverworts (*Marchantia*, *Plagiochila*, and *Pellia*) as outgroup, and Euphyllophytes were enforced as monophyletic according to generally accepted views of land plant phylogeny (e.g., Pryer et al. 2001; Qiu et al. 2007; Rydin et al. 2002).

### Estimation of Divergence Times of Clades

Phylogeny and absolute divergence times of clades were estimated using a Bayesian approach as implemented in

software BEAST v. 1.8 (Drummond et al. 2012). The analyses were set up using software BEAUti v. 1.8 (Drummond et al. 2012). A full description of the methodological approach is given in the Supplementary Material Online (supplementary file S1). Two identical analyses were run, each for 100 million generations, with parameters logged every 10,000 generations. In addition a third analysis was run, identical to the other two but with sampling from priors only. Values sampled for different parameters and convergence of runs were examined and assessed in Tracer v. 1.5 (Rambaut and Drummond 2007). Results were summarized using LogCombiner v. 1.8 and Tree Annotator v. 1.8 (Drummond et al. 2012). Graphic output was produced in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>; last accessed June 8, 2014).

### Gene Trees

The *H. selago* transcriptome data set was cleaned and translated with Transdecoder (Haas et al. 2013). It was then merged with the predicted proteomes of *Amborella trichopoda* (AmbTri), *Arabidopsis thaliana* (AraTha; Arabidopsis Genome Initiative 2000), *Chlamydomonas reinhardtii* (ChlRei; Merchant et al. 2007), *Oryza sativa* (OrySat; Goff et al. 2002), *Ostreococcus lucimarinus* (OstLuc; Palenik et al. 2007), *Physcomitrella patens* (PhyPat; Rensing et al. 2008), *Populus trichocarpa* (PopTri; Tuskan et al. 2006), *Selaginella moellendorffii* (SelMoe; Banks et al. 2011), *Vitis vinifera* (VitVin; Jaillon et al. 2007) and *Zea mays* (ZeaMay; Schnable et al. 2009), downloaded from Phytozome v. 11 (Goodstein et al. 2012); and the genomes of *Pinus taeda* (PinTae) (v. 1.0, downloaded from [http://dendrome.ucdavis.edu/ftp/Genome\\_Data/genome/pinerefseq/Pita/v1.0/](http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pinerefseq/Pita/v1.0/); last accessed August 18, 2017) and *Picea abies* (PicAbi) (v. 1.0, downloaded from <http://congenie.org/start>; last accessed August 18, 2017) (Nystedt et al. 2013). We used blastp (v. 2.2.28; Camacho et al. 2009) to find similar sequences; AT1G62360.1 (Arabidopsis STM) was used as the query for KNOX and AT1G08465.1 (Arabidopsis YAB2) as the query for YABBY. To achieve better coverage of the Lycopodiophyta, transcriptome data sets for species that contained both *class I* and *class II* KNOX transcripts were obtained from the OneKP database, namely, *Lycopodium deuterodensum*, *Isoetes tegetiformans*, *Selaginella apoda*, and *Selaginella kraussiana* (Matasci et al. 2014; Wickett et al. 2014).

Protein domain analysis was performed with Pfam-A (Finn et al. 2016) and hmmer (v. 3.1b2) (Mistry et al. 2013). All YABBY orthologs contained the characteristic YABBY domain whereas the most strict requirement for KNOX protein demanded that they are full-length homologs—that they have two KNOX domains, a homeobox and ELK domain (supplementary fig. S1, Supplementary Material online). Another, less strict set of requirements allowed for any sequence with at least one of the two KNOX domains or a homeobox domain (supplementary fig. S1, Supplementary Material online).

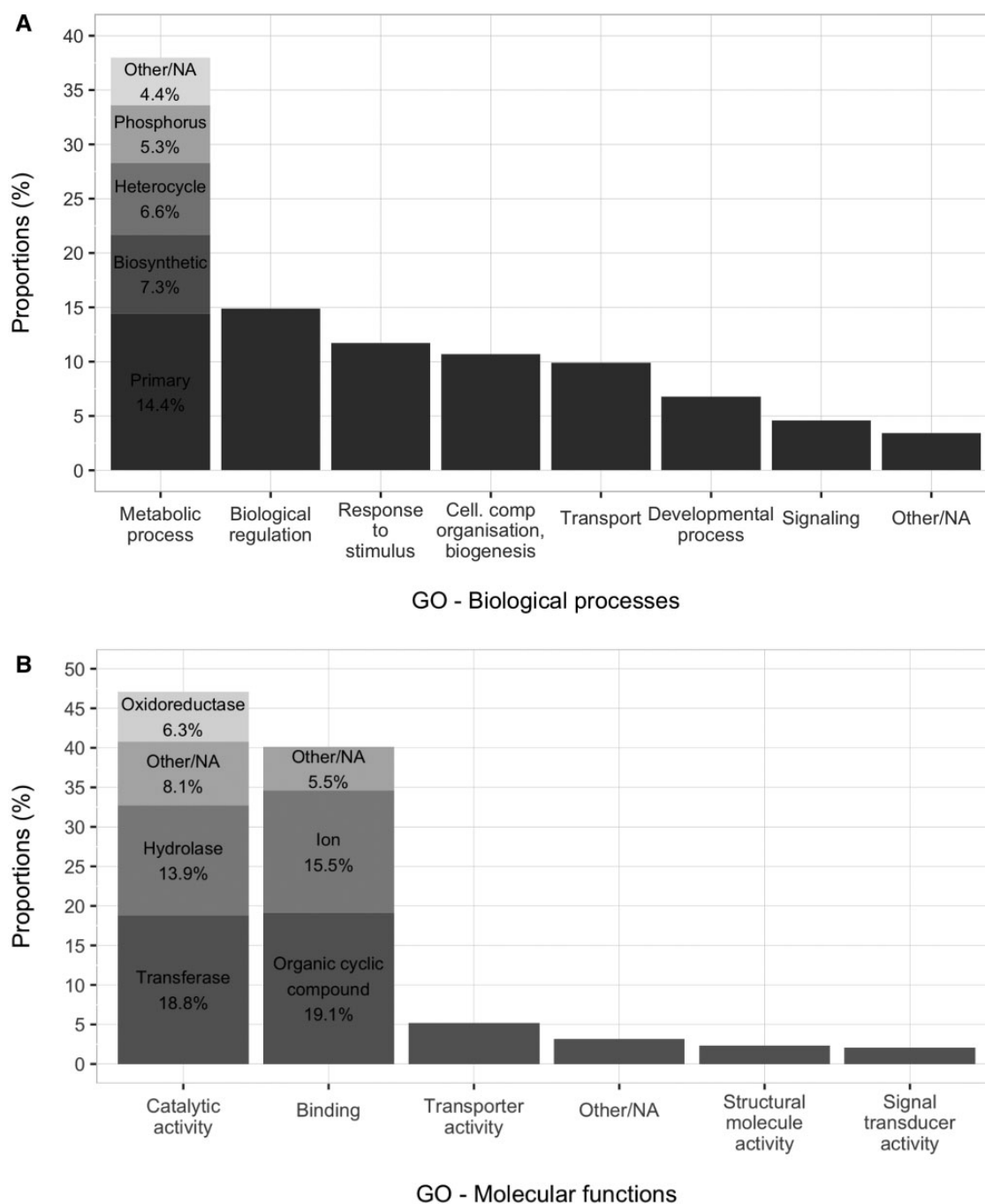
Protein sequences were aligned with mafft (v7.123b; parameter genafpair, maximum number of cycles of iterative refinement = 1000; Katoh and Standley 2013), columns with >20% gaps were removed with trimAl (version 1.4.rev15; Capella-Gutiérrez et al. 2009) and trees were inferred with RAxML (v. 8.2.4, parameters: rapid bootstrap analysis and search for the best-scoring ML tree in one program run (-f a), 100 bootstraps, protein model JTTIG) (Stamatakis 2014). Protein model was determined using ProtTest 3 (Darriba et al. 2011). Trees were visualized with iTol v3 (Letunic and Bork 2016).

