

RESEARCH

Open Access



# In silico characterization of putative gene homologues involved in somatic embryogenesis suggests that some conifer species may lack *LEC2*, one of the key regulators of initiation of the process

Sonali Sachin Ranade\* and Ulrika Egertsdotter

## Abstract

**Background:** Somatic embryogenesis (SE) is the process in which somatic embryos develop from somatic tissue in vitro on medium in most cases supplemented with growth regulators. Knowledge of genes involved in regulation of initiation and of development of somatic embryos is crucial for application of SE as an efficient tool to enable genetic improvement across genotypes by clonal propagation.

**Results:** Current work presents in silico identification of putative homologues of central regulators of SE initiation and development in conifers focusing mainly on key transcription factors (TFs) e.g. *BBM*, *LEC1*, *LEC1-LIKE*, *LEC2* and *FUSCA3*, based on sequence similarity using BLASTP. Protein sequences of well-characterised candidate genes from *Arabidopsis thaliana* were used to query the databases (Gymno PLAZA, Congenie, GenBank) including whole-genome sequence data from two representative species from the genus *Picea* (*Picea abies*) and *Pinus* (*Pinus taeda*), for finding putative conifer homologues, using BLASTP. Identification of corresponding conifer proteins was further confirmed by domain search (Conserved Domain Database), alignment (MUSCLE) with respective sequences of *Arabidopsis thaliana* proteins and phylogenetic analysis (Phylogeny.fr).

**Conclusions:** This in silico analysis suggests absence of *LEC2* in *Picea abies* and *Pinus taeda*, the conifer species whose genomes have been sequenced. Based on available sequence data to date, *LEC2* was also not detected in the other conifer species included in the study. *LEC2* is one of the key TFs associated with initiation and regulation of the process of SE in angiosperms. Potential alternative mechanisms that might be functional in conifers to compensate the lack of *LEC2* are discussed.

**Keywords:** *ABSCISIC ACID INSENSITIVE3*, *BABYBOOM*, Conifer, *FUSCA3*, *LEAFY COTYLEDON*, *PICKLE*, *SERK*, Somatic embryogenesis, *WUSCHEL*, *WOX2*

\* Correspondence: [Sonali.Ranade@slu.se](mailto:Sonali.Ranade@slu.se)

Department of Forest Genetics and Plant Physiology, Umeå Plant Science Center (UPSC), Swedish University of Agricultural Science (SLU), 901 83 Umeå, Sweden



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Somatic embryogenesis (SE) is the process of non-sexual reproduction in which the embryos develop from somatic tissue *in vitro* on medium in most cases supplemented with growth regulators. Somatic embryos morphologically resemble the zygotic embryos; SE in conifers involves the formation of early-stage somatic embryos, so-called pro-embryogenic masses (PEM), followed by somatic embryo maturation, partial drying with desiccation and germination, giving rise to plants [1]. SE has gained importance not only for its use as a model system in basic studies related to molecular genetics and developmental biology but largely due its application for the large-scale vegetative propagation of plants of uniform quality with selected characters, for commercial purposes [2]. This is of particular interest to the important part of the forest industry based on conifers where the majority of species with large commercial potential are difficult to propagate by traditional cloning methods. In addition, conifers have slow growth, long generation time and very large genome size that makes their genetic improvement difficult and time consuming. SE allows genetic improvements from conifer breeding programs to be captured at an earlier stage and large numbers of high-value plants can be produced [3]. Clonal propagation by SE was successfully demonstrated in coniferous species in the 1980s in *Picea abies* (*P. abies* [L.] Karst, Norway spruce) [4, 5], then in other genera in the family Pinaceae namely *Abies*, *Larix*, *Picea*, *Pinus* and *Pseudotsuga* [6] and only few species from other conifers belonging to the families Cupressaceae, Taxaceae, Cephalotaxaceae, and Araucariaceae [1]. However, regardless of major technical advances in clonal propagation by SE in conifers, some biological bottlenecks remain. A key step of concern is the limited initiation of SE across genotypes where only a subsection of the seeds can be induced to form a culture of somatic embryos. Furthermore, there are losses in each step during the subsequent development from PEMs to plant lowering the yields. There is only limited information available on the regulation of the SE processes in conifers. Therefore, the identification of key proteins controlling SE with reference to their structural domains deserves primary attention from the conifer perspective. Most investigations in conifers have been focused on the domain characterisation of the WUSCHEL (WUS) and WUS-related homeobox (WOX) protein family [7, 8]. Although expression profiles of some genes associated with SE initiation in conifers have been reported [9], the genetic and molecular interactions in the regulatory network associated with SE development has not been investigated in these species. In addition, there is no information available in conifers regarding the genes involved in suppression of SE (e.g. *PICKLE*). The

motivation for the present study is therefore to summarize information on conifer homologues for the most relevant key regulatory genes involved in SE in model species with the aim to provide a foundation for further detailed studies into functional regulation of the SE process in conifers.

The current work presents *in silico* identification of putative homologues in conifers to the key regulators of SE based on sequence similarity using BLASTP. These key regulators of SE have been previously identified in *Arabidopsis thaliana* (*A. thaliana*). The analysis includes the identification of putative functional domains of the respective genes, again based on sequence similarity. In addition, relevant information available in the literature with reference to genes associated with SE in conifers has also been reviewed. The analysis focuses mainly on the transcription factors that are demonstrated to be directly involved in the initiation of the SE in the model plants, primarily *A. thaliana*. A few other genes that are known to play significant role during the SE process were also included in the analysis, e.g. *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE* (*SERK*) which is associated with the initiation of SE and genes like *PICKLE* (*PKL*) that are involved in the suppression of SE.

Although PEMs is generally initiated from the immature zygotic embryos in conifers [10], recent studies have also reported SE initiation from the primordial shoot explants and matured embryos of SE plants in *P. abies* [11] and *Picea glauca* (*P. glauca*, White spruce,) [12]. Protocols are well established for the induction of SE in various angiosperm plants and gymnosperms including coniferous tree species, yet the information on underlying genetic regulatory mechanism is largely missing. SE in conifers can in most cases be induced by treating the primary explants with plant growth regulators e.g. auxin (2,4-dichlorophenoxyacetic acid) and cytokinin (N6-benzyladenine), or also by wounding or other stress factors e.g. temperature, heavy metal ions, starvation or osmotic stress [13]. Molecular mechanisms governing the regeneration in the explants of coniferous forest tree species with a focus on interaction between auxin and stress conditions have been reviewed [14]. Ectopic and/or over expression of the key transcription factors involved in the development of SE might also give rise to somatic embryos (discussed in the later part of “Background”).

### Genes involved in SE initiation

The core of understanding the SE process lies in the recognition of signals that change the genetic program of somatic tissue to induce the formation of a somatic embryo. This process involves the regulation of gene expression in the somatic tissue that form a somatic

embryo and also in its surrounding tissue. The role of the genes involved in the process of initiation and the regulation of development of the somatic embryos is well characterised in model plants like *A. thaliana*. The key transcription factors (TFs) which regulate this process include BABYBOOM (*BBM*), *EMBRYOMAKER* (*EMK*), *LEAFY COTYLEDON* (*LEC1*, *LEC2*), *LEC1-LIKE* (*LIL*), *ABSCISIC ACID INSENSITIVE 3* (*ABI3*) or *VIVI PAROUS* (*VPI*), *FUSCA3* (*FUS3*), *WUSCHEL* (*WUS*) and the *WUSCHEL-related homeobox* (*WOX*) 2 [15, 16]. These TFs share a complex association with auxin signalling pathways involving a number of gene regulatory networks where various crosstalk and feedback loops play a major role [15, 16]. Seed maturation is synchronised by the complex LAFL regulatory network, which includes *LEC1* and *LIL* of the NF-YB gene family, and the *ABI3/VPI*, *FUS3* and *LEC2* containing the B3 DNA-binding domain and belonging to the B3-AFL gene family. This network positively controls genes involved in embryo/seed development and maturation and represses those required for the transition from embryonic to vegetative development, suppressing premature germination [17].

*LECs* (*LEC1*, *LEC2*, *LEC1-LIKE*) are among the key regulators that promote the initiation of SE and are involved in the process of early embryo development and maturation [18]. *LECs* induce formation of somatic embryos when expressed ectopically [19]. Ectopic expression of *LIL* marked the embryogenic competence in epiphyllous plants [20], while ectopic over-expression of *LEC1* [21] and *LEC2* [22] was found to be associated with formation of somatic embryos in *A. thaliana*. By contrast, in conifers, the over-expression of *LEC1* homolog gene did not induce ectopic somatic embryo formation in *P. glauca* but abundance of *LEC1* transcripts was detected in PEMs but not in (non-embryogenic) callus; however in *Pinus contorta* (*P. contorta*, Lodgepole pine) [23] and *Pinus strobus* (*P. strobus*, White pine) [12], callus also showed expression of the *LEC1* homolog. A conifer *LEC1*-type gene (*PaHAP3A*) that is active during embryo development in *P. abies*, did not stimulate embryonic features in vegetative tissues; however, expression of *PaHAP3A* was observed during early to late embryo development and overexpression of *PaHAP3A* during the maturation stage leading to the differentiation of ectopic embryos from maturing somatic embryos [24]. Expression of *LEC1/LEC1-LIKE* gene was found to be associated with early to late embryo development in *Pinus sylvestris* (*P. sylvestris*, Scots pine) [25], *Pinus pinaster* (*P. pinaster*, Maritime pine) [26] and *Araucaria angustifolia* (*A. angustifolia*, Brazilian pine) [27].

*FUSCA3* regulates gene expression during late embryogenesis and it acts together with *LEC1* and *LEC2* controlling the plant embryo development; embryos

carrying *LEC1*, *LEC2* and *FUS3* loss-of-function mutants partially lose their embryo identity and enter post-germinative programs [28]. *VPI* is homologous to the *A. thaliana* *ABI3* which is essential for seed maturation; *ABI3* regulates the transition between embryo maturation and early seedling development and is the central regulator of ABA signalling pathway [29]. *FUSCA3* and *ABI3* do not induce SE on overexpression in *A. thaliana* [30, 31]. Differential expression of *FUS3* was observed in *P. glauca* during late SE development due to the inclusion of polyethylene glycol (PEG) in the maturation medium, which is proposed to improve the number and quality of the embryos produced [32]. Gene expression studies of SE in conifer species revealed the expression of *ABI3/VPI* during early to late somatic embryogenesis in *P. abies* [25, 33] and *P. sylvestris* [25, 34], and during initiation and early SE in *P. glauca* [12]. *VPI* is functionally conserved in *P. abies* and seed plants, considering not only the development of embryos, but also the later stages of plant life [35].

