Original Article

De novo Transcriptome Analysis and Gene Expression Profiling of Corylus Species

(*Corylus* spp. / *de novo* / transcriptome / hazelnut / RNA-seq)

SALIH ULU[1], ZEHRA OMEROGLU ULU[1,2], AYSUN AKAR[3], NEHIR OZDEMIR OZGENTURK[1]

[1]Department of Molecular Biology and Genetics, Faculty of Art and Science, Yildiz Technical University, Istanbul, Turkey
[2]Department of Genetics and Bioengineering, Faculty of Engineering, Yeditepe University, Istanbul, Turkey
[3]Hazelnut Research Institution, Ministry of Food, Agriculture and Livestock, Giresun, Turkey

Abstract. Hazelnut (*Corylus*), which has high commercial and nutritional benefits, is an important tree for producing nuts and nut oil consumed as ingredient especially in chocolate. While *Corylus avellana* L. (European hazelnut, Betulaceae) and *Corylus colurna* L. (Turkish hazelnut, Betulaceae) are the two common hazelnut species in Europe, *C. avellana* L. (Tombul hazelnut) is grown as the most widespread hazelnut species in Turkey, and *C. colurna* L., which is the most important genetic resource for hazelnut breeding, exists naturally in Anatolia. We generated the transcriptome data of these two *Corylus* species and used these data for gene discovery and gene expression profiling. Total RNA from young leaves, flowers (male and female), buds, and husk shoots of *C. avellana* and *C. colurna* were used for two different libraries and were sequenced using Illumina HiSeq4000 with 100 bp paired-end reads. The transcriptome data 10.48 and 10.30 Gb of *C. avellana* and *C. colurna*, respectively, were assembled into 70,265 and 88,343 unigenes, respectively. These unigenes were functionally annotated using the TRAPID platform. We identified 25,312 and 27,051 simple sequence repeats (SSRs) for *C. avellana* and *C. colurna*, respectively. *TL1*, *GMPM1*, *N*, *2MMP*, *At1g29670*, *CHIB1* unigenes were selected for validation with qPCR. The first *de novo* transcriptome data of *C. colurna* were used to compare data of *C. avellana* of commercial importance. These data constitute a valuable extension of the publicly available transcriptomic resource aimed at breeding, medicinal, and industrial research studies.

Received July 31, 2023. Accepted November 20, 2023.

This study was supported by Yildiz Technical University Scientific Research Projects Coordination Unit under project number FBA-2017-3134.

Corresponding author: Nehir Ozdemir Ozgenturk, Department of Molecular Biology and Genetics, Faculty of Art and Science, Yildiz Technical University, Davutpaşa Kampus, D1004 Esenler, Istanbul, Turkey; e-mail: nehirozdemir@yahoo.com

Abbreviations: GO – Gene Ontology, KEGG – Kyoto Encyclopaedia of Genes and Genomes, NGS – next-generation sequencing, ORF – open reading frame, qPCR – quantitative polymerase chain reaction, RNA-seq – RNA sequencing, SSRs – simple sequence repeats.

Introduction

Hazelnuts (*Corylus* spp.) are among the most beneficial and significant plants known to humanity for their nut crops and woody oil (Li et al., 2020). *Corylus* species have been specifically reproduced to cultivate comestible nuts, but almost all parts of the plant, including leaves, husks, bark, roots, nuts, shells and wood, are used. Uses include animal husbandry, food, food production, health products, alcoholic beverages, fertilizers, biofuels, industrial chemicals and products, craftwork, oils, farm implements, pharmaceuticals, and wildlife habitat (Sullivan et al., 2014). *Corylus avellana* L. (*C. avellana* L.) is the European hazelnut and is the most widely distributed and economically major species in Turkey. It thrives in moderate, moist climates and is naturally found in the Turkey's Black Sea region, where it contributed around 65 % of the world's supply in 2020 (FAO, 2020; Lucas et al., 2021). *C. avellana* L. contains several compounds with beneficial biological activity, such as flavonoids, tannins, phenolic acids, diarylheptanoids, taxanes, and lignans (Boccacci et al., 2006). *Corylus colurna* L. (*C. colurna* L.) is the Turkish hazelnut, is grown naturally in Anatolia and one of the significant hazelnut species in Turkey and used as rootstocks. It has the ability to adapt to the extreme climate and poor soil conditions (Ayan et al., 2018). *C. colurna* L. could also be considered as a resource of pharmaco-

logically substantial biological compounds. Nevertheless, its chemical structure has not yet been studied in detail. According to the previous literature data, the extract of *C. colurna* leaves has remarkable antioxidant activity (Benov and Georgiev, 1994; Riethmüller et al., 2014).