## Results

### The *H. selago* Shoot Tip Transcriptome

RNA was isolated from shoot tips of field samples of *H. selago*. After reduction of one isoform per gene, and removal of contigs <300 bp, the assembly contained 78,760 contigs. When the sequences with an RPKM value of 0 were removed, 63,908 contigs were left. For annotation, blastx searches were run in May 2014. From 63,908 sequences, only 22,212 yielded a hit against NCBI NR (plants) and/or Swissprot. For those contigs that got a hit in the Gene Ontology database with Blast2GO, the associated GO terms for biological processes and molecular functions are shown in figure 2. TransDecoder (v 2.0.1; Haas et al. 2013) was run on the 41,696 transcripts that did not get a blastx hit against NCBI NR plant. Among them, 7,681 sequences have a coding region of at least 75 aa long.

However, only ca. 35% of the assembled cDNAs showed similarity with database sequences, or in other words, the transcriptome contained ca. 65% of orphans. Such a high percentage of orphans is unusual, but not unheard of; a functional transcriptome of the lone star tick had been shown to contain 71% of orphans, a fact ascribed to taxonomic isolation (Gibson et al. 2013). Still, this was a very high number of orphans. Since a field sample had been used for isolating RNA for sequencing, blobtools (version 0.9.19; Kumar et al. 2013) was used to examine the transcriptome for potential contaminants/endophytes. Blobtools was run based on four blast (Version 2.4.0+) searches: a blastn megablast on the NCBI nucleotide database, a blastx on the SwissProt database, a blastn megablast on the SILVA database (RNAcentral SILVA collection) and a blastx search against NR (NCBI). The reads were remapped on the final transcriptome (63,908 contigs). The read\_cov plot (supplementary fig. S2; Supplementary Material Online) shows that 19% of the reads mapped to transcripts annotated as “no hits”, whereas 75% of the reads mapped to transcripts annotated as belonging to the Streptophyta clade. Almost no reads mapped to known sequences belonging to other organisms from various clades (bacteria, chordata, ascomycota). The percentage of “no hits” reads is not surprising given the high number of



**Fig. 2.**—Gene ontology (GO) mapping results for *H. selago* contigs. (A) Biological processes, (B) Molecular functions. The chosen GO categories are at the same “high level” in the GO hierarchy. The best represented categories, especially for molecular functions, are similar to those found in a whole plant transcriptome of *Selaginella moellendorffii* (Weng et al. 2005).

transcripts annotated as orphans. These orphans likely represent lineage-specific genes that are currently absent from the databases used for the searches, rather than unknown contaminants. Thus, a problem of contamination by known organisms can be excluded.

To ascertain the quality of the assembly, the 63,908 contigs were compared with the BUSCO plant set (Simão et al. 2015,

at the time only available on request). This set contains single-copy genes from the Embryophyta clade, based on the OrthoDB database of orthologs (<http://www.orthodb.org/>; last accessed October 31, 2016). Of 956 BUSCO groups, 780 were found complete in the transcriptome (590 groups were found once and 190 more than once). Altogether, 81.5% of the BUSCO groups were complete, 5.3% were

fragmented and 13.2% were missing. This compares well with other transcriptomes analyzed using BUSCO (spinach, Xu et al. 2015; *Pinus patula*, Visser et al. 2015; several sea grasses, Lee et al. 2016).

We next examined the phylogenetic position of *H. selago* to determine whether it could explain the high percentage of orphans.

### Phylogenetic Analysis of the Position of *H. selago* Relative to *Selaginella* sp. and to Angiosperms

In order to provide a context of the high amount of orphans in the transcriptome, the evolutionary position of *H. selago* was examined phylogenetically using Bayesian inference. The topological results (fig. 3) are generally well supported and consistent with accepted views of land plant phylogeny. Effective age prior distributions (as assessed by a prior only analysis) were consistent with those specified (table 1) for all nodes. *Huperzia* diverged from its closest relative *Lycopodium* about 93 Ma, from the heterosporous lycopods (*Isoetes* and *Selaginella*) 376 Ma, and the most recent common ancestor of *Huperzia* and the angiosperm *Arabidopsis* dates to 419 Ma. All node ages and confidence intervals are given in table 1. In summary, the divergence of Selaginellales and Lycopodiales within the lycopods took place in the Late Devonian before the evolution of megaphylls, and *Huperzia* and *Lycopodium* diverged during the Late Cretaceous, during the early radiation of angiosperms (Beerling et al. 2001; Pires and Dolan 2012).

### Phylogeny of a Family of Noncell-Autonomous Transcription Factors: KNOX

To trace the evolution of the KNOX protein family in *H. selago* we inferred a KNOX gene tree with orthologs from a wide range of plant species with sequenced genomes (see Material and methods for an overview). We supplemented these with transcriptome sequences from the OneKP project to more precisely determine the time of duplication events. The gene tree shows a clear distinction between *KNOX I* and *KNOX II* genes (fig. 4). In both groups, Lycopodiophyta sequences represent sister groups to the seed plant sequences. No class II KNOX protein was found in the *Ostreococcus lucimarinus* genome (Palenik et al. 2007) which is consistent with the conclusion of Gao et al. (2015) that algal KNOX proteins have higher similarity with KNOX I.

The KNOX tree was rooted based on the duplication event that took place before the last common ancestor of the chlorophytes and streptophytes. The transcriptome of *H. selago* contains five *KNOX I* and *KNOX II* genes. Three *KNOX* gene duplications took place in the ancestor of Lycopodiophyta, one in class I and two in class II (fig. 4; supplementary fig. S1, Supplementary Material online). The duplication in class I and one of the duplications in class II took place within the

Lycopodiales, represented in the tree by *H. selago* and *Lycopodium deuterodensum*.

### The *H. selago* Transcriptome Does Not Contain an ARP Homolog

No ARP homologs were found in the *H. selago* transcriptome using the *Arabidopsis* protein sequence (GenBank accession O80931.1) and tBlastN. To confirm the lack of ARP homologs in the *H. selago* transcriptome, we mapped the *Huperzia* reads on the ARP mRNA sequence from *S. kraussiana* (GenBank accession number AAW62520.1). Two different mapping programs, bwa-mem (Li and Durbin 2009) and Stampy (Lunter and Goodson 2011), were used, but despite the high conservation of the first 110 amino acids even between *S. kraussiana* and *Arabidopsis* (for comparison on the protein and RNA level, see supplementary fig. S1; Supplementary Material online) no reads mapped to the ARP sequence (data not shown). Hence, no ARP gene seems to be expressed in shoot tips of *H. selago*.