The *AINTEGUMENTA-LIKE* (*AIL*) gene clade coding for TFs with the APETALA2 domain (AP2-domain) includes *AINTEGUMENTA* (*ANT*) and *AIL* or *PLET HORA* (*PLT*) genes to which *BBM* (*PLETHORA4*, *PLT4*) and *EMK* (*PLETHORA5*, *PLT5*) belong [36]. The *A. thaliana* genome contains eight *AIL/PLT* genes that are expressed in the embryo and root/shoot meristems; they are required for stem cell maintenance and the functioning of meristems as well as for embryo development [36]. *BBM* is one of the central regulators of the developmental potency of plant cells having diverse functions in plant cell proliferation, growth and development, and is found to be expressed in embryos and lateral root primordia [36, 37]. *BBM* acts upstream of other major TFs involved in plant embryo identity as it triggers the *LEC1-ABI3-FUS3-LEC2* network to induce SE [38]. Ectopic expression of *BBM* induces SE in *A. thaliana* [39]. With reference to conifers, *BBM* studies have been confined to larch species and *P. glauca*. Increased expression of *BBM* was identified during later developmental stages of embryo development in *Larix decidua* (*L. decidua*, European larch) [40]. In *P. glauca*, *BBM* was observed to be involved in the initiation of SE and was found to be expressed specifically in the early stages of embryo development [12]. *BBM* along with *LEC* were proposed to be potential molecular markers for embryogenicity by these investigations. Apart from its involvement in the process of SE in conifers, *BBM* expression was proposed as a molecular marker for root primordia in hybrid larch (*Larix kaempferi* × *Larix olgensis*); *BBM* showed root-specific expression compared to the gene expression levels in the stem, stem tip and leaf, which indicated that *BBM* plays a vital role in regulating the development and growth of root during adventitious

rooting in larch [41]. Yet another study concluded the role of *BBM* (*LkBBM1* and *LkBBM2*) in the regulation of adventitious root development in the same larch hybrid [42].

*EMK* or *AIL5* codes for members of the AP2/ethylene-responsive element binding protein (AP2/EREBP) superfamily having the AP2 DNA-binding domain. *EMK* is involved in germination and seedling growth, and is essential for the developmental transition between the embryogenic and vegetative phases; over-expression of *EMK* resulted in the formation of somatic embryos on cotyledons in *A. thaliana* [43]. Early embryo development is associated with cleavage polyembryony in *Pinus* species but not in *Picea*, a process where the proembryo undergoes a cleavage process giving rise to multiple embryos; only one of these embryos develops to a dominant embryo that matures to a cotyledonary embryo, while the other embryos (subordinate embryos) are degraded [44, 45]. Genome-wide transcript expression profiling of early stages of zygotic embryo development in *P. sylvestris* showed transcript abundance of *AIL5* (*PsAIL5*) along with low expression of *VPI* (*PsVPI*) in subordinate embryos, while *PsAIL5* was down-regulated along with up-regulation of *PsVPI* in the dominant embryo. This indicated that the transition from the morphogenic phase to the maturation phase was not completed in the subordinate embryos [34].

The *WOX* family of TFs is comprised of multiple members, of which *WUS* and *WOX2* are associated with the initiation of SE. *WUS* promotes embryonic identity and vegetative-to-embryonic transition; ectopic *WUS* expression induces SE in *A. thaliana* [46]. In *P. glauca*, PEMs transformed with *A. thaliana WUS* produced severe phenotypes by disrupting the development of somatic embryos on the maturation medium and inhibiting germination; however *WUS* did not induce ectopic somatic embryogenesis even in the presence of plant growth regulators [47]. One of the early events in angiosperm embryogenesis is the asymmetric cell division that results in formation of an apical cell which forms the majority of the embryo, and a basal cell which forms the suspensor. *WOX2* becomes confined to the apical cell, thus marking the apical descendants of the zygote in *A. thaliana* involved in its further development [48]. In conifers, there is no corresponding early asymmetric cell division. However, the embryonic region of the early stage conifer embryos constitutes the corresponding tissue responsible for further development of the embryo. High expression of *WOX2* is associated with the early growth stages of somatic embryo in *P. glauca* [12], *P. abies* [8], *P. contorta* [23] and *P. pinaster* [26] and during late embryogeny in *P. abies* [49]. *WOX2* shows evolutionary conserved function related to protoderm formation early during embryo development among seed

plants; in addition, it also plays a unique role in suspensor expansion in gymnosperms [49]. Upregulation of a *WOX* gene was observed during the early to late stages of SE in *A. angustifolia* [27]. *WOX2* expression was much lower at the later embryonal stages in *P. abies* and it was not detected in non-embryogenic cell culture, therefore it can be used as a marker for embryogenic potential [8, 50]. *WOX2* transcripts were found not only in the early to late embryo developmental stages but also in the vegetative tissues of seedlings and mature/older trees in *P. abies* [8, 51] and *P. contorta* [23]. Interestingly, *WOX2* was found to be expressed in all developmental stages of somatic embryos in *P. sylvestris* where polyembryony exists, but significantly higher levels of *WOX2* expression was detected in subordinate embryos, which might be related to the blocked development of the subordinate embryos [34]. In *Cunninghamia lanceolata* (*C. lanceolata*, Chinese fir), however, the *WOX2* expression was not associated with the development of the embryos, instead *WOX13* transcripts showed high correlation with the transition of PEMs to proembryos [52].

*SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE* (*SERK*) belongs to the leucine-rich repeat receptor-like family of kinases (LRR-RLKs) that are involved in multiple processes in plant development. *SERK* contributes significantly to the process plant embryogenesis and is also found to be involved in diverse plant processes related to cell differentiation, growth and development, and plays important role in plant defense and plant responses to environmental cues [53]. Five *SERK* genes (*SERK1–5*) are identified in *A. thaliana*, where *SERK1* forms a component of the embryogenesis signalling pathway [54]. Overexpression of *SERK1* enhanced embryogenic competence in tissue cultures of *A. thaliana* [54]. Expression of *SERK1-like* was associated with initiation and early SE in *P. glauca*, as its expression was detected to be higher in the PEMs than in callus, however it was lower in the PEMs than in young shoot buds [12]. A putative homolog of *SERK1* gene was found to be expressed in *P. sylvestris* specifically at the very early stage of embryo development [34]. In *A. angustifolia*, *SERK1* transcripts initially accumulated in the groups of cells at the periphery of the PEMs and were then restricted to the developing embryo [55]. *SERK1–3* and *SERK1–4* in *C. lanceolata* share a high similarity with *A. thaliana SERK1*, and are predominantly expressed in PEMs indicating that both have functions during SE [52].

#### Genes involved in suppression of SE

*PICKLE* (*PKL*) codes for a chromatin re-modeling factor that belongs to the chromodomain-helicase-DNA-binding (CHD) subfamily II. CHD complexes regulate the assembly and organization of mature nucleosomes along



the DNA. CHD proteins are members of ATP-dependent chromatin remodeling complexes that are characterized by presence of the chromo (chromatin organization modifier) domains, SNF2-related helicase/ATPase domain and a DNA-binding domain [56]. *PKL* is necessary to repress expression of embryonic traits during germination and it regulates the transition from embryonic to vegetative development in *A. thaliana* [57]. In particular, *PKL* is necessary for repression of *LEC1*, a transcription factor, which is one of the key regulators that initiates embryo development [58]. However, there is a lack of information on the function of *PKL* in conifers.

VP1/ABI3-LIKE (VAL) proteins belong to the plant specific B3 TF superfamily the members of which contains the conserved B3 DNA-binding domain [17]. VALs in *A. thaliana* contain PHD-L (plant homeodomain-like), Zf (Zinc finger), B3, CW-Zf (named CW for its conserved cysteine and tryptophan residues) and EAR (ethylene response factor [ERF]- associated repression) domains [59]. The B3 domain of VAL mediates the repression of genes of the LAFL network; B3 domain of *VAL1* and *VAL2* is more similar when compared to *VAL3* [60] and all key residues involved in direct DNA contacts are conserved among *VAL1* and *VAL2* [61]. The PHD and CW-Zf domains are the histone modification readers which are involved in recognition, and the EAR motifs mediate the transcriptional repression [62]. In *A. thaliana*, *VAL1* and *VAL2* have been reported as suppressors of somatic embryogenesis [63]; nevertheless, the VAL genes function as suppressors of the LAFL genes during germination, but not during seed development [64]. Knock-down mutations in genes encoding the VAL proteins led to increased expression of *LEC* genes that resulted in the formation of ectopic somatic embryos on seedlings [65]. Similarly to *PKL*, the functional mechanism of the action of VAL genes has not been investigated in conifers.

## Results and discussion

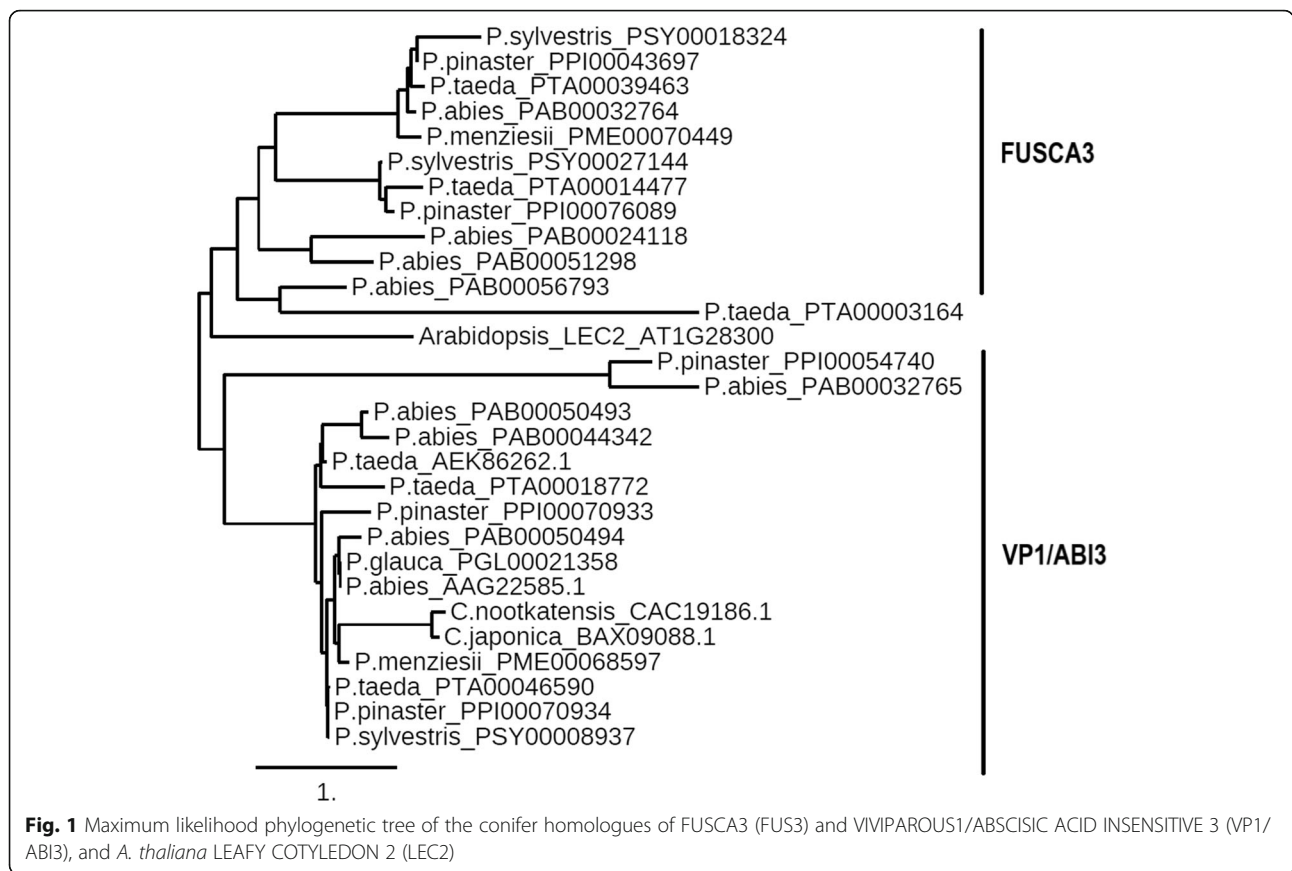
Homologues for all the candidate genes considered as involved in the initiation of SE were detected in the conifer species included in the analysis except one of the key regulators - *LEC2*. This in silico analysis suggests absence of *LEC2* in *P. abies* and *Pinus taeda* (*P. taeda*, Loblolly pine), the conifer species whose genome has been sequenced. Based on available sequence data to date, *LEC2* was not detected in the other conifer species included in the study. The details regarding the conifer homologues such as sequence ID, length of the protein etc., are included in the supplementary information (Additional file 1.xlsx). Full-length homologues of the candidate genes were detected in most conifer species with few exceptions; however, our results include all the

partial homologues as well, as this aspect is expected to improve with technological advances in the future through availability of elaborate and accurate data e.g. longer reads with PacBio sequencing. In few instances, more than one homologous sequence was detected for a specific candidate gene in a particular conifer species e.g. two *BBM* gene loci were detected in *P. abies* and *P. taeda*. These loci considerably differed in their protein sequences, which can be inferred from the alignment results (Additional file 2.pdf). This phenomenon is also observed with other genes and tree species, e.g. *Populus trichocarpa* (Torr. & Gray) has one *PHYA* locus and two *PHYB* loci, which were designated as *PHYB1* and *PHYB2* [66]. All homologues from conifer species for a specific candidate gene were aligned along with the corresponding gene from *A. thaliana* and the characteristic motifs/domains of the respective genes in the conifer homologues are highlighted with different colours and named accordingly with the specific domain names based on the scientific convention as referred from the literature (Additional file 2.pdf - Additional file 11.pdf).

