Hinc patriam sustinet
Instituto Superior de Agronomia
Universidade Técnica de Lisboa

An integrated BAC approach to the genomics of cork formation

Maria Joana Amado Teixeira Pinto

Dissertação para obtenção do Grau de Mestre em

Biologia Funcional

Orientador: Doutor Jorge A. Pinto Paiva
Co-orientador: Prof. Doutora Leonor Morais Cecílio

Júri:
Presidente: Doutora Sara Barros Queiroz Amâncio, Professora Associada com Agregação do Instituto Superior de Agronomia da Universidade Técnica de Lisboa.

Vogais: Doutor Jorge Almiro Barceló Caldeira Pinto Paiva, Investigador Auxiliar do Instituto de Investigação Cientifica Tropical, I.P.;
Doutora Célia Maria Rodrigues Miguel, Investigadora Auxiliar do Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa.

Lisboa, 2013

Acknowledgements

I would like to thank my supervisor Dr. Jorge Paiva and Professor Leonor Morais for their precious contribution for the design of the experiments, on the guidance during this work and for the helpful advices, corrections and suggestions that improved the manuscript. I acknowledge Professor Pedro Fevereiro for allowing to develop part of this work in the BCV laboratory and for the support and advices in critical moments. I am also very grateful to Victor Carocha for the kind and valuable help in solving so many difficulties and doubts during the work. I would like to express my gratitude to my colleagues of the BCV laboratory, in particular to Cátia Nunes for being present at the most challenging times; to Susana Araújo, Rita Morgado and Tomás for their precious help in the qPCR data analysis and to my colleagues Mara, Inês, Susana Leitão, Susana Neves, Nuno, Cláudio and Marco, for the companionship and valuable support in the laboratory work. In addition I am very thankful to my colleagues from the Genetic laboratory at ISA, in particular for the great help of Vera Inácio during the FISH experiments and for the assistance of Diana, Carlos and Ana in the laboratory. To Fundação João Lopes Fernandes for providing the plant material for this work, and to Eng. Pedro Marques for kindly providing images and photos for this manuscript. I also acknowledge Professor José Graça for the sampling and sample characterization. I am very grateful specially to my family and friends, in particular to my mother, to my sister Ana and my dear grandmother Maria, for the unconditional affection in the good and bad moments and without whom none of this would be possible.

The work was partially supported by a grant from the Fundação para a Ciência e Tecnologia (FCT), project SuberGene (PTDC/AGR-GPL/101785/2008, Genomics of Cork Formation: an integrated approach of cork quality). The author aknowledges FCT for the fellowship grant provided in the frame of the project SuberGene.

Abstract

Cork is produced by *Quercus suber*, an evergreen oak with major economic and environmental importance in Portugal. Suberin is the main component of cork but little is known about the molecular processes underlying suberin biosynthesis.

We screened a *Quercus suber* bacterial artificial chromosome (BAC) library for the presence of relevant genes for suberin biosynthesis through high-density BAC filters hybridizations and we validated hybridization positive clones by Colony PCR. We analyzed BAC-end sequences (BESs) to characterize the structure and composition of the cork oak genome. And we hybridized BAC DNA with mitotic cork oak chromosomes to obtain a cytogenetic map of *Q. suber*. Finally, we analyzed expression profiles of candidate genes of cork samples from cork oaks with different cork qualities, using quantitative real-time PCR.

We identified several genes involved in suberin biosynthesis pathways. BESs analysis showed high identity with coding regions and suggest the existence of conserved regions within the *Fagacea* family. We found strong evidence that the genes Cyp 86A1 and GDSL involved in lipid suberin biosynthesis pathway are closely located in the genome. However, we did not succeed in mapping genes of interest in the cork oak genome and we did not find any relation between the genes relative expression and cork quality.

Keywords: Cork; suberin; BAC library; BESs; FISH (fluorescente *in situ* hybridization); quantitative real-time PCR (qPCR).

Resumo

A cortiça é produzida pelo *Quercus suber*, um carvalho de grande importância económica e ambiental em Portugal. A suberina é a principal constituinte da cortiça, no entanto muito pouco é conhecido sobre os processos moleculares que estão na base da biosíntese da suberina.

A biblioteca BAC de *Quercus suber* foi examinada para encontrar a presença de genes relevantes para a biosíntese de suberina através de hibridizações com filtros BAC de alta densidade. Seguidamente, os clones BAC positivos foram validados usando PCR de colónias BAC. Analisaram-se *BAC-end sequences* (BESs) para caracterizar a estrutura e composição do genoma do sobreiro. Amostras de DNA isolado de clones BAC, foram usadas como sondas para hibridar cromossomas mitóticos de sobreiro. Tendo como obectivo o mapeamento de genes relevantes para a formação de cortiça no genoma do sobreiro e assim desenvolver o primeiro mapa citogenético desta espécie. Finalmente, analisou-se o perfil de expressão dos genes candidatos, usando o PCR quantitativo em tempo-real, em amostras de tecido de cortiça em desenvolvimento recolhidas de sobreiros com diferentes qualidades de cortiça.

Identificaram-se vários genes envolvidos nas vias de biosíntese de suberina. A análise das BESs mostrou elevada indentidade com regiões codificantes e sugere a existência de regiões conservadas dentro da família das *Fagacea.* Encontraram-se fortes indícios que os genes Cyp 86A1 e GDSL, envolvidos na via lipídica da biosíntese da suberina, estão proximamente localizados no genoma. No entanto, não foi possível mapear os genes de interesse no genoma do sobreiro e também não não foi encontrada nenhuma relação entre expressão relativa dos genes candidatos e a qualidade da cortiça.

Palavras-chave: Cortiça; suberina; biblioteca BAC; BESs; FISH (fluorescente *in situ* hybridization); PCR quantitativo em tempo real (qPCR).

Resumo alargado

Introdução

A cortiça é produzida pela espécie *Quercus suber,* um carvalho com grande importância económica e ambiental para Portugal pois é o maior produtor mundial deste material. A suberina é o principal componente das células da cortiça e é devido à sua presença que as células de cortiça são impermeáveis à água e a gases, tornando a cortiça um material com características únicas e múltiplas aplicações. A qualidade de cortiça determina a sua aplicação industrial e o seu valor económico. No entanto este fenótipo é muito diverso, verificando-se frequentemente que sobreiros crescendo lado a lado no mesmo ambiente, produzem cortiça de diferentes qualidades o que sugere que a qualidade de cortiça dependerá essencialmente do genótipo das mesmas árvores. Contudo, o conhecimento genómico desta espécie é ainda muito escasso e pouco se sabe acerca dos processos moleculares envolvidos na biossíntese e diferenciação da suberina e da cortiça. As bibliotecas BAC (*Bacterial Artificial Chromosome*) são ferramentas importantes que têm sido muito utilizadas no mapeamento físico, mapeamento comparativo e em estudos de evolução de genomas. Anteriormente a este trabalho, foi construída a primeira biblioteca BAC de *Q. suber* contendo insertos com tamanho médio de 120kb e uma cobertura do genoma de 5x.

O objectivo geral deste trabalho foi expandir o nosso conhecimento da genética da formação e da qualidade da cortiça utilizando clones BAC de sobreiro em diversas metodologias.

Material e Métodos

Para aprofundar o conhecimento sobre a biossíntese da suberina e os mecanismos genéticos envolvidos na qualidade de cortiça, caracterizámos genes candidatos presentes em clones BAC de sobreiro. Estes genes foram previamente descritos como estando envolvidos nas vias de biossíntese da suberina e como sendo preferencialmente expressos em tecidos de felema em desenvolvimento. Selecionámos 16 genes candidatos envolvidos em duas das principais vias de biossíntese da suberina, a via da formação dos ácidos gordos e a via dos fenilpropanoides, assim como genes de regulação e de resposta ao stress descritos como estando implicados na formação de cortiça.

A identificação de clones BACs contendo genes de interesse foi realizada através da hibridização de filtros de alta densidade de colónias com sondas de cDNA. Posteriormente, os clones BAC positivos foram validados usando PCR de colónias. Dos clones identificados, 24 foram escolhidos para sequenciação da extemidades dos insertos (*BAC-end sequences; BESs*).

Para mapear genes relevantes na formação de cortiça no genoma do sobreiro, foi isolado DNA de seis clones BAC validados. Este DNA foi usado como sonda para hibridar cromossomas mitóticos de sobreiro, através do método de *´BAC landing´*. Finalmente, para estudarmos a relação de alguns genes com a qualidade de cortiça, analisámos e caracterizámos o padrão de expressão de 15 genes candidatos, usando PCR quantitativo (qPCR). Para esse efeito, usámos um conjunto de amostras recolhidas de sobreiros que apresentavam diferentes qualidades de cortiça.

Resultados e Discussão

Através do PCR de colónias, 69% dos clones positivos foram validados para 14 dos 16 genes candidatos em estudo. Verificámos uma maior representação de clones contendo os genes Dolaccase e HCBT, da via dos fenilpropanoides, e do gene GDSL da via dos ácidos gordos da biossíntese da suberina. Foram também validados quatro clones nos quais se verificou a amplificação de dois genes em simultâneo (DHCA-S-AT e 4CL; PAL e ELTFP). Estes resultados sugerem que estes pares de genes estão proximamente localizados no genoma do sobreiro.

A análise das BESs revelou que os retrotransposões da Classe I, nomeadamente os elementos *Ty1/Copia* são os retrolementos mais abundantes. Esta análise permitiu também identificar uma presença significativa de microssatélites; uma elevada identidade com regiões codificantes; e sugere a existência de regiões conservadas dentro da família das *Fagacea.* Encontraram-se fortes indícios que os genes Cyp86A1 e GDSL, envolvidos na via da formação dos ácidos gordos da biosíntese de suberina, estão proximamente localizados no genoma. Estes resultados permitiram uma caracterização preliminar da organização e composição do genoma do sobreiro e podem ser utilizados futuramente para o desenvolvimento de marcadores para a construção e integração de mapas físicos e genéticos do genoma de *Q. suber*.

Quanto à análise citogenética, apesar de termos observado sinal de hibridação das sondas de DNA ribossomal não foi possível obter sinal das sondas BAC. Estes resultados indicam que será necessária uma optimização do método de *´BAC landing´* para espécies de árvores florestais, como o sobreiro.

Os resultados da análise do perfil de expressão dos genes candidatos revelaram uma correlação positiva significativa entre a expressão relativa dos genes que codificam enzimas envolvidas nas vias de biossíntese de suberina. Adicionalmente, encontrámos uma correlação positiva significativa entre o factor de transcrição WRKY e os 14 genes candidatos da via de biossíntese de suberina, indicando um possível papel deste gene na regulação desta via metabólica, de acordo com o descrito em estudos anteriores. Os nossos resultados mostraram também uma expressão relativa elevada de genes que codificam enzimas envolvidas no stress oxidativo, em concordância com trabalhos anteriores que demonstram que o metabolismo das células de cortiça produz espécies reactivas de oxigénio. Contudo, neste estudo não foi possível demonstrar a existência de uma relação entre a expressão relativa dos genes analisados e a qualidade da cortiça determinada por análise de imagem.

Conclusões

Este trabalho foi a primeira abordagem a utilizar e integrar clones BAC em diversas metodologias, desde a genética à citogenética molecular, com o propósito de investigar os mecanismos genéticos da biossíntese da suberina e da cortiça. Este trabalho pode também servir como base para o desenho experimental de novos estudos nesta área, nomeadamente para o desenvolvimento de novas ferramentas para o estabelecimento de relações entre o genoma do sobreiro e a qualidade de cortiça. Futuramente, a integração de dados da análise fenotípica e de diferentes níveis de expressão génica poderá contribuir para um maior conhecimento dos processos moleculares que se encontram na base da qualidade da cortiça e que poderá ser aplicado em programas de melhoramento do sobreiro com vista à produção de cortiça de elevada qualidade.

Palavras-chave: Cortiça; suberina; biblioteca BAC; BESs; FISH (fluorescente *in situ* hybridization); PCR quantitativo em tempo real (qPCR).

Contents

List of figures

List of tables

List of abbreviations

BAC - Bacterial artificial chromosome
BES - BAC end sequence
BLASTn - BLAST program algorithm which search a nucleotide database using a nucleotide query
BLASTx - BLAST program algorithm which search proteindatabase using a translated nucleotide query
bp, kb, Mb - Base pair(s), kilobase, megabase
BSA - Bovine serum albumin
cDNA - Complementary DNA
INRA- CNRGV - Institut National de la Recherche Agronomique - Centre National de Ressources Genomiques Vegetales
CT - Cycle threshold
DAPI - 4',6-diamidino-2-phenylindol
dATP - Deoxyadenosine triphosphate
dCTP - Deoxycytidine triphosphate
dGTP - Deoxyguanosine triphosphate
DGRF - Direcção Geral dos Recursos Florestais
DNA - Deoxyribonucleic acid
DNase - Deoxyribonuclease, enzyme degrading DNA
dNTP - Ducleotides mixture
dsDNA - Double - stranded DNA
dTTP - Deoxythimidine triphosphate
dUTP - Deoxyuracyl triphosphate
EC - Enzyme code
EST - Expressed sequence tag
EcoR I - Restriction enzyme isolated from some strains of *Escherichia coli*
EDTA - Ethylenediaminetetracetic acid
FISH - Fluorescence *in situ* hybridization
FITC - Fluorescein isothiocyanate
FJLF - Fundação João Lopes Fernandes
Fwd – Forward
g - Gravitational acceleration (relative centrifuge force)
g, mg, µg, ng, **pg** - Gram, milligram, microgram, nanogram, picogram
h(s), min, sec - Hour(s), minutes, seconds
HMW - High molecular weight
ISA - Instituto Superior de Agronomia
LB - Luria Broth medium
LINE - Long Interspersed repetitive Elements
l, ml, µl - Litres, millilitres, microlitres
LTR - Long terminal repeat of a retrotransposon
µm, nm - Micrometer, nanometer
M, mM, µM - Molar, millimolar, micromolar
mRNA - Messenger ribonucleic acid
NMR - Nuclear magnetic resonance
Nr - Non-redundant
Nt - Nucleotide
PCR - Polymerase chain reaction
qPCR - Quantitative real time polymerase chain reaction
RNA - Ribonucleic acid
ROS - Reactive oxygen species
rRNA – Ribosomal ribonucleic acid
RNase - Ribonuclease
rpm - Revolutions per minute
spp – Species
RT – Room temperature
SDS - Sodium dodecyl sulphate
SSC - Saline sodium citrate
SSR - Simple sequence repeat
Streptavidin-Cy3 - Streptavidin conjugated to fluorochrome cyanine 3

TAE – Tris/acetate/EDTA
Taq thermostable - DNA polymerase from the thermophilic bacterium *Thermus aquaticus*
TBE - Tris/Borate/EDTA
Tm – Temperature of melting
UTL - Universidade Técnica de Lisboa
UV – Ultraviolet
v/v - Volume per volume
w/v - Weight per volume
YAC - Yeast Artificial chromosome

General Introduction

1. The cork oak

Cork oak (*Quercus suber L.*) is a medium size, evergreen broad-leaved tree species that belongs to the Section *Suber* (subgenus Cerris) of the genus *Quercus* (see *Figure 1*). The genus *Quercus* is included in the *Fagaceae* family which includes other important genus for the Northern Hemisphere such as *Castanea* and *Fagus* (Toribio et al., 2005). Cork oak dates from the Oligocene epoch of the Tertiary period. The oldest fossil, dating 10 million years old, is identical to modern cork oak and it was found in Portugal (http://science.jrank.org). The cork oak is a forest species endemic to the Western Mediterranean region, growing in different European and African countries such as: Portugal, Spain, South France, Morocco, Algeria, Tunisia, Corsica, Sardinia, Italy, and Slovenia (Toribio et al., 2005).

Figure 1. *Quercus suber* or cork oak, from the genus *Quercus* is a forest tree species endemic from the Western Mediterranean region (image cortesy of Dr. Jorge Paiva).

The world area occupied by the cork oak is estimated to be 2.3 million hectares and the annual world cork production is around 340 thousand tons. Portugal has approximately 30% of the total world area and produces around 190 thousand tons of cork, which corresponds to more than 50% of the world cork production, followed by 30% of Spain, Italy, France, Morocco, Tunisia and Algeria (Silva J., 2007). According to the latest National Forest Inventory (2005-2006) from Direcção Geral dos Recursos Florestais (DGRF), the cork oak covers about 736,700 hectares and is ranked first species in terms of occupied area in Portugal, constituting also the major cork oak area in the world.

The tree has a broad round crown that reaches up to 20m tall, rarely up to 25m (under ideal conditions) and is long lived from 200 to 250 years. The stem diameter at breast height can reach more than 200cm. The bark is up to 20cm thick, porous and furrowed, with deep longitudinal fissures, which constitutes a thick dermal system important in the protection from forest fires. The leaves are alternate and simple, with the margin entire or with 4-7 pairs of acute teeth and they fall during the second year. Cork oak has predominantly separate male and female flowers on the same plant and is wind pollinated. This specie requires an annual mean temperature of 13-18°C, not tolerating temperatures below -10°C and is found under a wide range of annual rainfall. Cork oaks grow mainly on non- calcareous substrates, preferring sandy and lightly structured soils, but are occasionally found on decarbonated soils.

Naturally, cork oak can occur in mixed forests, sharing the arboreal stratum with other evergreens and

deciduous oaks, with pines and other conifers, and with a few other hardwoods. These agroforestry systems are open woodlands with low tree density (50-300 trees/ha) widely known as montados in Portugal, and dehesas in Spain (see *Figure 2*). In these systems, forage species are commonly grown under the trees and grazed by cattle during the summer.

Figure 2. Cork oaks are usually found together with other species in mixed forests in open woodlands with low tree density commonly known as Montados (image kindly provided by Eng. Pedro Marques, FJLF).

Montados are of great environmental importance in soil conservation, in the hydrological cycle regulation and water quality, in oxygen production and consequently in the carbon recover from the atmosphere and also in the biodiversity conservation. These ecosystems are considered highly rich in biodiversity and one of the most important for nature conservation at national and European level. The network Natura 2000 classified montados (habitat 6310) and cork oak forests (habitat 9330) as very important for the conservation of biodiversity. These are the habitat of some endangered and vulnerable species such as Bonelli's Eagle, Iberian Imperial Eagle, Black Stork and they are also the preferred habitat of the Iberian lynx, the feline most critically threatened in the world.

These forest ecosystems are also very important for the rural development of the interior part of the country, because cork industry generates high local incomes. The management of the montados is a key social-economic factor in the maintenance of the rural employment and equilibrium. Besides cork production, the montados not only allow the breeding of cattle for a good quality meat and milk production, which is the bases of regional and local food industry, but also exploiting hunting resources, apiculture, collection of edible mushrooms as well as and agro and eco-tourism.

1.1 Cork production

The primary use of cork oak is as a source of cork. This is due to the fact that cork oaks produce continuously abundant cork that can be stripped, without causing severe damage, at regular intervals providing commercial cork (Toribio et al., 2005).

The first harvest (virgin cork) is made when the tree is approximately 25-30 years old and the subsequent harvests are usually made each 9 to 12 years, depending on the geographical area (Graça and Pereira, 1992; Graça and Santos, 2007). The cork bark is stripped off for commercial use when it has a thickness of at least 25mm (see *Figure 3*). The cork thickness is achieved by the

accumulation of annual cork rings of about 2-3mm thick that adhere to one another (Caritat et al., 2000). The radial growth of cork starts in the beginning of the spring and continues until autumn, with a peak of growth rate in the late spring (June) (Natividade, 1950; Costa et al., 2001; Fialho et al., 2001).

Figure 3. Images representing the harvesting of cork. The cork bark is stripped off the tree in planks when it has a thickness of at least 25mm (Images kindly provided by Eng. Pedro Marques, FJLF).

Two main types of cork may be distinguished (*Figure 4*): the virgin cork, which is the first- generation cork produced by the original phellogen of the tree and the reproduction cork, which is produced by regenerated phellogens and corresponds to the subsequent cork layers (Pereira H., 1988). The virgin cork is a low quality cork due to its irregular texture. Within the reproduction cork type, it can be distinguished the *secundeira* cork which is the cork resulting from the second harvest and the *amadia* cork which corresponds to the subsequent harvests. The *amadia* cork has the most regular texture and therefore the highest quality and economic value. The cork yield is determined by the perimeter of the tree trunk, the harvesting frequency and the length of bole and main branches that can be stripped off. An old cork oak with large perimeter and branches can yield more than 455 kg of cork in a single harvest. Cork oaks with an age between 35 and 45 years typically yield around 91 kg per year, and trees aging 50 to 60 years can yield 150 kg of cork.

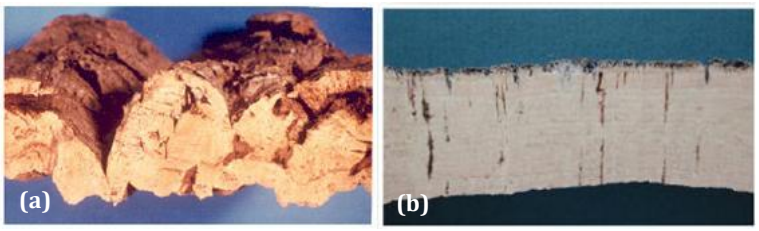


Figure 4. Cross-section of cork planks from different harvests. (a) Virgin cork, which corresponds to the first harvested cork and presents an irregular texture and therefore low quality. (b) Amadia cork (reproduction cork), which is produced after several harvesting cycles. This cork has the most regular texture and thus the highest quality and economic value (Photos kindly provided by Prof. José Graça, UTL-ISA).

1.2 Cork quality

The cork quality is determinant on the technical performance and on the economical value of the cork. The presence of lenticels and other defects of unsuberized tissue in the cork, like woody inclusions, decrease its quality and therefore its economic value. Also, corks constituted by a large number and high dimension of pores are classified as highly porous, and are considered of lower quality than lower porous corks. As an example, cork stoppers of wine bottles with few and small lenticular channels are more than five times more expensive than those with larger pores. In sum, the cork

classification by its porosity is very important in the cork industry (Pereira et al., 1996). Cork quality is a phenotype widely variable, and cork oaks producing high and low-quality cork can be found next to each other, in the same environment. These observations suggests that cork quality is under strong genetic control, although the genomic basic knowledge of this species is still very scarce

1.3 Cork applications

The unique properties of cork conferred by the suberized cell walls, their cellular structure and chemical composition allows cork to be used worldwide in technologically demanding industrial applications. Cork properties include low permeability to gases and water, low density, low temperature conductivity, and high chemical stability with high elasticity and low weight (Gandini et al., 2006; Graça and Santos, 2007).

In 2008, Portugal produced and industrially transformed half of the worlds cork, with an annual export value of 10^9 euros (www.realcork.org/artigo/292.htm).The most common use of cork is in cork-stoppers although there are a myriad of other products that also uses cork, such as floats, innersoles for shoes, gasket, washers, fishing poles grips, bulletin boards, cork paper for cigarette tips among many others. Other important applications of cork are in the production of composite materials which have numerous important applications as sealants and insulators. Cork composite materials are obtained by agglutination of cork granules with diverse binding agents. Cork can be combined with phenolic and polyurethane resins to produce cork panels for walls and floors. Cork granules combined with synthetic rubbers like Neoprene, Hypalon, and Vamac can be to use as sealant in motors (http://www.corkcomposites.amorim.com/). In addition, the granules combined with epoxide resins (Graça and Santos, 2007) are being used as thermal shields in space-craft (www.p2pays.org/ref/34/33161.pdf).

2. The cork tissue

The cork oak most distinctive characteristic is the presence of a cork tissue in the trunk and branches. Cork or phellem, is a secondary tissue that protects woody plant organs and healing tissues from dehydration, solar irradiation and pathogens (Soler et al., 2008). Cork in cork oak grows as a continuous and thick layer formed by the division of the phellogen (cork cambium) on the external side as part of the periderm, which replaces the epidermis in plant secondary growth.

Immature cork cells are parenchymatous with thin primary walls and it is between the fifth to the seventh years of growth that the phellem cells get the characteristics typical of mature cork cells (Pereira et al., 1987). Mature cells are small and have a thick layer of suberin deposited on the walls, resulting in the death of the cells. The cell protoplast is lost and the lumen becomes filled with air, with resiniferous or with tanniniferous substances (Graça and Pereira, 2004). The cells are stacked in the radial direction, with a regular radial arrangement (Graça and Pereira, 2004) and without intercellular voids (Pereira et al., 1987) as shown in *Figure 5*. Cork consists in multiple layers of phellem cells whose walls are highly suberized containing approximately 40% suberin (Pereira H., 1988). It is due to the

presence of suberin and waxes that cork cells present great insulating properties and high impermeability to water and gases (Graça and Pereira, 2004).

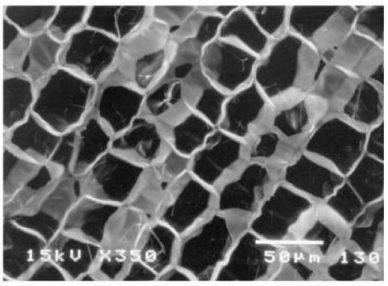


Figure 5. Electron microscopy of cork cells. Image adapted from http://www.realcork.org/artigo/84.htm.

2.1 Phellogen

The phellogen is a plant´s secondary meristem formed by cellular dedifferentiation. In the cork oak, the phellogen differentiates during the first year of growth in the cell layer immediately under the epidermis and divides to form 3-6 suberized phellem cells, as represented in *Figure 6* (Graça and Pereira, 2004). The phellogen forms a continuous layer that surrounds stems and branches and in cork oak lives as long as the tree (Natividade, 1950) with a high meristematic activity (Graça and Pereira, 2004). The division of the phellogen occurs only after suberization of the previous divided cell.

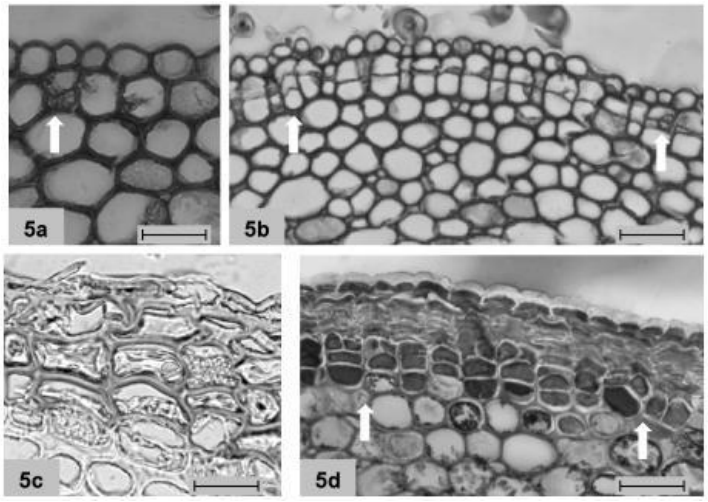


Figure 6. Transverse sections of a1-year-old cork oak shoot, showing the epidermis. 5a) The arrow points to the division of the first cell in the subepidermic layer, which is a precursor for the phellogen formation. 5b) The phellogen formation extends with cells division along the subepidermic layer; the arrow shows phellogen mother cells already divided. 5c) The figure shows the first cells that result from the division of the phellogen. 5d) After 1 year of growth it is observed (pointed by the arrows) a layer of 6 to 8 phellem cells aligned in rows. Scale bar of 5a = 30µm; of 5b and d = 40µm; of 5c = 20µm. Images adapted from Graça and Pereira, 2004.

When the cork layer is removed, the original phellogen is destroyed and a new phellogen differentiates in the outer phloem. In the following years, the new phellogen retains the cylindrical spatial development and a high meristematic activity (Natividade, 1950; Pereira et al., 1992). Cork oak trees may remain in production for at least 200 years and the successive formation of new phellogens following the removal of cork at periodic intervals makes from cork a renewable forest resource that can be harvested on a sustainable basis (Graça and Pereira, 2004).

2.2 Lenticels

Under the stomata there is a lenticular phellogen that produces unsuberized cells, which after three years of the tree growth originate the lenticels, the cork pores (Graça and Pereira, 2004). The lenticels form lenticular channels that radially cross the phellem and they are in part open and in part filled with the loosened filling tissue as may be seen in *Figure 7* (Graça and Pereira, 2004). The lenticular channels create discontinuities in the layer of suberized cells, compromising the cork quality and its commercial value.

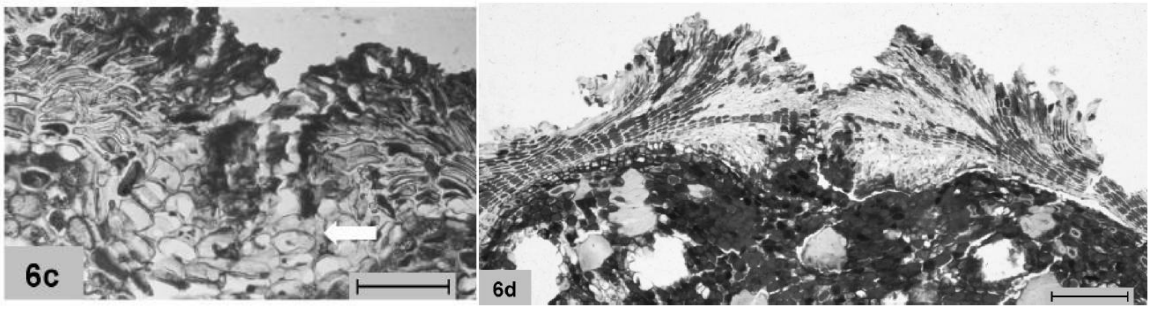


Figure 7. This figure presents a lenticel with cells with a loosened arrangment, which are produced by the lenticular phellogen. 6c) It can be observed sclereids, as the arrow points. 6d) The figure represents a lenticular channel in a cork oak branch after 3 years of growth. It can be seen, in the cortex, large sclerenchymatous nodules. Images adapted from Graça and Pereira, 2004.

3. Cork chemical composition

The cork cellular structure, chemical composition, physical and mechanical properties have been well described (Silva et al., 2005). Although the amounts of the cork different components can show significant variations (Pereira H., 1988; Lopes et al., 2001), on average cork contains 15% extractives, which includes 7.5% waxes (mostly terpenes and sterols) and 7.5% tannins, 41% aliphatic suberin (referred to as suberin), 22% aromatic suberin (also referred to as cork lignin), 20% polysaccharides, and 2% ashes (Pereira H., 1988; Soler et al., 2007). Cork originated from cork oak is a plant tissue with one of the highest suberin content, each year cork oak produces about 60 layers of almost pure cork cells with a suberin content of more than 50% of its dry weight (Pereira H., 1988). Thus, cork oak is the paradigm of cork and suberin research.

3.1 Suberin

Suberin is the plant cell-wall biopolymer that defines suberized cells. Suberin and suberized cells provide a barrier to water loss, attacks by pathogens, excessive heating and they are also formed as a defense towards physical injury and biological infection (Graça J., 2010).

Suberin is defined in literature as a complex biopolymer found in suberized cells, which consists of a polyaliphatic cutin-like domain cross-linked with a polyaromatic lignin-like domain that is derived from ferulic acid (Kolattukudy P., 2001; Bernards M., 2002). The aliphatic domain is a glycerol-bridged polyester with associated esterified phenolics (Moire et al., 1999; Graça and Santos, 2006). The aromatic domain is a polyphenolic substance mostly composed of hydroxycinnamic acid derivatives and is presumably involved in linking the aliphatic domain to the cell wall (Kolattukudy P., 1980; Bernards and

Lewis, 1998).

Although suberin has great biological importance, very little is known about the molecular processes on the bases of its biosynthesis and deposition. This results from difficulties in working with appropriate experimental model plants. Despite the cork oak is the model for cork and suberin production, its low transformation efficiency and its extraordinarily long regeneration time, makes functional genetic studies in this tree very challenging. On the other hand, most of what is known about the suberin structure and the molecular mechanisms of its deposition have been studied in the potato periderm (*Solanum tuberosum*), in the wound periderm (*e.g*. Chaves et al., 2009). This plant can be a good model for such studies as it can be easily transformed by *Agrobacterium tumefaciens* and grown in the laboratory (Soler M., 2008).

3.2 Suberin Applications

Significant amounts of plant wastes rich in suberin are produced each year, for example, the bark of birch used in pulp production in northern countries or in cork industries in southern countries. The Portuguese cork industry produces about 40, 000 ton/year of cork powder as a by-product (Eckerman et al., 1985). Efforts to develop new applications with suberin materials are old, but only recently the field has experienced a considerable thrust, greatly due to the advances of new isolation techniques and the search for bio-based materials.

Most of the products derived from suberin are at this time used for energy production. The potential of recovering the suberin monomers from these sources, namely α, β-diacids and ω-hydroxyacids, has not yet been fully explored. Suberized plant tissues and suberin are expected in the future to be a source of specialty chemicals and to serve as the basis for bioinspired materials (Graça and Santos, 2007).

Products derived from cork are valuable for composite material applications such as sealants and insulators, and plant-produced suberin monomers have numerous potential uses as polyester foams and biolubrificants (Gandini et al., 2006). In particular, promising results were obtained in the synthesis of polyurethane foams, after the oxypropylation of cork suberin. The unique structure of suberin and their constituent monomers are also thought to enable several potential uses in pharmaceuticals and cosmetics. For instance, it was found that suberin inhibits mutagenesis (Krizková et al., 1999) and that it can act as an absorbent of carcinogens (Harris and Ferguson, 1999), and that cork suberin hydrolysates can act as anti-ageing agents with smoothing anti-wrinkle action in the human skin (Coquet et al., 2005).

3.3 Suberin chemical composition

In the last decade, great progress has been achieved in the structural characterization of suberin. Suberin *in situ* is an insoluble heteropolymer deposited within the plant cell wall and its removal from the suberized tissues is achieved by depolymerization reactions which can be performed by reactions that break ester bonds, like hydrolysis, alcoholysis, or hydrogenolysis (Franke and Schreiber, 2007;

Graça and Santos, 2007).

Upon suberin depolymerization the main monomers released include glycerol, long-chain ω -hydroxyacids (ω-OH-acids) and α, ω-dicarboxylic acids (α, ω-diacids) ranging in chain length from C16 to C30. Small quantities of alcohols, ferulates and unsubstituted fatty acids are also components present among the suberin depolymerization products (Franke and Schreiber, 2007; Graça and Santos, 2007). The aromatic network of suberin is a polymer of C-C and ether-linked hydroxycinnamic acids, N-feruloyltyramine and monolignols (*Figure 9*), and therefore these monomers are not released by polyester depolymerization methods. Suberin seems to have the general basic pattern of glycerol, α, ω -diacids, and ω-hydroxyacids, although the relative abundance of saturated and mid- chain substituted monomers can vary depending on the plant species (Franke and Schreiber, 2007; Graça and Santos, 2007). It is the case of suberin from the cork oak bark which is characterized by in-chain epoxides and vicinal diols (Graça and Pereira, 2000). Specifically associated with suberin are saturated aliphatic monomers with chain lengths beyond C20, higher levels of aromatics and primary alcohol synthesis (Pollard et al., 2008). Suberin-associated waxes are in part similar to the structural elements of the polyester and also include components that are biosynthetically and chemically closely related with the suberin monomers (Pollard et al., 2008). The suberin waxes include alkanes, primary alcohols, fatty acids, alkyl ferulates and monoacylglycerols.

3.4 Suberin in the cell wall

In general suberized cell walls are very thin (less than 1mm thick) and present a primary wall, a secondary wall and sometimes a tertiary wall which confines the cell lumen. Topochemical studies have shown that the primary wall contains the polysaccharides and part of the polyaromatics, which are also present in the tertiary walls. The suberin and part of the polyaromatics are located in the secondary wall, which contributes for most of the cell wall thickness (Graça and Santos, 2007). Some works have shown that in the inner side of the cell wall there are deposited large quantities of extractives associated with suberin as represented in *Figure 8* (Sitte, 1962).

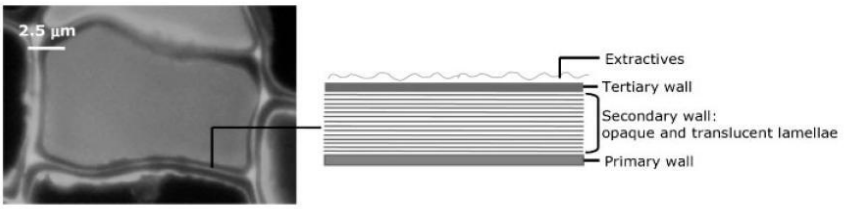


Figure 8. Cross-section of a suberized cork cell by UV fluorescence microscopy (left) and a schematic representation of the cork cell wall with the distribution of its components (right). Image from Graça and Santos, 2007**.**

The study of Graça and Santos (2007), with transmission electron microscopy (TEM), showed that the secondary walls containing suberin present a lamellate structure at the ultrastructural level and in suberized cork cells it were counted between 30 to 60 alternate opaque and translucid lamellae. Within the translucid membrane it was found structures perpendicular to the plane of the lamellae. It has

been proposed that the translucent lamellae correspond to aliphatic constituents and the dense lamellae correspond to aromatic constituents (Schmutz et al., 1993; Graça and Santos, 2007).

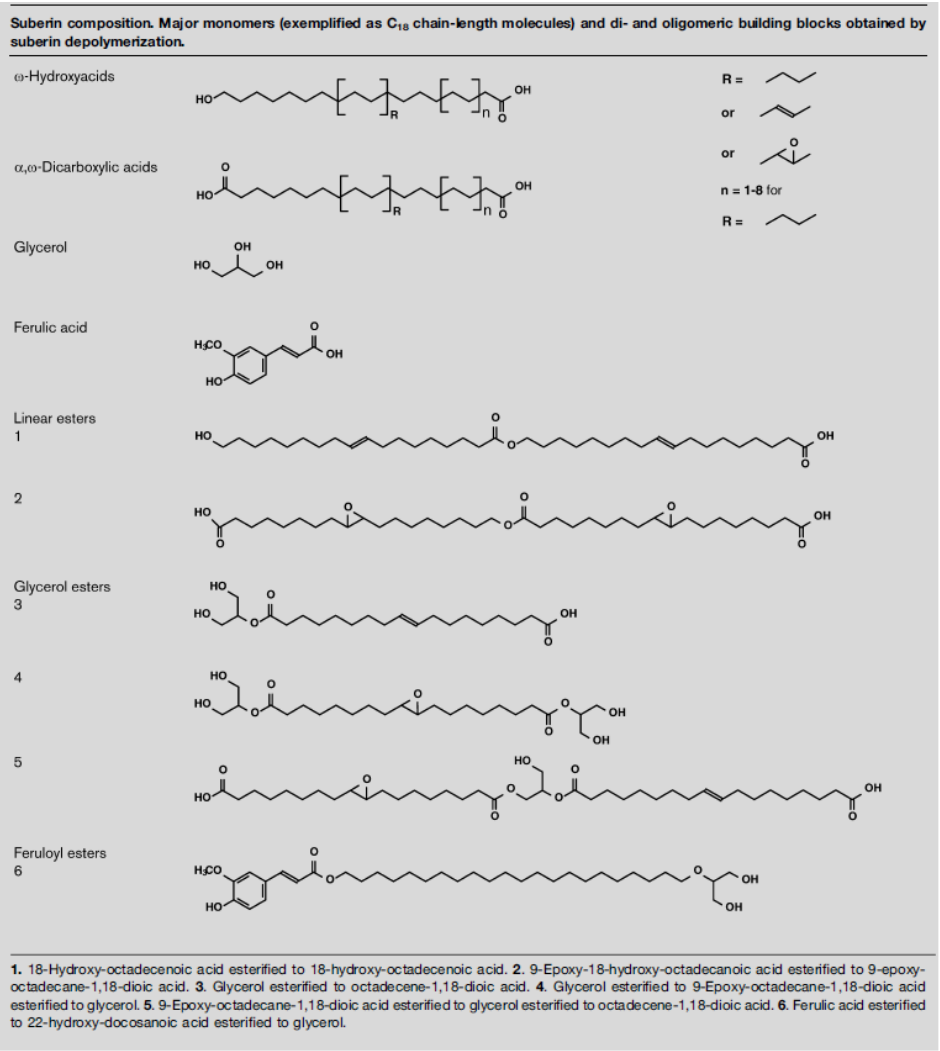


Figure 9. Suberin chemical composition adapted from Franke and Schreiber, 2007.

3.5 Suberin macromolecular structure

The macromolecular structure of suberin and its relation to other cell wall components are still poorly understood (Graça J., 2010). Although several models for the three-dimensional structure have been presented, the molecular architecture of suberin remains hypothetical (Pollard et al., 2008; Graça and Santos, 2007; Bernards M., 2002). The several proposed models have been based on lipid biochemistry, suberin structural studies and suberin chemical analysis. These models present possible linkages within the suberin suggesting linkages between suberin and the lignin/carbohydrate cell wall matrix (Soler et al., 2007). The knowledge of the suberin structure comes from two main sources: the oligomeric blocks obtained by partial depolymerization and the observations on the original biopolymer by solid-state NMR spectroscopy.

In the study of Graça and Santos (2007) it is proposed a suberin macromolecular model that is shown in *Figure 10*. Even though this model is represented in two dimensions this macromolecular

structure can be built in three dimensions. In this model it is proposed that the translucent lamellae of the suberized cell walls correspond to the suberin aliphatic polyester, where the α, ω-diacid long-chain monomers present the chains stretched, perpendicular to the lamellae plane and linked on both sides with glicerol. The opaque lamellae correspond to the polyaromatics that are covalently linked to the translucent lamellae (aliphatic suberin polyester) by the ferulic acid. Linear chains of ω -hydroxy-acids attach the polyaromatics to the aliphatic polyester and correspond to the translucent lamellae (Graça and Santos, 2007).

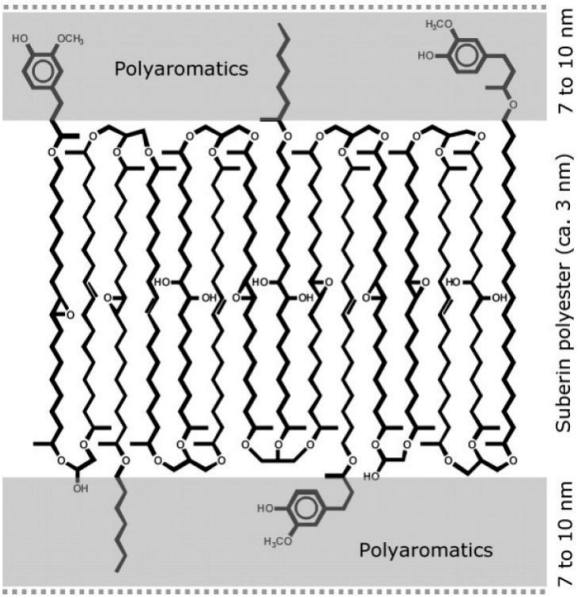


Figure 10. Suberin macromolecular model proposed by Graça and Santos (2007).

