RESEARCH ARTICLE

Gene expression analysis of bud burst process in European hazelnut (*Corylus avellana* L.) using RNA-Seq

Musa Kavas[1] · Aslıhan Kurt Kızıldoğan[1] · Hüseyin İrfan Balık[2]

Abstract The control of bud burst process depending on temperature is crucial factor in woody perennial plants to survive in unfavorable ecological conditions. Although it has important economic and agronomic values, little information is available on the molecular mechanism of the bud burst process in *Corylus avellana*. Here for the first time, we conducted a de novo transcriptome-based experiment using eco-dormant leaf bud tissues. Four transcriptome libraries were constructed from the leaf bud tissues and sequenced via Illumina platform. Transcriptome analysis revealed 86,394 unigenes with a mean length of 1189 nt and an N50 of 1916 nt. Among these unigenes, 63,854 (73.78%) of them were annotated by at least one database. De novo assembled transcripts were enriched in phenylpropanoid metabolism, phytohormone biosynthesis and signal transduction pathways. Analyses of phytohormone-associated genes revealed important changes during bud burst, in response to gibberellic acid, auxin, and brassinosteroids. Approximately 2163 putative transcription factors were predicted, of which the largest number of unique transcripts belonged to the MYB transcription factor family. These results contribute to a better understanding of the regulation of bud burst genes in perennial plants.

✉ Musa Kavas
  musa.kavas@omu.edu.tr

1 Department of Agricultural Biotechnology, Faculty of Agriculture, Ondokuz Mayıs University, Samsun, Turkey

2 Giresun Hazelnut Research Station, Ministry of Food, Agriculture and Livestock, Giresun, Turkey

Introduction

Hazelnut (*Corylus avellana* L.) is a diploid (2n = 22), monoecious and allogamous tree species belonging to the Betulaceae family. Among 25 *Corylus* species, *C. avellana* (European hazelnut) is the most important crop with a major impact on industry and global agricultural sector (Rowley et al. 2014). In ancient times, approximately 2500 years ago, hazelnut was cultivated in the Anatolian region and about 400 clonal cultivars were distributed across Europe at that time (Boccacci et al. 2013). The plant is highly rich in valuable nutrients such as oleic acid, phytosterols (especially β-sitosterol), vitamin E, and squalene. Additionally, the nut possesses several antioxidants such as tocopherols (mainly α-tocopherol) and phenolics (especially tannins, found in the brown skin) (Delgado et al. 2010).

One of the most important ecological factor in deciduous plant survival is the control of bud burst depending on a temperature-related mechanism. In perennial plants, on-time leaf bud burst allows them to start growing early enough to take advantage of the spring conditions, at the same time to grow late enough to escape from late frost to reduce the risk of tissue damage (Azad 2012). Hazelnut and other perennial plants sense environmental signals like photoperiodicity and temperature to adjust timing of actively growing and dormancy stages (Shim et al. 2014). Leaf bud burst time is associated with bud dormancy release which depends on suppression of growth inhibitors, and accumulation of growth promoters. Bud burst of perennial plants generally rely on exposure to a particular duration of

non-freezing temperatures to release dormancy, followed by a warm temperature to allow growth in the spring (Faust et al. 1997). There are three types of bud dormancy; (1) para-dormancy (inhibition of meristem growth in buds due to unfavorable conditions), (2) endo-dormancy (inhibition of meristematic growth by itself, chilling requirement is required to overcome) and (3) eco-dormancy (a state of dormancy in the presence of harsh conditions for cell division, insufficient temperature increment for bud break) (Lang 1987). Leaf bud of hazelnut requires certain amount of chilling for the transition from endo-dormancy to eco-dormancy stage and from eco-dormancy to bud and sprouting (Lang GA 1987; Howe et al. 2015). The chilling requirement varies among genotypes and plant's organs and is an important consideration in choosing cultivars adapted to local climatic conditions. For example, leaf buds of hazelnut have a chilling requirement of 365–1690 h below 7 °C, but the duration of chilling is variable between cultivars (Mehlenbacher 1991; Ma et al. 2016).

Turkey is the largest hazelnut producer in the world. Roughly, 70% of the world's hazelnut supply is grown in the north side of country. In 2012, a period of severe frost in Turkey reduced the output from 660,000 tonnes (728,000 tons) to 549,000 tonnes. Another instance of cold temperatures in the following year caused the yield to tumble to 381,000 tonnes in 2013. These types of disasters have been also reported in other hazelnut producer countries. Only way to overcome these types of disasters is the development of new cultivars that have more cold hardiness and late bud burst. However, selection and improvement of these types of hazelnut cultivars strongly depends on identification of bud burst related genes and their regulation networks that could be used as molecular markers in breeding studies.

The current study was aimed to conduct a transcriptome analysis for the identification of genes and related regulatory networks responsible for leaf bud burst time in *C. avellana* genotypes that have different bud burst time. The results of expression profiling generated through transcriptome sequencing will increase the knowledge about molecular mechanism of leaf bud burst and will contribute to identifying important genes involved in bud burst for the ultimate improvement of hazelnut.

Materials and methods

Plant material

Transcriptome sequencing and de novo assembly was carried out with Çakıldak (late vegetative bud burst) and Palaz (early vegetative bud burst) for comparative transcriptome analysis. Eco-dormant leaf buds of these hazelnut genotypes have been previously reported to differ for 15–20 days in terms of their bud-burst times (Köksal 2002). All leaf buds were collected in same developmental stage before the start of swelling. Eco-dormant buds containing two main branches for each genotype were collected at the same time from 22-year-old C. *avellana* L. trees grown on a field site in Giresun, Turkey and stored in liquid nitrogen until RNA isolation. Ten buds were isolated from each branch using sterile scalpel blades and pooled for RNA analysis. By this way, two biological replicates for each genotype were used for transcriptome sequencing.

RNA isolation and sequencing and assembly

Total RNA was extracted from approximately 0.3 g of frozen sample using a modified method of Chang et al. (1993). A modified protocol including CTAB, Qiagen RNeasy kit and Phase Lock Gel (5 Prime) was preferred because of high polyphenol content in hazelnut buds. Kits and columns were used according to the manufacturer's protocols. Genomic DNA contamination in extracted total RNA was eliminated using RNase-free DNase I (Thermo Scientific). The genomic DNA contamination and integrity of RNA samples were evaluated by an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). High quality samples with RNA integrity number ≥ 6.5, 28S:18S > 1, OD260/280 ≥ 2 and OD260/230 ≥ 2 were selected to construct cDNA library.

The poly(A)-containing mRNA was isolated from total RNA to carry out fragmentation. The first-strand cDNA synthesis was achieved with random hexamer primers by using approximately 170 bp fragments as templates. DNA polymerase I, dNTPs, and RNase H were used to synthesize the second-strand cDNA. To collect detailed information about the whole expression profiles at mRNA level, RNA-Seq was conducted in four sequencing libraries (2 genotypes X 2 biological replicates). Transcriptome sequencing was performed at Beijing Genomics Institute (http://www.genomics.cn) with Illumina Hi-Seq 4000 platform using a 100 bp paired-end sequencing approach. The standard sequencing protocols recommended by Illumina was followed to construct sequencing library and sequencing. After removal of adaptors, unknown nucleotides larger than 5%, reads that were shorter than 50 base pairs (bps) and low-quality reads with more than 20% low-quality bases (base quality ≤ 10), reads were obtained in FASTQ format and submitted to NCBI Sequence Read Archive (SRX1665375- SRX1665376- SRX1665377- SRX1665378). Assembly of all cleaned reads were carried out on Trinity (Grabherr et al. 2011), a software package optimized for de novo assembly of short-read data with default K-mers = 25. Assembled transcript file was deposited into NCBI (TSA) database under the

SUB4195016 accession number. The resulting sequences of Trinity was called as transcripts (raw sequences). Transcriptome completeness were evaluated using BUSCO v3.0 using the plantae linage with default parameters (Simao et al. 2015). The assembled reads were further processed for redundancy removal and clustered with TGICL (Pertea et al. 2003) in order to obtain final non-redundant unigenes (consensus sequences) as long as possible. The non-redundant unigenes were separated into two class which were singletons and clusters. The prefix CL were assigned to the latter group called as clusters. In one cluster, there were several Unigenes with more than 70% similarity among them. The other one was singletons, for which the prefix was unigene.

Functional annotation

The final unigenes were aligned through BLASTx (E value $< 10^{-5}$) to the publicly available protein databases, including NR, Swiss-Prot, Interpro, the Kyoto Encyclopedia of Genes and Genomes (KEGG), and Clusters of Orthologous Groups (COG). Gene names were assigned based on the best BLAST hit (Altschul et al. 1997). The Open reading frames (ORFs) finder software called as "GetORF" was used for prediction of ORF (http://www.bioinformatics.nl/cgi-bin/emboss/getorf). The longest ORF obtained from each unigene was used to determine coding sequence (CDS). Amino acid sequences of the CDSs were predicted by using the standard codon table. To determine coding sequences of unigenes with no hits in these databases, ESTScan software (Christian et al. 1999) was used. GO annotations of the unigenes was carried out by using Blast2Go program with an E value $\leq 10 - 5$ on the basis of molecular function, biological process, and cellular component categories (Conesa et al. 2005), and InterProScan software.

Determination of gene expression levels

To estimate the expression levels of genes, we mapped clean reads to unigenes with Bowtie 2.0 (Langmead et al. 2009). Expression profile of each unigenes was calculated using RSEM software (Li and Dewey 2011). Reads that could be uniquely mapped to a gene by using SAMtools (http://samtools.sourceforge.net/) were used to calculate expression levels. Gene expression levels were measured according to the number of uniquely mapped fragment per kilobase of exon region per million mapped reads (FPKM). Since FKPM eliminates the effect of gene length on expression pattern, FPKM values was used directly for comparing differences in gene expression among samples.

Screening and pathway enrichment of DEGs

To detect the differentially expressed genes (DEGs) between Çakıldak (control) and Palaz, DeSeq 2 R software package (which use a negative binomial distribution to determine differential expression in digital gene expression data) was used as described by Love et al. (2014). More than two-fold change of expression levels and the expression with significant difference ($P < 0.05$) were considered differentially expressed between samples. Then, DEGs across samples were further annotated by GO enrichment analysis and KEGG pathway enrichment analysis. GO terms and KEGG pathways fulfilling the criterion of a Bonferroni-corrected P value ≤ 0.05 were defined as significantly enriched in DEGs. Based on the NR annotation, the Blast2GO program was used (Conesa et al. 2005) to get the GO annotation of all consensus sequences.