Recent improvements in sequencing and computational tools have substantially aided research into reference-guided or *de novo* assembly (Han et al., 2015; Thole et al., 2019; Raghavan et al., 2022). Because there is no external information available to guide the developmental process, *de novo* transcriptome assembly is a 'reference-free' process. The reference guide can be achieved from the information contained in the reads alone. Therefore, *de novo* transcriptome data of the plant parts supplies very important information for understanding the functional and metabolic pathways and identifying simple sequence repeats (SSRs). Furthermore, it allows discovery of novel genes and bioactive compounds (Peona et al., 2018; Thakur and Randhawa, 2018; Raghavan et al., 2022).

In this report, a *de novo* transcriptome assembly approach was used for *C. avellana* cv. Tombul, one of the most commercially important European hazelnut cultivars for Turkey, and the Turkish hazelnut *C. colurna*. At present, few resources are accessible for *C. avellana* cv. Tombul, and these are limited to the chloroplast genome and SSRs (Kahraman and Lucas, 2019). The *C. colurna* genome has not yet been published. The objectives of this research were to establish a protocol for determining a wide range of plant tissue specificity of assembled transcripts: (1) RNA-seq assembly and annotation for the functional understanding of *C. avellana* cv. Tombul and *C. colurna*, (2) identification of *C. avellana* cv. Tombul and *C. colurna* specific SSRs, and (3) exploration and comparison of *C. avellana* cv. Tombul and *C. colurna* with available genetic resources.

Material and Methods

Material

Plant materials were obtained from young hazelnut trees in the plantation of Hazelnut Research Institution located at Giresun, Turkey. Young leaves, flowers (male and female), buds, and husk shoots from three different trees of *C. avellana* and *C. colurna* were collected in fruit development stage, on April 30. For RNA extraction, these materials were promptly frozen in liquid nitrogen and stored at −80 °C.

RNA extraction and de novo transcriptome sequencing

Total RNA was isolated from young leaves, flowers (male and female), buds, husk shoots of *C. avellana* and *C. colurna* with an RNeasy Plant Mini Kit (Qiagen, Valencia, CA). The concentration and purity of total RNA was determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific) and the integrity number was calculated by Agilent 2100 Bioanalyzer® (Agilent Technologies, Santa Clara, CA). Construction of mRNA libraries was conducted with a TruSeq RNA Sample Preparation kit, according to manufacturer's instructions. RNA-seq libraries were sequenced with Illumina HiSeq4000 with 100 bp paired-end reads at Beijing Genomics Institute (Shenzhen, China) following recommended protocols.

Transcriptome assembly, functional annotation and SSR identification

The quality control of the clean reads was done using the FastQC tool (Andrews et al., 2012). *De novo* assembly of *C. avellana* and *C. colurna* transcriptome was performed using the Trinity software (Haas et al., 2013). Using the CD-HIT-EST software, full-length transcriptomes were clustered to represent the lowest levels of redundancy (Fu et al., 2012). The CD-HIT-EST software combined contigs with greater than 80 % identity and greater than 80 % coverage to generate a single transcript. Functional annotation (Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG)), ORF prediction and searching similarity species for each of *C. avellana* and *C. colurna* unigenes was performed by the TRAPID platform (corrected P value < 0.05) (Van Bel et al., 2013). The identification of SSRs was done using the MISA software (Beier et al., 2017).

Quantitative real time PCR (qPCR)

Total RNA was extracted from young leaves, flowers (male and female), buds, husk shoots of *C. avellana* and *C. colurna* using an RNeasy Plant Mini Kit (Qiagen). After RNA extraction, RNA samples were pooled and cDNA libraries were constructed using a First Stranded cDNA Synthesis Kit (GeneMark, Taichung, Taiwan). The primers were designed for the six selected unigenes (*TL1*, *GMPM1*, *N*, *2MMP*, *At1g29670*, *CHIB1*) related to the development, defence response, and lipid metabolism. Primer3 was used for the primer design and the primers were controlled by the BLAST tool against nr database (NCBI). Selected primer sequences are given in Table 1. Quantitative real-time PCR (qPCR) was carried out by the Roche LightCycler® 480 Real-Time PCR System (Roche, Penzberg, Germany). The ΔΔCt comparative method (Livak and Schmittgen, 2001) was used for estimation of the change of gene expression between the two hazelnut species.

Results

Transcriptome sequencing using Illumina Hiseq4000

In the study, a total of 41,854,426 and 41,140,316 clean reads 2 × 100 bp in length were generated in *C. avellana* and *C. colurna* after Illumina Hiseq4000 sequencing, respectively. Data quality assessment was per-

Table 1. Selected primer sequences

Unigene	Forward/ Reverse	Sequence (5‘-3‘)
TL1	F	ACTCTTCGGCCTTGTCATGG
TL1	R	GGGTTTCTGGTCCGCAGTTA
GMPM1	F	CATATGCAGGGAGGAAAGAA
GMPM1	R	CTCGAGTTAAGTACCCCCAG
N	F	TTATGCGAAGGACCCACACC
N	R	CGACTCTGCTGACTGAACCA
2MMP	F	TACTACATATTCCCGCCCGT
2MMP	R	TTCCCTAGCATCTCTCCAGG
At1g29670	F	AGGTGGGCTGGACTCAATTC
At1g29670	R	AATGGCTCTGCATGTGTGGA
CHIB1	F	TTGAGAGCTTCCTTGCCACC
CHIB1	R	TACTGTGACCGTCCGTTTGG

formed through the FASTQC (version 0.11.9) tool. After quality control, we continued clean reads of *C. avellana* and *C. colurna* for further study.

