biology and life sciences forum

MDPI

Proceeding Paper

Quercus suber Transcriptome Analyses: Identification of Genes and SNPs Related to Cork Quality †

Bruna Mendes [1], Ana Usié [1,2,*], Tiago Capote [1,3], Brígida Meireles [1,4], Tânia Almeida [1,5], Liliana Marum [1,2,*], Sónia Gonçaves [1,6] and António Ramos [1,2]

1 CEBAL-Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo/IPBeja, Instituto Politécnico de Beja, 7801-908 Beja, Portugal; bruna.mendes@cebal.pt (B.M.); tcc414@nyu.edu (T.C.); bmeireles@medicina.ulisboa.pt (B.M.); taniaralmeida@ua.pt (T.A.); sonia.goncalves@sanger.ac.uk (S.G.); marcos.ramos@cebal.pt (A.R.)
2 MED-Mediterranean Institute for Agriculture, Environment and Development, 7801-908 Beja, Portugal
3 Center for Genomics and Systems Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi P.O. Box 129188, United Arab Emirates
4 GenoMed-Diagnóstico de Medicina Molecular SA. Edifício Egas Moniz, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal
5 Department of Chemistry, CICECO–Aveiro Institute of Materials, University of Aveiro, 3810-193 Aveiro, Portugal
6 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK
* Correspondence: ana.usie@cebal.pt (A.U.); liliana.marum@cebal.pt (L.M.); Tel.: +351-284-314-399 (A.U.)
† Presented at the 2nd International Electronic Conference on Plant Sciences—10th Anniversary of Journal Plants, 1–15 December 2021; Available online: https://iecps2021.sciforum.net/.

Abstract: Cork is an ancestral natural material derived from the cork oak tree (*Quercus suber* L.) with multiple industrial applications. During the recent years, this material has been the subject of several studies. The recent sequencing of the *Q. suber* genome opened the possibility to make new studies regarding cork quality. In this study, the transcriptomes of cork with superior and poor quality are compared to highlight new molecular pathways and identify SNPs that can be associated to cork traits, which remain one of the main concerns of the cork industry.

Keywords: cork oak; cork quality; transcriptome; SNPs

Citation: Mendes, B.; Usié, A.; Capote, T.; Meireles, B.; Almeida, T.; Marum, L.; Gonçaves, S.; Ramos, A. *Quercus suber* Transcriptome Analyses: Identification of Genes and SNPs Related to Cork Quality. *Biol. Life Sci. Forum* **2022**, *11*, 76. https://doi.org/10.3390/IECPS2021-11916

Academic Editor: Iker Aranjuelo Michelena

Published: 20 April 2022

1. Introduction

Cork oak (*Quercus suber* L.) is an evergreen broad-leaved tree that belongs to the genus *Quercus* (oaks) of the *Fagaceae* family and is one of the most important Mediterranean forest tree species. It plays an important environmental, social, and economic role in the Mediterranean ecosystems known as "Montado" in Portugal and "Dehesa" in Spain. Cork is obtained from the extraction of the outside layer of the cork oak tree, which is composed by suberized cells. The first cork extraction occurs when the tree is between 20–25 years old. After that, cork is extracted at regular intervals of at least nine years [1]. The chemical cork composition has already been extensively described, and it is known that it can change depending on environmental and genetic conditions [2,3]. Parameters such as thickness and structural discontinuities have been used to evaluate the cork quality [4]. Based on quality, the cork's economic value can vary, which is one of the main concerns in the cork industry. Scientific studies have been performed to identify conditions and processes that can help to stabilize the cork quality and, therefore, ensure more profitable and sustainable production [5,6].

Recent advances in sequencing technology, and subsequent sequencing of the cork oak genome, open the opportunity to perform new studies regarding the cork formation process [7]. The aim of this study was to analyze the dynamic profile of differential expressed genes associated with cork quality, highlighting for the first time, to our knowledge, a set of SNPs that could be involved in cork differentiation.