## Conifer homologues of genes involved in SE initiation

### Homologue of *LEC2* was not found in conifer species included in the analysis

In the current work, *LEC2*, which is a TF that plays a key role in the initiation and regulation of SE, was found to be absent from the genomes of the *P. abies* and *P. taeda*. *LEC2* was not detected in the other conifer species included in the study, based on available sequence data in those conifers to date. This observation is strongly supported by the fact that the searches were performed on full genomes of two conifer species involved in this analysis, one each from the genus *Picea* (*P. abies*) and *Pinus* (*P. taeda*). The phylogenetic tree constructed with conifer homologues of the B3 domain containing TFs (*FUS3* and VP1/ABI3, as *LEC2* is absent in conifers) and the *A. thaliana* *LEC2*, indicates that conifer *FUS3* and VP1/ABI3 form separate clusters and the *A. thaliana* *LEC2* forms a distinct clade (Fig. 1). In addition, all the *LEC*-like conifer homologues showed better alignment with *A. thaliana* *LEC1/LEC1-LIKE* than *A. thaliana* *LEC2* (Additional file 3.pdf). This suggests that *LEC2* may be absent in conifers, at least in the two conifer species whose whole-genome sequence data is available (*P. abies* and *P. taeda*). Furthermore, several transcriptomic investigations related to SE development in conifer species have been conducted but none of them reported the expression of *LEC2* [34, 67–69], whereas *LEC2* expression is commonly reported in transcriptome analyses in model systems e.g. *A. thaliana* [70, 71]. Likewise, an earlier study reported that *ABI3* homologues were found in all land plant genomes, but the *FUS3* homologues were present only in seed plants,



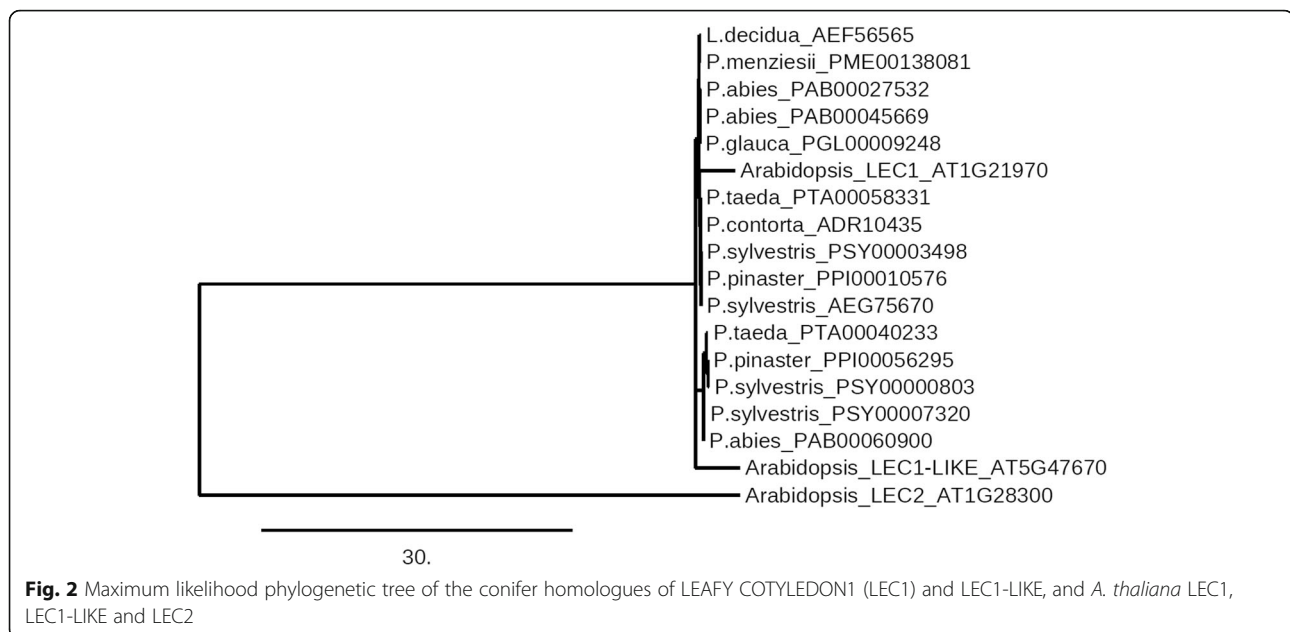
while the *LEC2*-like sequences were detected only in dicot genomes [72]. Phylogenetic and gene structure analyses of AFL genes (*ABI3/VP1*, *FUS3* and *LEC2*) in land plant species revealed loss of *LEC2* type genes in monocots [17]. This further supports our hypothesis that *LEC2* may be broadly absent in conifers.

From the context of loss of genes during evolution, eukaryotic plastid genome has lost many genes during the early events of endosymbiosis; some of these genes were lost totally, while others were found to be relocated and got functionally integrated to the host nuclear genomes during plant evolution [73]. In conifers, loss of *ndh* genes from several species is evident from plastid genome sequencing projects [74] but later, the presence of non-functional plastid *ndh* gene fragments was confirmed in the nuclear genome of *P. abies* [75]. Likewise, there is a specialization of the photosynthetic apparatus in Pinaceae; comparative analysis of the gene families reported gains and losses of genetic networks associated with photosynthesis in *Pseudotsuga menziesii* (*P. menziesii*, Douglas-fir) from family Pinaceae [76]. The current analysis suggests loss of *LEC2* gene from *P. abies* and *P. taeda*, and also from other conifer species included in the study, based on the available sequence information. Embryo development in gymnosperms including

conifers is very different from angiosperms in several aspects. For example, the endosperm of gymnosperm is haploid as there is no double fertilization. The conifers possess multiple cotyledons which is a distinctive phenotypic character compared to the monocots and dicots. Few such mechanisms/phenomenon in gymnosperms which are different from the angiosperms, could explain the lack of a master embryogenesis regulator such as *LEC2* gene from the conifers.

#### **LEAFY COTYLEDON1 (LEC1) and LEC1-LIKE**

BLASTP with *A. thaliana* *LEC1* and *LEC1-LIKE*, resulted in finding the conifer genes characterised as the *LEC*-like CCAAT-box binding factor HAP3 or the *LEC1*-type HAP3 subunit coding protein. Congenie displayed *A. thaliana* *LEC1-LIKE* as the best match for the conifer homologues detected. The phylogenetic tree constructed with conifer homologues of the *LEC* sequences precisely indicates that all the *LEC/LEC-LIKE* conifer homologues either cluster together with *A. thaliana* *LEC1* or *A. thaliana* *LEC1-LIKE* sequences (Fig. 2). Here *A. thaliana* *LEC2* is an outgroup which forms a separate clade. All the *LEC*-like conifer homologues showed alignment with *A. thaliana* *LEC1* and *LEC1-LIKE* sequences (Additional file 3.pdf). The Asp (D) residue is



critical for the LEC function [77, 78] was found to be conserved in conifers. In addition, the residues unique to *LEC1* and *LEC1-LIKE* HAP3 subunits in the B-domain were found to be conserved in conifers (Figure S57, Additional file 3.pdf); these residues were absent from the B-domain of other HAP3 proteins [79, 80].

#### ***FUSCA3 (FUS3)***

The intact B3 domain is essential for the regulation of seed maturation by *FUS3* [81]. Similar to the angiosperms [72, 82], the B domain of *FUS3* was more conserved among the conifer species as compared to the N-terminal domain and transcription-activating domain (Figure S16, Additional file 4.pdf). The transcription-activating domain contains conserved *FUS3*-specific fragments in dicots and in monocots respectively [72]. Likewise, the transcription-activating domain of the *FUS3* sequences was found to contain conifer-specific fragments as the transcription-activating domain shows good alignment within the conifer species included in the study but not with *A. thaliana*.

#### ***Viviparous 1 (VP1)***

The *VP1* protein contains four domains – A1, B1, B2 and the B3 [35]; the *VP1* gene with all the four domains were detected for all the conifer species included in this study. Three homologues of *VP1* were detected in *P. taeda* that contained all four domains, while in case of *P. pinaster*, only one sequence (PPI00070933) out of the two with all four domains seems to be the precise homologue of *VP1* as the other sequence (PPI00070934) did not show good alignment with the *A. thaliana* *VP1* (Additional file 5.pdf). Four homologues of *VP1* were

detected in *P. abies* but only one sequence (AAG22585.1) showed all four domains (Figure S19, Additional file 5.pdf). One sequence from *P. abies* contained the A1 and the B1 domains (PAB00050494), while the other contained B2 and B3 domains (PAB00050493). We propose that these two sequences may be parts of the same gene but are indicated as separate genes possibly due to annotation and/or sequencing issues. The putative nuclear localization signal (RKNR) of the B2 domain [35] was found to be conserved in all conifer species that showed presence of the B2 domain. The B3 DNA-binding domain of the *VP1* genes is well conserved as reported earlier [17] among all conifer homologues and also shows high similarity with *A. thaliana* (Figure S19, Additional file 5.pdf).

