Gene 850 (2023) 146935

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Analysis of the *ASR* and *LP3* homologous gene families reveal positive selection acting on *LP3-3* gene

Jonathan Lecoy [1], Sonali Sachin Ranade [1], María Rosario García-Gil [*]

Department of Forest Genetics and Plant physiology, Umeå Plant Science Centre, Swedish University of Agricultural Sciences, SE-901 83 Umeå, Sweden

ARTICLE INFO

Edited by: Andre van Wijnen

Keywords:
ABA/WDS
ASR
Codon usage
Drought resistance
GC-content
LP3
Selective pressure

ABSTRACT

Drought has long been established as a major environmental stress for plants which have in turn developed several coping strategies, ranging from physiological to molecular mechanisms. *LP3* that was first discovered in loblolly pine (*Pinus taeda* L.) is a homolog of the Abscisic Acid, Stress and Ripening (*ASR*) gene belonging to the ABA/WDS gene family that was first detected in tomato. *LP3* has been shown to be present in four different paralogs in loblolly pine called *LP3-0, LP3-1, LP3-2* and *LP3-3*. LP3 in loblolly pine has not been as extensively studied as the ASR in tomato. Similar to *ASR*, the different *LP3* paralogs have been shown to be upregulated in response to water deficit stress and to act as transcription factors for genes likely involved in hexose transport. In the current study, we have investigated the evolutionary history of *LP3* gene family, with the aim of relating it to that of *ASR* from a phylogenetic perspective and comparing the differences in selective pressure and codon usage. Phylogenetic trees revealed that *LP3* is less divergent across species than *ASR* even when the trees were solely based on the different sub-sections of the gene. Phylogenetic, GC content, codon usage and selective pressure analyses suggest that *LP3-3* is undergoing positive selection.

1. Introduction

Land colonisation by plants during the Paleozoic has forced these plants to adopt several adaptive strategies to survive desiccation (Edwards and Selden, 1992). These strategies led to the development of organs such as roots for taking up water and the implementation of water stress management tactics like the closure of stomata and the modulation of osmotic pressures within the plant cell in order to maintain the plants' water potential (Chaves et al., 2003). Today, many plant species have adapted strategies to cope up with drought through millennia of evolution, yet anthropogenic climate change is expected to dramatically affect the growth conditions of most plant species, notably through increased drought occurrence and aridity around the world.

Drought is a major hazard to the survival and development of commercially important plants, from crops to forest tree species. In recent years, there has been an observed increase in drought occurrences notably in southern Europe, sub-Saharan Africa and many other areas around the world and this trend will increase with time as climate change continues to progress (Gudmundsson and Seneviratne, 2016; Ruosteenoja et al., 2018). It is therefore important to understand in depth the mechanisms by which plants have adapted to water deficit stress, to produce drought resistant varieties.

Water deficiency, as a major stress for plants, is detected in many ways, with the signalling component being mediated largely through the phytohormone Abscisic Acid (ABA) which is involved in stress response in plants, notably via its effects on gene expression and osmotic pressure adjustment within the plant cell (Bray, 1993). Since the 1990 s, several research projects have focused on a drought responsive gene called ABA, Stress and Ripening (*ASR*), first detected in tomato (*Solanum lycopersicum*) leaves yet remarkably absent from *Arabidopsis thaliana* (Iusem et al., 1993). This research has led to the discovery of many different *ASR* orthologs and paralogs, with tomato having five different *ASR* genes, and rice up to six to date (Frankel et al., 2006; Dominguez and Carrari, 2015). Transgenic expression of the *ASR1* gene in *Arabidopsis* produced a phenotype similar to what is observed in *abi4* mutants, including a lack of sensitivity to ABA-mediated germination inhibition,

Abbreviations: ABA, Abscisic Acid; AICc, Akaike Information Criterion; *ASR*, ABA, Stress and Ripening; CAI, Codon Adaptation Index; CDD, Conserved sequence database; CESA, Cellulose synthases; CUB, Codon Usage Bias; ML, Maximum Likelihood; MYA, Million Years Ago; NLS, Nuclear Localisation Signal; RCSU, Relative Synonymous Codon Usage.

* Corresponding author.
E-mail addresses: jonathanlecoy@gmail.com (J. Lecoy), Sonali.Ranade@slu.se (S.S. Ranade), M.Rosario.Garcia@slu.se (M.R. García-Gil).

[1] These authors contributed equally.

https://doi.org/10.1016/j.gene.2022.146935
Received 27 April 2022; Received in revised form 20 September 2022; Accepted 26 September 2022
Available online 29 September 2022

in addition to an increased tolerance to salt, cold and other stresses (Yang et al., 2005; Gonzalez and Iusem, 2014).

The *ASR* gene family contains a highly conserved midsection gene domain called the ABA/WDS domain (Pfam reference: PF02496). The ABA/WDS domain is also expressed in mushrooms from the genus *Fomitopsis*, a membrane protein of *Pseudomonas* and angiomotin found in fern (Padmanabhan et al., 1997; Wang et al., 2002; Gonzalez and Iusem, 2014). ASR in its native state is a disorganised, highly hydrophilic protein that requires two zinc ions to bind to lysine located in its *N*-terminal region to adopt its functional conformation, which leads to a protein dimerization and in turn bind to the plants' DNA sequence (Goldgur et al., 2007; Gonzalez and Iusem, 2014). ASR proteins act as transcription factors that induce the expression of aquaporines, cellulose synthases (CESA) and glucanases. *ASR1*, the most studied of the *ASR* genes, is for example involved in sugar metabolism in response to drought (Dominguez and Carrari, 2015).

An *ASR* homolog called *LP*3 was first discovered in the roots of loblolly pine (*Pinus taeda* L) whose background constitutive expression was significantly upregulated in drought conditions.

(Padmanabhan et al., 1997; Wang et al., 2002; Gonzalez and Iusem, 2014). For the purposes of the current study, while both genes are homologous, the terms *LP*3 and *ASR* will be used to describe the gymnosperm and angiosperm sequences, respectively. The ABA/WDS domain is highly conserved in LP3; LP3 differs from ASR in that it contains a 35 amino acid insertion between its *N*-terminal and ABA/WDS regions. The *LP3* gene family consist of four members, *LP3-0*, *LP3-1*, *LP3-2* and *LP3-3*, where *LP3-2* and *LP3-3* have been only partially sequenced, therefore only partial coding sequences are available in conifers and other gymnosperms (Chang et al., 1996; Padmanabhan et al., 1997). *LP*3 transport into the nucleus is mediated by the putative C-terminal Nuclear Localisation Signal (NLS) of sequence KKESKEEEKEAEGKKHHH, although a previous evidence has shown that this NLS may not be required for transport through the nuclear envelope (Padmanabhan et al., 1997; Wang et al., 2002; Ricardi et al., 2012). LP3 has not been as extensively studied as ASR, probably due to the difficulty of genetic studies within gymnosperms, a fact that has motivated the current study.

The goal of the current research work is to study the evolutionary history of *LP3* gene family in conifers. We aim to compare the rate and mode of evolution of the *LP3* gene with its homologue in the angiosperms, the *ASR* gene. To achieve our goal we have conducted the following analysis: (i) retraced the phylogeny of *LP3* as a member of the ABA/WDS family and relate it to the *ASR* genes, (ii) estimated the GC content and Codon Usage Bias (CUB), and (iii) determined the mode of evolution of different subsections of the *ASR/LP3* genes.

2. Material and methods

2.1. Identification of LP3 and ASR genes

Loblolly pine *LP3-0*, *LP3-1*, *LP3-2* and *LP3-3* CDSs were downloaded from NCBI (Padmanabhan et al., 1997). Homologous angiosperm *ASR* and gymnosperm *LP3* whole gene CDSs were then extracted from NCBI using these sequences as queries via BLASTN in the NCBI database (Boratyn et al., 2013) with an e-value of 1^{e-10} as a threshold, using the standard databases (Nucleotide collection nr/nt), while keeping other search parameters as default (Supplementary Fig. 1). Complete gymnosperm homologous sequences were also extracted via BLASTN in the Gymno Plaza database (v1.0). For naming sequences in the instances where the gymnosperm sequences were uncharacterized, comparison with loblolly pine *LP3* sequences with the NEEDLE alignment tool were done and whichever alignment had the highest score was used to determine to which paralog the uncharacterized sequence was most likely to be similar. These sequences were then numbered successively as *LP3-0.1, LP3-0.2, LP3-0.3* and so on, in a particular species. The CDS accession number was added after this number. For example in *Pinus pinaster*, the LP3 sequences were named as LP3-0.1:PPI00000827, LP3-0.2:PPI00033797 and so on. With reference to *ASR* genes, the gene name was followed by the CDS accession number e.g. *ASR1:*NM_001247208, *ASR2:*NM_001320991, *ASR3:*NM_001309371 and *ASR4:* NM_001282319 in *Solanum lycopersicum*. For simplicity, the same naming convention was maintained for the corresponding protein sequences, for both LP3 and ASR. Since LP3 and ASR are members of the ABA/WDS induced protein superfamily (Chang et al., 1996; Padmanabhan et al., 1997; Gonzalez and Iusem, 2014), care was taken to ensure that all sequences retrieved contained the ABA/WDS domain in order to be the part of the ABA/WDS family, using PFAM v.31 (Bateman and Finn, 2007; Mistry et al., 2007; Schaeffer et al., 2017). LP3 orthologs were also confirmed for the presence of ABA/WDS domain using the conserved sequence database (CDD) (Marchler-Bauer et al., 2015). Species-wise accession numbers used for the phylogenetic tree reconstruction are presented in the Supplementary Table S1. To assess further, the LP3 orthologs were subjected to Blast2GO analysis with SWISS-PROT as the reference database (Blast2GO version 6.0.3) (Conesa et al., 2005).

