Genome-Wide Identification and Analysis of Cell Cycle Genes in *Betula pendula*

Yijie Li¹, Song Chen¹, Yuhang Liu¹, Haijiao Huang*¹

¹ State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin, China

* Correspondence: Haijiao Huang, e-mail: haijiao_sea@163.com

**Abstract:** Research Highlights: This study identified the cell cycle genes in birch that likely play important roles during plant growth and development. This provides a basis for understanding the regulatory mechanism of various cell cycles in *Betula pendula*. Background and Objectives: The cell cycle factors not only influence cell cycle progression together, but also regulate accretion, division and differentiation of cells, and then regulate growth and development of plant. In this study, we identified the putative cell cycle genes in *B. pendula* genome, based on the annotated cell cycle genes in *A. thaliana*. It could serve as a foundation for further functional studies. Materials and Methods: The transcript abundance was determined for all the cell cycle genes in xylem, root, leaf and flower tissues using RNA-seq technology. Results: We identified 59 cell cycle gene models in the genome of *B. pendula*, 17 highly expressed genes among them. These genes were *BpCDKA1*, *BpCDKB1.1*, *BpCDKB2.1*, *BpCKS1.2*, *BpCYCD1.1*, *BpCYCD3.1*, *BpCYCD3.5*, *BpDEL1*, *BpDpa2*, *BpE2Fa*, *BpE2Fb*, *BpKRP1*, *BpKRP2*, *BpRb1* and *BpWEE1*. Conclusions: We identified 17 core cell cycle genes in the genome of birch by combining phylogenetic analysis and tissue specific expression data.

**Keywords:** *Betula pendula*; cell cycle; Cyclin; RNA-seq

---

1. Introduction

Many important life processes are closely related to mitosis in higher organisms. The regulation mechanism of eukaryotic cell division cycle is one of the hot topics in cell biology and molecular biology. Research on the regulation of plant cell cycle started later than that of mammals and yeast. Great progress has been made in the research of cell cycle in higher plant in recent years [1-4]. The procedure of cell cycle is the result of interaction between the gene expression and the external factors. The cell cycle in higher plant is strictly regulated in the course of its growth and development.

The concept of cell cycle was brought forward by Howard and Pelcin 1953, which was divided into the intermitotic phase (G1, S, and G2) and mitotic phase (M). Growth and development of plant depended on accretion, division and differentiation of cells, while cell cycle involved into these process. Recent studies have shown that, during regulation of hormone, nutriment substance and other growth signals, Cyclin D (CYCD) was expressed firstly, and binds to cyclin dependent kinase A (CDKA) to form a complex. The complex is activated by the action of CDK activating kinase (CAK) and cyclin-dependent kinase inhibitor (CKI) or KIP-related proteins (KRPs). The activated complex attenuates the inhibitory effect of retino blastomaprotein-related (RBR) and E2F (E2 factor) a-b/DP through phosphorylation, and release transcript factor E2Fa-b/DP[5]. While E2F/DPs could promote the expression of genes required for G1 conversion to S phase (DNA synthesis phase). After entering the S phase, CYCA binds to CDKA, and it was combined with CDK subgroup cyclin-dependent kinase subunit (CKS) and CYCB synthesized during the development to G2 phase. To remove the inhibitory phosphate group from the tyrosine phosphatase, activate the CDKB, and enter the M phase. At the end of M phase, cyclin proteins are hydrolyzed through the anaphas-promoting complex (APC) protein pathway, and exit the mitosis. A whole cell cycle is completed [6,7].
Since the cell cyclins have been found in sea urchins by Hunt in the 1980s, tremendous advances have been made in the molecular mechanisms of the cell cycle. This provides a positive direction for the study of tumors and other physiological diseases caused by cell cycle regulation[8]. The most significant molecular structure feature of cyclin is its conserved domain sequence, known as cyclin box, which consists of about 100 amino acid residues. The cyclin framework is the core structure of cyclin. During the cell cycle, specific cyclins rely on their own unique cyclin frames to recognize specific cyclin-dependent kinase (CDK), and form a complex with it, thus showing specific CDK kinase activity[9]. Many different cyclins have been found, which have different expression patterns in different organs, tissues, and cell types of various organisms[10].

*Betula pendula* is a pioneer boreal tree that can be induced to flower within one year [11,12]. As an important timber tree, understanding how cell cycle genes regulate growth and development in *B. pendula* will greatly contribute to its use in industrial production and ornamental. Fortunately, the genome sequence of *B. pendula* [11] has become available in the last few years, which can help us to accurately identify the genes related to cell cycle. In this study, we identified cell cycle genes that likely play a very important role during plant growth and development. This provides a basis for understanding the expression processes and regulatory mechanism of various cell cycles in *B. pendula*, and may serve as a foundation for further functional studies.