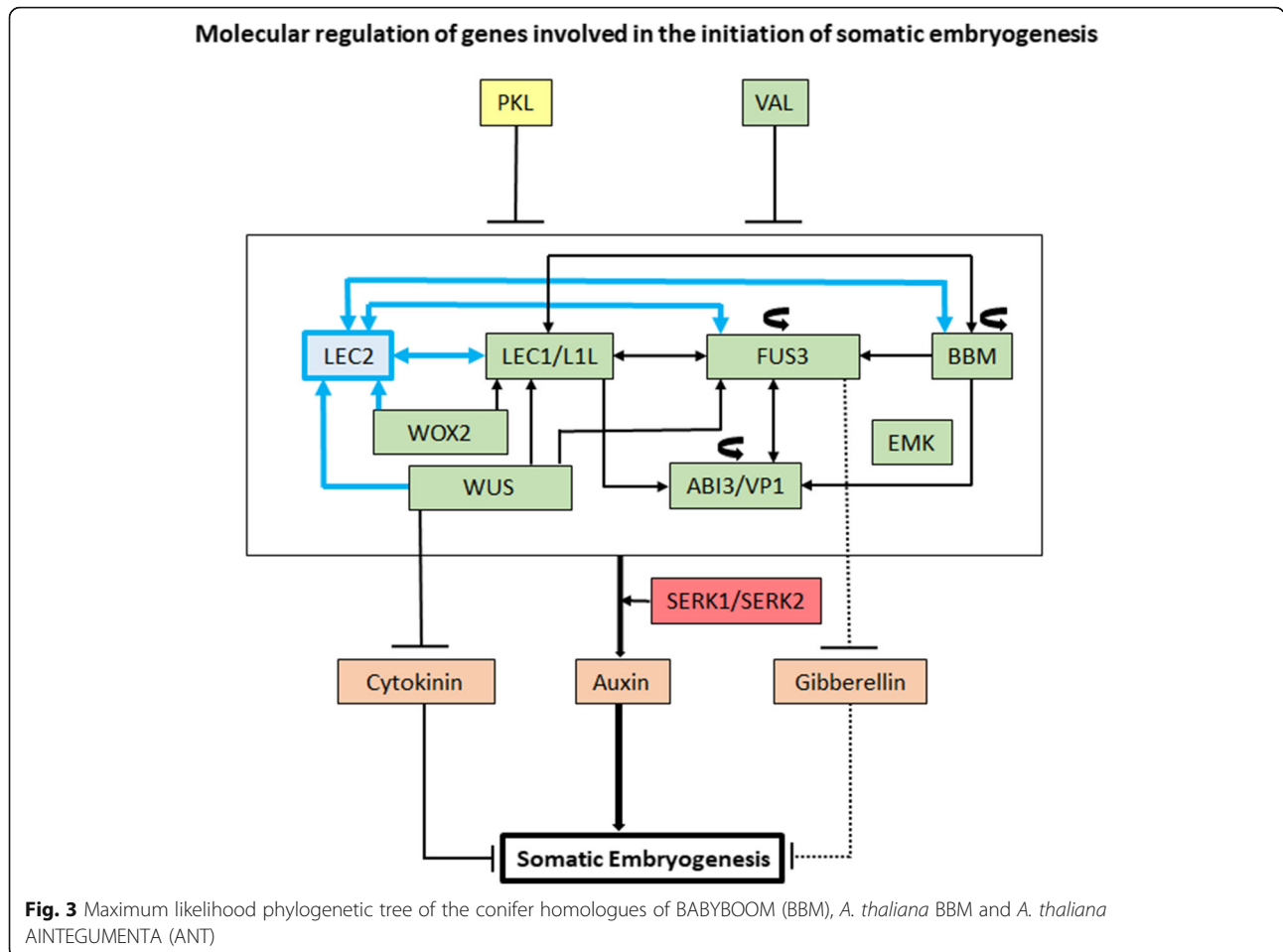
#### ***BABYBOOM (BBM)***

*BBM* is similar to *AINTEGUMENTA (ANT)*, but *BBM* possesses the characteristic conserved *BBM*-1 motif (GLSMIKTW); *ANT* lacks the *BBM*-1 or the *BBM*-1 like motif but contains SLSMSPGS motif [83] in *A. thaliana*. The significance of *BBM*-1 motif was demonstrated in *A. thaliana* where the plants overexpressing *BBM* gene with a mutated *BBM*-1 domain failed to produce somatic embryos on cotyledons as compared to the plants bearing the complete CDS of the *BBM* transgene [84]. Gene structure analysis of *LkBBM1* and *LkBBM2* in hybrid larch revealed that *LkBBM2* protein contained two AP2 DNA binding domains and a *BBM* specific motif as the *LkBBM1*, but lacked the euANT5 motif common to AP2 family members [42]. However, *LkBBM1* and *LkBBM2* showed similar behaviour with reference to regulation of adventitious root development. These

findings provide concrete evidence regarding the importance of *BBM* specific motif. *BBM* proteins of various plant species e.g. *A. thaliana* (NM\_121749, GenBank), *Brassica napus* (*BBM1*: AAM33802, *BBM2*: AAM33801), poplar (*BBM1*: XM\_002316143, *BBM2*: XM\_002311223, GenBank), hybrid larch (*BBM1*: AHH34920, *BBM2*: QEL52760, GenBank) and *L. decidua* contain the GLSMIKTW motif (AEF56566, GenBank). However, variations of the *BBM* specific motif occur in the maize (*Zea mays*) and rice (*Oryza sativa*) proteins. *Zea mays* *BBM* contains the ELSMIKTW motif (NP\_001147535, GenBank). In rice, three additional genes, Os02g0614300 (*OsBBM2*), Os01g0899800 (*OsBBM3*) and Os04g0504500 (*OsBBM4*) were referred to be homologous to *Oryza sativa* *BABY-BOOM LIKE 1* (*Os-BBML1*, Os11t0295900) [85] (<http://rapdb.dna.affrc.go.jp/>). *Os-BBML1* and the homologues contain the *BBM*-1 like motif; *Os-BBML1* and *OsBBM3* possess the GLSMIKNW motif and, *OsBBM2* and *OsBBM4* contain the ELSMIKTW motif. *OsBBM2* and *OsBBM3* function redundantly with *OsBBML1* [85]. Fern is a non-seed plant where the *BBM* gene is absent; it has the *ANT* gene which lacks the *BBM*-1 or *BBM*-1 like motif but

possesses the SLSMITGS motif at the same particular location similarly to the *A. thaliana* *ANT* (AT4G37750) protein. This *ANT* gene in fern functionally mimics the *BBM* gene promoting apogamy [86]; the expression pattern of fern *ANT* is similar to that of the *A. thaliana* *BBM* during early stages of embryo development [39, 86].

The *BBM*-1 motif and a *BBM*-1 like motif were detected in the current analysis in conifer proteins – the GLSMIKTW (*BBM*-1 motif) was found in *P. abies*, *P. taeda* and *P. sylvestris*, and the ELSDFKTW (*BBM*-1 like) motif was found in *Thuja koraiensis* (*T. koraiensis*) (Figure S21, Additional file 2.pdf). The phylogenetic tree constructed with the sequences of conifer homologues of *BBM* (Fig. 3) shows that the sequences from *P. menziesii* (PME00019482), *P. abies* (PAB00065438), *P. pinaster* (PPI00013750) and *P. thunbergii* (BAD16602.1) are closer to *A. thaliana* *ANT*. BAD16602.1 and PAB00065438 (MA\_98095g0010) are characterised as *AINTEGUMENTA*-like in the respective databases from where the sequences were obtained. Likewise, PPI00013750 is predicted as AP2-like ethylene-responsive transcription factor *ANT* in Gymno PLAZA,





whereas there is no annotation information available for PME00019482. The *BBM-1* motif was not detected in these four sequences; instead they show presence of the GLSALKTW motif. The GLSALKTW motif has higher similarity to the *BBM-1* motif (GLSMIKTW) than the motif found in the *ANT* gene (SLSMSPGS). The requirement of the *BBM-1* motif for the proper functioning of the *BBM* is demonstrated earlier [84]. We propose that the gene homologues found in the conifers, which show homology with the *ANT* or are characterised as *ANT*-like genes in either Congenie or Gymno PLAZA but contain the *BBM-1* like motif (GLSALKTW), may be the potential *BBM*-like genes. However, further analysis is required to confirm the functional conservation of these proteins in conifers and angiosperms.

#### **EMBRYOMAKER (EMK)**

The AP2 subfamily members that are involved in stress responses contain a single copy of the AP2 domain whereas two copies of the AP2 domains are present in the members which play a role in plant development [87]; *EMK* contains two AP2 domains [43]. In the current study, five conifer sequences were detected with the BLAST searches; three (*P. abies*, *P. sylvestris*, *P. pinaster*) showed two AP2 domains, which may be the precise putative homologues of the *EMK* genes, while two sequences were found with only one AP2 domain (*P. taeda*, *P. menziesii*). The two AP2 domains appear to be conserved among all the sequences in the alignment (Figure S6, Additional file 6.pdf).

#### **WUS and WUSCHEL-related homeobox (WOX) 2**

*WUS* homologue was reported by previous investigations related to the analysis of the *WOX* gene family in *P. abies* [51] and also in *P. pinaster* [7]. Although *WUS* and *WOX5* have similar domains (Homeodomain [HD], *WUS*-box [TL-X-L-F-P] and EAR domain [L-X-L-X-L]), *WUS* has an extra Y residue in the homeobox domain which is conserved in several plant species [48]. This conserved extra Y residue was earlier reported to be found in the HD of *WUS* in conifer species e.g. *P. abies* [51] and also in *P. pinaster* [7] (Figure S6, Additional file 7.pdf). Only one new *WUS* homologue was detected in the current analysis in Gymno PLAZA in *P. taeda*, which also possessed the extra Y residue (PTA00030527). This particular sequence is annotated as *WOX4* in the Gymno PLAZA, but since it has the highly conserved extra Y residue in HD, we propose that this is actually the *WUS* gene. *WOX2* contains the HD and *WUS*-box [48], which was found to be conserved in the conifer species included in this work similar to the earlier studies in conifers [7, 51] (Figure S17, Additional file 8.pdf).

#### **Somatic embryogenesis receptor kinases (SERK)**

Most of the conifer sequences retrieved with the BLASTP searches with *A. thaliana SERK1* were categorised as *SERK1* by the respective databases, however a few sequences were referred to as *SERK1-like* or *SERK2* (Additional file 9.pdf). We have included all these sequences in our analysis because *SERK1* and *SERK2* share 90% identity [88] and these two genes function redundantly while playing a major role in somatic and reproductive cell differentiation as reported during early anther development in *A. thaliana* [89]. The different domains of *SERK1-like* or *SERK2* were found to be well conserved in all the conifer homologues (Figure S31, Additional file 9.pdf). The *SERK1-like* or *SERK2* conifer homologues contained signal peptide domain, Leucine zipper domain with four conserved Leucine residues, five Leucine rich repeats, the Serine-Proline-Proline (SPP) domain with conserved SPP motifs, the transmembrane domain, the 11 subdomains of the protein kinase domain and the C-terminal domain [90]. Two pairs of cysteine residues were present in the Leucine zipper and SPP domain of the conifer homologues respectively, which were reported to be conserved [91]. The Arginine and Aspartate residues of the subdomain VI of the protein kinase domain, were found to be conserved in all the conifer homologues that contained this domain [90].

#### **Conifer homologues of genes involved in SE suppression Pickle (PKL)**

*PKL* acts as a repressor not only for the expression of embryonic traits but also represses the seedling de-etiolation pathway; *PKL* acts additively with SUPPRESSOR OF PHYTOCHROME A1 (SPA1) to repress seedling de-etiolation and inhibits the protein and transcript levels of ELONGATED HYPOCOTYL 5 (HY5) which is one of the important transcription factors that positively regulates the process of photomorphogenesis [92]. *PKL* physically interacts with HY5 and also with HY5-HOMOLOG (HYH), the close homolog of HY5 to regulate the hypocotyl cell elongation in *A. thaliana* and interestingly, the ATPase domain of *PKL* is essential and sufficient for the interaction with both HY5 and HYH [93]. However, a point mutation (Lysine to Alanine) at the position Lys-304 in *PKL* terminates this interaction [93]. Lys-304 in *A. thaliana PKL*, is an evolutionarily conserved amino acid that is predicted to bind to ATP within the ATPase domain of *PKL*. This amino acid was found to be conserved in all the *PKL* sequences of the conifer species where the ATPase domain was detected, which includes *P. abies*, *P. taeda*, *P. sylvestris*, *P. pinaster* and *P. menziesii* (Figure S19, Additional file 10.pdf). Only two *PKL* sequences, one from *P. menziesii* and one from *P. taeda* were found to possess all the known domains of *PKL*. The detection of partial *PKL* sequences in

the other conifers maybe because of either lack of data availability due to sequencing quality and/or poor annotation. Yet, it could also be argued that *PKL* in conifers with only some specific domains acts in a different fashion from what is known in the more advanced angiosperm species, as conifers are known to possess certain specialized pathways compared to the angiosperms e.g. specialization of photosynthetic apparatus in *P. menziesii* [76]; however, further detailed molecular studies are required to confirm this.