Quantitative reverse transcription (qRT)-PCR validation

Sixteen selected unigenes differentially expressed in phenylpropanoid and plant hormone synthesis pathways were chosen for validation using real time qPCR. Total RNA from leaf buds belonging to Çakıldak and Palaz was extracted with optimized protocol. Reverse transcription to cDNA was performed on 200 ng total RNA using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) by following the manufacturer's instructions. qRT-PCR was performed with cDNA template, 1 μl; 5 × HOT FIREPol® EvaGreen® qPCR Supermix, 4 μl; forward primer (10 μM), 0.4 μl; reverse primer (10 μM), 0.4 μl and dH$_2$O, 14.2 μl in a BioRad CFX96 instrument. The qRT-PCR reactions were started with an initial denaturation step (12 min at 95 °C) and proceeded with 40 cycles of amplification (15 s at 95 °C and 20 s at 60 °C). Melting curve analyses were performed to check for specificity of the amplifications. 18S rRNA gene was chosen as reference gene in relative quantification of gene expression. All reactions were performed with triple biological replicates and repeated two times. Transcript fold-changes describing the change in expression of the target gene in samples from Çakıldak (control) to that of the Palaz transcript were calculated using the $2^{-\Delta\Delta Ct}$ method described by Schmittgen and Livak (2008). The primers used in qRT-PCR analysis are listed in Table S1.

Results and discussion

Illumina sequencing and de novo sequence assembly

Dormancy and leaf bud burst is a highly complex physiological phenomenon used by perennial plants to survive in unfavorable environmental conditions. There are many woody plants that require a period of non-freezing temperature to break the dormancy of their buds. The physiological and molecular aspects of bud dormancy release in plants have been previously examined in several studies (Besford et al. 1996; Fuchigami and Wisniewski 1997; Hakkinen et al. 1998; Pletsers et al. 2015; Bhandawat et al. 2017; Singh et al. 2017). This is the first transcriptome study carried out on hazelnut genotypes to identify transcripts and gene regulation networks behind early and late bud burst.

To gather information about genes expressed at the time of bud burst period, RNA-Seq was performed on leaf buds of *C. avellana* L. cv Çakıldak and Palaz hazelnut genotypes having altered bud burst time. A total of 192,340,000 raw reads were generated after Illumina Hiseq sequencing from four sequencing libraries. A total of 179,830,250 clean reads with about 17.99 Gb bases was obtained with a Q30 percentage of 96.32% (Table 1).

De novo assembly of cleaned reads were carried out via Trinity software with default K-mers = 25. Redundant sequences in assembled sequences were removed and final sequences were clustered by using the TIGR Gene Indices clustering tools (TGICL). The result of Trinity assembly revealed 74,342, 75,000, 81,667 and 74,234 raw sequence reads (transcripts) from four RNA-Seq libraries (Table S2). The most abundant raw sequences were 300 bp (a total of 106,642 from four sequencing libraries) and the least

abundant were 3000 bp (1885); sequences over 3000 bp were grouped together (Figure S1).

Based on gene family clustering analysis, consensus sequences (non-redundant unigenes) were separated into two classes which were singletons and clusters. The prefix CL were assigned to the latter group called as clusters. In one cluster, there were several Unigenes having more than 70% similarity. The other one was singletons, for which the prefix was unigene. Altogether, considering both repetitions, 86,542 unigenes, including 40,387 clusters and 46,155 singletons, with a mean length of 1189 nt and an N_{50}-value of 1916 bps (Table 2) was obtained within the current study. We further quantified the completeness of our assembled transcriptome by comparing their sequences to a core set of plantae genes using BUSCO. The result revealed that 95.2% of BUSCO genes were "complete", 2.5% were "fragmented", and the remaining 2.3% were "missing". The unigene size distribution is summarized in Fig. 1, which shows that 78.97% of the total unigenes are shorter than 2000 nucleotides. Differences in the length of assembled unigenes ranged from 300 to 12,227 bps, with 47,556 unigenes being shorter than 1000 bps, and 5702 unigenes sequence lengths being more than 3000 bps (Fig. 1).

Consensus sequences were aligned with BLASTdb using BLASTx (e value < 0.00001). Sequence orientations were determined according to the best hit in the database. ESTScan was used to predict the orientation of CDS for the unigenes that were unannotated with BLAST search. Prediction summary is shown in Table S3. A total of 60,453 CDS was identified with BLAST search and ESTScan. Small amount of CDS (3858) was predicted by using ESTScan software.

Table 1 Summary of sequencing reads after filtering

Sample	Total raw reads (Mb)	Total clean reads (Mb)	Total clean bases (Gb)	Clean reads Q20 (%)	Clean reads Q30 (%)	Clean reads ratio (%)
Çakıldak1	47.52	44.95	4.50	99.08	97.00	94.61
Çakıldak2	49.78	45.59	4.56	97.91	93.34	91.59
Palaz1	47.52	44.58	4.46	99.00	96.78	93.81
Palaz2	47.52	44.71	4.47	99.03	96.90	94.09

Table 2 Quality metrics of Unigenes

Sample	Total number	Total length	Mean length	N50	N70	N90	GC (%)
Çakıldak1	54,614	52,261,443	956	1604	1040	375	42.54
Çakıldak2	56,014	54,163,749	966	1630	1055	379	42.64
Palaz1	60,940	62,134,645	1019	1747	1120	397	42.45
Palaz2	55,899	55,898,340	999	1676	1098	399	42.57
All-Unigene	86,542	102,932,762	1189	1916	1329	538	42.17

Fig. 1 Unigene length distribution in assembly. X axis represents the length of unigenes. Y axis represents the number of unigenes
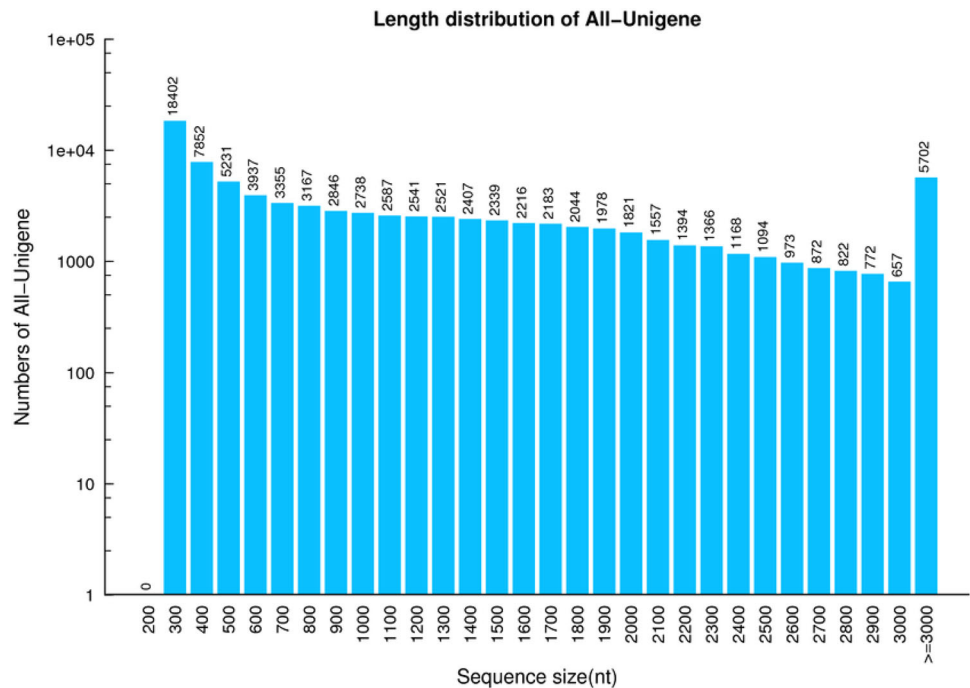
Annotation of predicted proteins

To identify the maximum number of similar genes, the unigene sequences were searched in different databases including NR, NT, KEGG, COG, SwissProt, Interpro and GO databases with a cutoff E value < 0.00001. According to the functional annotation result, 56,748; 55,919; 43,571; 26,394; 39,424; 44,965 and 17,379 unigenes were annotated, respectively (Table S4). Top-hit species distribution analysis based on BLASTx results indicated that a total of 10,080 sequences (18%) had top matches to sequences from *Vitis vinifera* (Fig. 2). The other closest matches with hazelnuts are *Prunus persica* (18%), *Theobroma cacao* (13%), *Morus notabilis* (7%), *Populus trichocarpa* (6%) *and Ricinus communis* (5%). Maximum transcripts showed similarity with three species of tree implying that the transcripts were assembled adequately. The KEGG pathway database is a collection of databases which shows networks of molecular interactions in cells for particular organisms. Analysis based on pathway gives additional information about the biological roles of target genes. KEGG pathway annotations were carried out for all annotated sequences. To determine active biological pathways responsible for differential bud break time of *C. avellena* genotypes, sequences were mapped to the reference canonical pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG). In total, 43,571 sequences were assigned to 137 KEGG pathways. The most enriched pathways were "metabolic pathways" (9880 unigenes), "biosynthesis of secondary metabolites" (5669 unigenes),
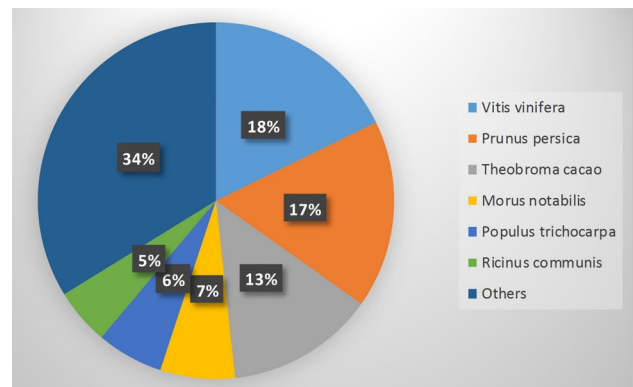

Fig. 2 Distribution of annotated species. This figure shows the species distribution of unigene BLASTX results against the NR protein database with a cutoff E value $< 10^{-5}$ and also shows the proportions of each species. Different colors represent different species. Species with proportions of less than 5% are not shown (colour figure online)

"plant-pathogen interaction" (1761 unigenes), ribosome (1605 unigenes) and "RNA transport" (1584 unigenes) in two different bud burst stages (Table S5). Our results were in consensus with previous work on grape, in which comparison of transcriptomic analysis of bud dormancy release period showed the most enriched pathways were plant-pathogen interaction and biosynthesis of secondary metabolites (Khalil-Ur-Rehman et al. 2017).