3.6 Suberin biosynthesis

Despite the physiological importance of suberin, there is still a great gap in the knowledge of suberin biosynthesis and deposition. In the work of Franke et al. (2005) it is proposed a biosynthethic pathway and suberin candidate genes in Arabidopsis, in which are described the enzyme activities necessary for the synthesis of suberin precursors.

There are two major metabolic pathways involved in the biosynthesis of the suberin, one is the acyl-lipid pathway for the aliphatic domain which consists mainly in fatty acid elongation or acyl elongation. The other one is the phenylpropanoid pathway for the biosynthesis of the aromatic domain that includes ω-carbon oxidation or acyl oxidation (Franke and Schreiber, 2007).

In the work of Soler et al. (2007), are reported a list of genes potentially important for the biosynthesis, transport and polymerization of suberin. The list of genes consists mainly of enzymes involved in the biosynthesis of suberin and some putative cork regulatory genes that could be important to shed some light in this field that is poorly studied. The described set of genes is, in general, in accord with the biosynthetic pathways suggested by evidences from other works such as the ones from Bernards (2002), Franke and Schreiber (2007) and Graça and Santos (2007). Based on this list of candidate genes it was possible to design a plan of the suberin biosynthesis which is summarized in the next subsection.

For a better understanding, a schematic representation of the biosynthetic pathways of the suberin monomers is presented in *Figure 11*.

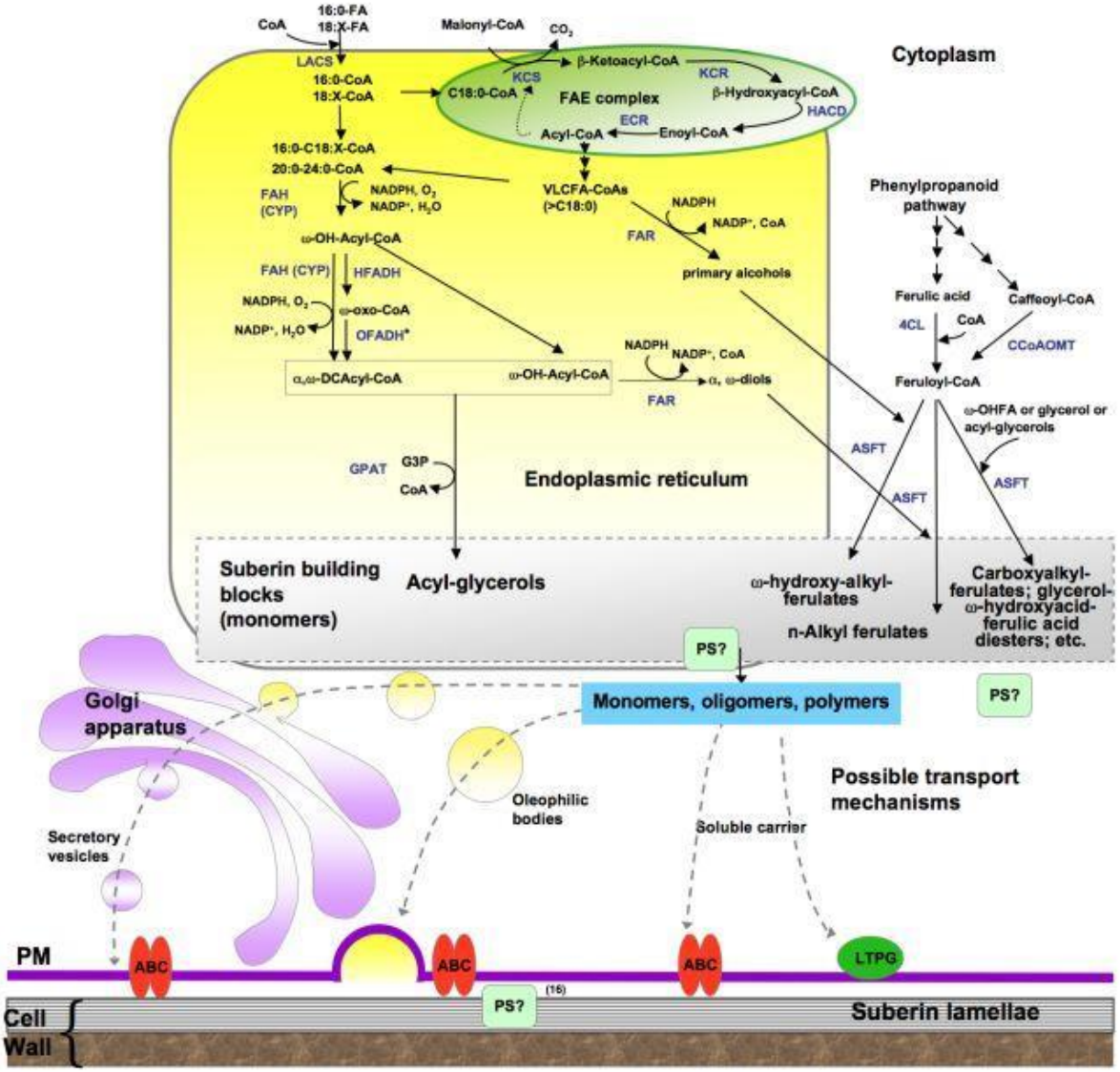


Figure 11. Simplified model with proposed pathways for the suberin monomers biosynthesis. It are represented the two main metabolic pathways involved in the biosynthesis of the suberin, on the left side is the acyl-lipid pathway for the biosynthesis of the aliphatic domain, which consists mainly in fatty acid elongation(acyl elongation). On the right side of the figure is represented the phenylpropanoid pathway for the biosynthesis of the aromatic domain that includes ω -carbon oxidation (acyl oxidation) (Franke and Schreiber, 2007). Abbreviations: 4CL, 4-coumarate ligase; ABC, ATP binding cassette transporter; ASFT, aliphatic suberin feruloyl transferase; CCoAOMT, caffeoyl coA O-methyltransferase; CYP, cytochrome P450 monooxygenase; DCA, dicaboxylic acid; ECR, enoyl-CoA reductase; FAE, fatty acid elongation; FAH, Fatty acyl ω -hydroxylase; FAR, alcohol-forming fatty acyl-CoA reductase; G3P, glycerol-3-phosphate; GPAT, glycerol 3-phosphate acyltransferase; HCD, β –hydroxyacyl-CoA dehydratase; HFADH, ω-hydroxy fatty acyl dehydrogenase; KCR, β –ketoacyl-CoA reductase; KCS, β -ketoacyl-CoA synthase LACS, long chain acyl-coA synthase; LTP, lipid transport protein; LTPG, glycosylphosphatidyl-inositol (GPI)-anchored protein; OFADH, ω -oxo fatty acyl dehydrogenase; OHFA, ω -hydroxy fatty acid; PS, polyester synthase. Figure from The AOCS Lipid Library (http://lipidlibrary.aocs.org/plantbio/polyesters/index.htm).

3.6.1 Candidate genes for the synthesis of the aliphatic polymer (acyl-lipid pathway)

1. Biosynthesis *de novo* of fatty acids in the chloroplast, involving the enzymes dihydrolipoamida S-acetyltransferase (DHCA-S-AT, EC: 2.3.1.12) and biotin carboxyl carrier protein.

2. Exportation of the fatty acids from the chloroplast by the PATE/FAT (palmitoyl-acyl carrier protein thioesterase) and LACS (long-chain acyl CoA synthase, EC: 6.2.1.3). Bonaventure et al. (2003) showed that the Arabidopsis orthologous gene to cork FAT encodes an enzyme that regulates the saturated fatty acids fluxes.

3. Elongation of the fatty acids, in the endoplasmic reticulum, by the enzymes of the fatty acid elongation complex (FAE) to produce very long chain fatty acids. Usually the acyl precursors for lipid polyesters are regular C16 and C18 fatty acids and it is known that the FAE complex extends the C16 and C18 acyl chains to very long chain fatty Acyl-CoAs. In this work were identified two genes encoding condensing enzymes of these complexes, a KCS (β-ketoacyl CoA synthase, EC: 2.3.1.119) which catalyses the first step of this reaction and KCR (β-ketoacyl CoA reductase, EC: 1.1.1.-) (*e.g.* Dietrich et al., 2005).

4. Hydroxylation of the terminal methyl of aliphatics, or the ω-terminal position, is usually catalyzed by NADP-dependent cytochrome P450 monooxygenases (P450). In particular the gene Cyp86A1, encoding the enzyme fatty acid ω-hydroxylase (EC: 1.14.-.-), is thought to provide fatty acid precursors for the suberin polyester (Franke and Schreiber, 2007).

3.6.2 Candidate genes for the assembly and transport of the aliphatic polyester

1. Assembly of the aliphatic polyester. The polymerization of the suberin glycerol polyester is thought to take place in the apoplast although it can also occur in the simplast. Glycerol is esterified with α, ω-diacids and ω-hydroxyacids by the enzymes acyltransferases (AT). The gene GPAT encodes the enzyme glycerol-3-phosphate acyltransferases (an acyl-CoA) (EC: 2.3.1.15) and the Arabidopsis ortholog of cork GPAT has been confirmed to catalyze the formation of ester bonds between fatty acids and glycerol (Beisson et al., 2007). Hydroxycynamoil-CoA/benzoyltransferase (HCBT, EC2.3.1.144) or GDSL-motif lipase/hidrolase protein like (GDSL, EC: 3.1.1.-) may also be implicated in esterifications of suberin monomers involving ferulates and hydroxylated fatty acids. In general, the fatty acyl chains undertake three main modification reactions: activation to coenzyme A thioesters by LACS enzymes (EC: 6.2.1.3) (acyl activation), ω -oxidation of the acyl-CoAs catalyzed by NADP-dependent cytochrome P450 monooxygenases (P450) in particular the gene CYP86A1, encoding the enzyme fatty acid ω -hydroxylase. And the third reaction is the esterification to produce glycerol-3-phosphate (acyl glycerols) by the GPAT enzymes (http://lipidlibrary.aocs.org/). These reactions may be ordered in several ways to produce ω -oxidized acylglycerols, which is a presumed structural element of lipid polymers. Pollard et al. (2008) purposed three possible pathways for the order of these reactions.

2. Transport of aliphatic materials to the apoplast. It is thought that after esterification the suberin aliphatic building blocks are transported to the apoplast via the plasma membrane ATP-binding cassette (ABC) transporters or via Golgi derived vesicles. The ABC transporters are membrane proteins known for translocating a wide variety of substances across biological membranes, such as sterols, drugs and lipids.

3.6.3 Candidate genes for the synthesis and assembly of the aromatic polymer (phenylpropanoid pathway)

1. Synthesis of aromatic monomers. The biosynthetic pathway of the aromatic monomers is based on the phenylpropanoid metabolism. The first step is the deamination of phenylalanine by the enzyme PAL (phenylalanine ammonia-lyase, EC: 4.3.1.25) followed by a hydroxylation with the C4H (cinnamate 4-hydroxylase, EC: 1.14.13.11) enzyme. Other enzymes seem to be also very important for the phenylpropanoid pathway such as COMT (caffeic acid O-methyltranferase, EC: 2.1.1.68), CCOMT (caffeoyl CoA 3-O- methyltranferase EC: 2.1.1.104), 4CL (4-coumarate: CoA ligase, EC: 6.2.1.12), F5H (ferulate-5-hydroxylase, EC: 1.14.-.-) and CCR (Cinnamoyl-CoA reductase, EC: 1.2.1.44) (Bernards and Lewis, 1998). From this metabolism derives two types of products, the hydroxycinnamic acids and monolignols.

The genes HCBT encode a hydroxycinnamoyl transferase, an enzyme crucial for the synthesis of the suberin aromatic (phenolic) monomers (Bernards, 2002). The HCBT genes belong to the BAHD gene family, which is known to encode acyl transferases. These enzymes catalyze the synthesis for example of feruloyltyramine and there are evidences that this is a component of the aromatic suberin of cork oak (D´Auria, 2006).

2. Assembly of the aromatic polymer. In the studies of Kolattukudy (2001) and Bernards (2002), it is proposed that class III peroxidases are involved in the assembly of the aromatic monomers. Although Soler et al. (2007) in his work was not able to find any class III peroxidase, it were found three genes coding for laccases which are extracellular oxidases capable of coupling phenylpropanoids.

3.6.4 Candidate genes for the linkage of the aromatic and aliphatic domains

It is thought that the linkage of the two suberin domains occurs due to the esterification between the ferulates and the fatty acids (Bernards and Lewis 1992, Lofty et al. 1994, Graça and Santos 2007). In the study of Soler et al. (2007) it were found two HCBT genes and one gene encoding an acyltranferase. These genes belong to the BAHD acyltranferase superfamily, a family which contains members that catalyze esterification of hydroxycinnamates with fatty acids (Lotfy et al., 1995) and so they could be responsible for the linkage of the suberin aromatic and aliphatic domains.

3.6.5 Candidate genes for regulation of cork formation

Despite the importance of cork tissue the regulatory mechanisms controlling suberin, cork formation and its differentiation are practically unknown. However there have been numerous studies in the regulatory factors involved in lignin deposition and wood differentiation. There is also a limited knowledge about the hormonal control of cork formation but in this work it was found the ethylene forming enzyme aminocyclopropane-carboxylate (ACC) oxidase (EC: 1.14.17.4) highly expressed in cork and wood. The ethylene role in these tissues is not clear (Andersson-Gunnerås et al., 2003; Lulai and Suttle, 2004) but it is possible that it could be a general regulator in cork and wood tissues.

In this work, it was also found highly induced in cork a WRKY transcription factor. The WRKY factors bind to W-boxes (cis-regulatory elements) modulating the gene expression of some stress induced genes, including P450s (Mahalingam et al., 2003; Narusaka et al., 2004). The WRKY transcription factors are a wide family of plant-specific regulators mostly controlling senescence, stress and defense responses (Eulgem et al., 2000). Although in this work, were indentified some regulatory candidate genes for cork regulation there is still a lack of knowledge on their specific role.

4. BAC libraries

Bacterial artificial chromosomes (BAC) are bacterial vectors constructed for cloning and stable preservation of large DNA fragments with insert sizes ranging from 30 to 300 kb. Species of the *Quercus* genus are diploid (2n = 24) and the haploid DNA may vary among species ranging from 539 Mb in *Q. velutina* to 921 Mb in *Q. coccifera* and *Q. ilex*, 740 Mb in *Q. robur* (Kremer A., 2007) and 1000 Mb in *Q. suber.* BAC vectors can carry large DNA inserts, although insert sizes between 80 and 200 kb are more typical.

Development of BAC vectors

BACs, contrary to their names are not in fact artificial chromosomes since they are derived from the F´ Factor (fertility factor) plasmid that naturally occurs in *Escherichia coli*. The F factor is an incompatibility group that can exist as an extra-chromosomal element, usually with 100 kb, which is involved in the chromosomal transfer and conjugation processes in *E. coli*. The mini-F plasmid pMBO131 was the base for the construction of the original BAC vector pBAC108L. Meanwhile, other BAC vectors were constructed, such as pBeloBAC11 and pBeloBACe3.6, as modified versions of the original BAC vector pBAC108L designed for specific uses.

The BAC vectors encode genes derived from the mini-F plasmid, for autonomous replication (*oriS* and *repE*), for copy-number control (*parA* and *parB*) that preserve the copy number to one or two in each host cell, and for plasmid partitioning. Additionally, it was added to the vector *cosN* and *loxP* sites to allow plasmid linearizarion for restriction mapping and multiple cloning sites, sided by T7 and SP6 universal promoters that are flanked by restriction enzyme sites for insert excision. Moreover, BAC vectors include a chloramphenicol resistance gene as a negative selectable marker of non-transformed bacteria. In this

study, was used the pIndigoBAC-5 BAC vector which is derived from the pBeloBAC11 and pIndigo BAC vectors (*Figure 12*). The pBeloBAC11 main characterisc is the presence of the *lacZ* gene in the multiple cloning site, which is inactivated with the successful ligation of the insert to the vector allowing the blue/white screening of recombinants.

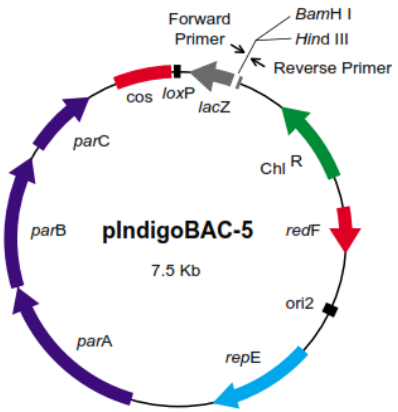


Figure 12. Representation of the pIndigoBAC-5 vector. This BAC is derived from pBeloBAC11 and pIndigoBAC vectors and can stably carry DNA inserts higher than 100 kb. This vector is linearized at either the unique *Bam*H I or *Hin*d III site, completely dephosphorylated, and highly purified. The genes *parA, parB, and parC* are required for partitioning. Additionally, *parB* and *parC* are required for incompatibility with other F factors. The *repE* gene encodes a protein essential for replication from the *ori2*. Chl^R is a chloramphenicol resistance gene incorporated for antibiotic selection of transformants. pIndigoBAC-5 has a polycloning site with recognition sequences for two different restriction enzymes (*Hind*III and *Bam*HI). The polycloning site is located within the *lac*Z gene allowing identification of recombinants by alphacomplementation (Adapted from Peterson et al., 2000 and http://www.epibio.com/item.asp?ID=567).

BAC clones advantages

BACs appeared in the early 1990s as an alternative to Yeast Artificial chromosomes (YACs) (Shizuya et al., 1992) that had several disadvantages such as: presenting high percentage of insert rearrengements and formation of chimeric clones (Burke, 1990; Neil et al., 1990; Green et al., 1991; Anderson, 1993; Venter et al., 1996; Cai et al., 1998) and were very difficult to handle (O'Conner et al., 1989; Woo et al., 1994). Contrary to YAC clones, BACs present low percentages of insert rearrangements and chimerism. In addition, the resulting BAC clones are simple to manipulate and stable albeit containing large inserts of genomic DNA of up to 200 kb (Villalobos et al., 2004). The BAC inserts stability seems to be due, partly to F factor genes *parA* and *parB* which avoid the possibility of a bacterium to be inhabited by more than one BAC (Willetts and Skurray, 1987; Shizuya et al., 1992; Cai et al., 1998).

BAC clones applications

Thus, BAC vectors are appropriate to construct large-insert DNA libraries for extensive cloning required by large-scale genome-sequencing projects and currently BAC libraries are key genomic tools that enable positional cloning of important traits, synteny evaluation, and the development of genome framework physical maps for genetic linkage and genome sequencing. Additionally, recently has been developed the BAC-end sequencing a method based on Sanger dideoxy DNA sequencing and DNA amplification (Rosenblum et al., 1997) and which can be combined with chromosome walking for contig construction. Furthermore, BACs have been successfully used as probes in fluorescence *in situ* hybridization (FISH), a technique called BAC ´landing´.

Due to their characteristics, the BAC vectors have become increasingly popular for the producing of libraries from a vast diversity of genomes. BAC libraries have been constructed for microbial, animal and plant species. BAC libraries have been higly used for the creation of physical genetic maps of model species (e.g. *Homo sapiens, Mus musculus* and *Arabidopsis thaliana*) and of economically important species including *Oryza sativa* (rice) and *Zea Mais* (corn). BAC libraries have been widely developed as genomic tools for diverse plant species, since many different crop plants (Frijters et al., 1997; Marek and Shoemaker, 1997; Vinatzer et al,. 1998; Woo et al., 1994; Zhang et al., 1996) to woody plants, including forest trees species such as *Liriodendron tulipifera* (Liang et al., 2007), *Coryllus avellana* (Kremer et al., 2010), *Pinus pinaster* (Bautista et al., 2007 and 2008), *Pinus taeda* (Magbanua et al., 2011), *Picea glauca* (Hamberger et al., 2009), *Populus trichocarpa* (Stirling et al., 2001; Tuskan et al., 2004), *P. tremuloides* (Tuskan et al., 2004; Fladung et al., 2008) and *Eucalyptus grandis* (Paiva et al., 2011a). Within the *Fagaceae* family, were developed BAC libraries for *Castanea mollissima* (Kremer et al., 2010) and *Quercus robur* (Rampant et al., 2011) enabling the construction of physical maps. In both species, BAC libraries and physical maps have been important sources for ongoing genome sequencing projects.

Objectives

The scarcity of studies in cork genomics it is far from allowing a full understanding of the molecular processes for suberin biosynthesis and cork formation. As such, the general objective of this work was to expand our knowledge on the genomics of cork formation and cork quality.

As specific obejectives:

a) to screen the *Q. suber* BAC library for the presence of several candidate genes relevant for suberin and cork biosynthesis and to provide some hints on the structure and composition of the cork oak (Chapter 1);
b) to test the BAC landing methodology in order to develop a *Q. suber* cytogenetic map (Chapter 2);
c) to analyse and characterize the gene expression variability of candidate genes associated with suberin and cork biosynthesis (Chapter 3).

Previously to this study, was constructed and characterized the first deep-coverage *Quercus suber* BAC library. For this we started by selecting a set of candidate genes from two of the main biosynthetic pathways involved in suberin biosynthesis, as well as stress related genes and transcription factors involved in cork formation. Then we performed high-density BAC filters hybridizations followed by PCR validation to evaluate the presence of the candidate genes in the BAC library. A second aim was to characterize the structure and composition of the cork oak genome, and for this we sequenced and analyzed BAC end sequences (BES) from BAC clones which included some of the candidate genes (Chapter 1).

In a second part of our study (Chapter 2), we attempted to map six candidate genes for cork formation in the cork oak genome and thus to develop the first *Q. suber* cytogenetic map. For that, was selected a group of BAC clones containing genes of interest to hybridize with mitotic cork oak chromosomes, using the BAC landing methodology.

In the last part of this work (Chapter 3), the aim was to analyse and characterize the gene expression of candidate genes associated with suberin and cork biosynthesis, in cork-differentiating tissue, from cork oak trees producing cork with different qualities. For this purpose we used quantitative PCR (qPCR) to identify and create a list of candidate genes associated with the cork quality phenotype.

Chapter 1 – Screening cork biosynthesis genes on the first Quercus suber BAC library and preliminary BAC end sequencing analysis

1. Introduction

One of the main reasons for cork oak genomic research is to identify genes involved in cork formation and cork quality. This would help to build a basic genomic knowledge to establish relations between the cork oak genome and the cork quality trait as a tree phenotype. These studies are intended to enhance the basic knowledge in molecular genetics and to be used, in the future, as tools for cork oak tree-breeding programs for production of cork with higher quality. Nowadays, many researchers in the field of genomic research have found the use of deep-coverage large-insert genomic libraries, such as bacterial artificial chromosomes (BACs) as essential tools to achieve these objectives.

Construction and characterization of the first Quercus suber BAC library

Previously to this work, Paiva et al. (2011b) constructed and characterized the first *Quercus suber* BAC library (Qsu-B-HDL). The library was constructed from extracted cell nuclei of leaves from cork oak trees from Herdade dos Leitões (Portugal). High molecular weight (HMW) DNA was prepared and sent to the Centre National de Ressources Genomiques Vegetales from the Institut National de la Recherche Agronomique (INRA-CNRGV) for construction and characterization of the *Q. suber* BAC library. The HMW DNA was digested with *Hind*III and the DNA restriction fragments were size selected. The DNA fragments were ligated to the dephosphorylated pIndigoBAC-5 *Hind*III vector to transform *Escherichia coli* cells, cultivated in a medium with chloramphenicol as selective antibiotic. Later, were performed quality tests and the library was characterized by insert size analysis of random clones by pulsed-field gel electrophoresis.

The *Q. suber* BAC library had a total of 180 plates with 348 wells and an estimated genome coverage of 5x haploid genome equivalents from a total of 41,472 clones, which increases the probability to find single copy genes in the library. Though genome size of cork oak is large, 1,000 Mbp, the insert size analysis showed that the constructed BAC library contains large inserts with an average size of 120 kb. The results of the BAC library characterization are summarized in *Table 1*.

Whithin the *Fagaceae* family has been developed several large insert libraries resources for different species. It is reported the construction of one *Eco*RI and two *Hind*III BAC libraries for *Castanea molissima* (Chinese chestnut), containing insert sizes with an average ranging from 90 to 128 kb, and with genome coverages from 10 to 12×. In addition, for comparative genomic studies was constructed a *Hind*III BAC library for *Castanea dentata* (American chestnut), with 12x genome coverage and with an average insert size of 140 kb. Moreover, were made *Eco*RI 12x (average insert size 135 kb) and *Hind*III 14x (average insert size 120 kb) BAC libraries from *Quercus robur*, which enabled the construction of genetic and QTL maps.

Table 1. Characterization summary of the *Quercus suber* BAC library (Qsu-B-HDL).

Qsu-B-HDL: *Quercus suber* BAC "Herdade dos Leitões" Library			
International name	Qsu-B-HDL	**Host**	DH10B TIR
CNRGV name	Qsu-B-HDL	**Number of clones**	41,472
Species	*Quercus suber*	**Number of plates**	108
Common name	Cork oak	**Plates type**	Genetix 384 wells
Ecotype	Herdade dos Leitões (Portugal)	**Average insert size**	120 Kb
Restriction enzyme	*Hind*III	**Genome size**	1,000 Mb
Vector	pIndigoBAC-5 *Hind* III	**Genome equivalents**	5
Selective antibiotic	Cloramphenicol		

Quercus suber BAC end sequences analysis

BAC end sequencing

BAC end sequences (BES) are sequences adjacent to the insert sites of BAC clones. Recently, has been developed a technique called BAC-end sequencing which is based on Sanger dideoxy DNA sequencing. BAC libraries have been widely selected for high throughput genomic sequencing projects mainly due to the BAC vector high stability when compared to YACs or cosmid vectors. The *E. coli* F-factor replicon allows copy number control, limiting the number of BACs to one or two copies per cell. As a result, the probability of DNA rearrangements and deletions is reduced, and the toxic effects of the cloning in *E. coli* cells decrease. Additionally large-insert genomic libraries allow the division of complex genomes into DNA segments. However, BAC end sequencing has proven to be difficult due to the low abundance of BAC DNA and the difficulty to obtain sufficient quantities of high quality template from standard minipreps (Kelley et al., 1999). Furthermore, the large molecular weight and secondary structure of BACs are a challenge for high throughput direct sequencing of the BAC clone ends.

BAC end sequences applications

BAC end sequences information has enhanced the BACs value as genomic resources and have been widely used in a variety of applications from gene discovery to genome sequencing. The collection of BESs in large databases has allowed: the identification of minimally overlapping clones for sequencing large genomic regions (Boysen et al., 1997); to find adequate clones for fluorescence *in situ* hybridization mapping; to find clones for restriction fingerprints for building overlapping clones sets; or to select BAC clones containing genes of interest (Poulsen and Johnsen, 2004; Paiva et al. 2011a). BESs are sometimes the first sight of sequence composition from genomes not yet sequenced (Mao et al., 2000; Zhao et al., 2001; Hong et al., 2004). Their analyses are very useful for determination of genome content and architecture (Seghal et al., 2012). Furthermore, BES are also an important source for the development of molecular markers, such as: simple sequence repeats (SSRs); genic sequence-based markers and, inserted transposable elements junction based ISBP (Insertion Site Based Polymorphisms) markers (Paux et al., 2006; Bartos et al., 2008), that can be very useful for genetic and physical mapping, as well as for breeding (Tomkins et al., 2004). Physical maps, based on BACs, have been constructed for the complete genomes of rice (Zhang and Wing, 1997; Chen et al., 2002), *Arabidopsis* (Mozo et al., 1999), maize (Coe et al., 2002),

Brassica rapa (Mun et al., 2008), bean (Schlueter et al., 2008), *Brachypodium* (Gu et al., 2009), papaya (Yu et al., 2009), melon (Gonzalez et al., 2010), and soybean (Wu et al., 2004).

In this part of the study, we had two main objectives:

i. to screen the BAC library for genes of interest for cork formation, through high-density BAC filters hybridizations and validation of hybridization positive clones by Colony PCR. The set of genes studied included genes involved in suberin biosynthesis pathways (Acyl lipids and Phenylpropanoid pathways) and in regulatory mechanisms of cork formation.
ii. the other aim of this work, was to characterize the composition of the cork oak genome by sequencing and analyzing 24 BAC end sequences. The BAC end sequences were selected from the group of BAC clones validated by Colony PCR containing candidate genes for cork formation. To analyze the cork oak BESs we searched for repetitive DNA sequences, SSRs and performed homology searches against diverse databases using bioinformatic tools.

1.1 Materials and Methods

1.1.1 BAC library screening for genes of interest

For screening genes of interest in the *Q. suber* BAC library (Qsu-B-HDL) (Paiva et al., 2011b). BAC clones were spotted onto a a nylon membrane to produced four high density filters using a Genetix Q-bot (Genetix, New Hamilton, Hampshire, UK). The filters (22x22 cm) contained 36,864 independent clones in a 6x6 pattern and following a 384-well microplate organization. The membranes were hybridized with pools of probes of the selected list of genes of interest.

1.1.1.1 Selection of candidate genes for cork formation and cork quality

The UniCork database (Paiva et al. unpublished) was used for searching candidate genes associated with cork formation and cork quality. This database contains about 70,000 unigenes, including the 6,500 ESTs publicly available at GenBank. These 6,500 ESTs were obtained by sequencing both ends of the inserts of 5,000 cDNA clones from a non-normalized composite phellem cDNA library constructed by Paiva et al. (unpublished). BlastN program from NCBI (www.ncbi.nlm.nih.gov/) was used for searching the cork genes (Soler et al., 2007 and 2008) against the Unicork database.

In order to evaluate the size of the inserts of cDNA clones identified by Blast, we proceed with an amplification of the inserts by Colony PCR. The Colony PCR reaction was prepared in a volume of 50µl, containing 1µl of bacteria suspension, 0.2µM M13 (-21) universal forward primer (5' TGTAAAACGACGGCCAGT 3'), 0.2µM M13 reverse sequencing primer (5' TTGTCGATACTGGTAC 3'), 0.2mM of each dNTP, 0.8U GoTaq® Flexi DNA Polymerase (Promega®, Madison, WI, USA), 5mM MgCl2, 1X Green GoTaq® Flexi Buffer (Promega®, Madison, WI, USA) and sterile water to complete 50µL. The PCR plate was placed on a C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the following cycling conditions: 1 cycle at 95°C for 5 min and 34 cycles at 95°C for 30 sec, 60°C for 15 sec, 72°C for 1 min and

1 cycle at 72°C for 5 min. To evaluate the Colony PCR products and their size, it was done an electrophoresis on a 2% denaturing agarose gel with 0.5X TBE buffer. Later, the clones of genes of interest that presented the highest insert size were selected for sequencing.

Before sending for sequencing, the Colony PCR products of the selected clones were purified with the QIAquick® PCR Purification Kit (Qiagen, Hilden, Germany). Purified PCR products concentration and purity was determined using the NanoDropTM 1000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA). Colony PCR products were sequenced by Sanger sequencing technology at STAB VIDA (Investigação e Serviços em Ciências Biológicas, Lda, Caparica, Portugal) with the forward primer T7 Fwd (5' TAATACGACTCACTATAGGG 3').

To check the PCR products sequence identity, we performed a BlastN of the sequences of PCR products against the non-redundant sequence database of GeneBank (http://www.ncbi.nlm.nih.gov/genbank/). From the list of the selected cDNA clones, we confirmed these clones contained the expected genes. From the total 26 cDNA clones sequenced, 16 were selected to be used as probes for membrane hybridization (*Table 2*). Additionally, 15 out of 16 of these candidate genes were selected for a study of gene expression by qPCR, described in Chapter 3. These 16 genes correspond to functional categories of the main metabolic pathways involved in cork formation, such as genes from acyl-lipids and phenylpropanoids pathways, genes encoding regulatory transcription factors and genes encoding stress-related proteins.

1.1.1.2 Probe construction of candidate genes for cork formation

For each of the 16 selected clones, a sterile toothpick was dipped into the corresponding well in the cDNA library plate to inoculate 1ml of LB agar medium with 30µg/ml of chloramphenicol, on a sterile eppendorf tube (2ml). The medium was solidified on a tilted position and the plugs were incubated overnight at 37°C. Later, the probes were sent to INRA-CNRGV for membrane hybridization.

1.1.1.3 Probe construction for chloroplastidial and mitochondrial DNA content estimation

The BAC library was constructed from nuclear DNA, although it is possible to present contaminations from genomes of other organelles. In order to verify if the BAC library had extra-nuclear genome contamination and to evaluate the contamination degree of the BAC library, were constructed chloroplastidial and mitochondrial probes at our laboratory.

The probes were obtained by amplification of *Q. suber* DNA using specific *E. globulus* chloroplastidial primers psbA, ndhA (Paiva et al., 2011a) and mitochondrial primers cox3 and ccb256 (Duminil et al., 2002). The PCR reaction was prepared in 20 µl, containing 20 ng of genomic DNA, 0.5µM of forward primer, 0.5µM of reverse primer, 2U GoTaq® Flexi DNA GoTaq® Flexi DNA Polymerase (Promega®, Madison, WI, USA), 1x Reaction Buffer [2x GoTaq® Flexi Buffer (Promega®, Madison, WI, USA), 2 mM $MgCl_2$ and 0.25 mM dNTP mixture] and sterile water to complete 20µL. The PCR plate was placed on a

C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the following cycling conditions: 1 cycle at 95°C for 5 min and 34 cycles at 95°C for 30 sec, 50°C for 15 sec (mitochondrial primers) and 55°C for 15 sec (chloroplastidial primers), 72°C for 1 min and 1 cycle at 72°C for 5 min.

To evaluate the PCR products, it was done an electrophoresis on a 2% agarose gel with 0.5x TBE buffer. Later, the products of amplification from chloroplastidial and mitochondrial genes were purified in column with the QIAquick® PCR Purification Kit (Quiagen, Hilden, Germany). After purifying the PCR products, the DNA concentration and purity was determined using the NanoDrop™ 1000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA). After, the PCR products were sequenced by Sanger sequencing technology at STAB VIDA (Investigação e Serviços em Ciências Biológicas, Lda, Caparica, Portugal) with the forward primer T7 Fwd (5' TAATACGACTCACTATAGGG 3').

To confirm the PCR products sequence identity, was performed a BlastN of the sequences of PCR products against the non-redundant sequence database of GeneBank (http://www.ncbi.nlm.nih.gov/genbank/). Purified PCR products were sent to INRA-CNRGV for membrane hybridization.

1.1.2 Membrane hybridizations

After constructing the cDNA probes, the hybridization of the high-density filters was performed at INRA-CNRGV. Each of the four membranes produced were hybridized with a pool of probes from the selected 16 genes for cork formation. In addition, to evaluate the chloroplastidial and mitochondrial DNA content of the BAC library, one membrane was hybridized with a pool of two chloroplastidial probes (psbA, ndhA) and other membrane hybridized with a pool of mitochondrial probes (cox3, ccb256). After hybridization, images of the membranes generated by a PhosphoImager were processed by an automatic analysis program (High Density Filter Reader). This program generates a list of coordinates for the positive clones identified and the corresponding signal intensity.

1.1.3 Validation of the positive clones by BAC colony PCR

1.1.3.1 Primer design for colony PCR

After the screening for genes of interest and membrane hybridization, the resulting positive clones were validated by Colony PCR at our laboratory. This primers were also used for a gene expression analysis by qPCR, described in chapter 3. To amplify the selected genes by Colony PCR was necessary to design specific primers. For this study, besides the specific new primer design some of the primer sequences were also obtained from the work of Soler et al. (2008). New primers pairs were designed with *Primer-BLAST* from NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) with the following parameters:

- PCR product size varying from 100 to 150 nt.
- Product position with the primer located near the 3' end of the gene to increase the gene specificity.

- The GC content between 45 and 60%.
- Primer Melting Temperatures (Tm) with the minimum temperature of 57°C, the maximum of 62°C, an optimum Tm of 60°C and a Maximum Tm difference of 3°C.

1.1.3.2 Preparation of BAC colony PCR reaction and cycle thermal parameters

The BAC colony PCR reaction was prepared in a volume of 20µl, containing 1µl of bacteria suspension, 0.1µM of the respective forward primer, 0.1µM of the respective reverse sequencing primer, 0.2mM of each dNTP, 0.8U GoTaq® Flexi DNA Polymerase (Promega®, Madison, WI, USA), 5mM $MgCl_2$, 1x Green GoTaq® Flexi Buffer (Promega®, Madison, WI, USA) and sterile water to complete 20µL. The PCR plate was placed on a C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the following cycling conditions: 1 cycle at 95°C for 5 min and 34 cycles at 95°C for 30 sec, 60°C for 15 sec, 72°C for 1 min and 1 cycle at 72°C for 5 min. To evaluate the Colony PCR products and their size, it was done an electrophoresis on a 2% agarose gel with 0.5x TBE buffer.

1.1.4 BAC end sequences analysis

Twenty four BAC clones were used for BES analysis. The selected BAC clones contained candidate genes involved in both suberin biosynthesis pathways and in regulatory processes during cork formation: HCBT and Cyp 84A1 involved in the Phenylpropanoids pathway; GDSL and Cyp 86A1 involved in the Acyl lipids pathway and the gene WRKY a transcription factor with a possible regulatory function (see *Table 16*, *Appendix 1*).

The cork oak BESs were first analyzed and trimmed with BioEdit Sequence Alignment Editor software version 7.1.3.0. The sequences were then checked for the presence of vector contamination and the segments of vector origin were removed from the BAC end sequences with VecScreen (NCBI).

1.1.4.1 Identification of known repeats

The sequences were screened and masked for repeat elements and low complexity DNA sequences with RepeatMasker version open 3.3.0. The repeats were identified by searches for similarity to sequences in the Viridiplantae section of the RepBase database (version 20-09-2011), the most commonly used database of repetitive DNA elements. Additionally, to investigate the presence and diversity of mobile genetic elements (MGEs) in our BESs it was used the Gypsy Database of Mobile Genetic Elements (GyDB 2.0) with an E-value threshold of $1\ E^{-05}$.

1.1.4.2 Mining of simple sequence repeats (SSRs)

Simple sequence repeats were mined from the BESs using the SciRoKo (3.4) program. SciRoKo is a program that allows SSRs search, investigation and statistics on a group of sequences until an entire genome.

1.1.4.3 Homology search

In order to identify the redundancy between the sequences, the cork oak BESs were compared with themselves using BLASTn and Megablast. Then it was performed homology search with BLASTn against the NCBI GenBank non-redundant nucleotide collection (nr/nt) database; the genomic survey sequences

(gss) database of the *Quercus* (Taxid: 3511) organisms and against the expressed sequence tags (est) database of the *Fagaceae* (Taxid: 3503) organisms. Further homology search was done with BLASTx against the NCBI GenBank non-redundant proteinprotein sequences (nr) and against the nr database with the organisms *Fagaceae* (Taxid: 3503); *Quercus* (Taxid: 3511) and *Quercus suber* (Taxid: 58331). In both BLASTn and BLASTx analysis, similarities to known sequences were considered significant when the expected values (E- value) were lower than 1 E^{-05}.

1.2 Results

1.2.1 BAC library screening for genes of interest

After the construction and characterization of the BAC library, an additional objective of this work was to screen the library for the presence of genes involved in cork formation and quality.

1.2.1.1 Selection of the candidate genes for cork formation and cork quality

After selecting the cDNA clones, from a phellem cDNA library (Paiva et al. unpublished), 250 clones containing genes of interest for cork formation were amplified by Colony PCR and an example of the gel electrophoresis results are illustrated in

*Figure **13***.

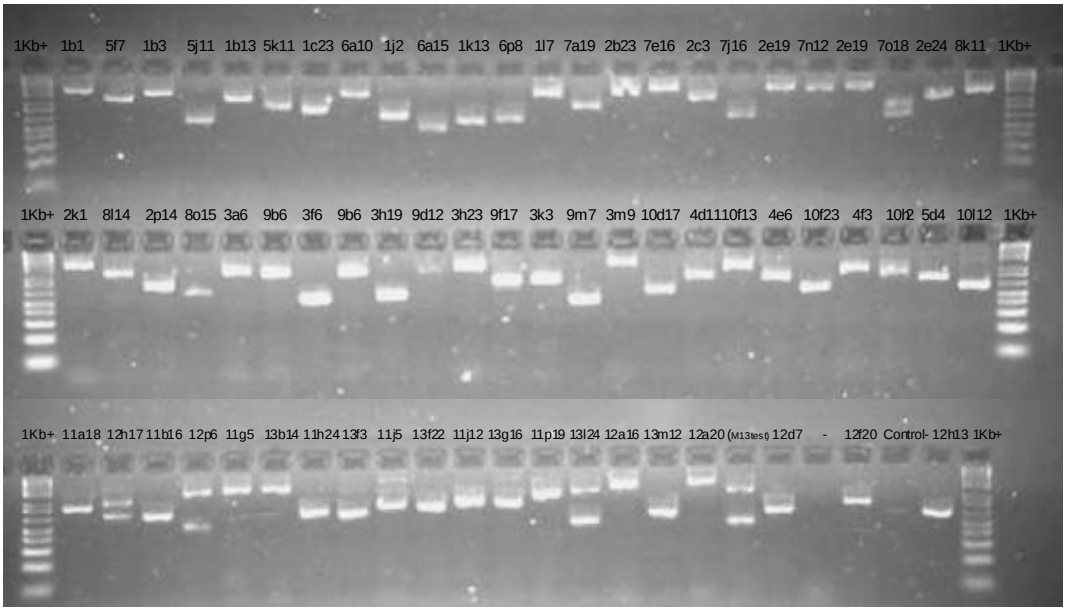


Figure 13. Evaluation of the quality of the Colony PCRs amplification products on a 2% agarose gel in 0.5x TBE and using 1Kb+ ladder (Invitrogen, Grand Island, NY, USA).

The amplification products correspond to the amplification of the inserts of the selected cDNA clones plus the product amplified from the vector's polylinker. After amplification by Colony PCR, the number and size of the fragments obtained were reported. On average the majority of the inserts presented only one fragment but some inserts such as that from the QS-12H17 cDNA clone presented two fragments. The inserts showed different sizes ranging from 500 to 2,000 base pairs. cDNA clones presenting the highest

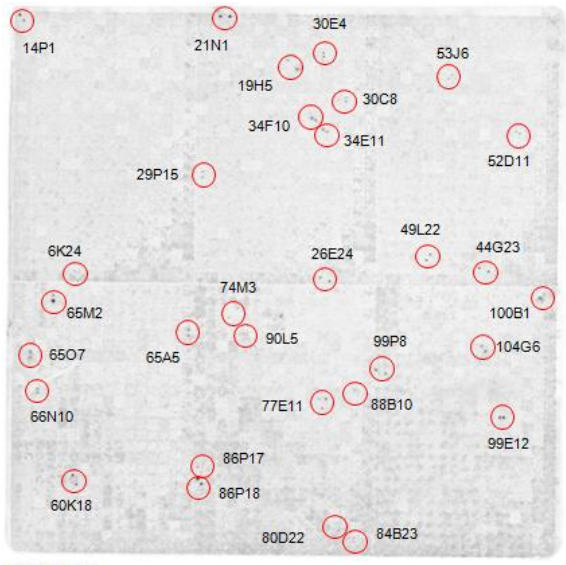
insert size and and unique fragment amplification product were selected for sequencing validation. The sequence identity of 26 cDNA and chloroplast and mitochondrial PCR products were confirmed, and 16 cDNA clones were selected for containing genes relevant for cork formation (
Table 2).