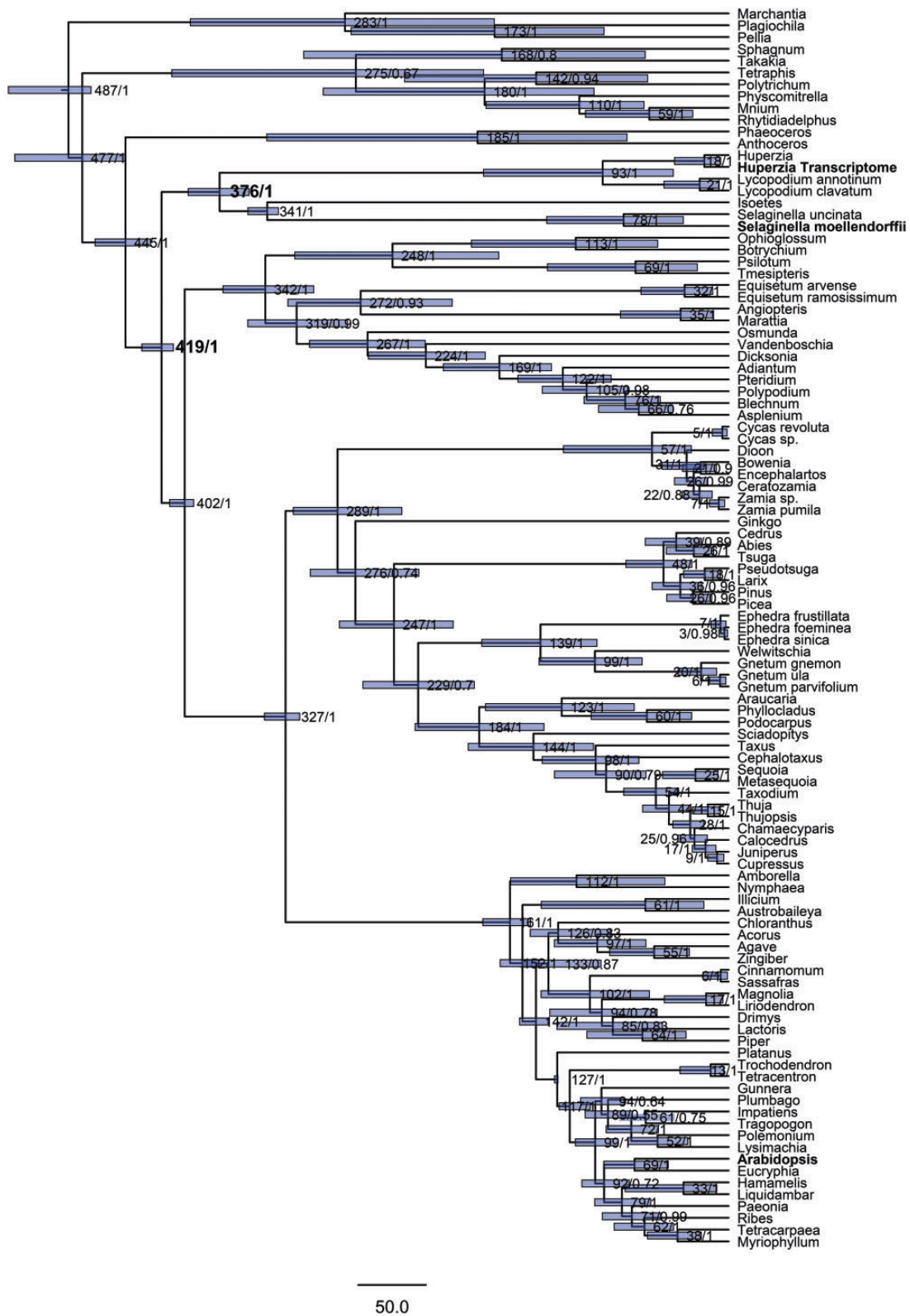
### The *H. selago* Shoot Tip Transcriptome Contains a YABBY Homolog

YABBY transcription factors are thought to be specific to seed plants with one or two YABBY genes present in the last common ancestor of extant seed plants (Finet et al. 2016), although a precursor has been found in a marine alga, *Micromonas* (Worden et al. 2009). Surprisingly, we were able to identify a YABBY domain protein in the *H. selago* transcriptome. It has a clear similarity to the YABBY domain as found in *Arabidopsis* YABBY proteins (fig. 5). The genomic sequence corresponding to the cDNA was amplified and sequenced (GenBank accession MF175244); the intron–exon structure of *HsYABBY* is similar to that of *Arabidopsis* YABBY genes (Siegfried et al. 1999; Supplementary fig. S1 and table S1; Supplementary Material online). Phylogenetic analysis showed that *HsYABBY* is a sister to YABBY proteins from ephyllphytes (supplementary fig. S1 and table S1; Supplementary Material online). These results suggest that the last common ancestor of Lycopodiophyta and Ephyllphyta thus already contained a YABBY-domain protein that was likely lost in Sellaginellaceae.

### *HsYABBY* Is Expressed in the Surface Cell Layer of the SAM and the Leaf Primordia

In situ RNA–RNA hybridization experiments showed that *HsYABBY* is expressed in the SAM with the strongest signal in the surface layer where the apical initials (AI) of the simplex SAM type are located, and weaker signal in their subsurface derivatives. However, much weaker or no expression was detected in the AI themselves (fig. 6A and C). In early stages of leaf development, *HsYABBY* expression could be detected in all cells of the incipient leaf primordia and of sporangia





**FIG. 3.**—Dated phylogeny inferred from analysis of three chloroplast genes (*atpB*, *rbcl*, and *rps4*), using a Bayesian framework and a relaxed molecular clock. Values represent node ages (median heights), followed by posterior probabilities of clades. Blue bars illustrate the node height 95% HPD intervals. The most recent common ancestor of *Huperzia* and *Selaginella* is estimated to have existed about 376 Ma, and that of *Huperzia* and *Arabidopsis* about 419 Ma. Accession numbers of all sequences used in the phylogeny are given in supplementary table S1, Supplementary Material online.

**Table 1**

Age Priors, and Posterior Distributions and Probabilities for Selected Clades

Nodes	Prior Source (distribution)	Monophyly Enforced	Prior Ages: Median (95% CI)	Posterior Ages: Median (95% HPD)	Posterior Prob. of Clades	Prior Setting References (see text for details)
Land plants (root)	Fossil <sup>a</sup> (lognormal)	—	480.2 (472.8–497.1)	487 (471–532)	1	Taylor et al. (2009), Kenrick (2003)
Land plants except Marchantiophyta	Tree prior only	yes	—	477 (446–527)	1	—
Anthoceroophyta and Tracheophyta	Tree prior only	—	—	445 (427–468)	0.94	—
Tracheophyta	Fossil <sup>a</sup> (lognormal)	yes	423.1 (412.4–459.8)	419 (410–433)	1	Kidston and Lang (1920), Hao (1988), Kenrick and Crane (1997)
Lycopodiophyta	Tree prior only	—	—	376 (354–399)	1	—
Lycopodiaceae	Tree prior only	—	—	93 (41–181)	1	—
Selaginellaceae and Isoetaceae	Fossil <sup>a</sup> (lognormal)	yes	346.1 (335.4–382.8)	341 (332–355)	1	Bateman and DiMichele (1994), Korall et al. (1999), Rowe (1988)
Euphyllophyta	Fossil <sup>a</sup> (lognormal)	yes	401.5 (396.2–419.9)	402 (395–413)	1	Granoff et al. (1976), Kenrick and Crane (1997), Hoffman and Tomescu (2013)
Monilophyta	Tree prior only	—	—	342 (306–373)	1	—
Spermatophyta	Fossil <sup>a</sup> (lognormal)	yes	329.6 (319.8–352.2)	327 (317–343)	1	Hilton and Bateman (2006), Doyle (2008), Taylor et al. (2009)
Angiospermae	Fossil <sup>a</sup> (lognormal)	yes	146.3 (138.3–173.9)	161 (146–181)	1	Friis et al. (2011)
Eudicots	Fossil <sup>a</sup> (exponential)	yes	126.7 (126.1–129)	127 (126–129)	1	Friis et al. (2011)
Cycadophyta	Tree prior only	—	—	57 (28–121)	1	—
Pinaceae	Tree prior only	—	—	48 (27–78)	1	—
Gnetales	Tree prior only	—	—	139 (97–182)	1	—
Cupressophyta	Tree prior only	—	—	184 (136–231)	1	—

<sup>a</sup>Absolute ages of fossil strata are inferred based on stratigraphic interpretations in The Geological Time Scale (Gradstein et al. 2012).