Another functional annotation study was completed by using COG database, where the proteins have been classified on the basis of their orthology concept with the help of completely sequenced genomes (Tatusov et al. 1997).

Fig. 3 Functional distribution of COG annotation. X axis represents the number of Unigenes. Y axis represents the COG functional category
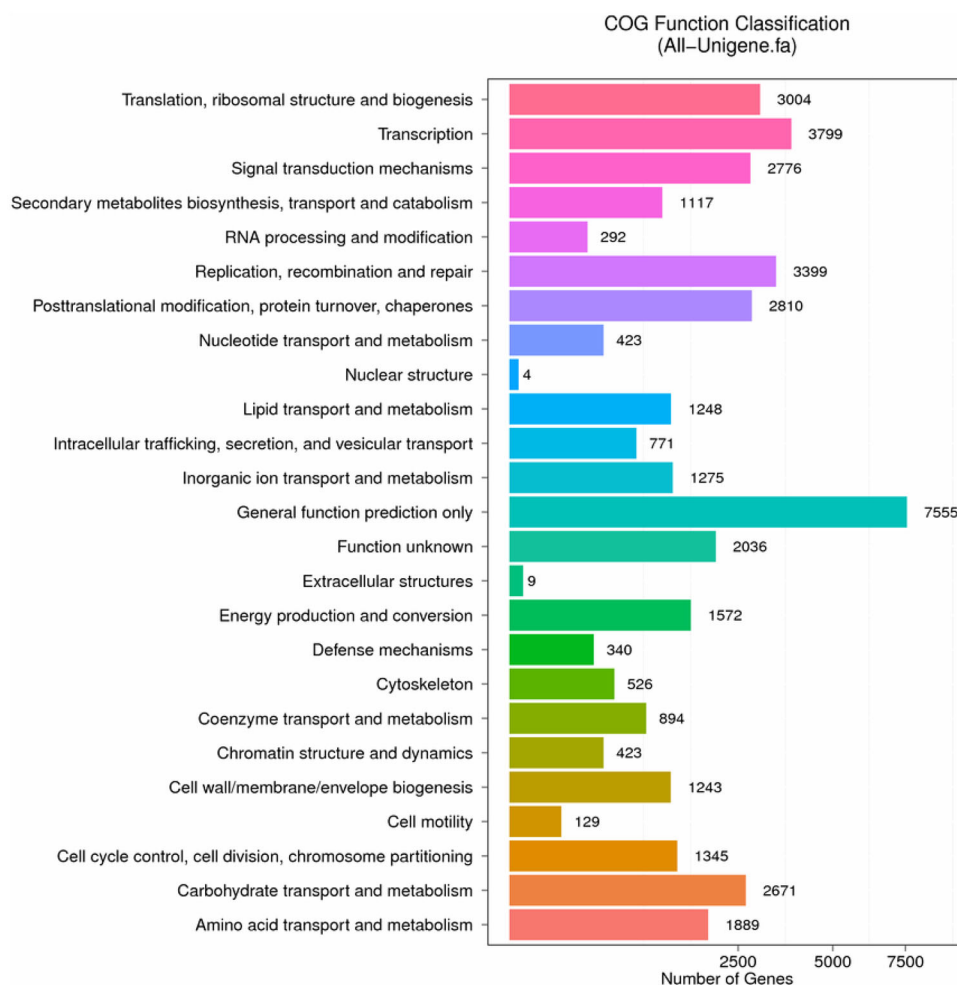
The COG database is built on genes from sequenced genomes and also show evolutionary relationships between organisms including bacteria, algae and eukaryotes. To predict possible roles of all unigenes of *C. avellena*, sequences were searched in the COG database. On the basis of NCBI COG classification, the 26,394 unigenes were divided into 25 clusters. The groups with the highest representation were found in the clusters R "general function prediction only (7555)", K "transcription (3799)", L "replication, recombination and repair (3399)" and "MJ translation, ribosomal structure and biogenesis (3004) (Fig. 3).

Based on GO analysis of all unigenes obtained by using Blast2GO, 17,379 *C. avellena* sequences were grouped into 54 GO categories; molecular function (15), biological process (22), and cellular component (17). Unigenes with GO annotation accounted for 20.08% of all the unigenes. Results indicated that a big portion of genes was from the categories of "cellular process (8395)", "metabolic process (10,020)", "single-organism process (6497), "cell (6121)", "cell part (6121)", "membrane (3565)" "organelle (4316)", "catalytic activity (9963)" and "binding activity (8389)" (Fig. 4). Enriched GO terms recognized in our study, response to stimulus, response to stress, oxidation–reduction process and transmembrane transport, were also in agreement with previous reports (Horvath et al. 2008; Khalil-Ur-Rehman et al. 2017). A total of 39,424 unigenes out of 86,542 has been successfully annotated in Swissprot database by using BLASTx (E value $\leq 1.0E^{-5}$) search. Using a five-way Venn diagram, we studied the overlap of the annotated unigene sequences within each of the databases (NR, Interpro, Swissprot, KEGG ve COG) (Fig. 5). Interestingly, a total of 20,010 unigenes had hits in all five databases, whereas there were overlaps of 2221, 1136, 3287, 111 and 155 unigenes between the NR and KEGG, Swissprot and NR, NR and Interpro, Nr and COG and COG and Interpro databases, respectively. We assume that unigenes having no hits in all databases were small transcripts lacking recognizable protein domains. The valuable information related with specific processes, structures, functions, and pathways obtained from these annotations will facilitate the research on *C. avellena*.
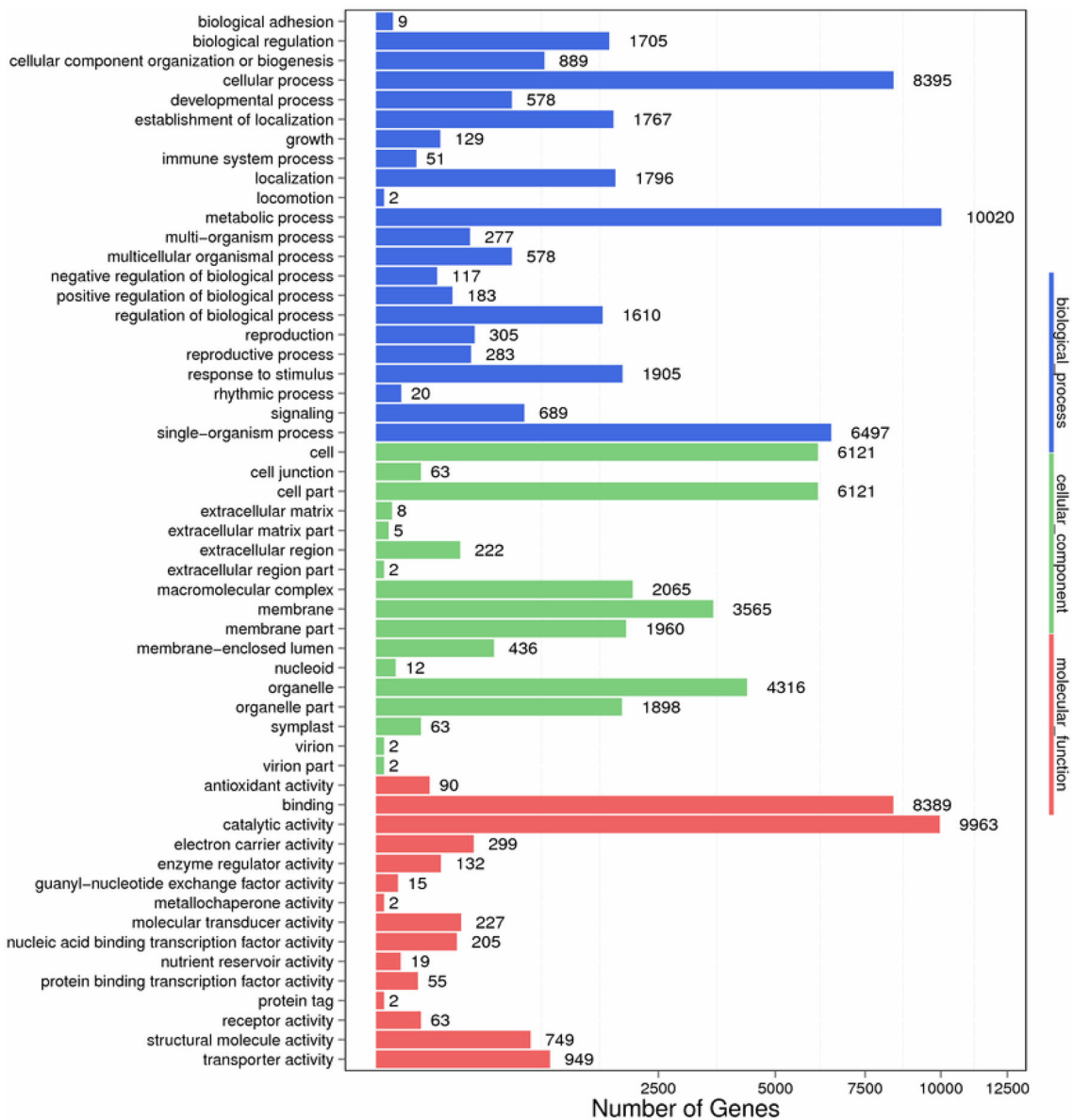
Fig. 4 Gene Ontology (GO) categories of unigenes. Unigenes were annotated in three categories: cellular components, molecular functions, and biological processes