Table 2. List of the selected genes and respective cDNA clones used as probes for membrane hybridization

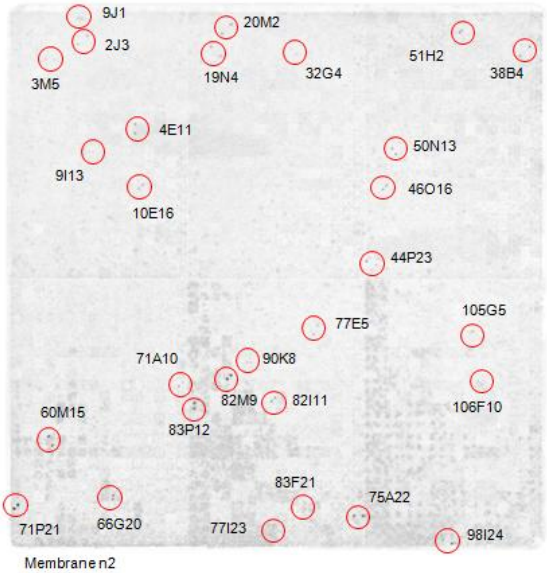
Gene functional categories	Gene name	cDNA clone	Description	SUBERGENE database
Acyl lipids				
Fatty acid synthesis	DHCA-S-AT	QS - 3A4	Dihydrolipoamide S-acetyltransferase	GGFEK7Q01CYTRE.l.qs.1
Fatty acid elongation	LACS	QS - 1E1	Long-chain acyl CoA synthase	GGFEK7Q01A06VB.l.qs.1
Lipid metabolism	ELTFP	QS - 3B6	Esterase/lipase/thiosterase family protein	GGFEK7Q01A59SZ.l.qs.1
Fatty acid hydroxylation	Cyp86A1	QS - 3C8	Fatty acid W-hydroxylase	GGFEK7Q01A0IF3.l.qs.1
Fatty acid esterification	GPAT	QS - 2D8	Glycerol-3-phosphate acyltransferase	GGFEK7Q01A03W4.l.qs.1
Lipid metabolism	GDSL	QS - 3E8	GDSL-motif lipase/hidrolase protein like	GGFEK7Q01A0ZV8.l.qs.1
Phenylpropanoids				
Phenylpropanoid pathway	CCR	QS - 3B10	Cinnamoyl CoA reductase	GGFEK7Q01A36GL.l.qs.1
	PAL	QS - 3E5	Phenylalanine ammonia-lyase	qscs0011.j05_3.1. l.qs.1
	4CL	QS - 3E4	4-Coumarate: CoA ligase	GGFEK7Q01A407A.l.qs.1
Phenylpropanoid derivatives	Cyp84A1 (F5H)	QS - 3D9	Ferulate-5-hydroxylase	GGFEK7Q01A0U9Z.l.qs.1
Acyltransferase	HCBT	QS - 1A1	N-hydroxycinnamoyl-CoA/benzoyltransferase	GGFEK7Q01A01QL.l.qs.1
Cross linking/oxidase	DoLaccase	QS - 3B1	Diphenol oxidase laccase	GGFEK7Q01A0R5F.l.qs.1
Regulatory proteins				
Transcritpion factor	WRKY	QS - 3B5	WRKY transcription factor	GGFEK7Q01A0IGO.l.qs.1
Ethylene forming	ACC-OX	QS - 3B3	1- aminocyclopropane -1 - carboxylic acid oxidase	GGFEK7Q01A65Y0.l.qs.1
Miscellaneous				
Monoxygenase	Cyp87A2	QS - 1E12	Cytochrome P450 family protein	GGFEK7Q01A0FQG.l.qs.1
Stress				
Reactive oxygen species scavenging	APX	QS - 3B11	Ascorbate peroxidase	GGFEK7Q01A2DPE.l.qs.1

1.2.1.2 BAC Library screening for genes of interest

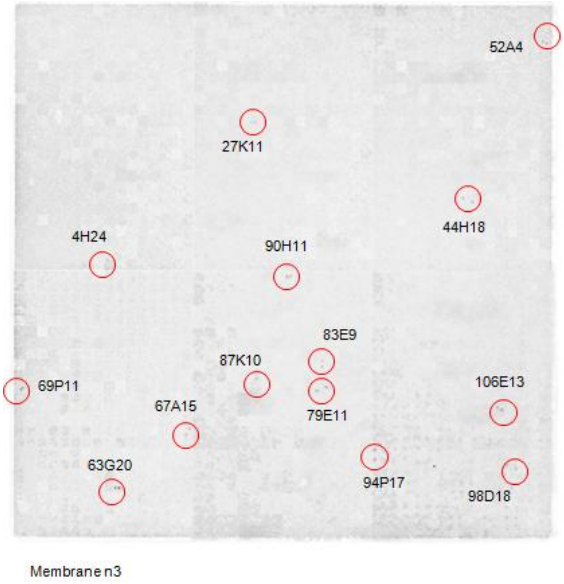
We selected and produced 16 cDNA probes of candidate genes for cork formation. Four membranes produced at INRA-CNRGV, were hybridized with pools of this probes and later screened for the presence of positive clones. After hybridization of the BAC high-density filters and image analysis, 88 positive clones were identified and the results are shown in *Figure 14*.



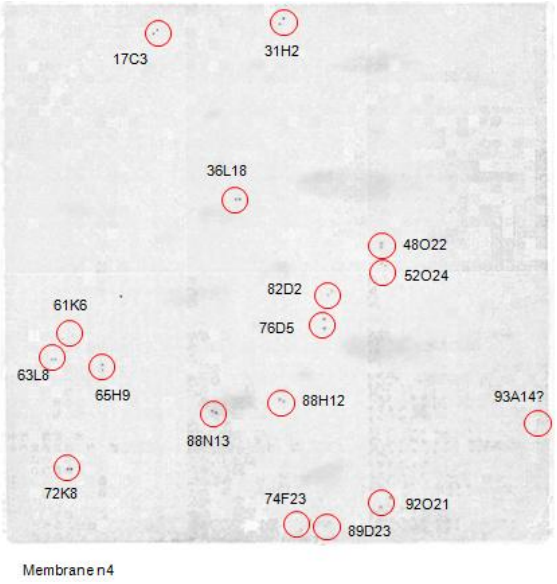
(a) Membrane 1 hybridized with the probes Dolaccase, ACC-OX, LACS, GDSL and CYP87A2.



(b) Membrane 2 hybridized with the probes DHPA-S-AT, ELTFP, 4CL, CCR and PAL.



(c) Membrane 3 hybridized with the probes HCBT, CYP86A1 and GPAT.



(d) Membrane 4 hybridized with the probes CYP84A1, AXP and WRKY.

Figure 14. Panels (a), (b), (c) and (d) correspond to membranes hybridized with different pools of probes, of the selected 16 genes of interest. ◯ Indicates the position of positive clones with the respective coordinates on the membrane (eg: 27K11).

1.2.1.3 Validation of positive clones by BAC colony PCR

Hybridization positive clones were validated by BAC colony PCR. Primers used for validation are presented in *Table 3*. Ten primers were developed in this study whereas primers for HCBT, GPAT, APX, Cyp86A1, WRKY, and Cyp84A1 genes were those designed by Soler et al. (2008).

Table 3. List of the genes selected and their respective primers designed for validation of positive clones by Colony PCR. Primers were designed with the software *Primer-BLAST* from NCBI, excepting the primers Cyp86A1, Cyp84A1, GPAT, HCBT, WRKY and APX were developed by Soler et al. (2008).

Primers name	Soler at al. 2008 annotation	Primer Forward Sequence (5'->3')	Primer Reverse Sequence (5'->3')	Reference
HCBT	HCBT	TGGCATTGGTGCTATGGAGTT	GGGTTGCGGGCCTTAAGTA	Soler et al. 2008
GPAT	GPAT	TTGACTCGGAGAAAATCAAGCA	GCTACCGGCACAATTCGATC	Soler et al. 2008
LACs	LACS, putative	GCCCACGCGAATGGCTGCATC	AGAGGAACCCCAGCAACCCAATGC	This study
Cyp87A2	Cytochrome P450 family protein, Cyp87A2	GCTGGCTAGTGTTGCTCCGGG	AGGAACCACCAAAAGACACCAGCCT	This study
4CL	4CL	TCGATGCGATTCCAAGTCGC	ATCAGGGCTTGTTGCAATACGTGT	This study
APX	Ascorbate peroxidase, putative	GGTCCAAAGCACACAAAAATCC	GGCTGCTGTCTGGTGTCTAGTCT	Soler et al. 2008
Cyp86A1	FA -hydroxylase, Cyp86A1	CCACGTGGACTTACAGGATTTG	ACAGTGTTTCCGGGTCTTTACC	Soler et al. 2008
Dolaccase	Diphenol oxidase laccase	ACACCAACAAATTCACAAGCCCAGT	AACTGGGTCCCCAAATGCAGGG	This study
ELTFP	Esterase/lipase/thioesterase family protein	CGACGGTTGTGTGCTTTTCAAGCCA	GGGCACTGTATGAGCGAGCTGG	This study
WRKY	WRKY transcription factor	CGGCCTAGGTTTGCATTTCA	TGCCCATACTTTCTCCATCGA	Soler et al. 2008
ACC-OX	ACC oxidase, putative	TTGCACCGTGTCATGGCTAACG	TGGGATATAGAGGTTGGGGAGCTGG	This study
Cyp84A1	F5H, Cyp84A1, putative	CAAGTGTTTGGACTGAGCCAGAT	CGGGATAAGCTCAAAGTCATGTC	Soler et al. 2008
CCR	Cinnamoyl CoA reductase	TCAAGAGCTTGCAGGAGAAGGGTGT	ACGAACACAGCCCATGACGCTT	This study
GDSL	GDSL-motif lipase/hydrolase, putative	ACAGATTGACAGCTTGTTGGGTCCA	CTTGCCCCAATGGAGAGGACTGGTA	This study
DHCA-S-AT	Dihydrolipoamide *S*-acetyltransferase, putative	AGGCACTCTGTCAAGTAGACACACC	AGAACCCCGAAAGCCTGACATTGT	This study
PAL	PAL	GTGGGTACGACCTGCACTCCTT	CATATGCCCGGGAATTCGGCCA	This study

Based on Colony PCR analysis we were able to validate 61 clones out of a total of 88 hybridization positive clones distributed along the four membranes, the results are summarized in *Table 4*. From the 16 genes analysed, the genes LACS and GPAT did not present any amplification by Colony PCR.

Our results show that in membrane #1, hybridized with probes Dolaccase, ACC-OX, LACS, GDSL and Cyp87A2, all positive clones (31) were validated except one clone (QS-H-HDL_06K24) containing the LACS gene. Membrane #2 was hybridized with the probes DHCA-S-AT, ELTFP, 4CL, CCR and PAL and all genes were validated in 18 clones out of 29 positive clones. Contrary to the observed in validated clones from the other three membranes, four clones of membrane #2 presented amplification with two different primers: clones QS-H-HDL_03M05 and QS-H-HDL_32G04 with primers DHCA-S-AT and 4CL; clones HDL44P23 and QS-H-HDL_98I24 with PAL and ELTFP. Membrane #3 was hybridized with the probes HCBT, Cyp86A1 and GPAT and were validated all positive clones (14) excepting two (QS-H-HDL_4H24, QS-H-HDL_69P11) that corresponded to the gene GPAT. Similarly to the observed in membrane #2, in membrane #4 hybridized with probes Cyp84A1, AXP and WRKY, we validated positive clones containing the 3 genes. Although, in membrane #4 were validated half of the positive clones, while on the other three membranes almost the totality of the positive clones were validated.

Table 4. Results for the positive clones validation by Colony PCR. For each membrane were validated all the positive clones and the corresponding primers for which they presented amplification. It is also indicated the group of cDNA probes hybridized to each membrane

Membrane 1 **Probes QS-3B1; 3B3; 3E8; 3E12**		**Membrane 2** **QS-3A4; 3B6; 3E4; 3B10; 3E5**		**Membrane 3** **QS-1A1; 3C8; 2D8**		**Membrane 4** **QS-3D9; 3B11; 3B5**	
Positive Clones	**Gene amplification**	**Positive Clones**	**Gene amplification**	**Positive Clones**	**Gene amplification**	**Positive Clones**	**Gene amplification**
06K24	-	**02J03**	DHCA-S-AT	**4H24**	-	**17C03**	AXP
14P01	Do Laccase	**03M05**	DHCA-S-AT, 4CL	**27K11**	HCBT	**31H02**	Cyp84A1
19H05	Do Laccase	**04E11**	-	**44H18**	HCBT	**36L18**	-
21N01	GDSL	**09I13**	4CL	**52A04**	HCBT	**48O22**	-
26E24	Do Laccase	**09J01**	4CL	**63G20**	Cyp86A1	**52O24**	WRKY
29P15	Cyp87A2	**10E16**	4CL	**67A15**	HCBT	**63L08**	-
30C08	Cyp87A2	**19N04**	4CL	**69P11**	-	**65H09**	-
30E04	GDSL	**20M02**	-	**79E11**	HCBT	**72K18**	-
34E11	Do Laccase	**32G04**	DHCA-S-AT, 4CL	**83E09**	Cyp86A1	**76D05**	AXP
34F10	Do Laccase	**38B04**	-	**87K10**	Cyp86A1	**82D02**	WRKY
44G23	Do Laccase	**44P23**	PAL, ELTFP	**90H01**	HCBT	**88H12**	AXP
49L22	ACC OX	**46O16**	CCR	**94P17**	HCBT	**88N13**	-
52D11	GDSL	**50N13**	CCR	**98D18**	HCBT	**92O21**	WRKY
53J06	Cyp87A2	**51H02**	-	**106E13**	HCBT	**93A14**	-
60K18	ACC OX	**60M15**	-	-	-	-	-
65A05	Cyp87A2	**66G20**	4CL	-	-	-	-
65M02	GDSL	**71A10**	-	-	-	-	-
65O07	Do Laccase	**71P21**	DHCA-S-AT	-	-	-	-
66N10	Cyp87A2	**75A22**	-	-	-	-	-
74M03	Do Laccase	**77E05**	CCR	-	-	-	-
77E11	GDSL	**77I23**	-	-	-	-	-
80D22	Do Laccase	**82I11**	DHCA-S-AT	-	-	-	-
84B23	Do Laccase	**82M09**	DHCA-S-AT	-	-	-	-
86P17	GDSL	**83F21**	-	-	-	-	-
86P18	GDSL	**83P12**	DHCA-S-AT	-	-	-	-
88B10	GDSL	**90K08**	ELTFP	-	-	-	-
90L05	ACC OX	**98I24**	PAL, ELTFP	-	-	-	-
98P08	ACC OX	**105G05**	-	-	-	-	-
99E12	Do Laccase	**106F10**	-	-	-	-	-
100B01	GDSL	-	-	-	-	-	-
104G06	GDSL	-	-	-	-	-	-

1.2.2 BES analysis results

1.2.2.1 Identification of known repeats

The presence of repetitive DNA in cork oak BESs was evaluated with RepeatMasker, in a total of 48 BAC end sequences corresponding to forward and reverse sequences of 24 BAC clones. The total length of the sequences was 24,211 bp, ranged from 335 to 743 bp, presented an average length of 504 bp and a GC content of 37.76% (*Table 5*). The GC content percentage of each sequence is presented in *Table 17*, *Appendix 2.* In addition**,** 2,001 bp annotated repeats were masked, which corresponded to 8.26% of the query sequences.

Table 5. Summary of RepeatMasker analysis.

Sequences	48
Total length	24,211 bp (24211 bp excl N/X-runs)
GC level	37.76 %
Bases masked	2001 bp (8.26 %)

Comparisons of the BESs to the plant repeat database, performed by RepeatMasker, revealed that 5.61% of the nucleotides in the cork oak BESs presented homologies to known plant repeat elements (*Table 6*). Class I retroelements (retotransposons) represent the most abundant repeats with a total of four retroelements (1,262 bp) corresponding to 5.21% of the BESs. BAC end sequences homologous to retrotransposons includeinclude LINEs (Long Interspersed Nuclear Elements) (2.04%) and LTR (Long Terminal Repeats) elements (3.18%). In the LTR elements group were found *Ty1/Copia* elements which were the most abundant retroelements (2.26%) and *Ty3 Gypsy/DIRS*1 (0.91%). Class II DNA transposable elements or transposons were the repeats with lower representation on the cork oak BESs (0.4%). The BAC end sequences containing the repeat elements are presented in *Table 7*. The Class II element, a DNA/MULE-MuDR, repeat was identified in sequence QS-H-HDL_98D18xT7. This transposon belongs to the MuDR family which is included in the Mutator superfamily, and whose elements are also commonly known as Mutator-like elements (MULEs).

Table 6. Classification and distribution of known plant repeat DNA elements in the cork oak BAC end sequences identified by RepeatMasker.

Class	Number of elements	Length (bp)	% of sequence
Retroelements	4	1262	5.21
LINEs:	2	493	2.04
L1/CIN4	2	493	2.04
LTR elements:	2	769	3.18
Ty1/Copia	1	548	2.26
Gypsy/DIRS1	1	221	0.91
DNA transposons	1	97	0.4
Total interspersed repeats	-	1359	5.61
Simples repeats	9	254	1.05
Low complexity	8	388	1.60

Table 7. List of the cork oak BAC end sequences containing known plant repeat elements by Sciroko.

BAC end sequence	Repeat	Repeat class	BES size (bp)	Begin	End	Total length	Score
QsuHDLx21N01xT7	Gypsy-192_ZM-I-int	LTR/Gypsy	668	38	258	221	189
QsuHDLx27K11xM13rev	Copia-79_Mad-I-int	LTR/Copia	584	3	550	548	701
QsuHDLx52O24xM13rev	LINE1-24_Sbi	LINE/L1	671	369	533	165	286
QsuHDLx82D02xM13rev	SHALINE10_MT	LINE/L1	561	6	333	328	562
QsuHDLx98D18xT7	MERMITE18C	DNA/MULE-MuDR	400	147	243	97	256

Further investigation of the presence and diversity of mobile genetic elements (MGEs) in the cork oak BESs, was performed using BLASTn and BLASTx against the Gypsy Database 2.0 (GyDB). This allowed searching for homologies of the BAC end sequences to the GENOMES, LTRs and CORES databases. A summary of the BLASTn and BLASTx results is shown in *Table 18Table 19* (*Appendix 3*), respectively.

From the BLASTn analysis against the GENOMES database were identified ten BAC end sequences which presented significant alignments with other sequences. Six BESs showed significant high similarities with GINO1, GINO2 and Ginger1 transposon sequences. One BAC end sequence had significant homology with SCAND3, a cellular integrase (C-INT). In addition, two BAC end sequences presented significant high identity with *Tdd-4* and *Tdd-5* DNA transposons sequences, SCAND3 and Nomad (*Ty3/Gypsy* LTR retroelement).

BLASTx analysis revealed that only two BAC end sequences presented significant alignments. The sequence QS-B-HDL_27K11xM13 showed significant high similaritiy with *Ty1/Copia* LTR retrotransposons such as: Melmoth, Retrofit, Poco, among others. The other BESBES, QS-H-HDL_21N01xT7 presented significant high similarity with sequences from the *Ty3/Gypsy* LTR retrotransposons plant lineage (*Tat*) of the cluster *Tat/Athila* like: RetroSor1, Tat4-1 and Cinful-1.

1.2.2.2 Mining simple sequence repeats (SSRs)

The repeat search and analysis of 48 cork oak BESs (from the forward and reverse reads), revealed a total of 14 different simple sequence repeats (SSRs) with a mean length of 26 nucleotides (*Table 8*). The frequency of SSRs in the cork oak BES corresponded to one SSR per 1.73 kb or 57.82 SSRs per 100 kb.

Table 8. SSRs analysis statistics of cork oak BESs by Sciroko

SSRs in Subset	14
Average mismatches	0,21
Average length	26
Average Standard deviation length	7,66
counts/Mbp	591,62

In the cork oak BESs, were detected SSRs with motif lengths of mono-, di-, penta- and hexanucleotide tandem repeats (*Table 9*). The most abundant class of microsatellites was the dinocluneotide motifs found in eight sequences (57.14%) with poly(AT), poly(AC) and poly(AG) identified in four (50%), three (37.5) and one (12.5%) sequences, respectively. In addition, the longest SSR identified was the poly(AG) dinucleotide with 36 bases. The second most abundant microsatellites were the mononucleotides, poly(A), identified in 4 BESs (28.57%). The penta- and hexanucleotide motifs were located in two distinct BAC end sequences (7.14%).

Since GC motifs were not identified, AT-rich SSRs were the most abundant motifs found in the *Q. suber* BESs. The 14 SSRs were found in 11 different BAC end sequences (*Table 10*). The sequence QS-H-

HDL_104G06xT7 contained two SSRs of different motifs, di- and hexanucleotide repeats. In addition, in QS-H-HDL_98D18xT7 sequence was found three SSRs, two mononucleotide repeats and one dinucleotide.

Table 9. SSRs identified in *Quercus suber* BESs by Sciroko.

Motif		Counts	Average length	Average mismatches	Counts/Mbp
Mononucleotide:	Poly(A)	4	21,5	0	169,03
Dinucleotide:		8	29,25	0,38	338,07
	Poly(AT)	4	22,75	0	169,03
	Poly(AC)	3	35,67	0	126,77
	Poly(AG)	1	36	3	42,26
Pentanucleotide:	Poly(AAAAG)	1	26	0	42,26
Hexanucleotide:	Poly(AAAAAG)	1	18	0	42,26

Table 10. Identification of simple sequence repeats (SSRs) in BAC-end sequence (BES) of the *Quercus suber* BAC library by SciRoko.

BAC end sequence	Motif	Motif standardized	Repeat class	SSR start	SSR end	Total length	Score	Mismatches
QsuHDLx100B01xM13rev	AT	AT	Simple repeat	337	357	21	21	0
QsuHDLx104G06xT7	CTTTTT	AAAAAG	Low complexity	338	355	18	18	0
	AT	AT	Simple repeat	358	382	25	25	0
QsuHDLx21N01xM13rev	TG	AC	Simple repeat	317	355	39	39	0
QsuHDLx52D11xT7	TG	AC	Simple repeat	317	355	39	39	0
QsuHDLx52O24xT7	AC	AC	Simple repeat	163	191	29	29	0
QsuHDLx79E11xM13rev	A	A	Low complexity	325	340	16	16	0
QsuHDLx79E11xT7	AT	AT	Simple repeat	162	187	26	26	0
QsuHDLx82D02xT7	T	A	Low complexity	124	141	18	18	0
QsuHDLx87K10xM13rev	TA	AT	Simple repeat	206	224	19	19	0
QsuHDLx90H01xM13rev	TCTTT	AAAAG	Simple repeat	328	353	26	26	0
QsuHDLx98D18xT7	A	A	Simple repeat	91	110	20	20	0
	A	A	Simple repeat	129	160	32	32	0
	TC	AG	Simple repeat	208	243	36	19	3

1.2.2.3 Homology search

BLASTn searches against entries in the NCBI GenBank non-redundant nucleotide collection (nr/nt) database showed that three sequences (QS-H-HDL_52D11xT7; QS-H-HDL_44H18xT7; QS-H-HDL_30E04xT7) had significant homologies with complete sequences of different clones from *Populus trichocarpa*, formerly published in GenBank (AC212845.1, AC213416.1, AC212845.1) and one BES (QS-H-HDL_27K11xM13)with a complete sequence from a *Medicago truncatula* clone (AC151462.30), see *Table 20* (*Appendix 4*). Two BAC end sequences presented significant homologies with a predicted protein, mRNA from *P. trichocarpa* (XM_002298612.1, XM_002318461.1) and one BES had significant high identity to an uncharacterized mRNA from *Vitis vinifera* (XM_002272559.1). Two cork oak BESs (QS-H-HDL_98D18xT7; QS-H-HDL_79E11xT7) had significant similarity with genome shotgun sequences from *V. vinifera* contigs (AM459017.2, AM471831.2). One sequence (QS-H-HDL_27K11xT7) presented significant high homology with a supposed receptor protein kinase ZmPK1-like mRNA from *Glycine max* (XM_003523930.1) and another (QS-H-HDL_21N01xM13) one with a contig from a *Triticum aestivum* chromosome 3B-specific BAC library (FN564429.1). The sequences QS-H-HDL_82D02xM13 and QS-H-HDL_52O24xM13 showed

significant high identity with LINE-type retrotransposons from *Beta vulgaris* (FR852835.1, FR852855.1) and the sequence QS-H-HDL_52O24xT7 had significant high homology with an SSR described in a clone of *Pinus pinaster* (CR377470.1). Finally, the sequence QS-H-HDL_63G20xM13 presented significant identity with mRNA of GDSL esterase/lipase EXL1-like from *Glycine max* (XM_003544976.1). Results from the nucleotide searches against the GSS database of the *Quercus* organism showed that 23 of the cork oak BESs presented significant high homologies with BAC end sequences of *Q. robur* (*Table 21*, *Appendix 4*). The nucleotide homology search with all the BESs against the expressed sequence tags (EST) database of the *Fagaceae* organisms showed that 15 BAC end sequences presented significant high similarities with mRNA sequences from cDNA clones of different species: nine BESs had significant homologies with *Q. robur*, three with *Fagus sylvatica*, two with *Quercus mongolica*, one with *Castanopsis sieboldii* and one BES showed significant high identity with a mRNA sequence from *Q. suber* cDNA clone (*Table 22*, *Appendix 4*).

BLASTx searches against the NCBI GenBank non-redundant protein sequences (nr) database revealed that fourteen BESs presented significant homologies with nucleotide sequences and the E-values ranged from $6E^{-73}$ to $1E^{-06}$ (*Table 23*, *Appendix 4*). Significant homologies were identified with uncharacterized or hypothetical proteins from *Glycine max*, *Ricinus communis*, *Medicago truncatula* and *Vitis vinifera*. It was also found significant identities with an oligopeptidase from *R. communis* (QS-H-HDL_52D11xT7) and with a non-LTR retroelement reverse transcriptase from *Arabidopsis* thaliana (QS-H-HDL_87K10xT7). Similarly to BLASTn search results, the sequence QS-H-HDL_63G20xM13 showed identity with a GDSL esterase/lipase *V. vinifera*. In addition, the sequence QS-H-HDL_2O24xM13 presented significant homologies with a hypothetical protein from *B. vulgaris* subsp. *vulgaris* and with a retrotransposon protein from *Oryza sativa*, similarly to BLASTn results that revealed high homology of this sequence with a LINE-type retrotransposons from *B. vulgaris* subsp. *vulgaris.* The sequence QS-H-HDL_27K11xT7 showed significant homologies with a serine-threonine protein kinase from *R. communis* and with BLASTx search for nr database of the *Fagaceae* organisms, it presented identity with a S-receptor kinase (SRK) from *Fagus sylvativa*. A similar picture was found for the sequence QS-H-HDL_82D02xM13, which presented in the nr database search significant homology with a retrotransposon protein (*O. sativa*) and in the nr database of the *Fagaceae* organisms, significant identity with a reverse transcriptase from *F. sylvativa*. The sequence QS-H-HDL_21N01xT7 showed significant homology with a GagPol3 protein (*A. thaliana*) and with a gag protein from *Quercus suber* (EMBL AAQ94320.1). A summary of all the homology searches performed in this work is presented in *Table 24* (*Appendix 5*).

1.3 *Discussion*

As part of the characterization of the first *Quercus suber* high-quality, deep-coverage and large-insert *Hind*III BAC library (Qsu-B-HDL), constructed at CNRGV-INRA laboratory, was screened for the present of contamination with extra-nuclear genomes. The *Q. suber* BAC library presented a low level of total contamination with mitochondrial and chloroplastidial genomes.

The *Q. suber* BAC library was then screened for the presence of genes involved in cork formation using a combined methodology of high density BAC filters hybridizations and PCR validation of the hybridization positive clones. High density filters are important tools for an exhaustive structural and functional study of genomes as they allow the identification of sequences of interest among thousands of clones. The membranes were hybridized with a series of cDNA pooled probes corresponding to a selection of 16 candidate genes related with suberin biosynthesis pathway, stress related genes and transcription factors. The selection of the candidate genes was based on their relevance for cork formation (Soler et al., 2007; 2008).

From membrane hybridization were identified 88 positive clones. Based on the Colony PCR analyses, we were able to validate sucessfully 69% of positive clones that included 14 out of the 16 candidate genes in study. In our study, we could not validate positive clones containing the LACS and GPAT genes, thus further studies with a better primer design would be necessary to validate the positive clones harbouring these genes. Our results from the positive clones Colony PCR validation suggest an over-representation of positive clones containing the genes Dolaccase (11 clones) and HCBT (9 clones) from the phenylpropanoids pathway, and the gene GDSL (10 clones) gene involved in the acil lipids pathway.

In both positive clones QS-H-HDL_03M05 and QS-H-HDL_32G04 were validated two genes, DHCA-S-AT and 4CL. The same was observed for clones QS-H-HDL_44P23 and QS-H-HDL_98I24 that were validated with both PAL and ELTFP genes. The pairs of genes validated in the same BAC clone suggest that these genes are closely located in the cork oak genome.

With this study we validated the presence of the candidate genes related with suberin biosynthesis in the constructed *Q. suber* BAC library.

BAC end sequences analysis

Identification of known repeats

In our study, the cork oak BAC end sequences analyzed presented a GC with an estimated mean value of 37.76%. While in a study of Zoldos et al. (1998), where seven *Quercus* species including *Q. suber* were evaluated for DNA content, the mean base composition found in *Q. suber* was 39.7% GC. This GC content value was reported as being characteristic of many higher plants (Marie and Brown, 1993; Martel et al., 1997; Cerbah et al., 1998; Le Thierry d'Ennequin et al., 1998; Zoldos et al., 1998). Through flow cytometry measures, they observed that the GC percentage presented small variations between the species (38.8% - 40.4%) and that a GC content of 39.9% is typical for oak. However, in BAC end sequences studies it was found a GC content of 35.33% in *Quercus robur* (Rampart et al., 2011), 35% in papaya (Lai et al., 2006) and 36% in the complete genome of *Arabidopsis thaliana* (The Arabidopsis Genome Iniative), 41% the yellow poplar (Liang et al., 2007) and 35% in grapevine (Jaillon et al., 2007).

The eukaryotic genomes are composed in its majority (between 50 and 90% of all DNA) by repetitive sequences (Heslop-Harrison, 2000). In plants, the amounts of repetitive sequences have high impact on the variation of the genome size (Flavell et al., 1974) and therefore on the evolution of plant genomes. In this study, the percentage of interspersed repeats found in the total BESs raw sequence (5.61%) was similar to the percentage found in *Q. robur* (5.51%) (Rampant et al., 2011) but lower when compared to *Arabidopsis* (10%) (The Arabidopsis Genome Iniative'), *Carica papaya* (16%) (Lai et al., 2006) or to *Citrus clementina* (25%) (Terol et al., 2008). Transposable elements found in many other plant genomes are also represented in *Q. suber*, including long interspersal nuclear elements (LINEs) and copia- and gypsy-like long terminal repeat (LTR) retrotransposons.

Our results show that, Class I retrotransposons, including LINE and LTR (*gypsy*-like and *copia*-like elements) retroelements, were the predominant repeat sequences found in the BESs analyzed, similarly to what has been reported in *Q. robur*, *C. clementine*, carrot (Cavagnaro et al., 2009) and grapevine (Moisy et al., 2008). In particular, our results show that *Ty1/Copia* elements were the most abundant retrolements in cork oak BESs, similarly to what has previously been found in the genomes of *Q. robur* (Rampant et al., 2011), banana (Hribová et al., 2010), grapevine (Moisy et al., 2008) and carrot (Cavagnaro et al. 2009). This is in contrast with other studies reporting that *Gypsy* elements are the most abundant retroelements in *Arabidopsis*, poplar, rice, clementine (Terol et al., 2008) and *Quercus rotundifolia* (Alves et al., 2012). However, in the last study the authors explain that these results could be biased by the small number of sequences analyzed and the restriction enzymes selected for the AFLP methodology. Nevertheless, our observations are in agreement with other studies showing that Class I retrotransposons contribute for most of the nucleotide content in a variety of plants with large genomes (SanMiguel et al., 1996).

In this study we analyzed a small number of BESs. For that reason, the estimates of the amounts of repetitive sequences found in the cork oak BESs may not be generalized to the entire *Q. suber* genome. Additionally, the selection of BACs including certain genes can bias the analysis. The Gypsy Database homology search revealed to be very valuable in this study. Searches in Gypsy database revealed that some cork oak BESs have significant high similarities with GINO1, GINO2 and Ginger1 sequences which encode for a transposase (TR). It were also found significant homologies with a SCAND3 sequence which contains a domain described as being involved in transcriptional regulation of growth factors and genes involved in metabolism, cell survival and differentiation (Sander et al., 2003; Collins and Sander 2000). Furthermore, were identified significant high homologies with: Tdd-4 and Tdd-5 sequences, DNA transposons described in *Dictyostelium discoideum* genome (Wells, 1999; Glockner et al., 2001); Nomad which is a *Ty3/Gypsy* LTR retroelement described as a retrovirus in *Drosophila melanogaster* (Whalen and Grigliatti 1998).

The BLASTx homology search confirmed the results obtained with the RepeatMasker program, for two BESs sequences containing high similarities with *Ty1/Copia* and *Ty3/Gypsy* LTR retrotransposons. It was the case of the sequence QS-H-HDL_27K11xM13 with significant high homology with *Ty1/Copia* LTR retrotransposons reported in several plant species such as: Melmoth, originally described in Brassica

(Pastuglia et al., 1997) and also found in other plant organisms such as Arabidopsis (Llorens et al., 2009); Retrofit, found in a disease resistance gene family of rice (*Oryza longistaminata*) genome (Song et al., 1997); Poco (Populus Copia element), found in the genome of the "black cottonwood" (*Populus trichocarpa*) (Llorens et al., 2009) among others. And also the sequence QS-H-HDL_21N01xT7 which presented significant high identity with sequences from the *Ty3/Gypsy* LTR retrotransposons plant lineage (*Tat*) of the cluster *Athila/Tat*, from different species like: RetroSor1, described in *Sorghum bicolor* (Llaca et al., 1998); Tat4-1, characterized in *Arabidopsis thaliana* genome (Peleman et al., 1991) and Cinful-1, described in *Zea mays* genome (San Miguel et al., 1998). Similarly to our results, in *Q. rotundifolia* were identified sequences homologous with *gypsy*-like retrotransposons from the *Tat/Athila* clade, in particular the Tat4 subtype element (Alves et al., 2012) which is in accordance with the genomic conservation detected between these two genomes.

LTR-retrotransposons are the most abundant transposable elements found in plants and comprise a significant part of many plant genomes. Evidences show that genome expansion in plants is mainly due to the accumulation of LTR retrotransposons. The abundance and diversity of retrotransposons found in our work and others (Rampant et al., 2011; Alves et al., 2012) suggest that these elements played an important role in the evolution of *Quercus* genus.

Mining simple sequence repeats (SSRs)

The BAC end sequences have been a powerful source for identification of simple sequence repeats (SSRs or microsatellites) in a wide variety of plant species, such as *Q. robur* (Rampart et al., 2011); *Eucalyptus grandis* (Paiva et al., 2011a) *Citrus clementina* (Terol et al., 2008); *Carica papaya* (Lai et al., 2006); *Solanum lycopersicum* and *Solanum tuberosum* (Datema et al., 2008) amongst others.

In this work, the SSR occurrence in the cork oak BES had a frequency of 57.82 SSRs per 100 kb. This SSR frequency value is higher than the frequency observed in *Quercus robur* (29.43 per 100 kb) (Rampant et al., 2011) and in other plant species such as *C. papaya* (26.04 per 100 kb) (Lai et al., 2006) or *C. clementina* (16.65 SSRs per 100 kb) (Terol et al., 2008). However, the SSRs frequency values from the other studies were obtained from analysis of a much higher number of BAC end sequences, ranging from 20,056 (*Q. robur*) to 50,661 (*C. papaya*) sequences contrasting with the small number (48) of cork oak BESs analysed. Our results show that dinucleotide repeats were the most abundant motifs, followed by the mononucleotide repeats, while the penta- and hexanucleotide motifs were less represented. Dinucleotide repeats were also the most abundant class of microsatellites found in other species, such as *Q. robur* (47.35%) (Rampant et al. 2011) and *C. papaya* (37.8%) (Lai et al., 2006). Poly(AT) dinucleotides, were the most frequent motifs (50%) found in cork oak BESs as*Q. robur* (60.71%) and *C. papaya* (40.3%). Similarly to our findings, the poly(AG) was one of the longest microsatellites found in *C. papaya* (Lai et al., 2006). In our study were not found GC motifs as well as in the BESs analysed of *Q. robur* (Rampant et al. 2011). The low GC conten estimated in the sequences for both species can explain these results. Possibly, the low abundance of GC dinucleotide repeats in cork oak BESs may result from the trend of these repeats to form secondary structures (hairpins), resulting on a selective pressure which avoid GC accumulation in

genomes (Eustice et al. 2008). As in our findings, mononucleotide repeats were the second most frequent SSRs as in *C. papaya* (22.9%); and the poly(A) motif was also the second most abundant homopolymer (43.1%) in *C. papaya*. The pentanucleotide motif was less abundant in *Q. suber* BESs than in *Q. robur* (16.71%) and hexanucleotide SSRs were the less frequent motifs found in *Q. robur* (9.03%).

BAC-end sequence based SSRs have been successfully used for developing genetic maps in several plant species such as: soybean (Shultz et al., 2007), rice (Ammiraju et al., 2006, Kim et al., 2007), *Medicago truncatula* (Mun et al., 2006), grape (Troggio et al., 2007) and cotton (Frelichowski et al., 2006). During the construction of genome physical maps, the most important step is the integration of physical and genetic maps.For this purpose, it has been widely used the BAC library screening method however this is a time consuming procedure and it may create false positives (Mahairas et al., 1999). Hence alternatively, developing BES based markers is a powerful tool for generating integrated physical and genetic maps. This study was a preliminary approach for SSRs identification in cork oak BAC end sequences and therefore may contribute for future construction and integration of physical and genetic maps of the *Quercus suber* genome.

Homology search

During this study we performed several homology searches against different databases, however only one BAC end sequence presented a significant homology with a *Quercus suber* sequence. This finding is an indication of the limited knowledge of the cork oak genome and the scarcity of genetic studies in this species. From different homologies searches, the BAC end sequence QS-H-HDL_21N01xT7 presented significant homology with: a gag protein (EMBL AAQ94320.1) described in *Q. suber*; a GagPol3 reported in *A. thaliana* and with sequences from the cluster *Tat/Athila* of the *Ty3/Gypsy* LTR retrotransposons plant lineage (*Tat*). *Tat*-like LTR retrotransposons comprise an internal region with the genes PBS, *gag* and *pol* (Havecker, Gao and Voytas 2004). The homologies found for this sequence revealed significant identity with two of these three genes: *gag* and *pol*.

The sequences QS-H-HDL_52O24xM13 and QS-H-HDL_82D02xM13 showed significant high homologies with LINE-type retrotransposons from *Beta vulgaris* subsp. *vulgaris* with the BLASTn search. With the BLASTx search, the sequence QS-H-HDL_52O24xM13 presented significant homologies with a hypothetical protein from *B. vulgaris* subsp. *vulgaris* and with a retrotransposon protein from *Oryza sativa* (nr database), and the sequence QS-H-HDL_82D02xM13 had significant homologies with a retrotransposon protein (*O. sativa*) (nr database) and with a reverse transcriptase from *F. sylvativa* (nr database of the *Fagaceae* organisms). These results are in agreement with the microsatellite search performed with Sciroko and RepeatMasker, which identified the presence of retrotransposons of the LINE (L1/CIN4) class for both sequences.

In this study we identified different families of mobile genetic elements (MGEs). This result suggests more MGEs families would be encountered in an analysis with a higher number of cork oak BESs. It also suggests that the cork oak genome may contain a wide diversity of retrotransposons. Previous studies in *Q. suber* have found *Gypsy* retroelements such as Corky, isolated throughout genome walking

(Rocheta et al., 2012) and env-like sequences identified in both *Gypsy* and *Copia*-like retroelements (Carvalho et al., 2010). Our together with other works show that the high diversity of retrotransposon families encountered cork oak is in agreement with the actual knowledge about the importanceimportance of thesethese elements in the structure of eukaryotic organisms.

With BLASTn search for the QS-H-HDL_52O24xT7 BES was identified with significant homology a SSR from *Pinus pinaster*. A result which corroborate the Sciroko result which identified in this sequence a dinucleotide SSR with a poly(AC) motif. The results from BLASTx searches against the nr database showed that the sequences QS-H-HDL_27K11xM13 and QS-H-HDL_30E04xT7 had a significant identity with a putative protein from *Vitis vinifera* and that the sequence QS-H-HDL_52D11xT7 had a significant homology with oligopeptidase B from *Ricinus communis*.

Both BLASTn and BLASTx searches identified in the sequence QS-H-HDL_63G20xM13 significant homologies with a GDSL esterase/lipase although in different species *Glycine max* and *V. vinifera*, respectively. This BAC end sequence was sequenced from the correspondent QS-H-HDL_63G20 BAC clone which comprises the gene Cyp86A1. These genes (Cyp86A1 and GDSL) encode for enzymes previously described as being involved in the processes of fatty acid hydroxylation (Cyp86A1) and lipid metabolism (GDSL) of the acyl lipids suberin biosynthesis pathway. Our findings suggest that due to their close locations these two genes may be linked. In the BAC sequence QS-H-HDL_27K11xT7 were identified homologies with a serine-threonine protein kinase from *R. communis* and with a S-receptor kinase (SRK) from *Fagus sylvatica*. Receptor-like protein kinases (RLKs) are receptors located in the cell surface of the plasma membrane. RLKs are involved in the initiation of signaling processes in a wide variety of biological systems. In plants, RLKs present an inherent protein kinase activity with serine/threonine specificity in the kinase domain, contrary to most animal receptors protein kinases that are tyrosine kinases. Serine/threonine kinase receptors have important roles in the regulation of cell proliferation, cell differentiation, programmed cell death (apoptosis), and embryonic development. RLKs present three categories based on structural similarities of their extracellular domains and the S-domain class is one of them. The S-receptor kinase (SRK) was described as being determinant on the self-incompatibility in *Brassica* (Takasaki et al., 2000).

Nearly half of the cork oak BAC end sequences analyzed showed significant high homologies with BESs from *Q. robur* what is an indication of the high similarity present between the *Q. suber* and *Q. robur* genomes. This has been formerly reported in several studies (Ohri and Ahuja, 1990; Favre and Brown, 1996; Zoldos et al., 1998), in which were revealed remarkable interspecific genome similarities, namely in size and organization, in several oak species, including *Q. suber* and *Q. robur* (Zoldos et al., 1999).