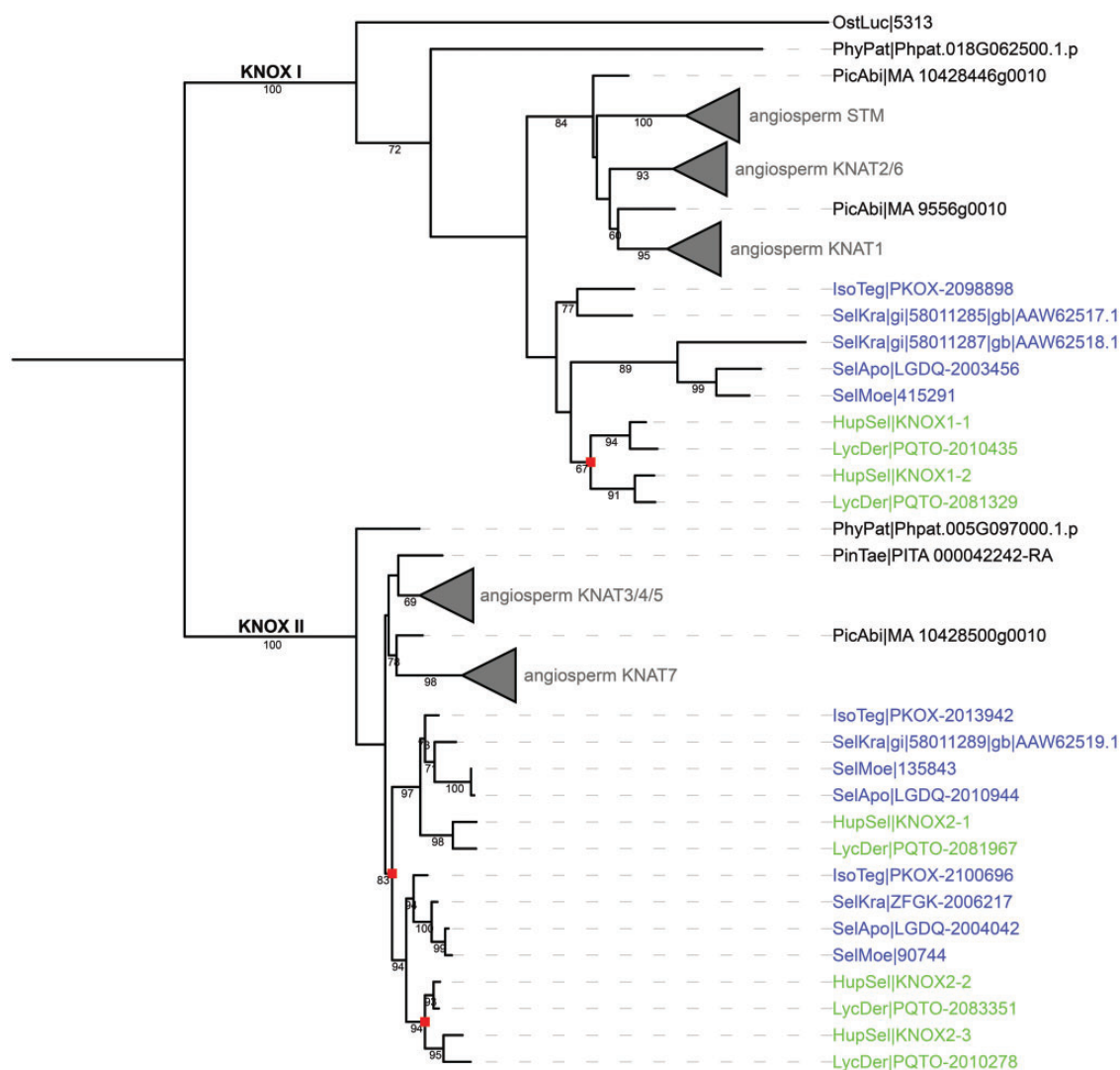
primordia, whereas in older leaf primordia its expression is confined to the surface cells of both the abaxial and adaxial sides (fig. 6A, C, and D). *HsYABBY* expression could also be detected in the cauline procambium as well as in the leaf trace procambium (fig. 6A).

## Discussion

The sequencing of the shoot tip transcriptome of *Huperzia selago* revealed in the first place an unusually high amount of orphan sequences, which could be explained by taxonomic isolation. Our phylogenetic studies revealed that the two branching points in the evolution of Lycopodiophyta could be dated to 376 (Lycopodiales vs. Selaginellales/Isoetales) and 341 Ma (Selaginellales vs. Isoetales), respectively, that is, they preceded the separation between gymnosperms and angiosperms. Therefore, considerable diversity can be expected between Lycopodiophyta from different orders. This assumption is supported by fossil discoveries that document a substantial historical lycophyte diversity, including groups that had gone extinct already by the end of the

Paleozoic (e.g., Kenrick and Crane 1997). Lycophytes were a common feature of Earth's vegetation during the late Paleozoic, the large tree-lycophods of the Isoetales being one of the most well-known examples (see, e.g., Larsén and Rydin 2016 for a brief summary of isoetalean diversity). Thus, the treatment of *Selaginella* sp. as general model species of the Lycopodiophyta is likely to lead to erroneous conclusions.

Lycopodiales and Isoetales contain simplex SAMs whereas Selaginellales contain monoplex SAMs. Simplex SAMs require the formation of secondary plasmodesmata (PD) whereas monoplex SAMs require only primary PD. One aim of this study was to find out whether the family of class I *KNOX* genes encoding noncell autonomous transcriptional regulators expanded during the evolution of simplex SAMs. A considerable expansion of *KNOX* genes took place at the basis of angiosperm evolution (Banks et al. 2011; Gao et al. 2015). In angiosperms, at least two different subgroups of class I *KNOX* proteins exist, STM (only in eudicots), KNAT1 and KNAT2/6 (Gao et al. 2015; fig. 4; supplementary fig. S1, Supplementary Material online). The *H. selago* shoot tip transcriptome described in this study contained five *KNOX* genes, whereas

