Genome-Wide Identification and Analysis of Class III Peroxidases in Betula pendula

Kewei Cai
  Northeast Forestry University

Huixin Liu
  Northeast Forestry University

Song Chen
  Northeast Forestry University

Yi Liu
  Northeast Forestry University School of Forestry

Xiyang Zhao
  Northeast Forestry University

Su Chen
  State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University

Research article

Keywords: Betula pendula, class III peroxidases, phylogenetic analysis, chromosomal location, expression pattern

Abstract

Background: Class III peroxidases (POD) proteins are widely present in the plant kingdom that are involved in a broad range of physiological processes including stress responses and lignin polymerization throughout the plant life cycle. However, little is known about the *POD* genes in *Betula pendula*, although it has been characterized in *Arabidopsis*, rice, poplar, maize and Chinese pear. The *POD* genes remain to be determined in *Betula pendula*.

Results: A total of 90 nonredundant *POD* genes were identified in *Betula pendula*. (designated *BpPODs*). These *POD* genes were divided into twelve groups based on their phylogenetic relationships. The *BpPODs* are unevenly distributed on the 14 chromosomes. In addition, some *BpPODs* were located sequentially in tandem on chromosomes, inferred that tandem duplication contributes to the expansion of the *POD* gene family in *Betula pendula*. Analysis of the distribution of conserved domains of BpPOD proteins showed that all these proteins contain highly conserved motifs. We also investigated their expression patterns in different tissues, the results show that some *BpPODs* might play significant roles in root, xylem, leaf and flower. Furthermore, under low temperature conditions, some *BpPODs* showed different expression patterns at different times.

Conclusions: Comprehensive study of the *POD* genes suggests that their functional diversity during *Betula pendula* growth and development. Our findings provide a basis for further functional analysis on *POD* gene family in *Betula pendula*.

Background

Peroxidases or peroxide reductases (POD, EC number 1.11.1.x) are a large group of oxidases existing in animals, plants and microorganisms, which catalyzes the oxidation of a particular substrate by hydrogen peroxide [1]. Among them, class III peroxidases, belonging to the haem peroxidase subfamily, exist only in plants and have an extremely widespread presence in the plant kingdom [2]. The Class III peroxidase in plants are also reported as POX [3, 4], GPX [5], Prx [6], ClassIII PRX [7], and POD [8, 9]. Most plant species contain dozens of Class III peroxidases, for example, switchgrass [7] genome contains more than 200 *POD* coding genes, and *Populus* [10], rice and *Arabidopsis* contain 93, 138 and 73 members of POD family, respectively [6, 11] .

POD are involved in a broad range of physiological processes throughout the plant life cycle, probably due to the high number of enzymatic isoforms (isoenzymes) and to the versatility of their enzyme-catalysed reactions [12]. Recent studies indicate that POD has two most important functions in plants: on the one hand, it is related to the normal morphogenesis of plants and plays a role in the growth and development of plants. On the other hand, it is related to the resistance of plants, including disease resistance, cold resistance, drought resistance, *etc.*, and it is one of the important protective enzymes in plants [13, 14]. Although it is known that POD play a key role in cell growth and response to abiotic stress,

the specific function of each member of the family is still elusive. The comprehensive researches are necessary to explore the role of POD in plant growth and defense [15].

The gene family always arose from multiple ways including tandem duplication, duplicative transposition, and whole genome duplication (WGD), which was followed by mutation and divergence [16]. During the last decade, several molecular biology approaches have been developed to isolate, characterize and study the expression of *POD* gene family in plants [6]. *Betula pendula* is a pioneer boreal tree that can be induced to flower within one year [17], it plays an important role in people life [18, 19]. Up to now, however, no genome-wide characterization of the POD family in *B. pendula*. It has been shown that POD is related to the synthesis of lignin [20] and cork [21, 22], and lignin is considered as an important defense means against invasion and expansion of pathogens [23, 24]. At the same time, a large number of experimental evidences of stress treatment showed that under the stress of drought and low temperature, the expression of POD increased significantly [25, 26].

Since *Betula pendula* is a widespread species and has many applications in the pulp and paper industry, it is necessary to study its development and physiology [27]. Understanding the role of POD family in lignin synthesis and resistance to biotic and abiotic stresses in *B. pendula* will greatly contribute to its application in industrial production [28]. Fortunately, *B. pendula* has attracted much attention, particularly by the availability of its genome sequences [29], which gave us an opportunity to carry out a comprehensive genome-wide analysis for exploring the potential functions of the *POD* gene family in *B. pendula.*

In the present study, a genome-wide analysis of *POD* gene family from *B. pendula* was conducted via genomic sequence, including *BpPOD* gene models, phylogenetic relationship, conserved motif, chromosome location and other structural features [30]. we performed for the first time the comprehensive analysis to the *POD* genes involved in lignin synthesis and abiotic stress response in *B. pendula.* Our study provides important insights for further study of the potential role of *POD* gene family in *B. pendula* growth and development.

Results

Identification of *POD* genes

To identify members of *POD* family in *B. pendula*, we used the 73 *POD* genes of *Arabidopsis* to obtain the best hits in the *B. pendula* genome by BLASTP. A total of 90 putative *PODs* were identified in the *B. pendula* genome. We further examined the conserved domains of proteins encoded by these genes using Pfam [31] and SMART [32] databases. The results revealed that all the genes have classical *POD* domain structures, which demonstrate the reliability of the results. The *B. pendula* genome contains more *PODs* than *Arabidopsis* (73) [6], but fewer than *Populus trichocarpa* (93) [33], *Pyrus bretschneideri* (94) [30], and rice (138) [11]. We defined the *BpPODs* as *BpPOD1* to *BpPOD90*. The isoelectric point (PI) varied from 4.28 to 9.6 with a mean of 7.25 and >7.0 of 52.2% POD proteins. In addition, subcellular locations of these *BpPODs* are mainly in the cytoplasm, cell membrane, vacuole, chloroplast and nucleus. Their

detailed information, including chromosome location, gene name, subcellular location and molecular weight (MW) gene size of each *BpPOD* gene/protein, was listed in **Table 1**.

Phylogenetic analyses of POD proteins in *B. pendula*

To investigate the evolutionary history and phylogenetic relationships among the members of POD family in *B. pendula*, a phylogenetic tree was constructed with the Neighbor-Joining method based on multiple sequence alignment of the 90 BpPODs, with 1000 bootstrap replicates (**Figure 1**). The BpPOD proteins were divided into twelve major subgroups with high bootstrap probabilities, designated group I to group XII. The *POD* genes of each subgroup is unevenly distributed, with the number of members varies from 4 to 15. Subgroup VIII contains the most members (15), subgroup X, XI, XII contains the least number of members, with only 4 members.

Gene structures

To further gain insights into the structural diversity of the *POD* genes, we subsequently performed exon-intron analysis in *BpPODs*. The result reveals several variations (**Figure 2**), including five *BpPODs* (*BpPOD9, BpPOD11, BpPOD16, BpPOD57 and BpPOD61*) lacking an intron in their gene structures. In the remaining *BpPODs*, the number of introns varies from one to six, while the major members have one to three introns. In addition, *BpPOD76 and BpPOD87* have the most introns (6), followed by *BpPOD24* and *BpPOD51* (5). Furthermore, genes in the same subfamily display similar exon/intron structures. For example, *BpPOD20, BpPOD22 and BpPOD82* have three exons and two intron, both of which belong to Group V; *BpPOD73 and BpPOD74* have two exons and one intron, both of which belong to Group XI [34].