2.2. Phylogenetic analysis

The phylogenetic history of LP3 and ASR was determined using MEGA X (Kumar et al., 2018). Firstly, this was done by looking at the whole sequences, then by looking only at the conserved ABA/WDS

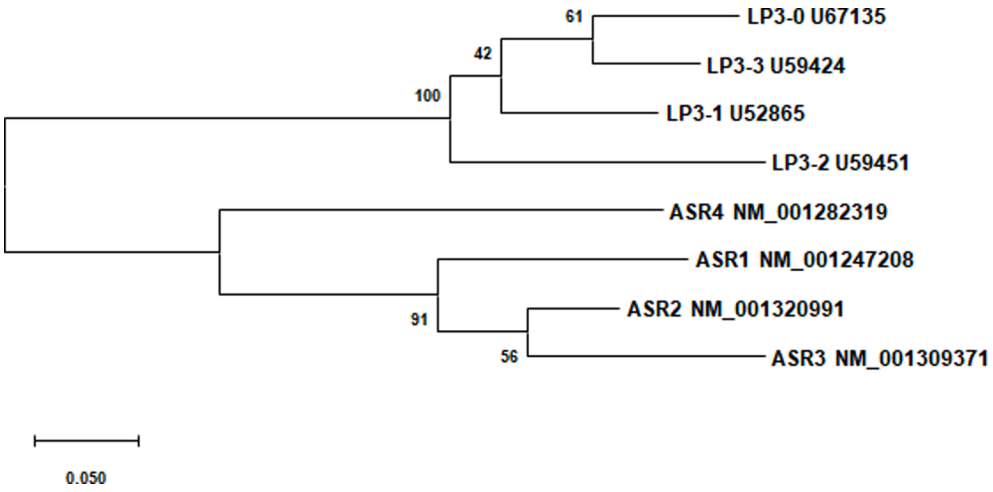

Fig. 1. Phylogenetic tree of *Pinus taeda* LP3 and *Solanum lycopersicum* ASR sequences.

region, the conserved *N*-terminal zinc binding region before the gymnosperm insertion, and then finally by looking only at the variable C-terminal NLS/DNA binding region of the sequences, producing a total of four different trees called FullSeq-tree, ABA/WDS-tree, *N*-tree and C-tree, respectively. The species phylogenetic tree was constructed using the Timetree software (Kumar et al., 2017). Furthermore, site substitution saturation tests were performed on fully resolved sites of ASR and LP3 genes using DAMBE (Xia et al., 2003; Xia, 2017).

For FullSeq-tree, sequences were aligned using the MUSCLE algorithm (Edgar, 2004), checked for errors and the multiple sequence alignment was exported for further analysis. The Maximum Likelihood (ML) Phylogenetic best-fit model was determined in MEGA X by log-likelihood analysis of each model and the one with the highest corrected Akaike Information Criterion (AICc) score was used (Hurvich and Tsai, 1989). The resulting best-fit model for the sequences used in tree FullSeq-tree was the Kimura 2 parameter model. The phylogenetic tree was created using the Maximum likelihood method combined with the Kimura 2 parameter model with a gamma parameter of 1.30 with 1000 permutations.

ABA/WDS-tree was constructed by isolating the highly conserved ABA/WDS domains from the Multiple Sequence Alignments of LP3 and ASR and exporting those for phylogenetic analysis. The ML phylogenetic model was determined in the same manner as previously described, with the resulting best model by log-likelihood analysis being the Kimura 2 parameter model. The tree was constructed using the ML method with the Kimura 2 parameter with a Gamma parameter equal to 1.0846 with 1000 permutations.

Tree *N*-tree was constructed by extracting the *N*-terminal nucleotide region of LP3 and ASR. Due to the presence of incomplete sequences that did not cover this section of the gene, the following sequences were excluded from the phylogenetic analysis: LP3-0 *Cupressus sempervirens;* LP3-1 *Pinus sylvestris* L.*;* LP3-1 *Pinus hwangshenensis;* LP3-0 *Pinus massoniana* Lamb*;* LP3-2 *Pinus taeda;* LP3-2 *Pinus sylvestris* and all LP3-3 sequences. The ML phylogenetic model was determined in the same manner as previously described, with the resulting best model by log-likelihood analysis result being the Kimura 2 parameter model. The tree was therefore constructed using the ML method with the Kimura 2 parameter model with a gamma parameter equal to 1.27 with 1000 permutations.

Tree C-tree was constructed by focusing on the C-terminal NLS/DNA binding regions of the sequences. The ML phylogenetic model was determined in the same manner as previously described, with the resulting best model by log-likelihood analysis result being the Kimura 2 parameter model. The following partial sequences were excluded from the phylogenetic analysis due to too short C-terminal sequences: LP3-0 *Cupressus sempervirens,* LP3-0 *Pinus massoniana;* LP3-2 *Pinus taeda;* LP3-2 *Pinus sylvestris* and all LP3-3 sequences. The tree was constructed using the ML method with the Kimura 2 parameter model with a gamma parameter equal to 1.5588 with 1000 permutations. Analysis of gene duplication events and construction of the corresponding gene duplication tree was also performed in MEGA X, using FullSeq-tree as a template.

2.3. GC and codon usage analyses

Species-wise accession numbers for determining GC% and Relative Synonymous Codon Usage (RCSU) are presented in the Supplementary Table S2. RSCU is a measure of the codon usage bias for a particular amino acid. The RSCU of the ABA/WDS genes found in majorly represented species (species in which at least three different ABA/WDS genes are present) was extracted using MEGA X. The average RSCU values per amino acid were then calculated and used to determine which codon was on average the most used in a particular gene. Codons with RSCU values above one are abundant, whilst codons with RSCU values below one are less abundant. Codons for Methionine and Tryptophane were not included since these amino acids are encoded by only one codon. Stop codons were not included either as they are only involved in transcription termination and therefore not in transcription efficiency.

The GC content at the first, second, and third codon position is referred to as GC1, GC2, and GC3 respectively. Individual gene parameters from majorly represented species such as total GC, GC1, GC2 and GC3 contents were computed using CAICal (http://genomes.urv.cat/CAIcal/). The highly expressed genes used for determining Codon Adaptation Index (CAI) in this study can be found at http://www.kazusa.or.jp/codon/. Statistical analyses were done in R using $p = 0.05$ as the significance threshold. Gene parameters were first compared using a Levene test to ensure variance equality among the different genes, followed by either ANOVA or a Kruskal-Wallis test. If either of these analyses indicated the presence of a significantly different group then a Tukey HSD test was performed in the case of ANOVA and a pairwise Wilcoxon test in the case of a Kruskal-Wallis. Pearson correlations between the different gene parameters were also done on *ASR* and *LP3* genes grouped together.

2.4. Mode of evolution of different subsections of the ASR/LP3 gene

A codon-by-codon selective pressure analysis provides insights regarding the amino acids undergoing selective constraints in a protein. This in turn allows predicting the amino acids that are likely to change over time. Amino acids under positive selection may be indicative of molecular adaptation, which may in turn provide the organism with an evolutionary advantage. Visualisation of codon selective pressure as defined by the ratio of synonymous to non-synonymous codons $\omega = \frac{dN}{dS}$ on the *LP3-0* gene was done using the complete CDS of all *LP3-0* orthologous sequences. Sequences were uploaded to the Selecton server (Stern et al., 2007) and selective pressure analysis was performed using the M8 model (Yang et al., 2000) that allows for positive selection operating on the protein. Multiple Sequence Alignments and phylogenetic tree for *LP3-0* orthologs created using MEGA X were uploaded to the server to get the results. Statistical analysis of the calculated selective pressure was also performed using Selecton with default settings by calculating the log likelihood ratio between M8 and the null model M8a. The same procedure was repeated for *LP3-1, LP3-3, ASR1, ASR2, ASR3* and *ASR4* orthologs, respectively. *Pinus taeda LP3-0, LP3-1, LP3-3* and *Solanum lycopersicum ASR1, ASR1, ASR2, ASR3* and *ASR4* sequences were set as references to visualise sites of selective pressure. There were not enough *LP3-2* homologous sequences for this analysis.

2.5. Norway Spruce LP3 expression data

Expression graph for different *LP3* sequences in Norway Spruce [*Picea abies* (L.) H. Karst] was plotted using exPlot from the ConGenIE.org. The highest homology of the sequences was determined through a BLAST analysis against *Pinus taeda* on the NCBI database.

3. Results

3.1. Phylogenetic analysis of LP3 and ASR

On a broad level, LP3 shows two lineages; one branch gave rise to LP3-2 and the other one diverse further to form LP3-0, LP3-1 and LP3-3. Likewise, the *ASR* phylogeny shows that an early diversification that led to the formation of ASR4, while the other branch gave rise to ASR1, ASR2 and ASR3 (Fig. 1). Fully resolved *ASR* sites on codon position 1 + 2, and position 3 show no saturation occurring on these sites (Supplementary Figs. 3 and 4). This is confirmed by the high number of fully resolved sites in *ASR* genes.