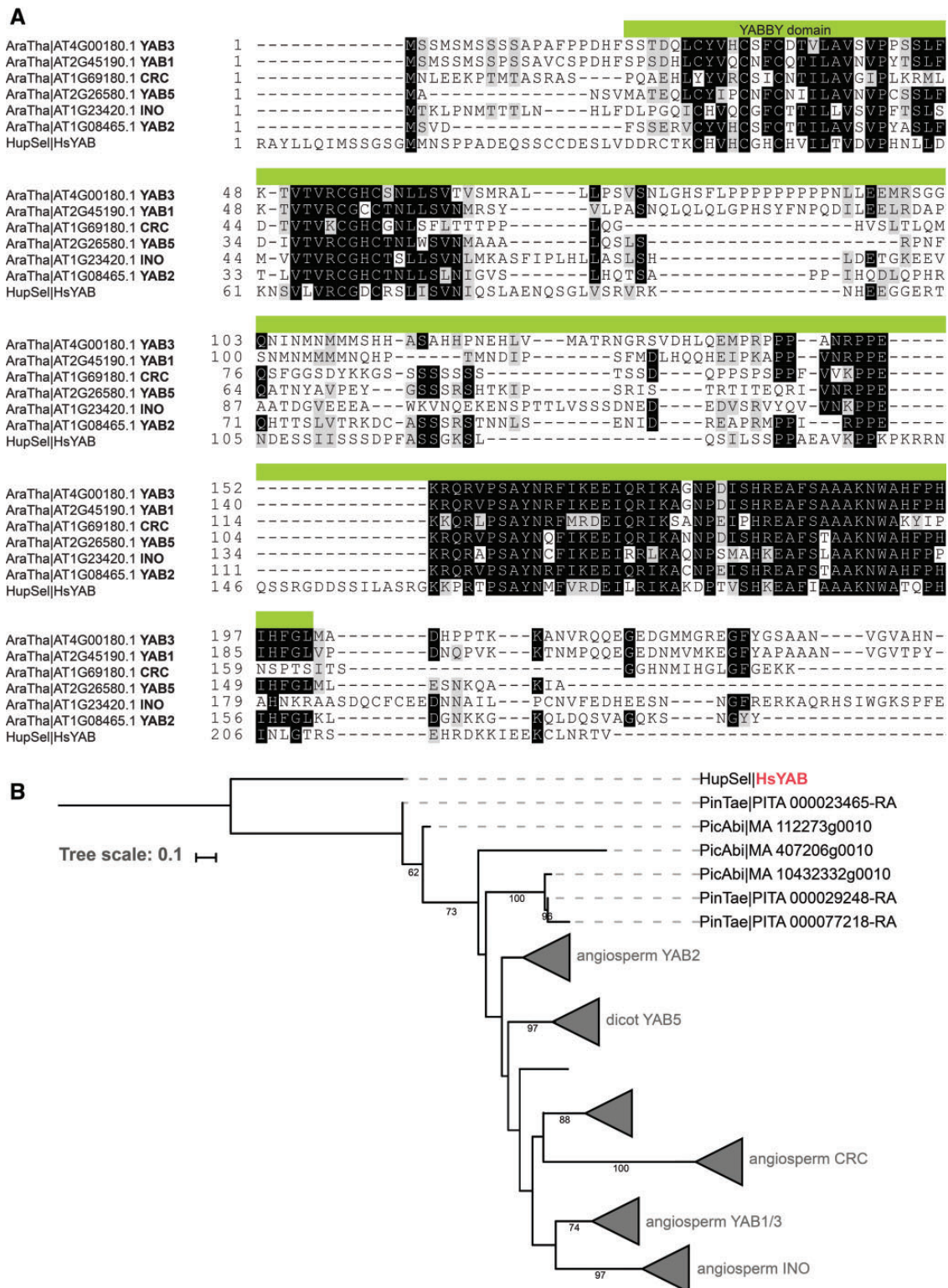


**Fig. 4.**—Gene tree of KNOX proteins. The gene tree of KNOX proteins shows a clear separation of sequences into KNOX I (top branch) and KNOX II (bottom branch) classes, with Lycopodiophyta sequences in both classes. Lycopodiophyta sequences are highlighted; species from class Isoetopsida are shown in blue (*Selaginella* sp. and *Isoetes tegetiformans*) whereas sequences from class Lycopodiopsida (*Huperzia selago* and *Lycopodium deuterodensum*) are shown in green. Duplication nodes for Lycopodiophyta are marked with red squares. Bootstrap support is only shown for values >60. Angiosperm sequences are collapsed for clarity. Complete trees are shown in supplementary figure S1, Supplementary Material online. All sequences used are listed in supplementary table S1, Supplementary Material online.

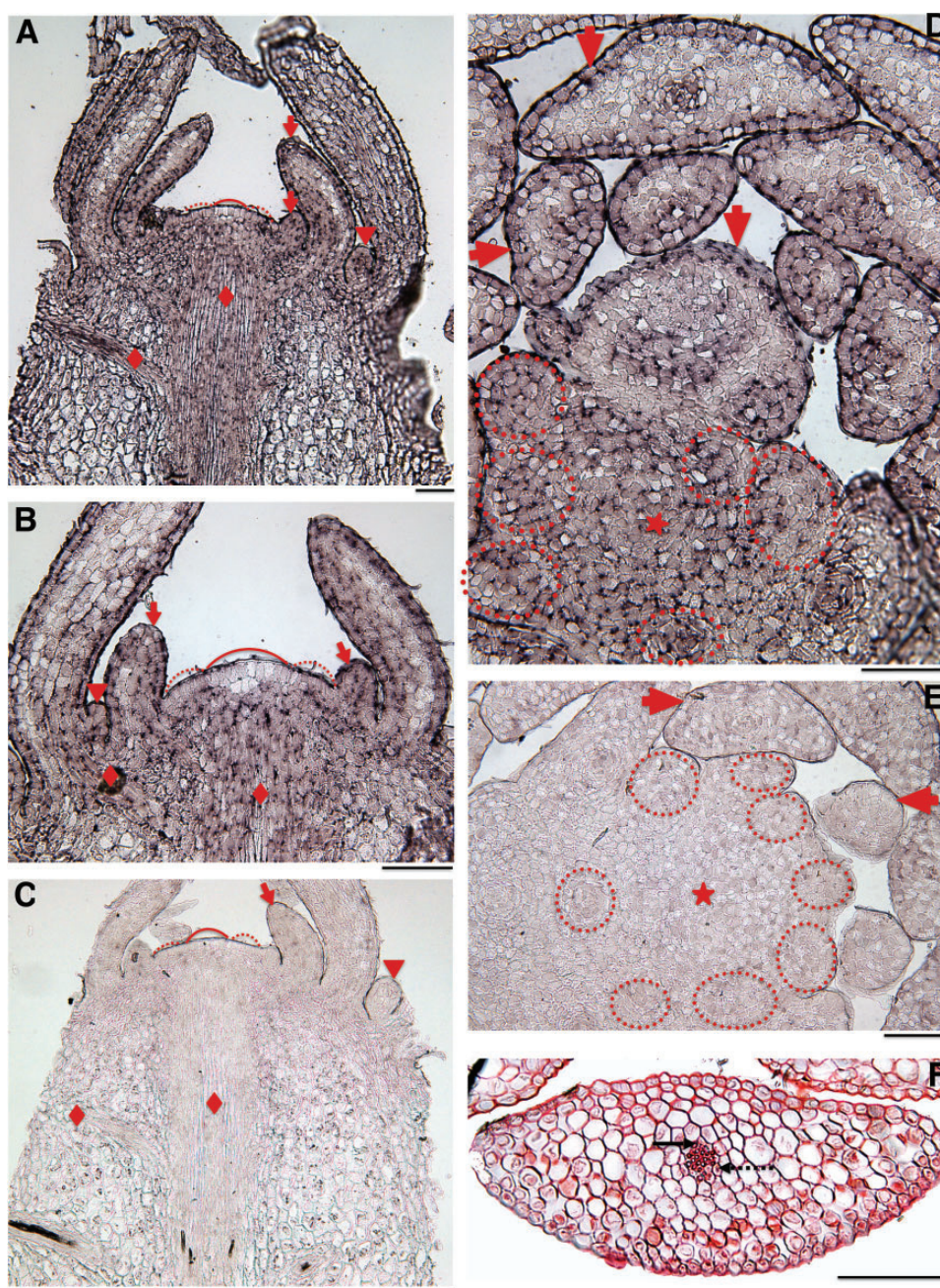
four KNOX genes (two class I and two class II genes) had been found in the *Selaginella moellendorffii* genome (Gao et al. 2015). Our analysis showed a gene duplication of class I KNOXes in the genomes of Lycopodiales (fig. 4). The situation for Isoetales is not clear—either there was no gene duplication, or not all class I KNOX gene sequences of *Isoetes* sp. are available, which might be due to the fact that no shoot tip transcriptome was analyzed. No clear conclusion can be drawn with regard to the ability of lycophyte KNOX proteins to traffick through PD (supplementary fig. S1; Supplementary Material online).

Two groups of transcription factors can repress the expression of class I KNOX genes. The group of MYB orthologues

collectively called ARP functions in the regulation of the proximal-distal leaf length by epigenetically repressing the expression of class I KNOX genes (Machida et al. 2015). The fact that no ARP transcripts could be found in the *H. selago* shoot tip transcriptome, might be taken to suggest that ARP, while present in Selaginellales, is missing in Lycopodiales. Conifer homologs of ARP can be found in the shotgun transcriptomes in GenBank using tBlastN (supplementary fig. S1; Supplementary Material online). Thus, although no ARP protein is annotated in GenBank for gymnosperms—which, like Lycopodiales, form simplex meristems—it should be pointed out that the absence of ARP genes, or ARP expression, cannot be a feature of plants with simplex meristems.



**Fig. 5.**—Identification of *H. selago* YABBY. (A) Alignment of *H. selago* YABBY with angiosperm YABBY proteins. Alignment of *H. selago* YABBY with Arabidopsis YABBY orthologs. The YABBY domain is marked in green. (B) A schematic tree of YABBY proteins. The full tree is available in supplementary figure S1, Supplementary Material online; all sequences used are listed in supplementary table S1, Supplementary Material online. The angiosperm orthologous groups are collapsed and named after the Arabidopsis gene. *H. selago* YABBY, belonging to the earliest branching species in the tree, is the outgroup to all YABBY orthologous groups.