2. Materials and Methods

2.1. Plant Material and RNA Isolation

Amadia cork planks were collected from physiologically active cork oak during the period July-August. Cork samples from eight genotypes were classified into Good and Bad quality cork (GQ and BQ, respectively) according to the main traits of thickness and structural discontinuities (porosity and inclusion of woody cells). GQ and BQ cork samples were harvested in the Barranco Velho and Cercal regions of Portugal, respectively. The inner part of harvested cork planks, corresponding to phellogen and phellem cells, was scraped and stored according to Soler et al. 2007 for RNA extractions (Figure 1) [3]. Total RNA from GQ and BQ samples (four cork oak trees per condition) was isolated according to Almeida et al. (2013) and sequenced by Illumina HiSeq 2000, producing PE reads of 100 bp in length [8].

Figure 1. Collection of sample material: the inner part from cork planks was harvested, corresponding to phellogen and phellem cells. Photographer: Leandra Rodrigues (www.liarodriguesphotography.com, accessed on 28 November 2021).

2.2. Reads Pre-Processing and Mapping

The raw reads were preprocessed keeping only reads with a minimum quality of 20 and minimum length of 80 bp using Trimmomatic (v0.38) [9]. Then, the pre-processed reads were mapped against the *Q. suber* genome with STAR (version 2.5.2 b) using the multi-sample 2-pass mapping mode, according to its user guidelines [10]. The unique mapped reads (UMR) were then filtered and extracted using SAMtools [11].

2.3. Differential Expression Analysis

The differential expression analyses were performed using edgeR, a Bioconductor package [12]. To avoid issues with the incompatibility of some tools such as edgeR in using biological and technical replicates at the same time for differential expression analyses, technical replicates were merged using the function "sumTechReps" in edgeR. Then, genes with low counts were excluded and a Trimmed Mean of M-values (TMM) normalization was applied. In the end, only genes with a log fold change (logFC) $\geq |2|$ and with a False Discovery Rate (FDR) ≤ 0.05 were considered as differentially expressed genes (DEGs). At last the interactions between DEGs associated with KEGG pathways and Gene Onthology (GO) terms were visualized and analyzed with Cytoscape. Additionally, BinGO (Cytoscape plugin) was used to identify GOs overrepresented over the set of genes up- or down-regulated against the GO database [13,14].

2.4. Variant Calling and Annotation

For Single Nucleotide Polymorphism (SNP) identification, a variant calling analyses was performed using mpileup from SAMtools. The raw variants were then filtered by SNP quality (≥30) and minimum depth coverage per genotype (≥8); indels were removed and only bi-allelic SNPs were maintained, resulting in a high-quality SNP set which was annotated using ANNOVAR [15].

3. Results and Discussion

The sequencing of RNA from good and bad quality cork samples resulted in a total of 708,803,510 reads. After pre-processing, 98% of the reads were kept. Pre-processed reads were then mapped against the cork oak genome [7] and the UMR obtained, representing 78% of the mapped reads, and used for the differential gene expression analyses.

3.1. Differential Expression Analysis

A total of 172 genes were differentially expressed, of which 66 were more expressed in GQ, and 106 in BQ. Within the set of the identified DEGs, only 73 were associated with at least one GO term. For each sub-ontology (biological process (BP), molecular function (MF) and cellular components (CC)), a total of 50, 20 and 18 terms were associated with the DEGs (Table 1). After performing the over-represented analyses with BinGO, a total of 23 GO terms were over-represented in the set of genes more expressed in GQ (BP:19; CC:1; MF:3), while only one GO term (BP: response to stress process) was over-represented in the set of genes more expressed in BQ.