Assembly of transcriptome

De novo transcriptome assembly was performed using high-quality paired reads with Trinity (version 2.9.1), and 89,019 and 110,667 transcripts were identified in *C. avellana* and *C. colurna*, respectively (Table 1).

Then, these assembled transcripts were clustered according to their similarity and pooled together using CD-HIT-EST (version 4.8.1) (Fu et al., 2012).

Thanks to clustering, the count of transcripts was decreased to about 70,265 transcripts in *C. avellana* and 88,343 transcripts in *C. colurna*, with an average length of 1,045.2 bp in *C. avellana* and 915.5 bp in *C. colurna* (Table 2), respectively.

Meanwhile, the clustered transcripts of *C. avellana* and *C. colurna*, 70,202 and 88,228 transcripts with open reading frame (ORF), were in agreement, respectively (Table 2). Furthermore, the top 10 species similarity was shown in Fig. 1. According to these results, the best similar organism is *Prunus persica* for *C. avellana* and *C. colurna* (Fig. 1).

The percentage of high-quality mapped reads was found to be 78.9 % and 79.8 % for *C. avellana* and *C. colurna*, separately. The GC rate was determined to be 41.38 %, 14,515 (20.65 %) contigs were determined to be in the range of 200–499 bp, and 15,620 (22.23 %) contigs in the range of 500–1999 bp in *C. avellana*. However, the GC rate was determined to be 41.43 %, 22,330 (25.27 %) contigs were determined to be in the range of 200–499 bp, and 16,665 (18.86 %) contigs in the range of 500–1999 bp in *C. colurna* (Fig. 2A).

Transcriptome annotation

To further explore the molecular and biological functionality of *C. avellana* and *C. colurna* unigenes, we used a GO analysis categorized into biological processes, molecular functions, and cellular component classes, and KEGG pathway analysis. Nearly 70,265 and 88,343 unigenes from *C. avellana* and *C. colurna*, respectively, were assigned to GO terms and KEGG pathways. GO and KEGG enrichment analysis of *C. avellana* and *C. colurna* unigenes was performed through Fisher’s exact test at the P value 0.05 in the TRAPID software.

The “cellular process”, “metabolic process”, “organic substance metabolic process”, “cellular metabolic process”, “single-organism process”, “primary metabolic process”, “nitrogen compound metabolic process”, “macromolecule metabolic process”, “response to stimulus”, and “cellular macromolecule metabolic process” were the most highly enriched in the biological process GO terms for *C. avellana* and *C. colurna* unigenes, respectively (Fig. 3A). The highly enriched cellular com-

Table 2. Summary of de novo transcriptome assembly and clustering of C. avellana and C. colurna

	C. avellana	***C. colurna***
Number of Transcripts	89,019	110,667
Transcriptome Length (bp)	98,784,610	111,791,404
Average Transcript Length (bp)	1,109.70	1,010.16
N50	1,742	1,714
Clustered Transcripts	70,265	88,343
Clustered Transcriptome Length (bp)	73,439,646	80,875,074
Minimum Transcript Length (bp)	201	201
Maximum Transcript Length (bp)	11,562	25,356
Average Transcript Length (bp)	1,045.18	915.47
N50	1,703	1,627
Transcripts with ORF	70,217	88,251
Average ORF Length	239.4	216.7

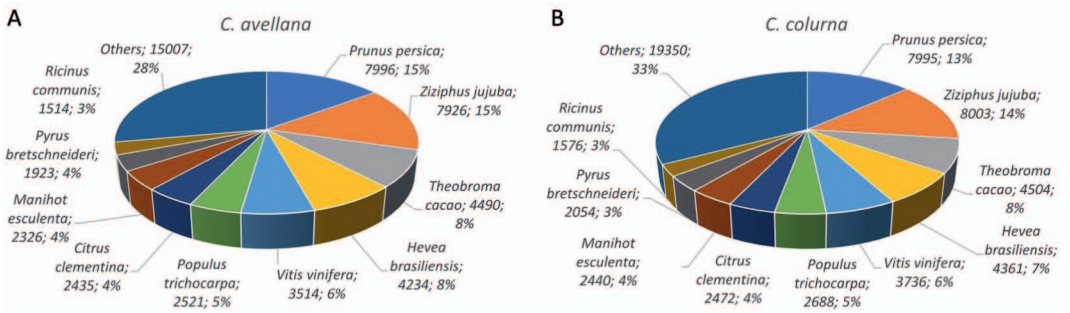

Fig. 1. BLASTX search analysis against the nr database. (**A**) *C. avellana* (**B**) *C. colurna.*