On the other hand, there were very few fully resolved sites in the *LP3* genes when observing both codon positions 1 + 2 and codon position 3 which, while displaying little saturation, cannot be trusted to represent the entirety of the *LP3* genes (Supplementary Figs. 5 and 6). The close

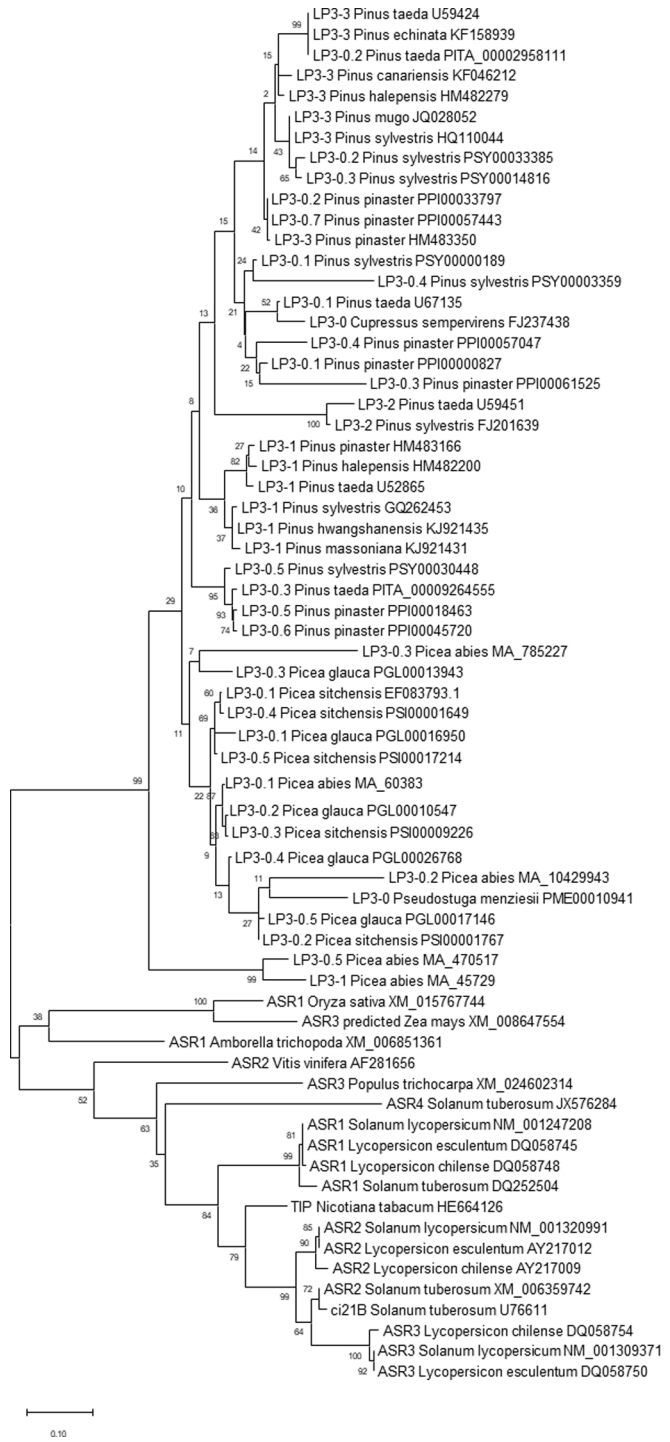

Fig. 2. FullSeq-tree corresponding to the entirety of available CDS of ASR/LP3 sequences. Constructed in MEGAX using the ML method and the K2 model + gamma = 1.2882. Genetic distance is given in number of substitutions per site. Bootstrap values are shown next to each node.

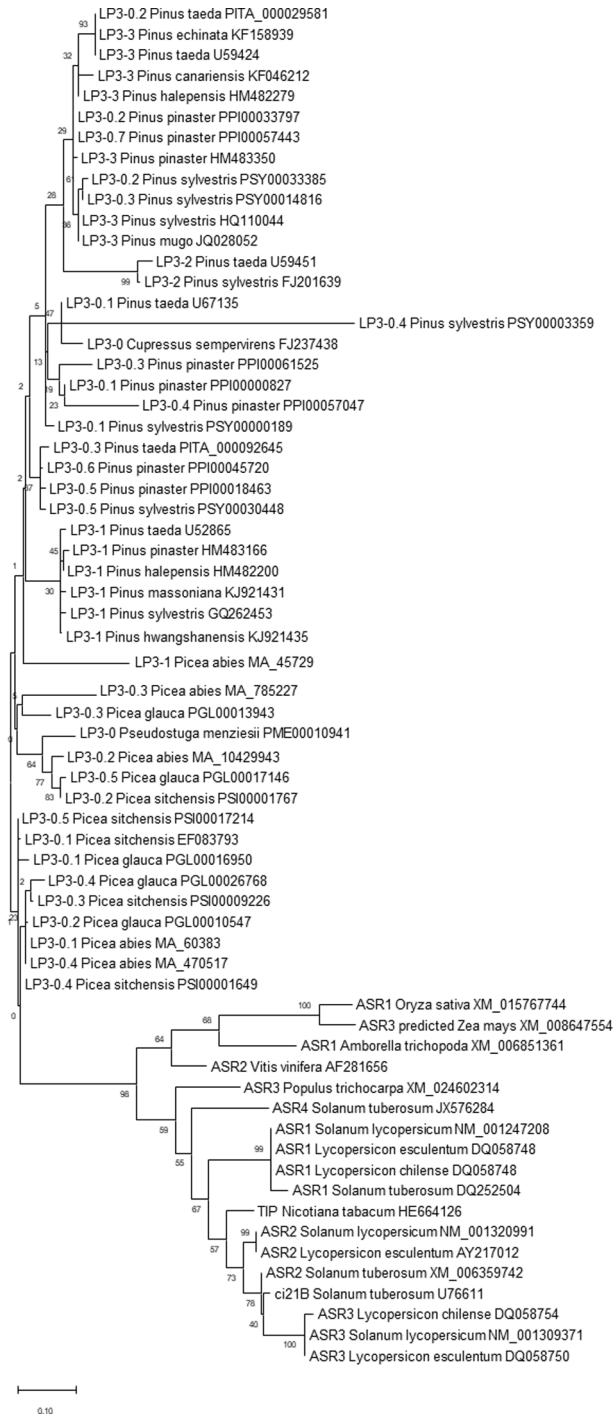

Fig. 3. ABA/WDS-tree focusing on the conserved ABA/WDS domain within each gene. Constructed in MEGAX using the ML method and the K2 model with gamma = 1.0846. Genetic distances are in number of substitutions per site. Bootstrap values are shown next to each node.

phylogenetic distance between the LP3 (conifer) species can account for the low number of fully resolved sites. In trees involving closely related taxa, the number of mutational differences is typically very low. This causes stochastic effects that could bias a saturation analysis due to a bias on the molecular clock (Walsh et al., 1999). For this reason, the saturation results in LP3 need to be interpreted with caution.

FullSeq-tree (Fig. 2) shows that there is a clear divergence between the angiosperm ASR and the gymnosperm LP3 sequences. This could be attributed to certain extent to the sequence insertion present between the *N*-terminal zinc binding domain and ABA/WDS regions in gymnosperm. Low bootstrap values within the gymnosperm nodes could be accounted by the general high similitude between sequences which would result in them easily swapping positions during different bootstrap analyses. It can be observed that the LP3 sequences cluster together according to genus and paralog. There is also perfect clustering of LP3-1 and LP3-2 sequences within the *Pinus* cluster, however this is not the case for the LP3-3 cluster. In that instance, clustering of LP3-3 occurs with LP3-0–2 and LP3-0–3 of *Pinus sylvestris*. While this may be

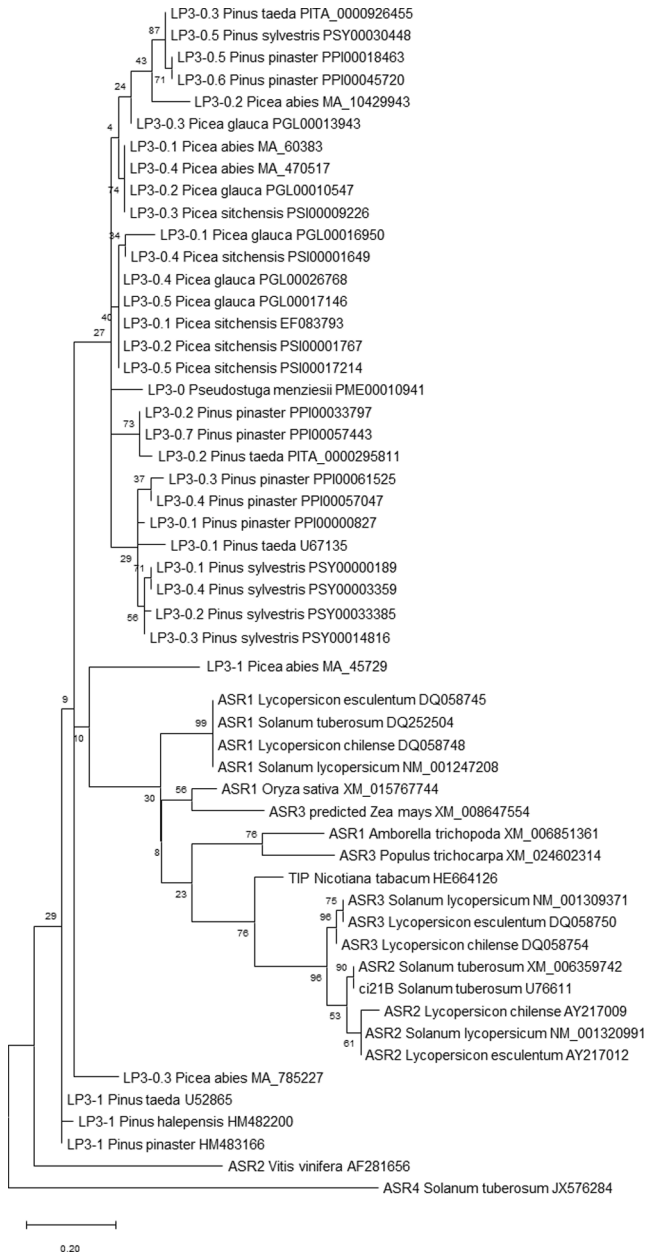

Fig. 4. *N*-tree focusing on *N*-terminal of ASR and LP3 genes. Constructed with MEGAX and using the ML method and K2 model with gamma = 1.3701. Genetic distances are in number of substitutions per site. Bootstrap values are shown next to each node.