### 2. Materials and Methods

#### 2.1. Identification of *B. pendula* cell cycle genes and physical and chemical properties analysis

The *Betula pendula* genome was used for the identification of the cell cycle genes according to the previous publication[13]. The putative cell cycle genes were first identified by BLASTP [14] with threshold of E-value less than 1e-5 using *A. thaliana* cell cycle genes as queries. In addition, all the genes were further manually examined using the Conserved Domain Database of NCBI [15] to confirm if they were correctly annotated. We then divided them into eight subgroups based on their functional type in *A. thaliana*. Then, we used ExPASy-ProtParam Tool (http://web.expasy.org/protparam/) to determine the physical and chemical parameters of cell cycle genes, including the number of amino acids, molecular weight and isoelectric point (PI).

#### 2.2. Chromosome distribution of *B. pendula* cell cycle genes

According to the starting position of *B. pendula* cell cycle gene on the birch chromosome, use the TBtools software to determine the chromosome location image of the birch cell cycle gene.

#### 2.3. Phylogenetic analyses of *B. pendulacell* cycle genes, Gene structure and Conserved sequence and specific motif analysis

To investigate the phylogenetic relationships of the cell cycle genes of *B. pendula*, a phylogenetic tree was constructed for every subgroup according to the previous publication[16]. We performed a multiple sequence alignment. Then, the phylogenetic trees of each subgroup were built using MEGA 5.05 with 500 bootstrap trials. Representative trees were selected using the Neighbor-Joining method.

In order to understand the structural diversity of *B. pendula* cell cycle genes, we performed exon/intron analysis. In order to understand the functional regions of birch cell cycle proteins and analyze the structural differences of birch cell cycle genes, we used the online software MEME (Multiple Em for Motif Elicitation, Version 5.4.1, http://meme-suite.org/tools/meme) to analyze birch cells. The conserved amino acid motifs of cyclins were analyzed. And use TBtools for visualization. The CDS sequence of Betula platyphylla was extracted from the genomic structure information of the genome(https://phytozome-next.jgi.doe.gov/report/gene/Bplatyphylla_v1_1), and its intron and exon structure were visualized with TBtools.
2.4. RNA-seq expression analysis of *B. pendula* cell cycle genes

To investigate the expression patterns of *B. Pendula* cell cycle genes in different tissues, transcriptome data (PRJNA535361) was downloaded from the public database of NCBI SRA. 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 using bowtie2. The RNA-seq (RNA-sequencing) data were then analyzed using the RSEM (RNA-seq by Expectation-Maximization) pipeline and the data were processed using a paired-end sequencing mode. RSEM could compute transcript abundance, estimating the number of RNA-seq fragments corresponding to each gene. Draw a heat map of the differential expression of cell cycle genes with the value of Log2 (TPM+1).

3. Results.

3.1. Identification of *B. pendula* cell cycle genes and physical and chemical properties analysis

The annotated genes in *B. pendula* genome were used to identify putative cell cycle genes, based on the annotated cell cycle genes in *A. thaliana*. In total, 59 gene models (Table 1) were identified as putative cell cycle genes in *B. pendula* genome. The 59 genes contain 15 cyclin-dependent kinases (CDKs), 2 cyclin-dependent kinase subunit (CKSs), 27 Cyclins (CYCs), 3 E2 factor (E2Fs), 2 DP-E2F-like (DELS), 4 KIP-related proteins (KRPs), 2 Rbs, and 2 WEEs, respectively. Among these cell cycle genes, CYC is the largest family that contains 27 members, while CKS, DEL, Rb and WEE are all the smallest families containing only two members. Rb and WEE are also the smallest families in *A. Thaliana* containing only one member. Analysis of protein characteristics showed that the size of the cell cycle gene protein ranges from 69 amino acids (*Bpev01.c0457.g0045*) to 1316 amino acids (*Bpev01.c1113.g0001*), and the relative molecular mass ranges from 7 kDa to 14 kDa. The predicted isoelectric point also varies greatly from 4.42 (*Bpev01.c0579.g0010*) to 9.69 (*Bpev01.c1061.g0010*), which indicates that different cyclins may work in different microenvironments. The detailed information of the protein molecular weight, isoelectric point and amino acid number of the gene family are shown in Table 1.

Table 1. Putative cell cycle genes in *Betula pendula*.