Analysis of conserved amino acid motifs

To understand the functional regions of BpPODs, conserved amino acid motifs analyses of BpPOD proteins were performed. A total of eight conserved amino acid motifs were identified in the BpPOD proteins (**Figure 3**). All BpPOD proteins contain at least one conserved amino acid motif. For example, BpPOD55 only contains motif 8, BpPOD83 contains motif 1 and 7, while BpPOD10 proteins contain all the eight conserved amino acid motifs.

Most of the closely related members have the same motif compositions, suggesting that there are functional similarities between POD proteins within the same subgroup [35]. We found that motifs 1, 2, 3, 4, 5 and 7 appeared in nearly all members of BpPOD proteins, these motifs might be important for the functions of BpPOD proteins.

Chromosomal location and evolution analysis of *BpPODs*

To investigate the genome organization and distribution of *BpPODs* on different chromosomes of *B. pendula*, a chromosome map was constructed. The results show that the 90 *BpPODs* were distributed among 14 chromosomes, as shown in **Figure 4**, the physical locations of these *BpPODs* on chromosomes were scattered and uneven. Chromosome 1 and 8 contains the most *BpPODs* (14), followed by chromosome 13 (10). Eight *BpPODs* were simultaneously distributed on chromosomes 5 and 7, whereas chromosomes 14 had only one and chromosomes 11 does not include the *POD* gene. In addition, some chromosomes exhibit a relatively high density of *BpPODs*, such as the bottoms of chromosomes 13 and the top of chromosome 8.

Gene duplication, including segmental and tandem duplication, is considered to be one of the primary driving forces in the evolution of genomes [36, 37]. In this study, among the 90 *BpPODs* identified, a large number of *BpPODs* have the same duplicated regions (**Figure 5**). Generally, a gene cluster is the result of gene tandem duplication [38]. In this study, we found that some *BpPODs* were adjacent to each other (**Figure 4**). For instance, *BpPOD17-20*, *BpPOD22-29* and *BpPOD11-15* were located sequentially in tandem on chromosomes 5, 8, and 13, respectively, implying that these genes might arise from recent tandem duplication events [39]. The result indicated that tandem duplications play main contributors in the expansion of the *BpPOD* gene family. The result was consistent with the previously reported for *Populus trichocarpa POD* genes, tandem duplications also contributed significantly to the expansion of *POD* gene family in *Populus trichocarpa* [33]. However, in previous studies, many species also have produced some different results. For example, segmental duplication events were the major contributors to the expansion of the pear *POD* family [30] and segmental duplication and tandem duplication were identified in maize *POD* family [35]. These results indicate that there are significant differences in the *POD* genes expansion pattern in *B. pendula*, maize and Chinese pear, which strongly implied that POD family members have different expansion patterns among different species. It may be the reason why the *POD* family members (90) in *B. pendula* were less than those in the maize (119) [35].

To explore the selection pressures among *BpPOD* duplicated genes, we calculated the *Ka*, *Ks* and *Ka/Ks* values for the 23 gene pairs (**Table 2**). In the general, Ka/Ks > 1 indicates positive selection, Ka/Ks = 1 indicates neutral selection and Ka/Ks < 1 indicates negative or purifying selection [34]. The Ka/Ks ratios of most *BpPOD* gene pairs were < 1, suggesting that these gene pairs evolved under negative selection in *B. pendula.* The results of this Ka/Ks analysis suggest that negative selection was vital to the functional divergence of *BpPODs.*

To further elucidate the evolution mechanisms of *BpPODs* family, we constructed the comparative syntenic maps of birch associated with three representative species containing *Arabidopsis thaliana*, *Populus trichocarpa* and *Vitis vinifera*. (**Figure 6; Table 3**). A total of 17 (*Arabidopsis thaliana*), 49 (*Populus trichocarpa*) and 43 (*Vitis vinifera*) orthologous pairs were found between birch and the other species, among these gene pairs, some *BpPOD* genes (*BpPOD3*, *BpPOD7*, *BpPOD16*, *BpPOD21*, *BpPOD40*, *BpPOD4*, *BpPOD48*, *BpPOD52*, *BpPOD57* and *BpPOD84*) were shown to have

collinear relationships with all of the above three species. Interestingly, two or more *POD* genes from *Arabidopsis thaliana*, *Populus trichocarpa* and *Vitis vinifera* matched one birch *POD* gene. These genes, which are probably paralogous gene pairs, may have played a crucial role in the expansion of the *POD* gene family during evolution. For example, AT1G24110.1.TAIR10, AT3G28200.1.TAIR10 and AT5G40150.1.TAIR10 are orthologous to *BpPOD16*, Potri.006G107000.1.v4.1 and Potri.016G132800.1.v4.1 are orthologous to *BpPOD22*, VIT_206s0004g01180.1 and VIT_208s0007g06650.1 are orthologous to *BpPOD41* (**Table 3**).

Tissue-specific expression of *BpPODs*

To better understand the functions of *POD* genes in the growth and development of *Betula Platyphylla × Betula Pendula*, their expression profiles in different tissues (including root, xylem, young leaf and flower) were analyzed with publicly available transcriptome data. Of the 90 *BpPODs*, 69 genes were expressed in one or more birch tissues, while 21 *BpPOD* genes exhibited no expression in various individual tissues (Relative expression value > 0 as basal expression) [40]. The heat map (**Figure 7**) demonstrated that most *BpPODs* had tissue-specific or preferential expression patterns. As shown in **Figure 7**, *BpPOD6*, *BpPOD21* and *BpPOD37* were highly expressed in xylem. Several *BpPODs* were expressed in root during development, such as *BpPOD62, BpPOD63* and *BpPOD65*. *BpPOD78* and *BpPOD19* showed higher expression levels in young leaf and flower, respectively. The expression level of *BpPOD6* was high in xylem and low in root, leaf and flower. In contrast, *BpPOD67, BpPOD68, BpPOD80* and *BpPOD81* had no expression in any of the investigated tissues. *BpPOD21, BpPOD59* and *BpPOD62* were highly expressed in developing xylem, root, leaf and flower. In conclusion, the variations in the expression of *BpPODs* in different tissues revealed that *POD* genes may be involved in several processes during *B. pendula* growth and development.

Responses of *BpPODs* expression to cold treatment

Several roles have been attributed to plant peroxidases in response to biotic and abiotic stresses [41]. In recent years, the number of studies on *POD* genes response to abiotic stress have been reported [35]. For example, *Arabidopsis* overexpressing *AtPOD3* showed an increase in dehydration and salt tolerance, whereas the antisense suppression of *AtPOD3* exhibited dehydration and salt sensitive phenotypes [42]. The expression of *POD* genes is induced by various environmental stresses, such as metal, pathogens, humidity, temperature, anoxia and potassium deficiency [41], suggesting that *POD* genes are involved in plant defense. In this study, we examined the expression levels of the *BpPODs* in response to low temperature stress. As shown in **Figure 8**, the result indicated that the expression of *BpPODs* was altered under cold treatment, some of *BpPODs* are induced but most of them not or slightly induced. After cold treatment, the expression levels of *BpPOD4, BpPOD13, BpPOD15, BpPOD17* and *BpPOD21* were significantly induced at a relatively early stage (0.5 h after treatment), and with the increase of cold

treatment time, the relative expression level of these genes was also at a high level. The **Figure 8** shows that the expression levels of *BpPOD19*, *BpPOD21, BpPOD39* and *BpPOD47* were increased after 1.5 h treatment of low temperature. *BpPOD50* and *BpPOD58* did not respond to cold treatment at the beginning (0.5 h), and were slightly increased after 2 h exposure to low temperature. In addition, other genes are also induced by cold stress, such as *BpPOD14*, *BpPOD16*, *BpPOD59*, etc. In general, the low temperature responsive *BpPODs* may play important roles in birch under cold stress.