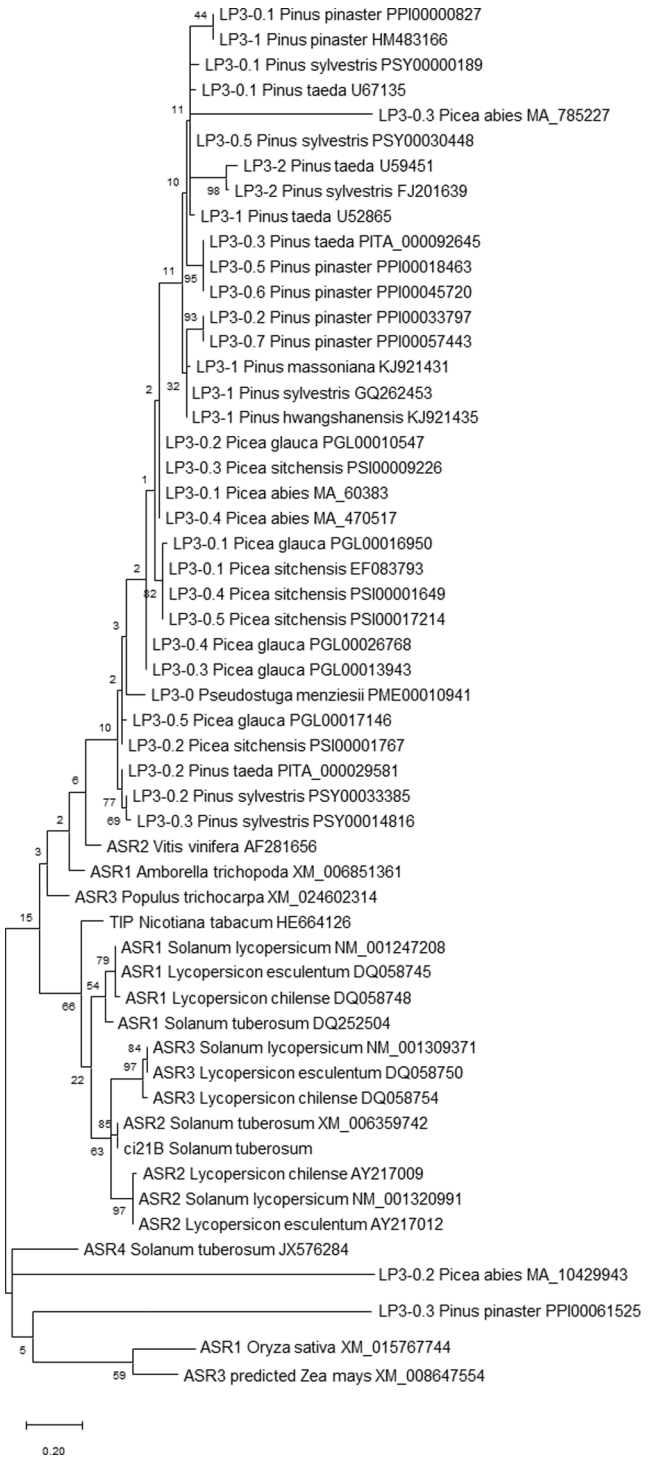

Fig. 5. C-tree focusing on the C-terminal region of ASR and LP3 genes. Constructed with MEGA X and using the ML method with K2 model and gamma = 1.5588. Genetic distances are in number of substitutions per site. Bootstrap values are shown next to each node.

accounted by the sequence similarity between those two paralog genes, it cannot be discarded the possibility of an erroneous gene annotation. For example, the grouping together of *Pinus taeda* LP3-3 and *Pinus taeda* LP3-0–2 with a very high bootstrap score could suggest that LP3-0–2 (PITA_00002958) might be actually a complete sequence of LP3-3. In *Picea* most of the sequences were annotated as LP3-0, except one sequence that was annotated as LP3-1 (MA45729). No LP3-2 or LP3-3 were available. The ASR sequences also show clear clustering together of ASR1, ASR2 and ASR3 sequences. The ASR1 of *Oryza sativa*, ASR3 of *Zea mays*, ASR2 of *Vitis vinifera* and ASR3 of *Populus trichocarpa* show signs of divergence from the other ASR sequences. There are big genetic distances between the different ASR clusters, indicative of substantial divergence between the *ASR* genes in the angiosperms.

Looking at ABA/WDS-tree (Fig. 3) which focuses on the central ABA/WDS domain, there is yet again another clear grouping of angiosperm and gymnosperm sequences together. Within the gymnosperm grouping, one can observe the sub-grouping of LP3-1 and LP3-2 together; while similar to FullSeq-tree, the LP3-3 sequences are again broadly grouped together with some LP3-0 sequences grouped alongside them. The angiosperm sequences also show clear clustering of ASR1, ASR2 and ASR3 sequences with the *Oriza sativa* ASR1 and *Zea mays* ASR3 seeming to diverge from the other ASR sequences. As in FullSeq-tree, there is also a clear clustering according to genus, with *Picea* and

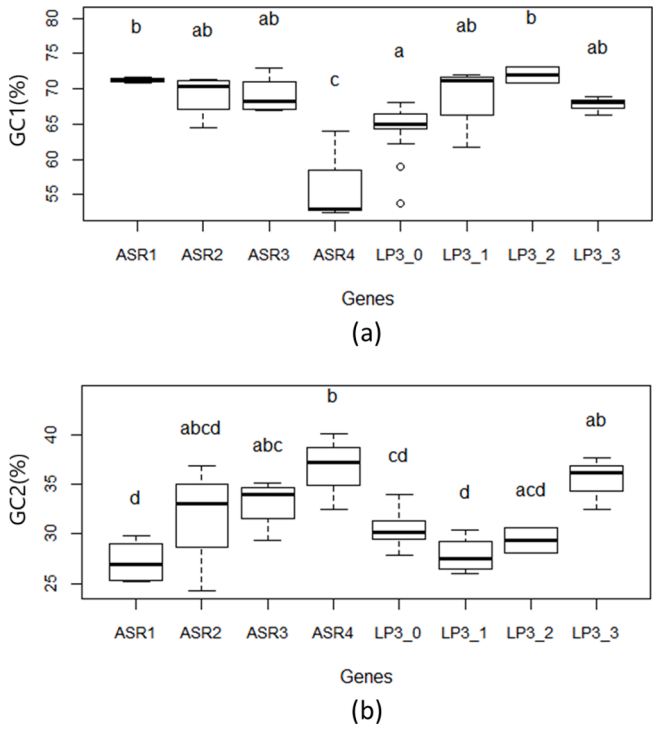

Fig. 6. The GC content at the first and second codon position is referred to as GC1 and GC2 respectively (a) Comparison of the average GC1% per ABA/WDS gene, using a one-way ANOVA followed by Tukey HSD test; (b) Comparison of the average GC2% per ABA/WDS gene, using a one-way ANOVA followed by Tukey HSD test.

Pinus sequences clustering together, respectively.

N-tree focuses on the *N*-terminal region of ASR and LP3, to which zinc ions are theorised to bind (Fig. 4). Here it can be observed that there is not a definite separation between angiosperm and gymnosperm sequences, with ASR2 from *Vitis vinifera* and ASR4 from *Solanum tuberosum* being the most divergent sequences. LP3-1 sequences in *Pinus* formed a different cluster although with low support. After this node, ASR and LP3 sequences mostly segregate according to their order. Like the trees previously analysed, the ASR sequences cluster according to ASR1, ASR2 and ASR3, with the *Oriza sativa* ASR1 and *Zea mays* ASR3 diverging together. In contrast to previous trees, there is a less clear clustering of LP3 sequences according to genus. There remains however a very low genetic distance between each of the LP3 sequences, indicative yet again of a high level of conservation between the sequences.

C-tree, which focuses on the variable C-terminal regions of ASR and LP3 which contains the NLS/DNA binding domain (Fig. 5), again shows partial segregation of ASR and LP3 sequences. Here the tree is rooted by the ASR4 of *Solanum tuberosum*, LP3-0–3 *Pinus pinaster* (Ait.) and *Picea abies* LP3-0–2 and LP3-0–3 sequences, indicative that these sequences are the most divergent in the C-terminal NLS region. Subsequently, the ASR and LP3 sequences do segregate according to their order (similar to what is be observed for the other trees). Within the ASR cluster the ASR1, ASR2 and ASR3 groups show low genetic distances, which suggests that the C-terminal region of those sequences is well conserved. The LP3 C-terminal sequences also do not show much divergence between them as seen by the small genetic distances between the sequences. The exception are LP3-0–3, LP3-0–4 and LP3-0–2 sequences of *Picea abies*, which show a high level of divergence from the other LP3 sequences. The segregation within the LP3 sequences showed a similar clustering mostly according to the genus.

3.2. GC and codon bias

Levene test results showed that only GC1 and GC2 followed the ANOVA equal variance among groups assumption ($p > 0.05$). Total GC and GC3 were tested using a Kruskal-Wallis test. In all cases, the average GC2 of the ABA/WDS genes was lower than both average GC1 and GC3. Except for *ASR1*, all ABA/WDS genes had on average higher GC1 than GC3 (Table 1). Unlike GC and GC3, GC1 and GC2 showed significant differences between the ABA/WDS genes ($p < 0.05$). Analysis of GC1 suggests that on average *ASR4* has a significantly lower GC1 value compared to the rest of the genes (Fig. 6a). GC2 analysis revealed that *ASR4* has a significantly higher GC2 than *ASR1*, while *LP3-3* has significantly higher GC2 content than the rest of the *LP3* genes (Fig. 6b).