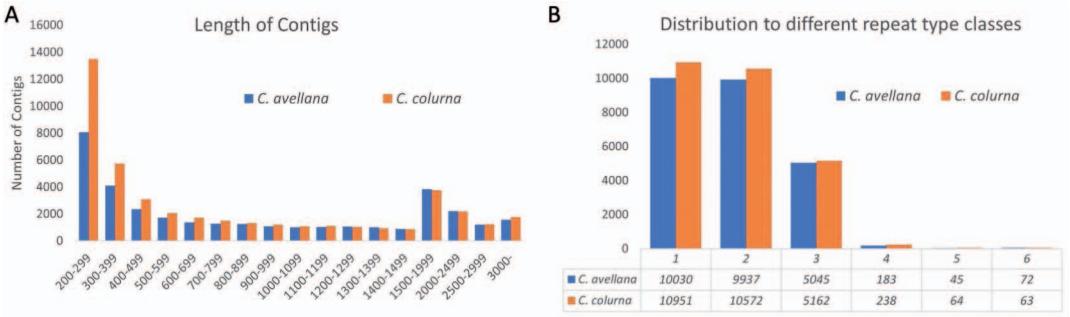

Fig. 2. (**A**) Length distribution of contigs of *C. avellana* and *C. colurna* transcriptomes. (**B**) The number of SSRs identified per transcript.

ponent GO terms were "cell", "cell part", "intracellular", "intracellular part", "organelle", "intracellular organelle", "membrane-bounded organelle", "intracellular membrane-bounded organelle", "cytoplasm", and "cytoplasmic part" for *C. avellana* and *C. colurna* (Fig. 3B). Also, the mostly enriched molecular function GO terms in *C. avellana* and *C. colurna* were "binding", "catalytic activity", "protein binding", "organic cyclic compound binding", "heterocyclic compound binding", "ion binding", "small molecule binding", "anion binding", "nucleoside phosphate binding", and "nucleotide binding" (Fig. 3C).

The KEGG pathway analysis revealed that transcription factor MYB plant, peptide/histidine transporter, interleukin 1 receptor-associated kinase 1, heat-shock 70 kDa protein, glutathione S-transferase, disease resistance protein RPS2, multidrug resistance protein, MATE family, disease resistance protein RPM1, peroxidase and senataxin-related pathway were mostly enriched in *C. avellana* unigenes. The unigenes in *C. colurna* were also significantly involved in transcription factor MYB plant, peptide/histidine transporter, interleukin 1 receptor-associated kinase 1, heat-shock 70 kDa protein, glutathione S-transferase, disease resistance protein RPS2, multidrug resistance protein, disease resistance protein RPM1, pectin esterase, and leucine-rich repeat protein in the KEGG pathways (Fig. 4).

Identification of simple sequence repeats (SSRs)

We identified SSRs from the expressions of *C. avellana* and *C. colurna*. According to the MISA software, 25,312 SSRs were detected with at least one SSR in 70,265 (36 %) *C. avellana* transcripts. Similarly, in *C. colurna*, around 27,051 SSRs were identified in 88,343 transcripts (30 %).

Most common SSR types were di-, tri, tetra- and pentanucleotide repeats. These were respectively 39.62 % in *C. avellana* and 40.48 % in *C. colurna*, 39.25 % in *C. avellana* and 39.08 % in *C. colurna*, 19.93 % in *C. avellana* and 19.08 % in *C. colurna*, and 0.72 % in *C. avellana* and 0.88 % in *C. colurna* (Fig. 2B).

Validation by qPCR

In order to validate the accuracy and reproducibility of the RNA-seq, qPCR was performed with development, defence response, and lipid metabolism-related genes, including *TL1*, *GMPM1*, *N*, *2MMP*, *At1g29670*, *CHIB1*. As shown in Fig. 4, TL1, GMPM1, N, 2MMP, At1g29670, CHIB1 in *C. avellana* and *C. colurna* transcriptomes were up-regulated in qPCR analysis (Fig. 5).