Results of homology search against the expressed sequence tags (EST) database of the *Fagaceae* organisms revealed that from the 15 BESs which presented significant identities, the majority of the sequences had homologies with *Q. robur*, followed by *F. sylvatica*, *Q. mongolica*, *Castanopsis sieboldii* and *Q. suber* ESTs. Expressed sequence tags are short sequences with a length ranging from 200 to 800 bp,

that result from a single-pass read from complementary DNA (cDNA) clones (Nagaraj et al., 2007). cDNA clones consist in DNA complementary to mRNA thus ESTs represent portions of expressed genes and ESTs databases constitute a sample of the mRNAs or the genes expressed in a given tissue. Therefore our results are an indication that the BAC inserts from which BESs were analyzed contained additional coding regions then the genes previously identified by membrane hybridization and validated by Colony PCR. Furthermore, homologies identified with different *Quercus* species and with species from *Fagus* and *Castanea* genera are an indication of the existence of conserved sequences among the *Quercus* genus and the *Fagaceae* family.

In this study, we characterized forty eight cork oak BAC end sequences based on homology analysis. These findings provide a preliminary understanding into the organization and composition of the *Quercus suber* genome. Therefore, further studies with a higher number of BESs are required for a better knowledge of the cork oak genome.

Chapter 2 - Cytogenetic analysis of Quercus suber chromosomes using 'BAC landing'

2.1 Introduction

Cytogenetics

Cytogenetics is a science that emerged from the association of Cytology and Genetics in the end of 19th century and in the beginning of the 20th century. The cytogenetic term in classic literature is defined as the correlation of genetic and chromosomal characterization, while in modern literature is defined as the discipline which studies the genetic implications of chromosome structure and behavior (Gil and Friebe, 1998). In sum, cytogenetics is the combination of molecular and genetic analysis of genome structure and function.

In plants, cytogenetics has been applied in identification of chromosomes in genomes and in karyotype construction. Several methods have been developed and contributed to a high impact on the research and knowledge of plant genome organization. These methods include techniques of chromosome banding (Caspersson et al., 1968), *in situ* hybridization (ISH) (Gall and Pardue, 1969) and in particular fluorescent *in situ* hybridization (FISH) (Pinkel et al., 1986). Recently, have been developed molecular cytogenetic techniques that are mainly based on different modifications of ISH, which allow the analysis of somatic and meiotic chromosomes. It is the case of fluorescent *in situ* hybridization (FISH), genomic *in situ* hybridization (GISH) and multicolor fluorescent *in situ* hybridization (mFISH). These methods are used for detection of chromosomal DNA in cytological preparations allowing physical gene mapping and localization of different non-coding DNA sequences in chromosomes or interphase nuclei (Maluszynska, 2002).

Fluorescent in situ hybridization (FISH)

Initially, to detect chromosomal DNA in cytological preparations it was used isotopic labeled RNAs or DNAs. Although, the introduction of fluorescent labels linked to DNA probes which were visualized under the fluorescence microscope, became an enormous advance in cytological analysis (Jong et al., 1999). Since the 1990s, *in situ* hybridization has become an important method for molecular cytogenetics, because it enables the localization of nucleic acids in fixed biological material. The basic procedure of *in situ* hybridization consists on the specific annealing of labeled nucleic acid probes to complementary sequences in the target material, in cytological preparations, and finally the detection of the probes by light or electron microscopy (Jiang and Gill, 2006).

The development of fluorescent *in situ* hybridization (FISH), in the early 80s, allowed direct mapping of unique sequences and isolated repetitive DNA sequences on chromosomes. Therefore, this cytogenetic

tool has contributed to the advancement of chromosome studies as it enables construction of physical maps; diverse genome analyses and evolutionary studies (Heslop-Harrison, 2000; Jiang and Gill, 1994, 1996; Schwarzacher and Heslop-Harrison, 2000). Isolated repetitive DNA sequences, such as dispersed and tandem repeats, represent the comprise 40-95% of the genomic DNA in higher plants (Devi et al., 2005) and thus have been used as cytological markers. Tandem repetitive sequences include ribosomal DNA (rDNA) and other satellite DNA (satDNA) and are positioned as blocks in certain regions of the chromosomes (Walling et al., 2005). Hence, the three types of repetitive sequences (dispersed, rDNA and satDNA) can produce distinctive FISH patterns on individual chromosomes for karyotyping and phylogenetic analysis (Jiang and Gill, 2006), as it is known that these repetitive sequences can change significantly after hybridization or polyploidization events (Walling et al., 2005).

FISH is usually applied to metaphase chromosomes, although it has also been applied to meiotic pachytene chromosomes and this has become an effective technique for the creation of physical maps in plant species (Jiang and Gill, 2006). Other variation, is the application of FISH on extended DNA fibers in a method called fiber FISH, which has been used for high resolution characterization and measure of large genomic loci (Walling et al., 2005).

Due to the difficulty of mapping small DNA probes, large insert DNA clones from genomic libraries such as bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs), and cosmids, have been highly used as probes and this has become a powerful alternative for FISH mapping (Walling et al., 2005; Jong et al., 1999).

BAC 'landing'

BAC landing consists in fluorescent tagging of BACs and their hybridization *in situ* to chromosome or nuclei substrates. This technique has been successfully used to advance the knowledge on the organization of a wide range of animal and plant genomes (Hasterok et al., 2006). Specific clones can be isolated from artificial chromosome (BAC); cosmid; or yeast artificial chromosome (YAC) libraries and used as FISH probes. However, BAC libraries are usually preferred than other large insert libraries because BAC clones are stable and easily handled and present a low percentage of chimerism when compared to YACs (Shizuya et al., 1992; Woo et al., 1994; Yu et al., 2000; Peterson et al., 2002). Additionally, BAC libraries are also important sources for genome sequencing, genome structural analysis, comparative genomics, and for physical mapping (Hasterok et al., 2006).

BAC 'landing' has been used in chromosomal mapping, structural genomics and in integrated karyotyping studies in a wide variety of species such as rice (Jiang et al., 1995), barley (Lapitan et al., 1997), sorghum (Islam-Faridi et al., 2002; Kim et al., 2002), and wheat (Zhang et al., 2004a,b). This technique has also been used for positional mapping and cloning of particular genes (e.g. bacterial blight resistance gene in rice [Jiang et al., 1995]) and for combination of cytogenetics, genetics, and physical maps for instance in rice

(Cheng et al., 2001). Moreover, BAC landing can also be applied in comparative genomics and colinearity experiments as it was done in a study between *Arabidopsis thaliana* and *Brassicae* species (Jackson et al., 2000; Ziolkowski and Sadowski, 2002). Furthermore, this methodology has been used in chromosome "painting" which has enabled the study of the nuclear architecture and the chromosomes performance throughout somatic and meiotic divisions (Lysak et al., 2001; Lysak et al., 2003).

Cytogenetic studies in trees are scarse, and despite the importance of *Fagaceae* family and *Quercus* genus, very little is known about oak chromosomes. The chromosome number is not known in all members of the *Fagaceae* family, but so far the studies done on species of the genus *Castanea*, *Castanopsis*, *Fagus*, *Lithocarpus* and *Quercus* show that with only few exceptions, these species are in general diploid with 2n=24 chromosomes (Darlington and Wylie, 1955; Jaynes, 1962; Gallois et al., 1999; Zoldoš et al., 1999; Chokchaichamnankit et al., 2008). The DNA content is variable in the *Fagaceae* genera and the 2C DNA values, obtained by cytometric analysis, range from 1.11 pg in *Fagus* and 2.0 pg in *Quercus*. Within the *Quercus* genus, 2C DNA values vary from 1.17 pg in *Q. velutina* (Arumuganathan et al., unpublished), to 2.00 pg in *Q. coccifera* and *Q. ilex*, in *Q. suber* it is 1.98 pg (Zoldos et al., 1998).

In the past, most of the oak studies where performed using classical cytogenetic methods and mainly to identify their chromosome number. The study of oak chromosomes has been hindered mainly due to their small size, similar shape and to the difficulty in obtaining good spreads (Zoldos et al., 1999). In recent years, it has been used modern cytogenetic methods such as the construction of karyotypes in *Q. petraea*, *Q. robur* and *Q. rubra* (Ohri and Ahuja, 1990); and heterochromatin characterization in *Q. petraea* and *Q. robur* (Besendorfer et al., 1996). Additionally, in 2011 Ribeiro presented a study of the rDNA FISH patterns evolution in the *Fagaceae.* To our knowledge, apart from this work there has been no other study using the BAC landing technique onto root tip meiotic chromosomes of *Quercus suber.*

In this work we attempted to map genes involved in the formation and quality of the cork tissue in mitotic chromosomes of *Q. suber* using the BAC landing method.

2.2 Materials and Methods

2.2.1 Isolation and purification of plasmid DNA

In this work, the clones used for the BAC landing technique were obtained from the a cork oak BAC library constructed by Paiva et al. (2011b) described in Chapter 1.

Growing BAC clones

To obtain isolated bacterial colonies of the selected BAC clones, these were collected from the 384-well plates of th BAC collection, using a sterile toothpick to make a streak on Petri dishes with 2x Luria Broth (LB) solid medium and cloramphenicol 12,5 μg/ml. The Petri dishes were sealed and incubated overnight at 37°C in inverted position. A single colony of each BAC clone was collected from the Petri dishes with a sterile toothpick to inoculate 1 ml of sterilized 2x LB liquid medium in 2 ml eppendorf tubes, containing cloramphenicol 12,5 μg/ml. The tubes were incubated on a position which could maximize the air interface, at 37°C from 6 hours to overnight with constant shaking (220 rpm) in order to allow cell growth. After confirming the growth of the bacterial colonies, the total volume of 1 ml was transferred into a Falcon tube with 20 ml of sterilized 2x LB liquid medium and cloramphenicol 12,5 μg/ml. The tubes were incubated at 37°C from 6 hours to overnight at constant shaking (220 rpm). Later, 20 ml of the culture were used to inoculate 100 ml of sterilized 2x LB liquid medium, with cloramphenicol 12,5 μg/ml, in Erlenmeyers of 500 ml and incubated overnight at 37°C with constant shaking (220 rpm).

BAC DNA isolation and purification

BAC DNA was isolated by standard protocols using two kits in different extraction essays: QIAGEN Plasmid Midi Kit (QIAGEN, Hilden, Germany) and JETSTAR 2.0 Plasmid Midi Kit (GENOMED GmbH, Lohne, Germany) according to the manufacturer´s instructions. For the extraction of plasmid BAC DNA it was also used the alkaline lyses method adapted from Sambrook et al. (1989) which is described below.

A volume of 50 ml from a total of 100 ml of bacterial cell culture were transferred to a Falcon tube and centrifuged at 3,500 rpm for 15 min at 4°C. The supernatant mediums were discarded and the remaining 50 ml of the bacterial cell culture was transferred to the same Falcon tube and the bacterial cells were harvested again by centrifugation at 3.500 rpm for 15 min at 4°C. Cells were resuspended, by adding to the pellet 4 ml of Solution I (50 mM glucose, 25 mM Tris-HCL pH 8 and 10 mM EDTA) and 100 μg/ml of RNase, until the suspension was homogeneous and incubated for 1 h at –20°C. For the cell lysis 8 ml of Solution II (0.2 M NaOH and 1% SDS), was added and the mixture was mixed by gentle inversion of the tube until the lysate appeared to be homogeneous and incubated on ice 5 min. Afterwards, was added 6 ml of Solution III (5M potassium acetate pH 4.8) for neutralization and precipitation of cellular remains. The mixture was mixed immediately by multiple inversions until a homogeneous suspension was obtained and incubated 5 min on ice. The mixture was centrifuged at

room temperature and 7,000 rpm for 10 min and the aqueous supernatant (containing the nucleic acids) was transferred to new Falcon tubes. For removal of proteins and contaminants, two extractions with equal volume of chloroform: isoamylic alcohol (24:1) and centrifugations at 7,000 rpm for 10 min, were performed. For precipitation of the nucleic acids, were added two volumes of chilled isopropanol and incubated for 1 hour at –20°C, after centrifuged 30 min at 4°C and at maximum speed. The pellet was washed in 70% ethanol and resuspended in 40 to 100 µl of Milli-Q water and stored at -20°C until its use. The quality of the BAC DNA extracted was evaluated on a 0.6% agarose gel electrophoresis in 0.5x TBE buffer.

Verification of the BAC clones identity by Colony PCR

To verify the identity of the BAC clones, was performed a BAC colony PCR for each isolated BAC culture. The primers used for the PCR reaction were the same as the ones designed and used for the validation of the positive clones resulting from membrane hybridization. The PCR reaction was prepared in a volume of 20µl, containing 1µl of the BAC culture sample, 0.1µM of the respective forward primer, 0.1µM of the respective reverse sequencing primer, 0.2mM of each dNTP, 0.8U GoTaq® Flexi DNA Polymerase (Promega®, Madison, WI, USA), 5mM $MgCl_2$, 1x Green GoTaq® Flexi Buffer (Promega®, Madison, WI, USA) and sterile water to complete 20µL. The PCR plate was placed on a C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the following cycling conditions: 1 cycle at 95°C for 5 min and 34 cycles at 95°C for 30 sec, 60°C for 15 sec, 72°C for 1 min and 1 cycle at 72°C for 5 min. To evaluate the Colony PCR products and their size, it was done an electrophoresis on a 2% agarose gel with 0.5X TBE buffer.

Validation of the isolated BAC DNA by PCR

To validate the BAC DNA extracted was performed a PCR. The procedure was the same as described in the previous paragraph with the exception that was used, for the PCR reaction, 1µl of the extracted BAC DNA sample instead of BAC culture.

Quantification of plasmid BAC DNA

The quantification and purity of the DNA was evaluated on a 1% agarose gel electrophoresis in 1x TAE buffer using λ*Hind*III ladder (50 ng/µl) as size marker. The image capture and treatment was achieved by using the software Gel Doc 1000 (Bio-Rad, Hercules, CA, USA).

2.2.2 Preparation of mitotic chromosome and nuclei substracts for FISH

Roots collection

Acorns of *Quercus suber* germinated in sterile sand on a plant growth chamber at 21°C, with 80% of relative humidity and a photoperiod of 16h day. The germination time varied depending on the quality of seeds but the seedlings were maintained in humid sand until the development of adventitious roots.

C-mitotic treatment

Seminal and adventitious roots with length around 5 cm were excised and immediately washed in distilled water. Afterwards, to induce c-metaphases the roots were placed into eppendorf tubes with 1 ml of a saturated solution of α-bromonaphthalene and incubated 3 hours in total obscurity at RT. Later, the roots were fixed with a solution of fresh absolute ethanol: glacial acetic acid (3:1 v/v) and incubated in agitation for 10 min. The roots were then placed in eppendorf tubes with fresh fixative 3:1 and incubated in agitation overnight at room temperature (RT). On the following day the roots were washed several times with freshly made fixative 3:1 and then stored with new fixative at -20 °C until being used.

Coating slides

Coating of slides is an important step as it prevents material loss in following treatments and enhances adhesion of nuclei and chromosomes to the glass surface. Therefore, glass microscope slides 76 x 26 mm were washed with water and detergent for 30 min, rinsed vigorously and dried at 37°C. Afterwards the slides were immersed in freshly prepared 2% (v/v) solution of APTES (3-aminopropyltriethoxysilane) in acetone for 10 sec, briefly immersed in a solution of pure acetone and finally washed in distillated water. The slides were air-dried at room temperature and stored in a slide box until needed.

Preparation of chromosome and meristematic nuclei squashes by the drop technique

The drop technique methodology used in this work was performed according to the protocol described in Ribeiro et al. (2011). The fixed material was washed twice in 1x enzymatic buffer (EB) (1:10 dilution of a stock solution 10x EB: 40ml de monohydrated citric acid 0.1M and 60ml of dehydrated Trisodium citrate 0,1M; pH 4.8) for 10 min with agitation. The root meristem was isolated and the root cap was excised with a scalpel using a stereo microscope. The isolated root meristems were placed in 1 ml of enzyme misture for digestion of the cellular wall and incubated at 37°C for 3h. The enzyme mixture contained 2% (w/v) cellulase 'Onozuka' R10 (from *Trichoderma viride*, Serva), 3% pectinase (v/v) (from *Aspergillus niger*, solution in glycerol) (Sigma-Aldrich Co., MO, USA), 0.3% (w/v) pectolyase Y-23 (from *Aspergillus japonicus*) (Sigma-Aldrich Co., MO, USA), dissolved in 0.03% EDTA in 2x SSC, pH 4.2 (Zoldos et al., 1999).

After digestion, a pellet was obtained by centrifugation at 3,500 rpm for 4 min and the supernatant carefully removed. It was added to the pellet 400µl of 1x EB, centrifuged 4 min at 3,000 rpm and the supernatant discarded, this step was performed twice. Later, was added to the pellet 400µl of fresh and ice cold absolute ethanol: glacial acetic acid (3:1 v/v), centrifuged 4 min at 2,500 rpm and the supernatant discarded, this step was also performed twice. Finnaly, was added between 20 to 100µl of 3:1 to resuspend the pellet. Around 10µl of the suspension was dropped onto an APTES coated slide, gently blown to spread the drop and and allowed to air-dry.

After air drying, the slides were analysed under the phase contrast microscop to check the presence of isolated nuclei in a reasonable concentration and the absence of cytoplasm. In the case of cytoplasm presence, the suspension was again centrifuged for 4 min at 2,500 rpm, the supernatant was discarded and replaced by fresh 60% glacial acetic acid during 2 min to remove the cytoplasm. The suspension was centrifuged 4 min at 2,500 rpm and resuspended again with an appropriate volume of fixative. In the absence of cytoplasm, the drop technique continued to be performed into several slides until the suspension was finished. Finally, the slides were scanned by phase contrast microscopy to verify the squash quality, the nuclei concentration and the presence of metaphase plates. When necessary it was added to each slide 8µl of Vectashield-DAPI (4',6-diamidino-2-phenylindole) (2µg/ml) mounting staining medium and after an incubation of 90 min at 4°C the slides were observed with the epifluorescence microscope Axioskop 2 (Zeiss, Oberkochen, Germany), the images were captured using an AxioCam digital camera (Zeiss, Oberkochen, Germany) controlled by AxioVision 3.0 (Zeiss, Oberkochen, Germany). The selected slides were stored at 4°C until being used for fluorescent *in situ* hybridization (FISH).

2.2.3 BAC 'landing'

DNA probes

In this work were used probes obtained from cloned DNA sequences and polymerase chain reaction (PCR) products which are listed and described below:

- pTa71, containing a 9 kb EcoRI fragment from the highly repeated 45S rDNA sequence isolated from wheat (Gerlach and Bedbrook, 1979), cloned into the plasmid pUC19.
- 5S Qsu, a 353 bp fragment (including 89 bp of the conserved gene sequence and 264 bp of the non-transcribed spacer) of the repeated 5S rDNA sequence obtained by polymerase chain reaction from genomic DNA of *Q. suber*.
- BAC clones plasmid DNA from the *Q. suber* BAC library.

Labelling of pTa71 (45S r DNA) or BAC DNA by nick translation

For digoxigenin labelling was used the DIG-Nick Translation Mix (Roche Diagnostics, Mannheim, Germany) according to manufacturer 's instructions and the protocol used is described below.

In ice cold 0.5 ml thin-walled PCR tubes, it was added 1µg of the plasmid DNA (pTa71 or BAC DNA) to Milli-Q water for a final volume of 16 µl and 4 µl of Digoxigenin Nick Translation Mix (Roche Diagnostics, Mannheim, Germany) containing 5x concentrated stabilized reaction buffer in 50% glycerol (v/v) and DNA-polymerase I/ DNase I enzyme mixture, 0.25 mM dATP, 0.25 mM dCTP, 0.25 mM dGTP, 0.17 mM dTTP and 0.08 mM of labelled dNTP: DIG-11-dUTP. Then, the labeling mix was incubated at 15°C for 90 min. The reaction was stopped by adding 1µl of 0.5 EDTA (pH 8.0) and heating at 65°C for 10 min.

For biotin labelling used Nick Translation Mix (Roche Diagnostics, Mannheim, Germany) as recommend by the following standard protocol. In ice cold 0.5 ml thin-walled PCR tubes it was added 1µg of plasmid DNA (pTa71 or BAC DNA), Milli-Q water to a final volume of 12µl, 4µl of dNTP mixture (0.25 mM dATP, 0.25 mM dCTP, 0.25 mM dGTP, 0.17 mM dTTP and 0.08 mM of labelled dNTP: biotin-11-dUTP) and finally 4 µl of Nick Translation Mix (Roche Diagnostics, Mannheim, Germany) containing 5x concentrated stabilized reaction buffer in 50% glycerol (v/v) and DNA-polymerase I/ DNase I enzyme mixture. The reaction occurred for 90 min at 15°C and was stopped by adding 1 ml 0.5 M EDTA (pH 8.0) and heating at 65°C for 10 min.

Labelling of 5S Qsu by polymerase chain reaction (PCR)

In order to amplify and label the 5S Qsu sequence from *Q. suber* genomic DNA with biotin or digoxigenin, a polymerase chain reaction (PCR), was performed. The PCR reaction was prepared in a volume of 50µl, containing 1µl of genomic DNA from *Q. suber* (100ng/µl), 0.5 mM Primer 5SQsu cw (5' ATCCCATCAGAACTCCG 3'), 0.5 mM Primer 5SQsu ccw (5' GCAACGATGCTCCTTAA 3'), 0.1 mM of dNTPs, 1.5 mM $MgCl_2$, 1X PCR buffer, 0.03 mM Biotin-16-dUTP or Digoxigenin-11-dUTP, sterile water to complete 50µl and 2U NZYTaq DNA polymerase (NZYTech Lda, Lisboa, Portugal). The PCR tubes were placed on a DNA Engine Peltier Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the following cycling conditions: 30 cycles at 94°C for 1 min, 45°C for 45 sec, 72°C for 1.30 min and 1 cycle at 72°C for 6 min.

Pretreatments for DNA in situ hybridization

After selecting the best drop preparations, it was added to each slide 200 µl of pepsin solution (6.7 µg/µl in HCL 0.01M) for cytoplasm digestion. Afterwards, the slides were covered with a plastic cover slip (24 x 24 mm) and incubated in a humid chamber at 37°C for 10 min. The cover slips were carefully removed by peeling off with forceps and the slides were washed twice for 5 min each in 2x SSC (Saline Sodium Citrate, 1:10 dilution of a stock solution 20x SSC: 3M NaCl, 0.3 M sodium citrate in distilled water, pH 7.0) with agitation and at RT. Then was added 200 µl of RNase solution [0.1 mg/ml of DNase-free RNase (Sigma-Aldrich Co., St. Louis, MO, USA) in 2x SSC] to each slide, these were covered with a plastic cover slip (24 x 24 mm) and incubated on a humid chamber at 37°C for 1 h. The cover slips were carefully removed by peeling off with forceps and after, the slides were washed three times for 5 min each in 2x SSC in agitation and at RT. The slides were dehydrated in successive washes with 70% and 100% ethanol (3 min with agitation each wash) and air dried at RT.

Denaturation and hybridization

During this work, the hybridization mixtures were prepared for a stringency of 76%, which means the percentage of sequence identity or the percentage of nucleotides that properly match in the hybrid DNA molecule. The hybridization mixtures were prepared on ice and in 0.5 ml microcentrifuge tubes. The hybridization mixtue contained 50% deionized formamide, 2x SSC, 10% (w/v) dextran sulfate, 1% (w/v) sodium dodecyl sulphate (SDS), 1.5 µg/µl sonicated salmon sperm DNA, 200 ng/slide BAC DNA probe, 2.5 ng/µl for both pTa71 and 5SQsu and sterile water for a final volume of 40 µl. The

hybridization mixtures were denaturated at 80°C for 10 min and after cooled in ice for 5 min. Then, 40 µl of the hybridization mixture were added to each slide and immediately covered with a plastic cover slip. The slides were placed on a MJ Research PTC-100™ Thermal Cycler (MJ Research Inc., Quebec, Canada) and subjected to the following program: 85°C for 5 min and after the temperature was gradually reduced in five steps to 37°C. Later, the slides were incubated overnight on a humid chamber at 37°C.

Stringent washing

Stringency washes of 84% were performed as the follows: slides were washed in 2x SSC at 42°C for 3 min in a washing box and the cover slips carefully peeled off. Then, the slides were incubated in 20% (v/v) formamide in 0.1x SSC, for 10 min at 42°C. Afterwards, the slides were washed twice in 2x SSC for 5 min at 42°C with agitation and finally washed in 2x SSC for 5 min at RT also with agitation.

Immunodetection of DNA labelled probes

The slides were incubated in 4x SSC/0.2% Tween®-20 (Sigma-Aldrich Co., MO, USA) twice for 5 min at RT. Later were added 200 µl of 5% bovine albumin serum (BSA) (Sigma-Aldrich Co., MO, USA) solution in 4x SSC/0.2% Tween®-20 to each slide, covered with plastic cover slips and incubated for 5 min at RT. A detection solution was prepared containing BSA solution (5% in 4x SSC/0.2% Tween-20), 0.2 µg/µl of anti-digoxigenin antibody conjugated with FITC (Boeringer Mannheim) and 1 µg/µl streptavidin-biotin-cy3 (Sigma-Aldrich Co., MO, USA) to an appropriate final volume. It was added 50 µl of detection solution to each slide, the cover slips were replaced and the slides were incubated for 1h at 37°C in a humid chamber. Afterwards, the slides were washed with 4x SSC/0.2% Tween®-20, three times for 5 min at RT with agitation.

Counterstaining and mounting

To each slide, were applied 8 µl of Vectashield-DAPI (4',6-diamidino-2-phenylindole) (2 µg/ml) mounting staining medium. DAPI is a common DNA stain, which allows the chromosomes and nuclei visualization under UV light. Vectashield as an antifade compound prevents photobleaching of the fluorophores while under ilumination. The plant material was covered with a thin glass cover slip and the excess of mounting medium was removed. The slides were stored in the dark at 4°C for at least 24hs.

Image capture and processing

The slides were observed and analysed with the epifluorescence microscope Axioskop 2 (Zeiss, Oberkochen, Germany), the images were captured using an AxioCam digital camera (Zeiss, Oberkochen, Germany) controlled by AxioVision 3.0 (Zeiss, Oberkochen, Germany). Images were captured with appropriate filters, then the three channels of each image were superimposed and when necessary it was done the appropriate adjustments.

2.2.3.1 Signal redetection and amplification

Slides washing

For DAPI removal, the slides were placed in a washing box and rinsed with fresh absolute ethanol: glacial acetic acid (3:1 v/v) until detachment of the glass cover slips. The fixative was replaced and the slides were washed at least 3x in distilled water and once in 4x SSC/0.2% Tween®-20 (Sigma-Aldrich Co., MO, USA).

Imunodetection of digoxigenin labeled probes

For hybridization of the primary antibody, it was prepared for several slides 5% bovine albumin serum (BSA) (Sigma-Aldrich Co., MO, USA) solution in 4x SSC/0.2% Tween®-20 (Sigma-Aldrich Co., MO, USA). Later, the slides were washed twice for five minutes in 4x SSC/0.2% Tween®-20. Then, it was prepared a detection solution containing BSA solution (5% in 4x SSC/0.2% Tween-20) and 0.2 μg/μl of anti-digoxigenin antibody from sheep conjugated with FITC (Boeringer Mannheim). The following steps of this procedure were the same as decribed in section 'Immunodetection of DNA labelled probes'.

After hybridization of the primary antibody, was performed the hibridization of the secondary antibody using in the 5% bovine albumin serum (BSA) solution previously prepared and a detection solution containing BSA solution (5% in 4x SSC/0.2% Tween-20), 0.2 μg/μl of anti-sheep antibody conjugated with FITC (Boeringer Mannheim) and 1 μg/μl streptavidin-biotin-cy3 (Sigma-Aldrich Co., MO, USA). The solutions were added in the same manner as for the primary antibody hybridization and the subsequent steps, including the counterstaining and mounting of the slides, were the same as previously described.

2.3 Results

2.3.1 Isolation and purification of plasmid DNA

Growing BAC clones

For this study were selected 8 BAC clones containing genes involved in the formation and quality of the cork tissue: DHCA-S-AT, Cyp86A1, GDSL, CCR, HCBT, Dolaccase, WRKY and ACC-OX. These genes are described as being important in the two major metabolic pathways of suberin biosynthesis and in regulatory mechanisms of cork formation (see *Table 2*). The genes DHCA-S-AT, Cyp86A1 and GDSL are involved in the acyl-lipids pathway; CCR, HCBT and Dolaccase are involved in the phenylpropanoid pathway; WRKY and ACC-OX are possible regulatory proteins in the cork tissue.

BAC DNA isolation and purification

The isolation of the BAC DNA was performed using the alkaline lyses method and two different extraction kits. Using an adaption of the alkaline lyses protocol from Sambrook et al. (1989) we were able to successfully extract DNA from the BAC clones. Based on the electrophoresis analysis (*Figure 15*),

we observed that the extracted BAC DNA presented good quality in the majority of the samples. The BAC DNA samples presented a single bright band with high molecular weight (> 12,000 bp) and almost no signs of smear, except for the Dolaccase sample for which was not observed any band, indicating that the extraction failed in this sample or the quantity of DNA extracted was so low that could not be detected on the electrophoresis profile. The clones harboring HCBT, CYP86A1, DHCA-S-AT and CCR genes showed an intense band, whereas the CCR is the most intense and thick band followed by the bands of the clones harboring Cyp86A1, DHCA-S-AT and HCBT genes. The isolated BAC DNA from WRKY, ACC-OX and GDSL samples presented thinner bands and with lower intensity, nevertheless with good quality.

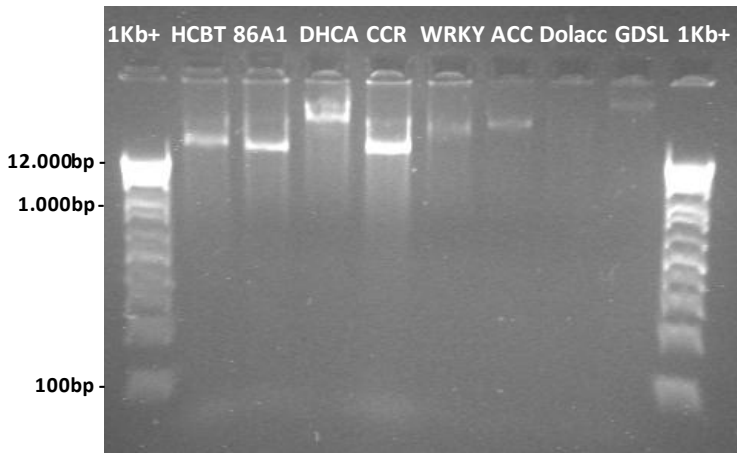


Figure 15. Evaluation of the quality of the BAC DNA extracted from BAC clones by the alkaline lyses method adapted from Sambrook et al. (1989), in a 0.6% agarose gel in 0.5x TBE using 1Kb+ ladder (Invitrogen, Grand Island, NY, USA).

Later, we performed extractions with two different plasmid DNA isolation kits. The BAC DNA was successfully isolated using both kits, JETSTAR 2.0 Plasmid Midi Kit and QIAGEN Plasmid Midi Kit. It was obtained BAC DNA with good quality for all samples. Although, despite the use of three different methodologies to extract and purify the BAC DNA, was always observed the presence of bacterial genomic DNA in the DNA samples extracted.

When comparing the three methods, the JETSTAR 2.0 Plasmid Midi Kit (GENOMED GmbH, Lohne, Germany) protocol resulted in lower contamination of the samples with genomic DNA. However, the JETSTAR kit originated BAC DNA samples containing higher levels of residual RNA. From the isolation methods used in this work, the JETSTAR kit presented the easier and faster protocol. By using the modified alkaline lysis protocol and the two Plasmid Midi Kits, but mainly the JETSTAR kit, we were able to consistently isolate BAC DNA that was later successfully validated by PCR.

Validation of the BAC clones identity and of the BAC DNA extracted

After the BAC clones isolation and growth was necessary to confirm the identity of the BAC clones by by Colony PCR. In *Figure 16* is represented an example of a Colony PCR result for confirmation of the BAC clones identity and validation of the BAC DNA extracted from these cultures, simultaneously.

The identity of the BAC clones was confirmed for all samples, since it was observed amplification of a PCR product from the BAC culture samples, except for the DHCA sample. The isolated BAC DNA was validated for all samples, as it was observed amplification of PCR products for all the BAC DNA samples. In the case of the DHCA sample, the absence of amplification product from the bacteria culture may be due to some procedure failure during the electrophoresis analysis. Nevertheless, the isolated DHCA BAC DNA was validated which indicates that the bacteria culture was from the correct BAC clone.

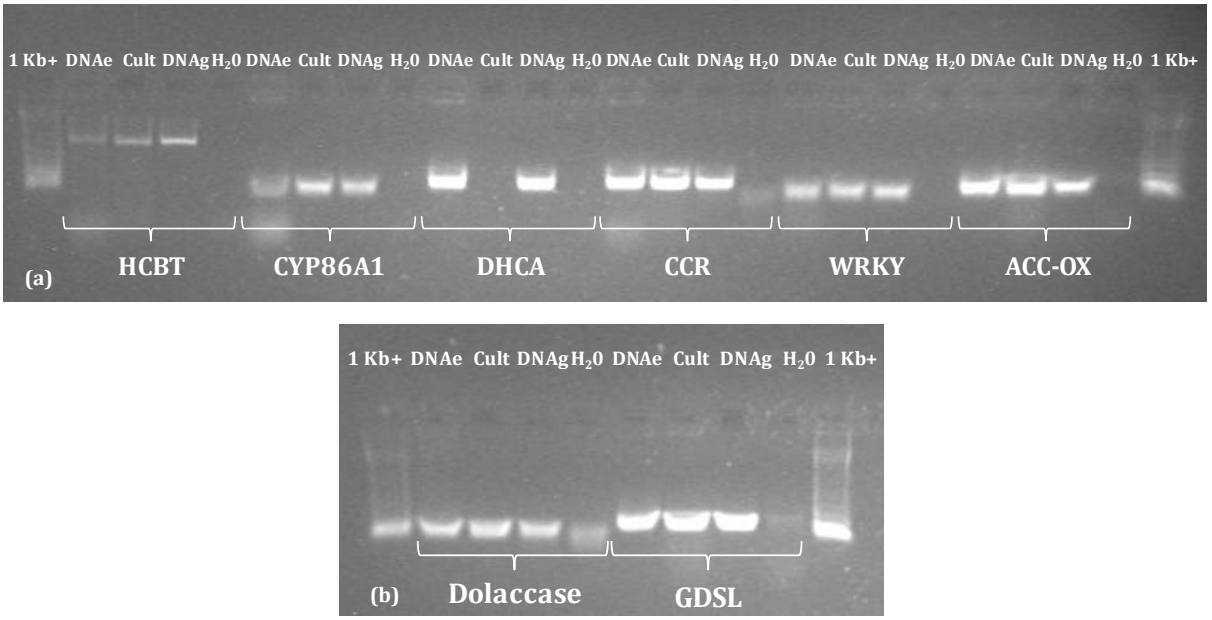


Figure 16 (a) and (b). Evaluation of the quality of the Colony PCR amplification products on a 2% agarose gel in 0.5x TBE and using 1Kb+ ladder. (DNAe), extracted BAC DNA sample; (Cult), Bacteria culture from which BAC DNA was isolated; (DNAg), positive control with *Q. suber* genomic DNA and (H_20), negative control with sterilized Milli-Q water.

Quantification of plasmid BAC DNA

To quantify the isolated BAC DNA through electrophoresis analysis, was used the λ *Hind*III DNA ladders as a size and amount marker to compare the bands intensity and to calculate the concentration of the samples (*Figure 17*).

The BAC DNA isolated with the three different methods presented in general low yields. However, we observed that BAC DNA extractions performed with the alkaline lysis method and the JETSTAR kit resulted in higher yields. Therefore, these two extraction methods were preferly used to achieve the sufficient quantity of good quality BAC DNA required for the BAC landing experiments.

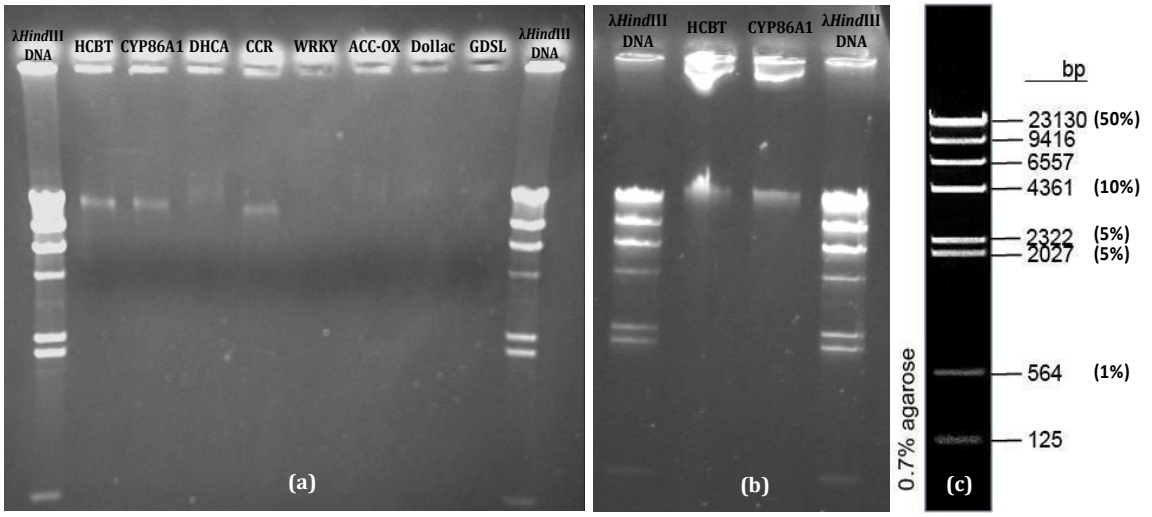


Figure 17. Evaluation of the quality and quantity of the BAC DNA extracted from BAC clones, on 1% agarose gel in 1x TAE and with λ *Hind*III DNA ladder. (a) Extraction by the alkaline lyses method adapted from Sambrook et al. (1989); (b) Extraction with QIAGEN Plasmid Midi Kit; (c) λ *Hind*III DNA size marker, indicating the fragments base pairs (bp) and the percentage of the final concentration to which similar band intensities correspond.

Preparation of chromosome and meristematic nuclei squashes by drop technique

The drop technique revealed to be a successful method as it provided metaphase spreads with good quality and with small amounts of cytoplasm, as depicted in

Figure 18. In all the cork oak samples used during this study was observed the same diploid chromosome number number (2n=24).

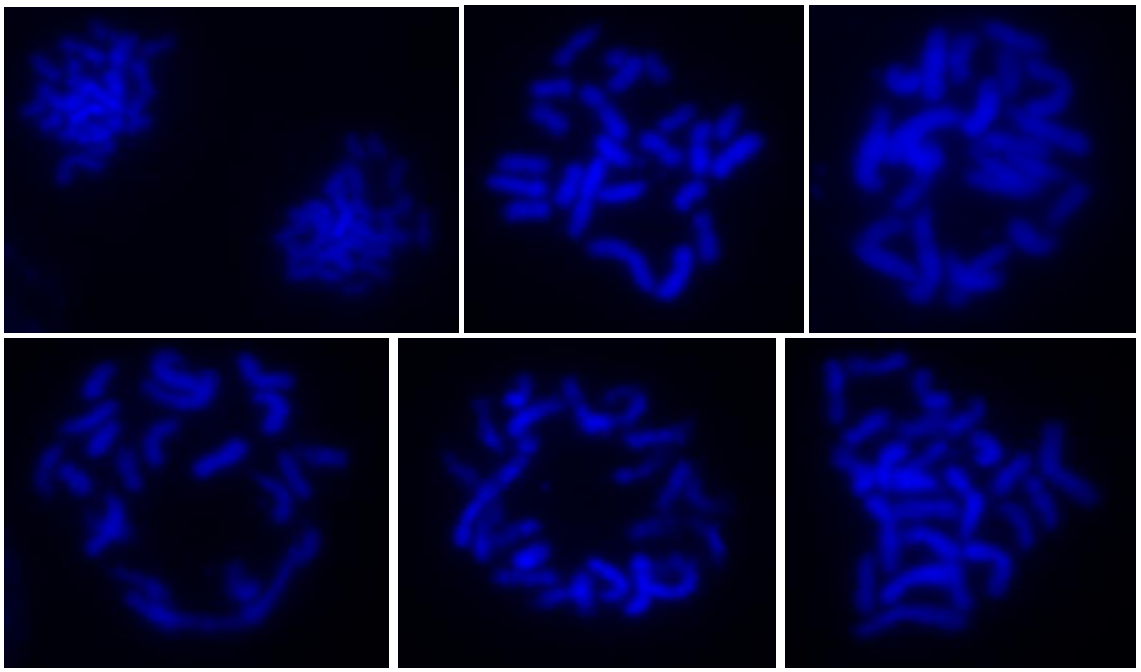


Figure 18. Meristematic root-tip metaphase chromosomes of *Quercus suber* counterstained with DAPI.

2.3.2 BAC 'landing'

During this work several BAC landing assays were performed. In *Table 11* are presented the hybridization results obtained from the carried out experiments. In this work were performed six

BAC landing assays using different combinations of BAC DNA and rDNA probes. In the diverse hybridization experiments, probes were labeled with different fluorochromes and for the assay III the BAC DNA was sonicated to obtain smaller fragments of approximately 1,000 pb. In all BAC landing experiments was observed hybridization signal for the respective rDNA probes used. However, it was never observed any hybridization signal of the BAC DNA used, even after performing signal redetection and amplification as occurred in the assays II and VI. The BAC landing results of the assay I, are shown in *Figure 19*. As depicted in this figure, it was not observed the green signal resulting from the landing or hybridization of the digoxigenin-labeled GDSL and WRKY BAC DNAs to the cork oak chromosomes.

However, was observed the hybridization signal corresponding to the 45S rDNA probe. The rDNA probes were used as a control to evaluate the success of the hybridization. Hence, the presence of signal of the rDNA probes indicates that the fluorescent *in situ* hybridization (FISH) technique was successfully performed.

Table 11. Summary list presenting the results of the BAC landing assays performed. Are indicated the BAC DNA and rDNA probes used and the respective fluorochromes selected for each hybridization experiment.Are also presented the slides in which were performed the signal redetection and amplification procedure. * Sonicated BAC DNA.