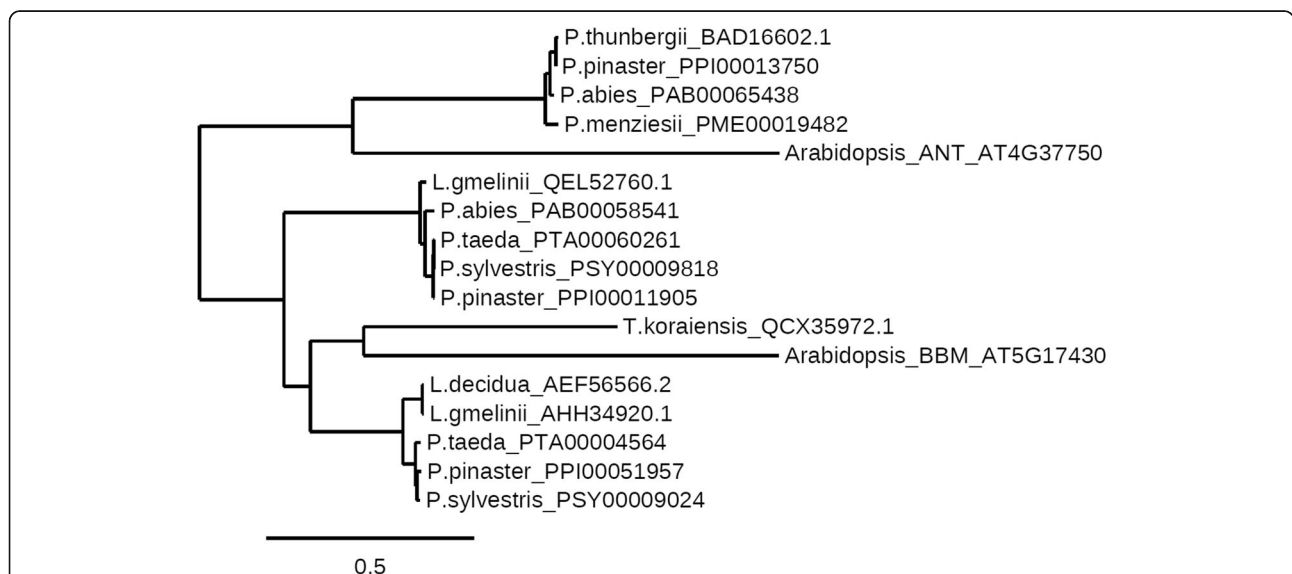
**VP1/ABI3-like (VAL)**

BLASTP to GenBank with *A. thaliana* VAL1/VAL2/VAL3 did not give significant matches in Pinidae, while Gymno PLAZA resulted in similar hits with *A. thaliana* VAL1/VAL2/VAL3 in all the conifer species included in this analysis (with the available sequence data), except in case of *P. menziesii* where BLASTP searches with VAL1/VAL2 resulted in similar hits but searches with VAL3 gave different matches. There were no significant matches found for *A. thaliana* VAL1/VAL2/VAL3 in *Picea sitchensis* (Sitka spruce, *P. sitchensis*). *A. thaliana* VAL1 shares 47% identity with *A. thaliana* VAL2 and 34% identity with *A. thaliana* VAL3, while *A. thaliana* VAL2 shares 44% identity with *A. thaliana* VAL3 as

observed from the BLASTP (Additional file 11.pdf). Since similar hits were detected in all the conifer species (included in the analysis), we considered all the sequences together to make the alignments and marked the different domains of the conifer VAL homologues, which were found to be conserved within the conifer species included in the analysis (Figure S67, Additional file 11.pdf). The B3 domain, in particular was found to be highly conserved.

**Molecular regulation of genes involved in the initiation of somatic embryogenesis**

A schematic model for the mechanism of regulation of initiation of somatic embryogenesis in plants with reference to the key genes involved in the process is summarised in Fig. 4 [13, 16, 18, 94, 95]. Homologues for all the candidate initiation genes except for *LEC2* were detected in the conifers. The knowledge in conifers with reference to the initiation of SE is limited to the information regarding the expression patterns of the genes involved and there is lack of evidence for regulation of the process through a gene network. We propose a putative alternative mechanism of the molecular regulation of the process of SE initiation, which may be functional in conifers in absence of *LEC2* (Fig. 4) assuming that the



**Fig. 4** Molecular regulatory network of genes involved in the initiation of somatic embryogenesis: Gene indicated in yellow is a chromatin remodeling factor; *PICKLE* (*PKL*). Gene indicated in blue, is a transcription factor (TF) *LEAFY COTYLEDON 2* (*LEC2*); *LEC2* is absent in conifers. Genes indicated in green are TFs - *LEAFY COTYLEDON 1* (*LEC1*), *LEAFY COTYLEDON 1 LIKE* (*L1L*), *FUSCA3* (*FUS3*), *BABYBOOM* (*BBM*), *EMBRYOMAKER* (*EMK*), *ABSCISIC ACID INSENSITIVE 3* (*ABI3*) or *VIVIPAROUS 1* (*VP1*), *WUSCHEL* (*WUS*), *WUSCHEL-related homeobox 2* (*WOX2*). Curved arrows for *FUS3*, *BBM* and *ABI3/VP1* indicate that these genes regulate themselves through feedback loops. Hormones involved in the process are indicated in orange (Cytokinin, Auxin and Gibberellin). Genes in red are *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE 1* and 2 (*SERK1/SERK2*). Lines ending with arrow indicate transcriptional regulation and lines ending with bars indicate transcriptional repression. Solid lines indicate transcriptional regulation by molecular evidence and dotted lines indicate molecular mechanisms that are not clear. Blue lines indicate the regulation that is absent in conifers because of the absence of *LEC2*. The regulation represented here is summarized from the investigations done in angiosperms. In conifers, only the information regarding expression data of the genes with reference to initiation of SE has been reported that includes the genes - *LEC1*, *FUS3*, *BBM*, *WUS*, *WOX2*, *ABI3/VP1* and *SERK1/SERK2*

overall functions of the other genes involved are conserved in conifers. *LEC2* is one of the central players in the process of seed and embryo development in plants [18, 19, 22]. In *A. thaliana*, although *LEC2* regulates SE through stimulation of auxin synthesis [96], one of the major roles of *LEC2* is to upregulate *FUS3* and *ABI3*; however, *ABI3* and *FUS3* positively regulate themselves and each other to achieve a uniform expression in the embryo through the feedback loops [97]. *LEC1* has also been shown to positively regulate *ABI3* and *FUS3* expression [97, 98]. Although the expression levels of *ABI3* and *FUS3* were lowered in *LEC2* mutants, constitutive expression of *ABI3* or *FUS3* was able to rescue the *LEC2* phenotypes in *A. thaliana* [97]. Further, *ABI3*, *LEC2*, and *FUS3* were proposed to work in parallel pathways and also, *FUS3* and *LEC2* were shown to act in a partially redundant manner [99]. From this context, the action of *ABI3/FUS3* or both may compensate the absence of *LEC2* in conifers. Similar to *LEC2*, *LEC1* mediates not only the up-regulation of the auxin synthesis [100] but also facilitates effects of auxin to promote embryonic cell identity [101]. Although *BBM* and *LEC2* regulate each other through a feedback loop, *LEC1* and *BBM* also regulate each other in a similar way [95]. Moreover, *BBM* also stimulates its own expression through a positive feedback loop to control its own activity [102]. With these assumptions, we propose that in conifers, *LEC1* (possibility along with *LEC1-LIKE*) regulates the network in order to make up for the loss of *LEC2*. To summarise, SE regulation in conifers may include action of *ABI3/FUS3* or both to compensate the absence of *LEC2*, and the conifer *LEC1* along with *LEC1-LIKE* might be capable of performing adequate functions that are carried out by *LEC2*. However, further molecular work is required to confirm such associated alternative pathways in the conifer species. In this context, it is worth mentioning again that conifers are known to follow alternative pathways e.g. networks associated with photosynthesis [76] and proposed molecular mechanisms involved with etiolation/de-etiolation [103].

## Conclusions

This *in silico* analysis suggests absence of *LEC2* in *P. abies* and *P. taeda*, the conifer species whose genomes have been sequenced. Based on available sequence data to date, *LEC2* was also not detected in the other conifer species included in the study. The presence of a haploid endosperm due to the absence of a double fertilization event and presence of multiple cotyledons in conifers, could be associated with the lack of a master embryogenesis regulator such as *LEC2* gene from the conifers. Based on existing expression data, SE regulation in conifers may include action of *ABI3/FUS3* or both to compensate for the absence of *LEC2*, and the conifer *LEC1*

along with *LEC1-LIKE* might be capable of performing adequate functions that are otherwise carried out by *LEC2*. However, further molecular analyses are required to confirm such associated alternative pathways in conifers. Furthermore, conifers exhibit characteristic mechanisms with reference to somatic embryo development such as the presence of cleavage polyembryony in *Pinus* but broadly not in *Picea*. Analyses of PEMs from more or less polyembryogenic species with respect to known transcription factors involved in somatic embryo regulation in model species could offer insights to regulatory processes active during conifer embryo development. The current work presents fundamental information to support applied studies into underlying molecular mechanisms of conifer somatic embryo initiation and development.

## Methods

In this article, we have identified the conifer homologues potentially involved in the initiation of SE. Protein sequences of the candidate genes from *A. thaliana* were used to query the databases for finding the conifer homologues, using standard protein BLAST (BLASTP) (Basic Local Alignment Search Tool). *A. thaliana* was chosen as the reference species to detect the conifer homologues as this is the most widely used and the most well-documented model plant species. Databases included in the searches were Gymno PLAZA, 1.0 (<https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/>) [104], Congenie (<http://congenie.org/>, v1.0) [105, 106] and GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) [107]. BLASTP searches in GenBank were executed by selecting the non-redundant protein sequence database along with selection of Subclass Pinidae (taxid:3313) under the organism option for performing conifer specific searches. Congenie and Gymno PLAZA are platforms for plant comparative genomics; these databases perform the homology searches using BLAST and include the information regarding the best homologues (e.g. best *A. thaliana* homologue) in the results. Congenie is integrated with gene prediction software e.g. AUGUSTUS and EuGene, which identifies a gene and, it provides the gene description based on Blast2GO, the functional characterization of the gene and the best BLAST homologues. Gymno PLAZA provides the structural and functional annotation of a particular gene, the associated gene family data and phylogenetic trees. However, identity of the particular conifer gene was further confirmed with domain search, alignment and phylogenetic analysis. Specific domains of the particular conifer homologue were identified by performing the search with the Conserved Domain Database (CDD, <https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) [108]. In addition to CDD search, domains of the particular

**Table 1** Gene-wise references used for detecting the different domains in the respective genes involved in initiation of somatic embryogenesis in conifers

Genes involved in initiation of somatic embryogenesis	References
<i>BABYBOOM</i> ( <i>BBM</i> )	[41, 42, 83, 84, 109]
<i>LEAFY COTYLEDON</i> ( <i>LEC</i> )	[77–80, 110]
<i>FUSCA3</i> ( <i>FUS3</i> )	[72, 82]
<i>ABSCISIC ACID INSENSITIVE 3</i> ( <i>ABI3</i> ) or <i>VIVIPAROUS</i> ( <i>VP1</i> )	[35, 111, 112]
<i>EMBRYOMAKER</i> ( <i>EMK</i> )	[43]
<i>WUSCHEL</i> ( <i>WUS</i> ) and <i>WUSCHEL-related homeobox</i> ( <i>WOX</i> ) 2	[7, 8, 48, 51]
<i>Somatic embryogenesis receptor kinases</i> ( <i>SERK</i> )	[90, 91, 113]
<b>Genes involved in suppression somatic embryogenesis</b>	
<i>PICKLE</i> ( <i>PKL</i> )	[58, 108]
<i>VP1/ABI3-LIKE</i> ( <i>VAL</i> )	[59, 60, 65]

conifer gene were also confirmed by referring to the gene specific sequence information available from the literature (Table 1). Furthermore, particular conifer protein sequence of a gene was aligned with the protein sequence of the respective *A. thaliana* gene using MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle/>) [114]. MUSCLE was selected for making the alignments as it uses both global and local alignment algorithms as compared to ClustalW, which uses only global alignment that creates more gaps. Only in case of *WOX2*, Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) [115] was used which resulted into better alignment related to the WUS box. Phylogenetic trees of protein sequences were constructed for further validation, wherever required, using *Phylogeny.fr* in the ‘one click mode’ using default settings (<https://www.phylogeny.fr/>) [116]. In brief, the alignment was done with MUSCLE [114], phylogeny was done using PhyML [117] which is based on the maximum-likelihood principle and the phylogenetic tree was prepared using TreeDyn [118].