Validation of DEGs identified by RNA-seq using qRT-PCR

We used qRT-PCR (Quantitative Real-time PCR) to validate the expression profiles of *BpPODs* after low temperature treatment (6°C). A total of 6 cold-responsive *BpPODs* were randomly selected for qRT-PCR analysis. As shown in **Figure 9**, the qRT-PCR results indicated that all the 6 *BpPODs* were up-regulated by low-temperature stress, which was consistent with the results derived from the RNA-seq data. *BpPODs* including *BpPOD15*, *BpPOD47* and *BpPOD49* showed the highest transcript level when exposed to a low temperature for 1.5h. *BpPOD4*, *BpPOD17* and *BpPOD26* showed the highest transcript level at 3h. In general, this qRT-PCR result supports the reliability of the RNA-seq analysis.

Discussion

It is reported that members of Class III Peroxidases gene family are involved in the regulation of a variety of processes [6, 33], and play a key role in biological and abiotic stress responses during plant growth and development [35]. Systematic and comprehensive analyses of *POD* gene families have been published for *Arabidopsis thaliana* [6], *Populus trichocarpa* [33], *Zea mays* [35] and *Oryza sativa* [11], but a genome-wide study of the POD family has not previously been reported in *B. pendula*. The published genome data of *B. pendula* [29] provides a useful tool for analysis of the *POD* gene family in *B. pendula*.

In the present study, 90 non-redundant *POD* genes were identified in *B. pendula*, the number is higher than that in *Arabidopsis* (73), but lower than that *Populus trichocarpa* (93) and *Pyrus bretschneideri* (94), which indicates that the *POD* genes in *B. pendula* have expanded compared to those in *Arabidopsis*. Subsequently, we performed analyses of the phylogenic relationships, gene structure, chromosomal locations, conserved motifs and expression profiles [43].

In the process of genome evolution, tandem duplication and segmental duplication were the main factors that led to the expansion of gene family [37]. Certain studies have shown that tandem duplication was largely responsible for the expansion of   birch gene family [35], such as tandem duplication are the main reason for the expansion of *B. pendula* NAC gene family [16]. By contrast, segmental duplication has contributed significantly to the expansion of this gene family in *Pyrus bretschneideri* [30]. Interestingly, in this study, we found that some *BpPODs* were adjacent to each other, suggesting tandem duplications play main contributors to the expansion of the *BpPOD* gene family. However, in maize, the segmental duplication and tandem duplication almost identically contributed to the *POD* gene family expansion [35].

These results also explain why the number of *POD* genes in *B. pendula* (90), pear and *populus* were less than those in maize (119). According to the above analysis, we speculated that the expansion of the *POD* gene family differed between different plants. In addition, the Ka/Ks < 1 of the most *BpPOD* duplicated pairs showed that negative selection may be largely responsible for maintaining the functions of birch POD enzymes. Furthermore, the Ka/Ks ratios for the *BpPOD5/-6*, *BpPOD24/-25* and *BpPOD24/-27* gene pairs are relatively high, suggesting that these genes experienced rapid evolutionary diversification following duplication. We also constructed the comparative syntenic maps of birch associated with *Arabidopsis thaliana*, *Populus trichocarpa* and *Vitis vinifera*. We identified 49 orthologous gene pairs between *Populus trichocarpa* and birch, while only 17 orthologous gene pairs between *Arabidopsis thaliana* and birch were found, perhaps due to the closer relationship between *Populus trichocarpa* and birch *versus Arabidopsis thaliana* and birch.

The 90 BpPOD proteins possess ten highly conserved motifs. MEME analysis revealed that different conserved motifs are present in each of the BpPOD proteins. Notably, most BpPOD proteins contain all the conserved motifs, while only a few BpPOD proteins contain one or two motifs, which means that these motifs may be involved in the important basic function of the POD protein. However, a few motifs with unknown functions are present in nearly every subgroup, these motifs might play important roles in the BpPOD family. Additionally, gene structures were also investigated in *BpPODs*. We found that 90 *BpPODs* contained different numbers of exons or introns, with most *BpPODs* containing more than two introns, indicating that there is a great diversity in *POD* gene family of *B. pendula*. Many studies have shown that structural diversification of genes plays an important role in the evolution of multi-gene families [44, 45]. In our research, we found that they presented different characteristics of *BpPODs* from different subfamilies, suggesting that *BpPODs* members are functionally diversified. In addition, many studies have reported that introns could be specifically inserted into the plants and were retained in the genome during the course of evolution [10, 46]. Therefore, we inferred that loss or gain of introns may be caused by specific approach.

Gene expression patterns are an important aspect of the study of gene function [47]. High-throughput microarray technology provides a good platform for the study of genome-wide gene expression patterns [45]. We used publicly available genome-wide transcript profiling data from *B. pendula* tissues as a resource to investigate the expression patterns of *BpPODs*. Most *BpPODs* exhibited variable expression patterns, suggesting functional diversification of *BpPODs*. Twenty-one *BpPODs* exhibited no expression in four individual tissues, indicating that *BpPODs* are expressed under specific conditions or at specific developmental stages. In this study, we found that of the 90 *BpPODs*, the most abundant expression was in the root, followed by the xylem. The result showed that most highly expressed *POD* genes might play significant roles in root. The highest expression levels of *BpPOD6, BpPOD21* and *BpPOD37* genes were found in xylem. It was suggested that these three genes were participated in regulation of the xylem synthesis in *B. pendula. BpPOD59* is most expressed in flowers and leaves, suggesting that it may be related to leaf spreading and flowering formation in *B. pendula.* In addition, several *BpPODs* were expressed in all tissues, suggesting that they might play basic roles in *B. pendula*. In conclusion, the

expression profiling in this study provides an important basis for further studying of expression and biological functions of the *POD* gene family in *B. pendula.*

The growth and development of plants are usually affected by abiotic stress, such as drought, low temperature and high salinity [45]. A lot of stress-related genes were induced to adapt to these abiotic stresses [48, 49]. A large number of experimental studies [35] on stress treatment showed that under the stress of low temperature and other conditions, *POD* genes expression increased significantly [25, 26]. However, no *POD* genes responding to cold treatment have been reported in *B. pendula*. Thus, we performed a survey of the expression patterns of the *POD* genes in *B. pendula* under cold treatment. The results suggested that some of *BpPODs* are induced but most of them not or slightly induced. A small number of *BpPODs* were highly expressed at a relatively early stage (0.5 h after treatment), and with the extension of time, the expression reached the highest level. This indicated that these genes might play an important role in *B. pendula*. By contrast, the expression level of *BpPOD30* and *BpPOD8* gradually increased at 2 h after treatment, indicating that these genes are involved in the late reaction of cold treatment. In addition, the expression of a few *BpPODs* decreased under cold treatment, we speculate that these genes may also have defense and other specific functions in *B. pendula.* These results indicated that most *POD* genes were induced by low temperature and might contribute to the defense against abiotic stresses in *B. pendula*.