| Gene family | Gene name | Gene ID  | Deduced number of amino acids | Molecular weight (Da) | Isoelectric point (pl) | Instability index | Grand average of hydropathicity |
|-------------|-----------|----------|-------------------------------|-----------------------|------------------------|--------------------|-------------------------------|
| CDK         | CDKA.1    | *Bpev01.c0957.g0013* | 295 | 33777.93 | 6.42 | 39.45 | -0.247 |
| CDKB.1      | *Bpev01.c0224.g0013* | 305 | 34519.94 | 8.16 | 30.49 | -0.272 |
| CDKB.2      | *Bpev01.c0480.g0058* | 319 | 36190.12 | 9.04 | 30.26 | -0.297 |
| CDKC.1      | *Bpev01.c0000.g0179* | 515 | 57319.57 | 9.22 | 44.26 | -0.810 |
| CDKC.2      | *Bpev01.c0275.g0056* | 649 | 71959.85 | 9.11 | 47.61 | -0.579 |
| CDKC.3      | *Bpev01.c0344.g0012* | 721 | 80003.75 | 9.28 | 47.51 | -0.657 |
| CDKC.4      | *Bpev01.c0349.g0031* | 698 | 77679.08 | 9.30 | 43.51 | -0.563 |
| CDKC.5      | *Bpev01.c0420.g0019* | 563 | 62760.24 | 9.36 | 54.61 | -0.681 |
| CDKC.6      | *Bpev01.c0745.g0005* | 711 | 79499.18 | 9.30 | 51.47 | -0.664 |
| CDKC.7      | *Bpev01.c1061.g0010* | 711 | 79560.55 | 9.69 | 48.78 | -0.634 |
| CDKC.8      | *Bpev01.c1202.g0053* | 568 | 63441.50 | 9.63 | 51.02 | -0.575 |
| CDKD.1      | *Bpev01.c1443.g0002* | 415 | 46691.88 | 9.36 | 36.70 | -0.391 |
| CDKE.1      | *Bpev01.c0263.g0012* | 111 | 12348.09 | 6.03 | 34.85 | -0.374 |
| CDKE.2      | *Bpev01.c0390.g0015* | 478 | 53271.81 | 9.30 | 41.51 | -0.461 |
| DEL      | Bpev01.c0389.g0056 | 474 | 53297.39 | 4.51 | 53.09 | -0.434 |
|---------|-------------------|-----|----------|------|-------|--------|
| Rb1     | Bpev01.c0118.g0029 | 498 | 56182.49 | 8.17 | 49.57 | -0.364 |
| Rb2     | Bpev01.c0706.g0005 | 238 | 27110.95 | 5.35 | 52.20 | -0.202 |
| E2Fa    | Bpev01.c1588.g0004 | 493 | 54391.90 | 6.43 | 56.98 | -0.220 |
| E2Fb    | Bpev01.c0167.g0006 | 521 | 59705.00 | 8.99 | 48.34 | -0.263 |
| E2Fc    | Bpev01.c0207.g0010 | 491 | 55055.52 | 8.63 | 46.18 | -0.243 |
| DEL1    | Bpev01.c1398.g0012 | 365 | 41875.90 | 5.20 | 61.96 | -0.336 |
| DEL2    | Bpev01.c1588.g0005 | 514 | 56762.03 | 8.19 | 46.44 | -0.234 |
| KRP1    | Bpev01.c1764.g0001 | 361 | 40479.12 | 9.29 | 39.11 | -0.247 |
| KRP2    | Bpev01.c1028.g0001 | 381 | 43109.82 | 8.83 | 43.20 | -0.355 |
| KRP3    | Bpev01.c1764.g0001 | 361 | 40479.12 | 9.29 | 39.11 | -0.247 |
| CYC1.1  | Bpev01.c1009.g0008 | 459 | 50545.59 | 9.00 | 38.21 | -0.207 |
| CYC1.2  | Bpev01.c0645.g0033 | 427 | 47430.69 | 8.73 | 50.72 | -0.264 |
| CYC1.3  | Bpev01.c0022.g0129 | 435 | 49791.13 | 5.39 | 50.14 | -0.365 |
| CYC1.4  | Bpev01.c0455.g0011 | 394 | 45186.84 | 4.82 | 46.83 | -0.117 |
| CYC1.5  | Bpev01.c0134.g0104 | 435 | 49391.85 | 5.63 | 48.64 | -0.269 |
| CYC1.6  | Bpev01.c1259.g0013 | 221 | 26057.57 | 6.39 | 32.87 | 0.011 |
| CYC1.7  | Bpev01.c0848.g0042 | 325 | 36316.28 | 5.31 | 61.70 | -0.215 |
| CYC1.8  | Bpev01.c0157.g0019 | 382 | 43607.70 | 5.19 | 62.70 | -0.238 |
| CYC1.9  | Bpev01.c0506.g0013 | 128 | 13598.62 | 9.30 | 71.59 | 0.009 |
| CYC1.10 | Bpev01.c0106.g0013 | 141 | 14557.40 | 9.10 | 89.39 | 0.343 |
| CYC1.11 | Bpev01.c0229.g0031 | 140 | 14728.15 | 7.89 | 58.46 | 0.184 |
| CYC1.12 | Bpev01.c0015.g0054 | 374 | 42291.38 | 5.08 | 64.20 | -0.111 |
| CYC1.13 | Bpev01.c0640.g0020 | 374 | 42444.17 | 5.22 | 52.89 | -0.295 |
| CYC1.14 | Bpev01.c0469.g0009 | 309 | 35275.72 | 6.03 | 44.03 | -0.081 |
| CYC1.15 | Bpev01.c1653.g0004 | 352 | 40349.92 | 9.27 | 53.39 | 0.023 |
| CYC1.16 | Bpev01.c1947.g0006 | 520 | 59565.08 | 8.40 | 40.98 | -0.418 |
| CYC1.17 | Bpev01.c1113.g0001 | 1316 | 148157.79 | 6.70 | 47.53 | -0.523 |
| CDKF.1  | Bpev01.c0389.g0056 | 474 | 53297.39 | 4.51 | 53.09 | -0.434 |
| CYC1.1  | Bpev01.c0118.g0029 | 498 | 56182.49 | 8.17 | 49.57 | -0.364 |
| CYC1.2  | Bpev01.c0706.g0005 | 238 | 27110.95 | 5.35 | 52.20 | -0.202 |
| CYC1.3  | Bpev01.c1588.g0004 | 493 | 54391.90 | 6.43 | 56.98 | -0.220 |
| CYC1.4  | Bpev01.c1398.g0012 | 365 | 41875.90 | 5.20 | 61.96 | -0.336 |
| CYC1.5  | Bpev01.c1588.g0005 | 514 | 56762.03 | 8.19 | 46.44 | -0.234 |
| CYC1.6  | Bpev01.c1764.g0001 | 361 | 40479.12 | 9.29 | 39.11 | -0.247 |
| CYC1.7  | Bpev01.c1028.g0001 | 381 | 43109.82 | 8.83 | 43.20 | -0.355 |
| CYC1.8  | Bpev01.c1009.g0008 | 459 | 50545.59 | 9.00 | 38.21 | -0.207 |
| CYC1.9  | Bpev01.c0645.g0033 | 427 | 47430.69 | 8.73 | 50.72 | -0.264 |
| CYC1.10 | Bpev01.c0022.g0129 | 435 | 49791.13 | 5.39 | 50.14 | -0.365 |
| CYC1.11 | Bpev01.c0455.g0011 | 394 | 45186.84 | 4.82 | 46.83 | -0.117 |
3.2. Chromosome distribution of cell cycle genes in B. pendula