Congenie and Gymno PLAZA include the whole genome sequence data from the two representative species from genus *Picea* (*P. abies*, v1.0) [105] and *Pinus* (*P. taeda*, v1.0) [119] from the Pinaceae family. Other conifers species included in the current in silico analysis were *P. abies*, *P. glauca*, *P. sitchensis*, *P. taeda*, *P. sylvestris*, *P. pinaster*, *P. contorta*, *Pinus massoniana* (*P. massoniana*, Chinese red pine), *P. menziesii*, *A. angustifolia*, *C. lanceolata*, *Thuja koraiensis* (*T. koraiensis*, Korean arborvitae), *L. decidua* and *Larix gmelinii* var. *olgensis* x *Larix kaempferi* (Hybrid larch, *L. gmelinii*). There were no specific criteria applied for the choice of a particular conifer species included in this analysis, the availability of the data was the prime factor; therefore, all the relevant sequences obtained in the BLASTP results were included in the current analysis.

#### Abbreviations

ABI3: ABSCISIC ACID INSENSITIVE 3; BBM: BABYBOOM; EMK: EMBRYOMAKER; FUS3: FUSCA3; LEC: LEAFY COTYLEDON; PKL: PICKLE; PEM: Pro-embryogenic masses; SE: Somatic embryogenesis; SERK: Somatic embryogenesis receptor kinases; VP1: VIVIPAROUS; VAL: VP1/ABI3-LIKE; WUS: WUSCHEL; WOX: WUSCHEL-related homeobox

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07718-8>.

**Additional file 1.** Details of conifer homologues involved in somatic embryogenesis – Table S1-Table S10.

**Additional file 2.** Alignments of BBM gene.

**Additional file 3.** Alignments of LEC gene.

**Additional file 4.** Alignments of FUS3 gene.

**Additional file 5.** Alignments of VP1 gene.

**Additional file 6.** Alignments of EMK gene.

**Additional file 7.** Alignments of WUS gene.

**Additional file 8.** Alignments of WOX2 gene.

**Additional file 9.** Alignments of SERK gene.

**Additional file 10.** Alignments of PKL gene.

**Additional file 11.** Alignments of VAL gene.

#### Acknowledgements

The authors acknowledge the funding from MULTIFOREVER granted to UE.

#### Authors' contributions

SSR was involved in conceptualization, methodology, analysis, representation of the results and manuscript writing – original draft preparation and review & editing. UE was involved in conceptualization, manuscript writing – review & editing, resources and funding acquisition. Both authors have read and approved the manuscript.

#### Funding

This work was supported by MULTIFOREVER granted to UE; Project MULTIFOREVER is supported under the umbrella of ERA-NET Cofund Forest-Value by ANR (FR), FNR (DE), MINCYT (AR), MINECO-AEI (ES), MMM (FI) and VINNOVA (SE). ForestValue has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°773324. Open Access funding provided by Swedish University of Agricultural Sciences.



**Availability of data and materials**

All data and materials with reference to this work are contained within the article or supplementary material.

**Declarations****Ethics approval and consent to participate**

Not applicable for the study.

**Consent for publication**

The authors confirm their consent for publication of their work.

**Competing interests**

The authors declare that they do not have any competing interests.

Received: 5 February 2021 Accepted: 12 May 2021

Published online: 26 May 2021

**References**

- von Arnold S, Clapham D, Abrahamsson M. Embryology in conifers. *Mol Physiol Biotechnol Trees*. 2019;89:157–84. <https://doi.org/10.1016/bs.abr.2018.11.005>.
- Egertsdotter U, Ahmad I, Clapham D. Automation and scale up of somatic embryogenesis for commercial plant production, with emphasis on conifers. *Front Plant Sci*. 2019;10. <https://doi.org/10.3389/fpls.2019.00109>.
- Rosvall O. Using Norway spruce clones in Swedish forestry: general overview and concepts. *Scand J For Res*. 2019;34(5):336–41. <https://doi.org/10.1080/02827581.2019.1614659>.
- Hakman I, Fowke LC. Somatic embryogenesis in *Picea glauca* (white spruce) and *Picea mariana* (black spruce). *Can J Botany*. 1987;65(4):656–9.
- Chalupa V. Somatic embryogenesis and plantlet regeneration from cultured immature and mature embryos of *Picea abies* (L.) karst. *Commun Institut Forest Czechosloveniae*. 1985;14:57–63.
- Salaj T, Matusova R, Salaj J. Conifer somatic embryogenesis - an efficient plant regeneration system for theoretical studies and mass propagation. *Dendrobiology*. 2015;74:69–76. <https://doi.org/10.12657/denbio.074.007>.
- Alvarez JM, Bueno N, Canas RA, Avila C, Canovas FM, Ordas RJ. Analysis of the WUSCHEL-RELATED HOMEODOMAIN gene family in *Pinus pinaster*: new insights into the gene family evolution. *Plant Physiol Biochem*. 2018;123:304–18. <https://doi.org/10.1016/j.plaphy.2017.12.031>.
- Palovaara J, Hakman I. Conifer WOX-related homeodomain transcription factors, developmental consideration and expression dynamic of WOX2 during *Picea abies* somatic embryogenesis. *Plant Mol Biol*. 2008;66(5):533–49. <https://doi.org/10.1007/s11103-008-9289-5>.
- Trontin JF, Klimaszewska K, Morel A, Hargreaves C, Lelu-Walter MA. Molecular aspects of conifer zygotic and somatic embryo development: a review of genome-wide approaches and recent insights. *Methods Mol Biol*. 2016;1359:167–207. [https://doi.org/10.1007/978-1-4939-3061-6\\_8](https://doi.org/10.1007/978-1-4939-3061-6_8).
- von Arnold S, Egertsdotter U, Ekberg I, Gupta P, Mo H, Nörsgaard J. Somatic Embryogenesis in Norway Spruce (*Picea abies*). In: Jain SM, Gupta PK, Newton RJ, editors. *Somatic Embryogenesis in Woody Plants Forestry Sciences*. Dordrecht: Springer; 1995. p. 44–6.
- Varis S, Klimaszewska K, Aronen T. Somatic embryogenesis and plant regeneration from primordial shoot explants of *Picea abies* (L.) H. Karst. *Somatic Trees*. *Front Plant Sci*. 2018;9:1551.
- Klimaszewska K, Overton C, Stewart D, Rutledge RG. Initiation of somatic embryos and regeneration of plants from primordial shoots of 10-year-old somatic white spruce and expression profiles of 11 genes followed during the tissue culture process. *Planta*. 2011;233(3):635–47. <https://doi.org/10.1007/s00425-010-1325-4>.
- Ikeuchi M, Favero D, Sakamoto Y, Iwase A, Coleman D, Rymen B, et al. Molecular mechanisms of plant regeneration. *Ann Rev Plant Biol*. 2019;70:377–406.
- Diaz-Sala C. Molecular dissection of the regenerative capacity of forest tree species: special focus on conifers. *Front Plant Sci*. 2019;9. <https://doi.org/10.3389/fpls.2018.01943>.
- Wojcik AM, Wojcikowska B, Gaj MD. Current perspectives on the auxin-mediated genetic network that controls the induction of somatic embryogenesis in plants. *Int J Mol Sci*. 2020;21(4):1333.
- Horstman A, Willemsen V, Boutilier K, Heidstra R. AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends Plant Sci*. 2014;19(3):146–57. <https://doi.org/10.1016/j.tplants.2013.10.010>.
- Han JD, Li X, Jiang CK, Wong GKS, Rothfels CJ, Rao GY. Evolutionary analysis of the LAFL genes involved in the land plant seed maturation program. *Front Plant Sci*. 2017;8. <https://doi.org/10.3389/fpls.2017.00439>.
- Kumar V, Jha P, Van Staden J. LEAFY COTYLEDONS (LECs): master regulators in plant embryo development. *Plant Cell Tiss Org*. 2020;140(3):475–87. <https://doi.org/10.1007/s11240-019-01752-x>.
- Braybrook SA, Harada JJ. LECs go crazy in embryo development. *Trends Plant Sci*. 2008;13(12):624–30. <https://doi.org/10.1016/j.tplants.2008.09.008>.
- Chiappetta A, Fambrini M, Petrarulo M, Rapparini F, Michelotti V, Bruno L, et al. Ectopic expression of LEAFY COTYLEDON1-LIKE gene and localized auxin accumulation mark embryogenic competence in epiphyllous plants of *Helianthus annuus* x *H-tuberosus*. *Ann Bot*. 2009;103(5):735–47. <https://doi.org/10.1093/aob/mcn266>.
- Lotan T, Ohto M, Yee KM, West MAL, Lo R, Kwong RW, et al. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell*. 1998;93(7):1195–205. [https://doi.org/10.1016/S0092-8674\(00\)81463-4](https://doi.org/10.1016/S0092-8674(00)81463-4).
- Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, et al. LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *P Natl Acad Sci USA*. 2001;98(20):11806–11. <https://doi.org/10.1073/pnas.201413498>.
- Park SY, Klimaszewska K, Park JY, Mansfield SD. Lodgepole pine: the first evidence of seed-based somatic embryogenesis and the expression of embryogenesis marker genes in shoot bud cultures of adult trees. *Tree Physiol*. 2010;30(11):1469–78. <https://doi.org/10.1093/treephys/tpq081>.
- Uddenberg D, Abrahamsson M, von Arnold S. Overexpression of PaHAP3A stimulates differentiation of ectopic embryos from maturing somatic embryos of Norway spruce. *Tree Genet Genomes*. 2016;12(2):18.
- Uddenberg D, Valladares S, Abrahamsson M, Sundstrom JF, Sundas-Larsson A, von Arnold S. Embryogenic potential and expression of embryogenesis-related genes in conifers are affected by treatment with a histone deacetylase inhibitor. *Planta*. 2011;234(3):527–39. <https://doi.org/10.1007/s00425-011-1418-8>.
- Arrillaga I, Morcillo M, Zanon I, Lario F, Segura J, Sales E. New approaches to optimize somatic embryogenesis in maritime pine. *Front Plant Sci*. 2019;10. <https://doi.org/10.3389/fpls.2019.00138>.
- Schlögl PS, dos Santos ALW, Vieira LDN, Floh EIS, Guerra MP. Gene expression during early somatic embryogenesis in Brazilian pine (*Araucaria angustifolia* (Bert.) O. Ktze). *Plant Cell Tiss Org*. 2012;108(1):173–80. <https://doi.org/10.1007/s11240-011-0023-7>.
- Harada JJ. Role of Arabidopsis LEAFY COTYLEDON genes in seed development. *J Plant Physiol*. 2001;158(4):405–9. <https://doi.org/10.1078/0176-1617-00351>.
- Suzuki M, Ketterling MG, Li QB, McCarty DR. Viviparous1 alters global gene expression patterns through regulation of abscisic acid signaling. *Plant Physiol*. 2003;132(3):1664–77. <https://doi.org/10.1104/pp.103.022475>.
- Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P. The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid. *Dev Cell*. 2004;7(3):373–85. <https://doi.org/10.1016/j.devcel.2004.06.017>.
- Parcy F, Valon C, Raynal M, Gaubiercomella P, Delseny M, Giraudat J. Regulation of gene-expression programs during Arabidopsis seed development - roles of the ABI3 locus and of endogenous abscisic-acid. *Plant Cell*. 1994;6(11):1567–82. <https://doi.org/10.1105/tpc.6.11.1567>.
- Stasolla C, van Zyl L, Egertsdotter U, Craig D, Liu WB, Sederoff RR. The effects of polyethylene glycol on gene expression of developing white spruce somatic embryos. *Plant Physiol*. 2003;131(1):49–60. <https://doi.org/10.1104/pp.015214>.
- Fischerova L, Fischer L, Vondrakova Z, Vagner M. Expression of the gene encoding transcription factor PaVP1 differs in *Picea abies* embryogenic lines depending on their ability to develop somatic embryos. *Plant Cell Rep*. 2008;27(3):435–41. <https://doi.org/10.1007/s00299-007-0469-6>.
- Merino I, Abrahamsson M, Sterck L, Craven-Bartle B, Canovas F, von Arnold S. Transcript profiling for early stages during embryo development in Scots pine. *BMC Plant Biol*. 2016;16(1):255. <https://doi.org/10.1186/s12870-016-0939-5>.
- Footitt S, Ingouff M, Clapham D, von Arnold S. Expression of the viviparous 1 (Pavp1) and p34(cdc2) protein kinase (cdc2Pa) genes during somatic embryogenesis in Norway spruce (*Picea abies* [L.] karst). *J Exp Bot*. 2003;54(388):1711–9. <https://doi.org/10.1093/jxb/erg178>.