Unigene TF prediction

Transcription factors (TFs) play important roles in various biological processes by regulating the gene expression. Determination of conserved domains located in deduced polypeptide can provide valuable information about the function, regulation and localization of the predicted protein. In order to increase our knowledge about control and regulation of gene expression in *C. avellena*, all transcription factors were predicted according to the family assignment rules illustrated in PlantTFDB (Riano-Pachon et al. 2007). A total of 2163 unigenes distributed into 59 transcription factor families were predicted to be involved in the regulation of transcription (Figure S1). Among these TF gene families, MYB, MYB-related and bHLH were the most abundant TF families. MYB proteins constitute a diverse class of DNA-binding proteins, which are important to the control of proliferation and differentiation in a number of cell types and regulation of secondary metabolism (Jin and Martin 1999). In hazelnut transcriptome, we identified 105 unigenes encoding MYB TFs. The expression profiles of these unigenes were compared between Palaz and Çakıldak genotypes to elucidate the roles of MYB TFs in dormancy release period. It was found that a total of nine unigenes have down-regulated expression pattern and 18 unigenes up-regulated (Table S6). Similar
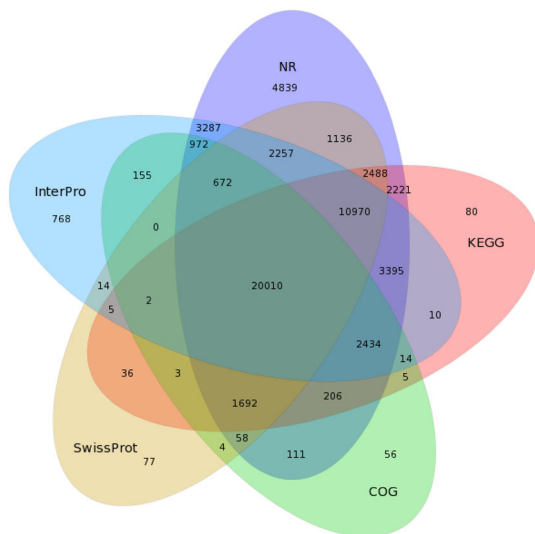
Fig. 5 Venn diagram between NR, COG, KEGG, Swissprot and Interpro. Venn diagram showing the BLASTX results of the *C. avellena* transcriptome against five protein databases. Using BLASTX search, de novo reconstructed unigene sequences were queried against the following public databases: NCBI-NR, Swiss-Prot, Interpro, KEGG and COG. The number of transcripts that have significant hits (E value ≤ 1e−5) against the five databases is shown in each intersection of the Venn diagram

results about possible function of MYB TFs in bud burst have been reported by several researchers (Xu et al. 2016; Bhandawat et al. 2017; Hao et al. 2017). It was shown that the expression of genes encoding MYB TFs was increased between endo-dormancy, eco-dormancy and bud flush in bamboo (Bhandawat et al. 2017). bHLH proteins responsible for the controlling of a diversity of processes including cell proliferation constitute the second largest TF family in plants (Heim et al. 2003). By this context, it was reported that bHLH TFs may be involved in dormancy release by regulating CBF transcription expression (Ren et al. 2016). In this study, we identified 62 bHLH TF encoding unigenes in hazelnut transcriptome. Another TF family highly related to dormancy release called as Dormancy Associated MADS-box (DAM), has been described in several other perennial species, including poplar (Ruttink et al. 2007; Yıldırım and Kaya 2017), leafy spurge (Horvath et al. 2008) and peach (Jiminez et al. 2010). It was reported that MADS-box TFs have crucial roles in signaling and dormancy transition in several species (Mazzitelli et al. 2007; Hedley et al. 2010). They show high homology to two *Arabidopsis* genes, SHORT VEGETATIVE PHASE (SVP) and AGAMOUSLIKE 24 (AGL24), responsible for the regulation of flowering. By this context, mutation in DAM genes cause incomplete dormancy in peach (*Prunus persica*) (Horvath et al. 2008). In contrast to previous studies, there are just three differentially expressed MADS-box unigenes between Palaz and Çakıldak genotypes. Among them, the expression of

unigene32100 and CL3418 were up-regulated and CL4239 were down-regulated (Table S6). In addition to these TFs, a total of 53 SBP/SPL transcription factors related unigenes have been identified in hazelnut genome. A unigene (CL2147) annotated to squamosa promoter-binding like TF was differentially expressed among palaz and çakıldak hazelnut. According to the comparative transcriptome analysis between two genotypes, it was found that expression of this unigene was increased 11.62-fold between Çakıldak and Palaz genotypes. Since squamosa promoter-binding-like (SBP/SPL) transcription factor family members have role in diverse biological process including development, our findings showing increased expression of SBPs during leaf bud dormancy release supports previous studies and suggests that they could be involved in the formation or development of new leaf primordia (Schwarz et al. 2008; Howe et al. 2015). Unigenes encoding AP2 transcription factors were also up-regulated during leaf bud burst. Our study revealed that there were 76 AP2 unigenes and one of the AP2 member (CL5542) showed 11.85-fold increased expression between Çakıldak and Palaz. This result support general information about the role of AP2 transcription factor in dormancy release process and re-initiation of development by inhibiting the synthesis of abscisic acid (ABA) (Pandey et al. 2005).

Variations in gene expression between Çakıldak and Palaz

In order to determine the variations in gene expression between leaf buds of Çakıldak and Palaz, two Çakıldak (Cakildak1 + Cakildak2) and two Palaz (Palaz1 + Palaz2) libraries were compared. More than two-fold change of expression levels and the expression with significant difference ($P < 0.05$) were considered differentially expressed between samples. Results obtained from comparison of two samples showed that there were 8125 significantly differentially expressed unigenes. Among these unigenes, 4348 were upregulated and 3777 were downregulated. The significantly differentially expressed unigenes were further analyzed to determine their biological and molecular functions. Their functional categories were determined according to information in GO and KEGG databases.

In present study, we also focused on DEGs that showed 10-fold and more increased expression between two genotypes. One of the highly expressed unigene identified in hazelnut transcriptome was responsible from the production of serpin-ZX-like protein. Serpins have a highly-conserved core structure that is critical for their functioning as serine protease inhibitors (Huntington 2011). Hazelnut CL976 having homology with *Citrus sinensis* serpin-ZX-

like (LOC102629272) gene showed 11.8-fold increased expression pattern in Palaz genotype. Another highly expressed and annotated as actin related protein like-4 unigene was CL5247 (11.98-fold change). It was reported that plant actin genes have roles in stomatal closure, cell division and differentiation by regulating actin filaments in plant actin cytoskeleton (Kandasamy et al. 2004; Li et al. 2004). Although, there was no study revealing the functions of plant actin genes in dormancy release process, homolog of AT1G18450 gene (CL5247) was required to study in more detail because of its highly-increased expression level. According to comparative transcriptome analysis of two genotypes, it was determined that Unigene12494 encoding an ankyrin repeat-containing protein was differentially up-regulated (10.79-fold increase) in Palaz genotype. In previous studies, it was revealed that this group of proteins carried out crucial functions in cell cycle, differentiation, lateral root development, leaf formation, plant microbe interaction, pollen germination and pollen tube development by controlling the protein–protein interactions (Mou et al. 2013).

Functional classification of DEGs

GO and KEGG assignments were used to classify the functions of the DEGs identified in the pairwise comparisons of the Çakıldak and Palaz cDNA libraries. Then, we performed Gene Ontology (GO) classification and functional enrichment analysis. Based on GO analysis of DEGs between Çakıldak and Palaz, 8125 *C. avellena* sequences were categorized into 47 functional groups, belonging to three main GO ontologies: molecular function (13), cellular component (14) and biological process (20). Results indicated that a big portion of genes was from the categories of "cellular process (1174)", "metabolic process (1422)", "single-organism process (968), "cell (755)", "cell part (755)", "membrane (466)" "organelle (521)", "catalytic activity (1429)" and "binding activity (1166)" with only a few genes related to "developmental process (94)", "growth (27)", "nucleoid (2)", "electron carrier activity (37)", "enzyme regulator activity (19)" (Fig. 6). "Metabolic Process" and "Cellular Process are the two most abundant GO terms among Biological Process groups.

Fig. 6 GO classification of DEGs. X axis represents number of DEG. Y axis represents GO term
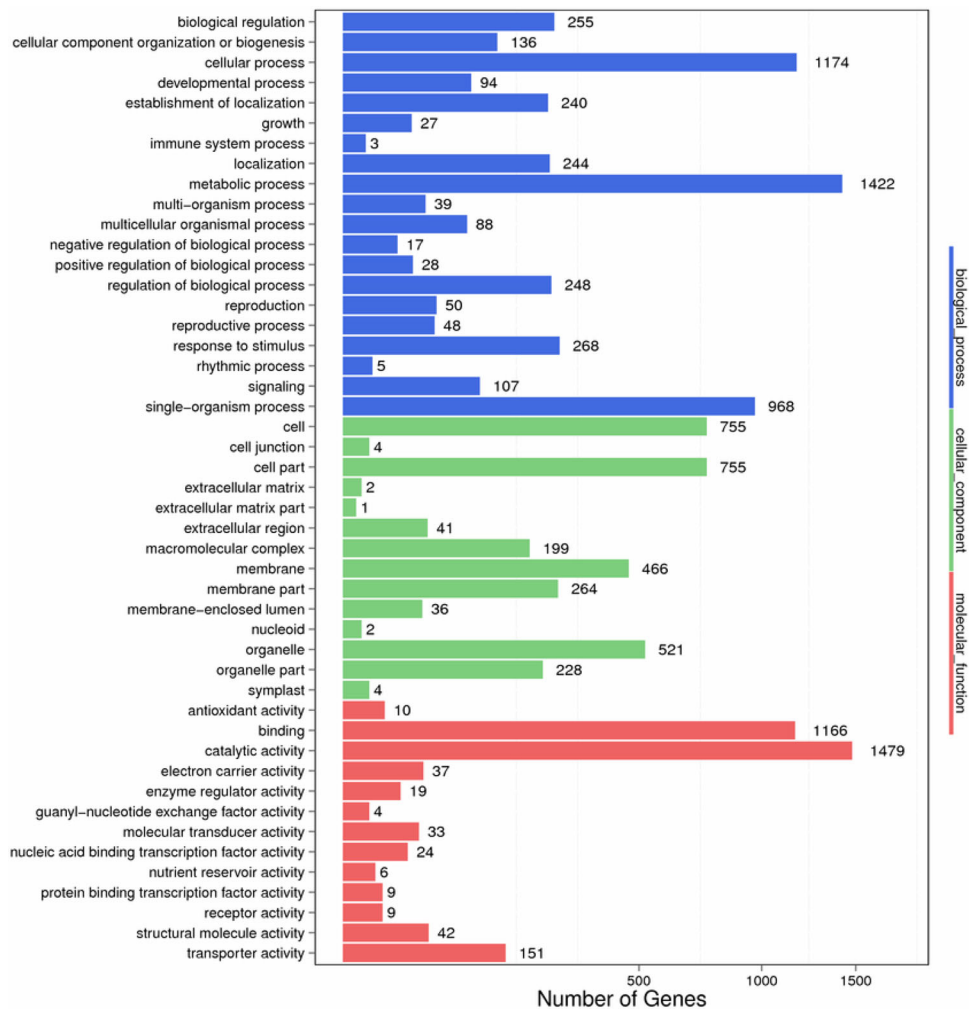
In fact, both metabolic and cellular processes containing genes involved in the most basic of life processes and metabolic changes have been widely reported in dormancy releasing process (Lesur et al. 2015; Qi et al. 2015). The increased energy and biosynthetic needs of dormancy releasing process and bud burst are generally provided by metabolic regulation (Lesur et al. 2015). The heading "Metabolic Process" includes nitrogen compound metabolic process, catabolic process, single organism metabolic process, and biosynthetic process. "Cellular Process" includes cellular aromatic compound metabolic process, organic cyclic compound metabolic process, and organonitrogen compound metabolic process. "Biological Regulation" includes the regulation of apoptosis, metabolism, cell cycle, translation, catalytic activity, and homeostasis. For the molecular function categories, several GO categories including binding, catalytic and transporter activity were enriched. Our GO annotation of *C. avellena* transcriptome, therefore, suggests presence of several transcripts associated with active metabolism and dormancy release process. GO functional enrichment analysis were carried out for three GO categories separately. L-Phenylalanine metabolic process (11 unigenes), erythrose 4-phosphate/phosphoenolpyruvate family amino acid metabolic process (11 unigenes), phenylpropanoid metabolic process (17 unigenes), L-phenylalanine catabolic process (6 unigenes) and cinnamic acid biosynthetic process (6 unigenes) are the most enriched biological processes. Based on analysis of cellular component terms, MCM complex, microtubule, Golgi apparatus part, microtubule cytoskeleton, cytoskeletal part and microtubule associated complex are identified as significantly enriched in Palaz libraries (Figure S2). By influencing the deposition direction of cellulose microfibrils, the plant microtubule cytoskeleton have a crucial role in growth and development of plants (Gardiner et al. 2012).