Discussion

Corylus species plants, commonly referred to as hazelnut, are very important economically and nutrition-

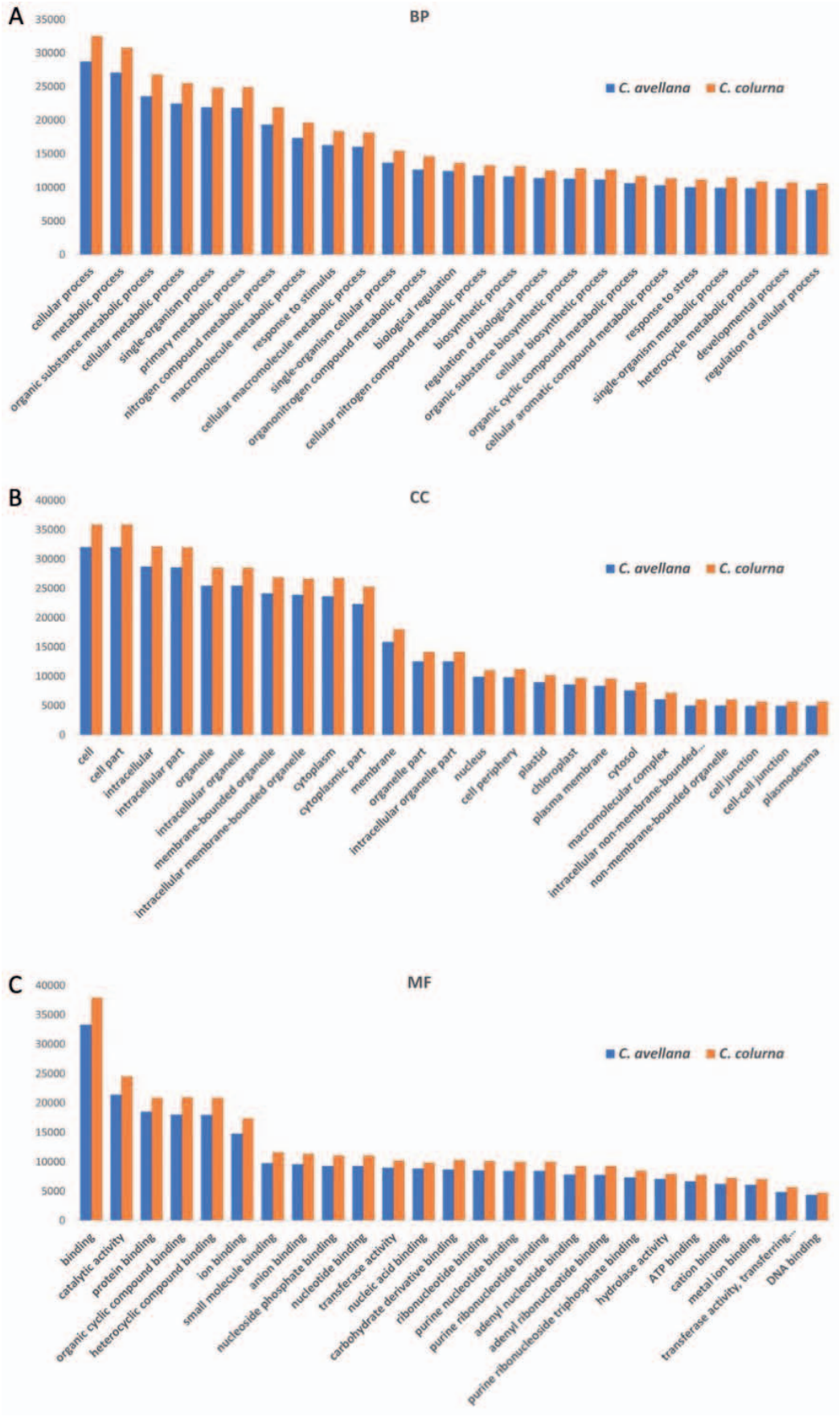

Fig. 3. The top 25 Gene Ontology categories in *C. avellana* and *C. colurna* unigenes. (**A**) Biological Process (BP), (**B**) Cellular Component (CC), (**C**) Molecular Function (MF).