Conclusion

In short, a total of 90 *POD* genes were identified in *B. pendula* and divided into twelve major subgroups. A total of eight conserved amino acid motifs were identified in the BpPOD proteins. Chromosomal location and microsynteny analysis suggested that these *BpPODs* were unevenly distributed in fourteen chromosomes. Tandem duplication were identified as the main patterns contributors to the expansion of *POD* genes expansion in *B. pendula.* Finally, expression patterns analysis revealed that some *BpPODs* might play significant roles in root, xylem, leaf and flower. Furthermore, under low temperature conditions, some *BpPODs* showed different expression patterns at different times. This present study increases our understanding of POD genes in *B. pendula* and lays the foundation for further clarify of the biological functions of these POD proteins in other plants.

Methods

Identification of *B. pendula* peroxidase genes

To identify *B. pendula* peroxidase genes, the *B. pendula* genome sequences were downloaded [29]. The local BLAST database was established with protein sequences of the *B. pendula* whole genome. The protein sequences of POD family members in the genome of *Arabidopsis* were retrieved from the TAIR database. The candidate sequences of POD in *B. pendula* were acquired by BLASTP with each of the 73 different amino acid sequences of *Arabidopsis* POD gene family as query sequences. To verify the reliability of the results, all the acquired candidate sequences were examined for the presence of the POD

domain using PFAM [50] and SMART [51]. Finally, all candidate *POD* sequences were aligned using ClustalW [52] and checked manually to remove potentially redundant genes, and all of the non-redundant *POD* genes were used for further analysis. The theoretical molecular weights (MWs) and isoelectric points (pIs) of the BpPOD protein sequences were analyzed by the ExPaSy Compute pI/Mw tool [53].

Phylogenetic analyses of *B. pendula* peroxidase genes

To investigate the phylogenetic relationships of the peroxidase genes of *B. pendula*, a phylogenetic tree was constructed. Prior to phylogenetic analysis, multiple sequence alignments were generated using MUSCLE [54]. Subsequently, the RAxML [55] was employed to construct an unrooted phylogenetic tree based on alignments using the Neighbor-Joining (NJ) method with the following parameters: model (p-distance), bootstrap (1000 replicates), and gap/missing data (pairwise deletion).

Gene structure and conserved motif analysis

The genomic and CDS sequences of *B. pendula* PODs, extracted from *B. pendula* genome databases, were compared by using the Gene Structure Display Server [56] program to infer the exon/intron organization of *POD* genes. To determine conserved motifs structures of the BpPOD proteins , the conserved motifs were detected using the online MEME Tool [57]. The conserved motifs were analyzed with the SMART and PFAM programs.

Analysis of chromosomal location

Information about the physical locations of all *POD* genes on chromosomes was obtained from the Phytozome database. According to *BpPODs* starting positions on the birch chromosomes, TBtools software was used to determine the chromosome location image of the *BpPODs* [58]. In addition, to explore the selection pressures among *POD* duplicated genes, we calculated the nonsynonymous mutation rate (Ka), synonymous mutation rate (Ks), and Ka/Ks values for the duplicated gene pairs with TBtools [58]. The ratio of nonsynonymous to synonymous nucleotide substitutions (Ka/Ks) between paralogs was analyzed to detect the mode of selection. Syntenic maps of birch associated with three representative species were visualized by MCScanX [59].

Differential expression profile of *BpPOD* gene family

To determine the expression patterns of *BpPODs* in different tissues in *Betula platyphylla* × *Betula pendula*, we downloaded the sequencing data from the NCBI SRA database with an accession number of

PRJNA535361 [16]. To identify the expression of *BpPODs* during cold treatment in *Betula platyphylla × Betula pendula*, we designed a time series experiment including six time points, according to Yang's method [60]. Two-month-old *Betula platyphylla × Betula pendula* plants grown in the greenhouse of Northeast Forestry University at a constant temperature of 25°C with a photoperiod of 16 h of light and 8 h of dark were exposed to low temperatures (6°C) for 0.5 h, 1 h, 1.5 h, 2 h, 2.5 h, and 3 h, respectively. After exposure to low temperature, all young leaves were harvested at the same time to avoid falsely differentially expressed genes caused by diurnal changes. Plants without low temperature treatment were used as the control for this study. We generated three biological replicates for time points of 0.5h, 1 h, 1.5h, 2 h, 2.5 h, 3 h, and CK (control check). Total RNA of the samples were extracted using the traditional CTAB method [61]. The constructed cDNA libraries were sequenced using the Illumina HiSeq platform at Biomarker Technologies Corporation (Beijing, China). We can downloaded the sequencing data from the NCBI SRA database with an accession number of PRJNA532995 [16]. The clean reads of each sample were obtained by filtering out reads of low quality. All the clean reads were aligned to the *B. pendula* reference genome[29] using bowtie2 [62]. The RNA-seq (RNA-sequencing) data were analyzed using the RSEM (RNA-seq by Expectation-Maximization) pipeline [63] and the data were processed using a paired-end sequencing mode. RSEM [63] could compute transcript abundance, estimating the number of RNA-seq fragments corresponding to each gene, and normalized expression values as TMM (trimmed mean of M-values).

Quantitative RT-PCR

The two month old plants of birch were treated with low temperature stress (6°C) for 1.5 h and 3 h, and three biological replicates were prepared at each time point. Total RNA of young leaves was then isolated for the qRT-PCR experiment. Total RNA was extracted from the young leaves with the help of Trizol (Invitrogen) according to the manufacturer·s instructions. Then, the first-strand cDNA synthesis was performed with the resulting RNA using a PrimeScript RT Master Mix Perfect Real-Time kit (TAKARA, Japan). Quantitative real-time RT-PCR was performed on an ABI 7500 Real-Time system (Applied Biosystems). Primer pairs for real-time quantitative PCR were designed using A plasmid Editor v1.11 (**Table 4**). Each reaction contained 10 μL of THUNDERBIRD SYBR qPCR Mix (QPS-201, TOYOBO, Japan), 2 μL of cDNA, 1 μL of forward primer, 1 μL of reverse primer, and 6 μL of double-distilled water (ddH2O). The thermal cycle was used as follows: 95°C for 10 min, which was followed by 45 cycles of 95°C for 30 s and 60°C for 10s [35]. The relative mRNA level for each gene was calculated using the $2^{-\Delta\Delta CT}$ method. The real-time PCR experiment was carried out at least three times under identical conditions [16].

Abbreviations

POD: class III peroxidases *B. pendula*: *Betula pendula* *BpPODs*: *POD* genes in *Betula pendula* RNA-seq: RNA sequencing

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article.

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the National Natural Science Foundation of China (31870659), the Fundamental Research Funds for the Central Universities (2572019CG08) and Heilongjiang Touyan Innovation Team Program (Tree Genetics and Breeding Innovation Team).

Authors' contributions

CKW was a major contributor in writing the manuscript. LHX was a major contributor to the revision of the manuscript language. CS analyzed the data and make figures. LY participated in RNA extraction and performed RT-qPCR assay. ZXY participated in the design of the study and analyzed data. CS conceived of the study, participated in its design and data interpretation, and revised the manuscript critically.