Based on the genomic information of B. pendula, the chromosomal distribution of 59 B. pendula cell cycle genes was simply analyzed. According to chromosome location analysis, 59 B. pendula cell cycle genes are unevenly distributed on 14 chromosomes of birch (Figure 1). Chromosome 11 contains the most cell cycle genes (9), followed by chromosome 6 (8). There are 6 cell cycle genes on chromosomes 1 and 3, and only 1 cell cycle gene on chromosomes 2, 8 and 12.

3.3. CDK

There are many regulators of cell cycle in plants, most of them have special serine/threonine protein kinase activity, because they bind to cyclins to function, and are named as cyclin dependent kinases (CDKs). According to their structural and functional similarities with animal and yeast CDKs and their conserved PSTAIRE domains that bind cyclins, plant CDKs were divided into 8 groups: CDKA, CDKB, CDKC, CDKD, CDKE, CDKF, CDKG and CDKLIKE [4, 18]. In this study, we identified 5 groups: BpCDKA, BpCDKB, BpCDKC, BpCDKD, BpCDKE and BpCDKF. CDKA.1 plays a key role in the process of leaf cell division and differentiation and the development of leaf [19]. CDKB1.1 can lengthen hypocotyl cells, promote cotyledon cell development, and regulate stomatal development of Arabidopsis thaliana [2, 20]. The mutation of CDKB2 has been shown to impact meristem seriously [21].

We identified 15 BpCDKS in the B. pendula genome, Constructed a phylogenetic tree for the BpCDKs (Figure 2a). Reveal the evolutionary relationships within these groups. Seven different conserved domains and special motifs of BpCDKS protein were located using MEME tool (Figure 2c). All BpCDKs proteins contain at least one conserved amino acid motif. For example, BpCDKE1.1 only contains motif 2, while the rest of BpCDKs proteins contain 1, 2, and 3 conserved amino acid motifs. The conserved motifs of each BpCDKs protein branch are similar in composition, indicating that these members have a close evolutionary relationship [22]. In addition, most members of the BpCDKs protein contain motif 1, motif 2, motif 3, and motif 6 conservative motifs, these motifs may have an important influence on the function of BpCDKs protein. The gene structure helps to further understand the gene family. In the BpCDK family, there are at most 13 introns (BpCDKC1.1 and
**BpCDKE1.2**, and at least one intron (BpCDKC1.8 and BpCDKE1.1). Most genes in the BpCDKs family contain 7-8 introns (Figure 2b), and the fact that most members of the same subfamily share a similar exon/intron structure strengthens the observed phylogenetic distribution.

![Figure 2](image)

**Figure 2.** Phylogenetic analysis, exon/Intron genomic structure and protein motif organization of CDK in B. pendula. The phylogenetic tree was made by MEGA5.05 using the neighbor-joining algorithm, and the confidence was evaluated by 500 bootstraps. Exons and introns of the BpCDKs genes are represented by yellow boxes and grey lines respectively. Use TBtools software to compare gene structure. Use MEME Web server to analyze the distribution of conserved motifs in BpCDKs protein. And visualized by TBtools software. (a) Phylogenetic analysis of BpCDKs; (b) Exon/Intron genomic structure of BpCDKs; (c) Protein motif organization of BpCYCs.

### 3.4. CYC

Monomeric CDKs have no kinase activity and must associate with regulatory proteins called cyclins to be activated. There is common molecular structure among various cyclins, which contain a rather conservative amino acid sequence called cyclin frame to mediate the binding to CDK and regulate the activity of CDK. In plant, cyclins can be grouped into M-cyclin (containing A- and B-type cyclins) and G1- specific cyclins (designated D-type cyclins). C-cyclin and H-cyclin have been confirmed, and only CYCH.1 could activate CDK[23].