Pathway analysis of DEGs

KEGG pathway classification and functional enrichment for DEGs were carried out to determine the important pathways having roles in dormancy release and development. Based on gene set enrichment analysis of DEGs, biosynthesis of secondary metabolites, starch and sucrose metabolism and phenylpropanoid biosynthesis pathways were the most represented pathways in de novo assembled transcriptome of Palaz (Fig. 7). The starch and sucrose metabolism and phenylpropanoid biosynthesis pathways, important pathways for the plant development and growth, were assigned to 182 and 46 of differentially expressed unigenes, respectively (Table S7).

Pathways of DEGs was classified into six categories. These are cellular processes, environmental information processing, genetic information processing, metabolism and organismal systems. The largest proportion of the DEGs belonged to metabolism such as global and overview maps (1468 unigenes), carbohydrate metabolism (513 unigenes) and amino acid metabolism (296 unigenes) (Fig. 8). A total of 133 DEGs was assigned to glycan biosynthesis and metabolism, which have an important function in plant cell wall biosynthesis (von Schaewen et al. 2008). These findings were consistent with GO analysis since in plant cells, glycans attached to asparagine (N) residues of proteins undergo various modifications in the endoplasmic reticulum and the Golgi apparatus. One of the most enriched enzyme was pectin methyl esterase (PME) (EC 3.1.1.11) a member of starch and sucrose metabolism pathway. PME catalyzes the hydrolysis of methylester groups of cell wall pectins. Ren and Kermode (2000) reported that PME activity was positively correlated to the degree of dormancy breakage of yellow cedar seeds. In our study, a total 22 unigenes encoding PME showed differentially upregulated expression pattern between two samples (Table 3).

As a consequence of low temperature, there will be an increase of reactive oxygen species (ROS) formation due to the damage of membrane lipids and proteins (Prasad 1996). Likewise, the bud dormancy release period coincides with up-regulation of the antioxidant system (Faust and Wang 1993). There are many studies reporting that bud dormancy release is associated with activated antioxidant system including catalase (CAT), ascorbate peroxidase (APX), superoxide dismutase (SOD), glutathione-*S*-transferase (GS) and glutathione reductase (GR) (Bartolini et al. 2006; Ben Mohamed et al. 2012; Viti et al. 2012). It was found that the expression of several genes whose products are related with antioxidant mechanism, including APX (CL2105), CAT (Unigene21988, Unigene22264), GS (Unigene8147), SOD (Unigene19084) was increased during bud burst. One of the most prominent antioxidant enzyme was APX (EC:1.11.1.11) in ascorbate and aldorate metabolism. The APX enzyme-encoding gene (CL2105) was differentially up-regulated with 5.2-fold change between Çakıldak and Palaz (Table 3). The other antioxidant enzyme encoding genes also showed similar expression patterns during leaf bud burst period.

Transcripts related to phenylpropanoid biosynthesis and flavonoid biosynthesis pathway

A variety of structural and defense compounds in plants are synthesized through the phenylpropanoid pathway (Ververidis et al. 2007). Each part of this pathway is responsible for the synthesis of the different phenolic secondary metabolites, such as flavonoids, coumarins, and lignin. These compounds have different roles in metabolism

Fig. 7 Pathway functional enrichment of DEGs. X axis represents enrichment factor. Y axis represents pathway name. Coloring indicate q value (high: white, low: blue), the lower q value indicate the more significant enriched. Point size indicate DEG number (more: big, less: small) (colour figure online)
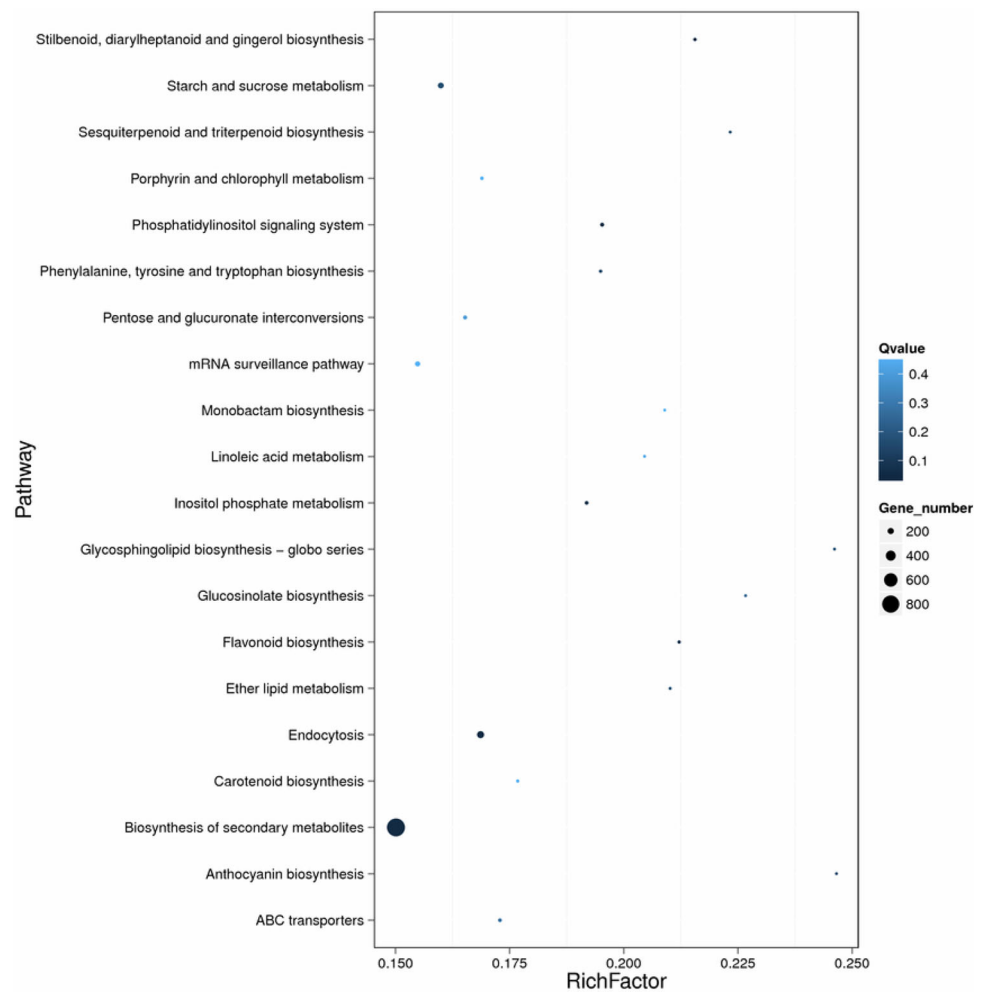

including protecting against oxidative stress, participating in developmental and stress signaling, mediating plant–microbe interactions, and defense against pathogens and pests (Barber and Mitchell 1997; Weisshaar and Jenkins 1998; Vogt 2010; Jones et al. 2012; Fan et al. 2015). Based on the KEGG annotated sequences, enzymes involved in phenylpropanoid biosynthesis pathway that are considered to be involved in development and dormancy release were visualized and colored in the pathway-maps obtained from KEGG database (Figure S3). There are 140 DEGs that translate to 12 enzymes in the phenylpropanoid biosynthesis pathway. The three most important enzymes in the phenylpropanoid biosynthesis are phenylalanine ammonia lyase (PAL, EC:4.3.1.24), cinnamoyl-CoA reductase (CCR, EC:1.2.1.44), 4-coumarate-CoA ligase (4CL, EC:6.2.1.12), which were represented by 3, 8, and 4 DEGs respectively. All PAL encoding genes were differentially up-regulated between Çakıldak and Palaz. For instance, the expression of CL2676 was increased significantly between Çakıldak and Palaz transcriptome. Several copies of PAL-genes are found in all plant species, comprising four genes

in Arabidopsis, five genes in willow and five in poplar (Cochrane et al. 2004; Tsai et al. 2006; de Jong et al. 2015). In contrast to PAL encoding genes, there are both up and down-regulated expression pattern in 4CL and CCR pathway related genes. Additionally, we identified that the unigene27109 annotated to 4CL enzyme was highly expressed between Çakıldak and Palaz genotypes (Table 3).