Acknowledgements

Not applicable

References

1. Dunford HB, Stillman JS: **On the Function and Mechanism of Action of Peroxidases**. *Coordination Chemistry Reviews* 1976, **19**(3):187–251.

2. Passardi F, Theiler G, Zamocky M, Cosio C, Rouhier N, Teixera F, Margis-Pinheiro M, Ioannidis V, Penel C, Falquet L: **PeroxiBase: The peroxidase database**. *Phytochemistry* 2007, **68**(12):1605-1611.

3. Mei W, Qin Y, Song W, Li J, Zhu Y: **Cotton GhPOX1 encoding plant class III peroxidase may be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation**. *Journal of Genetics and Genomics* 2009, **36**(3):141-150.

4. Martínez-Rubio R, Acebes JL, Encina A, Krknen A: **Class III peroxidases in cellulose deficient cultured maize cells during cell wall remodeling**. *Physiol Plant* 2018, **164**(1).

5. Zhang Z, Xin W, Wang S, Zhang X, Wang Q: **Xylem sap in cotton contains proteins that contribute to environmental stress response and cell wall development**. *Funct Integr Genomics* 2014, **15**(1):17-26.

6. Tognolli M, Penel C, Greppin H, Simon P: **Analysis and expression of the class III peroxidase large gene family in Arabidopsis thaliana**. *Gene* 2002, **288**(1-2):0-138.

7. Moural TW, Lewis KM, Barnaba C, Zhu F, Kang CH: **Characterization of Class III Peroxidases from Switchgrass**. *Plant Physiology* 2016, **173**(1).

8. Delannoy E, Jalloul LA, Assigbetsé K, Marmey P, Nicole M: **Activity of Class III Peroxidases in the Defense of Cotton to Bacterial Blight**. *Molecular Plant Microbe Interactions* 2003, **16**(11):1030-1038.

9. Rácz A, Hideg É, Czégény G: **Selective responses of class III plant peroxidase isoforms to environmentally relevant UV-B doses**. *Journal of Plant Physiology* 2018, **221**:101-106.

10. **Comparative genomic analysis of the WRKY III gene family inpopulus, grape,arabidopsisand rice**. *Biology Direct* 2015, **10**(1):48.

11. Passardi F, Longet D, Penel C, Dunand C: **The class III peroxidase multigenic family in rice and its evolution in land plants**. *Phytochemistry* 2004, **65**(13):1879-1893.

12. Passardi F, Cosio C, Penel C, Dunand C: **Peroxidases have more functions than a Swiss army knife**. 2005.

13. Hiraga S, Sasaki K, Ito H, Ohashi Y, Matsui H: **A large family of class III plant peroxidases**. *Plant and Cell Physiology* 2001, **42**(5):462-468.

14. Ritonga FN, Chen S: **Physiological and Molecular Mechanism Involved in Cold Stress Tolerance in Plants**. *Plants* 2020, **9**(5):560.

15. Hu Y, Peuke AD, Zhao X, Yan J, Li CJPP, Biochemistry: **Effects of simulated atmospheric nitrogen deposition on foliar chemistry and physiology of hybrid poplar seedlings**. 2019, **143**.

16. Chen S, Lin X, Zhang D, Li Q, Chen S: **Genome-Wide Analysis of NAC Gene Family in Betula pendula**. *Forests* 2019, **10**(9):741.

17. Alonso-Serra J, Safronov O, Lim KJ, Fraser-Miller SJ, Blokhina OB, Campilho A, Chong SL, Fagerstedt K, Haavikko R, Helariutta Y: **Tissue-specific study across the stem reveals the chemistry and transcriptome dynamics of birch bark**. *New Phytologist* 2019, **222**.

18. Hedrick UP: **Sturtevants Edible Plants of the World**: Dover Publications; 1919.

19. Johnson CP, Sowrby JE: **useful plants of Great Britain**. 1899.

20. Huystee RBV: **Some Molecular Aspects of Plant Peroxidase Biosynthetic Studies**. *Annual Review of Plant Biology* 2003, **38**(1):205-219.

21. Mittler R, Zilinskas BA: **Molecular-Cloning And Characterization Of A Gene Encoding Pea Cytosolic Ascorbate Peroxidase**. *Journal of Biological Chemistry* 1992, **267**(30):21802-21807.

22. Christensen JrH, Bauw G, Welinder KG, Montagu MV, Boerjan W: **Purification and Characterization of Peroxidases Correlated with Lignification in Poplar Xylem**. *Plant Physiology* 1998, **118**(1):125-135.

23. Lewis NG: **Lignin: Occurrence, Biogenesis And Biodegradation**. *Annu Rev Plant Physiol Plant Mol Biol* 1990, **41**(1):455-496.

24. Agostini E, Forchetti SMD, Tigier HA: **Production of peroxidases by hairy roots of Brassica napus**. *Plant Cell Tissue & Organ Culture* 1997, **47**:177-182.

25. Botella MA, Quesada MA, Kononowicz AK, Bressan RA, Valpuesta V: **Characterization and in situ localization of a salt-induced tomato peroxidase mRNA**. *Plant Molecular Biology* 1994, **25**(1):105-114.

26. Chittoor JM, Leach JE, White FF: **Differential Induction of a Peroxidase Gene Family During Infection of Rice by Xanthomonas oryzae pv. oryzae**. *Molecular Plant Microbe Interactions* 1997, **10**(7):861-871.

27. Zhang R, Yang C, Wang C, Wei Z, Xia D, Wang Y, Liu G, Wang Y: **Time-course analyses of abscisic acid level and the expression of genes involved in abscisic acid biosynthesis in the leaves of Betula platyphylla**. *Molecular biology reports* 2012, **39**(3):2505-2513.

28. Hu Y, Chen BJM: **Arbuscular mycorrhiza induced putrescine degradation into γ-aminobutyric acid, malic acid accumulation, and improvement of nitrogen assimilation in roots of water-stressed maize plants**. 2020, **30**:329-339.

29. Salojrvi J, Smolander OP, Nieminen K, Rajaraman S, Kangasjrvi J: **Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch**. *Nature Genetics* 2017, **49**(6):904.

30. Cao Y, Han Y, Meng D, Li D, Jin Q, Lin Y, Cai Y: **Structural, Evolutionary, and Functional Analysis of the Class III Peroxidase Gene Family in Chinese Pear (Pyrus bretschneideri)**. *Frontiers in Plant Science* 2016, **7**:1874-1874.

31. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J: **The Pfam protein families database**. *Nucleic Acids Research* 2004, **28**(1):263-266.

32. Ivica L, Tobias D, Peer B: **SMART 7: recent updates to the protein domain annotation resource**.

33. Ren LL, Liu YJ, Liu HJ, Qian TT, Qi LW, Wang XR, Zeng QY: **Subcellular Relocalization and Positive Selection Play Key Roles in the Retention of Duplicate Genes of Populus Class III Peroxidase Family**. *Plant Cell* 2014, **26**(6):2404-2419.

34. Duan P, Wang G, Chao M, Zhang Z, Zhang B: **Genome-Wide Identification and Analysis of Class III Peroxidases in Allotetraploid Cotton (Gossypium hirsutum L.) and their Responses to PK Deficiency**. *Genes* 2019.