36. Horstman A, Bemer M, Boutilier K. A transcriptional view on somatic embryogenesis. *Regeneration*. 2017;4(4):201–16. <https://doi.org/10.1002/reg2.91>.
37. Jha P, Kumar V. BABY BOOM (BBM): a candidate transcription factor gene in plant biotechnology. *Biotechnol Lett*. 2018;40(11–12):1467–75. <https://doi.org/10.1007/s10529-018-2613-5>.
38. Horstman A, Li MF, Heidmann I, Weemen M, Chen BJ, Muino JM, et al. The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis. *Plant Physiol*. 2017;175(2):848–57. <https://doi.org/10.1104/pp.17.00232>.
39. Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang LM, et al. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell*. 2002;14(8):1737–49. <https://doi.org/10.1105/tpc.001941>.
40. Rupps A, Raschke J, Rummeler M, Linke B, Zoglauer K. Identification of putative homologs of Larix decidua to BABYBOOM (BBM), LEAFY COTYLEDON1 (LEC1), WUSCHEL-related HOMEBOX2 (WOX2) and SOMATIC EMBRYOGENESIS RECEPTOR-like KINASE (SERK) during somatic embryogenesis. *Planta*. 2016;243(2):473–88. <https://doi.org/10.1007/s00425-015-2409-y>.
41. Li KP, Sun XM, Han H, Zhang SG. Isolation, characterization and expression analysis of the BABY BOOM (BBM) gene from Larix kaempferi x L. olgensis during adventitious rooting. *Gene*. 2014;551(2):111–8. <https://doi.org/10.1016/j.gene.2014.08.023>.
42. Wang HM, Li KP, Sun XM, Xie YH, Han XM, Zhang SG. Isolation and characterization of larch BABY BOOM2 and its regulation of adventitious root development. *Gene*. 2019;690:90–8. <https://doi.org/10.1016/j.gene.2018.12.049>.
43. Tsuwamoto R, Yokoi S, Takahata Y. Arabidopsis EMBRYOMAKER encoding an AP2 domain transcription factor plays a key role in developmental change from vegetative to embryonic phase. *Plant Mol Biol*. 2010;73(4–5):481–92. <https://doi.org/10.1007/s11103-010-9634-3>.
44. Buchholz J. Embryo development and polyembryony in relation to the phylogeny of conifers. *Am J Bot*. 1920;7(4):125–45. <https://doi.org/10.1002/j.1537-2197.1920.tb05570.x>.
45. Filonova LH, von Arnold S, Daniel G, Bozhkov PV. Programmed cell death eliminates all but one embryo in a polyembryonic plant seed. *Cell Death Differ*. 2002;9(10):1057–62. <https://doi.org/10.1038/sj.cdd.4401068>.
46. Zuo JR, Niu QW, Frugis G, Chua NH. The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant J*. 2002;30(3):349–59. <https://doi.org/10.1046/j.1365-3113X.2002.01289.x>.
47. Klimaszewska K, Pelletier G, Overton C, Stewart D, Rutledge RG. Hormonally regulated overexpression of Arabidopsis WUS and conifer LEC1 (CHAP3A) in transgenic white spruce: implications for somatic embryo development and somatic seedling growth. *Plant Cell Rep*. 2010;29(7):723–34. <https://doi.org/10.1007/s00299-010-0859-z>.
48. Haecker A, Gross-Hardt R, Geiges B, Sarkar A, Breuning H, Herrmann M, et al. Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. *Development*. 2004;131(3):657–68. <https://doi.org/10.1242/dev.00963>.
49. Zhu T, Moschou PN, Alvarez JM, Sohlberg JJ, von Arnold S. WUSCHEL-RELATED HOMEBOX 2 is important for protoderm and suspensor development in the gymnosperm Norway spruce. *BMC Plant Biol*. 2016;16(1):19. <https://doi.org/10.1186/s12870-016-0706-7>.
50. Palovaara J, Hallberg H, Stasolla C, Hakman I. Comparative expression pattern analysis of WUSCHEL-related homeobox 2 (WOX2) and WOX8/9 in developing seeds and somatic embryos of the gymnosperm Picea abies. *New Phytol*. 2010;188(1):122–35. <https://doi.org/10.1111/j.1469-8137.2010.03336.x>.
51. Hedman H, Zhu TQ, von Arnold S, Sohlberg JJ. Analysis of the WUSCHEL-RELATED HOMEBOX gene family in the conifer Picea abies reveals extensive conservation as well as dynamic patterns. *BMC Plant Biol*. 2013;13(1):89. <https://doi.org/10.1186/1471-2229-13-89>.
52. Zhou X, Zheng R, Liu G, Xu Y, Zhou Y, Laux T, et al. Desiccation treatment and endogenous IAA levels are key factors influencing high frequency somatic embryogenesis in *Cunninghamia lanceolata* (Lamb.) Hook. *Front Plant Sci*. 2017;8:2054. <https://doi.org/10.3389/fpls.2017.02054>.
53. Kumar V, Van Staden J. Multi-tasking of SERK-like kinases in plant embryogenesis, growth, and development: current advances and biotechnological applications. *Acta Physiol Plant*. 2019;41(3):1–6.
54. Hecht V, Vielle-Calzada JP, Hartog MV, Schmidt EDL, Boutilier K, Grossniklaus U, et al. The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiol*. 2001;127(3):803–16. <https://doi.org/10.1104/pp.010324>.
55. Steiner N, Santa-Catarina C, Guerra M, Cutri L, Dornelas M, Floh E. A gymnosperm homolog of SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE 1 (SERK1) is expressed during somatic embryogenesis. *Plant Cell Tissue Organ Cult*. 2012;109(1):41–50. <https://doi.org/10.1007/s11240-011-0071-z>.
56. Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*. 2017;18(7):407–22. <https://doi.org/10.1038/nrm.2017.26>.
57. Henderson JT, Li HC, Rider SD, Mordhorst AP, Romero-Severson J, Cheng JC, et al. PICKLE acts throughout the plant to repress expression of embryonic traits and may play a role in gibberellin-dependent responses. *Plant Physiol*. 2004;134(3):995–1005. <https://doi.org/10.1104/pp.103.030148>.
58. Ogas J, Kaufmann S, Henderson J, Somerville C. PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. *P Natl Acad Sci USA*. 1999;96(24):13839–44. <https://doi.org/10.1073/pnas.96.24.13839>.
59. Jia H, Suzuki M, McCarty DR. Regulation of the seed to seedling developmental phase transition by the LAF1 and VAL transcription factor networks. *Wiley Interdiscip Rev Dev Biol*. 2014;3(1):135–45. <https://doi.org/10.1002/wdev.126>.
60. Tsukagoshi H, Saijo T, Shibata D, Morikami A, Nakamura K. Analysis of a sugar response mutant of Arabidopsis identified a novel B3 domain protein that functions as an active transcriptional repressor. *Plant Physiol*. 2005;138(2):675–85. <https://doi.org/10.1104/pp.104.057752>.
61. Sasnauskas G, Kauneckaitė K, Siksnys V. Structural basis of DNA target recognition by the B3 domain of Arabidopsis epigenome reader VAL1. *Nucleic Acids Res*. 2018;46(8):4316–24. <https://doi.org/10.1093/nar/gky256>.
62. Kagale S, Rozwadowski K. EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. *Epigenetics*. 2011;6(2):141–6. <https://doi.org/10.4161/epi.6.2.13627>.
63. Guan Y, Li SG, Fan XF, Su ZH. Application of somatic embryogenesis in woody plants. *Front Plant Sci*. 2016;7. <https://doi.org/10.3389/fpls.2016.00938>.
64. Jia H, McCarty DR, Suzuki M. Distinct roles of LAF1 network genes in promoting the embryonic seedling fate in the absence of VAL repression. *Plant Physiol*. 2013;163(3):1293–305. <https://doi.org/10.1104/pp.113.220988>.
65. Suzuki M, Wang HHY, McCarty DR. Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol*. 2007;143(2):902–11. <https://doi.org/10.1104/pp.106.092320>.
66. Howe GT, Bucciaglia PA, Hackett WP, Furnier GR, Cordonnier-Pratt MM, Gardner G. Evidence that the phytochrome gene family in black cottonwood has one PHYA locus and two PHYB loci but lacks members of the PHYC/F and PHYE subfamilies. *Mol Biol Evol*. 1998;15(2):160–75. <https://doi.org/10.1093/oxfordjournals.molbev.a025912>.
67. Merino I, Abrahamsson M, Larsson E, von Arnold S. Identification of molecular processes that differ among Scots pine somatic embryogenic cell lines leading to the development of normal or abnormal cotyledonary embryos. *Tree Genet Genomes*. 2018;14(2):1–7.
68. Dobrowolska I, Businge E, Abreu IN, Moritz Z, Egertsdotter U. Metabolome and transcriptome profiling reveal new insights into somatic embryo germination in Norway spruce (*Picea abies*). *Tree Physiol*. 2017;37(12):1752–66. <https://doi.org/10.1093/treephys/tpx078>.
69. Elbl P, Campos RA, Lira BS, Andrade SCS, Jo L, dos Santos ALW, et al. Comparative transcriptome analysis of early somatic embryo formation and seed development in Brazilian pine, *Araucaria angustifolia* (Bertol.) Kuntze. *Plant Cell Tiss Org*. 2015;120(3):917. <https://doi.org/10.1007/s11240-015-0730-6>.
70. Hofmann F, Schon MA, Nodine MD. The embryonic transcriptome of Arabidopsis thaliana. *Plant Reprod*. 2019;32(1):77–91. <https://doi.org/10.1007/s00497-018-00357-2>.
71. Wickramasuriya AM, Dunwell JM. Global scale transcriptome analysis of Arabidopsis embryogenesis in vitro. *BMC Genomics*. 2015;16(1):301. <https://doi.org/10.1186/s12864-015-1504-6>.
72. Li Y, Jin K, Zhu Z, Yang J. Stepwise origin and functional diversification of the AFL subfamily B3 genes during land plant evolution. *J Bioinf Comput Biol*. 2010;8(suppl1):33–45. <https://doi.org/10.1142/S0219720010005129>.
73. Eckardt NA. Genomic hopscotch: gene transfer from plastid to nucleus. *Plant Cell*. 2006;18(11):2865–7. <https://doi.org/10.1105/tpc.106.049031>.