Table 1. Number of GO terms, for category, that were associated only with genes more expressed in BQ (gBQ) and genes in GQ (gGQ), and in common.

	gBQ	gGQ	Common	Total
Biological Process	24	34	7	50
Cellular Component	8	17	5	20
Molecular Function	13	11	6	18

Several proteins such as thiamine thiazole, late embryogenesis abundant protein lea5, and dehydrin erd10 were associated with the response to stress process in BQ (Table 2). Additionally, heat shock proteins were also found differentially expressed: HSP17.5-E, HSP17.6C, HSP26.5 and HSP22.7 more expressed in BQ and HSP70-15 more expressed in GQ. This is not the first time that proteins from the heat shock group were identified in phellem of cork oak trees [5]. Regarding the heat shock proteins more expressed in BQ, the only information available so far is that its expression confers resistance to heat stress. [16,17]. Contrarily, HSP70-15, which was more expressed in GQ, belongs to the 70-kDA heat shock proteins group, a well-known group frequently found highly expressed in tissues under stress [18]. This group of proteins is actively involved in the folding of de novo synthesized proteins, translocation of precursor proteins into organelles and degradation of damaged proteins under stress conditions [18].

During the differential gene expression analysis, genes uniquely expressed in both conditions were identified. For instance, SRG1, which is involved in oxidation-reduction processes, and PIP2-2, an important aquaporin involved in the transport of water and other small solutes across the cell membrane [19], were the only genes exclusively expressed in BQ. On the other hand, 14 genes were found uniquely expressed in GQ. It is important to highlight the presence of genes such as EMF2 and FYPP-3 that are involved in regulation of the flowering process [20,21], and KUA1, which is a transcript factor from the MYB-like protein family that acts as a repressor and promotes response to auxin, ethylene, and abscisic acid [22]. These plant growth regulators (PGRs) are involved in the regulation of plant growth and development, and abscisic acid is also associated with the increase of resistance of plants to different stresses. Likewise, EMF2 and FYPP-3 are also associated with the regulation of abscisic acid [20,23]. These results could indicate a correlation

between hormonal regulation and cork quality; however more studies are necessary to assess its influence.

Table 2. List of genes referred to in the results and discussion section. Genes with negative values of logFC are more expressed in GQ while genes with positives values are more expressed in BQ.

Annotation	LogFC	Exclusive Expression
thiamine thiazole synthase, chloroplastic (THI 1)	5.53	NO
late embryogenesis abundant protein (LEA5)	2.46	NO
dehydrin (ERD10)	2.56	NO
17.5 kda class I heat shock protein (HSP17.5-E)	3.73	NO
17.6 kda class I heat shock protein 3 (HSP17.6C)	2.27	NO
26.5 kda heat shock protein, mitochondrial (HSP26.5)	2.18	NO
22.7 kda class iv heat shock protein (HSP22.7)	2.57	NO
heat shock 70 kda protein 15 (HSP70-15)	−7.18	NO
Protein srg1 (SRG1)	8.21	YES
Aquaporin PIP2.2 (PIP2-2)	8.32	YES
Polycomb group protein embryonic flower (2EMF2)	−8.63	YES
Phytochrome-associated serine/threonine-protein phosphatase 3 (FYPP-3)	−9.57	YES
Transcription factor KUA1	−9.56	YES
Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (AccD)	−4.31 and −4.52	NO

Two genes, both annotated as acetyl-coenzyme A carboxylase carboxyl transferase (beta subunit) (AccD), were found more expressed in GQ. The AccD protein is responsible to produce malonyl-CoA from acetyl-CoA, a starting unit of the fatty acid biosynthetic process. The fatty acid pathway is responsible for very long and long-chain fatty acids synthesis, which are important precursors of waxes and some suberin monomers, two important compounds of cork. The expression of genes involved in fatty acids biosynthesis in phellogen cork tissue from GQ have already been reported [5], which reinforce the hypothesis that genes involved in the regulation and activity of cell wall assembly can affect cork quality [24].

3.2. SNPs analysis

The variant calling resulted in the identification of 1,296,640 raw variants, of which 159,248 were considered high-quality SNPs (Table 3). The high-quality SNPs were further evaluated to confirm if some of them were located in genes identified as differentially expressed. As a result, 8078 SNPs were identified in 148 DEGs from which only the exonic and non-synonymous SNPs (879) were analysed.

Table 3. Summary of the number of SNPs identified.