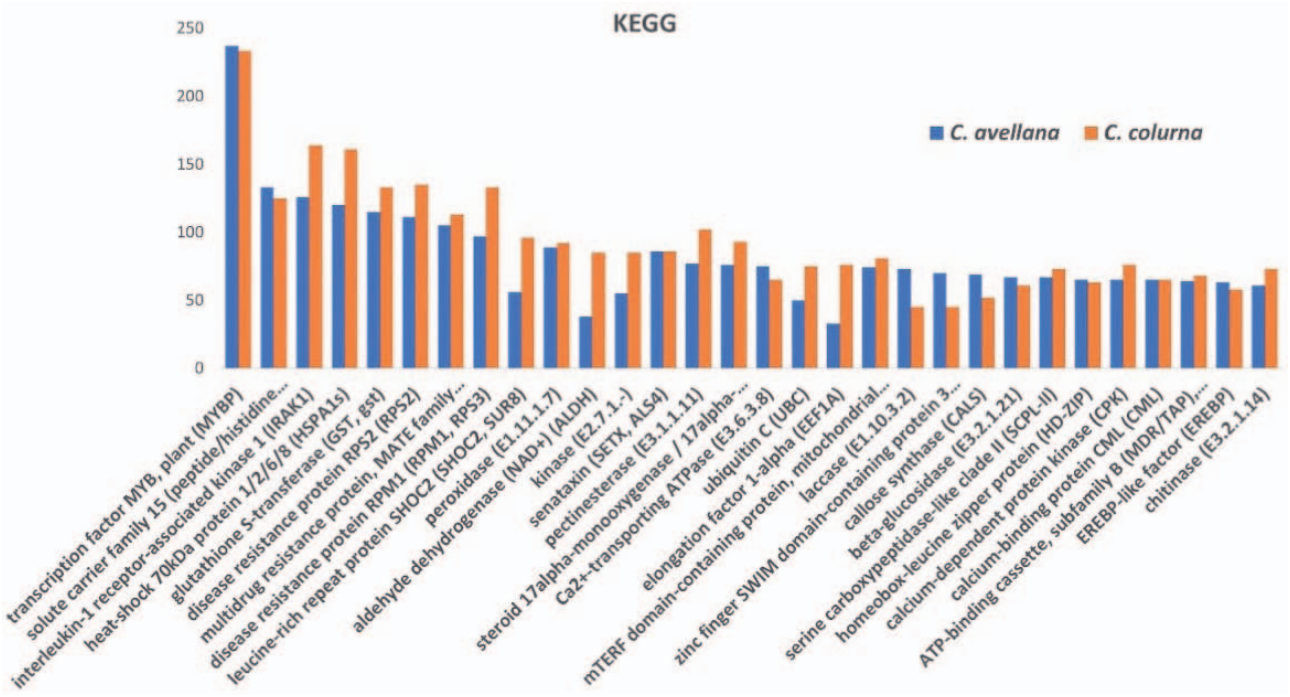

Fig. 4. The 30 most represented KEGG pathway terms in *C. avellana* and *C. colurna* unigenes

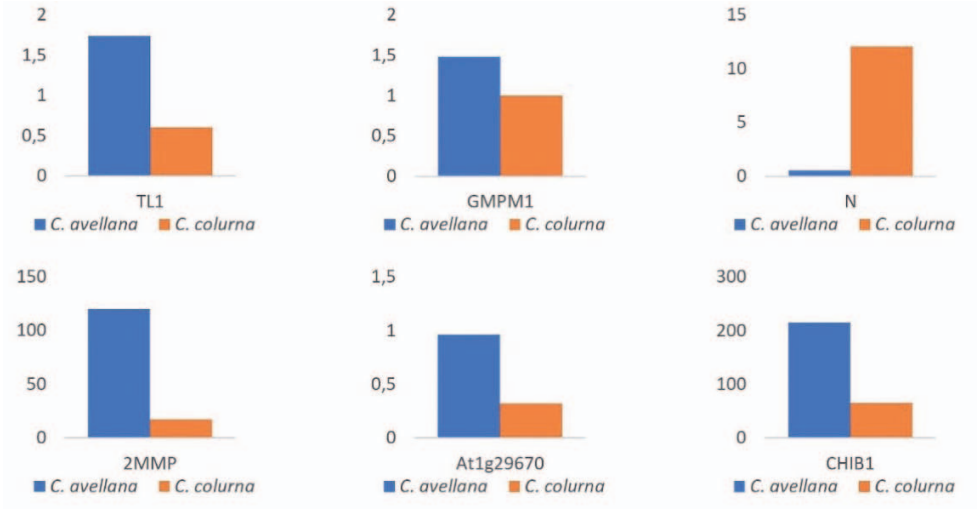

Fig. 5. qPCR expression profiles of selected genes for *C. avellana* and *C. colurna* (the x-axis shows gene names, the y-axis shows relative expression).

ally around the world. *C. avellana* L. (European hazelnut) is the most common and economically significant species and is widely distributed in the Black Sea region of Turkey (Rowley et al., 2018). *C. avellana* L. has many varieties such as "Tombul", "Çakıldak", "Jefferson", and "Palaz" (Rowley et al., 2012; Kavas et al., 2019). *C. avellana* cv. Tombul is considered to be very superior in quality (İslam, 2020). Also, *C. colurna* L. (Turkish hazelnut) is grown naturally in Turkey and is one of the significant hazelnut species (Ayan et al., 2016). *De novo* transcriptome analysis is carried out to elucidate the molecular mechanisms in plants (Jo et al., 2015; Ban et al., 2019; McGarvey et al., 2020). In the genome database, a few genomic resources are available for *C. avellana* cv. Jefferson, Çakıldak, and Palaz (Rowley et al., 2012; Kavas et al., 2019). However, the *C. avellana* cv. Tombul is limited to the chloroplast genome and SSR resources (Boccacci et al., 2006; Kahraman and Lucas, 2019; Tanhuanpää et al., 2019). The *C. colurna* genome study has not yet been published.

In this study, we used *de novo* transcriptome assemblies of *C. avellana* cv. Tombul and *C. colurna* from pooled young leaves, flowers (male and female), buds, and husk shoots. We generated comprehensive RNA-seq data sets from which we assembled 70,265 and 88,343 unigenes of *C. avellana* cv. Tombul and *C. colurna*, respectively, using a combination of *de novo* transcriptome assembly and redundancy reduction. These unigenes were functionally annotated using the TRAPID platform. The MISA software was used for identification of SSRs of *C. avellana* cv. Tombul and *C. colurna* contigs.