There appear to be differences in average preferential codon usage between the different ABA/WDS genes (Table 2, Fig. 7, Supplementary Table S3 and S4). In both *LP3-0* and *LP3-3,* the same most highly used codon is AGG, encoding arginine, at RSCU values 3.7 and 6.0, respectively. *LP3-2* and *ASR2* both preferentially use CCA, encoding proline, at RSCU values 4.0 and 2.8, respectively. Furthermore, *ASR2* has a second codon AGC, encoding serine, at RSCU 2.8. *LP3-1*, *ASR3* and *ASR4* all possess codons encoding for serine as their most used codons with codons UGC, AGC and AGU at RSCU values 3.9, 3.3 and 2.8, respectively. There were similarities between *LP3* and *ASR* genes in least used codons, such as the arginine encoding CGG and leucine encoding CUA. Furthermore, *LP3-3* showed the highest individual RSCU value (Fig. 8).

Significant correlations were found between the different gene parameters in *ASR* (Table 3) and *LP3* (Table 4). In both *ASR* and *LP3*, gene length was significantly negatively correlated with total GC, GC1. Total GC was in turn positively correlated with GC1 and GC3 in both *ASR* and *LP3* genes. GC1 was negatively correlated with GC2 in both *ASR* and *LP3* and positively correlated with GC3 in *ASR* only. Levene test results showed unequal variance among groups assumption for ANOVA ($p > 0.05$).

3.3. Mode of evolution of different subsections of the ASR/LP3 gene

The reference sequences of LP3-0, LP3-1, ASR1, ASR2, ASR3 and ASR4 show signs of purifying selection, mainly on amino acids K, H and E, and do not show signs of positive selection (Fig. 9a, b; Fig. 10 a, b, c, d; Supplementary Figure S7, S8, S9, S10, S11, S12; Supplementary Table S3, S4, S5, S6, S8, S9, S10, S11). In contrast, LP3-3 presents signs of positive selection on amino acids N, S and T at positions 34, 37 and 61, respectively (Fig. 9c). These amino acids tend to turn into E, T and A, respectively (w-values = 2.8, 2.9, 2.9; p-values = $2.1\ 10^{-14}$, 1.0, $2\ 10^{-19}$; $1.4\ 10^{-19}$, Fig. 11, Supplementary Table S7). LP3-3 orthologs from different species were used for the multiple sequence alignment (MUSCLE) represented in Fig. 11. Only partial coding sequences are available were available for this alignment; however all sequences contain the highly conserved ABA/WDS region (CDD: pfam02496, PFAM: PF02496, InterPro: IPR003496) which is the primary prerequisite for it being a *LP3* gene, which was confirmed using the CDD and PFAM. Blast2GO analysis further confirms the annotation (Description: Abscisic stress-ripening protein) for all the LP3 orthologs included in this alignment

Table 1
Average GC content of ABA/WDS genes. (GCn: where n is either 1, 2 or 3 that indicates the codon position being considered).

Gene	Length(bp)	GCtot	GC1	GC2	GC3
ASR1	323.3	56.83	71.25^{b}	27.20^{d}	72.05
ASR2	336.0	55.90	69.13^{ab}	31.80^{abcd}	66.70
ASR3	337.5	56.43	69.05^{ab}	33.10^{abc}	67.18
ASR4	629.0	43.47	56.50^{c}	36.60^{b}	37.33
LP3-0	435.7	50.69	64.76^{a}	30.44^{cd}	56.89
LP3-1	300.8	51.25	68.98^{ab}	27.88^{d}	56.98
LP3-2	306.0	51.60	71.95^{b}	29.35^{acd}	53.50
LP3-3	272.0	52.83	67.73^{ab}	35.43^{ab}	55.43

Table 2
Average most used codon in ABA/WDS genes.

Gene	AA	Codon	RSCU
ASR1	Leucine	CUC	2.6
ASR2	Proline, Serine	CCA, AGC	2.8
ASR3	Serine	AGC	3.3
ASR4	Serine	AGU	2.8
LP3-0	Arginine	AGG	3.7
LP3-1	Serine	UCG	3.9
LP3-2	Proline	CCA	4.0
LP3-3	Arginine	AGG	6.0

with one of the GO being response to water deprivation (Supplementary Table S12). Overall, the ABA/WDS genes investigated here are all undergoing purifying selection, and all tend to have similar amino acid sequences and codon usage as shown by the low Ka/Ks values between gene pairs (Table 5 and Table 6).

3.4. Norway Spruce LP3 expression patterns

The expression graph of LP3 from Congenie shows that there is a variability in expression patterns between the different LP3 genes (Fig. 12). MA_10429943 (84 % identity with *Pinus taeda LP3-3*, U59424.1) shows the highest overall expression, whereas MA_8225804 (84 % identity with *Pinus taeda LP3-3*, U59424.1) has the lowest overall expression, seemingly only expressed in June vegetative shoots. All other sequences appear to be upregulated in pineapple galls with MA_10429943 and MA_60383 (84 % identity with *Pinus taeda LP3-0*, U67135.1) being the most highly expressed in these conditions. MA_47051 (84 % identity with *Pinus taeda LP3-0*, U67135.1) appears to be most highly expressed two-year-old needles; MA_785227 (84 % identity with *Pinus taeda LP3-3*, U59424.1) has its highest expression in June vegetative shoots and MA_45729 (84 % identity with *Pinus taeda LP3-0*, U67135.1) has its highest expression during the establishment of August buds.

4. Discussion

In this study, we have characterized the *ASR* and *LP3* gene families by investigating their phylogenetic relationship, differences in GC content, codon usage and mode of evolution in angiosperms and gymnosperms.

While there is ample evidence that *ASR* and *LP3* belong to the same gene family (Padmanabhan et al., 1997; Frankel et al., 2006; Gonzalez and Iusem, 2014), the nucleotide phylogeny reveals a clear segregation between gymnosperm and angiosperm, even when only the highly conserved domain like the ABA/WDS domain is considered. The separation between LP3 and ASR sequences are consistent with previous phylogenetic knowledge about the divergence between gymnosperms and angiosperms around 313 Million Years Ago (MYA) between the Pennsylvanian and Permian periods, thereby giving the *ASR* and *LP3* genes within both orders ample time to diverge (Supplementary Fig. 2,

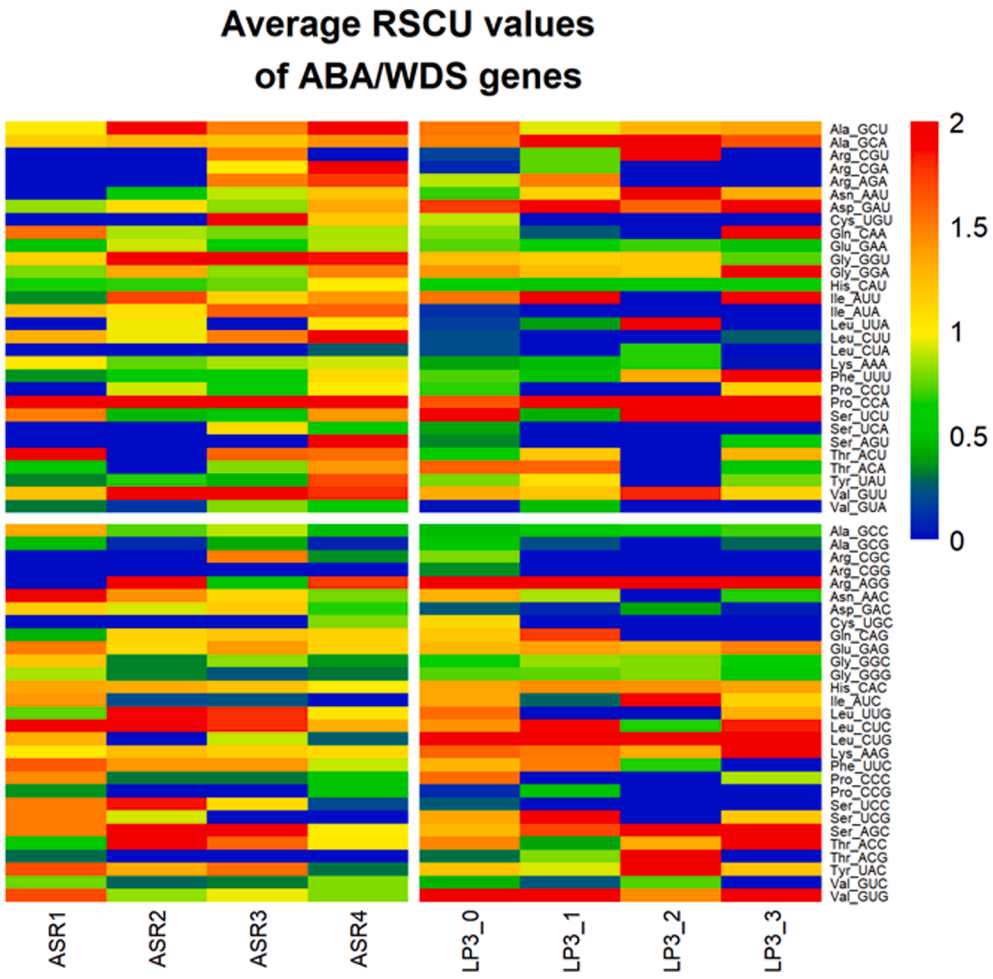

Fig. 7. Heatmap of average codon usage per ABA/WDS genes. Codons separated according to whether they are AU or GC ended. Average codon usage above 1.5 are indicated in red while codon usage lower than 0.5 are indicated in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