Hybridization assay	**Slides**	**Labelled probes**		**Hybridization signal**		**Signal amplification**
		Biotin-rDNA	**Digoxigenin-BAC**	**rDNA**	**BAC**	
I	**1**	pTa71 (45S)	GDSL	+	-	No
	2	pTa71 (45S)	WRKY	+	-	No
II	**3**	5S	HCBT	+	-	Yes
	4	5S	CYP86A1	+	-	Yes
III	**5**	5S	CCR *	+	-	No
	6	5S	DHCA-S-AT *	+	-	No
		Biotin-BAC	**Digoxigenin-rDNA**			
IV	**7**	HCBT	5S	+	-	No
	8	HCBT	5S	+	-	No
	9	CYP86A1	5S	+	-	No
	10	CYP86A1	5S	+	-	No
V	**11**	HCBT	5S	+	-	No
	12	CYP86A1	5S	+	-	No
VI	**13**	HCBT	5S	+	-	Yes
	14	CYP86A1	5S	+	-	Yes
	15	CYP86A1	5S	+	-	Yes

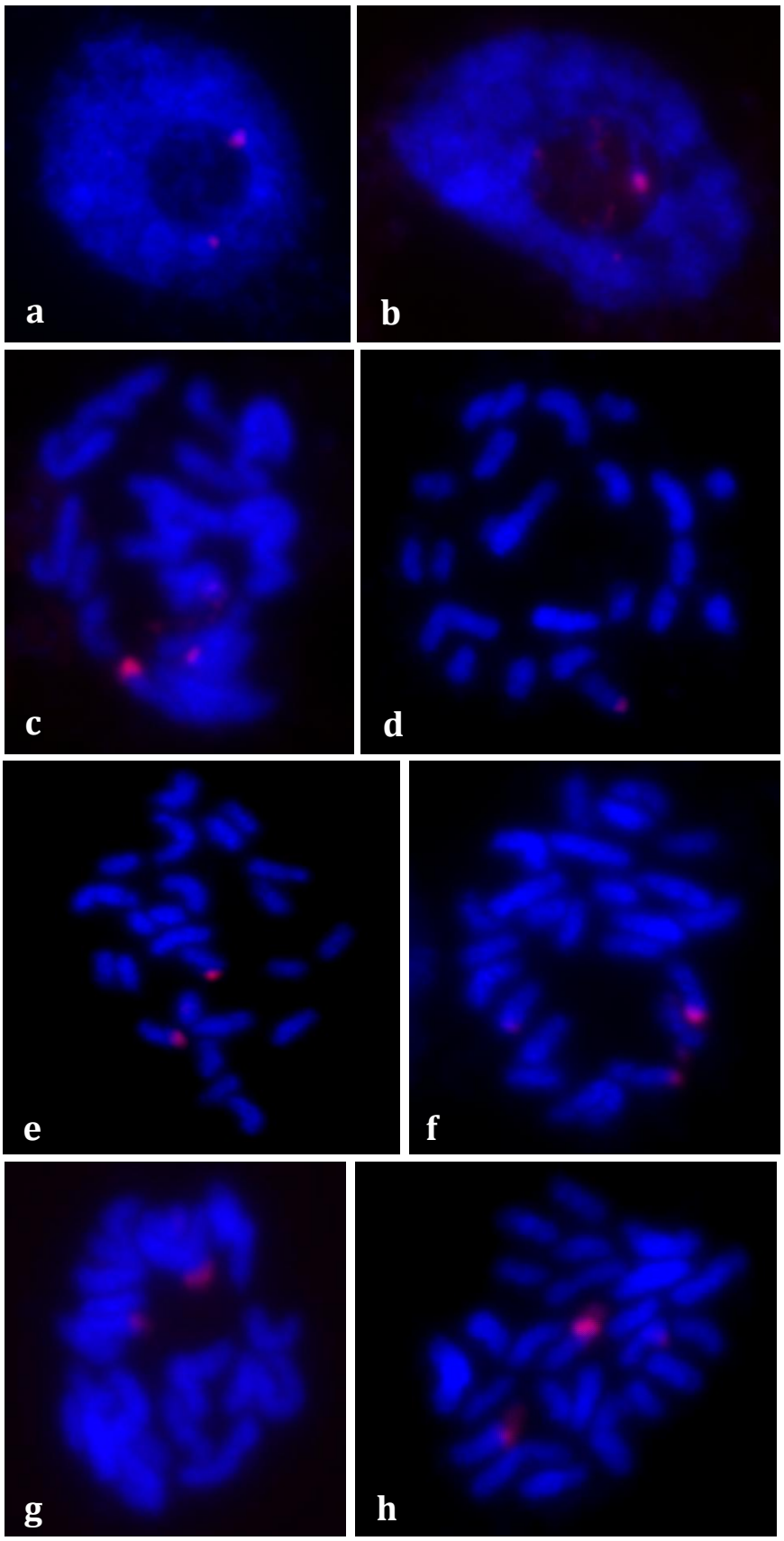


Figure 19. Landing of digoxigenin-labeled BAC DNA and biotin-labeled rDNA probe pTa71 (45S rDNA) on meristematic root-tip nuclei and c-metaphase chromosomes of *Quercus suber L*. **(a)**, **(b)**, **(c)** and **(d)** BAC landing was performed using DNA from GDSL BAC clone. **(e)**, **(f)**, **(g)** and **(h)** And using DNA from WRKY BAC clone. Chromosomes were counterstained with DAPI and the red signal corresponds to the hybridized pTa71 (45S rDNA) probe.

2.4 Discussion

Isolation and purification of plasmid DNA

In this work we were able to successfully isolate BAC DNA with reasonable good quality, from BAC clones of the *Quercus suber* BAC library, using the alkaline lysis protocol adapted from Sambrook et al. (1989) and the extraction kits QIAGEN Plasmid Midi Kit (QIAGEN, Hilden, Germany) and JETSTAR 2.0 Plasmid Midi Kit (GENOMED GmbH, Lohne, Germany). The isolated BAC DNA presented contaminationcontamination with bacterial genomic DNA and low yields. The difficulty in obtaining sufficient quantities of high quality BAC DNA from standard midipreps, created a challenge to achieve the adequate quantity of good quality BAC DNA for BAC landing.

Molecular cytogenetic analysis

For the success of cytogenetic studies is fundamental to have properly spread metaphase chromosomes. Therefore, the main problem in cytogenetic studies remains in the difficulty to obtain consistently high quality chromosome spreads (Barch et al., 1997). During this work, it was used the drop technique which provided good metaphase spreads that allowed unifocal views of *Quercus suber* chromosomes. In general, using this technique, we were able to obtain chromosome spreads with almost no presence of cytoplasmic debris, nuclear membranes and cell walls, thus providing low background hybridization and good accessibility for the probes.

The diploid chromosome number observed in all *Q. suber* spreads was 2n=24. This observation supports previous studies suggesting that the haploid chromosome number of the *Quercus* genus is n=12 (Duffield 1940; Stairs 1964; Ohri and Ahuja, 1990; D' Emerico et al., 1995-2000; Zoldos et al., 1998; Kurokawa and Yonezawa, 2004; Yilmaz et al., 2008; Yilmaz et al., 2011) and that the chromosome number in the genus is stable.

BAC landing

To our knowledge, this work was the first attempt to land BAC DNA onto *Q. suber* chromosomes. For this purpose, BAC clones from the cork oak BAC library, which included candidate genes for cork formation, were selected and used as probes for BAC landing. In addition, two genes encoding the 45S (45S rDNA) and 5S ribosomal genes (5S rDNA) were also used as probes. These highly conserved and tandemly repeated ribosomal gene sequences are currently the most common used FISH markers in a variety of species (Ribeiro et al., 2008).

The results of six distinct BAC landing assays consistently showed strong fluorescent signals emitted from the rDNA probes, which is an indication of the success of the fluorescent *in situ* hybridization (FISH) technique. However, no visible signal was observed from hybridized BAC DNA probes. The lack of visible fluorescent signal from the BAC DNA probes suggests that the BAC landing experiments were not efficient.

There are several possible reasons for this result. First, mapping low or single-copy sequences in plant chromosome preparations produce an inconsistent fluorescent signal (Jiang et al., 1995). That is the main motive why *in situ* hybridization in plant species have been mostly used for mapping multicopy gene families and repetitive DNA sequences (Jiang and Gill 1994). Second, when compared to animal tissues, the sensitivity of *in situ* hybridization in plants (including FISH) is lower due to the presence of rigid cell wall fragments and cytoplasmic debris in the chromosome preparations. Third, the quality of BAC DNA extracted was not satisfactory for the BAC landing experiments. Possibly due to the high volume of cells in the starting culture that increased the presence of proteins and cellular debris in the sample and compromised the achievement of clean DNA (Kelley et al., 1999). In addition, most of the BAC DNA samples contained some bacterial genomic DNA and this may have interfered in the BAC landing experiments.

In this study, we were not able to map BACs of interest for cork formation in mitotic cork oak chromosomes through the BAC landing methodology. The protocol used was originally customized for *Brachypodium distachyon* (Jenkins and Hasterok, 2006) and possibly, further troubleshooting will be necessary for the optimization of the protocol in forest trees species. In particular, the BAC DNA isolation and purification method could be improved for yielding enough good quality BAC DNA for flurorescent *in situ* hybrization.

Chapter 3 - Expression analysis of suberin biosynthesis genes by qPCR

3.1 Introduction

3.1.1 What is Quantitative Real-time PCR?

The Real-time PCR or Quantitative Real-time PCR (qPCR) is a molecular biological technique for quantitative nucleic acid analysis that makes possible to quantify the initial amplified product, since the amount identified at a certain point of the assay is directly related to the initial amount in the sample. This technique allows achieving information for relative and absolute measurements of a certain starting material and nowadays as a great value in many fields of biological research. Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQUE Guidelines, Bustin et al., 2009) should be used to ensure precision and traceability of the qPCR results.

3.1.2 PCR versus qPCR

During classical DNA polymerase chain reaction (PCR), the double-stranded DNA containing the target sequence is denaturated by cycles of heating followed by a decreasing in the reaction temperature for primer annealing. In this step, the two primers (forward and reverse) flank the target sequence and the thermostable DNA polymerase Taq polymerase (from the *Thermus aquaticus* bacterium) synthesizes complementary new strands originating two double-stranded DNA molecules identical with the parental double-stranded molecule. The repeated cycles of denaturation, annealing and synthesis result on an exponential amplification of the number of the segments replicated.

Therefore the classical PCR can be divided in three phases, initially it has an exponential phase where the product is precisely doubled at each cycle (assuming the reaction has 100% efficiency). At this stage the reaction is very specific and precise. With the progression of the amplification reaction, some of the reaction components are being consumed. As a result, the reaction starts to slow down and the doubling of the PCR products cease. This corresponds to the linear phase of the reaction. Ultimately the reaction stops and therefore the amplification of PCR products comes to an end. This corresponds to the plateau phase, where each reaction reaches this stage at different points even if they have the same starting target concentration. When the PCR reaction is completed, is performed the analysis and measurement of the amplification product by gel electrophoresis, a step also known as the end-point detection.

In contrast, in Real-Time PCR end-the point detection by gel electrophoresis can be avoided because it allows the immediate detection and quantification of the accumulated amplified product as the reaction progresses, in real time. The Real-time detection methods of the PCR products is based on the inclusion in the reaction of a fluorescent molecule that increases its fluorescent signal proportionally to the increase in the amount of DNA amplified. Fluorescence changes are monitored at each PCR cycle; as a result, the measured fluorescence reflects the

amount of amplified product in each cycle allowing the user to be able to follow in real time the reaction. Real-time PCR follows the same principle as PCR, which is the assumption that at the exponential stage the amount of target doubles at each cycle. The qPCR measurements are specifically made during the exponential phase (*Figure 20*), when the amplification is more efficient and less affected by the limitation of reagents or reaction conditions (Walker, 2002).

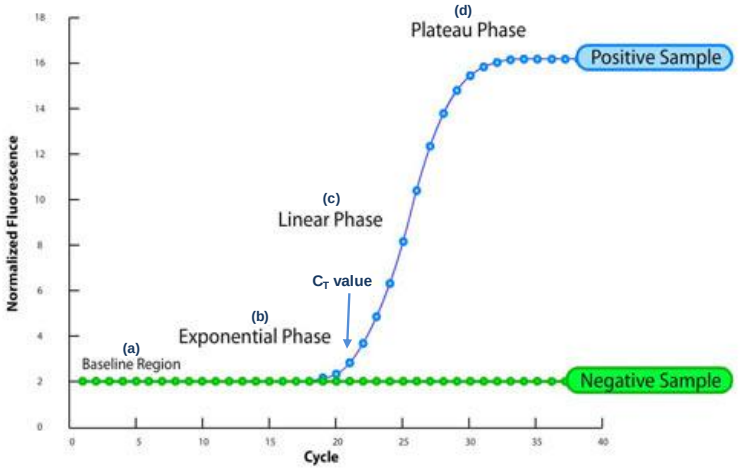


Figure 20. qPCR amplification plot. The amplification plot presents: (a) The baseline region corresponding to the initial cycles of qPCR where there is little change in fluorescence signal above the background; (b) The exponential phase in which the theoretical doubling of product at every cycle creates exponential signal growth and the reaction is very specific and precise; (c) The linear phase where the reaction components start to become limiting and the reaction efficiency is falling so that the signal no longer grows exponentially; (d) The plateau phase in which the reaction components have been exhausted and the reaction can not generate more fluorescence. Image adapted from http://www.abbottmolecular.com/technologies/real-time-pcr/maxratio-data-analysis.html.

The measurement is done when there is an accumulation of amplified product enough to produce a fluorescent signal that reaches a specific threshold level of detection. The cycle number where this takes place is entitled threshold cycle, or the CT value. The threshold is the level of fluorescence that is above the average background, the baseline (*Figure 21*). The baseline region corresponds to the initial cycles of PCR during which, despite the theoretical doubling of product, there is little change in fluorescence signal above the background.

The CT value is inversely correlated with the starting amount of nucleic acid present in the sample. Thus, the CT value is in inverse proportion to the expression level of the gene. So, if a sample has an initial large amount of template, it will be necessary few amplification cycles to produce sufficient product to produce a fluorescent signal above the threshold. This point is reached earlier and consequently it gives an earlier or lower CT value. And the opposite occurs when is obtained a high or late CT value, which is an indication that the initial amount of template was smaller and therefore it were necessary more amplification cycles for the fluorescent signal to rise above the threshold.

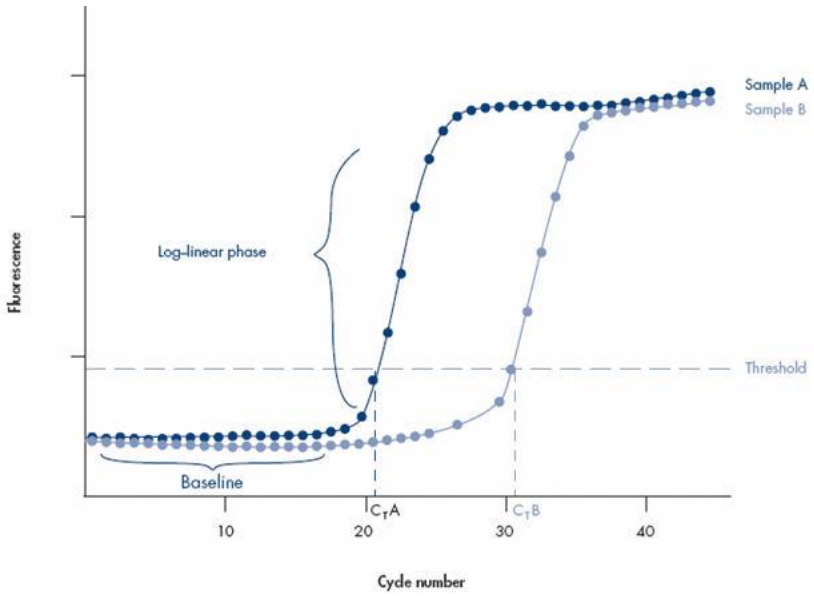


Figure 21. Real-time PCR growth curves. Real-time PCR data are produced as sigmoidal-shaped amplification plots (when using a linear scale), in which fluorescence is plotted against the number of cycles**.** The C_T value for each sample is determined from each curve as the cycle at which the fluorescence achieves a specific threshold value. The amplification plot shows the increases in fluorescence of 2 samples (A and B). Sample A contains a higher amount of starting template than sample B. Image from http://www.qiagen.com/resources/info/guidelines _rtpcr/dataanalysis_sybr.aspx.

3.1.3 Advantages of qPCR

Nowadays Real-time PCR (qPCR) is the method of election for gene expression analysis, because this is a technique that provides rapid and highly sensitivity detection of the PCR products at the same time as it also generates quantitative results (Guy et al., 2003; VanGuilder et al., 2008).

The most important advantage of real-time PCR when compared to the classical PCR is that real-time PCR allows the determination, with accuracy and high sensitivity, the starting template copy number. The analysis of real-time qPCR results can be either absolute (quantitative), *i.e*. it is calculated the number of copies of a specific DNA sample, or relative (qualitative), *i.e*. it is calculated a presence or absence of a sequence for instance a certain sample has twice as much mRNA of a specific gene as other sample.

Another advantage of qPCR is that while in the classical PCR the detection of the amplified products occurs after the reaction has finished by running an agarose gel electrophoresis, in real-time PCR the detection of the amplified products is done by the incorporation of a fluorescent probe in the PCR reaction mixture that produces a fluorescent signal proportional to the amount of amplified product obtained in each PCR cycle. This is an important benefit because it allows a more precise quantification and eliminates post-PCR processing and procedures which decreases significantly the risk of contamination and minimizes experimental errors. An additional advantage of the qPCR method is that it can reduce the experimental time as it eliminates a gel based analysis of the amplified products.

3.1.4 qPCR specificity analysis

The qPCR specificity analysis corresponds to the analysis of the capacity of the qPCR assay to detect the specific target sequence rather than other nonspecific target sequences that are present in a sample (Bustin et al., 2009). After the completion of the amplification of qPCR run it is recommended to run a melt-curve. The melting curve analysis is an important method for the identification of different reaction products, such as non-specific products and thus is a good measure for the specificity of the primers and the qPCR assay.

A melt-curve analysis is generated by a slow increase of the temperature, in small increments usually of 0.5°C, from 65°C to 95°C while the fluorescent signal is constantly checked at each step. With the increment of the temperature the amplified product denatures, *i.e* the double-stranded DNA dissociates or melts and therefore the fluorescent dye also dissociates from the dsDNA, resulting on the decreasing of the fluorescence. Consequently when the dsDNA melts into it is fully denaturated form it is detected a sudden decrease in the fluorescence. The dissociation temperature depends on the composition and length of the amplification product, so as a consequence it allows verifying the amount of amplification products present in the each reaction.

From the melt curve analysis is obtained a characteristic peak which corresponds to the amplification product melting temperature, the Tm. The Tm is defined as the temperature in the melting curve where 50% of the base pairs of the DNA are double-stranded, and 50% are single-stranded or melted. To visualize the Tm more clearly the negative first derivatives of the change in the fluorescence are often plotted as a function of the temperature, which makes the Tms of the amplification products appear as peaks (*Figure 22*).

An ideal melt curve presents a single dissociation peak with a certain Tm, representing the amplification specific product. If other peak with a different melting temperature is present in the melt curve it may be an indication of the presence of other non-specific products such as primer-dimers. Primer-dimers are products generated when two primers anneal to each other instead of the target sequence of interest. Usually the Tm values of the primer-dimers are considerably lower than the specific product of interest and the amplification product is shorter than the specific product of interest, although the primer dimer peak can sometimes be higher than the peak of the specific product. When primer dimers are amplified, generally it is observed a small product around the temperature of 70°C. Other non-specific products can also be generated when primers anneal to an area different from the region of interest. In these cases, the Tm value and the amplification product size may not differ significantly from the specific product of interest (http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/real-time pcr.html).

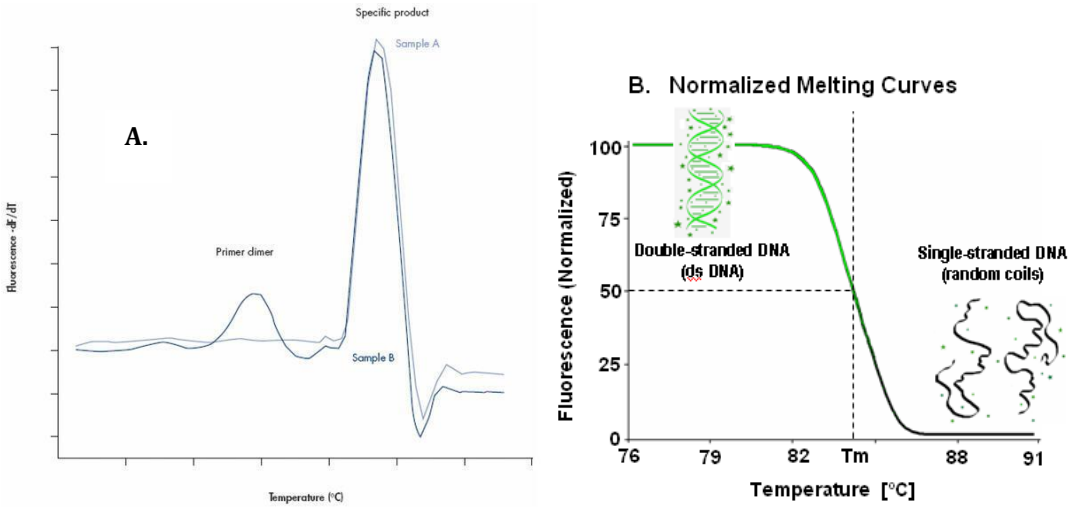


Figure 22. Melting curve analysis. A. Melting curve analysis of 2 samples (A and B). Sample A yields only 1 peak resulting from the specific amplification product (primer-dimers not coamplified). Sample B shows a peak from the specific product and a peak at a lower temperature from amplification of primer-dimers. B. When the sample is heated to high temperatures, the cDNA denatures and the fluorescent color fades away as the double stranded cDNA separates, generating a melting curve. Image adapted from http://www.qiagen.com/resources/info/guidelines_rtpcr/dataanalysis_sybr.aspx.

3.1.5 qPCR optimization

The quantification of the real-time qPCR data is based on the amount of the initial template and the CT value achieved during the amplification. For a precise and reproducible quantification of the samples is absolutely necessary to have an optimal qPCR assay. The precision and robustness of the qPCR assay is usually associated with high PCR efficiency (Bustin et al., 2009).

qPCR amplification efficiency

The PCR amplification efficiency or performance can be determined through calibration curves. These can be constructed from the results obtained by running serial dilutions of a template (*Figure 23*). The calibration curves are very important as they are simple, fast and reproducible indicatives of the PCR efficiency, of the analytical sensitivity and of the robustness of the test (Bustin et al., 2009).

The amplification efficiency, E, is calculated from the slope of the calibration curve using the equation $Efficiency\ (E) = 10^{(-1/slope)} - 1$, obtained from the linear regression between the mean Ct and the log of the starting quantity of template in the dilution series. The amplification efficiency can also be presented in percentage *(% Efficiency = 100*E)*.

An efficiency of 100% indicates that the amount of PCR product was perfectly doubled during each cycle of exponential amplification. The PCR efficiency is acceptable between 80% - 110% and efficiency close to 100% is an indication of a robust and reproducible assay (qPCR Technical Guide, Sigma-Aldrich Co 2008).

In the equation of the linear regression line, the coefficient of determination (R^2) can be used to evaluate the optimization of the qPCR. The R^2 value of a calibration curve is a parameter which

indicates how well the experimental data points lie on the regression line, it indicates how linear the data are and consequently it shows the linearity of the PCR assay. The linearity is a measure of the variability across assay replicates (and if the amplification efficiency is the same for different starting template copy numbers). $R^2 < 0.95$ may be an indication that there is no linear relation between the CT and the 10 log of the DNA concentration or that there must have been imprecise pipetting of the reactions.

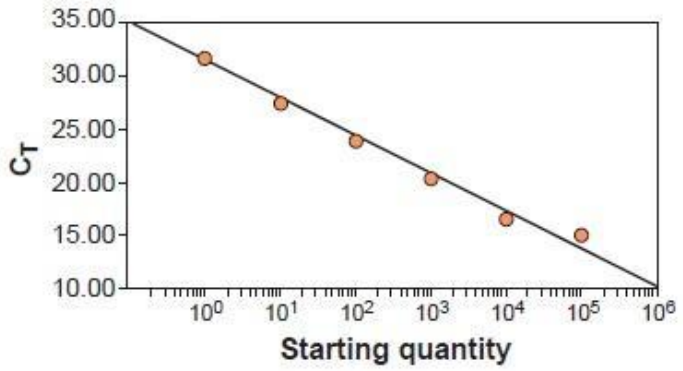


Figure 23. Dynamic range of real-time PCR. Fluorescent detection of the amplification of the gene product is linear over five orders of magnitude. Values were calculated on the basis of CT values from amplification figure. Slope = –3.538; y intercept, 31.806; correlation coefficient R^2 = 0.991. Image adapted from Walker (2002).

Analytical sensitivity

The analytical sensitivity of the assay corresponds to the minimum number of copies in a sample that can be measured in precision with an assay. The comparison of the sensitivity of the assay can be done by comparing the CT values between the same sample, under the same conditions and with the same threshold level. Therefore, the lower the CT value the higher is the assay sensitivity.

Reproducibility

The replicates of an assay are the indication of the reproducibility. It can be calculated the standard deviation between the replicates and when the replicates have a difference higher than 0.5 CT is preferable to avoid these replicates.

3.1.6 Real-Time qPCR Data Analysis

During the Real-time qPCR assays there may be present several uncontrolled variables, such us the amount of the starting material, differences between tissues, individuals or experimental conditions and others. When analyzing and comparing the results of the Real- time qPCR assays these uncontrolled variables must be taken in account and should be corrected using as a method the normalization.

There are two main methods of quantification: *absolute quantification*, where is calculated the exact number of copies of the gene of interest; and *relative quantification*, where the expression of

the gene of interest in a sample is calculated relatively to another gene or another sample, used as a reference.

The use of reference genes for normalization is based on the assumptions that they present an expression level that is constant between all test samples, that is not affected by the experimental conditions and that go through all the qPCR steps with the same kinetics as the target gene (VanGuilder et al., 2008). The most used reference genes are the tubulin and β-actin because these genes are ubiquitously expressed in cells and tissues and their transcription is usually resistant to the experimental conditions (VanGuilder et al., 2008).

The relative quantification is by far the most widely used method as it is easier to measure (VanGuilder et al., 2008). In this method the analysis result is a ratio where the gene expression level is determined by the amount of target gene and a reference gene. The most common method for relative quantification is the $2^{-\Delta\Delta C_T}$ Method, the Livak method (Livak and Schmittgen, 2001). This method is a direct comparison of the CT values between the target gene and the reference gene. Although this method is base the assumption that the PCR efficiencies of both the target genes and reference genes are near 100% and do not differ more than 5-10% of each other. If there is more than 10% of difference in PCR efficiencies, between the reference gene and the target gene, then the $2^{-\Delta\Delta C_T}$ Method should not be used.

When the efficiency of the amplification of the target and reference genes are not similar it should be used the Pfaffl Method (Pfaffl M, 2001). This method is strongly related with the $2^{-\Delta\Delta C_T}$ Method. Using relative quantification to compare multiple samples requires the choice of a calibrator sample. The calibrator sample can be any sample where the expression of the target gene is expressed as an increase or decrease relative to the calibrator. The equation used to determine the expression ratio between the sample and the calibrator was the following:

Ratio = (Etarget) ΔCt target (calibrator - sample) / (Ereference) ΔCt reference (calibrator - sample)

Where:

Etarget = PCR efficiency of the target gene

Ereference = PCR efficiency of the refence gene

ΔCt target (calibrator - sample) = (Ct target) in the calibrator - (Ct target) in the test sample.

ΔCt reference (calibrator - sample) = (Ct reference) in the calibrator - (Ct reference) in the test sample.

For the reference genes is recommended to use more than one because the expression of one reference gene may vary slightly. When several reference genes are used the normalization is made by using the geometrical average. The bases of this method are the use several reference genes to

compensate possible variations in the expression of these genes. In this way, this is a method that confers robustness for a more precise normalization.

Objective

In this research project we attempt to characterize and analyze the gene expression of candidate genes involved in suberin biosynthesis pathway, in cork-differentiating tissue, from *Quercus suber* trees with different cork qualities. To address this question it was performed the quantitative real-time PCR (qPCR) for the nucleic acid quantification, to evaluate the differential gene expression of 15 candidate genes.

3.2 Materials and Metods

3.2.1 Plant material and cork developing tissue sampling

Cork oak trees were sampled in Herdade do Rosal (Fundação João Lopes Fernandes, FJLF), near Montemor-o-Novo, in Portugal. The plant material was collected from cork oaks with about 60 years old, at the moment of the cork extraction, in July 10th of 2008.

Figure 24. Cork developing tissue sampling at Herdade do Rosal (image courtesy by Eng. Pedro Marques, FJLF).

The trees were chosen according to the cork quality, preliminary assessed by visual inspection. As a result, were selected trees presenting cork with good quality, trees with cork with high porosity and trees with other cork defects.

The phellogen was collected at the moment of highest phellogen activity, which coincides with the period of cork extraction. From each sampled tree, two cork planks with a minimum size of 50x50 cm were collected. The phellogen was collected immediately after the cork removal from the tree, by scrapping the underneath surface of the extracted cork planks (*Figure 24*). The collected material was immediately frozen and kept in liquid nitrogen. It were also sampled few young leaves and branches

of the year that were kept in liquid nitrogen. All plant material was stored at -80°C until use.

3.2.2 Cork quality assessment

Cork quality quantitative assessment was done by image analysis in a work performed by Professor José Graça (ISA, UTL). The cork planks were boiled and air dried and later analyzed by image analysis, in their transverse, radial and tangential sections using a CCD- IRIS camera in a Kaiser RS1 Board, with a controlled illumination apparatus, connected to a computer with the *Analysis*® *Software 3.2*. The pores and other defects were analyzed by their number, dimensions, form and distribution.

The overall cork quality classification (*QCor*) was achieved by the image analysis in the tangential section at the middle thickness. The quality parameters considered were: Porosity coefficient (CP), percent of the area of porosity/defects to total area, Average pore or defect area (AM) and Pore or defect frequency per area unit (FR). The formula for the determination of the cork quality (*QCor*) is presented in *Figure 25*.

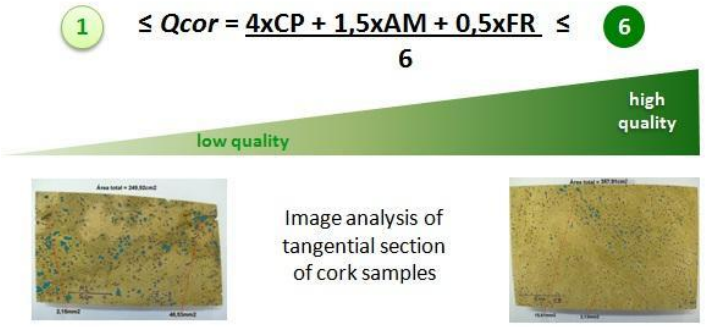


Figure 25. Cork quality assessment. To estimate the cork quality (*QCor*) it was used the presented formula. The values of the cork quality classification (*QCor*) have an interval ranging from 1 to 6. Corks with a low *QCor* value or equal to 1, are classified as corks of low quality. In contrast, higher *QCor* values or equal to 6 correspond to corks classified as being of high quality. The two pictures represent cork planks of the two types of cork qualities and the tangential section where the image analysis were attained. (CP) Porosity coefficient, (AM) Average pore or defect area and (FR) Pore or defect frequency per area unit.

For the study of the gene expression of candidate genes involved in suberin biosynthesis pathway and possible related to cork formation and cork quality, it were selected a set of six samples representing a panel of corks with different qualities, based on image analysis. The selected samples and the cork quality classification are presented in

Table 12.

Table 12. Results of the cork quality assessment of the cork samples using image analysis. The parameters used for the determination of the cork quality are presented together with the final estimation and classification of the cork quality (*QCor*). The cork quality parameters used (CP) Porosity coefficient, (AM) Average pore or defect area and (FR) Pore or defect frequency per area unit.

Samples	CP	AM	FR	*QCor*
C1	2	1	2	2
C2	2	1	3	2
C9	2	1	2	2
C7	1	1	2	2
C8	1	1	1	1
C11	5	2	3	5

3.2.3 Total RNA Extraction

The total RNA was extracted from cork developing tissue, phellem, using the modified RNA extraction protocol of Provost et al. (2007) described in the *Appendix 6*. The RNA quality was evaluated by a denaturing 2% agarose gel electrophoresis in 0.5x TBE buffer. The purity and quantity of RNA were evaluated by the NanoDrop™ 1000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA).

Before the synthesis of the cDNA, the presence of genomic DNA in the samples of the total RNA extracted was checked by PCR. PCR reaction was prepared for each total RNA sample in a final volume of 20µL, containing 1µl total RNA or 10ng of DNA (positive control), 0.2µM of each of the GPAT gene primers (Soler et al., 2008) (forward sequence: 5' TTGACTCGGAGAAAATCAAGCA 3' and reverse sequence: 5' GCTACCGGCACAATTCGATC 3'), 0.2mM of each dNTP, 0.8U GoTaq® Flexi DNA Polymerase (Promega®, Madison, WI, USA), 5mM $MgCl_2$, 1x Green GoTaq® Flexi Buffer (Promega®, Madison, WI, USA), and sterile water to complete 20µL. The thermal cycling parameters were as follows: 1 cycle at 95°C for 5 min and 34 cycles at 95°C for 30 sec, 60°C for 15 sec, 72°C for 1 min and 1 cycle at 72°C for 5 min, using the C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA). The remaining traces of genomic DNA were removed with the kit TURBO DNA-free™ DNase treatment (Ambion Inc., Grand Island, NY, USA), and the procedure followed is briefly described below:

One tenth of 10X TURBO DNase Buffer and 1 µl of TURBO DNase were added to a volume corresponding to a final concentration of 6µg of total RNA. The tubes were mixed gently and then

let incubated for 30 min at 37°C. To inactivate the DNase it were added 0.1 volumes of DNase Inactivation Reagent (resuspended), mixed well and incubated for 5 min at room temperature while mixed several times to redisperse the DNase Inactivation. The next step was the centrifugation for 1.5 min at 10,000 rcf to pellet the DNase Inactivation Reagent. After, the supernatant containing the purified RNA was carefully transferred to a new tube avoid touching the DNase Inactivation Reagent. The presence of genomic DNA was rechecked by PCR with the same conditions as described above. Finally, RNA quality, quantity and purity were checked by 2% denaturing agarose gel electrophoresis in 0.5x TBE buffer and by the NanoDrop ™ 1000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA).

The double-stranded cDNA was synthesized from each sample with a quantity of RNA of 0.54µg using the kit ImProm-II™ Reverse Transcription System from Promega (Promega®, Madison, WI, USA), according to the manufacturer´s instructions and the protocol used is detailed below:

Each reaction of primer combination and denaturation was performed on a final volume of 5µl. For each sample, 0.54µg of total RNA was combined (on ice) with the cDNA Random Primer (0.5µg) in Nuclease-Free Water. Then the mix of the total RNA with the Random primer was denaturated at 70°C for 5 min and directly chilled on ice. A reverse transcription mix was prepared, on ice, for a final volume of 15µl per each cDNA synthesis reaction. The components used were: ImProm-II™ 5x Reaction Buffer at a final concentration of 1x, $MgCl_2$ at 2.5mM, dNTP Mix with final concentration of 0.5mM per each dNTP, 20U of Recombinant RNasin® Ribonuclease Inhibitor, 1µl of ImProm-II™Reverse Transcriptase (Promega®, Madison, WI, USA) and Nuclease-Free Water for a final volume of reaction of 15 µl. This reaction mix was then added to the target RNA- primer combination on ice. Each reaction had a final volume of 20 µl. The tubes were placed on the C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the cycling parameters as the following: 25°C for 5 min (for the annealing of the primers), 42°C for 1 h (for the reaction of extension of the cDNA) and 70°C for 15 min (for the thermal inactivation of the reverse transcriptase). After the synthesis of the cDNA 5µl of each sample immediately stored at -20°C, until use.

3.2.4 Selection of candidate genes for cork formation and cork quality

The 15 candidate genes used for this qPCR study is a subgroup of the 16 genes used for membrane hybridization described in Chapter 1, section 1.1.1 ("BAC library screening for genes of interest").

3.2.5 Real-Time qPCR

3.2.5.1 Primer design

In this part of the study we used the same primers as those described in Chapter 1, section 1.1.3 ("Validation of positive clones by BAC Colony PCR"), except for the reference genes Actin and Tubulin whose pairs of primers were the same as from Soler et al. (2008) Tubulin: forward sequence 5' AAGAACATGATGTGCGCTGCT 3' and reverse sequence 5' TCCACCTCCTTGGTGCTCA 3', Actin: forward

sequence 5' GCCCCACGAGCTGTGTTC 3' and reverse sequence 5' TCTGGCCCATTCCAACCA 3'.

3.2.5.2 Preparation of qPCR reaction and thermal parameters of qPCR

Each reaction was prepared directly on a 96-well PCR plate, using as template 4µl diluted cDNA (1:5) of each sample (C1, C2, C9, C7, C8 and C11), 1X SsoFast™ EvaGreen® Supermix (Bio-Rad, Hercules, CA, USA), 0.5pmol of each forward and reverse primer and RNase/DNase Free Water added to a final volume of 20µl. Three technical replicates were performed for each sample. After sealing the plates with Microseal® 'B' Film (Bio-Rad, Hercules, CA, USA), the plates were placed on the iQ™5 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) with the following thermal parameters: a first step at 95°C for 30sec for the enzyme activation followed by 40 cycles of a denaturation step 1 at 95°C during 5 sec, and with annealing and extension step 2 at 60°C during 10 sec. The fluorescent signal was recorded after the annealing and extension step.

After the amplification cycles it was performed a melting-curve analysis for each gene, to confirm the presence of single specific amplification products. For that, were used the following thermal parameters: slow increase of the temperature in increments of 0.5°C, from 65°C to 95°C with the recording of the fluorescent signal after each temperature increment.

3.2.5.3 qPCR reaction efficiency

The PCR amplification efficiency was calculated using the method of constructing a dilution calibration curve. In this method, the Efficiency (E) of the PCR is based on the slope of a calibration curve of the starting quantity of template (or the dilution factor, for unknown quantities) against the CT value produced by the amplification of each dilution. Two technical replicates of 10-fold serial dilutions of an equimolar cDNA pool of all samples were prepared to build a calibration curve. From the calibration curve were calculated the slope and the coefficient of determination (R^2). The real-time PCR efficiency (E) was calculated according to the equation: $Efficiency\ (E) = 10^{(-1/slope)} - 1$ and later converted into percentage.

3.2.5.4 Quantitative Real-time PCR data analysis

The data results obtained from the qPCR assays were processed and analyzed with the methods available on the *GenEx 5.2.1.3 software* (MultiD Analyses AB, Göteborg, Sweden).

Relative gene expression

Before the relative gene expression calculation, the replicates presenting a difference higher than 0.5 CT were removed and the missing data replaced by the mean of the replicas of the same sample. For the relative gene expression analysis was used the Pfaffl method (Pfaffl, 2001). Efficiency correction for each gene and the normalization was done with the reference genes ACT (actin) and TUB (tubulin) as proposed by Soler et al. 2008. The relative expression of a gene of interest, for each

sample, was calculated relatively to the expression of the sample C8, the sample with the cork developing tissue which presented the lowest cork quality. The expression of sample C8 got an expression value of 1, the samples with higher/lower expression than C8 got and expression superior/inferior to 1 respectively.

Statistical Analysis

The *GenEx 5.2.1.3 software* (MultiD Analyses AB, Göteborg, Sweden) was used for the statistical analysis of the gene expression. It was performed one-way ANOVA to identify significant differences of the expression of a gene among the six samples tested. Were identified significant differences, and thus was performed a post hoc Tukey- Kramer's test for the pairwise comparison analysis.

In order to group genes by their co-expression levels and to group samples by their transcriptomic profile, we performed a hierarchical clustering of the samples and genes using the average linkage, as the clustering method, and the Euclidean distance measure. Both the sample hierarchical clustering and the heatmap were produced using the *GenEx 5.2.1.3 software* (MultiD Analyses AB, Göteborg, Sweden).

It was also determined the correlation among the genes relative expression using *Excel 2007* and the p-value of the correlation was calculated using the *p-Value Calculator for Correlation Coefficients* (http://www.danielsoper.com/statcalc3).

3.3 Results

3.3.1 Total RNA extraction

The total RNA was successfully extracted using the adapted RNA protocol of Provost et al. (2007). The total RNA was extracted from phellem developing tissue of a set of samples with low and high cork quality. It was obtained total RNA with good quality in all the samples, as it presented the typical two bands corresponding to the subunits 28S and 18S of the ribosomal RNA (rRNA) (*Figure* ***26***). The evaluation of the concentration and purity of the total RNA is presented in *Table 13*.

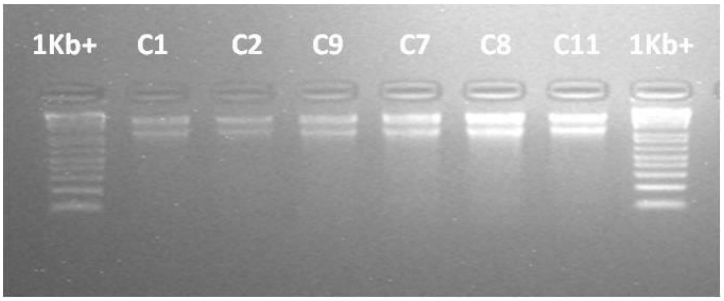


Figure 26. Evaluation of the quality of the total RNA extracted from samples of cork developing tissue of cork oaks from Herdade do Rosal on a 2% agarose gel in 0.5X TBE using 1Kb+ ladder (Invitrogen, Grand Island, NY, USA).

After extraction of the total RNA, in general it was observed the presence of genomic DNA in most of the samples. For that reason it was necessary to proceed with the elimination of genomic DNA from the samples total RNA extracted. The purification kit used to remove the genomic DNA was very efficient because no PCR products were detected after the treatment of the total RNA and only the positive control (genomic DNA) presented amplification (*Figure 27*).

The total RNA extracted presented an average yield of 7.96 µg per 100 mg of tissue. The sample C8 showed the highest RNA yield followed by the sample C11 and C7. The samples C1 and C9 had a yield very similar, while sample C2 presented the lowest yield of total RNA from all the extractions performed.

On average the majority of the samples presented an $A_{260/280}$ ratio close to 2.0 which is an indication that total RNA extracted presented a high level of purity. The ratio $A_{260/230}$ in all samples presented a mean value between 2.20 and 2.30 suggesting the absence of contaminants and therefore confirming the purity of the RNA extracted.