According to the *C. avellana* transcriptome analysis results, 89,019 transcripts were identified from 41,854,426 high quality clean reads with the Trinity software. After transcript assembly, we used the CD-HIT-EST program for clustering the assembled transcripts, and the number of transcripts was reduced to 70,265. In addition, in the *C. colurna* transcriptome analysis, we obtained 88,343 transcripts after clustering the assembled transcripts for redundancy reduction with the CD-HIT-EST program. We performed GO and KEGG pathway analysis of the 70,265 unigenes in *C. avellana* and 88,343 unigenes in *C. colurna* to determine the biological function of the unigenes. GO analysis was described into biological processes, cellular components, and molecular functions classes. According to the GO results, *C. avellana* and *C. colurna* unigenes were mostly enriched in the same GO terms in BP, CC, and MF (Fig. 3). Moreover, transcription factor MYB plant, peptide/histidine transporter, interleukin 1 receptor-associated kinase 1, heat-shock 70 kDa protein, glutathione S-transferase, disease resistance protein RPS2, multidrug resistance protein, MATE family, disease resistance protein RPM1, peroxidase and senataxin KEGG pathway terms were mostly enriched in *C. avellana* unigenes. The unigenes in *C. colurna* were also significantly involved in transcription factor MYB plant, peptide/histidine transporter, interleukin 1 receptor-associated kinase 1, heat-shock 70 kDa protein, glutathione S-transferase, disease resistance protein RPS2, multidrug resistance protein, disease resistance protein RPM1, pectin esterase, and leucine-rich repeat protein in KEGG pathways (Fig. 4).

In the present study, we also identified a total of 25,312 and 27,051 SSR motifs from transcriptome data of *C. avellana* and *C. colurna*, respectively. These results are the first research of the SSR motifs from the RNA-seq data of *C. colurna* with NGS-based sequencing technology. We characterized SSRs in the entire sequencing data of *C. avellana* and *C. colurna* and analysed their frequency and distribution in different scaffold/contig regions (Fig. 5). Most of the SSR motifs in *C. avellana* and *C. colurna* were dinucleotides and trinucleotides, approximately 80 % SSRs.

To experimentally confirm the transcriptome results, a subset of six unigenes related to the development, defence response, and lipid metabolism were chosen for qPCR analysis. According to the results, the expression of TL1 (Thaumatin-like protein 1), GMPM1 (18 kDa seed maturation protein), N (resistance protein N), 2MMP (metalloproteinase 2-MMP), At1g29670 (GDSL esterase/lipase), and CHIB1 (acidic endochitinase) was up-regulated in *C. avellana* and *C. colurna*. (Fig. 5). Furthermore, *TL1*, *GMPM1*, *2MMP*, *At1g29670*, and *CHIB1* unigene expression levels exhibited the highest increase in the *C. avellana* cv. Tombul variety compared to *C. colurna*. On the other hand, *N* unigene expressions was shown to be down-regulated in the *C. avellana* cv. Tombul variety as compared to that in the *C. colurna*. The qPCR results confirmed the data obtained from the transcriptome results.

Conclusion

This study of *C. avellana* cv. Tombul and *C. colurna* provided transcriptome, annotation, and SSR data, which is valuable for gene discovery and gene expression profiling experiments, as well as for continuing and upcoming genome annotation and marker development applications.

Data availability

Two sequencing data sets have been deposited in the NCBI Sequence Read Archive (SRA) under the accession numbers SRR19139694 for *C. avellana* and SRR19139693 for *C. colurna*.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval

Corylus colurna and *Corylus avellana* are not an endangered or protected species in the sampling area. Sampling activities were not performed at locations where specific permission is required. This article does not contain any other studies with human participants or animals performed by any authors.

References

Andrews, S., Krueger, F., Segonds-Pichon, A. et al. (2012) *FastQC*. Babraham Institute, Babraham, UK.

Ayan, S., Aydınözü, D., Yer, E. N. et al. (2016) Turkish Filbert (*Corylus colurna* L.): a new distribution area in Northwestern Anatolia Forests (Provinces of Müsellimler, Tunuslar in Ağlı, Kastamonu/Turkey). *Biol. Divers. Conserv.* **9**, 128-135. (in Turkish)

Ayan, S., Ünalan, E., İslam, A. et al. (2018) Kastamonu yöresinde yetişen Türk fındığının (*Corylus colurna* L.) yağ ve protein içeriği. *Artvin Çoruh Üniversitesi Orman Fakültesi Dergisi* **19**, 1-7. (in Turkish)

Ban, Y.-W., Roy, N. S., Yang, H. et al. (2019) Comparative transcriptome analysis reveals higher expression of stress and defense responsive genes in dwarf soybeans obtained from the crossing of *G. max* and *G. soja. Genes Genomics* **41**, 1315-1327.