35. Wang Y, Wang Q, Zhao Y, Han G, Zhu S: **Systematic analysis of maize class III peroxidase gene family reveals a conserved subfamily involved in abiotic stress response**. *Gene* 2015, **566**(1):95-108.

36. Moore, Richard, C., Purugganan, Michael, D.: **The early stages of duplicate gene evolution**. *Proceedings of the National Academy of Sciences of the United States of America* 2003.

37. Cannon SB, Mitra A, Baumgarten A, Young ND, May G: **The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana**. *Bmc Plant Biology* 2004, **4**:1-21.

38. Riechmann J, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe O, Samaha R: **Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes**. *Science* 2000, **290**(5499):P.2105-2110.

39. Elsheery NI: **Genome-wide characterization of aspartic protease (AP) gene family in Populus trichocarpa and identification of the potential PtAPs involved in wood formation**. *BMC Plant Biology* 2019, **19**(276):1-17.

40. Ning K, Ding C, Huang Q, Zhang W, Su XJPO: **Transcriptome profiling revealed diverse gene expression patterns in poplar (Populus × euramericana) under different planting densities**. 2019, **14**(5):e0217066-.

41. Claudia C, Christophe D: **Specific functions of individual class III peroxidase genes**. *Journal of Experimental Botany* 2008(2):2.

42. Llorente F, López-Cobollo RM, Catalá R, Martínez-Zapater JM, Salinas J: **A novel cold-inducible gene from Arabidopsis , RCI3 , encodes a peroxidase that constitutes a component for stress tolerance**. *Plant Journal* 2002, **32**(1):13-24.

43. Hu X, Liu C, Tian J, Zhang Y, Xin Q, Chen A, Li D, Liu X: **Identification, molecular characterization, and expression analysis of the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) gene family in Betula platyphylla Suk**. *Trees* 2020, **34**(1):229-241.

44. Mehanathan M, Rohit K, Bhan YC, Suresh BV, Yusuf K, Manoj P, Pandey GK: **Identification and Molecular Characterization of MYB Transcription Factor Superfamily in C4 Model Plant Foxtail Millet (Setaria italica L.)**. *Plos One* 2014, **9**(10):e109920-.

45. Han Y, Ding T, Su B, Jiang H: **Genome-Wide Identification, Characterization and Expression Analysis of the Chalcone Synthase Family in Maize**. *International Journal of Molecular Sciences* 2016, **17**(2):161-.

46. Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV: **Remarkable Interkingdom Conservation of Intron Positions and Massive, Lineage-Specific Intron Loss and Gain in Eukaryotic Evolution**. *Current Biology Cb* 2003, **13**(17):1512-1517.

47. Oakley T, Ostman B, Wilson A: **Repression and loss of gene expression outpaces activation and gain in recently duplicated fly genes**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(31):11637-11641.

48. **The calcium sensor CBL1 integrates plant responses to abiotic stresses**. *Plant Journal* 2003, **36**(4):457-470.

49. Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K: **Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor**. *Nature*

Biotechnology 1999, **17**(3):287-291.

50. Finn RD, Mistry J, Schuster-Bckler B, Griffiths-Jones S, Bateman A: **PFAM: clans, web tools and services**. *Nucleic Acids Research* 2006, **34**(Database issue):D247-251.

51. Letunic I, Doerks TT, Bork P: **SMART 6: Recent updates and new developments**. *Nuclc Acids Research* 2008, **37**(Database issue):D229-232.

52. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic acids research* 1994, **22**(22):4673-4680.

53. Elisabeth G, Alexandre G, Christine H, Ivan I, Appel RD, Amos B: **ExPASy: the proteomics server for in-depth protein knowledge and analysis**. *Nuclc Acids Research* 2003, **31**(13):3784-3788.

54. Publications M: **MUSCLE: multiple sequence alignment with high accuracy and high throughput – Edgar 32 (5): 1792 – Nucleic Acids Research**. *My Publications* 2010.

55. Alexandros, Stamatakis: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics* 2014.

56. Guo AY, Zhu QH, Chen X, Luo JC: **GSDS: A gene structure display server**. *Hereditas* 2007, **29**(8):1023-1026.

57. Bailey TL, Elkan C: **The value of prior knowledge in discovering motifs with MEME**. *Proceedings / International Conference on Intelligent Systems for Molecular Biology; ISMB International Conference on Intelligent Systems for Molecular Biology* 1995, **3**:21-29.

58. Chen C, Chen H, He Y, Xia R: **TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface**. *BioRxiv* 2018:289660.

59. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler BS, Guo H: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity**. *Nucleic Acids Research* 2012, **40**(7).

60. Chuanping, Yang, Hairong, Plant WJM: **Designing Microarray and RNA-Seq Experiments for Greater Systems Biology Discovery in Modern Plant Genomics**. 2015.

61. Chang S, Puryear J, Cairney JJPMBR: **A simple and efficient method for isolating RNA from pine trees**. 1993, **11**(2):113-116.

62. Langmead B, Salzberg SL: **Langmead B, Salzberg SL.. Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359**. *Nature Methods* 2012, **9**(4):357-359.

63. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome**. *Bmc Bioinformatics* 2011, **12**(1):1-16.

Tables

Due to technical limitations, the tables are available as downloads in the supplementary files section.

Figures
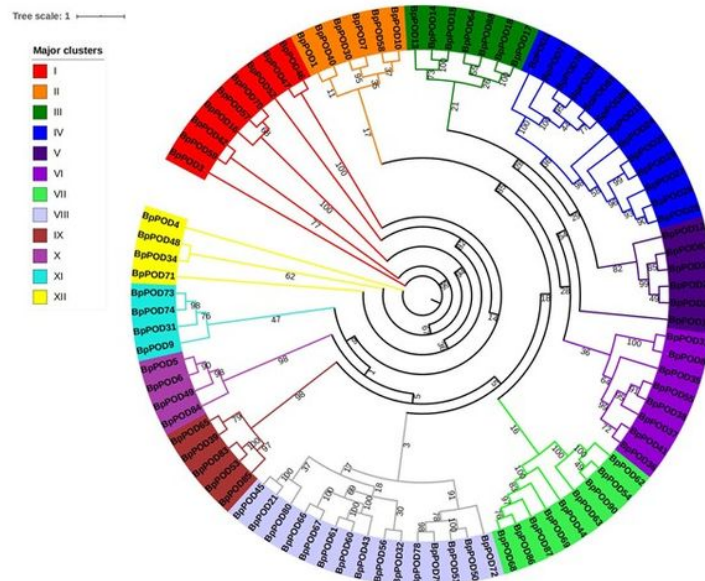
Figure 1

Figure 1

Phylogenetic relationship of the 90 identified BpPOD genes. Unrooted tree constructed using RAxML by the Neighbor-Joining (NJ) method. Bootstrap values from 1000 replicates are indicated at each node. The tree shows 12 major phylogenetic subgroups (subgroups I to XII).