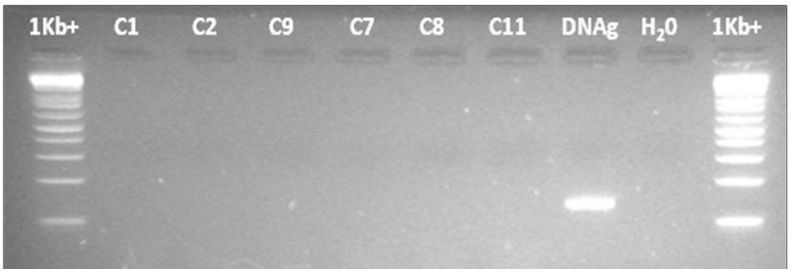


Figure 27. Evaluation of the presence of genomic DNA in total RNA samples after Turbo DNase purification, using 1Kb+ ladder (Invitrogen, Grand Island, NY, USA). (DNAg) corresponds to the positive control with the genomic DNA. (H_20) corresponds to the negative control which contained Milli- Q water instead of template RNA.

Table 13. Quantification of the total RNA yield (µg/100mg of tissue) and purity extracted from samples of cork developing tissue, in several extraction performances, using the NanoDrop.

Sample	Total RNA yield (µg/100mg of tissue)	Mean of ratio $A_{260/280}$	Mean of ratio $A_{260/230}$
C1	7.61	2.04	2.23
C2	6.00	2.06	2.29
C9	6.97	2.05	2.33
C7	8.00	2.05	2.3
C8	11.00	2.04	2.25
C11	8.22	2.08	2.33
Average	7.96	2.05	2.29

3.3.2 Gene expression analysis by qPCR

3.3.2.1 Testing q PCR primers

The qPCR amplification efficiency, of each primer pair tested, was calculated based on the calibration curves method. In *Figure 28*, are presented the dilutions curves constructed for the primers WRKY transcription factor (WRKY) and Diphenol oxidase laccase (Dolaccase). The calibration curves generated for the all primers tested are presented in *Appendix 8*.

The calculation results of the coefficient of determination (R^2) and of the qPCR amplification efficiency (E) for the genes in study is presented in the Table 5. In this study the R^2 values varied from 0.956 in the primer CYP84A1 and 0.995 in the primer GDSL, The PCR efficiency values varied between 80 and 110%. Excepting for the genes CCR (111.6%), CYP84A1 (114.1%), ACC-OX (113.1%) and ACT (111.2%) which presented a value slightly above 110%, the rest of the genes presented an efficiency close to 100%.

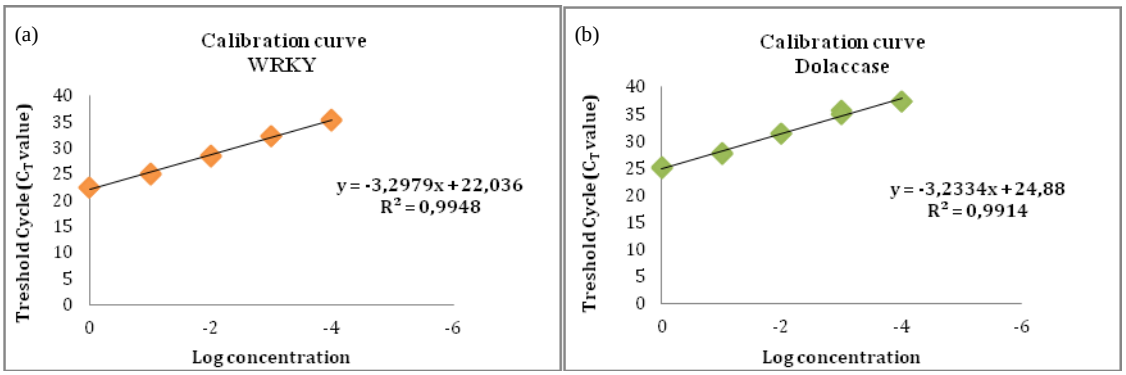


Figure 28. Calibration curves for the primers (a) WRKY transcription factor (WRKY) and (b) Diphenol oxidase laccase (Dolaccase). The Threshold Cycle (CT) is plotted as function of the 10 log concentration of the dilution series. The regression line from the dilution curve was used to determine the efficiency (using the slope value) and the optimization (using the R^2 value) of the qPCR assay.

Based on the melting curve analysis we could observe that the amplification of all the 17 genes was very specific, as only one peak was obtained and it was not found the amplification of other non-specific products. In

Table ***14*** is summarized the result of the evaluation of the melt curve analysis for all the genes. The Melt Peak Charts of allthe genes analyzed are presented in *Appendix 7*. As an example, in *Figure 29* is presented the result of the melt curve analysis of the gene Diphenol oxidase laccase (Dolaccase). This Melt Peak chart shows the presence of only one peak corresponding to the amplification of the specific product, with a melting temperature (Tm) of 77.0°C.

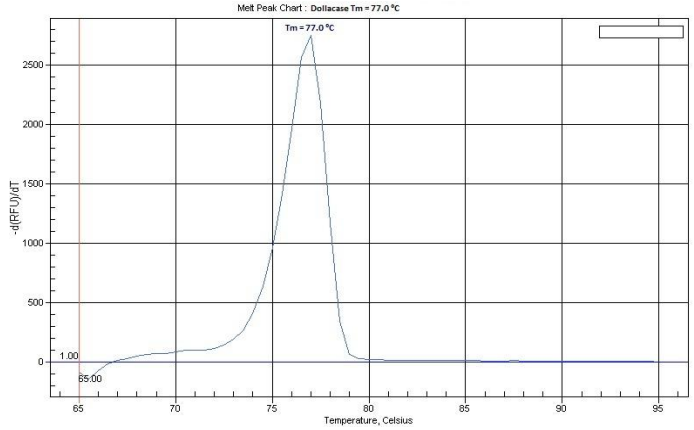


Figure 29. Melt Peak Chart of the primer Diphenol oxidase laccase (Dolaccase) with one peak indicating the melting temperature (Tm) value of one single PCR product. The peak corresponds to the specific amplification product with a temperature of melting (Tm) of 77.0°C. In this reaction did not occur co- amplification of other nonspecific PCR products such as primer-dimers. (T) Temperature and (RFU) Relative fluorescence unit.

Table 14. Optimization of qPCR reaction. This table presents the results from the calculation of the coefficient of determination (R^2) and the percentage of the qPCR amplification efficiency (E) obtained from the calibration curves generated for the primer testing. It also lists the results of the melting curve analysis with the melting temperature (Tm) and the amplicon size of the cDNA obtained from the tested primers.

Primer name	E (%)	R^2	Melting Temperature Tm (°C)	Amplicon size (bp)
DHCA-S-AT	97.8	0.991	79.50	150
LACS	80.4	0.989	84.50	150
ELTFP	84.3	0.988	82.50	200
CYP86A1	101.0	0.992	80.00	150
GPAT	93.9	0.990	80.00	200
GDSL	88.4	0.995	78.00	200
CCR	111.6	0.989	77.50	200
4CL	83.0	0.988	80.50	200
CYP84A1	114.1	0.956	77.50	200
HCBT	97.6	0.963	84.00	400
DoLaccase	103.8	0.991	77.00	200
WRKY	101.0	0.990	78.50	150
ACC-OX	113.1	0.991	81.50	150
CYP87A2	84.1	0.990	78.00	150
APX	109.4	0.990	80.00	150
Reference genes				
ACT	111.2	0.978	85.00	200
TUB	95.3	0.995	84.50	150

3.3.2.1 Real time qPCR data analysis

The bar plots in *Figure 30* to *Figure 33* represent the gene expression level of each sample relative to the sample C8. The statistical analysis results are presented in *Table 25 (Appendix 9).*

Dihydrolipoamide S-acetyltransferase (DHCA-S-AT) - This gene showed significant lowest expression values in the samples C1, C2 and C9 relative to C8. In contrast the samples C7 and C11 presented a gene expression with no significant difference from sample C8.

Long-chain acyl CoA syntahse (LACS) - For the gene LACS, the relative gene expression presented the lowest values in the samples C1 and C2. The samples C9 and C7 showed no significant differential expression relatively to C8. While the sample C11 presented the highest level of gene expression for this gene and with a significant difference.

Esterase/ lipase/ thiosterase family protein (ELTFP) - The lowest values of relative gene expression were showed by the samples C1 and C2 with significant difference, while the samples C9 and C7 did not present significant differential expression relatively to C8. The highest value of gene expression was presented by the sample C11 although with no significant difference when comparing to the sample C8.

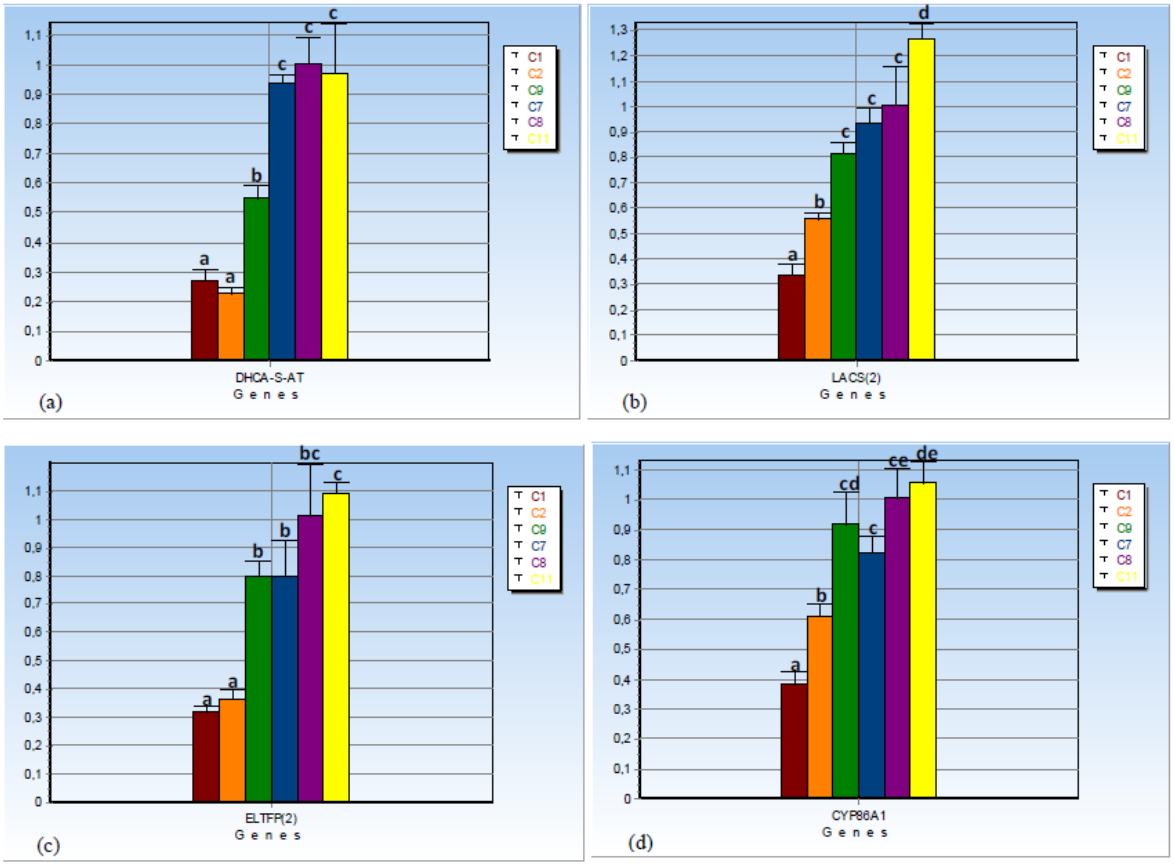


Figure 30. Bar plots of relative gene expression levels of the genes (a) Dihydrolipoamide S- acetyltransferase (DHCA-S-AT), (b) Long-chain acyl CoA syntahse (LACS), (c) Esterase/ lipase/ thiosterase family protein (ELTFP) and (d) Fatty acid ω-hydroxylase (CYP86A1). The bar plot values for each sample consist in three replicates, and the error bars correspond to their respective standard deviations. All plotted data represent gene expression levels relative to sample C8. Same letters over error bars indicate no significant differences of gene expression levels between pair of samples ($p < 0.05$).

Fatty acid ω-hydroxylase (CYP86A1) - In the analysis of the gene CYP86A1 the sampleC1 and C2 showed the lowest significant relative expression values while the rest of the samples did not present an expression level significantly different from C8.

Glycerol– 3 - phosphate acyltransferase (GPAT) - For the gene GPAT the samples C1, C2, C9 and C7 presented significant lower expression levels relatively to C8. The pair of samples C2 and C7, C9 and C7 did not present significant differences in the level of gene expression among them. In contrast the sample C11 did not present a significant differential expression from C8.

GDSL–motif lipase/hidrolase protein like (GDSL) - For this gene the samples C7 and C11 did not present differential expression values relatively to C8. The sample C1, C2 and C9 showed significant lower values of gene expression than C8.

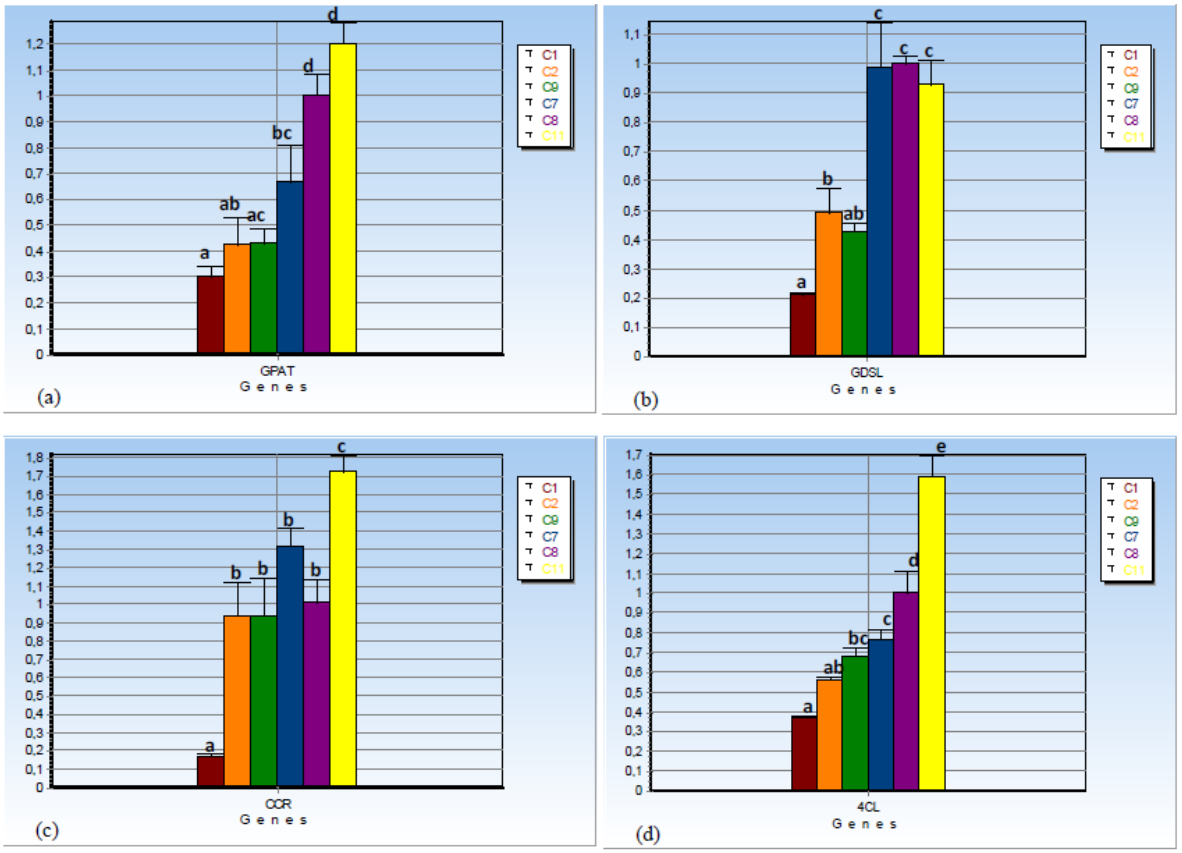


Figure 31. Bar plots of relative gene expression levels of the genes (a) Glycerol– 3 - phosphate acyltransferase (GPAT), (b) GDSL–motif lipase/hidrolase protein like (GDSL), (c) Cinnamoyl CoA reductase (CCR) and (d) 4-Coumarate: CoA ligase (4CL). The bar plot values for each sample consist in three replicates, and the error bars correspond to their respective standard deviations. All plotted data represent gene expression levels relative to sample C8. Same letters over error bars indicate no significant differences of gene expression levels between pair of samples ($p < 0.05$).

Cinnamoyl CoA reductase (CCR) - For the gene CCR, the samples C1 and C11 showed respectively the lowest and highest significant gene expression values relatively to C8. The rest of the samples did not present a significant differential gene expression from C8.

4-Coumarate: CoA ligase (4CL) - The highest significant relative expression of this gene was presented by the sample C11. The remaining samples had significantly lower levels of expression than C8, and C1 presented the lowest gene expression value not significantly different from C2. The pair of samples C2 and C9, C9 and C7 showed slight increase on the gene expression level but with no significant differences among them.

Cinnamoyl CoA reductase (CCR) - For the gene CCR, the samples C1 and C11 showed respectively the lowest and highest significant gene expression values relatively to C8. The rest of the samples did not present a significant differential gene expression from C8.

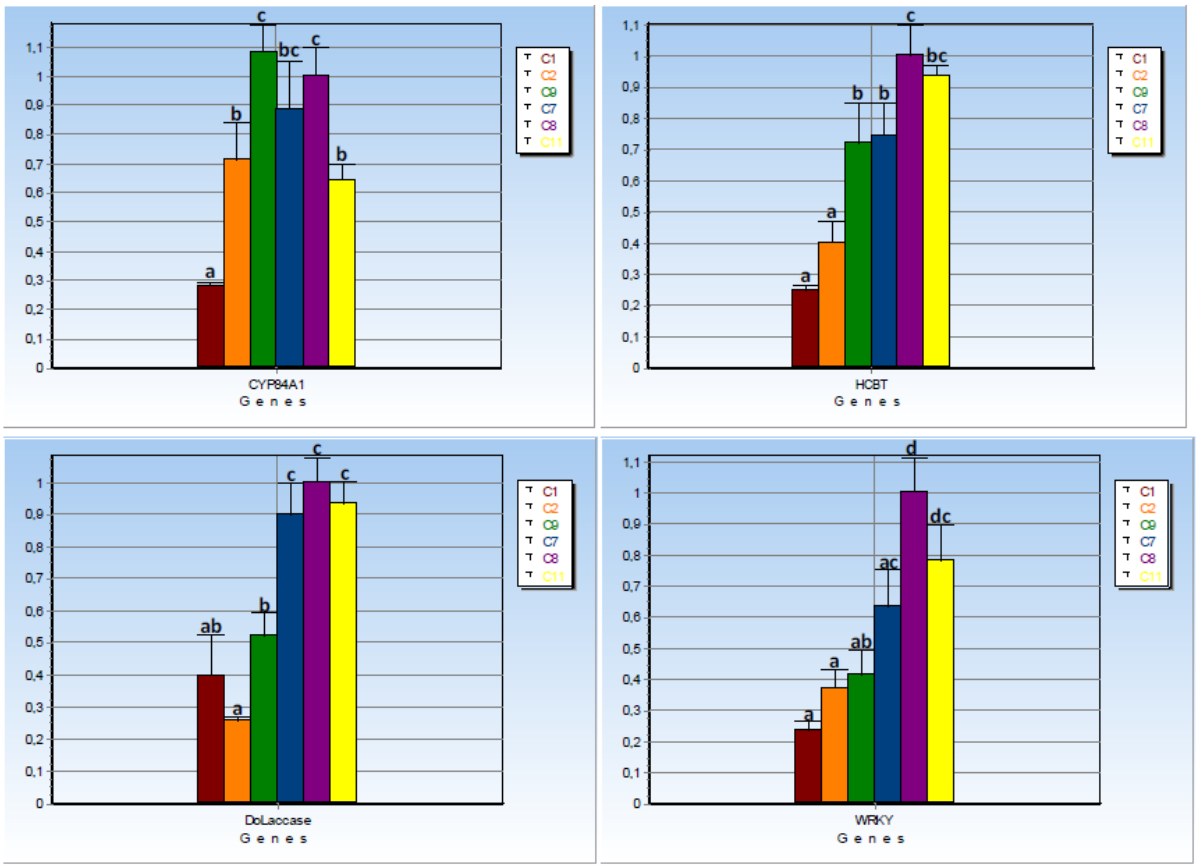


Figure 32. Bar plots of relative gene expression levels of the genes (a) Ferulate-5-hydroxylase (CYP84A1), (b) Hydroxycynamoil-CoA/benzoyltransferase (HCBT), (c) Diphenol oxidase laccase (Dolaccase) and (d) WRKY transcription factor (WRKY). The bar plot values for each sample consist in three replicates, and the error bars correspond to their respective standard deviations. All plotted data represent gene expression levels relative to sample C8. Same letters over error bars indicate no significant differences of gene expression levels between pair of samples ($p < 0.05$).

Ferulate-5-hydroxylase (CYP84A1) - For the gene CYP84A1 the sample C1 presented the smallest value of relative expression which was significantly different from all the other samples. The samples C2, C7 and C11 showed a significant lower expression level than C8. The sample C9 presented the highest values of the gene expression but not significantly different from the sample C8.

Hydroxycynamoil-CoA/benzoyltransferase (HCBT) - The lower values of gene expression are presented by the samples C1 and C2 with a significant difference when comparing with C8. The samples C9 and C7 have significantly higher levels of gene expression than the pair of samples C1 and C2, but significantly lower expression relatively to sample C8. The value of gene expression for the sample C11 was not significantly different from the sample C8.

Diphenol oxidase laccase (Dolaccase) – The samples C7 and C11 showed the highest value for the relative gene expression with no significant differences between them although the expression level was not significantly different from sample C8. On the opposite the sample C2 presented the lowest gene expression value not significantly different from the sample C1 and the sample C9 had a gene expression level similar to the sample C1. The group of the three samples C1, C2 and C9 has a significantly lower

expression of the gene Dolaccase relatively to C8.

WRKY transcription factor (WRKY) – For the gene WRKY all the samples showed significantly lower levels of expression than C8, with the exception of the sample C11 which did not present significant differential expression. The highest gene expression value is presented in the sample C11 followed by the sample C7 and between both were not found significant differences in the gene expression. The samples C1, C2 and C9 showed with no significant differences the lower values of expression of this gene.

Cytochrome P450 family protein (CYP87A2) – For this gene is notable the high values of expression showed by the samples C11 and C9 which are significantly higher relatively to sample C8. The rest of the samples (C1, C2 and C7) did not present differential expression of this gene when comparing with C8.

Ascorbate peroxidase (APX) – The lowest gene expression was presented by the sample C1 and the highest expression level was found in sample C7, both with values significantly different from C8. The other samples (C2, C9 and C11) showed expression levels not significantly different from C8.

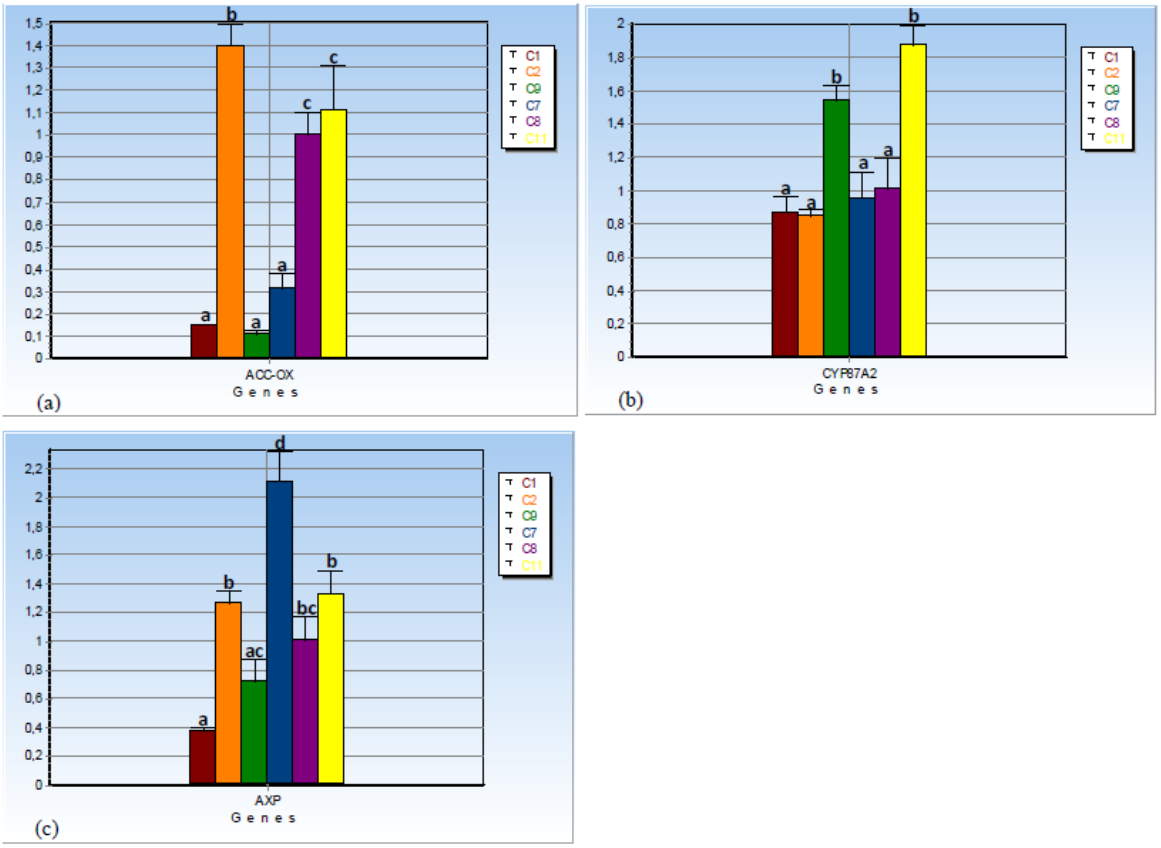


Figure 33. Bar plot of relative gene expression levels of the genes (a) Aminocyclopropane-1-carboxylic acid oxidase (ACC-OX), (b) Cytochrome P450 family protein (CYP87A2) and (c) Ascorbate peroxidase (APX). The bar plot values for each sample consist in three replicates and the error bars correspond to their respective standard deviations. All plotted data represent gene expression levels relative to sample C8. Same letters over error bars indicate no significant differences of gene expression levels between pair of samples ($p < 0.05$).

With the heatmap analysis is possible to have a visual way to identify patterns in the gene expression across various samples and groups of genes. The heatmap, created for the genes and

samples analyzed, is shown in *Figure 34*.

This analysis showed that for all samples used in the study, the three replicates of each sample were clustered together. Furthermore, samples were clustered into two larger groups. The Group I, grouping the samples C1 and C2 and Group II, clustering the remaining samples. This last group could be subdivided into two subgroups linking the pair of samples C9 and C7 in subgroup IIa and the other subgroup IIb grouping the pair of samples C8 and C11.

Genes were also clustered in two major groups, which are divided in subgroups. The first major group (Group A) links the genes Fatty acid ω-hydroxylase (CYP86A1), Long-chain acyl CoA syntahse (LACS), 4-Coumarate: CoA ligase (4CL), Cinnamoyl CoA reductase (CCR), Cytochrome P450 family protein (CYP87A2) and Ascorbate peroxidase (APX), while the second major group (Group B) contains the remaining genes. For purposes of simplicity Group B was divided in three main subgroups. In subgroup B1 the genes GDSL and GPAT are linked with Diphenol oxidase laccase (Dolaccase), Dihydrolipoamide S-acetyltransferase (DHCA-S-AT) and WRKY transcription factor (WRKY). Subgroup B2 clusters the genes Esterase/ lipase/ thiosterase family protein (ELTFP), Hydroxycynamoil-CoA/benzoyltransferase (HCBT) and Ferulate-5-hydroxylase (CYP84A1).Whereas subgroup B3 corresponds to the gene Aminocyclopropane-1-carboxylic acid oxidase (ACC-OX).

The results of the study of the correlation among the analyzed genes are presented in *Table 15*. The correlation of the expression of gene ACC oxidase (ACC-OX) and gene Ferulate-5-hydroxylase (CYP84A1) with the expression of the other genes varied between 0.077 to 0.456 and 0.206 to 0.715, respectively, but were not significant. While the relative expression of the gene APX only presented a significant correlation of 0.734 (p-value = 0.048) with the gene GDSL. Additionally the expression of the gene CYP87A2 showed a significant correlation with 4CL (0.761), p = 0.039. The expression of the gene CCR was significantly correlated with most of the genes excepting Diphenol oxidase laccase (Dolaccase) and WRKY. The gene CYP86A1 showed expression levels with a significant correlation with most of the genes excepting the genes CYP87A2 and AXP. For the gene 4CL, were found expression levels significantly correlated in addition to the genes described above, also with the genes Diphenol oxidase laccase (Dolaccase) (p = 0.046), Dihydrolipoamide S-acetyltransferase (DHCA-S-AT) (p = 0.038), Esterase/ lipase/ thiosterase family protein (ELTFP) (p = 0.014), LACS (p = 0.004), GPAT (p = 0.017), WRKY (p = 0.043) and HCBT (p = 0.025). The gene Diphenol oxidase laccase (Dolaccase) presented relative expression levels significantly correlated with all genes excepting the genes: APX, ACC-OX, CCR, CYP84A1 and CYP87A2.

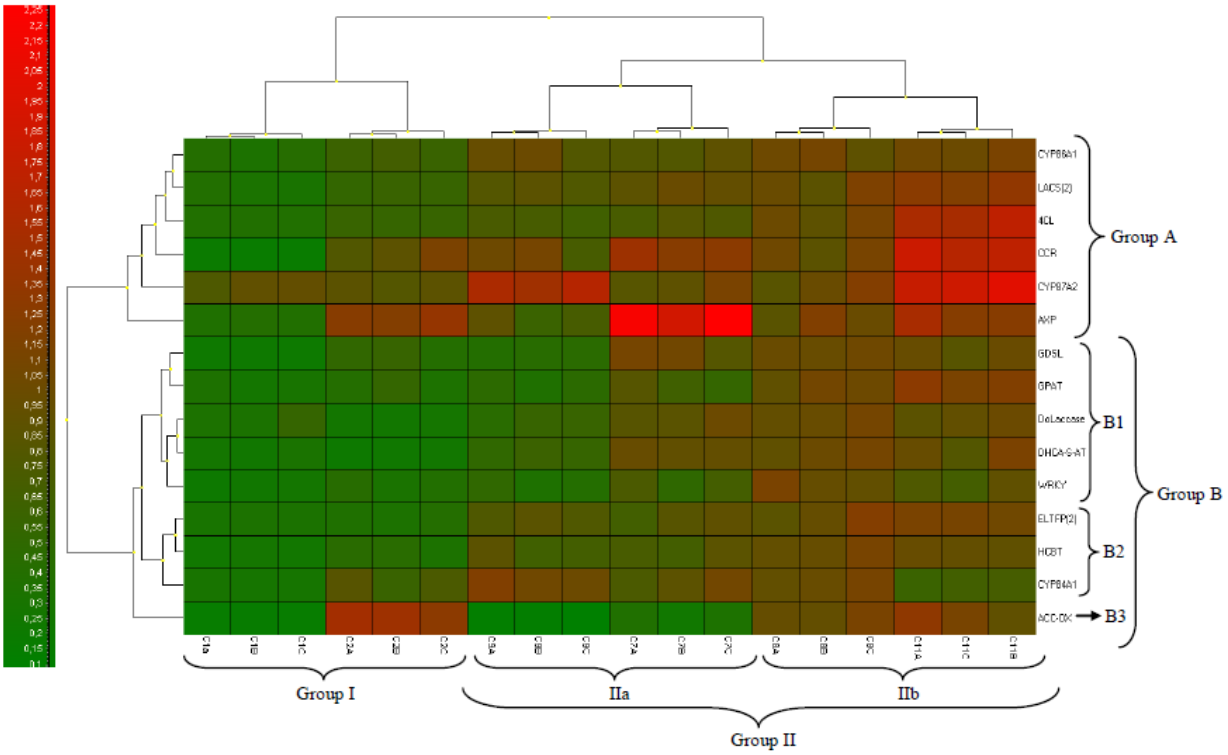


Figure 34. Heatmap of relative gene expression of developing phellem associated genes. The heatmap clusters different samples of cork developing tissue and the candidate genes potentially associated with cork quality and formation. Gene names are listed on the right and the sample names at the bottom. The columns represent replicates from cork developing tissue samples, and the rows represent the genes analyzed. Relative gene expression levels are depicted according to the color scale at the left. Green indicates down-regulation and red indicates up-regulation of the gene. On the top of the diagram is shown a hierarchical clustering order for individual samples and on the left for the candidate genes.

Table 15. Correlation matrix among the relative expression of the genes studied by qPCR (n=6, (*) p-value < 0.05).

Genes	*ACC-OX*	*CCR*	*CYP84A1*	*CYP86A1*	*CYP87A2*	*4CL*	*Dolaccase*	*DHCA-S-AT*	*ELTFP*	*LACS*	*GDSL*	*GPAT*	*WRKY*	*APX*	*HCBT*
ACC-OX	1														
CCR	0,437	1,000													
CYP84A1	0,018	0,450	1,000												
CYP86A1	0,257	0,824*	0,715	1,000											
CYP87A2	0,042	0,631	0,209	0,686	1,000										
4CL	0,456	0,852*	0,206	0,823*	0,761*	1,000									
Dolaccase	0,077	0,649	0,392	0,772*	0,360	0,739*	1,000								
DHCA-S-AT	0,097	0,729*	0,485	0,837*	0,417	0,764*	0,990*	1,000							
ELTFP	0,152	0,779*	0,575	0,963*	0,674	0,861*	0,894*	0,929*	1,000						
LACS	0,302	0,913*	0,519	0,952*	0,693	0,924*	0,853*	0,903*	0,965*	1,000					
GDSL	0,376	0,798*	0,481	0,768*	0,229	0,725	0,904*	0,926*	0,810*	0,857*	1,000				
GPAT	0,493	0,782*	0,243	0,811*	0,554	0,952*	0,857*	0,858*	0,877*	0,910*	0,848*	1,000			
WRKY	0,416	0,645	0,497	0,819*	0,284	0,751*	0,903*	0,902*	0,866*	0,835*	0,910*	0,901*	1,000		
APX	0,252	0,717	0,341	0,376	-0,017	0,341	0,475	0,537	0,351	0,515	0,734*	0,379	0,396	1,000	
HCBT	0,239	0,766*	0,671	0,971*	0,558	0,810*	0,886*	0,925*	0,981*	0,945 *	0,855*	0,864*	0,920*	0,393	1

3.4 Discussion

3.4.1 Nucleic Acid extraction

Total RNA was extracted from cork developing tissue of cork oaks with different cork qualities, classified after phenotypic characterization. Using the adapted protocol from Provost et al. (2007) we were able to perform successfully several extractions of total RNA which presented high quality. From all the samples C8 presented the highest total RNA yield. From the spectrometric analysis it was verified good values for both the ratios $A_{260/280}$ and $A_{260/230}$, indicating a high level of purity and the nonexistence of contaminants on the samples of extracted total RNA. Although as highlighted before when mature, the phellem is mainly constituted by dead cells and it contains high quantities of phenols and tannins, and this can make cork RNA extraction and downstream processes as well as some biochemical and molecular analyses become particularly difficult.

After the total RNA extraction it was observed the presence of genomic DNA in the majority of the samples. The methodology used to remove genomic DNA from total RNA samples revealed to be very efficient since it was not observed remaining traces of genomic DNA contamination.

3.4.2 qPCR analysis

3.4.2.1 Testing qPCR primers

The calibration curves generated allowed the assessment of the qPCR efficiency. The R^2 values calculated from the calibration curves of the primers were around 0.99 for almost of the primers, excepting CYP84A1 (0.956), HCBT (0.963) and ACT (0.978). These results suggest a high linearity of the PCR assay which means that there was low variability across the assay replicates. The PCR efficiency varied between the acceptable interval of values from 80 to 110%, with the exception of four primers which presented an efficiency value slightly superior than 110%. Nevertheless, the majority of the primers showed efficiency values close to 100%, which indicates that the amount of PCR product was nearly doubled during each cycle of exponential amplification and therefore the assay can be considered robust and reproducible. The melt curve analysis all the primers studied presented the amplification of a single specific product and it was not found the amplification of other non-specific products. Therefore, from the melt curve analysis the qPCR assay could be considered to have a reasonable good specificity to proceed with the study.

3.4.2.2 Expression profile of suberin related genes in developing phellem associated with different cork qualities

The developing phellem tissue samples were collected in a small homogeneous edapho- climatic area at Herdade do Rosal (Montemor-o-Novo). Although trees were located closer to each other, cork in these trees showed very contrasting macroscopic cork quality classes, suggesting some degree of genetic control of the cork quality parameters such as the porosity coefficient, the area of pores or other defects, and the pores or other defects frequency per area unit.

From the cluster analysis, based on the expression profile of the genes analyzed, the hierarchical clustering of the samples showed the three replicates of each sample clustered in the same group. This is an indication of a high reproducibility of the qPCR assay.

From the results of the cork quality assessment it would be expected that samples with similar scores for the cork quality parameters (*QCor*) would be clustered similarly as that by transcription profile. However, we observed that the sample clustering based on the expression profile of the genes analyzed in this study is not congruent with their cork quality image analysis classification. Indeed, the sample C8 (quality class = 1) and sample C11 (quality class = 5) were clustered together, which was an unexpected result. The genes used in this study were not able to discriminate the different quality classes obtained by image analysis.

Since clustering based on gene analysis results was very different from that based on image classification of cork class quality, it would be necessary further biochemical analysis of the cork developing tissue. Indeed, almost of the genes analyzed in this study are related to suberin biosynthesis. Therefore this clustering may reflect the differences/similarities on suberin biosynthesis/composition, and not the sample macroscopic cork quality class. The biochemical analysis of the cell wall of cork cells from developing cork tissue with different qualities could give a better knowledge on the different components of the cork cells of the different samples.

3.4.2.3 Gene expression correlation between suberin biosynthesis related gene

For the biosynthesis of suberin are involved two main metabolic pathways. The first is the acyl-lipid pathway involved in the biosynthesis of the suberin aliphatic domain that consists essentially in the fatty acid elongation (acyl elongation). The other one is the phenylpropanoid pathway, involved in the biosynthesis of the suberin aromatic domain, which includes the ω- carbon oxidation (acyl oxidation) (Franke and Schreiber, 2007).

The enzyme **dihydrolipoamide S-acetyltransferase (DHCA-S-AT)** is involved in the acyl- lipid pathway, in the synthesis of the aliphatic monomers specifically in *de novo* fatty acids biosynthesis in the chloroplast. The enzyme synthesizes long chain fatty acids (LCFA) and very long chain fatty acids (VLCFA), and it is known that VLCFA are precursors of waxes. The lipase role of **esterase/lipase/thioesterase family protein (ELTFP)** it has not been established in plants (Soler et al., 2007), although it can have a possible lipase function in cork lipid metabolism. In cork-oak, Soler et al. (2007) found that these genes were highly expressed in phellem when compared with xylem developing tissues. In both genes we found a strong and significant correlation of their relative expression with the other genes of the acyl-lipids pathway, as well as with the genes **CCR, 4CL, Diphenol oxidase laccase, and HCBT genes**

Long-chain acyl CoA synthase (**LACS), Fatty acid ω -hydroxylase** (**CYP86A1)** and **Glycerol-3-**

phosphate acyl transferase (GPAT) are candidate genes for the synthesis and assembly of the aliphatic polymer, in the acyl-lipid pathway of the biosynthesis of suberin, see *Figure 35*. Pollard et al. (2008) have proposed three different possible reaction pathways for the lipid polyester biosynthesis catalyzed by these genes. These genes are related with two of the three most important modification reactions of the fatty acyl chains, the acyl activation, the ω-oxidation and esterification. From the three reaction results the production of ω- oxidized acylglycerols, which are presumed structural elements of the lipid polymers (http://lipidlibrary.aocs.org/plantbio/polyesters/index.htm) The **LACS** enzymes are involved in the acyl activation to coenzyme A thiosterase. The gene LACS encodes the long-chain acyl CoA synthase enzymes, involved on the exportation of the fatty acids from the chloroplast during the synthesis of the aliphatic polymer (Soler et al., 2007). The gene CYP86A1 encodes the enzyme fatty acid ω -hydroxylase which is a hydroxylase able to catalyze the ω- hydroxylation of C12-C18 fatty acids (Benveniste et al., 1998) and is thought to provide fatty acid precursors for the suberin polyester (Franke and Schreiber, 2007). In the reaction of esterification is produced glycerol-3-phosphate (acyl glycerols) by the acyltransferases of the family of GPAT enzymes. The gene **GPAT** encodes the enzyme glycerol-3-phosphate acyltransferase and in Arabidopsis ortholog of cork GPAT has been confirmed to catalyze the formation of ester bonds between fatty acids and glycerol (Beisson et al., 2007) In this study, we found a strong and significant ($P < 0.05$) correlation on the expression of these three genes.

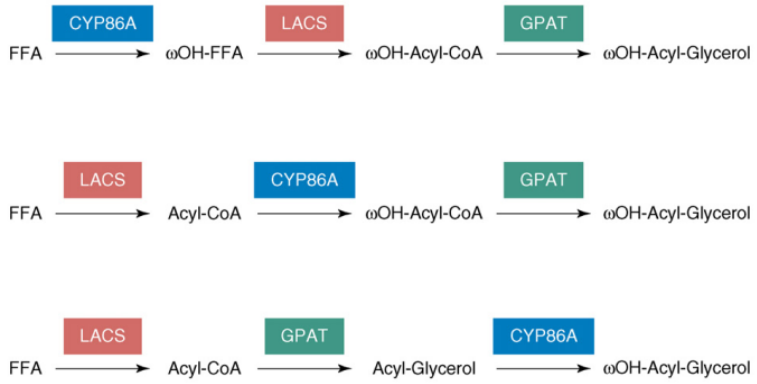


Figure 35. Possible reactions for the lipid polyester biosynthesis pathway catalyzed by LACS, GPAT and CYP86A1, as proposed by Pollard et al. (2008). The three schemes represent the possible metabolic pathways involved in the synthesis of a polyester building block. For reasons of simplicity, only a ω-hydroxy fatty acid-containing building block is represented.

GDSL-motif lipase/hydrolase family protein (GDSL) is involved in the assembly of the aliphatic polyester. Soler et al. (2007) analyzed the expression of 6 members of this family, and found that these genes were more expressed between 3-49 fold in cork tissues than in wood. The **hydroxycynamoil-CoA/ benzoyltransferase (HCBT)** belongs to the BAHD acyltransferase superfamily (D'Auria, 2006) which catalyzes the synthesis of phenylpropane derivatives, N- hydroxycinnamoyl amides, characteristic of suberin such as feruloyltyramine (Yang et al., 1997).There are evidences that feruloyltyramine is a component of the aromatic suberin of cork oak (Marques et al., 2002). In the study of Soler et al. (2007) were analyzed two members of this family and it was showed that these genes were 32 to 36 fold more expressed in cork than in wood. The genes GDSL and HCBT may also be

involved in esterifications of suberin monomers involving ferulates and hydroxylated fatty acids. And in this work we found a significant ($P < 0.05$) and strong correlation between these two genes.