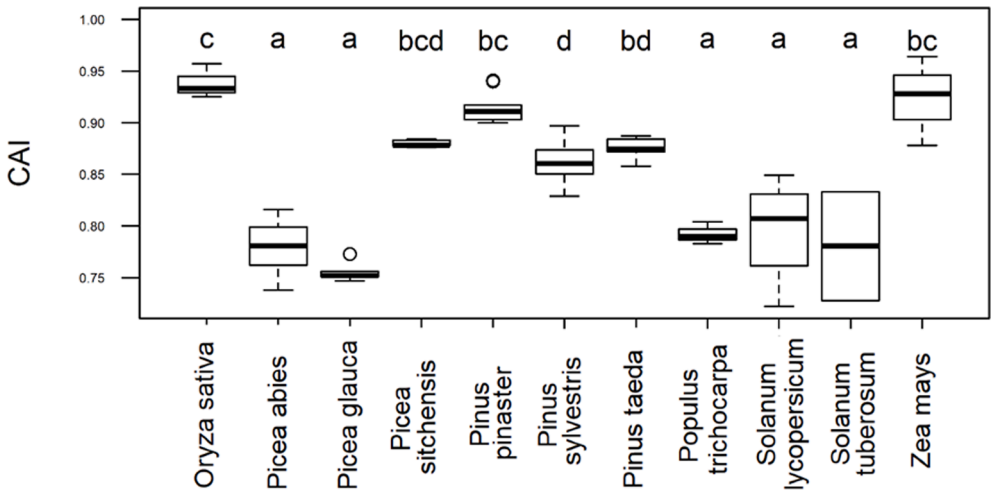

Fig. 8. Codon Adaptation Index (CAI) per species of the ABA/WDS genes obtained from the values of majorly represented species, using a Kruskal-Wallis test followed by a pairwise Wilcoxon test.

Table 3
Pearson correlation matrix of *ASR* genes. Correlation values indicated below diagonal and p-values indicated above diagonal. Significant correlations shown in bold.

	Length	GC	GC1	GC2	GC3
Length		0.0413	0.0000	0.0294	0.0602
CAI	−0.14	0.0001	0.4982	0.1384	0.0000
GC	**−0.50**		0.0182	0.5516	0.0000
GC1	**−0.89**	**0.56**		0.0048	0.0263
GC2	**0.53**	0.16	**−0.65**		0.5729
GC3	−0.46	**1.00**	**0.54**	0.15	
ENC	0.14	**−0.84**	−0.18	−0.47	**−0.84**

Table 4
Pearson correlation matrix of *LP3* genes. Correlation values indicated below diagonal and p-values indicated above diagonal. Significant correlations shown in bold.

	Length	GC	GC1	GC2	GC3
Length		0.0284	0.0020	0.7432	0.6149
CAI	−0.10	0.1385	0.0431	0.7252	0.6568
GC	**−0.35**		0.0000	0.8198	0.0000
GC1	**−0.48**	**0.58**		0.0110	0.9112
GC2	−0.05	0.04	**−0.40**		0.1903
GC3	0.08	**0.65**	−0.02	−0.21	
ENC	**0.48**	−0.17	−0.25	−0.09	0.11

www.timetree.org) (Hedges et al., 2015; Kumar et al., 2017). The FullSeq-tree is corroborated by a previous study on *ASR* genes (Frankel et al., 2006). The phylogenetic analyses of different sub-sections of the *ASR/LP3* genes provide insight into the different divergence rates occurring within each of them. Unlike in the ASR sequences, the ABA/WDS-tree shows very little divergence between the different LP3 sequences. This result is in line with previous analyses of homologous genes that also support a faster divergence of the angiosperm genes, which has been attributed to a lower rate of substitution in gymnosperms compared to angiosperms (Buschiazzo et al., 2012; De La Torre et al., 2017). This low level of phylogenetic divergence within LP3 can also be observed in the other two domain-based trees (*N*-terminal tree and C-terminal tree).

The *N*-terminal tree was focused on the putative zinc binding domains in the *N*-terminal region of *ASR/LP3* genes, therefore it would make sense that there is a low level of divergence occurring since ASR/LP3 requires Zinc ions to adopt their functional conformation (Gonzalez and Iusem, 2014; Dominguez and Carrari, 2015). Moreover, a less clear segregation between ASR and LP3 may be the result of a high purifying selective pressure, which adds to the argument in support of the essential and possibly similar role of the *N*-terminal domain in both clades. C-tree, which was focused on the putative NLS/DNA binding region of ASR/LP3, reveals similar topological properties as the *N*-tree but with a lower support to the branches discerning the different *ASR* gene family members. Both ASR and LP3 encode relatively small proteins and in the case of ASR1 it has been previously shown that it does not require its putative NLS to enter the nucleus (Rom et al., 2006; Ricardi et al., 2012). This low level of divergence might indicate that the selective pressures focusing on these parts of the genes stem mostly from their role as transcription factors. Following this reasoning, it could be suggested that the sequence corresponding to LP3-0–3 *Picea abies* targets a different gene somewhere along the genome other than the *LP3* gene family.

In addition to differences in the rate of evolution, GC content and codon bias (RSCU) also contribute to distinguish *ASR* and *LP3* gene families. In all cases GC2 was lower than both GC1 and GC3, a pattern which has also been reported in other studies (Smarda et al., 2014; Li et al., 2016; Song et al., 2017; Song et al., 2018). Serine shows the overall highest codon bias in both genes, while the amino acid affected by codon bias differ across individual genes, where leucine and arginine appear as in average most preferable amino acids in ASR and LP3, respectively. It has already been established that monocots and dicots differ in their codon usage for the homologous genes (Campbell and Gowri, 1990), therefore it is not improbable that gymnosperms would have different codon usages than angiosperms for homologous genes (De La Torre et al., 2017).

After gene duplication, the original and new copies can go down multiple evolutionary paths (Innan and Kondrashov, 2010). Previous studies have reported that *LP3-1* in *Pinus pinaster* and *LP3-3* in *Pinus halepensis* (Mill.) are undergoing purifying selection (Eveno et al., 2008; Grivet et al., 2011), whereas another study in *Pinus pinaster* did not find any traces of selection on these two (Gonzalez-Martinez et al., 2006). While *LP3-3* and *LP3-0* genes have seemingly diverged recently, they are affected by different modes of gene-duplication evolution. *LP3-3* presents signs of positive selection on three amino acids and the rest are evolving either under neutral or purifying selection, whereas *LP3-0* is only affected by purifying selection. In LP3-3 those three amino acids (mutation) may have introduced new beneficial functional aspects to the original LP3 copy resulting in their fixation and maintenance through positive selection. This mode of gene-duplication evolution could be attributed to the acquisition of a novel function for *LP3-3* gene. Notably purifying selection is acting on Lysine (K), Histidine (H) and Glutamate