Another group of genes found in phenylpropanoid pathways and differentially expressed during dormancy release period are associated with secondary metabolism, specifically those leading to flavonoid biosynthesis. This groups included the genes encoding chalcone flavone isomerase, chalcone synthase, chalcone isomerase and flavones. They have been reported to be significantly regulated during dormancy release in variety of plants including raspberry (Mazzitelli et al. 2007) and leafy spurge (Horvath et al. 2006). There are 5 different genes encoding these enzymes in hazelnut transcriptome. CL7570 and CL658 annotated to chalcone synthase were identified as differentially upregulated with 2 and 1.2-fold

Fig. 8 Pathway classification of DEGs. X axis represents number of DEG. Y axis represents pathway name
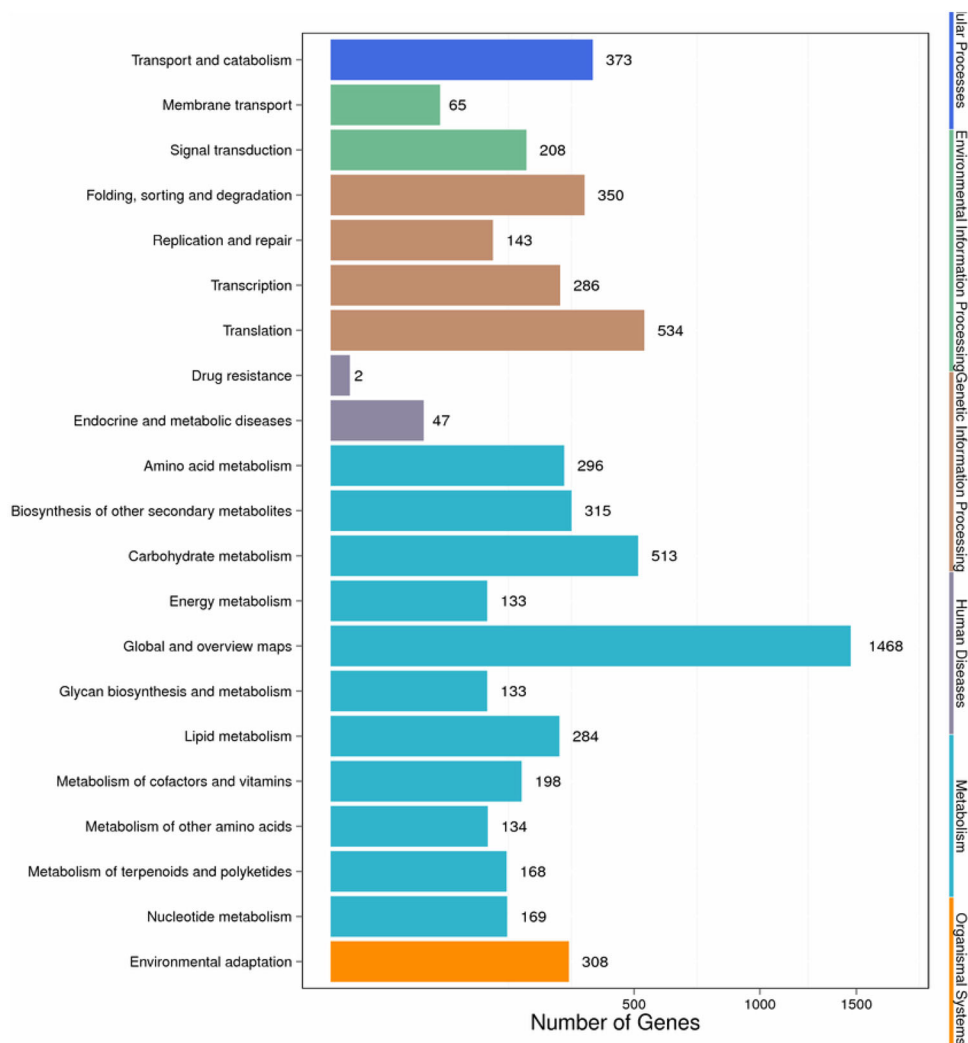
change expression value between Çakıldak and Palaz transcriptome, respectively. There were 2 homolog genes encoding for chalcone isomerase, CL8598 and unigene13134. Among them just CL8598 was differentially down-regulated in Palaz (Table 3).

Transcripts related to phytohormone biosynthesis and signaling

Phytohormones play important roles in plant growth and development. Bud burst process involves an interaction between plant growth inhibiting substances such as, abscisic acid (ABA) and other plant hormones including auxin, gibberellins (GA) and ethylene (Rodriguez and Sancheztames 1986). Gibberellin type of phytohormones have many roles in plant life cycle. It has been reported that they are required in germination of seed, triggering of vegetative growth, and development of floral and seed (Michalczuk 2005). GID1 and Della encoding genes were enriched in diterpenoid biosynthesis pathway when

compared two hazelnut transcriptomes (Figure S4). GID1 is a soluble protein that interacts directly with GA, triggering the gibberellin response (García-Martinez and Gil 2001). There were 4 DEGs encoding GID1 in hazelnut transcriptome. CL1560, CL3178, CL7872 and Unigene21953 have been upregulated between two hazelnut sequencing libraries. Similar results have been obtained with different plants including grape and leafy spurge (Mathiason et al. 2009; Dogramaci et al. 2010). Additionally, it was reported that GID1 encoding gene was overexpressed during eco-dormancy period in oak leaf bud (Ueno et al. 2013). The Della proteins, which have role in repressing of GA-dependent, are the most studied components of the GA signaling pathway (Silverstone et al. 1997; Mutasa-Göttgens and Hedden 2009; Li et al. 2016). There are four differentially expressed genes encoding Della proteins in hazelnut. Among them, unigene6641, unigene440 and CL1163 were differentially upregulated with 3.2, 1.5 and 1.0-fold change value between Çakıldak and Palaz transcriptome, respectively (Table 3). However,

Table 3 Expression values and annotation of bud burst related genes

Gene	Cakildak expression	Palaz expression	Log$_2$ fold change	P value	Up/down regulation	Annotation
CL1051	0	57.99	5.25692	2.19E−08	Up	PME
CL15	104.81	405.64	1.865676	1.47E−13	Up	PME
CL331	29.92	110.64	1.720042	0.000485	Up	PME
CL3818	148.66	434.29	1.459538	1.19E−07	Up	PME
CL4859	117.82	360.19	1.529959	5.13E−09	Up	PME
CL8545	0	41.25	4.837783	4.73E−07	Up	PME
CL9501	51.43	193.18	1.800725	1.52E−07	Up	PME
CL9979	3.98	35.65	2.41824	0.005588	Up	PME
CL9979	0.5	38.3	4.124141	1.06E−05	Up	PME
CL999	96.77	306.74	1.565104	1.55E−06	Up	PME
Unigene23379	3.48	46.89	2.738842	0.002824	Up	PME
Unigene36564	0	150.02	6.556023	4.70E−14	Up	PME
Unigene4443	8.98	58.57	2.377297	0.000211	Up	PME
Unigene4496	163.21	499.02	1.535606	1.53E−12	Up	PME
Unigene6981	166.15	439.91	1.324007	1.70E−07	Up	PME
Unigene7212	61.87	143.26	1.111535	0.003553	Up	PME
Unigene7396	127.27	488.77	1.859306	3.45E−16	Up	PME
Unigene9231	89.81	205.5	1.103132	0.000991	Up	PME
CL2105	0	55.45	5.187076	3.88E−08	Up	APX
Unigene21988	1	29.06	3.472702	0.000139	Up	CAT
Unigene22264	0	14.28	3.191483	0.002411	Up	CAT
Unigene8147	5	130.41	4.232768	1.24E−12	Up	GS
Unigene19084	10	144.77	3.557329	5.30E−12	Up	SOD
CL2676	18.97	77.59	1.818471	0.001306	Up	PAL
CL8386	102.31	338.03	1.635794	1.33E−09	Up	PAL
Unigene24019	79.34	176.43	1.060828	0.002363	Up	PAL
Unigene27684	11.49	129.98	3.231051	1.98E−10	Up	CCR
CL1734	95.85	363.33	1.835191	1.65E−12	Up	CCR
CL5628	94.37	227.72	1.186655	0.000135	Up	CCR
Unigene6632	45.43	0	− 5.08995	6.95E−08	Down	CCR
CL4563	2666.7	111.5	− 4.59943	1.43E−124	Down	CCR
CL2809	75.93	2.56	− 4.20982	1.04E−08	Down	CCR
Unigene6402	142.29	7.13	− 4.08167	9.16E−15	Down	CCR
Unigene27109	15.97	91.75	2.310078	6.49E−06	Up	4CL
CL1942	111.82	566.32	2.256321	3.59E−25	Up	4CL
CL9242	215.36	695.73	1.605326	3.95E−08	Up	4CL
Unigene9496	82.37	184.04	1.073989	0.000801	Up	4CL
CL7570	766.46	3322.74	2.046554	2.20E−48	Up	Chalcone synthase
CL658	2564.18	6232.62	1.217948	4.24E−30	Up	Chalcone synthase
CL8598	133.83	31.65	− 2.04304	5.64E−06	Down	Chalcone isomerase
CL1560	146.23	374.26	1.27641	2.83E−07	Up	GID1
CL3178	13.99	112.08	2.757481	1.11E−07	Up	GID1
CL7872	30.43	123.39	1.860273	5.70E−05	Up	GID1
Unigene21953	3774.65	10,372.5	1.395144	1.08E−34	Up	GID1
Unigene6641	11.98	133.52	3.225336	4.42E−11	Up	Della
Unigene440	133.68	405.2	1.507629	3.66E−07	Up	Della
CL1163	203.56	431.14	1.00698	7.43E−05	Up	Della
CL1210	60.9	13.77	− 2.02541	0.000531	Down	Della

Table 3 continued

Gene	Cakildak expression	Palaz expression	Log$_2$ fold change	P value	Up/down regulation	Annotation
CL7563	303.12	18.86	− 3.93939	1.11E−26	Down	TIR1
CL9080	156.66	1183.78	2.829942	1.38E−44	Up	TIR1
CL93	0	91.33	6.010007	1.31E−11	Up	Auxin response factor
CL283	21.02	211.2	2.961044	2.16E−06	Up	Auxin response factor
CL4416	10.98	89.19	2.746923	6.02E−07	Up	5NG4
CL3309	0	51.89	5.094513	7.97E−08	Up	5NG4
CL4303	454.34	181.86	− 1.36268	4.16E−09	Down	ARP
CL1901	263.68	74.42	− 1.84549	1.14E−09	Down	ABA biosynthesis