All four types of cyclins known in plants were identified. A total of 27 BpCYCs genes were detected in the B. pendula genome, for B. pedula, nine A-type, six B-type, eleven D-type, and one H-type cyclins have been identified. And build an evolutionary tree. The MEME tool was used to locate five different conserved amino acid motifs of the CYC protein (Figure 3c). All BpCYCs proteins contain at least one conserved amino acid motif. For example, BpCYCD3.4, BpCYCD3.2, and BpCYCD3.3 only contain motif 2, BpCYCA1.2 only contains motif 3, and most of the other BpCYCs proteins contain motifs 1, 2, 3, and 4 conservative amino acid motifs, indicating that these motifs may have an important influence on the function of BpCYCs protein. It can be seen from Figure 3b that the BpCYCs family has a similar intron structure (Figure 3b). Finally, the intron-exon organization of the BpCYCs family is similar to that of Arabidopsis, This indicates that CYC is highly conserved in plants in an evolutionary manner.

![Figure 3](image)
Figure 3. Phylogenetic analysis, exon/Intron genomic structure and protein motif organization of CYC in B. pendula. The phylogenetic tree was made by MEGA5.05 using the neighbor-joining algorithm, and the confidence was evaluated by 500 bootstraps. Exons and introns of the BpCYCs genes are represented by yellow boxes and grey lines respectively. Use TBtools software to compare gene structure. Use MEME Web server to analyze the distribution of conserved motifs in BpCYCs protein. And visualized by TBtools software. (a) Phylogenetic analysis of BpCYCs; (b) Exon/Intron genomic structure of BpCYCs; (c) Protein motif organization of BpCYCs;

3.5. CKS

CDK subunit (CKS) proteins act as docking factors that mediate the interaction of CDKs with putative substrates and regulatory proteins. There are two CDK subunit genes in Arabidopsis described previously [4]. In this study, we have identified two BpCKSs in the B. pendula genome. It can be seen that these two genes have the same motif, but their gene structures are quite different.

Figure 4. Exon/Intron genomic structure and protein motif organization of CKS in B. pendula. Exons and introns of the BpCKSs genes are represented by yellow boxes and grey lines respectively. Use TBtools software to compare gene structure. Use MEME Web server to analyze the distribution of conserved motifs in BpCKSs protein. And visualized by TBtools software. (a) Exon/Intron genomic structure of BpCKSs; (c) Protein motif organization of BpCKSs;

3.6. Rb and E2F/DP

RB regulates the expression of many essential genes in cell cycle progression by regulating the activity of E2F transcription factor. Only one Rb could be identified in the Arabidopsis genome [4]. We identified two BpRbs in the B. pendula genome. E2F transcription factors, composed of E2F and DP, play a decisive role in plant cell size control[24]. We identified three BpE2Fs in the B. pendula genome. Two BpDPs were identified in the B. pendula genome. Because of phylogenetic positions of E2F and DP, they form a distinct class, which we designated DP-E2F–like (DEL). We identified two BpDELS in the B. pendula genome. Four groups had emerged based on the phylogenetic analysis (Figure 5a). Through the analysis of conservative motifs, it can be seen that both E2F and DP families contain conservative motif 1 (Figure 5c), indicating that conservative motif 1 is highly conserved during evolution. Except for BpRb2 and BpDPa2, both intron and exon structures contain highly similar and numerous introns (Figure 5b).

Figure 5. Phylogenetic analysis, exon/Intron genomic structure and protein motif organization of E2F, DP, DEL and Rb in B. pendula. The phylogenetic tree was made by MEGA5.05 using the neighbor-joining algorithm, and the confidence was evaluated by 500 bootstraps. Exons and introns of the BpE2Fs, BpDPs, BpDELS and BpRbs genes are represented by yellow boxes and grey lines respectively. Use TBtools software to compare gene structure. Use MEME Web server to analyze the distribution of conserved motifs in BpE2Fs, BpDPs, BpDELS and BpRbs protein. And visualized by TBtools software. (a) Phylogenetic analysis of BpE2Fs, BpDPs, BpDELS and BpRbs; (b) Exon/Intron genomic structure of BpE2Fs, BpDPs, BpDELS and BpRbs; (c) Protein motif organization of BpE2Fs, BpDPs, BpDELS and BpRbs;
3.7. KRP and WEE

The activity of CYC-CDK is also regulated by an inhibitory protein CKI (also known as KRP). Seven CKI genes belonging to the group of Kip/Cip CKIs have been described previously for Arabidopsis, designated KRP1 to KRP7 [25]. In this study, we have identified four BpKRPs in the B. pendula genome. These four genes all have motif 1 (Figure 6c). BpKRP1 and BpKRP2 also contain the same motif 2, and both contain 3-4 introns (Figure 6b), and have similar structures.

CDK/cyclin activity is regulated negatively by phosphorylation of the CDK subunit by the WEE1 kinase and positively when the inhibitory phosphate groups are removed by the CDC25 phosphatase. Two BpWEEs were identified in the B. pendula genome. Their conserved motifs are similar in structure, while there are only two introns in BpWEE2.