74. Braukmann TWA, Kuzmina M, Stefanovic S. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Curr Genet.* 2009;55(3):323–37. <https://doi.org/10.1007/s00294-009-0249-7>.
75. Ranade SS, García-Gil MR, Rossello JA. Non-functional plastid *ndh* gene fragments are present in the nuclear genome of Norway spruce (*Picea abies* L. Karsch): insights from in silico analysis of nuclear and organellar genomes. *Mol Gen Genomics.* 2016;291(2):935–41. <https://doi.org/10.1007/s00438-015-1159-7>.
76. Neale DB, McGuire PE, Wheeler NC, Stevens KA, Crepeau MW, Cardeno C, et al. The Douglas-fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. *G3.* 2017;7(9):3157–67.
77. Lee HS, Fischer RL, Goldberg RB, Harada JJ. Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *P Natl Acad Sci USA.* 2003;100(4):2152–6. <https://doi.org/10.1073/pnas.0437909100>.
78. Alemanno L, Devic M, Niemenak N, Sanier C, Guilleminot J, Rio M, et al. Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L. *Planta.* 2008;227(4):853–66. <https://doi.org/10.1007/s00425-007-0662-4>.
79. Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, et al. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell.* 2003;15(1):5–18. <https://doi.org/10.1105/tpc.006973>.
80. Xie ZY, Li X, Glover BJ, Bai SN, Rao GY, Luo JC, et al. Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. *Mol Biol Evol.* 2008;25(8):1581–92. <https://doi.org/10.1093/molbev/msn105>.
81. Luerssen K, Kirik V, Herrmann P, Misera S. FUSCA3 encodes a protein with a conserved VP1/ABI3-like B3 domain which is of functional importance for the regulation of seed maturation in *Arabidopsis thaliana*. *Plant J.* 1998;15(6):755–64. <https://doi.org/10.1046/j.1365-313X.1998.00259.x>.
82. Sun FS, Liu XY, Wei QH, Liu JN, Yang TX, Jia LY, et al. Functional characterization of TaFUSCA3, a B3-superfamily transcription factor gene in the wheat. *Front Plant Sci.* 2017;8. <https://doi.org/10.3389/fpls.2017.011133>.
83. Bilchak A, Luu J, Jiang F, Eudes F. Identification of BABY BOOM homolog in bread wheat. *Agri Gene.* 2018;7:43–51. <https://doi.org/10.1016/j.aggene.2017.11.002>.
84. El Ouakfaoui S, Schnell J, Abdeen A, Colville A, Labbe H, Han SY, et al. Control of somatic embryogenesis and embryo development by AP2 transcription factors. *Plant Mol Biol.* 2010;74(4–5):313–26. <https://doi.org/10.1007/s11103-010-9674-8>.
85. Rahman MH, Toda E, Kobayashi M, Kudo T, Koshimizu S, Takahara M, et al. Expression of genes from paternal alleles in rice zygotes and involvement of OsASGR-BBML1 in initiation of zygotic development. *Plant Cell Physiol.* 2019;60(4):725–37. <https://doi.org/10.1093/pcp/pcz030>.
86. Bui LT, Pandzic D, Youngstrom CE, Wallace S, Irish EE, Szovenyi P, et al. A fern AINTEGUMENTA gene mirrors BABY BOOM in promoting apogamy in *Ceratopteris richardii*. *Plant J.* 2017;90(1):122–32. <https://doi.org/10.1111/tbj.13479>.
87. Liu CX, Zhang TZ. Expansion and stress responses of the AP2/EREBP superfamily in cotton. *BMC Genomics.* 2017;18:1–6.
88. Albrecht C, Russinova E, Hecht V, Baaijens E, de Vries S. The Arabidopsis thaliana SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. *Plant Cell.* 2005;17(12):3337–49. <https://doi.org/10.1105/tpc.105.036814>.
89. Li ZY, Wang Y, Huang J, Ahsan N, Biener G, Paprocki J, et al. Two SERK receptor-like kinases interact with EMS1 to control anther cell fate determination. *Plant Physiol.* 2017;173(1):326–37. <https://doi.org/10.1104/pp.16.01219>.
90. Nolan KE, Kurdyukov S, Rose RJ. Characterisation of the legume SERK-NIK gene superfamily including splice variants: implications for development and defence. *BMC Plant Biol.* 2011;11(1):44. <https://doi.org/10.1186/1471-2229-11-44>.
91. Sharma SK, Millam S, Hein I, Bryan GJ. Cloning and molecular characterisation of a potato SERK gene transcriptionally induced during initiation of somatic embryogenesis. *Planta.* 2008;228(2):319–30. <https://doi.org/10.1007/s00425-008-0739-8>.
92. Jing YJ, Lin RC. PICKLE is a repressor in seedling de-etiolation pathway. *Plant Signal Behav.* 2013;8(8):e25026. <https://doi.org/10.4161/psb.25026>.
93. Jing YJ, Zhang D, Wang X, Tang WJ, Wang WQ, Huai JL, et al. Arabidopsis chromatin remodeling factor PICKLE interacts with transcription factor HY5 to regulate hypocotyl cell elongation. *Plant Cell.* 2013;25(1):242–56. <https://doi.org/10.1105/tpc.112.105742>.
94. Jha P, Ochatt SJ, Kumar V. WUSCHEL: a master regulator in plant growth signaling. *Plant Cell Rep.* 2020;39(4):431–44. <https://doi.org/10.1007/s00299-020-02511-5>.
95. Su YH, Tang LP, Zhao XY, Zhang XS. Plant cell totipotency: insights into cellular reprogramming. *J Integr Plant Biol.* 2020;63:228–43.
96. Nowak K, Wojcikowska B, Gaj MD. ERF022 impacts the induction of somatic embryogenesis in *Arabidopsis* through the ethylene-related pathway. *Planta.* 2015;241(4):967–85. <https://doi.org/10.1007/s00425-014-2225-9>.
97. To A, Valon C, Savino G, Guilleminot J, Devic M, Giraudat J, et al. A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell.* 2006;18(7):1642–51. <https://doi.org/10.1105/tpc.105.03.9925>.
98. Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T. LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 2005;46(3):399–406. <https://doi.org/10.1093/pcp/pci048>.
99. Kroj T, Savino G, Valon C, Giraudat J, Parcy F. Regulation of storage protein gene expression in *Arabidopsis*. *Development.* 2003;130(24):6065–73. <https://doi.org/10.1242/dev.00814>.
100. Junker A, Monke G, Rutten T, Keilwagen J, Seifert M, Thi TM, et al. Elongation-related functions of LEAFY COTYLEDON1 during the development of *Arabidopsis thaliana*. *Plant J.* 2012;71(3):427–42.
101. Casson SA, Lindsey K. The turnip mutant of *Arabidopsis* reveals that LEAFY COTYLEDON1 expression mediates the effects of auxin and sugars to promote embryonic cell identity. *Plant Physiol.* 2006;142(2):526–41. <https://doi.org/10.1104/pp.106.080895>.
102. Passarinho P, Ketelaar T, Xing M, van Arkel J, Maliepaard C, Hendriks MW, et al. BABY BOOM target genes provide diverse entry points into cell proliferation and cell growth pathways. *Plant Mol Biol.* 2008;68(3):225–37. <https://doi.org/10.1007/s11103-008-9364-y>.
103. Ranade SS, Delhomme N, García-Gil MR. Global gene expression analysis in etiolated and de-etiolated seedlings in conifers. *PLoS One.* 2019;14(7):e0219272.
104. Proost S, Van Bel M, Vanechoutte D, Van de Peer Y, Inze D, Mueller-Roeber B, et al. PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* 2015;43(D1):D974–81. <https://doi.org/10.1093/nar/gku986>.
105. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature.* 2013;497(7451):579–84. <https://doi.org/10.1038/nature12211>.
106. Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin YC, Sjodin A, et al. The plant genome integrative explorer resource: PlantGenIE.org. *New Phytol.* 2015;208(4):1149–56. <https://doi.org/10.1111/nph.13557>.
107. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res.* 2013;41(D1):D36–42.
108. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu SN, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 2015;43(D1):D222–6. <https://doi.org/10.1093/nar/gku1221>.
109. Yang HF, Kou YP, Gao B, Soliman TMA, Xu KD, Ma N, et al. Identification and functional analysis of BABY BOOM genes from *Rosa canina*. *Biol Plant.* 2014;58(3):427–35. <https://doi.org/10.1007/s10535-014-0420-y>.
110. Brand A, Quimbaya M, Tohme J, Chavarriga-Aguirre P. Arabidopsis LEC1 and LEC2 orthologous genes are key regulators of somatic embryogenesis in Cassava. *Front Plant Sci.* 2019;10. <https://doi.org/10.3389/fpls.2019.00673>.
111. Lazarova G, Zeng Y, Kermod AR. Cloning and expression of an ABSCISIC ACID-INSENSITIVE 3 (ABI3) gene homologue of yellow-cedar (*Chamaecyparis nootkatensis*). *J Exp Bot.* 2002;53(371):1219–21. <https://doi.org/10.1093/jxbbot/53.371.1219>.
112. Barreto HG, Sagio SA, Chalfun A, Fevereiro P, Benedito VA. Transcriptional profiling of the AFL subfamily of B3-type transcription factors during the in vitro induction of somatic embryogenesis in the model legume *Medicago truncatula*. *Plant Cell Tiss Org.* 2019;139(2):327–37. <https://doi.org/10.1007/s11240-019-01687-3>.
113. Li YB, Li QW, Guo GM, He T, Gao RH, Faheem M, et al. Transient overexpression of HvSERK2 improves barley resistance to powdery mildew. *Int J Mol Sci.* 2018;19(4):1226.
114. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7. <https://doi.org/10.1093/nar/gkh340>.

115. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019;47(W1):W636–41. <https://doi.org/10.1093/nar/gkz268>.
116. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008;36(Web Server):W465–9. <https://doi.org/10.1093/nar/gkn180>.
117. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–21. <https://doi.org/10.1093/sysbio/syq010>.
118. Chevenet F, Brun C, Banuls AL, Jacq B, Christen R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics.* 2006;7(1). <https://doi.org/10.1186/1471-2105-7-439>.
119. Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* 2014;15(3):R59. <https://doi.org/10.1186/gb-2014-15-3-r59>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