CL1210 was down regulated (-2-fold change). Another important dormancy related phytohormone is auxin that is generally up-regulated during bud burst period. Auxin have role in regulation of plant growth and development. It regulates the expression of related genes via the degradation of a group of repressor proteins known as Aux/IAA proteins. Auxin induce three important gene families. They are SAUR, Aux/IAA, and GH3 (Guilfoyle et al. 1998). There were 11 DEGS unigenes encoding auxin receptor (TIR1), auxin/indole 3-acetic acid (AUX/IAA) and auxin response factor (ARF) proteins involved in plant hormone signal transduction pathway. TIR1/AFBs, members of F-box proteins, is responsible for the perception of auxin. These proteins have been classified into four groups such as TIR1, AFB2, AFB4, and AFB6 in flowering plants. Members of TIR1 and AFB2 groups have been reported as regulators of auxin signaling by promoting the degradation of the Aux/IAA transcriptional repressors (Prigge et al. 2016). In our study, genes that encode TIR1 were both down-regulated (CL7563) and up-regulated (CL9080) during bud dormancy release process. It was reported that there was another important auxin related element having roles in dormancy release process. It was called as monopteros homologue or auxin response factor and their expressions were increased during bud burst (Derory et al. 2006). There were two differentially expressed monopteros homologue gene in hazelnut transcriptome. In transcriptome of Palaz genotype, CL93 showed six-fold higher expression pattern when compared to Çakıldak genotype. Likewise, it was observed that expression of CL283 increased 3 times from Çakıldak to Palaz (Table 3). In addition to these genes, there were also other differentially expressed unigenes encoding auxin induced protein 5NG4. In previous study conducted with *Cunninghamia lanceolate,* the expression of this gene was increased in induced leaf bud during dormancy release period (Xu et al. 2016). We identified three homolog 5NG4 genes in *C. avellana* transcriptome. These were CL4416, CL3309 and CL124 and their expression increased 2.74, 5 and 5.7-fold between two genotypes, respectively. The other DEGs in this pathway was auxin repressed (ARP) protein like gene. It was reported that some of the homologs of this gene have roles in dormancy and inhibition of vegetative growth. For instance, there was negative correlation between shoot growth and expression of this gene in *Robinia pseudoacacia* (Park and Han 2003). In our study, the homolog of ARP (CL4303) was down-regulated (-1.3-fold) in Palaz early leaf developed genotype.

One of the most important player in dormancy process is ABA. This phytohormone have function starting of dormancy by interaction with other phytohormones. In this context, it was thought that enzymes responsible for degrading of ABA have roles in dormancy release process of leaf buds (Zheng et al. 2015). The enzyme abscisic-aldehyde oxidase (AAO3) have critical role in this process, since it is found in the last step of biosynthesis of ABA (Kushiro et al. 2004). According to the comparative expression analysis of Palaz and Çakıldak, it was found that CL1901 a homolog of Arabidopsis AAO3 has a reduced activity in Palaz (Table 3).

Confirmation of candidate DEGs by qRT-PCR analysis

To experimentally confirm the Illumina RNA-Seq results, a subset of 16 genes in phenylpropanoid (PP) and plant hormone (PH) synthesis pathways, which were found to be differentially up-regulated in Palaz libraries, was chosen for quantitative real-time PCR (qRT PCR) analysis. Although different optimization conditions were performed, amplification of CL6063 could not be achieved. According to the results, the expressions of CL1469, CL1560, CL3668, CL8248, unigene28358, unigene4384, unigene4817 and unigene6641 were upregulated in Palaz variety when compared to that in Çakıldak variety (Fig. 9). Furthermore, unigene 28358 expression levels exhibited the highest increase in Palaz variety compared to Çakıldak with a maximum of 16.2-fold more expression value. On the other hand, CL2140, CL5737, CL483, CL7825, unigene15907, and unigene3 gene expressions were shown to
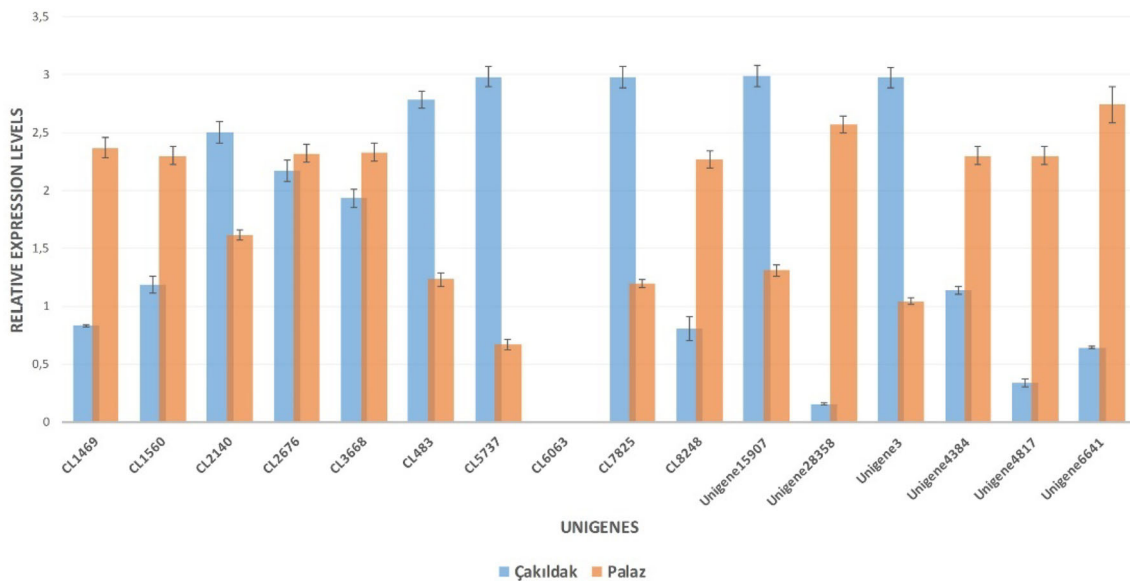
Fig. 9 Relative expression levels of unigenes from phenylpropanoid (PP) and plant hormone (PH) synthesis pathways in Çakıldak and Palaz genotypes

be downregulated in Palaz variety as compared to that in the Çakıldak. The most drastic expression decrease was recorded in CL5737 with a 4.96-fold reduced expression level in Palaz. Four of six down regulated genes in Palaz variety were found to be members of phenylpropanoid pathway. Among them, CL7825 is responsible for formation of xyloglucosyl transferase. However, five of eight up-regulated genes in Palaz variety belong to plant hormone synthesis pathway. Of these, Unigene6641 encodes for DELLA, unigene4383 encodes for SAUR, CL8248 encodes for JAZ (jasmonate zym domain) and CL1560 encodes for gibberellins. The qRT-PCR results that show upregulation of CL1469, CL1560, CL3668, CL8248, unigene4384 and unigene6641 gene expression in Palaz variety confirm the data obtained from transcriptome results.

Conclusion

Transcriptome sequencing analysis was carried to elucidate the major molecular mechanism in bud burst of hazelnut plant. The expression profiling analysis was done using buds in the ecodormancy and bud flush stages. Generally, phenylpropanoid and phytohormone biosynthesis pathway genes and gene sets that are significantly enriched have been indicated to exhibit crucial roles in regulation mechanisms of bud burst. By using transcription factor identification studies, the key transcription factors involved in bud burst process were highlighted, including MADS-box related, MYB and bHLH gene family. We further reported that bud burst process involves an interaction

between plant growth inhibiting substances such as, abscisic acid (ABA) and other plant hormones including auxin, gibberellins (GA) and ethylene. These results will increase the knowledge about the molecular mechanism of bud burst process in hazelnut plant.

Compliance with ethical standards

Data archiving statement The sequencing datasets were deposited in NCBI SRA database under project number accession PRJNA316492. The experiment accession numbers of cDNA libraries obtained from the Palaz 1, Palaz 2, Çakıldak 1 and Çakıldak are SRX1665375, SRX1665376, SRX1665377, and SRX1665378, respectively. Assembled transcript sequences was deposited in TSA (SUB4195016).

Conflict of interest The authors have no conflict of interest.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Azad MS (2012) Observations on bud burst phenology in a field trial established with Poplar (*Populus* spp.). For Stud China 14:251–259

Barber MS, Mitchell HJ (1997) Regulation of phenylpropanoid metabolism in relation to lignin biosynthesis in plants. Int Rev Cytol 172:243–293

Bartolini S, Zanol GC, Viti R (2006) Changes in antioxidant compounds in flower buds of two apricot cultivars during winter season. International Society for Horticultural Science (ISHS), Leuven, pp 69–74

Ben Mohamed H, Vadel AM, Geuns JMC, Khemira H (2012) Effects of hydrogen cyanamide on antioxidant enzymes' activity, proline and polyamine contents during bud dormancy release in Superior Seedless grapevine buds. Acta Physiol Plant 34:429–437

Besford RT, Hand P, Richardson CM, Peppitt SD (1996) Photoperiod effect on bud burst in Prunus is phase dependent: significance for early photosynthetic development. Tree Physiol 16:491–496

Bhandawat A, Singh G, Seth R, Singh P, Sharma RK (2017) Genome-widetranscriptional profiling to elucidate key candidates involved in bud burst and rattling growth in a subtropical bamboo (Dendrocalamus hamiltonii). Front Plant Sci 7:2038

Boccacci P et al (2013) Molecular and morphological diversity of on-farm hazelnut (Corylus avellana L.) landraces from southern Europe and their role in the origin and diffusion of cultivated germplasm. Tree Genet Genomes 9:1465–1480

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11(2):113–116

Christian I, Victor CJ, Philipp B (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol 7:138–148

Cochrane FC, Davin LB, Lewis NG (2004) The Arabidopsis phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL isoforms. Phytochemistry 65:1557–1564

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

de Jong F, Hanley SJ, Beale MH, Karp A (2015) Characterisation of the willow phenylalanine ammonia-lyase (PAL) gene family reveals expression differences compared with poplar. Phytochemistry 117:90–97

Delgado T, Malheiro R, Pereira JA, Ramalhosa E (2010) Hazelnut (Corylus avellana L.) kernels as a source of antioxidants and their potential in relation to other nuts. Ind Crops Prod 32:621–626

Derory J et al (2006) Transcriptome analysis of bud burst in sessile oak (Quercus petraea). New Phytol 170:723–738

Dogramaci M, Horvath DP, Chao WS, Foley ME, Christoffers MJ, Anderson JV (2010) Low temperatures impact dormancy status, flowering competence, and transcript profiles in crown buds of leafy spurge. Plant Mol Biol 73:207–226

Fan GQ, Xu EK, Deng MJ, Zhao ZL, Niu SY (2015) Phenylpropanoid metabolism, hormone biosynthesis and signal transduction-related genes play crucial roles in the resistance of Paulownia fortunei to paulownia witches' broom phytoplasma infection. Genes Genomics 37:913–929

Faust M, Wang SY (1993) Biochemical events associated with resumption of growth in temperate zone fruit trees. International Society for Horticultural Science (ISHS), Leuven, pp 257–264