![Figure 6](image)

**Figure 6.** Phylogenetic analysis, exon/Intron genomic structure and protein motif organization of KRP and WEE in B. pendula. The phylogenetic tree was made by MEGA 5.05 using the neighbor-joining algorithm, and the confidence was evaluated by 500 bootstraps. Exons and introns of the BpKRPs, BpWEEs genes are represented by yellow boxes and grey lines respectively. Use TBtools software to compare gene structure. Use MEME Web server to analyze the distribution of conserved motifs in BpKRPs, BpWEEs protein. And visualized by TBtools software. (a) Phylogenetic analysis of BpKRPs, BpWEEs; (b) Exon/Intron genomic structure of BpKRPs, BpWEEs; (c) Protein motif organization of BpKRPs, BpWEEs;

3.8. RNA-seq expression analysis of B. pendula cell cycle genes

We applied quantitative criteria to assign genes that are likely to be cell cycle genes based on transcript abundance and specificity. The tissue specific expression data include xylem, roots, leaves and flowers. We calculated the total expression of the 59 identified genes in xylem and selected 17 genes which have a high expression in leaves or xylem or flower (Figure 7). The 17 cell cycle genes were BpCDKA.1, BpCDKB1.1, BpCDKB2.1, BpCKS1.2, BpCYCB1.1, BpCYCB1.2, BpCYCB2.1, BpCYCD3.1, BpCYCD3.5, BpDEL1, BpDpa2, BpE2Fa, BpE2Fb, BpKRP1, BpKRP2, BpRb1 and BpWEE1.

In the BpCDK family, BpCDKA.1 is abundant in xylem. In addition, BpCDKA.1, BpCDKB1.1 and BpCDKB2.1 were highly expressed in leaves. BpCDKA.1 is highly expressed in all the four investigated tissues, which indicated BpCDKA.1 may play multiple roles in different tissues. The most similar genes to BpCDKA.1, BpCDKB1.1 and BpCDKB2.1 in A. thaliana are CDKA;1 (AT3G48750), CDKB1.1 (AT3G54180) and CDKB2.1 (AT1G76540).

A total of 27 BpCYCs were detected in the B. pendula genome, of which BpCYCD3.5 is abundant in flower and leaves. The gene most similar to BpCYCD3.5 in A. thaliana is AT3G50070. In addition to this, BpCYCB1.1, BpCYCB1.2, BpCYCB2.1 and BpCYCD3.1 are highly expressed in leaves.

In the CKS family of birch, BpCKS1.2 was most abundant in the leaf and expressed at moderate levels in the other three tissues. The gene most similar to BpCKS1.2 in A. thaliana is AT2G27960.

BpRb1 is abundant in leaves BpRb1 is most similar to AT3G12280. ZmRb1 binds to D-type cyclins in plants, is highly expressed in differentiated cells, and regulates leaf development in temporal and spatial level [26]. BpE2Fa and BpE2Fb are abundant in leaves. BpE2Fa and BpE2Fb are most similar to AT2G36010 and AT5G22220, respectively. Two BpDPs were identified in the B. pendula genome, of which BpDP2 is abundant in xylem, and this gene is similar to AT5G02470. BpDEL1 is abundant in leaves. This gene is similar to AT3G48160 in A. thaliana.
In the KRP family of birch, \textit{BpKRP1} was most abundant in the xylem and \textit{BpKRP2} also was expressed at a high level in the xylem. These two genes are most similar to \textit{AT2G23430} in \textit{A. thaliana}. Moreover, \textit{BpKRP1} and \textit{BpKRP2} are also highly expressed in flower and leaves. \textit{BpWEE1} is abundant in leaves. This gene is similar to \textit{AT1G02970}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{heatmap.png}
\caption{The heat map shows the expression of cell cycle genes in different parts of the birch tissue. Highly or lowly expressed genes are colored red or blue, respectively.}
\end{figure}

4. Discussion

Previous studies have identified many cell cycle genes [27], but the genetic and biochemical roles of the birch cell cycle genes needs to be better defined. In this study, we identified a total of 59 cell cycle genes in \textit{B. pendula}, which should help clarifying the molecular mechanism of plant growth and development in \textit{B. pendula}.

Plant cell cycle could be regulated by altered expression of some G1-S and G2-M checkpoints genes in cells [3]. G1-S phase was one of the most important checkpoints among all the cell cycle, and \textit{CycD} genes have been indicated as a sensor of extracellular growth condition [1]. Over expression of \textit{CycD3;1} in \textit{Arabidopsis thaliana} could induce B-type cyclin expression, resulting in not only an increase in endoreduplication but also in mitosis [28]. A further study revealed that \textit{CYCLIN B1;2} was the mitosis promoting factor [29]. \textit{CYCLIN B1;2} expression can promote nuclear and cellular division, which is sufficient to trigger endoreduplication to mitosis, but not sufficient enough to increase cell cycle rounds [29]. In contrast with our results, \textit{BpCYCB1.1}、\textit{BpCYCB1.2}、\textit{BpCYCB2.1} and \textit{BpCYCD3.1}
are highly expressed in leaves, and BpCYCD3.5 is abundant in flower and leaves (Figure 7). These genes with high expression levels in birch tissues contain CYCD3.1 and CYCB1.2, indicating that these two genes in birch may also play a very important role in cell division. Gene structure analysis found that the gene sequence structure of Bp CYCs family members is similar (Figure 3b), indicating that their gene structure is highly conserved during evolution. Both the pistil cell death and stamen cell arrest are involved in cell cycle regulation in maize sex determination, CYCA, CYCB and CDK were highly expressed in the developing pistil and stamen, while Wee1 and CKI were only expressed in the arresting stamen[30]. In our study, part of genes was highly expressed in flower, such as BpCYCD3.5, BpCKS1.2, BpCDKA.1 (Figure 7). However, birch has unisexual flowers on separate male and female inflorescences (catkins)[12,31,32]. How the cell cycle genes regulate the flower development process of birch needs our further research.