**Fig. 6.**—*YABBY* expression in shoot tips *H. selago* analyzed by in situ hybridization. (A, B) Median longitudinal sections through the SAM were hybridized with antisense (A, B) and sense RNA (C). The structure of the *H. selago* SAM is outlined in figure 1A. *HsYABBY* mRNA can be detected in the SAM (peripheral parts indicated by curved dashed lines, central part indicated by curved line), with the strongest signal in the outermost cell layer and in the peripheral zone, incipient and young leaf primordia (LPs; arrows), and sporangia primordia (arrowheads) and in the axial and leaf trace procambium (diamonds). Considerably weaker signal is seen in the central cells of the outer layer, the apical initials (AIs, indicated by the curved line) and the subapical cells which are part of the meristem. (D, E) Transverse section of the shoot tip hybridized with antisense (D) and sense (E) RNA of *HsYABBY*. *YABBY* expression is uniform in the youngest LPs, whereas in older LPs it is confined to both the abaxial and the adaxial epidermis and to the procambium of the leaf trace. The meristem is marked by an asterisk, incipient LPs are outlined by dashed lines; older LPs are marked with arrows. (F) A transversal section through a *H. selago* leaf stained with Safranin shows its uniform mesophyll cells and amphicribal vascular bundle. An arrow points to the xylem, a dashed arrow to the phloem. Bars denote 100  $\mu\text{m}$ .

Transcription factors of the YABBY family have been shown to be involved in the initiation of the outgrowth of the lamina, the maintenance of polarity, and the establishment of the leaf margin. Like ARP proteins, YABBY can repress the expression of *KNOX1* genes (Kumaran et al. 2002). So far, YABBY proteins have been believed to be restricted to Spermatophyta (Finet et al. 2016). In gymnosperms, YABBY proteins are grouped in four clades, whereas five clades are present in angiosperms. Based on phylogeny and expression patterns, Finet et al. (2016) concluded that either one or two YABBY genes were present in the last common ancestor of extant seed plants where they presumably were involved in the determination of polarity connected to the evolution of laminar leaves. Yet, we identified a YABBY homologue in the *H. selago* shoot transcriptome. Its expression pattern in shoot tips showed major differences with those of angiosperm YABBY genes (see e.g., Siegfried et al. 1999) in that *HsYABBY* transcripts are localized in the SAM with the strongest signal in the surface layer where the AIs of the simplex SAM are located, although not in the AIs themselves (fig. 6), whereas angiosperm YABBY transcripts are never present in the SAM. However, the expression pattern of *HsYABBY* showed striking similarity with that of *ARP* from *Selaginella kraussiana* (fig. 1J in Harrison et al. 2005), the expression of which in the SAM had been explained by its ancestral function being the control of shoot dichotomy. As *Huperzia* is characterized by dichotomous branching just like *Selaginella* (Gola and Jernstedt 2011), *HsYABBY* might have the same function here.

In incipient leaf primordia, *HsYABBY* expression was detected in all cells, whereas later its expression was confined to the surface layers on both the abaxial and the adaxial side. Again, this is different from the distinct abaxial expression of YABBY in angiosperms, but corresponds perfectly to the expression pattern of *ARP* in *S. kraussiana* (Harrison et al. 2005). A further similarity between the expression pattern of *SkARP* and *HsYABBY* is the presence of both mRNAs in the leaf trace procambium. *HsYABBY* is transcribed also in the cauline procambium which was not shown for *SkARP*. The expression of both *SkARP* and *HsYABBY* in procambium cells, where angiosperm *ARP* or YABBY genes are not expressed, might be linked to the different mechanisms employed for vascularization in lycophytes versus angiosperms (Floyd and Bowman 2006).

Mature microphylls of both *Selaginella* and *Huperzia* lack internal polarization into adaxial palisade and abaxial spongy mesophyll; they are composed of uniform spongy cells (Chu 1974; Harrison et al. 2007; fig. 6E). For *H. selago* microphylls, there is no difference between the upper and lower epidermis regarding stomata count (Chu 1974). Moreover, vascular bundles of both species are concentric and amphicribal (Smith 1938; fig. 6E). This is also the case for the symmetric “abaxialized” leaves of an *Antirrhinum majus* *ARP* (*PHAN*) loss-of-function mutant (Waites and Hudson 1995). Similarly, a gain-of-function mutant of Arabidopsis that constitutively represses the expression of a member of the YABBY

family (*fil*; McConnell and Barton 1998) has symmetric “adaxialized” leaves with a concentric vascular bundle (reviewed by Bowman et al. 2002). In summary, the resemblance between *Selaginella* sp. and *Huperzia* sp. microphylls on the one hand, and angiosperm leaves lacking dorsoventral polarization through combined action of ARP and YABBY on the other hand, suggests that either one of the two suppressors of *KNOX* expression is sufficient for the switch between determinacy and indeterminacy that results in microphyll development.

In summary, either YABBY evolved convergently in both Lycopodiales and seed plants, or the common ancestor of both must have contained both *ARP* and YABBY genes, both of which encode transcription factors able to suppress the expression of *class I KNOX* genes. Given the absence of YABBY in *Selaginella* and the apparent absence of *ARP* in *Huperzia*, and the strong similarity of their expression patterns in shoot tips, we propose that in Lycopodiales, the *ARP* gene or its expression in shoot apices was lost, while in Selaginellales, the YABBY gene was lost. And if YABBY genes evolved in the common ancestor of vascular plants and are expressed in the primordia of both leaves and sporangia of *Huperzia*, one possible scenario is that like *KNOX1* and *ARP* genes they were part of a sporangium-specific developmental program that was recruited for leaf development in lycophytes, monilophytes and seed plants (Worsdell 1905; Vasco et al. 2016).

Based on the fact that the SAM-like structure in *Chara* resembles a monoplex SAM (Graham et al. 2000) and that most seedless land plants analyzed have monoplex SAMs, it has been traditionally assumed that the monoplex SAM is plesiomorphic for land plants (Gifford and Foster 1989; Kenrick and Crane 1997). Kidston and Lang (1920) had observed several initials in the SAMs of Rhyniophytes, but due to poor tissue preservation this was not considered reliable proof of simplex SAM ancestry. However, stronger support of the plesiomorphy of SAMs with several, not single, apical initials came from research of Cooke et al. (1996) and Imaichi and Hiratsuka (2007) on the symplastic structure of SAMs in an evolutionary context. As mentioned earlier, monoplex SAMs (SAMs with a single AI) are characterized by containing exclusively primary PDs, whereas SAMs with several AIs have both primary and secondary PDs. Both groups of authors interpreted this as support for a scenario where the loss of the ability to form secondary PDs led to the development of monoplex SAMs from simplex SAMs. This would have happened independently in microphyllous *Selaginella* and most euphyllous monilophytes. Furthermore, recent molecular evidence (Frank et al. 2015) supports the idea that the simplex SAM might indeed be ancestral for the sporophytes of terrestrial plants and later was lost in the lycophyte *Selaginella* and in most monilophytes. The existence of YABBY homologues in Spermatophyta as well as in the shoot tips of *H. selago*, and therefore probably in the common ancestor of vascular plants,

also suggests that it might be more parsimonious to assume that simplex SAMs, and thus also secondary PDs, were present in the common ancestor of vascular plants.

## Conclusions

The common ancestor of lycophytes and euphyllophytes is likely to have contained two types of transcription factors that could repress the transcription of class I *KNOX* genes, *ARP* as well as *YABBY*. This supports the new paradigm of leaf evolution proposed by Vasco et al. (2016) that supposes a deep homology of all types of leaves. In due course, *YABBY* was lost in Selaginellales, whereas our data suggest that *ARP* may have been lost in Lycopodiales.