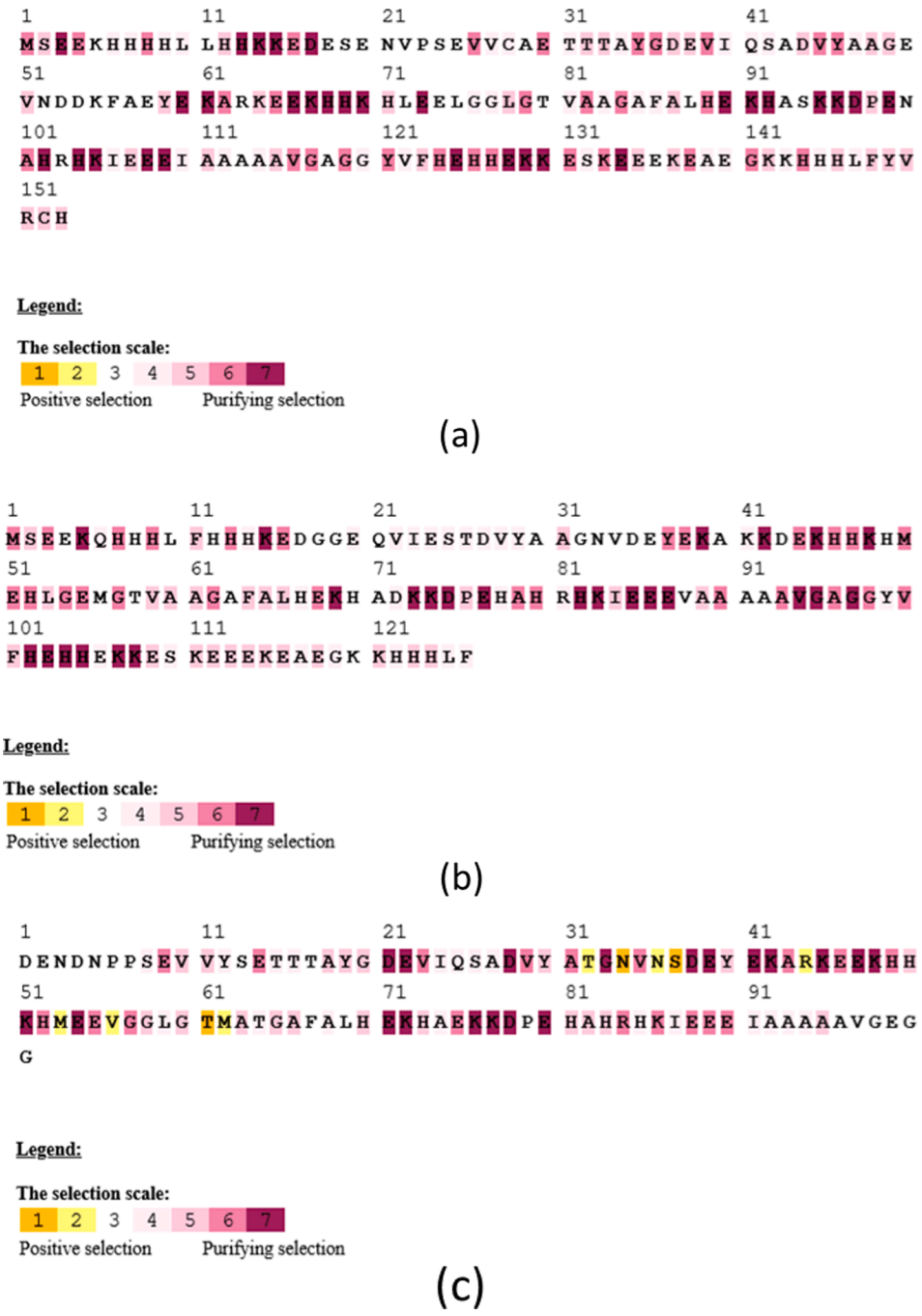

Fig. 9. Selective pressures on LP3 gene produced using Selecton: (a) LP3-0 (U67135) (b) LP3-1 (U52865) (c) LP3-3 (U59424). These specific sequences were used as references; analysis was performed using the available orthologous sequences of each gene. Yellow colours indicate sites of positive/diversifying selection; purple colours indicate sites of purifying selection. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(E). *LP3* and *ASR*, both belonging to the ABA/WDS gene family, are both highly hydrophilic protein groups therefore the presence of charged, polar amino acids is important in attaining their functional conformations (Rom et al., 2006; Gonzalez and Iusem, 2014). It has also been suggested that zinc ions bind to lysine residues in the *N*-terminal region of both sets of genes, this binding being required for the proteins to finally obtain their functional conformation, thereby further explaining why mutations affecting Lysine are purified. Evidence of purifying selection can also be observed in both the ABA/WDS domain and C-terminal containing a putative NLS/DNA binding amino acid sequence, implying that both are important for the overall functioning of the genes. This is logical when considering that expressions of both *LP3* and *ASR* are upregulated in presence of ABA and that both act as transcription factors in response to water-deficit stress (Padmanabhan et al., 1997; Wang et al., 2002). Interestingly, *LP3-3* gene is characterized by a significantly higher GC2 content than the *LP3-0* and *LP3-1* genes within the *LP3* family and highest codon usage (RCSU) for a particular codon. In this context, *LP3-3* from *Picea abies* shows the highest overall expression among all the *LP3* genes.

(Fig. 12). All together, these parameters strongly suggest this gene to have enhanced expression and possibly translation levels, which does not compete with the speculation of the acquisition or refinement of *LP3-3* gene.

5. Conclusion

Overall, our results propose strongly that while *ASR* and *LP3* may have originated from the same common ancestor, they have undergone significant shifts in codon usage, potentially due to different evolutionary constraints. Different *ASR* genes have already been studied in

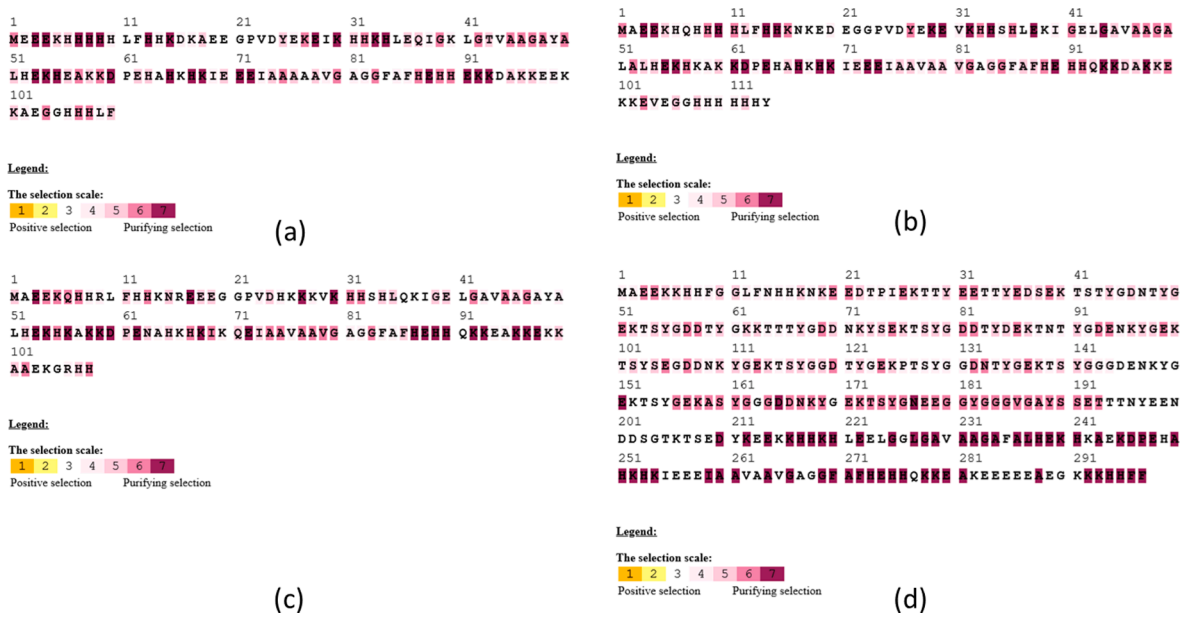

Fig. 10. Selective pressures on ARS gene produced using Selecton: (a) ASR1 (NM_001247208) (b) ASR2 (NM_001320991) (c) ASR3 (NM_001309371) (d) ASR4 (NM_001282319). These specific sequences were used as references; analysis was performed using the available orthologous sequences of each gene. Yellow colours indicate sites of positive/diversifying selection; purple colours indicate sites of purifying selection. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 11. Amino acid multiple sequence alignment (MUSCLE) of LP3-3 orthologs showing sites of diversifying selection. Sites of positive selection are indicated by an asterisk above the site position. Only partial coding sequences are available, for all the sequences included in this alignment; however all sequences contain the conserved ABA/WDS region which is prerequisite for it being a *LP3* gene, which is indicated by red box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Ka/Ks estimates between *LP3* genes in *Pinus taeda*.

		LP3-0	*LP3-1*	*LP3-2*
U67135	*LP3-0*			
U52865	*LP3-1*	0.21		
U59451	*LP3-2*	0.74	0.53	
U59424	*LP3-3*	0.35	0.28	0.74

Table 6
Ka/Ks estimates between *ASR* genes in tomato.

		ASR1	*ASR2*	*ASR3*
NM001247208	*ASR1*			
NM001320991	*ASR2*	0.15		
NM001309371	*ASR3*	0.25	0.06	
NM001282319	*ASR4*	0.16	0.1	0.21

depth, with *ASR1* being the most substantially studied of them, while *LP3* remains under-explored. Further studies into *LP3* mutants could produce similar phenotypes as observed in *ASR* mutants. The precise functions and genomic targets of LP3 could be suggested by homology with genes targeted by ASR, possibly genes involved in hexose transport. This could be done by a comparative RNASeq study of lp3 mutants (Dominguez et al., 2013; Dominguez and Carrari, 2015). LP3 may also act as a chaperone protein in the cytosol like ASR1 (Grivet et al., 2011; Neale et al., 2014), although the particular LP3 which does this precisely, will need to be experimentally verified. Precise mapping of the *LP3* genes onto the *Pinus taeda* genome should also be done. Further research could establish the presence of more paralog genes of *LP3* within *Pinus taeda* and their precise role in drought response. Since the responses of plants to cold stress are similar to their responses to drought stress, research into how *ASR* and *LP3* are affected in terms of expression and cellular function in conditions of cold stress would also be of interest. Finally, a precise expression network between *ASR/LP3* and downstream targets should be established since these have not yet been determined.

Author contributions

Curation of raw data, data analysis, visualizations (JL and SSR). Conceptualisation of the research project, funding acquisition and supervision (MRGG). All authors contributed to writing of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

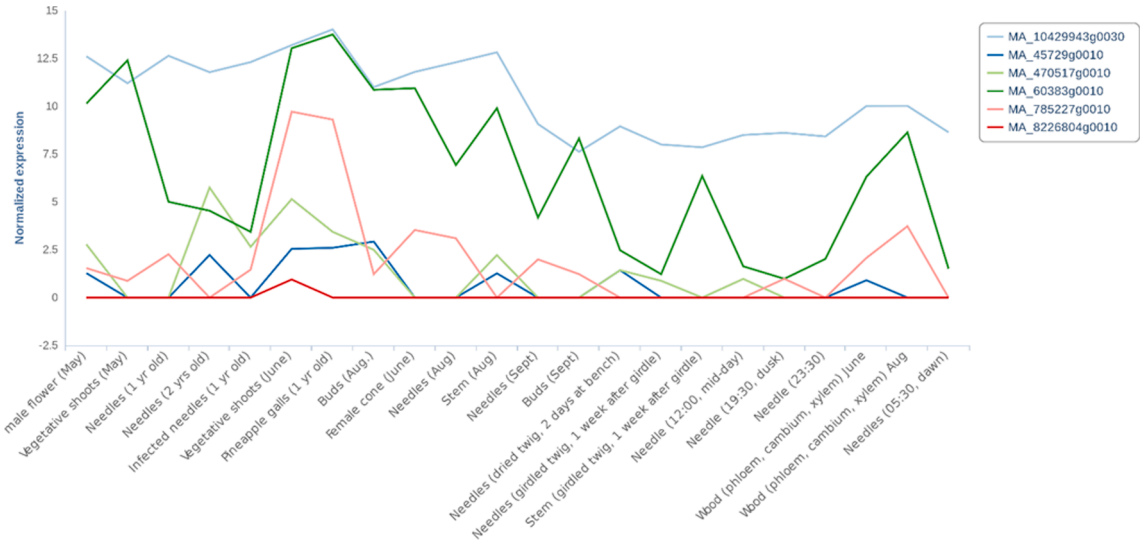

Fig. 12. Expression graph of Norway Spruce LP3 genes shows that there is a variability in expression patterns between the different LP3 genes. Accession numbers of the genes are indicated on the right of the graph.