4-Coumarate: CoA ligase (4CL) and Cinnamoyl CoA reductase (CCR) and Ferulate-5- hydroxylase (F5H, CYP84A1) are involved in the phenylpropanoid metabolism which is the base of the biosynthetic pathway of the aromatic monomers of suberin. From the phenylpropanoid metabolism results two types of products, the hydroxycinnamic acids and monolignols. In this work, it was found a significant ($P < 0.05$) high correlation between the genes 4CL and CCR, while it was not found a significant correlation of both genes with CYP84A1 (F5H). The enzyme F5H is thought to be important for the synthesis of the hydroxycinnamic acids which is one of the resulting products from the phenylpropanoid metabolim and a typical component of suberin (Bernards and Lewis, 1998). In this work we found high correlation value between F5H and CYP86A1, although this was not significant. The strong correlation found between the genes 4CL and CCR, CYP84A1 and CYP86A1 may suggest the importance of the phenylpropanoid pathway for the suberin biosynthesis.

The enzymes **diphenol oxidase laccases (Dolaccase)** are extracellular oxidases that are capable to combine the phenylpropapnoids. Based on the proposed role of laccases in lignin synthesis (Kiefer-Meyer et al., 1996; LaFayette et al., 1999; Ranocha et al., 2002; Ehlting et al., 2005) it can be supposed an analogous function in the synthesis of the aromatic part of suberin (Liang et al., 2006). In cork-oak, this gene was found highly expressed in developing cork when compared with wood (Soler et al., 2007). Here we found significant high correlation of this gene with 4CL and HCBT also from the phenylpropanoid pathway genes and with those of the lipid-acyl pathway.

Regulatory proteins

The gene **WRKY** encodes for a transcription factors that binding to W-boxes (cis-regulatory elements) can modulate the gene expression of some stress induced genes, including P450 monooxigenases (Mahalingam et al., 2003; Narusaka et al., 2004). The WRKY transcription factors constitute a large family of plant-regulators that in general control senescence, stress and defense responses (Eulgem et al., 2000). Nevertheless the knowledge of the specific function of these regulatory genes in cork regulation is still very scarce. Soler et al. (2007) found WRKY transcription factor highly expressed in cork when comparing with xylem developing tissue. In this work it was found a significant correlation on the relative expression of this transcription factor with the genes involved in the acyl lipids pathway (DHCA-S-AT, LACS, ELTFP, CYP86A1, GPAT and GDSL) and with 4CL and HCBT from the phenylpropanoid pathway. This may be an indication of a possible role of this gene as a regulator of the expression of genes involved in suberin biosynthesis.

The gene **ACC oxidase (ACC-OX)** encodes the ethylene forming enzyme aminocyclopropane-carboxylate (ACC) oxidase. Even though the function of ethylene in the cork tissue is not clear (Andersson-Gunnerås et al., 2003; Lulai and Suttle, 2004), it could have a possible role of regulation in cork. It was shown (Soler et al., 2007) a highly expression of this gene, but with no significant differences, both in cork and

wood. Although with our work, we did not find a significant relative expression correlated with any other genes analyzed and the correlation values obtained were very low.

Miscellaneous and Stress

The gene **CYP87A2** is described as a monoxygenase belonging to the cytochrome P450 superfamily. While the gene **APX** encodes the ascorbate peroxidade, an important enzyme for H_2O_2 detoxifying in plant cells (Davletova et al., 2005). In previous works (Pla et al., 1998, 2000) it was shown that cork cells present high levels of oxidative stress. Soler et al. (2007) also found these genes more highly expressed in cork than in wood. Therefore the up- regulation of these two genes may suggest that during the cork cell metabolism are produced high quantities of reactive oxygen species (ROS). In this work it was found a significant correlation on the relative expression of CYP87A2 with 4CL and of APX with GDSL.

Conclusions

In this study we characterized the expression of 15 genes associated with suberin byosintheis and cork formation, in six cork developing tissues samples with different cork quality classes, ranging from low quality to high quality, by qPCR. The robustness of the assay was high, as the three technical replicates of each sample were clustered together. It was found a significant correlation on the relative expression of the genes from the acyl lipids pathway with the genes from the phenylpropanoids pathway of the suberin biosynthesis. It was also found a significant correlation of the transcription factor WRKY with the genes from both suberin biosynthesis pathways supporting the possible function of this gene as a regulator of the expression of genes involved in the biosynthesis of suberin. In addition it was shown up-regulation of genes encoding for enzymes related with the oxidative stress and their relative expression was significantly correlated with some genes from both pathways of suberin biosynthesis, suggesting the possible production of reactive oxygen species (ROS) during the metabolism of the cork cells. However, in this work it was not possible to relate the classification of cork quality obtained from image analysis of the samples with the gene expression profiles derived from the selected group of genes analyzed.

Thus for a better characterization of the genes potentially involved in cork formation and cork quality traits it would be interesting to analyze other group of genes. One alternative could be to perform deep-sequencing of the transcriptomes of developing cork tissue associated to different cork quality classes. By this way, we could investigate the expression of other genes expressed during the formation of the cork cell wall, and their molecular regulation, for further development of molecular markers for the trait of cork quality. Furthermore, a detailed study of the metabolites and suberin composition, such as the study of lipids and proteins accumulated during cork formation and involved in suberization processes, could advance our knowledge on cork differentiating tissue from trees with contrasting cork qualities.

4. General conclusions

The *Quercus suber* BAC library was screened for the presence of several genes of interest, through hybridizations to high-density BAC filters followed by PCR validation, confirming the utility of use this new resource on gene identification and characterization. This library allowed to find genes involved in cork formation and suberin biosynthesis pathways, and presented a very low level of extranuclear genome contamination. In addition, this BAC library enabled analysis of BAC end sequences and cytogenetic studies using BAC clones. This *Q. suber* BAC library represents an important landmark for the advancement of *Fagaceae* genomics and forest tree species research, including genome sequencing, gene isolation, functional and comparative genomics.

BAC-end sequencing is an important tool for improving the value of BAC libraries as genomic resources. In this study, we carried out a preliminary analysis of cork oak BAC end sequences. Our results showed that the cork oak BESs presented a GC content of 37.76% and contained a relatively small proportion of the known repetitive DNA sequences (5.61%), similarly to the observed in other oak species (Zoldos et al., 1998; Rampant et al., 2011). In addition, Class I retrotransposons, in particular *Ty1*/copia elements were the most abundant retroelements in cork oak BESs. Although, in this study were analysed a small number of BESs which can introduce a bias in the results. Homology searches identified BESs with significant identity with *Ty3*/Gypsy LTR retrotransposons from the Athila/Tat clade, similarly to findings in other *Quercus* species. From the different homology searches performed were found significant identity with only two sequences described in *Quercus suber*. The search for SSRs showed that dinucleotide repeats, in particular the poly (AT) motif were the most frequent repeats. A remarkable finding, was the identification in the same BAC clone (HD_B-HDL_63G20) of two genes (Cyp 86A1 and GDSL) involved in the acyl lipids suberin biosynthesis pathway, strongly suggesting the linkage between these genes. Furthermore, during validation of the positive clones we identified BAC clones containing two different genes. It was the case of the clones QS-H-HDL_03M05 and QS-H-HDL_32G04 comprising the genes DHCA-S-AT and 4CL; and of clones QS-H-HDL_44P23 and QS-H-HDL_98I24 including the genes PAL and ELTFP. These results suggest that these group of genes involved in the suberin biosynthetic pathway are proximally located in the genome and thus indicating also the possible linkage between these genes. The BAC end sequence analysis showed high identity with coding regions and suggest the existence of conserved regions within the *Fagaceae* family. The BAC end sequences analysis results provided a first insight on the composition of the cork oak genome and could potentially be used for the development of molecular markers for genetic mapping of cork oak genome.

In a second part of our study, we attempted to map some candidate genes for cork formation in the cork oak genome and thus to develop the first *Q. suber* cytogenetic map. For that, was selected a group of BAC clones containing genes of interest to hybridize with mitotic cork oak chromosomes, using the BAC landing methodology. In this work we were able to isolate BAC DNA with reasonable good quality, from BAC clones of the *Quercus suber* BAC library, with three different methods. We also obtained good

chromosome metaphase spreads allowing low background hybridization and good accessibility for the probes. To our knowledge, this work was the first attempt to land BAC DNA onto *Q. suber* chromosomes. We observed strong fluorescent signals from the hybridized rDNA probes, although we did not observe any visible signal from the hybridized BAC DNA probes, suggesting that either the BAC DNA isolation or/and the BAC landing methods were not efficient. Further troubleshooting will be necessary for optimizing the isolation of BAC DNA and the BAC landing protocol in forest trees species.

In this work we also analyzed and characterized the expression of 15 genes related with suberin biosynthesis, in cork developing tissues, from cork oak trees with different cork qualities. The data suggests a high variability of gene expression for the different genotypes, but no significant differencial expression have been found for genotypes obtained from low and high cork quality. Further experiments are required to evaluate the differential expression of the candidate genes involved in cork quality phenotype. The combination of data from phenotype analysis and different genome expression levels could provide a better understanding of cork quality to be applied in the future as tools for cork oak tree-breeding programs for production of cork with higher quality.

In this work, we successfully used several molecular genetic methodologies to characterize genes involved in suberin biosynthesis and cork formation. To our knowledge this was the first BAC approach to improve the understanding of cork oak genomics and provide some of the basic tools which will be needed for the establishment of relations between the *Quercus suber* genome and cork quality as a tree phenotype.

Appendices

Appendix 1 – BAC clones selected for BAC end sequencing

Table 16. List of 24 BAC clones used for BAC end sequencing analysis and the genes found in these clones validated by Colony PCR. The BAC clones selected for BAC end sequencing contained genes involved in the main metabolic pathways of suberin biosynthesis (Acyl lipids and Phenylpropanoid pathways) and in regulatory mechanisms of cork formation.

Suberin biosynthesis pathway or function	Gene functional category	Gene name	BAC clones used for BAC end sequencing	Insert size (bp)
Phenylpropanoids:	Acyltransferase	HCBT	27K11 44H18 52A04 67A15 79E11 90H01 94P17 98D18 106E13	>2000
	Phenylpropanoids derivatives	Cyp 84A1	31H02	100
Acyl lipids:	Lipid metabolism	GDSL	21N01 30E04 52D11 77E11 86P17 86P18 88B10 100B01 104G06	150
	Fatty acid hydroxylation	Cyp 86A1	63G20 83E09 87K10	100
Regulatory proteins	Transcription factor	WRKY	52024 82D02	100

Appendix 2 – GC content of BAC end sequences

Table 17. GC content of the BAC end sequences analysed. The GC content was determined with the online software Oligo Calculator (http://www.mbcf.dfci.harvard.edu/docs/oligocalc.html).

BAC end sequence	GC content (%)
HEBE_QsuHDLx106E13xT7	55
HEBE_QsuHDLx106E13xM13rev	40
HEBE_QsuHDLx104G06xT7	31
HEBE_QsuHDLx104G06xM13rev	30
HEBE_QsuHDLx100B01xT7	31
HEBE_QsuHDLx100B01xM13rev	33
HEBE_QsuHDLx98D18xT7	28
HEBE_QsuHDLx98D18xM13rev	28
HEBE_QsuHDLx94P17xT7	36
HEBE_QsuHDLx94P17xM13rev	41
HEBE_QsuHDLx90H01xT7	38
HEBE_QsuHDLx90H01xM13rev	29
HEBE_QsuHDLx88B10xT7	48
HEBE_QsuHDLx88B10xM13rev	24
HEBE_QsuHDLx87K10xT7	45
HEBE_QsuHDLx87K10xM13rev	36
HEBE_QsuHDLx86P18xT7	40
HEBE_QsuHDLx86P18xM13rev	49
HEBE_QsuHDLx86P17xT7	40
HEBE_QsuHDLx86P17xM13rev	49
HEBE_QsuHDLx83E09xT7	44
HEBE_QsuHDLx83E09xM13rev	39
HEBE_QsuHDLx82D02xT7	32
HEBE_QsuHDLx82D02xM13rev	38
HEBE_QsuHDLx79E11xT7	33
HEBE_QsuHDLx79E11xM13rev	36
HEBE_QsuHDLx77E11xT7	27
HEBE_QsuHDLx77E11xM13rev	32
HEBE_QsuHDLx67A15xT7	36
HEBE_QsuHDLx67A15xM13rev	36
HEBE_QsuHDLx63G20xT7	32
HEBE_QsuHDLx63G20xM13rev	31
HEBE_QsuHDLx52O24xT7	44
HEBE_QsuHDLx52O24xM13rev	41
HEBE_QsuHDLx52D11xT7	40
HEBE_QsuHDLx52D11xM13rev	35
HEBE_QsuHDLx52A04xT7	39
HEBE_QsuHDLx52A04xM13rev	32
HEBE_QsuHDLx44H18xT7	36
HEBE_QsuHDLx44H18xM13rev	37
HEBE_QsuHDLx31H02xT7	35
HEBE_QsuHDLx31H02xM13rev	50
HEBE_QsuHDLx30E04xT7	39
HEBE_QsuHDLx30E04xM13rev	55
HEBE_QsuHDLx27K11xT7	37
HEBE_QsuHDLx27K11xM13rev	39
HEBE_QsuHDLx21N01xT7	41
HEBE_QsuHDLx21N01xM13rev	39
Average GC content (%)	**37,625**

Appendix 3 – Identification of mobile genetic elements (MGEs) *with BLAST homology searches against the Gypsy Database (GyDB)*

Table 18. List of the BLASTn results between *Q. suber* BESs and genome database obtained by Gypsy database (GyDB).

BLASTn_Genome database		
BAC end sequence	**Sequences producing significant alignments**	**E value**
QsuHDLx104G06xT7	GEN_HmGINO1	7E-08
	GEN_HmGINGER1_mRNA	7E-08
	GEN_HmGINO2_with_45ntTIRs	3E-07
	GEN_Nomad	2E-05
QsuHDLx100B01xM13rev	GEN_HmGINO2_with_45ntTIRs	7E-05
QsuHDLx98D18xT7	GEN_Tom	1E-11
	GEN_Tdd-5_with_TIRs	1E-11
	GEN_Tdd-4_withTIRs	1E-11
	GEN_Mdg1	6E-11
	GEN_Nomad	6E-11
	GEN_BtGIN1_mRNA	6E-11
	GEN_Yoyo	2E-10
	GEN_V12	9E-10
	GEN_Zam	1E-08
	GEN_SsGIN1_mRNA	5E-08
	GEN_HMS-Beagle	2E-07
	GEN_CfSCAND3	8E-07
	GEN_17.6	3E-06
	GEN_Gypsy	3E-06
	GEN_LsGINGER2-1_Repbaseaccession_Ginger2-1_LS	3E-06
	GEN_SsSCAND3	3E-06
	GEN_PvSCAND3	3E-06
	GEN_AcGINGER1-1_Repbaseaccession_Ginger1-1_AC_with_138ntTIRs	3E-06
	GEN_Gypsyvir	1E-05
	GEN_Osvaldo	1E-05
	GEN_Zeco1	5E-05

BLASTn_Genome database		
BAC end sequence	**Sequences producing significant alignments**	**E value**
QsuHDLx87K10xM13rev	GEN_Del	1E-05
	GEN_HmGINO2_with_45ntTIRs	1E-05
	GEN_ApGINGER1f	1E-05
	GEN_HmGINO1	5E-05
QsuHDLx82D02xT7	GEN_Tdd-5_with_TIRs	9E-07
	GEN_CfSCAND3	1E-05
	GEN_Nomad	6E-05
	GEN_Tdd-4_withTIRs	6E-05
	GEN_SsSCAND3	6,E-05
	GEN_LaSCAND3	6E-05
QsuHDLx79E11xT7	GEN_HmGINO1	7E-08
	GEN_HmGINO2_with_45ntTIRs	7E-08
	GEN_HmGINGER1_mRNA	3E-07
	GEN_Del	4E-06
	GEN_Nomad	2E-05
	GEN_ApGINGER1f	7E-05
QsuHDLx79E11xM13rev	GEN_SsSCAND3	6E-05
QsuHDLx52O24xT7	GEN_HmGINO2_with_45ntTIRs	1E-12
	GEN_HmGINO1	2E-11
	GEN_AcGINGER1-1_Repbaseaccession_Ginger1-1_AC_with_138ntTIRs	1E-09
	GEN_OcKRBA2	2E-05
QsuHDLx52D11xT7	GEN_HmGINO2_with_45ntTIRs	1E-18
	GEN_HmGINO1	2E-17
	GEN_AcGINGER1-1_Repbaseaccession_Ginger1-1_AC_with_138ntTIRs	9E-16
	GEN_OcKRBA2	3E-06
	GEN_HmGINGER1_mRNA	5E-05
QsuHDLx21N01xM13rev	GEN_HmGINO2_with_45ntTIRs	1E-18
	GEN_HmGINO1	2E-17
	GEN_AcGINGER1-1_Repbaseaccession_Ginger1-1_AC_with_138ntTIRs	1E-15
	GEN_OcKRBA2	4E-06
	GEN_HmGINGER1_mRNA	7E-05

Table 19. Results of BLASTx analysis between cork aok BESs and cores database by GyDB.

BLASTx_Cores database		
BAC end sequence	**Sequences producing significant alignments**	**E value**
QsuHDLx27K11xM13rev	RT_Melmoth	5E-20
	RT_Retrofit	6E-13
	RT_Koala	1E-12
	RT_ToRTL1	2E-11
	RT_Poco	3E-11
	RT_Sto-4	9E-11
	RT_Hopscotch	2E-10
	RT_Hydra1-2	3E-10
	RT_V12	3E-09
	RT_RTvr2	7E-09
	RT_Tto1	7E-09
	RT_Oryco1-1	9E-09
	RT_Xanthias	9E-09
	RT_Tricopia	3E-08
	RT_Batata	4E-08
	RT_Humnum	1E-07
	RT_TSI-9	1E-07
	RT_Endovir1-1	1E-07
	RT_Hydra1-1	1E-07
	RT_Tork4	2E-07
	RT_Fourf	2E-07
	RT_1731	2E-07
	RT_Koco	3E-07
	RT_Araco	1E-06
	RT_Copia	2E-06
	RT_Tnt-1	2E-06
	RT_Opie-2	3E-05
	RT_SIRE1-4	3E-05
	RT_Mtanga	5E-05
	RT_Vitico1-1	6E-05
QsuHDLx21N01xT7	INT_RetroSor1	2E-11
	INT_Tat4-1	5E-11
	INT_Cinful-1	3E-09
	INT_B1147A04.5	7E-08
	INT_Tft2	7E-08
	INT_RIRE2	2E-07
	INT_Ogre	1E-06
	INT_Grande1-4	1E-05

Appendix 4 - BLAST homology searches.

Table 20. BLASTn search results between cork oak BESs and the non-redundant nucleotide collection (nr/nt) database.

BAC end sequence	Size (bp)	Accession Number	Nucleotide Sequences in GenBank	E-value	Max. Identity (%)
QsuHDLx106E13xM13rev	346	XM_002272559.1	ref|XM_002272559.1| PREDICTED: Vitis vinifera uncharacterized LOC100265200 (LOC100265200), mRNA	1E-43	76
QsuHDLx98D18xT7	400	AM459017.2	emb|AM459017.2| Vitis vinifera contig VV78X254162.5, whole genome shotgun sequence	2E-07	93
QsuHDLx82D02xM13rev	561	FR852835.1	emb|FR852835.1| Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline17_5	2E-15	68
QsuHDLx79E11xT7	518	AM471831.2	emb|AM471831.2| Vitis vinifera contig VV78X127536.10, whole genome shotgun sequence	7E-08	100
QsuHDLx63G20xM13rev	559	XM_003544976.1	ref|XM_003544976.1| PREDICTED: Glycine max GDSL esterase/lipase EXL1-like (LOC100776769), mRNA	3E-05	76
QsuHDLx52O24xT7	557	CR377470.1	emb|CR377470.1| Pinus pinaster SSR, clone PPA6B11	4E-10	96
QsuHDLx52O24xM13rev	671	FR852855.1	emb|FR852855.1| Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR41 (Belline1_41)	4E-10	75
QsuHDLx52D11xT7	395	AC212845.1	gb|AC212845.1| Populus trichocarpa clone POP021-P21, complete sequence	1E-17	81
QsuHDLx44H18xT7	626	AC213416.1	gb|AC213416.1| Populus trichocarpa clone POP016-O23, complete sequence	3E-06	84
QsuHDLx44H18xM13rev	600	XM_002298612.1	ref|XM_002298612.1| Populus trichocarpa predicted protein, mRNA	1E-29	75
QsuHDLx31H02xT7	642	XM_002318461.1	ref|XM_002318461.1| Populus trichocarpa predicted protein, mRNA	1E-11	82
QsuHDLx30E04xT7	731	AC212845.1	gb|AC212845.1| Populus trichocarpa clone POP021-P21, complete sequence	4E-17	75
QsuHDLx27K11xT7	722	XM_003523930.1	ref|XM_003523930.1| PREDICTED: Glycine max putative receptor protein kinase ZmPK1-like (LOC100816469), mRNA	5E-66	74
QsuHDLx27K11xM13rev	584	AC151462.30	gb|AC151462.30| Medicago truncatula clone mth2-90j2, complete sequence	8E-07	72
QsuHDLx21N01xM13rev	510	FN564429.1	emb|FN564429.1| Triticum aestivum chromosome 3B-specific BAC library, contig ctg0382b	8E-13	96

Table 21. BLASTn results between cork oak BESs and the genomic survey sequences (GSS) database with the organism defined as *Quercus* (Taxid: 3511).

BAC end sequence	Size (bp)	Accession Number	Nucleotide Sequences in GenBank	E-value	Max. Identity (%)
QsuHDLx106E13xT7	501	gb|JS680224.1|	gb|JS680224.1| ATE0AAA7YF20FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 5', genomic survey sequence.	2E-32	75
QsuHDLx100B01xM13rev	566	gb|HN155880.1|	gb|HN155880.1| >TI0AQRHB12YF19RM1 Quercus robur, 3P genotype, BAC end sequences.Quercus robur genomic 3', genomic survey sequence.	3E-47	84
QsuHDLx98D18xM13rev	474	gb|JS682333.1|	gb|JS682333.1| ATE0AAA13YP15RM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 3', genomic survey sequence.	2E-57	69
QsuHDLx94P17xT7	433	gb|HR302411.1|	gb|HR302411.1| WZ0AQRGC4YK02EM1.SCF Quercus robur, WZ0AQRGC, random genomic. library Quercus robur genomic 5', genomic survey sequence.	1E-45	93
QsuHDLx90H01xT7	433	gb|HR302411.1|	gb|JS680105.1| ATE0AAA7YA12RM1 Pedunculate oak Quercus robur Qro-B-3Ph . Hind III Quercus robur genomic 3', genomic survey sequence.	2E-50	77
		gb|HN161697.1|	gb|HN161697.1| >TI0AQRHB22YI13RM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	4E-34	71
QsuHDLx88B10xT7	462	gb|HN160440.1|	gb|HN160440.1| >TI0AQRHB20YD14FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	2E-06	73
		gb|HN158909.1|	gb|HN158909.1| >TI0AQRHB18YF04RM1 Quercus robur, 3P genotype, BAC end sequences.Quercus robur genomic 3', genomic survey sequence.	6E-06	90
QsuHDLx87K10xT7	431	gb|HN170872.1|	gb|HN170872.1| >TI0AQRHB38YO03FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	1E-97	78
		gb|HN173761.1|	gb|HN173761.1| >TI0AQRHB9YD18RM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	4E-97	80
QsuHDLx86P18xM13rev	390	gb|JS683140.1|	gb|JS683140.1| ATE0AAA16YB13FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque . Hind III Quercus robur genomic 5', genomic survey sequence.	1E-14	66
QsuHDLx86P17xM13rev	403	gb|JS683140.1|	gb|JS683140.1| ATE0AAA16YB13FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque . Hind III Quercus robur genomic 5', genomic survey sequence.	2E-11	80
QsuHDLx83E09xT7	560	gb|JS676726.1|	gb|JS676726.1| ATE0AAA21YC16FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 5', genomic survey sequence.	3E-47	85

BAC end sequence	Size (bp)	Accession Number	Nucleotide Sequences in GenBank	E-value	Max. Identity (%)
QsuHDLx82D02xM13rev	561	gb|JS683332.1|	gb|JS683332.1| ATE0AAA16YJ08RM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 3', genomic survey sequence	8E-68	77
QsuHDLx79E11xT7	518	gb|HN161593.1|	gb|HN161593.1| >TI0AQRHB22YF09RM2 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	2E-63	79
QsuHDLx79E11xM13rev	427	gb|HN157482.1|	gb|HN157482.1| >TI0AQRHB15YH13RM2 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	9E-131	86
QsuHDLx77E11xT7	560	gb|HN163653.1|	gb|HN163653.1| >TI0AQRHB26YB01RM1 Quercus robur, 3P genotype, BAC end sequences . Quercus robur genomic 3', genomic survey sequence.	2E-24	75
QsuHDLx77E11xM13rev	525	gb|HN171680.1|	gb|HN171680.1| >TI0AQRHB5YF19FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	5E-26	82
QsuHDLx67A15xT7	532	gb|HN168194.1|	gb|HN168194.1| >TI0AQRHB33YL22FM2 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	2E-100	78
QsuHDLx63G20xT7	624	gb|HN157676.1|	gb|HN157676.1| >TI0AQRHB15YN18FM2 Quercus robur, 3P genotype, BAC end sequences . Quercus robur genomic 5', genomic survey sequence.	3E-10	85
QsuHDLx52O24xM13rev	671	gb|HN165204.1|	gb|HN165204.1| >TI0AQRHB28YM05FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	6E-25	71
QsuHDLx52D11xM13rev	440	gb|HN164821.1|	gb|HN164821.1| >TI0AQRHB28YB17FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	2E-120	86
QsuHDLx52A04xT7	367	gb|JS680105.1|	gb|JS680105.1| ATE0AAA7YA12RM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 3', genomic survey sequence.	7E-33	71
QsuHDLx44H18xT7	626	gb|HN163371.1|	gb|HN163371.1| >TI0AQRHB25YI20RM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	2E-33	84
QsuHDLx27K11xT7	722	gb|HN171004.1|	gb|HN171004.1| >TI0AQRHB4YB17FM1 Quercus robur, 3P genotype, BAC end sequences . Quercus robur genomic 5', genomic survey sequence.	3E-18	71
QsuHDLx21N01xT7	668	gb|JS681710.1|	gb|JS681710.1| ATE0AAA12YC18FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 5', genomic survey sequence.	3E-44	77

Table 22. BLASTn results between cork oak BESs and the expressed sequence tags (ESTs) database of the organism defined as *Fagaceae* (Taxid: 3503).

BAC end sequence	Size (bp)	Accession Number	Nucleotide Sequences in GenBank	E-value	Max. Identity (%)
QsuHDLx106E13xM13rev	346	FP053794.1	emb|FP053794.1| FP053794 LG0AAD Quercus robur cDNA clone LG0AAD50YF24RM1, mRNA	8E-127	85
QsuHDLx100B01xT7	631	FR616959.1	emb|FR616959.1| FR616959 WZOAFSAA Fagus sylvatica cDNA clone WZ0AFSAA21YB22FM1,mRNA sequence.	3E-05	91
QsuHDLx94P17xM13rev	348	FP072945.1	emb|FP072945.1| FP072945 LG0AAD Quercus robur cDNA clone LG0AAD15YC13RM1_2X, mRNA sequence.	3E-09	79
QsuHDLx88B10xM13rev	455	FP070663.1	emb|FP070663.1|FP070663 LG0AAD Quercus robur cDNA clone LG0AAD1YI14FM2_1T, mRNA sequence.	7E-10	73
QsuHDLx82D02xT7	441	DB996567.2	dbj|DB996567.2| DB996567 Quercus mongolica var. crispula inner bark cDNA library Quercus mongolica subsp. crispula cDNA clone Qm_001_03K12 5', mRNA sequence.	5E-10	100
QsuHDLx82D02xM13rev	561	FR617160.1	emb|FR617160.1| FR617160 WZOAFSAA Fagus sylvatica cDNA clone WZ0AFSAA22YD07FM1, mRNA sequence.	1E-07	84
QsuHDLx79E11xT7	518	FP034097.1	emb|FP034097.1| FP034097 LG0AAB Quercus robur cDNA clone LG0AAB3YI01RM1, mRNA sequence.	5E-07	81
QsuHDLx79E11xM13rev	427	FP051817.1	emb|FP051817.1| FP051817 LG0AAD Quercus robur cDNA clone LG0AAD8YJ02RM1, mRNA sequence.	2E-17	96
QsuHDLx77E11xT7	560	FP070663.1	emb|FP070663.1| FP070663 LG0AAD Quercus robur cDNA clone LG0AAD1YI14FM2_1T, mRNA sequence.	4E-15	77
QsuHDLx63G20xT7	624	FR665068.1	emb|FR665068.1| FR665068 WZ0AQSAA Quercus suber cDNA clone WZ0AQSAA6YN21CM1, mRNA sequence.	7E-10	83
QsuHDLx52D11xT7	395	FR601817.1	emb|FR601817.1| FR601817 WZOAFSCA Fagus sylvatica cDNA clone WZ0AFSCA22YN15FM1, mRNA sequence.	1E-07	97
QsuHDLx52D11xM13rev	440	DC601112.1	dbj|DC601112.1| DC601112 Castanopsis sieboldii inner bark cDNA library Castanopsis sieboldii cDNA clone Cc_TUM05_05C21 5', mRNA sequence.	1E-07	90
QsuHDLx44H18xM13rev	600	FP036592.1	emb|FP036592.1| FP036592 LG0AAB Quercus robur cDNA clone LG0AAB23YO07RM1 similar to PplGI.3|TC81072 similar to UP|Q9LSE0_ARATH Genomic DNA, chromosome 3, P1 clone: MOJ10, mRNA sequence.	2E-131	96
QsuHDLx31H02xT7	642	FP038492.1	emb|FP038492.1| FP038492 LG0AAB Quercus robur cDNA clone LG0AAB18YB24RM1 similar to PplGI.3|TC58622 weakly similar to UP|Q337M3_ORYSA Expressed protein, mRNA sequence.	2E-79	100
QsuHDLx27K11xT7	722	FP065588.1	emb|FP065588.1| FP065588 LG0AAD Quercus robur cDNA clone LG0AAD30YJ10RM1, mRNA sequence.	6E-25	73
QsuHDLx21N01xT7	668	DB998111.2	dbj|DB998111.2| DB998111 Quercus mongolica var. crispula inner bark cDNA library Quercus mongolica subsp. crispula cDNA clone Qm_001_08A04 5', mRNA sequence.	1E-08	72

Table 23. BLASTx results between cork oak BESs and the non-redundant protein sequences (nr) database; nr/nt (nr) database of the *Fagaceae* (Taxid: 3503) and *Quercus* (Taxid: 3511) organisms.

Database and defined organism	BAC end sequence	Size (bp)	Accession Number	Nucleotide Sequences in GenBank	E-value	Max. Identity (%)
non-redundant protein sequences (nr)	QsuHDLx106E13xM13rev	566	XP_003549067.1	PREDICTED: uncharacterized protein LOC100788210 [Glycine max]	3E-32	62
	QsuHDLx87K10xT7	431	AAD37019.2	putative non-LTR retrolelement reverse transcriptase [Arabidopsis thaliana]	1E-09	31
	QsuHDLx82D02xM13rev	561	ABA98491.1	gb|ABA98491.1| retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]	2E-46	45
	QsuHDLx63G20xM13rev	559	XP_002279353.2	ref|XP_002279353.2| PREDICTED: GDSL esterase/lipase EXL3-like [Vitis vinifera]	1E-08	59
	QsuHDLx52O24xM13rev	671	CCA66036.1	emb|CCA66036.1| hypothetical protein [Beta vulgaris subsp. vulgaris]	2E-21	31
			AAP54617.2	gb|AAP54617.2| retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]	4E-19	36
	QsuHDLx52D11xT7	395	XP_002522361.1	ref|XP_002522361.1| oligopeptidase B, putative [Ricinus communis]	1E-06	79
	QsuHDLx44H18xM13rev	600	XP_003622039.1	ref|XP_003622039.1| hypothetical protein MTR_7g026510 [Medicago truncatula]	4E-25	79
	QsuHDLx31H02xT7	642	XP_002525594.1	ref|XP_002525594.1| conserved hypothetical protein [Ricinus communis]	4E-17	72
	QsuHDLx30E04xT7	731	CBI40973.3	emb|CBI40973.3| unnamed protein product [Vitis vinifera]	3E-10	62
	QsuHDLx27K11xT7	722	XP_002510149.1	ref|XP_002510149.1| serine-threonine protein kinase, plant-type, putative [Ricinus communis]	6E-73	58
	QsuHDLx27K11xM13rev	584	CAN81494.1	emb|CAN81494.1| hypothetical protein VITISV_031968 [Vitis vinifera]	5E-34	50
	QsuHDLx21N01xT7	668	ABW81060.1	gb|ABW81060.1| GagPol3 [Arabidopsis lyrata subsp. lyrata]	8E-13	49
nr_*Fagaceae*	QsuHDLx82D02xM13rev	561	AAL78659.1	gb|AAL78659.1|AF405557_1 reverse transcriptase [Fagus sylvatica]	1E-08	50
	QsuHDLx27K11xT7	722	CAC09571.1	emb|CAC09571.1| S-receptor kinase (SRK) [Fagus sylvatica]	2E-17	46
nr_ *Quercus*	QsuHDLx21N01xT7	668	AAQ94320.1|	gb|AAQ94320.1| gag protein [Quercus suber]	7E-10	49

Appendix 5 – Summary of the homology search

Table 24. Summary of the homology searches results obtained for each BAC end sequence. Eight BESs did not present homology results and are highlighted in grey.

BAC end sequence	Homology search	Nucleotide Sequences in GenBank/ Sequences producing significant alignments	E-value	SSRs identification	
				Motif	Length
QsuHDLx106E13xT7	BLASTn vs. GSS database (*Quercus*)	gb|JS680224.1| ATE0AAA7YF20FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 5', genomic survey sequence.	2,00E-32		
QsuHDLx106E13xM13rev	BLASTn vs. nr/nt database	ref|XM_002272559.1| PREDICTED: Vitis vinifera uncharacterized LOC100265200 (LOC100265200), mRNA	1,00E-43		
	BLASTn vs. EST database (*Fagaceae*)	emb|FP053794.1| FP053794 LG0AAD Quercus robur cDNA clone LG0AAD50YF24RM1, mRNA	8,00E-127		
	BLASTx vs. nr/nt database	ref| XP_003549067.1| PREDICTED: uncharacterized protein LOC100788210 [Glycine max]	3,00E-32		
QsuHDLx104G06xT7	BLASTn vs. genome database (GyDB)	GEN_HmGINO1	7,00E-08	AAAAAG	18
		GEN_HmGINGER1_mRNA	7,00E-08		
		GEN_HmGINO2_with_45ntTIRs	3,00E-07	AT	25
		GEN_Nomad	2,00E-05		
QsuHDLx104G06xM13rev					
QsuHDLx100B01xT7	BLASTn vs. EST database (*Fagaceae*)	emb|FR616959.1| FR616959 WZOAFSAA Fagus sylvatica cDNA clone WZ0AFSAA21YB22FM1,mRNA sequence.	3,00E-05		
QsuHDLx100B01xM13rev	BLASTn vs. genome database (GyDB)	GEN_HmGINO2_with_45ntTIRs	7,00E-05	AT	21
	BLASTn vs. GSS database (*Quercus*)	gb|HN155880.1| >TI0AQRHB12YF19RM1 Quercus robur, 3P genotype, BAC end sequences.Quercus robur genomic 3', genomic survey sequence.	3,00E-47		
QsuHDLx98D18xT7	BLASTn vs. nr/nt database	emb|AM459017.2| Vitis vinifera contig VV78X254162.5, whole genome shotgun sequence	2,00E-07	A	20
	BLASTn vs. genome database (GyDB)	GEN_Tom	1,00E-11		
		GEN_Tdd-5_with_TIRs	1,00E-11		
		GEN_Tdd-4_withTIRs	1,00E-11	A	32
		GEN_Mdg1	6,00E-11		
		GEN_Nomad	6,00E-11		
		GEN_BtGIN1_mRNA	6,00E-11	AG	36
		GEN_Yoyo	2,00E-10		
		GEN_V12	9,00E-10		
QsuHDLx98D18xM13rev	BLASTn vs. GSS database (*Quercus*)	gb|JS682333.1| ATE0AAA13YP15RM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 3', genomic survey sequence.	2,00E-57		
QsuHDLx94P17xT7	BLASTn vs. GSS database (*Quercus*)	gb|HR302411.1| WZ0AQRGC4YK02EM1.SCF Quercus robur, WZ0AQRGC, random genomic. library Quercus robur genomic 5', genomic survey sequence.	1,00E-45		

QsuHDLx94P17xM13rev	BLASTn vs. EST database (*Fagaceae*)	emb|FP072945.1| FP072945 LG0AAD Quercus robur cDNA clone LG0AAD15YC13RM1_2X, mRNA sequence.	3,00E-09		
QsuHDLx90H01xT7	BLASTn vs. GSS database (*Quercus*)	gb|JS680105.1| ATE0AAA7YA12RM1 Pedunculate oak Quercus robur Qro-B-3Ph . Hind III Quercus robur genomic 3', genomic survey sequence.	2,00E-50		
		gb|HN161697.1| >TI0AQRHB22YI13RM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	4,00E-34		
QsuHDLx90H01xM13rev				AAAAG	26
QsuHDLx88B10xT7	BLASTn vs. GSS database (*Quercus*)	gb|HN160440.1| >TI0AQRHB20YD14FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	2,00E-06		
		gb|HN158909.1| >TI0AQRHB18YF04RM1 Quercus robur, 3P genotype, BAC end sequences.Quercus robur genomic 3', genomic survey sequence.	6,00E-06		
QsuHDLx88B10xM13rev	BLASTn vs. EST database (*Fagaceae*)	emb|FP070663.1|FP070663 LG0AAD Quercus robur cDNA clone LG0AAD1YI14FM2_1T, mRNA sequence.	7,00E-10		
QsuHDLx87K10xT7	BLASTx vs. nr/nt database	gb| AAD37019.2| putative non-LTR retrolelement reverse transcriptase [Arabidopsis thaliana]	1,00E-09		
	BLASTn vs. GSS database (*Quercus*)	gb|HN170872.1| >TI0AQRHB38YO03FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	1,00E-97		
		gb|HN173761.1| >TI0AQRHB9YD18RM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	4,00E-97		
QsuHDLx87K10xM13rev	BLASTn vs. genome database (GyDB)	GEN_Del	1,00E-05	AT	19
		GEN_HmGINO2_with_45ntTIRs	1,00E-05		
		GEN_ApGINGER1f	1,00E-05		
		GEN_HmGINO1	5,00E-05		
QsuHDLx86P18xT7					
QsuHDLx86P18xM13rev	BLASTn vs. GSS database (*Quercus*)	gb|JS683140.1| ATE0AAA16YB13FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque . Hind III Quercus robur genomic 5', genomic survey sequence.	1,00E-14		
QsuHDLx86P17xT7					
QsuHDLx86P17xM13rev	BLASTn vs. GSS database (*Quercus*)	gb|JS683140.1| ATE0AAA16YB13FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque . Hind III Quercus robur genomic 5', genomic survey sequence.	2,00E-11		
QsuHDLx83E09xT7	BLASTn vs. GSS database (*Quercus*)	gb|JS676726.1| ATE0AAA21YC16FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 5', genomic survey sequence.	3,00E-47		
QsuHDLx83E09xM13rev					
QsuHDLx82D02xT7	BLASTn vs. EST database (*Fagaceae*)	dbj|DB996567.2| DB996567 Quercus mongolica var. crispula inner bark cDNA library Quercus mongolica subsp. crispula cDNA clone Qm_001_03K12 5', mRNA sequence.	5,00E-10	T	18
	BLASTn vs. genome database (GyDB)	GEN_Tdd-5_with_TIRs	9,00E-07		
		GEN_CfSCAND3	1,00E-05		
		GEN_Nomad	6,00E-05		
		GEN_Tdd-4_withTIRs	6,00E-05		
		GEN_SsSCAND3	6,00E-05		
		GEN_LaSCAND3	6,00E-05		
QsuHDLx82D02xM13rev	BLASTn vs. nr/nt database	emb|FR852835.1| Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline17_5	2,00E-15		
	BLASTn vs. GSS database (*Quercus*)	gb|JS683332.1| ATE0AAA16YJ08RM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 3', genomic survey sequence	8,00E-68		

	BLASTn vs. EST database (*Fagaceae*)	emb|FR617160.1| FR617160 WZOAFSAA Fagus sylvatica cDNA clone WZ0AFSAA22YD07FM1, mRNA sequence.	1,00E-07		
	BLASTx vs. nr/nt database	gb|ABA98491.1| retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]	2,00E-46		
	BLASTx vs. nr/nt database (Fagaceae)	gb|AAL78659.1|AF405557_1 reverse transcriptase [Fagus sylvatica]	1,00E-08		
QsuHDLx79E11xT7	BLASTn vs. nr/nt database	emb|AM471831.2| Vitis vinifera contig VV78X127536.10, whole genome shotgun sequence	7,00E-08	AT	26
	BLASTn vs. GSS database (*Quercus*)	gb|HN161593.1| >TI0AQRHB22YF09RM2 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	2,00E-63		
	BLASTn vs. EST database (*Fagaceae*)	emb|FP034097.1| FP034097 LG0AAB Quercus robur cDNA clone LG0AAB3YI01RM1, mRNA sequence.	5,00E-07		
	BLASTn vs. genome database (GyDB)	GEN_HmGINO1	7,00E-08		
		GEN_HmGINO2_with_45ntTIRs	7,00E-08		
		GEN_HmGINGER1_mRNA	3,00E-07		
		GEN_Del	4,00E-06		
		GEN_Nomad	2,00E-05		
		GEN_ApGINGER1f	7,00E-05		
QsuHDLx79E11xM13rev	BLASTn vs. GSS database (*Quercus*)	gb|HN157482.1| >TI0AQRHB15YH13RM2 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	9,00E-131	A	16
	BLASTn vs. EST database (*Fagaceae*)	emb|FP051817.1| FP051817 LG0AAD Quercus robur cDNA clone LG0AAD8YJ02RM1, mRNA sequence.	2,00E-17		
	BLASTn vs. genome database (GyDB)	GEN_SsSCAND3	6,00E-05		
QsuHDLx77E11xT7	BLASTn vs. GSS database (*Quercus*)	gb|HN163653.1| >TI0AQRHB26YB01RM1 Quercus robur, 3P genotype, BAC end sequences . Quercus robur genomic 3', genomic survey sequence.	2,00E-24		
	BLASTn vs. EST database (*Fagaceae*)	emb|FP070663.1| FP070663 LG0AAD Quercus robur cDNA clone LG0AAD1YI14FM2_1T, mRNA sequence.	4,00E-15		
QsuHDLx77E11xM13rev	BLASTn vs. GSS database (*Quercus*)	gb|HN171680.1| >TI0AQRHB5YF19FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	5,00E-26		
QsuHDLx67A15xT7	BLASTn vs. GSS database (*Quercus*)	gb|HN168194.1| >TI0AQRHB33YL22FM2 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	2,00E-100		
QsuHDLx67A15xM13rev					
QsuHDLx63G20xT7	BLASTn vs. GSS database (*Quercus*)	gb|HN157676.1| >TI0AQRHB15YN18FM2 Quercus robur, 3P genotype, BAC end sequences . Quercus robur genomic 5', genomic survey sequence.	3,00E-10		
	BLASTn vs. EST database (*Fagaceae*)	emb|FR665068.1| FR665068 WZ0AQSAA Quercus suber cDNA clone WZ0AQSAA6YN21CM1, mRNA sequence.	7,00E-10		
QsuHDLx63G20xM13rev	BLASTn vs. nr/nt database	ref|XM_003544976.1| PREDICTED: Glycine max GDSL esterase/lipase EXL1-like (LOC100776769), mRNA	3,00E-05		
	BLASTx vs. nr/nt database	ref|XP_002279353.2| PREDICTED: GDSL esterase/lipase EXL3-like [Vitis vinifera]	1,00E-08		
QsuHDLx52O24xT7	BLASTn vs. nr/nt database	emb|CR377470.1| Pinus pinaster SSR, clone PPA6B11	4,00E-10	AC	29
	BLASTn vs. genome database (GyDB)	GEN_HmGINO2_with_45ntTIRs	1,00E-12		
		GEN_HmGINO1	2,00E-11		
		GEN_AcGINGER1-1_Repbaseaccession_Ginger1-1_AC_with_138ntTIRs	1,00E-09		