	Nr. of SNPs
SNPs	1,296,640
SNPs filtered Q30-DP7 (High-quality SNPs)	159,248
SNPs in DEGs	8078 (in 149 genes)
Exonic and Non-synonymous SNPs	879 (in 124 genes)
SNPs in DEGs in BQ	469 (in 67 genes)
SNPs in DEGs in GQ	410 (in 40 genes)

The identification of exclusive SNPs—an SNP is considered as exclusive if it is only present in at least 75% of the individuals from one group, GQ or BQ—was performed. Following this criterion for GQ, 121 exclusive SNPs were found in a total of 49 genes, while in BQ 68, exclusive SNPs were found among 44 genes. Regarding the SNPs found in DEGs, 18 were exclusive in 8 genes more expressed in GQ, while 5 SNPs were exclusive in 5 genes more expressed in BQ.

4. Conclusions

In this study, we identified a set of candidate genes for cork quality in *Quercus suber*. Some mechanisms associated with cork quality were revealed that allow us to hypothesis that the observed differences in cork quality could be directly related to specific PGRs increasing resistance to stress and involved in cell wall assembly. Additionally, several exclusive SNPs for individuals of contrasting phenotypes for cork quality were identified, although further studies will be needed to assess their phenotypic influence and potential usage as genetic markers for cork quality.

Author Contributions: This study was conceived by S.G. and A.R.; collection and identification of field material was performed by T.A., T.C. and S.G.; sample preparation and nucleic acid isolation were performed by T.A. and T.C.; bioinformatics data analyses were conducted by B.M. (Bruna Mendes), A.U., B.M. (Brígida Meireles) and A.R.; biological interpretation of the results was conducted by B.M. (Bruna Mendes), A.U., A.R. and L.M.; the manuscript was written by B.M. (Bruna Mendes), A.U. and L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by InAlentejo under the scope "GenoSuber–Cork oak genome sequencing" (ALENT-07-0224-FEDER-001754) and by Alentejo2020 through FEDER under the scope "Lentidev-A genomic approach to cork quality" (ALT20-03-0145-FEDER-000020). Contrato–Programa to L. Marum (CEECINST/00131/2018) and UIDB/05183/2020 were funded by FCT.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw read files are available at NCBI Sequence Read Archive under the accession PRJNA824660.

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

References

1. Silva, V.M.C.; Sabino, M.A.; Fernandes, E.; Correlo, V.M.; Boesel, L.F.; Reis, R.L. Cork: Properties, capabilities and applications. *Int. Mater. Rev.* **2005**, *50*, 345–365. [CrossRef]
2. Pereira, H. Variability of the chemical composition of cork. *BioResources* **2013**, *8*, 2246–2256. [CrossRef]
3. Soler, M.; Serra, O.; Molinas, M.; Huguet, G.; Fluch, S.; Figueras, M. A Genomic Approach to Suberin Biosynthesis and Cork Differentiation. *Plant Physiol.* **2007**, *144*, 419–431. [CrossRef] [PubMed]
4. Pereira, H.; Lopes, F.; Graça, J. The Evaluation of the Quality of Cork Planks by Image Analysis. *Holzforschung* **1996**, *50*, 111–115. [CrossRef]
5. Teixeira, R.T.; Fortes, A.M.; Pinheiro, C.; Pereira, H. Comparison of good- and bad-quality cork: Application of high-throughput sequencing of phellogenic tissue. *J. Exp. Bot.* **2014**, *65*, 4887–4905. [CrossRef] [PubMed]
6. Ricardo, C.P.; Martins, I.; Francisco, R.; Sergeant, K.; Pinheiro, C.; Campos, A.; Renaut, J.; Fevereiro, P. Proteins associated with cork formation in *Quercus suber* L. stem tissues. *J. Proteom.* **2011**, *74*, 1266–1278. [CrossRef] [PubMed]
7. Ramos, A.; Usié, A.; Barbosa, P.; Barros, P.M.; Capote, T.; Chaves, I.; Simões, F.; Abreu, I.; Carrasquinho, I.; Faro, C.; et al. The draft genome sequence of cork oak. *Sci. Data* **2018**, *5*, 180069. [CrossRef] [PubMed]
8. Almeida, T.; Pinto, G.; Correia, B.; Santos, C.; Gonçalves, S. QsMYB1 expression is modulated in response to heat and drought stresses and during plant recovery in *Quercus suber*. *Plant Physiol. Biochem.* **2013**, *73*, 274–281. [CrossRef]
9. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
10. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef] [PubMed]
11. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
12. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef] [PubMed]
13. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
14. Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **2005**, *21*, 3448–3449. [CrossRef] [PubMed]
15. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef] [PubMed]