Figure 2

Exon–intron structures of the 90 identified BpPOD genes. Exon-intron structure was generated using Gene Structure Display Server (GSDS). The exons and introns are indicated by yellow cylinder bars and black lines, respectively. The scale at the bottom is in kilobases.
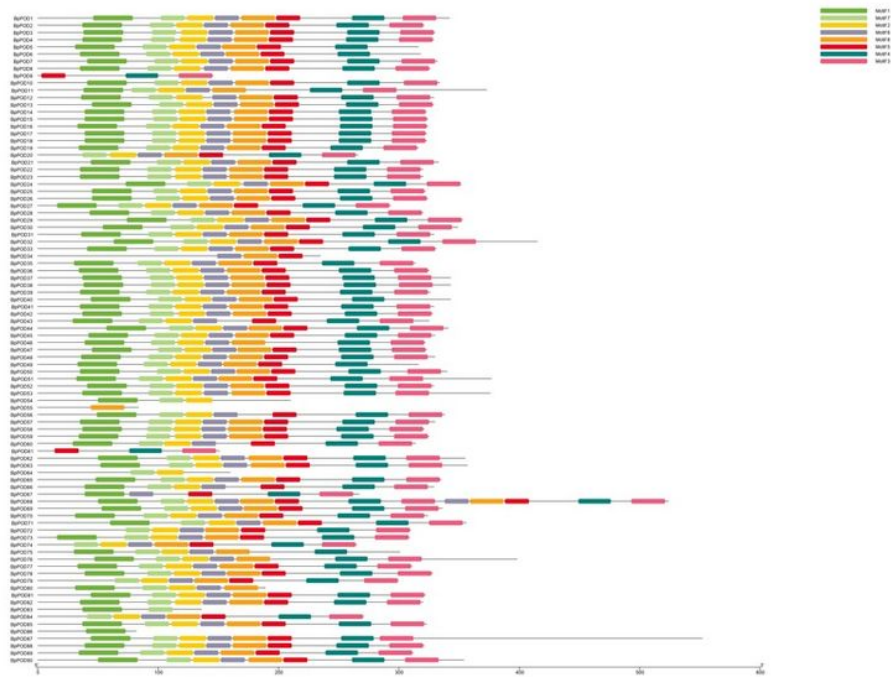
Figure 3

Distribution of eight putative conserved motifs in BpPOD proteins. Motifs in the putative BpPOD proteins were obtained using the Multiple Em for Motif Elicitation (MEME) web server. Conserved motifs are represented by different colored boxes while nonconserved sequences are shown by gray lines. Note that the length of each box does not represent the actual motif size, and the colored boxes were ordered manually according to the results of MEME analysis.
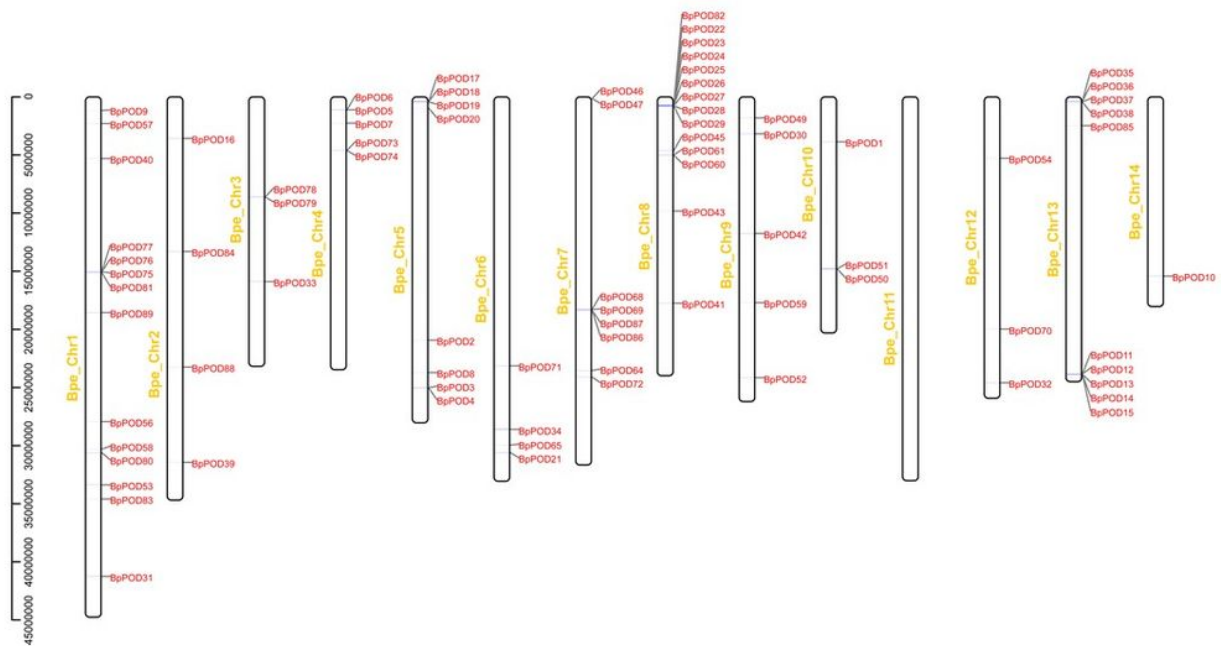
Figure 4

Figure 4

Chromosomal locations of 90 POD genes on 14 B. pendula chromosomes. According to BpPODs starting positions on the birch chromosomes, TBtools software was used to determine the chromosome location image of the BpPODs. Each was mapped to the chromosome based on its physical location. The number of chromosomes (chr01-chr14) is marked in yellow. The gene names on the right side of each chromosome correspond to the approximate locations of each POD gene.
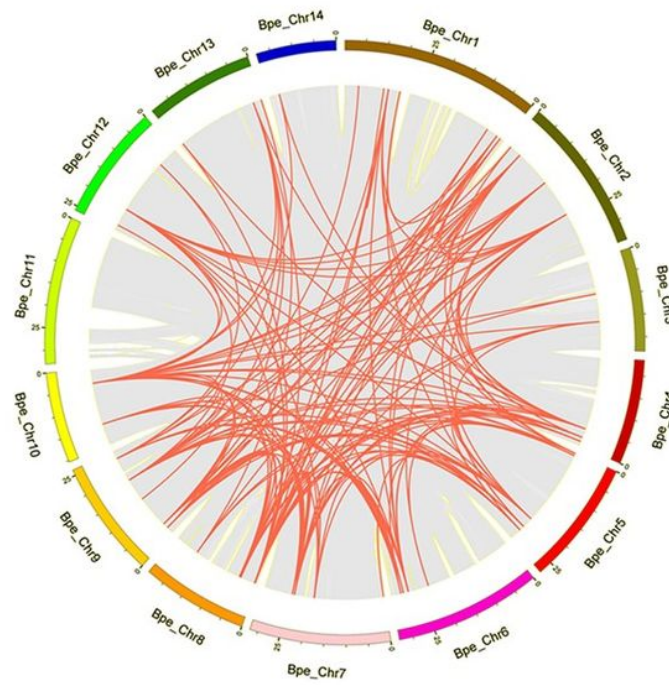
Figure 5