5. Conclusions

Cell cycle genes are closely related to all life activities of plants, we identified 17 core cell cycle genes in the genome of birch by combining phylogenetic analysis, gene structure analysis and tissue specific expression data, provide some help for better application of cell cycle genes and modern molecular breeding.

Author Contributions: Conceptualization, Haijiao Huang; software, Yijie Li and Song Chen; validation, Haijiao Huang; writing—original draft preparation, Yijie Li and Song Chen; writing—review and editing, Yijie Li; project administration, Haijiao Huang; funding acquisition, Haijiao Huang. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities (2572018BW06), the National Natural Science Foundation of China (31800556).

Acknowledgments: We thank the reviewers and editors who provided constructive comments on our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. De Veylder, L. The Discovery of Plant D-Type Cyclins. The Plant cell 2019, 31, 1194-1195, doi:10.1105/tpc.19.00277.
2. Boudolf, V.; Lammens, T.; Boruc, J.; Van Leene, J.; Van Den Daele, H.; Maes, S.; Van Isterdael, G.; Russinova, E.; Kondorosi, E.; Witters, E., et al. CDKB1;1 Forms a Functional Complex with CYCA2;3 to Suppress Endocycle Onset. Plant physiology 2009, 150, 1482-1493, doi:10.1104/pp.109.140269.
3. Boudolf, V.; Vlieghe, K.; Beemster, G.T.S.; Magyar, Z.; Acosta, J.A.T.; Maes, S.; Van Der Schueren, E.; Inze, D.; De Veylder, L. The plant-specific cyclin-dependent kinase CDKB1;1 and transcription factor E2Fa-DPa control the balance of mitotically dividing and endoreplicating cells in Arabidopsis. The Plant cell 2004, 16, 2683-2692, doi:10.1105/tpc.104.024398.
4. Vandepoele, K.; Raes, J.; De Veylder, L.; Rouze, P.; Rombauts, S.; Inze, D. Genome-wide analysis of core cell cycle genes in Arabidopsis. The Plant cell 2002, 14, 903-916, doi:10.1105/tpc.010445.
5. Inze, D.; De Veylder, L. Cell cycle regulation in plant development. Annual review of genetics 2006, 40, 77-105, doi:10.1146/annurev.genet.40.110405.090431.
6. Kosugi, S.; Ohashi, Y. Interaction of the Arabidopsis E2F and DP proteins confers their concomitant nuclear translocation and transactivation. Plant physiology 2002, 128, 833-843, doi:10.1105/tpc.010642.
7. Boniotti, M.B.; Gutierrez, C. A cell-cycle-regulated kinase activity phosphorylates plant retinoblastoma protein and contains, in Arabidopsis, a CDKA/cyclin D complex. The Plant journal : for cell and molecular biology 2001, 28, 341-350, doi:10.1046/j.1365-313x.2001.01160.x.
8. Drakare, S.; Lennon, J.J.; Hillebrand, H. The imprint of the geographical, evolutionary and ecological context on species-area relationships. Ecol Lett 2006, 9, 215-227, doi:10.1111/j.1461-0248.2005.00848.x.
9. Finlay, B.J. Global dispersal of free-living microbial eukaryote species. Science 2002, 296, 1061-1063, doi:10.1126/science.1070710.
10. Fenchel, T.; Finlay, B.J. The ubiquity of small species: Patterns of local and global diversity. Bioscience 2004, 54, 777-784, doi:10.1641/0006-3568(2004)054[0777:Tuossp]2.0.Co;2.
11. Salojarvi, J.; Smolander, O.P.; Nieminen, K.; Rajaraman, S.; Safronov, O.; Safdari, P.; Lamminmaki, A.; Immanen, J.; Lan, T.Y.; Tanskanen, J., et al. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. Nature genetics 2017, 49, 904-+, doi:10.1038/ng.3862.

12. Huang, H.J.; Wang, S.; Jiang, J.; Liu, G.F.; Li, H.Y.; Chen, S.; Xu, H.W. Overexpression of BpAP1 induces early flowering and produces dwarfism in Betula platyphylla x Betula pendula. Physiologia plantarum 2014, 151, 495-506, doi:10.1111/plp.12123.

13. Chen, S.; Lin, X.; Zhang, D.W.; Li, Q.; Zhao, X.Y.; Chen, S. Genome-Wide Analysis of NAC Gene Family in Betula pendula. Forests 2019, 10, doi:Artn 74110.3390/F10090741.

14. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST plus : architecture and applications. BMC bioinformatics 2009, 10, doi:Artn 42110.1186/1471-2105-10-421.

15. Marchler-Bauer, A.; Bryant, S.H. CD-Search: protein domain annotations on the fly. Nucleic acids research 2004, 32, W327-W331, doi:10.1093/nar/gkh454.

16. Gang, H.X.; Li, R.H.; Zhao, Y.M.; Liu, G.F.; Chen, S.; Jiang, J. Loss of GLK1 transcription factor function reveals new insights in chlorophyll biosynthesis and chloroplast development. Journal of experimental botany 2019, 70, 3125-3138, doi:10.1093/jxb/erz128.

17. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nature methods 2012, 9, 357-U53, doi:10.1038/Nmeth.1923.

18. Li, B.; Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 2011, 12, doi:Artn 32310.1186/1471-2105-12-323.

19. Guo, J.; Song, J.; Wang, F.; Zhang, X.S. Genome-wide identification and expression analysis of rice cell cycle genes. Plant molecular biology 2007, 64, 349-360, doi:10.1007/s11103-007-9154-y.

20. Adachi, S.; Nobusawa, T.; Umeda, M. Quantitative and cell type-specific transcriptional regulation of A-type cyclin-dependent kinase in Arabidopsis thaliana. Developmental biology 2009, 329, 306-314, doi:10.1016/j.ydbio.2009.03.002.

21. Boudolf, V.; Barroco, R.; Engler, J.D.; Verkest, A.; Beeckman, T.; Naudts, M.; Inze, D.; De Veylder, L. B1-type cyclin-dependent kinases are essential for the formation of stomatal complexes in Arabidopsis thaliana. The Plant cell 2004, 16, 945-955, doi:10.1105/tpc.021774.

22. 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 Cai et al. BMC Genomics (2021) 22:314 Page 18 of 19 stress response. Gene. 2015;566(1):95–108. https://doi.org/10.1016/j.gene.2015.04.041

23. Andersen, S.U.; Buechel, S.; Zhao, Z.; Ljung, K.; Novak, O.; Busch, W.; Schuster, C.; Lohmann, J.U. Requirement of B2-type cyclin-dependent kinases for meristem integrity in Arabidopsis thaliana. The Plant cell 2008, 20, 88-100, doi:10.1105/tpc.107.054676.

24. Huntley, R.; Healy, S.; Freeman, D.; Lavender, P.; Murray, J.A.H. The maize retinoblastoma protein homologue ZmRb-1 is regulated during leaf development and displays conserved interactions with G1/S regulators and plant cyclin D (CycD) proteins. Plant molecular biology 1998, 37, 155-169.

25. Sabelli, P.A.; Larkins, B.A. Regulation and function of retinoblastoma-related genes. Plant Science 2009, 177, 540-548, doi:10.1016/j.plantsci.2009.09.012.

26. Shimotohno, A.; Umeda-Hara, C.; Bisova, K.; Uchimiya, H.; Umeda, M. The plant-specific kinase CDKF;1 is involved in activating phosphorylation of cyclin-dependent kinase-activating kinases in Arabidopsis. The Plant cell 2004, 16, 2954-2966, doi:DOI 10.1105/tpc.104.025601.

27. De Veylder, L.; Beeckman, T.; Beemster, G.T.; Krols, L.; Terras, F.; Landrieu, I.; van der Schueren, E.; Maes, S.; Naudts, M.; Inze, D. Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis. The Plant cell 2001, 13, 1653-1668, doi:10.1105/tpc.010087.

28. Schnittger, A.; Schobinger, U.; Bouyer, D.; Weinl, C.; Stierhof, Y.D.; Hulskamp, M. Ectopic D-type cyclin expression induces not only DNA replication but also cell division in Arabidopsis trichomes. Proceedings of the National Academy of Sciences of the United States of America 2002, 99, 6410-6415, doi:10.1073/pnas.092657299.

29. Schnittger, A.; Schobinger, U.; Stierhof, Y.D.; Hulskamp, M. Ectopic B-type cyclin expression induces mitotic cycles in endoreduplicating Arabidopsis trichomes. Current biology : CB 2002, 12, 415-420, doi:10.1016/s0960-9822(02)00693-0.

30. Lee, J.R.; Kim, J.C. Temporal and Spatial Regulation of Cell Cycle Genes during Maize Sex Determination. Journal of Life Science 2006, 16, 828-833.

31. Wang, S.; Huang, H.J.; Han, R.; Liu, C.Y.; Qiu, Z.N.; Liu, G.F.; Chen, S.; Jiang, J. Negative feedback loop between BpAPI and BpPI/BpDEF heterodimer in Betula platyphylla x B. pendula. Plant Science 2019, 289, doi:ARTN 11028010.1105/j.plantsci.2019.110280.
32. Wang, S.; Huang, H.J.; Han, R.; Chen, J.Y.; Jiang, J.; Li, H.Y.; Liu, G.F.; Chen, S. BpAP1 directly regulates BpDEF to promote male inflorescence formation in Betula platyphyla x B. pendula. Tree physiology 2019, 39, 1046-1060, doi:10.1093/treephys/tpz021.