16. Pla, M.; Huguet, G.; Verdaguer, D.; Puigderrajols, P.; Llompart, B.; Nadal, A.; Molinas, M. Stress proteins co-expressed in suberized and lignified cells and in apical meristems. *Plant Sci.* **1998**, *139*, 49–57. [CrossRef]
17. Puigderrajols, P.; Jofré, A.; Mir, G.; Pla, M.; Verdaguer, D.; Huguet, G.; Molinas, M. Developmentally and stress-induced small heat shock proteins in cork oak somatic embryos. *J. Exp. Bot.* **2002**, *53*, 1445–1452. [CrossRef]
18. Lin, B.L.; Wang, J.S.; Liu, H.C.; Chen, R.W.; Meyer, Y.; Barakat, A.; Delseny, M. Genomic analysis of the Hsp70 superfamily in Arabidopsis thaliana. *Cell Stress Chaperones* **2001**, *6*, 201–208. [CrossRef]
19. Javot, H.; Lauvergeat, V.; Santoni, V.; Martin-Laurent, F.; Güçlü, J.; Vinh, J.; Heyes, J.; Franck, K.I.; Schäffner, A.R.; Bouchez, D.; et al. Role of a Single Aquaporin Isoform in Root Water Uptake. *Plant Cell* **2003**, *15*, 509–522. [CrossRef]
20. Yoshida, N.; Yanai, Y.; Chen, L.; Kato, Y.; Hiratsuka, J.; Miwa, T.; Sung, Z.R.; Takahashi, S. EMBRYONIC FLOWER2, a Novel Polycomb Group Protein Homolog, Mediates Shoot Development and Flowering in Arabidopsis. *Plant Cell* **2001**, *13*, 2471–2481. [CrossRef] [PubMed]
21. Kim, D.-H.; Kang, J.-G.; Yang, S.-S.; Chung, K.-S.; Song, P.-S.; Park, C.-M. A Phytochrome-Associated Protein Phosphatase 2A Modulates Light Signals in Flowering Time Control in Arabidopsis. *Plant Cell* **2002**, *14*, 3043–3056. [CrossRef] [PubMed]
22. Huang, C.-K.; Lo, P.-C.; Huang, L.-F.; Wu, S.-J.; Yeh, C.-H.; Lu, C.-A. A single-repeat MYB transcription repressor, MYBH, participates in regulation of leaf senescence in Arabidopsis. *Plant Mol. Biol.* **2015**, *88*, 269–286. [CrossRef] [PubMed]
23. Dai, M.; Xue, Q.; Mccray, T.; Margavage, K.; Chen, F.; Lee, J.-H.; Nezames, C.D.; Guo, L.; Terzaghi, W.; Wan, J.; et al. The PP6 Phosphatase Regulates ABI5 Phosphorylation and Abscisic Acid Signaling in Arabidopsis. *Plant Cell* **2013**, *25*, 517–534. [CrossRef] [PubMed]
24. Pinheiro, C.; Wienkoop, S.; de Almeida, J.F.; Brunetti, C.; Zarrouk, O.; Planchon, S.; Gori, A.; Tattini, M.; Ricardo, C.P.; Renaut, J.; et al. Phellem Cell-Wall Components Are Discriminants of Cork Quality in *Quercus suber*. *Front. Plant Sci.* **2019**, *10*, 944. [CrossRef] [PubMed]