The presence of *YABBY* in the common ancestor of vascular plants would also support the hypothesis that this common ancestor had a simplex SAM and thus also had to have evolved secondary PD. If this was the case, secondary PD in Lycopodiales and seed plants should be homologous, and lycophyte *KNOX* proteins should be able to traffick through secondary PD.

The fact that analyses of the *H. selago* transcriptome could change our views on different hypotheses about the ancestral SAM type in land plants shows that more genomic/transcriptomic data from plants from different ancestral lineages have to be studied.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

This study was supported by the institutional research projects no. 01201255613 and no. 0126-2016-0001 of the Komarov Botanical Institute RAS and by the Russian Foundation for Basic Research (grants #14-04-01397 and #17-04-00837 to M.A.R.). The Core Facility Center “Cell and Molecular Technologies in Plant Science” at the Komarov Botanical Institute RAS and the Research Resource Centre for Molecular and Cell Technologies of Saint-Petersburg State University are gratefully acknowledged for technical support. Bioinformatics support by BILS (Bioinformatics Services for Swedish Life Science) is gratefully acknowledged. We would also like to thank the two anonymous reviewers who improved the manuscript with their constructive criticism.

## Literature Cited

Ambrose B, Vasco A. 2016. Bringing the multicellular fern meristem into focus. *New Phytol.* 210(3):790–793.  
 Arabidopsis Genome Initiative 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815.

Banks JA, et al. 2011. *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332(6032):960–963.  
 Bateman RM, DiMichele WA. 1994. Heterospory: the most iterative key innovation in the evolutionary history of the plant kingdom. *Biol Rev.* 69:345–417.  
 Beerling DJ, Osborne CP, Chaloner WG. 2001. Evolution of leaf-form in land plants linked to atmospheric CO<sub>2</sub> decline in the Late Palaeozoic era. *Nature* 410(6826):352–354.  
 Bierhorst DW. 1977. On the stem apex, leaf initiation and early leaf ontogeny in filicalean ferns. *Am J Bot.* 64(2):125–152.  
 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30(15):2114–2120.  
 Bower FO. 1908. The origin of a land flora: a theory based upon the facts of alternation. London, UK: Macmillan.  
 Bowman JL, Eshed Y, Baum SF. 2002. Establishment of polarity in angiosperm lateral organs. *Trends Genet.* 18(3):134–141.  
 Buvat R. 1989. Ontogeny, cell differentiation, and structure of vascular plants. Berlin, Germany: Springer-Verlag.  
 Byrne M. 2012. Making leaves. *Curr Op Plant Biol.* 15(1):24–30.  
 Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.  
 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.  
 Carles CC, Fletcher JC. 2003. Shoot apical meristem maintenance: the art of a dynamic balance. *Trends Plant Sci.* 8(8):394–401.  
 Chen H, Ahmad M, Rim Y, Lucas WJ, Kim JY. 2013. Evolutionary and molecular analysis of Dof transcription factors identified a conserved motif for intercellular protein trafficking. *New Phytol.* 198(4):1250–1260.  
 Chen H, Jackson D, Kim JY. 2014. Identification of evolutionarily conserved amino acid residues in homeodomain of KNOX proteins for intercellular trafficking. *Plant Signal Behav.* 9(2):e28355.  
 Chenna R, et al. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31(13):3497–3500.  
 Chomczynski P, Mackey K. 1995. Short technical reports. Modification of the TRI Reagent procedure for isolation of RNA from polysaccharide and proteoglycan rich sources. *Biotechniques* 19(6):942–945.  
 Chu MCY. 1974. A comparative study of the foliar anatomy of *Lycopodium* species. *Am J Bot.* 61(7):681–692.  
 Coen ES, et al. 1990. *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* 63(6):1311–1322.  
 Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676.  
 Cooke TD, Tilney MS, Tilney LG. 1996. Plasmodesmatal networks in apical meristems and mature structures: geometric evidence for both primary and secondary formation of plasmodesmata. In: Smallwood M, Knox JP, Bowles DJ, editors. *Membranes: specialized functions in plants*. Cambridge: BIOS Scientific. p. 471–488.  
 Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.  
 Ding B, et al. 1992. Secondary plasmodesmata are specific sites of localization of the tobacco mosaic virus movement protein in transgenic tobacco plants. *The Plant Cell* 4(8):915–928.  
 Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus.* 12:13–15.  
 Doyle JA. 2008. Integrating molecular phylogenetic and paleobotanical evidence on origin of the flower. *Int J Plant Sci.* 169:816–843.  
 Drummond AJ, Suchard MA, Dong X-P, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29(8):1969–1973.