Data availability

Data used in this analysis was taken from public databases and this has been mentioned in the article

Acknowledgments

We would like to thank Nicolas Delhomme for bioinformatics support. We acknowledge Swedish Research Council (VR) and Swedish Governmental Agency for Innovation Systems (VINNOVA) for their support.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.146935.

References

Bateman, A., Finn, R.D., 2007. SCOOP: a simple method for identification of novel protein superfamily relationships. Bioinformatics 23, 809–814.
Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y., Raytselis, Y., Sayers, E.W., Tao, T., Ye, J., Zaretskaya, I., 2013. BLAST: a more efficient report with usability improvements. Nucleic Acids Res. 41, W29–W33.
Bray, E.A., 1993. Molecular responses to water-deficit. Plant Physiol. 103, 1035–1040.
Buschiazzo, E., Ritland, C., Bohlmann, J., Ritland, K., 2012. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evol. Biol. 12.
Campbell, W.H., Gowri, G., 1990. Codon usage in higher-plants, green-algae, and cyanobacteria. Plant Physiol. 92, 1–11.
Chang, S.J., Puryear, J.D., Dias, M.A.D.L., Funkhouser, E.A., Newton, R.J., Cairney, J., 1996. Gene expression under water deficit in loblolly pine (Pinus taeda): Isolation and characterization of cDNA clones. Physiol. Plant. 97, 139–148.
Chaves, M.M., Maroco, J.P., Pereira, J.S., 2003. Understanding plant responses to drought - from genes to the whole plant. Funct. Plant Biol. 30, 239–264.
Conesa, A., Gotees, S., García-Gómez, J., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal annotation and visualization tool in functional genomics research. Application note. Bioinformatics 21, 3674–3676.
De La Torre, A.R., Li, Z., Van de Peer, Y., Ingvarsson, P.K., 2017. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Mol. Biol. Evol. 34, 1363–1377.
Dominguez, P.G., Carrari, F., 2015. ASR1 transcription factor and its role in metabolism. Plant Signaling Behav. 10.
Dominguez, P.G., Frankel, N., Mazuch, J., Balbo, I., Iusem, N., Fernie, A.R., Carrari, F., 2013. ASR1 mediates glucose-hormone cross talk by affecting sugar trafficking in tobacco plants. Plant Physiol. 161, 1486–1500.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Edwards, D., Selden, P.A., 1992. The development of early terrestrial ecosystems. Botanical Journal of Scotland 42, 337–366.
Eveno, E., Collada, C., Guevara, M.A., Leger, V., Soto, A., Diaz, L., Leger, P., Gonzalez-Martinez, S.C., Cervera, M.T., Plomion, C., Garnier-Gere, P.H., 2008. Contrasting patterns of selection at Pinus pinaster Ait. drought stress candidate genes as revealed by genetic differentiation analyses. Mol. Biol. Evol. 25, 417–437.
Frankel, N., Carrari, F., Hasson, E., Iusem, N.D., 2006. Evolutionary history of the Asr gene family. Gene 378, 74–83.
Goldgur, Y., Rom, S., Ghirlando, R., Shkolnik, D., Shadrin, N., Konrad, Z., Bar-Zvi, D., 2007. Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. Plant Physiol. 143, 617–628.
Gonzalez, R.M., Iusem, N.D., 2014. Twenty years of research on Asr (ABA-stress-ripening) genes and proteins. Planta 239, 941–949.
Gonzalez-Martinez, S.C., Ersoz, E., Brown, G.R., Wheeler, N.C., Neale, D.B., 2006. DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in Pinus taeda L. Genetics 172, 1915–1926.
Grivet, D., Sebastiani, F., Alia, R., Bataillon, T., Torre, S., Zabal-Aguirre, M., Vendramin, G.G., Gonzalez-Martinez, S.C., 2011. Molecular footprints of local adaptation in two mediterranean conifers. Mol. Biol. Evol. 28, 101–116.
Gudmundsson, L., Seneviratne, S.I., 2016. Anthropogenic climate change affects meteorological drought risk in Europe. Environ. Res. Lett. 11.
Hedges, S.B., Marin, J., Suleski, M., Paymer, M., Kumar, S., 2015. Tree of life reveals clock-like speciation and diversification. Mol. Biol. Evol. 32, 835–845.
Hurvich, C.M., Tsai, C.L., 1989. Regression and time-series model selection in small samples. Biometrika 76, 297–307.
Innan, H., Kondrashov, F., 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11, 97–108.
Iusem, N.D., Bartholomew, D.M., Hitz, W.D., Scolnik, P.A., 1993. Tomato (Lycopersicon-Esculentum) transcript induced by water-deficit and ripening. Plant Physiol. 102, 1353–1354.
Kumar, S., Stecher, G., Suleski, M., Hedges, S.B., 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. 34, 1812–1819.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
Li, N., Li, Y.Y., Zheng, C.C., Huang, J.G., Zhang, S.Z., 2016. Genome-wide comparative analysis of the codon usage patterns in plants. Genes & Genomics 38, 723–731.
Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S.N., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z.X., Yamashita, R.A., Zhang, D.C., Zheng, C.J., Bryant, S.H., 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43, D222–D226.
Mistry, J., Bateman, A., Finn, R.D., 2007. Predicting active site residue annotations in the Pfam database. BMC Bioinf. 8.
Neale, D.B., Wegrzyn, J.L., Stevens, K.A., Zimin, A.V., Puiu, D., Crepeau, M.W., Cardeno, C., Koriabine, M., Holtz-Morris, A.E., Liechty, J.D., Martinez-Garcia, P.J., Vasquez-Gross, H.A., Lin, B.Y., Zieve, J.J., Dougherty, W.M., Fuentes-Soriano, S., Wu, L.S., Gilbert, D., Marcais, G., Roberts, M., Holt, C., Yandell, M., Davis, J.M., Smith, K.E., Dean, J.F.D., Lorenz, W.W., Whetten, R.W., Sederoff, R., Wheeler, N., McGuire, P.E., Main, D., Loopstra, C.A., Mockaitis, K., deJong, P.J., Yorke, J.A., Salzberg, S.L., Langley, C.H., 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol. 15.

Padmanabhan, V., Dias, D.M.A.L., Newton, R.J., 1997. Expression analysis of a gene family in loblolly pine (Pinus taeda L.) induced by water deficit stress. Plant Mol. Biol. 35, 801–807.
Ricardi, M.M., Guaimas, F.F., Gonzalez, R.M., Burrieza, H.P., Lopez-Fernandez, M.P., Jares-Erijman, E.A., Estevez, J.M., Iusem, N.D., 2012. Nuclear import and dimerization of tomato ASR1, a water stress-inducible protein exclusive to plants. PLoS ONE 7.
Rom, S., Gilad, A., Kalifa, Y., Konrad, Z., Karpasas, M.M., Goldgur, Y., Bar-Zvi, D., 2006. Mapping the DNA- and zinc-binding domains of ASR1 (abscisic acid stress ripening), an abiotic-stress regulated plant specific protein. Biochimie 88, 621–628.
Ruosteenoja, K., Markkanen, T., Venalainen, A., Raisanen, P., Peltola, H., 2018. Seasonal soil moisture and drought occurrence in Europe in CMIP5 projections for the 21st century. Clim. Dyn. 50, 1177–1192.
Schaeffer, R.D., Liao, Y.X., Cheng, H., Grishin, N.V., 2017. ECOD: new developments in the evolutionary classification of domains. Nucleic Acids Res. 45, D296–D302.
Smarda, P., Bures, P., Horova, L., Leitch, I.J., Mucina, L., Pacini, E., Tichy, L., Grulich, V., Rotreklova, O., 2014. Ecological and evolutionary significance of genomic GC content diversity in monocots. PNAS 111, E4096–E4102.
Song, H., Gao, H.J., Liu, J., Tian, P., Nan, Z.B., 2017. Comprehensive analysis of correlations among codon usage bias, gene expression, and substitution rate in Arachis duranensis and Arachis ipaensis orthologs. Sci. Rep. 7.
Song, X.M., Ma, X., Li, C.J., Hu, J.J., Yang, Q.H., Wang, T., Wang, L., Wang, J.P., Guo, D., Ge, W.N., Wang, Z.Y., Li, M.M., Wang, Q.M., Ren, T.Z., Feng, S.Y., Wang, L.X., Zhang, W.M., Wang, X.Y., 2018. Comprehensive analyses of the BES1 gene family in Brassica napus and examination of their evolutionary pattern in representative species. BMC Genomics 19.
Stern, A., Doron-Faigenboim, A., Erez, E., Martz, E., Bacharach, E., Pupko, T., 2007. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res. 35, W506–W511.
Walsh, H.E., Kidd, M.G., Moum, T., Friesen, V.L., 1999. Polytomies and the power of phylogenetic inference. Evolution 53, 932–937.
Wang, J.T., Gould, J.H., Padmanabhan, V., Newton, R.J., 2002. Analysis and localization of the water-deficit stress-induced gene (lp3). J. Plant Growth Regul. 21, 469–478.
Xia, X.H., 2017. DAMBE6: new tools for microbial genomics, phylogenetics, and molecular evolution. J. Hered. 108, 431–437.
Xia, X.H., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26, 1–7.
Yang, C.Y., Chen, Y.C., Jauh, G.Y., Wang, C.S., 2005. A lily ASR protein involves abscisic acid signaling and confers drought and salt resistance in Arabidopsis. Plant Physiol. 139, 836–846.
Yang, Z.H., Nielsen, R., Goldman, N., Pedersen, A.M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.