Faust M, Erez A, Rowland LJ, Wang SY, Norman HA (1997) Bud dormancy in perennial fruit trees: physiological basis for dormancy induction, maintenance, and release. HortScience 32:623–629

Fuchigami LH, Wisniewski M (1997) Quantifying bud dormancy: physiological approaches. HortScience 32:618–623

García-Martinez LJ, Gil J (2001) Light regulation of gibberellin biosynthesis and mode of action. J Plant Growth Regul 20:354–368

Gardiner J, Overall R, Marc J (2012) Plant microtubule cytoskeleton complexity: microtubule arrays as fractals. J Exp Bot 63:635–642

Grabherr MG et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–730

Guilfoyle T, Hagen G, Ulmasov T, Murfett J (1998) How does auxin turn on genes? Plant Physiol 118:341–347

Hakkinen R, Linkosalo T, Hari P (1998) Effects of dormancy and environmental factors on timing of bud burst in Betula pendula. Tree Physiol 18:707–712

Hao X, Yang Y, Yue C, Wang L, Horvath DP, Wang X (2017) Comprehensive transcriptome analyses reveal differential gene expression profiles of Camellia sinensis axillary buds at para-, endo-, ecodormancy, and bud flush stages. Front Plant Sci 8:553

Hedley PE et al (2010) Candidate genes associated with bud dormancy release in blackcurrant (Ribes nigrum L.). BMC Plant Biol 10:1–13

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol Biol Evol 20:735–747

Horvath DP, Anderson JV, Soto-Suarez M, Chao WS (2006) Transcriptome analysis of leafy spurge (Euphorbia esula) crown buds during shifts in well-defined phases of dormancy. Weed Sci 54:821–827

Horvath DP, Chao WS, Suttle JC, Thimmapuram J, Anderson JV (2008) Transcriptome analysis identifies novel responses and potential regulatory genes involved in seasonal dormancy transitions of leafy spurge (Euphorbia esula L.). BMC Genomics 12(9):536

Howe GT, Horvath DP, Dharmawardhana P, Priest HD, Mockler TC, Strauss SH (2015) Extensive transcriptome changes during natural onset and release of vegetative bud dormancy in populus. Front Plant Sci 6:989

Huntington JA (2011) Serpin structure, function and dysfunction. J Thromb Haemost 9(Suppl 1):26–34

Jiminez S, Reighard GL, Bielenberg DG (2010) Gene expression of DAM5 and DAM6 is suppressed by chilling temperatures and inversely correlated with bud break rate. Plant Mol Biol 73(1–2):157–167

Jin HL, Martin C (1999) Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol 41:577–585

Jones AMP, Chattopadhyay A, Shukla M, Zon J, Saxena PK (2012) Inhibition of phenylpropanoid biosynthesis increases cell wall digestibility, protoplast isolation, and facilitates sustained cell division in American elm (Ulmus americana). BMC Plant Biol 12:75

Kandasamy MK, Deal RB, McKinney EC, Meagher RB (2004) Plant actin-related proteins. Trends Plant Sci 9:196–202

Khalil-Ur-Rehman M, Sun L, Li C-X, Faheem M, Wang W, Tao J-M (2017) Comparative RNA-seq based transcriptomic analysis of bud dormancy in grape. BMC Plant Biol 17:18

Köksal İ (2002) Turkish hazelnut cultivars. Hazelnut Promotion Group, Ankara

Kushiro T et al (2004) The Arabidopsis cytochrome P450 CYP707A encodes ABA 8′-hydroxylases: key enzymes in ABA catabolism. EMBO J 23:1647–1656

Lang GAMG (1987) Endo-, para-, and eco-dormancy: physiological terminology and classification for dormancy research. HortScience 22:371–377

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Lesur I et al (2015) The oak gene expression atlas: insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. BMC Genomics 16(1):112

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 4(12):323

Li Y, Sorefan K, Hemmann G, Bevan MW (2004) Arabidopsis NAP and PIR regulate actin-based cell morphogenesis and multiple developmental processes. Plant Physiol 136:3616–3627

Li Z et al (2016) Transcriptome sequencing determined flowering pathway genes in *Aechmea fasciata* treated with ethylene. J Plant Growth Regul 35:316–329

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550

Ma Q-H, Luo Q-H, Aijier A, Cao Y-Q, Cui L-Z, Huo H-L, Wang G-X (2016) Chilling and heat requirements of main cultivars of Ping'ou hybrid hazelnuts (*Corylus heterophylla* × *Corylus avellana*). For Res 29:354–361

Mathiason K, He D, Grimplet J, Venkateswari J, Galbraith DW, Or E, Fennell A (2009) Transcript profiling in *Vitis riparia* during chilling requirement fulfillment reveals coordination of gene expression patterns with optimized bud break. Funct Integr Genomics 9:81–96

Mazzitelli L et al (2007) Co-ordinated gene expression during phases of dormancy release in raspberry (*Rubus idaeus* L.) buds. J Exp Bot 58:1035–1045

Mehlenbacher SA (1991) Chilling requirements of hazelnut cultivars. Sci Hortic 47:271–282

Michalczuk L (2005) Hormonal control of dormancy. Int J Fruit Sci 5:59–73

Mou S, Liu Z, Guan D, Qiu A, Lai Y, He S (2013) Functional analysis and expressional characterization of rice ankyrin repeat-containing protein, OsPIANK1, in basal defense against Magnaporthe oryzae attack. PLoS ONE 8:e59699

Mutasa-Göttgens E, Hedden P (2009) Gibberellin as a factor in floral regulatory networks. J Exp Bot 60:1979–1989

Pandey GK, Grant JJ, Cheong YH, Kim BG, Li LG, Luan S (2005) ABR1, an APETALA2-domain transcription factor that functions as a repressor of ABA response in Arabidopsis. Plant Physiol 139:1185–1193

Park S, Han KH (2003) An auxin-repressed gene (RpARP) from black locust (*Robinia pseudoacacia*) is posttranscriptionally regulated and negatively associated with shoot elongation. Tree Physiol 23:815–823

Pertea G et al (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652

Pletsers A, Caffarra A, Kelleher CT, Donnelly A (2015) Chilling temperature and photoperiod influence the timing of bud burst in juvenile *Betula pubescens* Ehrh. and *Populus tremula* L. trees. Ann For Sci 72:941–953

Prasad TK (1996) Mechanisms of chilling-induced oxidative stress injury and tolerance in developing maize seedlings: changes in antioxidant system, oxidation of proteins and lipids, and protease activities. Plant J 10:1017–1026

Prigge MJ et al (2016) The Arabidopsis auxin receptor F-box proteins AFB4 and AFB5 are required for response to the synthetic auxin picloram. G3 Genes Genomics Genet 6:1383–1390

Qi JJ, Sun P, Liao DQ, Sun TY, Zhu J, Li XN (2015) Transcriptomic analysis of American ginseng Seeds during the dormancy release process by RNA-Seq. PLoS ONE 10(3):e0118558

Ren CW, Kermode AR (2000) An increase in pectin methyl esterase activity accompanies dormancy breakage and germination of yellow cedar seeds. Plant Physiol 124:231–242

Ren Y, Zhao Q, Zhao X, Hao Y, You C (2016) Expression analysis of the MdCIbHLH1 gene in apple flower buds and seeds in the process of dormancy. Hortic Plant J 2:61–66

Riano-Pachon DM, Ruzicic S, Dreyer I, Mueller-Roeber B (2007) PlnTFDB: an integrative plant transcription factor database. BMC Bioinform 8:42

Rodriguez A, Sancheztames R (1986) Dormancy and seasonal-changes of plant-growth regulators in hazel buds. Physiol Plant 66:288–292

Rowley ER, Bryant DW, Fox SE, Givan SA, Mehlenbacher SA, Mockler TC (2014) Genome sequencing and resource development for European hazelnut. International Society for Horticultural Science (ISHS), Leuven, pp 75–78

Ruttink T et al (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. Plant Cell 19:2370–2390

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nat Protocols 3:1101–1108

Schwarz S, Grande AV, Bujdoso N, Saedler H, Huijser P (2008) The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. Plant Mol Biol 67:183–195

Shim D, Ko J-H, Kim W-C, Wang Q, Keathley DE, Han K-H (2014) A molecular framework for seasonal growth-dormancy regulation in perennial plants. Hortic Res 1:14059

Silverstone AL, Chang CW, Krol E, Sun TP (1997) Developmental regulation of the gibberellin biosynthetic gene GA1 in *Arabidopsis thaliana*. Plant J 12:9–19

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212

Singh RK, Svystun T, AlDahmash B, Jonsson AM, Bhalerao RP (2017) Photoperiod- and temperature-mediated control of phenology in trees—a molecular perspective. New Phytol 213:511–524

Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278:631–637

Tsai CJ, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan YN (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. New Phytol 172:47–62

Ueno S et al (2013) Transcriptional profiling of bud dormancy induction and release in oak by next-generation sequencing. BMC Genomics 14:1–15

Ververidis F, Trantas E, Douglas C, Vollmer G, Kretzschmar G, Panopoulos N (2007) Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: chemical diversity, impacts on plant biology and human health. Biotechnol J 2:1214–1234

Viti R, Bartolini S, Zanol GC (2012) Biological changes and active oxygen-scavenging enzymes activities in Apricot (*Prunus armeniaca* L.) flower buds during dormancy transitions. International Society for Horticultural Science (ISHS), Leuven, pp 331–339

Vogt T (2010) Phenylpropanoid biosynthesis. Mol Plant 3:2–20

von Schaewen A, Frank J, Koiwa H (2008) Role of complex N-glycans in plant stress tolerance. Plant Signal Behav 3:871–873

Weisshaar B, Jenkins GI (1998) Phenylpropanoid biosynthesis and its regulation. Curr Opin Plant Biol 1:251–257

Xu H et al (2016) Gene expression and proteomic analysis of shoot apical meristem transition from dormancy to activation in *Cunninghamia lanceolata* (Lamb.) Hook. Sci Rep 6:19938

Yıldırım K, Kaya Z (2017) Gene regulation network behind drought escape, avoidance and tolerance strategies in black poplar (*Populus nigra* L.). Plant Physiol Biochem 115:183–199

Zheng C, Halaly T, Acheampong AK, Takebayashi Y, Jikumaru Y, Kamiya Y, Or E (2015) Abscisic acid (ABA) regulates grape bud dormancy, and dormancy release stimuli may act through modification of ABA metabolism. J Exp Bot 66(5):1527–1542