Beier, S., Thiel, T., Münch, T. et al. (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583-2585.

Benov, L., Georgiev, N. (1994) The antioxidant activity of flavonoids isolated from *Corylus colurna*. *Phythother. Res.* **8**, 92-94.

Boccacci, P., Akkak, A., Botta, R. (2006) DNA typing and genetic relations among European hazelnut (*Corylus avellana* L.) cultivars using microsatellite markers. *Genome* **49**, 598-611.

FAO (2020) *Global Forest Resources Assessment 2020 Main Report*. Available at: https://www.fao.org/3/ca9825en/ca9825en.pdf

Fu, L., Niu, B., Zhu, Z. et al. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152.

Haas, B. J., Papanicolaou, A., Yassour, M. et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494-1512.

Han, Y., Gao, S., Muegge, K. et al. (2015) Advanced applications of RNA sequencing and challenges. *Bioinform. Biol. Insights* **9 (Suppl. 1)**, BBI.S28991.

İslam, A. (2020) Fındık ıslahında gelişmeler. *Akademik Ziraat Dergisi*. **8**, 167-174. (in Turkish)

Jo, Y., Lian, S., Cho, J. K. et al. (2015) De novo transcriptome assembly of two different Prunus salicina cultivars. *Genom. Data* **6**, 262-263.

Kahraman, K., Lucas, S. J. (2019) Comparison of different annotation tools for characterization of the complete chloroplast genome of *Corylus avellana* cv Tombul. *BMC Genomics* **20**, 874.

Kavas, M., Kurt Kızıldoğan, A., Balık, H. İ. (2019) Gene expression analysis of bud burst process in European hazelnut (*Corylus avellana* L.) using RNA-Seq. *Physiol. Mol. Biol. Plants* **25**, 13-29.

Li, Q., Zhao, T., Liang, L. et al. (2020) Molecular cloning and expression analysis of hybrid hazelnut (*Corylus heterophylla* × *Corylus avellana*) ChaSRK1/2 genes and their homologs from other cultivars and species. *Gene* **756**, 144917.

Livak, K. J., Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402-408.

Lucas, S. J., Kahraman, K., Avşar, B. et al. (2021) A chromosome-scale genome assembly of European hazel (*Corylus avellana* L.) reveals targets for crop improvement. *Plant J.* **105**, 1413-1430.

McGarvey, P., Huang, J., McCoy, M. et al. (2020) De novo assembly and annotation of transcriptomes from two cultivars of *Cannabis sativa* with different cannabinoid profiles. *Gene* **762**, 145026.

Peona, V., Weissensteiner, M. H., Suh, A. (2018) How complete are "complete" genome assemblies? An avian perspective. *Mol. Ecol. Resour.* **18**, 1188-1195.

Raghavan, V., Kraft, L., Mesny, F. et al. (2022) A simple guide to *de novo* transcriptome assembly and annotation. *Brief. Bioinform.* **23**, 1-30.

Riethmüller, E., Tóth, G., Alberti, A. et al. (2014) Antioxidant activity and phenolic composition of *Corylus colurna*. *Nat. Prod. Commun.* **9**, 679-682.

Rowley, E. R., Fox, S. E., Bryant, D. W. et al. (2012) Assembly and characterization of the European hazelnut "Jefferson" transcriptome. *Crop Sci.* **52**, 2679-2686.

Rowley, E. R., VanBuren, R., Bryant, D. W. et al. (2018) A draft genome and high-density genetic map of European hazelnut (*Corylus avellana* L.). *bioRxiv*. DOI:10.1101/469015

Sullivan, G. T., Ozman-Sullivan, S. K., Akbasli, O. et al. (2014) A tribute to the hazelnut plant (*Corylus* spp.) – the multiple uses of nature's magnificent gifts. *Acta Hortic.* **1052**, 371-376.

Tanhuanpää, P., Heinonen, M., Bitz, L. et al. (2019) Genetic diversity and structure in the northern populations of European hazelnut (*Corylus avellana* L.). *Genome* **62**, 537-548.

Thakur, O., Randhawa, G. S. (2018) Identification and characterization of SSR, SNP and InDel molecular markers from RNA-Seq data of guar (*Cyamopsis tetragonoloba*, L. Taub.) roots. *BMC Genomics* **19**, 951.

Thole, V., Bassard, J.-E., Ramírez-González, R. et al. (2019) RNA-seq, de novo transcriptome assembly and flavonoid gene analysis in 13 wild and cultivated berry fruit species with high content of phenolics. *BMC Genomics* **20**, 995.

Van Bel, M., Proost, S., Van Neste, C. et al. (2013) TRAPID: an efficient online tool for the functional and comparative analysis of de novo RNA-Seq transcriptomes. *Genome Biol.* **14**, R134.