Figure 5

BpPODs genomic distribution and collinear relationships. A total of 90 BpPODs were disproportionately mapped on the Betula pendula linkage groups using TBtools software. Red lines represent all homologous blocks in Betula pendula genome.
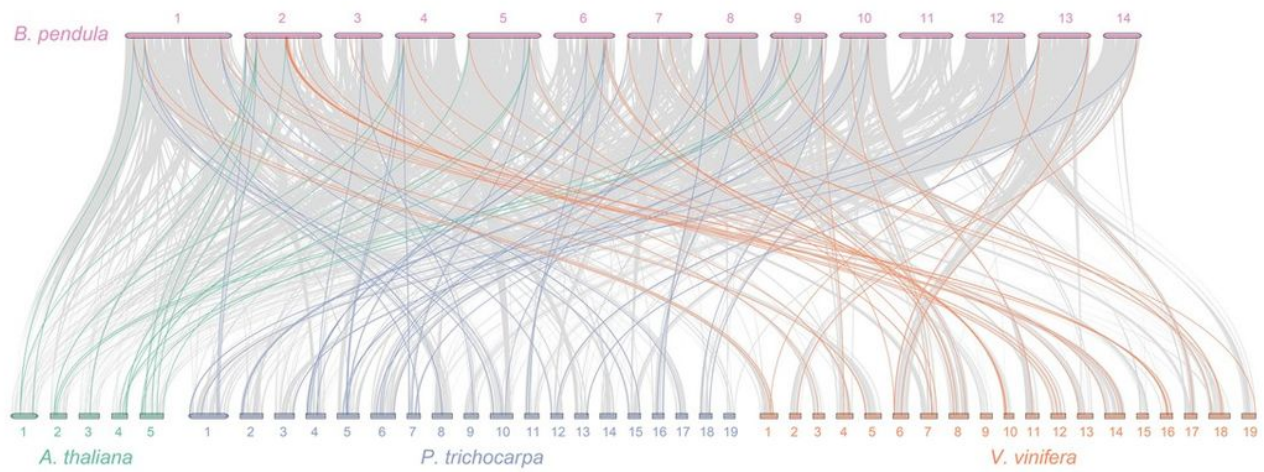
Figure 6

Figure 6

Syntenic relationships of POD genes between B. pendula and three representative species. Gray lines indicate all syntenic blocks among birch linkage groups or between birch and the other species. In three species, collinear pairs of BpPODs are connected by green, blue and orange lines. Species: Arabidopsis thaliana, Populus trichocarpa and Vitis vinifera.
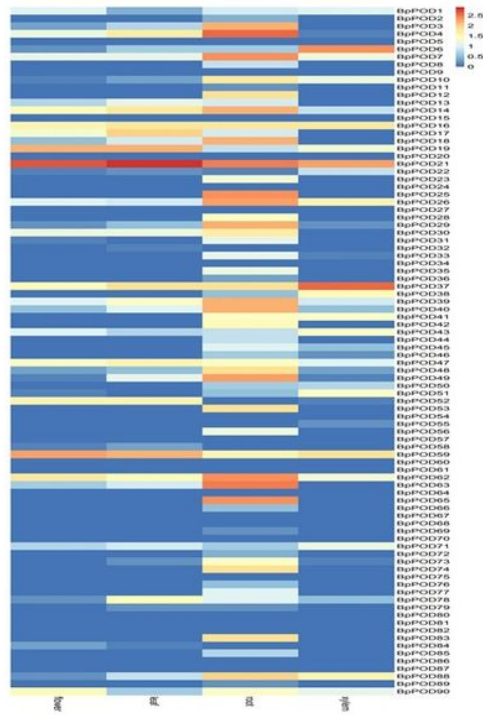
Figure 7

Figure 7

Expression profiles of BpPOD genes across different tissues. Different organs/tissues are exhibited below each column. The BpPOD genes were listed at the right of the expression array, and the colour box from blue (0) to orange (2.5) indicate an increased expression level is shown at the right of the figure. Color scale represent the normalized value of the expression. For the sake of unified comparison, the normalized value of the expression was log10 transformed.
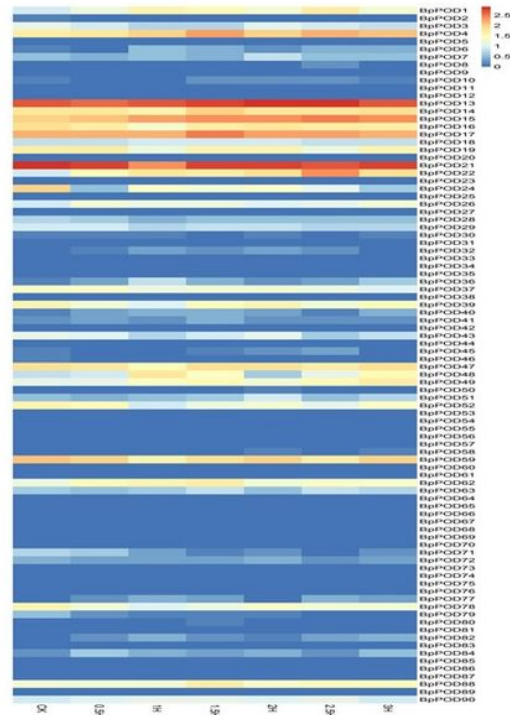
Figure 8

Figure 8

Responses of BpPOD genes expression to cold treatment. Different time are exhibited below each column. The BpPOD genes were listed at the right of the expression array, and the colour box from blue (0) to orange (2.5) indicate an increased expression level is shown at the right of the figure. Color scale represent the normalized value of the expression. For the sake of unified comparison, the normalized value of the expression was log10 transformed.
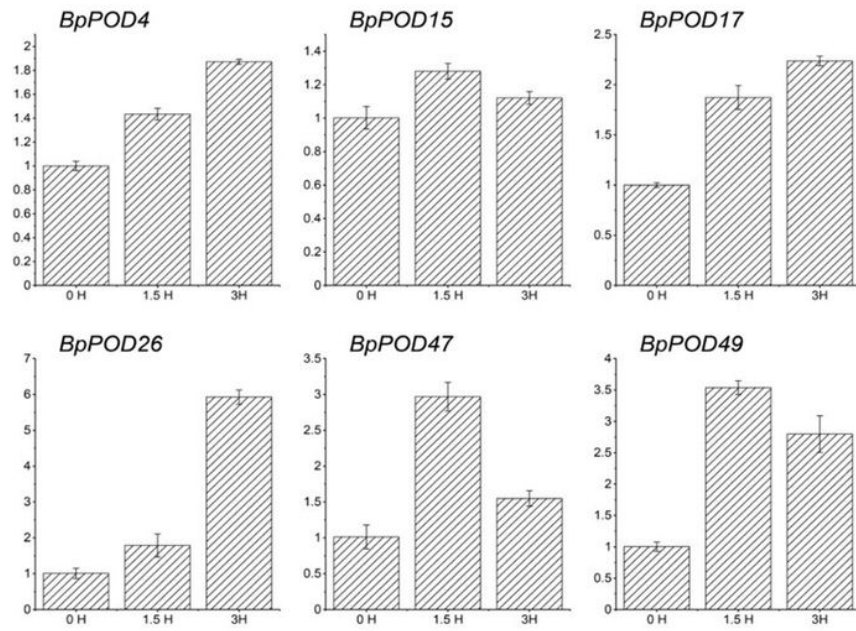
Figure 9

Figure 9

Expression patterns of six stress-responsive BpPOD genes under cold stress treatments. Relative expression levels of BpPOD genes in response to cold stress treatments were examined by RT-qPCR and normalized expression values use TMM. The y-axis represents the relative expression level and the x-axis represents the time course of stress treatment. Seedlings were sampled at 0, 1.5 and 3 after cold treatment. Data represent means ± SD in three replicates.