- Eisel D, Seth O, Grünewald-Janho S, Kruchen B. 2008. DIG application manual for nonradioactive *in situ* hybridization. 4th ed. Mannheim, Germany: Roche Diagnostics GmbH, Roche Applied Science.
- Evkaikina AI, Romanova MA, Voitsekhovskaja OV. 2014. Evolutionary aspects of non-cell-autonomous regulation in vascular plants: structural background and models to study. *Front Plant Sci*. 5:31.
- Finet C, et al. 2016. Evolution of the YABBY gene family in seed plants. *Evol Dev*. 18(2):116–126.
- Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 44(D1):D279–D285.
- Floyd SK, Bowman JL. 2006. Distinct developmental mechanisms reflect the independent origins of leaves in vascular plants. *Curr Biol*. 16(19):1911–1917.
- Frank MH, et al. 2015. Dissecting the molecular signatures of apical cell-type shoot meristems from two ancient land plant lineages. *New Phytol*. 207(3):893–904.
- Friis EM, Crane PR, Pedersen KR. 2011. Early flowers and angiosperm evolution. New York, NY: Cambridge University Press.
- Gaillochot C, Daum G, Lohmann JUO. 2015. cell, where art thou? The mechanisms of shoot meristem patterning. *Curr Opin Plant Biol*. 23:91–23.
- Gao J, Yang X, Zhao W, Lang T, Samuelsson T. 2015. Evolution, diversification, and expression of KNOX proteins in plants. *Front Plant Sci*. 6:882.
- Gene Ontology Consortium. 2001. Creating the gene ontology resource: design and implementation. *Genome Res*. 11:1425–1433.
- Gibson AK, Smith Z, Fuqua C, Clay K, Colbourne JK. 2013. Why so many unknown genes? Partitioning orphans from a representative transcriptome of the lone star tick *Amblyomma americanum*. *BMC Genomics* 14(1):135.
- Gifford EM, Foster AS. 1989. Morphology and evolution of vascular plants. New York, NY: Freeman.
- Gifford EM. 1983. Concept of apical cells in bryophytes and pteridophytes. *Annu Rev Plant Physiol*. 34(1):419–440.
- Goff SA, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296(5565):92–100.
- Gola EM, Jernstedt JA. 2011. Impermanency of initial cells in *Huperzia lucidula* (Huperziaceae) shoot apices. *Int J Plant Sci*. 172(7):847–855.
- Goodstein DM, et al. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 40(Database issue):D1178–D1186.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 29(7):644–652.
- Gradstein FM, Ogg JG, Schmitz M, Ogg G. 2012. The Geologic Time Scale 2012. Oxford, UK: Elsevier.
- Graham LE, Cook ME, Busse JS. 2000. The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proc Natl Acad Sci U S A*. 97(9):4535–4540.
- Granoff JA, Gensel PG, Andrews HN. 1976. A new species of *Pertica* from the Devonian of eastern Canada. *Palaeontographica B*. 155:119–128.
- Gunning BES. 1978. Age-related and origin-related control of the numbers of plasmodesmata in cell walls of developing *Azolla* roots. *Planta* 143:181–190.
- Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8(8):1494–1512.
- Hake S, Freeling M. 1986. Analysis of genetic mosaics shows that the extra epidermal cell divisions in *Knotted* mutant maize plants are induced by adjacent mesophyll cells. *Nature* 320:621–623.
- Hao S-G. 1988. A new Lower Devonian genus from Yunnan, with notes on the origin of leaves. *Act Bot Sinica*. 30:441–448.
- Harrison CJ, et al. 2005. Independent recruitment of a conserved developmental mechanism during leaf evolution. *Nature* 434(7032):509–514.
- Harrison CJ, Rezvani M, Langdale JA. 2007. Growth from two transient apical initials in the meristem of *Selaginella kraussiana*. *Development* 134(5):881–889.
- Heidstra R, et al. 1994. Root hair deformation activity of nodulation factors and their fate on *Vicia sativa*. *Plant Physiol*. 105(3):787–797.
- Hilton J, Bateman RM. 2006. Pteridospers are the backbone of seed-plant phylogeny. *J Torrey Bot Soc*. 133:119–168.
- Hoffman LA, Tomescu AMF. 2013. An early origin of secondary growth: *Franhuerberia gerriennei* gen. et sp. nov. from the Lower Devonian of Gaspé (Quebec, Canada). *Am J Bot*. 100:754–763.
- Holt AL, van Haperen JM, Groot EP, Laux T. 2014. Signaling in shoot and flower meristems of *Arabidopsis thaliana*. *Curr Opin Plant Biol*. 17:96–102.
- Imaichi R, Hiratsuka R. 2007. Evolution of shoot apical meristem structures in vascular plants with respect to plasmodesmatal network. *Am J Bot*. 94(12):1911–1921.
- Jaillon O, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161):463–467.
- Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ. 2002. Plant systematics: a phylogenetic approach, 2nd edn. Sunderland: Sinauer Assoc.
- Karlgen A, Carlsson J, Gyllenstrand N, Lagercrantz U, Sundström JF. 2009. Non-radioactive *in situ* hybridization protocol applicable for norway spruce and a range of plant species. *J Vis Exp*. 26:e1205.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Kenrick P. 2003. Palaeobotany: Fishing for the first plants. *Nature* 425(6955):248–249.
- Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. *Nature* 389(6646):33–39.
- Kessler S, Sinha N. 2004. Shaping up: the genetic control of leaf shape. *Curr Op Plant Biol*. 7(1):65–72.
- Kidston R, Lang WH. 1920. On Old Red Sandstone plants showing structure, from the Rhynie Chert Bed, Aberdeenshire. Part III. *Asteroxylon mackiei*, Kidston and Lang. *Trans R Soc Edinb*. 52(03):643–680.
- Korall P, Kenrick P, Therrien J. 1999. Phylogeny of Selaginellaceae: evaluation of generic/subgeneric relationships based on *rbcl* gene sequences. *Int J Plant Sci*. 160:585–594.
- Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. 2013. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet*. 4:237.
- Kumaran MK, Bowman JL, Sundaresan V. 2002. YABBY polarity genes mediate the repression of KNOX homeobox genes in Arabidopsis. *Plant Cell* 14(11):2761–2770.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10(3):R25.
- Larsén E, Rydin C. 2016. Disentangling the phylogeny of *Isoetes* (Isoetales), using nuclear and plastid data. *Int J Plant Sci*. 177(2):157–174.
- Lee H, et al. 2016. The genome of a Southern hemisphere seagrass species (*Zostera muelleri*). *Plant Physiol*. 172(1):272–283.
- Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 44(W1):W242–W245.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res*. 21(6):936–939.
- Machida C, Nakagawa A, Kojima S, Takahashi H, Machida Y. 2015. The complex of ASYMMETRIC LEAVES (AS) proteins plays a central role in antagonistic interactions of genes for leaf polarity specification in Arabidopsis. *Wiley Interdiscip Rev Dev Biol*. 4:655–671.



- Matasci N, et al. 2014. Data access for the 1, 000 Plants (1KP) project. *Gigascience* 3:17.
- McConnell JR, Barton MK. 1998. Leaf polarity and meristem formation in *Arabidopsis*. *Development* 125(15):2935–2942.
- Merchant SS, et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318(5848):245–250.
- Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41(12):e121.
- Newman IV. 1965. Pattern in the meristems of vascular plants. III. Pursuing the patterns in the apical meristem where no cell is a permanent cell. *J Linn Soc Bot.* 59(378):185–214.
- Nystedt B, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497(7451):579–584.
- Palenik B, et al. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci U S A.* 104(18):7705–7710.
- Perales M, Reddy V. 2012. Stem cell maintenance in shoot apical meristems. *Curr Opin Plant Biol.* 15(1):10–16.
- Philipson WR. 1990. The significance of apical meristems in the phylogeny of land plants. *Plant Sys Evol.* 173(1–2):17–38.
- Pires ND, Dolan L. 2012. Morphological evolution in land plants: new designs with old genes. *Philos Trans R Soc Lond B Biol Sci.* 367(1588):508–518.
- Popham RA. 1951. Principal types of vegetative shoot apex organization in vascular plants. *Ohio J Sci.* 51:249–270.
- Pryer KM, et al. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409(6820):618–622.
- Qiu YL, et al. 2007. A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *Int J Plant Sci.* 168(5):691–708.
- Rambaut A, Drummond AJ. 2007. Tracer. MCMC trace analysis tool. Version 1.4: Computer program distributed by the authors. Available from: <http://beast.bio.ed.ac.uk/> (last accessed June 8, 2014).
- Rensing SA, et al. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319(5859):64–69.
- Rowe NP. 1988. A herbaceous lycophyte from the Lower Carboniferous Drybrook Sandstone of the Forest of Dean, Gloucestershire. *Palaeontology* 31:69–83.
- Rydin C, Källersjö M, Friis EM, Kne of the Forriis EM 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems and the monophyly of conifers. *Int J Plant Sci.* 163(2):197–214.
- Ruzin SE. 1999. *Plant microtechnique and microscopy*. New York: Oxford University Press. p. 322.
- Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* 326(5956):1112–1115.
- Siegfried KR, et al. 1999. Members of the YABBY gene family specify abaxial cell fate in *Arabidopsis*. *Development* 126(18):4117–4128.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Smith GM. 1938. *Cryptogamic botany. Vol. II. Bryophytes and pteridophytes*. New York: McGraw-Hill Book Company, Inc. p. 380.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Steeves TA, Sussex IM. 1989. *Patterns in plant development*. 2nd ed. Cambridge, UK: Cambridge University Press.
- Swofford DL. 2002. PAUP\*: Phylogenetic analysis using parsimony (\*and other methods). Sunderland (MA): Sinauer Associates.
- Taylor TN, Taylor EL, Krings M. 2009. *Paleobotany - The biology and evolution of fossil plants*. New York, NY: Academic Press.
- Tuskan GA, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793):1596–1604.
- Vasco A, et al. 2016. Challenging the paradigms of leaf evolution: Class III HD-Zips in ferns and lycophytes. *New Phytol.* 212(3):745–758.
- Visser EA, Wegryn JL, Steenkamp ET, Myburg AA, Naidoo S. 2015. Combined de novo and genome guided assembly and annotation of the *Pinus patula* juvenile shoot transcriptome. *BMC Genomics* 16(1):1057.
- Waites R, Hudson A. 1995. *phantastica*: a gene required for dorsiventrality of leaves in *Antirrhinum majus*. *Development* 121:2143–54.
- Weng JK, Tanurdzic M, Chapple C. 2005. Functional analysis and comparative genomics of expressed sequence tags from the lycophyte *Selaginella moellendorffii*. *BMC Genomics* 6:85.
- Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci U S A.* 111:E4859–E4868.
- Worden AZ, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324(5924):268–272.
- Worsdell W. 1905. *The principles of morphology II. The evolution of the sporangium*. *New Phytol.* 4(7):163–170.
- Xu C, et al. 2015. De novo and comparative transcriptome analysis of cultivated and wild spinach. *Sci Rep.* 5(1):17706.
- Yamaguchi T, Nukazuka A, Tsukaya H. 2012. Leaf adaxial-abaxial polarity specification and lamina outgrowth: evolution and development. *Plant Cell Physiol.* 53(7):1180–1194.

Associate editor: Susanne S. Renner