		GEN_OcKRBA2	2,00E-05		
QsuHDLx52O24xM13rev	BLASTn vs. nr/nt database	emb|FR852855.1| Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR41 (Belline1_41)	4,00E-10		
	BLASTn vs. GSS database (*Quercus*)	gb|HN165204.1| >TI0AQRHB28YM05FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	6,00E-25		
	BLASTx vs. nr/nt database	emb|CCA66036.1| hypothetical protein [Beta vulgaris subsp. vulgaris]	2,00E-21		
		gb|AAP54617.2| retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]	4,00E-19		
QsuHDLx52D11xT7	BLASTn vs. nr/nt database	gb|AC212845.1| Populus trichocarpa clone POP021-P21, complete sequence	1,00E-17	AC	39
	BLASTn vs. EST database (*Fagaceae*)	emb|FR601817.1| FR601817 WZOAFSCA Fagus sylvatica cDNA clone WZ0AFSCA22YN15FM1, mRNA sequence.	1,00E-07		
	BLASTx vs. nr/nt database	ref|XP_002522361.1| oligopeptidase B, putative [Ricinus communis]	1,00E-06		
	BLASTn vs. genome database (GyDB)	GEN_HmGINO2_with_45ntTIRs	1,00E-12		
		GEN_HmGINO1	2,00E-11		
		GEN_AcGINGER1-1_Repbaseaccession_Ginger1-1_AC_with_138ntTIRs	1,00E-09		
		GEN_OcKRBA2	2,00E-05		
QsuHDLx52D11xM13rev	BLASTn vs. GSS database (*Quercus*)	gb|HN164821.1| >TI0AQRHB28YB17FM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 5', genomic survey sequence.	2,00E-120		
	BLASTn vs. EST database (*Fagaceae*)	dbj|DC601112.1| DC601112 Castanopsis sieboldii inner bark cDNA library Castanopsis sieboldii cDNA clone Cc_TUM05_05C21 5', mRNA sequence.	1,00E-07		
QsuHDLx52A04xT7	BLASTn vs. GSS database (*Quercus*)	gb|JS680105.1| ATE0AAA7YA12RM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 3', genomic survey sequence.	7,00E-33		
QsuHDLx52A04xM13rev					
QsuHDLx44H18xT7	BLASTn vs. nr/nt database	gb|AC213416.1| Populus trichocarpa clone POP016-O23, complete sequence	3,00E-06		
	BLASTn vs. GSS database (*Quercus*)	gb|HN163371.1| >TI0AQRHB25YI20RM1 Quercus robur, 3P genotype, BAC end sequences. Quercus robur genomic 3', genomic survey sequence.	2,00E-33		
QsuHDLx44H18xM13rev	BLASTn vs. nr/nt database	ref|XM_002298612.1| Populus trichocarpa predicted protein, mRNA	1,00E-29		
	BLASTn vs. EST database (*Fagaceae*)	emb|FP036592.1| FP036592 LG0AAB Quercus robur cDNA clone LG0AAB23YO07RM1 similar to PplGI.3|TC81072 similar to UP|Q9LSE0_ARATH Genomic DNA, chromosome 3, P1 clone: MOJ10, mRNA sequence.	2,00E-131		
	BLASTx vs. nr/nt database	ref|XP_003622039.1| hypothetical protein MTR_7g026510 [Medicago truncatula]	4,00E-25		
QsuHDLx31H02xT7	BLASTn vs. nr/nt database	ref|XM_002318461.1| Populus trichocarpa predicted protein, mRNA	1,00E-11		
	BLASTn vs. EST database (*Fagaceae*)	emb|FP038492.1| FP038492 LG0AAB Quercus robur cDNA clone LG0AAB18YB24RM1 similar to PplGI.3|TC58622 weakly similar to UP|Q337M3_ORYSA Expressed protein, mRNA sequence.	2,00E-79		
	BLASTx vs. nr/nt database	ref|XP_002525594.1| conserved hypothetical protein [Ricinus communis]	4,00E-17		
QsuHDLx31H02xM13rev					
QsuHDLx30E04xT7	BLASTn vs. nr/nt database	gb|AC212845.1| Populus trichocarpa clone POP021-P21, complete sequence	4,00E-17		
	BLASTx vs. nr/nt database	emb|CBI40973.3| unnamed protein product [Vitis vinifera]	3,00E-10		
QsuHDLx30E04xM13rev					
QsuHDLx27K11xT7	BLASTn vs. nr/nt database	ref|XM_003523930.1| PREDICTED: Glycine max putative receptor protein kinase ZmPK1-like (LOC100816469), mRNA	5,00E-66		
	BLASTn vs. GSS database (*Quercus*)	gb|HN171004.1| >TI0AQRHB4YB17FM1 Quercus robur, 3P genotype, BAC end sequences . Quercus robur genomic 5', genomic survey sequence.	3,00E-18		

	BLASTn vs. EST database (*Fagaceae*)	emb|FP065588.1| FP065588 LG0AAD Quercus robur cDNA clone LG0AAD30YJ10RM1, mRNA sequence.	6,00E-25		
	BLASTx vs. nr/nt database	ref|XP_002510149.1| serine-threonine protein kinase, plant-type, putative [Ricinus communis]	6,00E-73		
	BLASTx vs. nr/nt database (Fagaceae)	emb|CAC09571.1| S-receptor kinase (SRK) [Fagus sylvatica]	2,00E-17		
QsuHDLx27K11xM13rev	BLASTn vs. nr/nt database	gb|AC151462.30| Medicago truncatula clone mth2-90j2, complete sequence	8,00E-07		
	BLASTx vs. nr/nt database	emb|CAN81494.1| hypothetical protein VITISV_031968 [Vitis vinifera]	5,00E-34		
	BLASTn vs. cores database (GyDB)	RT_Melmoth	5,00E-20		
		RT_Retrofit	6,00E-13		
		RT_Koala	1,00E-12		
		RT_ToRTL1	2,00E-11		
		RT_Poco	3,00E-11		
QsuHDLx21N01xT7	BLASTn vs. GSS database (*Quercus*)	gb|JS681710.1| ATE0AAA12YC18FM1 Pedunculate oak Quercus robur Qro-B-3Ph BAC/banque. Hind III Quercus robur genomic 5', genomic survey sequence.	3,00E-44		
	BLASTn vs. EST database (*Fagaceae*)	dbj|DB998111.2| DB998111 Quercus mongolica var. crispula inner bark cDNA library Quercus mongolica subsp. crispula cDNA clone Qm_001_08A04 5', mRNA sequence.	1,00E-08		
	BLASTx vs. nr/nt database	gb|ABW81060.1| GagPol3 [Arabidopsis lyrata subsp. lyrata]	8,00E-13		
	BLASTx vs. nr/nt database (*Quercus*)	gb|AAQ94320.1| gag protein [Quercus suber]	7,00E-10		
	BLASTn vs. cores database (GyDB)	INT_RetroSor1	2,00E-11		
		INT_Tat4-1	5,00E-11		
		INT_Cinful-1	3,00E-09		
		INT_B1147A04.5	7,00E-08		
QsuHDLx21N01xM13rev	BLASTn vs. nr/nt database	emb|FN564429.1| Triticum aestivum chromosome 3B-specific BAC library, contig ctg0382b	8,00E-13	AC	39
	BLASTn vs. genome database (GyDB)	GEN_HmGINO2_with_45ntTIRs	1,00E-18		
		GEN_HmGINO1	2,00E-17		
		GEN_AcGINGER1-1_Repbaseaccession_Ginger1-1_AC_with_138ntTIRs	1,00E-15		
		GEN_OcKRBA2	4,00E-06		

Appendix 6 - Total RNA extraction protocol

The total RNA was extracted from cork tissue by a modification of the method described by Provost et al. (2007).

1. The "RNA extraction buffer" (2% CTAB, 2% PVP, 10 mM Tris-HCl pH 8.0, 25 mM EDTA, 2 M NaCl and 0.5 g/l spermidine) was previously heated at 65ºC for 10 minutes and then added with 3% 2-mercaptoethanol.

2. Roughly 2 to 5 g of tissue was ground in liquid nitrogen to a fine powder and then mixed with the buffer.

3. After a 15-min incubation at 65°C, were performed three chloroform: isoamyl alcohol 24:1 (v/v) extractions at room temperature.

4. The RNA in the upper phase was precipitated with 10% NaOAc 4M (pH = 5.8) and 2 volumes of 100% ethanol for 1 hour at -20°C.

5. The precipitate was collected by centrifugation and the pellet was firstly washed with 70% (v/v) ethanol and secondly with 100% (v/v) ethanol. After the drying of the pellet, it was resuspended in 30µl of Milli-Q fresh water and stored at -80°C until use.

6. For the extraction of m RNA, after washing the pellet with ethanol, the pellet was resuspended in 400µl of Milli-Q fresh water and added with the same volume of 12 M LiCl.

7. It was precipitated for 1 hour until overnight at -20°C.

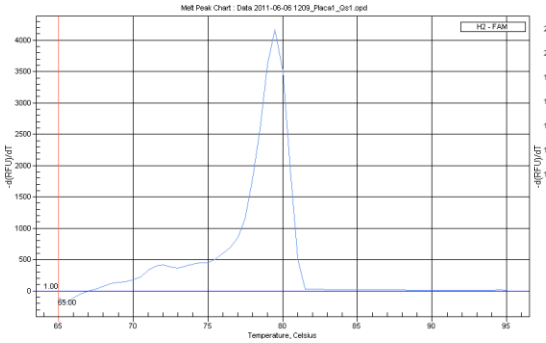
8. The precipitate was then collected by centrifugation for 30 minutes at 4°C and washed with 80% (v/v) ethanol. The pellet was left to dry at room temperature.

9. The pellet was then resuspended in 30µl of Milli-Q fresh water and stored at -80°C until use.

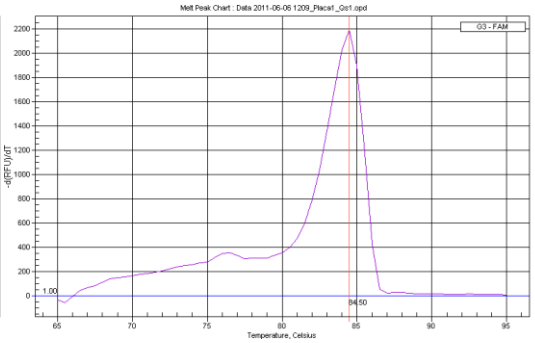
10. To verify the absence of degradation, RNA was separated by electrophoresis on a 2% denaturing agarose gel in 0.5X TBE buffer. The purity and quantity of RNA were checked with the NanoDrop ^{TM} 1000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA).

Appendix 7 - Melt-curve analysis

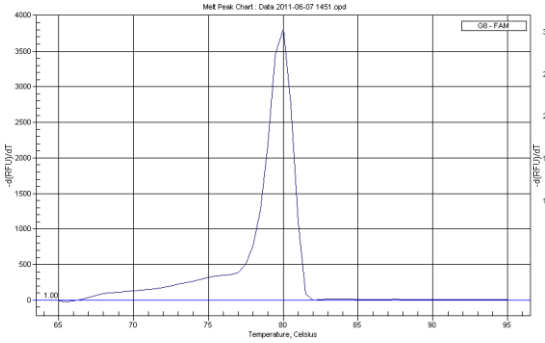
Melt curve charts of the genes analyzed by qPCR with the corresponding melting temperature (Tm).



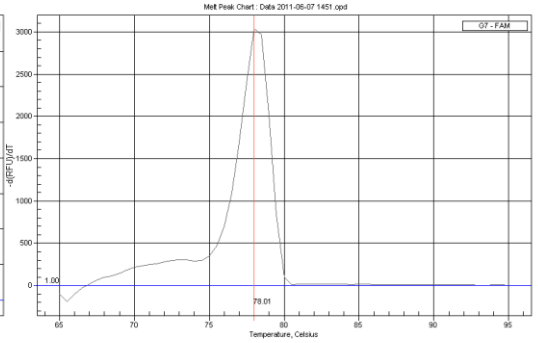
(a) Gene DHCA-S-AT with Tm of 79.50°C



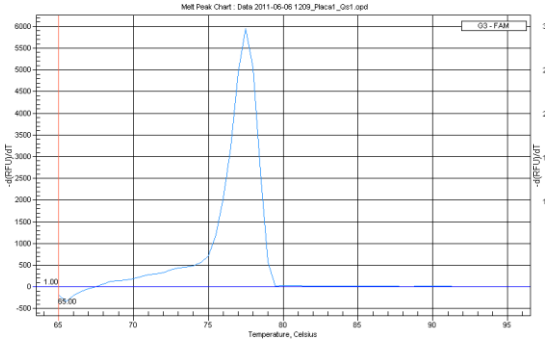
(b) Gene LACS with Tm of 84.50°C



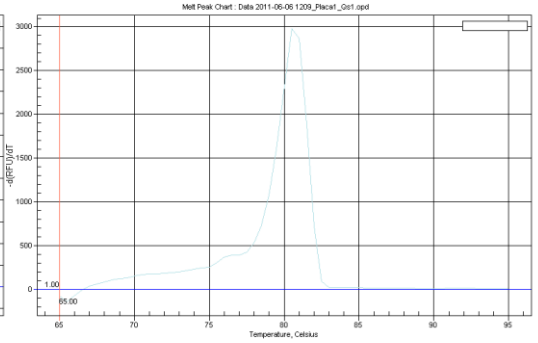
(c) Gene ELTFP with Tm of 82.50°C



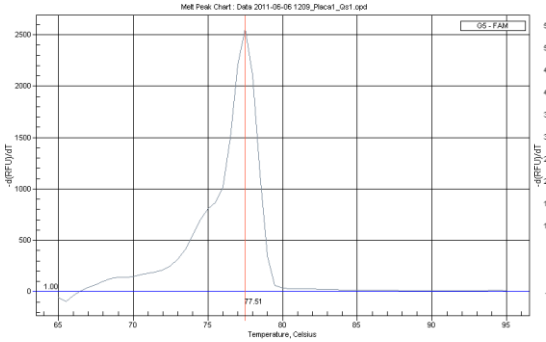
(d) Gene CYP86A1 with Tm of 80.00°C



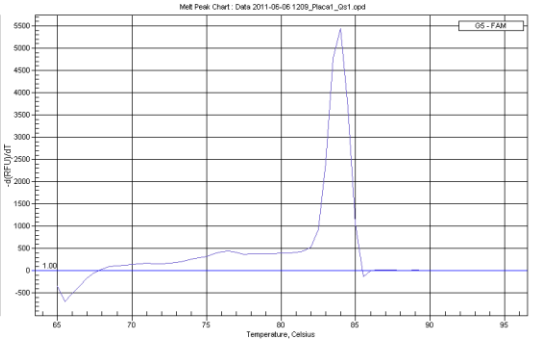
(e) Gene GPAT with Tm of 80.00°C



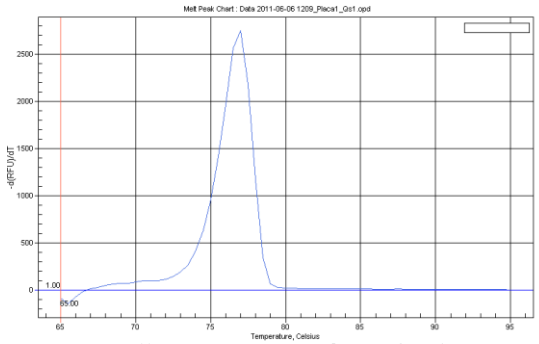
(f) Gene GDSL with Tm of 78.00°C



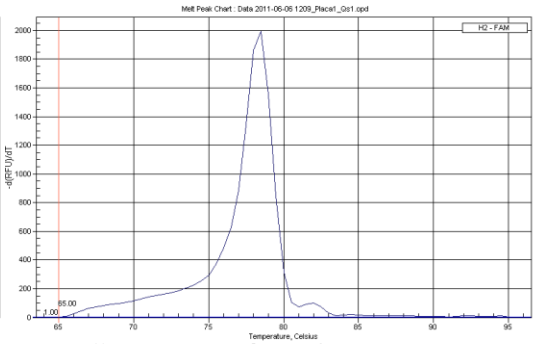
(g) Gene CCR with Tm of 77.50°C



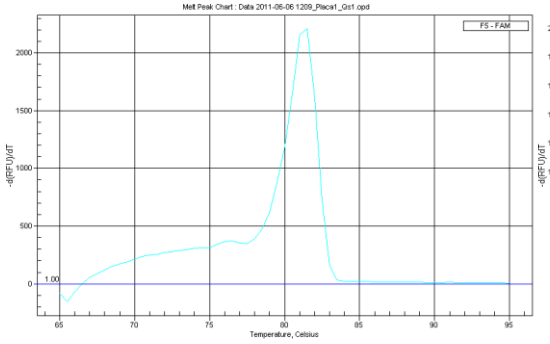
(h) Gene 4CL with Tm of 80.50°C



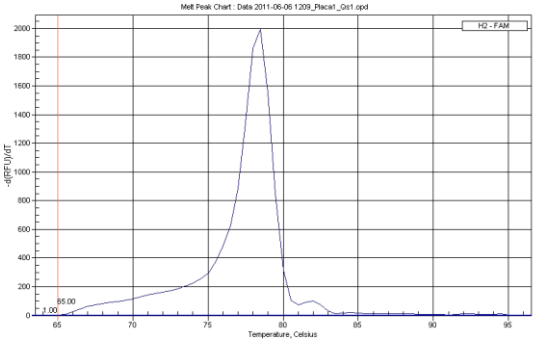
(i) Gene CYP84A1 with Tm of 77.50°C



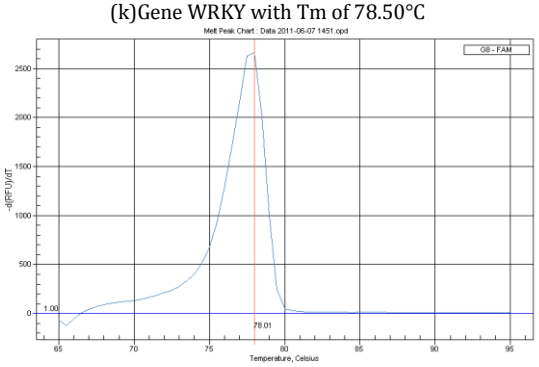
(j) Gene HCBT with Tm of 84.00°C



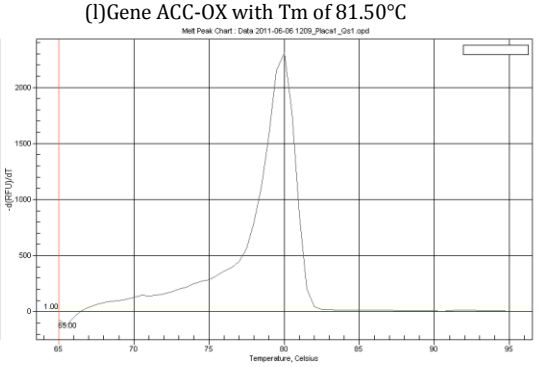
(k)Gene WRKY with Tm of 78.50°C



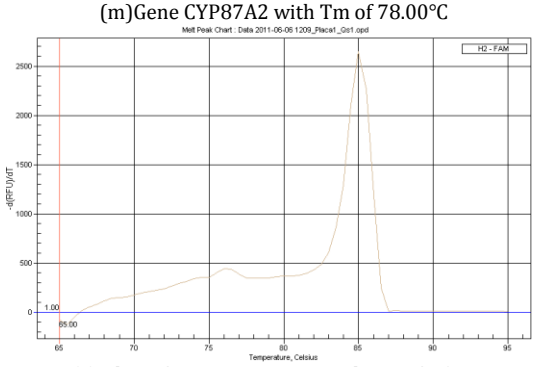
(l)Gene ACC-OX with Tm of 81.50°C



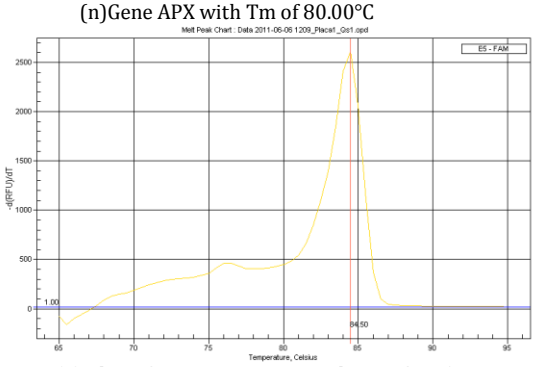
(m)Gene CYP87A2 with Tm of 78.00°C



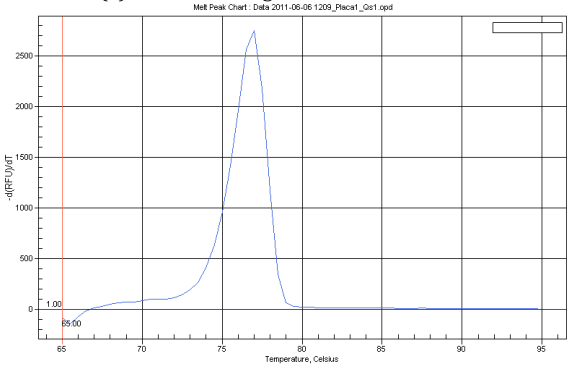
(n)Gene APX with Tm of 80.00°C



(o) The reference gene ACT with Tm of 85.00°C



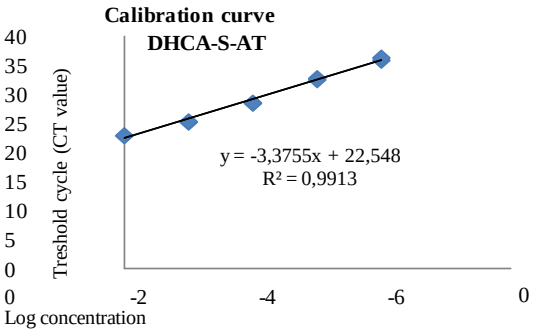
(p) The reference gene TUB with Tm of 84.50°C

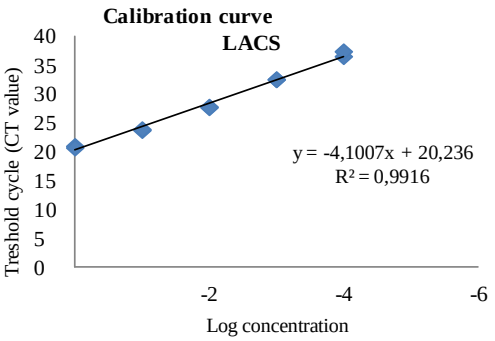


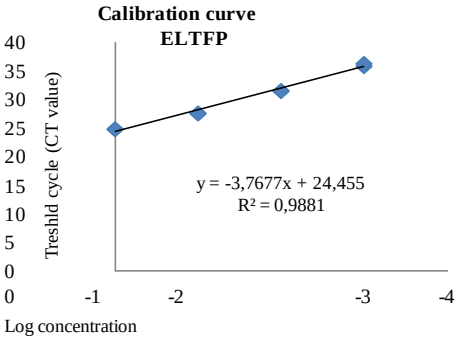
(q) Gene Dolaccase with Tm of 77.00°C

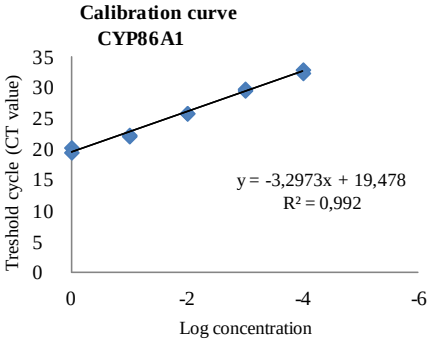
Appendix 8 - Calibration curves

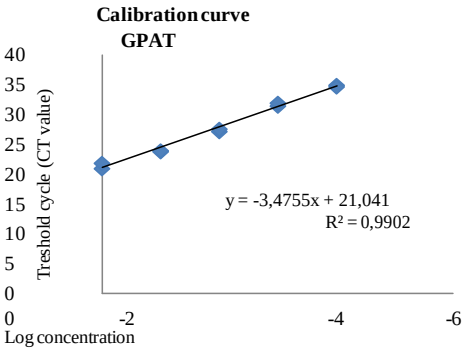
Dilution calibration curves constructed for the relative expression of the genes analyzed by qPCR, with the corresponding equation of the linear regression and the coefficient of determination (R^2).

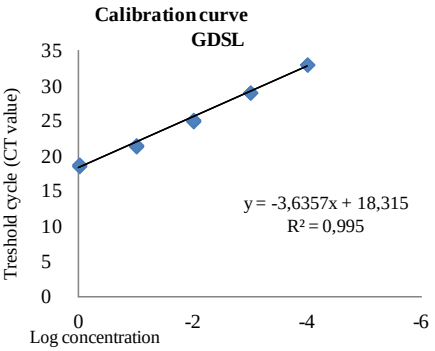


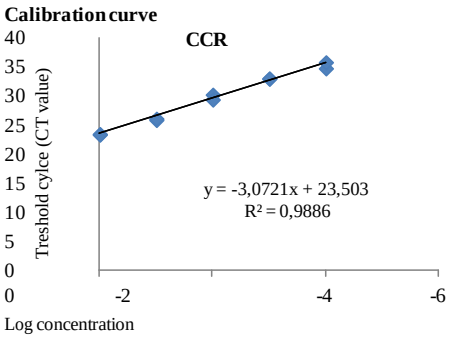


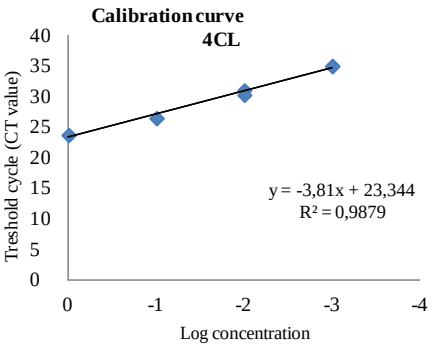


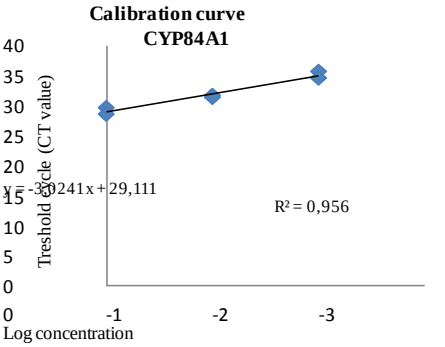


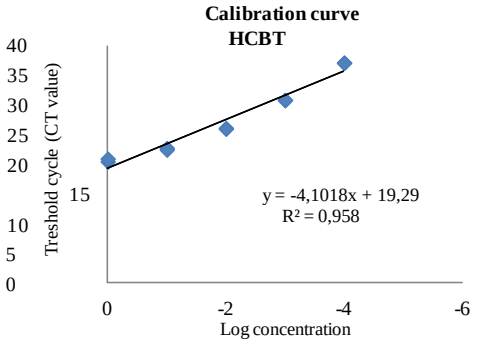


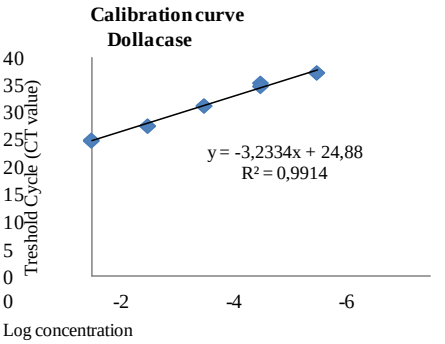


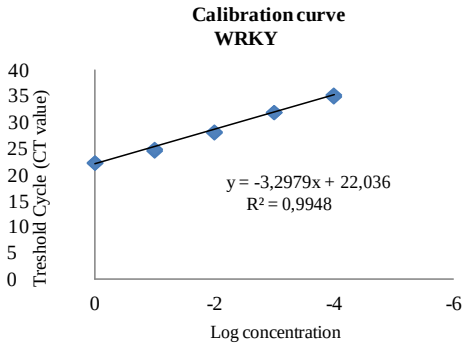


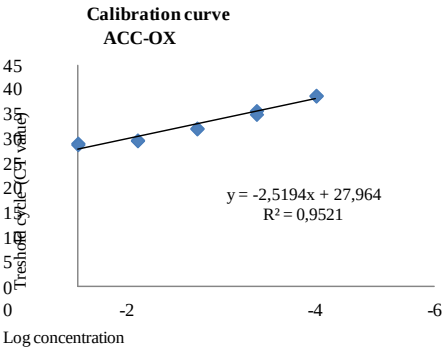


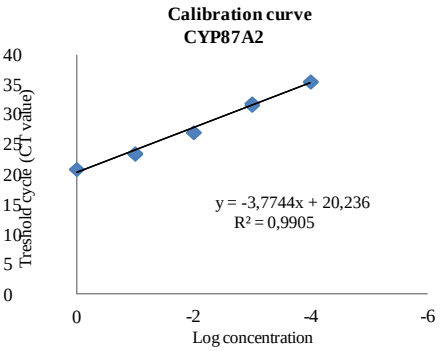

Calibration curve
CYP84A1
Treshold cycle (CT value)
y = -3,3241x + 29,111
R² = 0,956
Log concentration

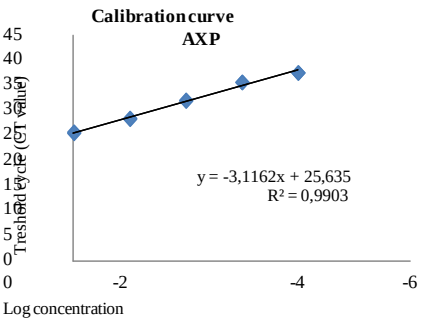

Calibration curve
HCBT
Treshold cycle (CT value)
y = -4,1018x + 19,29
R² = 0,958
Log concentration

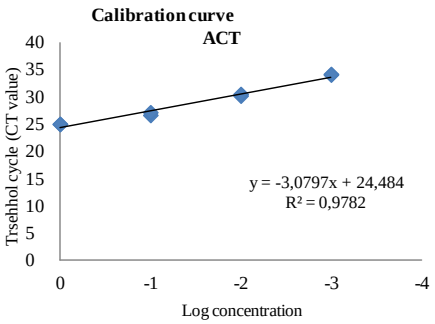

Calibration curve
Dollacase
Treshold Cycle (CT value)
y = -3,2334x + 24,88
R² = 0,9914
Log concentration

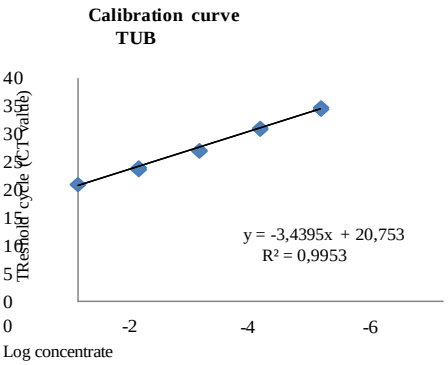

Calibration curve
WRKY
Treshold Cycle (CT value)
y = -3,2979x + 22,036
R² = 0,9948
Log concentration


Calibration curve
ACC-OX
Treshold cycle (CT value)
y = -2,5194x + 27,964
R² = 0,9521
Log concentration


Calibration curve
CYP87A2
Treshold cycle (CT value)
y = -3,7744x + 20,236
R² = 0,9905
Log concentration


Calibration curve
AXP
Treshold cycle (CT value)
y = -3,1162x + 25,635
R² = 0,9903
Log concentration


Calibration curve
ACT
Trsehhol cycle (CT value)
y = -3,0797x + 24,484
R² = 0,9782
Log concentration


Calibration curve
TUB
TReshold cycle (CT valdte)
y = -3,4395x + 20,753
R² = 0,9953
Log concentrate

Appendix 9 – Relative gene expression statistical analysis

Table 25. Statistical analysis of the relative expression of the genes analyzed in the samples of different cork quality developing tissues. It was performed one- way ANOVA to identify significant differences of expression of a gene among the six samples tested and later a post hoc Tukey-Kramer's test for the pairwise comparison analysis. The statistical analysis was done using the *GenEx 5.2.1.3 software* (MultiD Analyses AB, Göteborg, Sweden).

	Genes P-value														
Pair of samples	**DHCA-S-AT**	**LACS**	**ELTFP**	**CYP86A1**	**GPAT**	**GDSL**	**CCR**	**4CL**	**CYP84A1**	**HCBT**	**DoLaccase**	**WRKY**	**ACC-OX**	**CYP87A2**	**APX**
1 vs 2	0,98246298	0,03546884	0,98943066	0,0301415	0,57319623	0,01019125	0,00020908	0,05347043	0,00281538	0,26781886	0,35525537	0,47607384	7,00E-08	0,99996651	9,44E-05
1 vs 9	0,01376391	5,69E-05	0,00060565	1,95E-05	0,54197611	0,05655577	0,00019994	0,00168973	7,83E-06	0,00016881	0,48150011	0,22784434	0,99504646	0,0002856	0,1100622
1 vs 7	4,81E-06	5,91E-06	0,00062244	0,00014442	0,00356586	6,40E-07	3,39E-06	0,00021091	0,00013232	0,00010721	8,78E-05	0,00210124	0,46266917	0,95988033	8,00E-08
1 vs 8	1,85E-06	1,69E-06	1,77E-05	4,22E-06	7,74E-06	5,30E-07	9,00E-05	1,46E-06	2,40E-05	1,40E-06	1,37E-05	3,41E-06	4,58E-06	0,73120035	0,0022397
1 vs 11	3,04E-06	5,00E-08	5,48E-06	1,68E-06	5,00E-07	1,47E-06	1,20E-07	<1*10^-8	0,01121808	3,67E-06	4,68E-05	0,00011566	1,25E-06	4,91E-06	4,85E-05
2 vs 9	0,00454428	0,01167969	0,0015	0,0032934	0,99999991	0,90325394	0,99999999	0,35311411	0,00945088	0,00548588	0,01945299	0,99179065	5,00E-08	0,00022162	0,00711396
2 vs 7	2,38E-06	0,00056606	0,00154361	0,04486621	0,05803928	7,37E-05	0,04756037	0,04008069	0,36628315	0,00318639	6,98E-06	0,04469653	3,10E-07	0,91244721	0,00017094
2 vs 8	9,60E-07	0,00010477	3,65E-05	0,00041391	5,77E-05	5,72E-05	0,98741809	6,27E-05	0,04751898	1,68E-05	1,46E-06	2,79E-05	0,00597358	0,63008703	0,31811875
2 vs 11	1,54E-06	9,30E-07	1,07E-05	0,00011904	2,55E-06	0,00024107	0,00015738	<1*10^-8	0,95358767	5,33E-05	4,11E-06	0,00170456	0,0499981	4,05E-06	0,99466581
9 vs 7	0,00098926	0,44469177	1	0,63332374	0,06358951	2,06E-05	0,05017871	0,72599529	0,27026131	0,99929778	0,00121719	0,11292178	0,23927798	0,00094513	9,40E-07
9 vs 8	0,00025787	0,07704898	0,15764469	0,75453798	6,19E-05	1,63E-05	0,99007487	0,00121311	0,92251781	0,01437518	0,00013878	5,65E-05	2,69E-06	0,00233641	0,24711794
9 vs 11	0,00051935	0,00011308	0,02922549	0,29005646	2,70E-06	6,09E-05	0,00016447	3,00E-08	0,00239004	0,0696673	0,00058378	0,00419291	7,70E-07	0,05872543	0,00308045
7 vs 8	0,93574667	0,83089756	0,15322742	0,10410609	0,00727616	0,99994901	0,13095127	0,01171564	0,76514546	0,02524283	0,67450654	0,00363918	4,05E-05	0,99051809	1,11E-05
7 vs 11	0,99766035	0,0018217	0,02833267	0,0237046	0,00012208	0,94636494	0,03159849	7,00E-08	0,10618327	0,1198214	0,99584236	0,40701481	8,93E-06	1,22E-05	0,00035227
8 vs 11	0,99588162	0,01284927	0,90020279	0,94240245	0,14585855	0,88526039	0,00037695	3,46E-06	0,01151109	0,92860471	0,90641394	0,0981343	0,79743819	2,43E-05	0,15080043

5. References

Andersson-Gunneråsa S, Hellgren J, Björklund S. et al. Asymmetric expression of a poplar ACC oxidase controls ethylene production during gravitational induction of tension wood. The Plant Journal. 2003;34(3):339-349.

Adam-Blondon A, Bernole A, Faes G, Lamoureux D, Pateyron S, Grando M, Caboche M, Velasco R, Chalhoub B. Construction and characterization of BAC libraries from major grapevine cultivars. Theor Appl Genet. 2005; 110:1363–1371.

Alves S, Ribeiro T, Inácio V, Rocheta M. Genomic organization and dynamics of repetitive DNA sequences in representatives of three Fagaceae genera. Genome. 2012; 55:348–359.

Ammiraju J, Luo M, Goicoechea J. The Oriza bacterial artificial chromosome library resource: construction and analysis of 12 deep coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. Genome Research. 2006; 16:140-147.

Anderson, C. Genome shortcut leads to problems. Science. 1993; 259:1684-1687.

Bao W, Kapitonov V, Jurka J. Ginger DNA transposons in eukaryotes and their evolutionary relationships with long terminal repeat retrotransposons. Mob DNA. 2009; 1:3

Barch MJ, Knusten T, Spurbeck JL. The AGT cytogenetics laboratory manual. 3rd ed. New York: Raven Press.1997; 25–50

Bartos J, Paux E, Kofler R, Havrankova M, Kopecky D, Suchankova P, Safar J, Simkova H, Town CD, Lelley T, Feuillet C, Dolezel J. A first survey of the rye (Secale cereale) genome composition through BAC end sequencing of the short arm of chromosome 1R. BMC Plant Biol. 2008; 8(95).

Bautista R, Villalobos D, Diaz-Moreno S, Canton F, Canovas F, Claros M. Toward a Pinus pinaster bacterial artificial chromosome library. Annals of Forest Science. 2007; 64(8): 855-864.

Bautista R, Villalobos D, Diaz-Moreno S, Canton F, Canovas F, Claros M. New strategy for Pinus pinaster genomic library construction in bacterial artificial chromosomes. Investigacion Agraria-Sistemas Y Recursos Forestales. 2008; 17(3):238-249.

Beisson F, Li Y, Bonaventure G, Pollard M and Ohlrogge JB. The Acyltransferase GPAT5 Is Required for the Synthesis of Suberin in Seed Coat and Root of Arabidopsis. The Plant Cell Online. 2007;19(1): 351-368.

Besendorfer V, Zoldos V, Peskan T, Krsnik-Rasol M, Littvay T, Papes D. Identification of potential cytogenetical and biochemical markers in bioindication of common oak forests. Phyton. 1996; 36: 139–146

Bernards M. Demystifying suberin. Canadian Journal of Botany. 2002;80(3): 227-240.

Bernards M, Lewis N. Alkyl ferulates in wound healing potato tubers. Phytochemistry. 1992;31(10):3409-3412.

Bernards M, Lewis N. The macromolecular aromatic domain in suberized tissue: A changing paradigm. Phytochemistry. 1998;47(6): 915-933.

Bogunic F, Muratovic E, Brown S, Siljak-Yakovlev S. Genome size and base composition of five Pinus species from the Balkan region. Plant Cell Rep. 2003; 22: 59–63

Boysen C, Simon MI, Hood L. Analysis of the 1.1-Mb Human α/δ T-Cell Receptor Locus with Bacterial Artificial Chromosome Clones. Genome Res. 1997; 7: 330–338.

Burke, D, Carle G, Olsen M. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science. 1987; 236: 806-812.

Burke, D. YAC cloning: options and problems. GATA. 1990; 7: 94-99.

Cai W, Reneker J, Chow C, Vaishnav M, Bradley A. An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. *Genomics*. 1998; 54: 387-397.

Caritat A, Gutiérrez E and Molinas M. Influence of weather on cork-ring width. Tree Physiology. 2000; 893-900.

Carvalho M, Ribeiro T, Viegas W, Morais-Cecilio L, Rocheta M. Presence of env-like sequences in Quercus suber retrotransposons. Journal of applied genetics. 2010; *51*: 461–7.

Caspersson T, Farber S, Foley CE, Kudynowski J, Modest EJ, Simonsson E, Waugh U, Zech L Exp. Cell Res. 1968; 49, 219-222.

Cavagnaro PF, Chung SM, Szklarczyk M, Grzebelus D, Senalik D, Atkins AE, Simon PW. Characterization of a deep-coverage carrot (Daucus carota L.) BAC library and initial analysis of BAC-end sequences. Molecular Genetics and Genomics. 2009; 281(3):273-288.

Cerbah M, Coulaud J, Siljak-Yakovlev S. rDNA organization and evolutionary relationships in the genus Hypochaeris (Asteraceae). J Hered. 1998a; 89: 312–318

Cerbah M, Kevei Z, Yakovlev S, Kondorosi E, Trinh TH. rDNA organization and heterochromatin pattern in Medicago truncatula. (Abstr 13th Int Chromo Conf). Cytogenet Cell Genet. 1998b;81: 141

Chaves, I, Pinheiro C, Paiva JAP, Planchon S, Sergeant K, Renaut J, Graça JA, Ricardo PP, Costa G, Coelho AV. Proteomic evaluation of wound-healing processes in potato (Solanum tuberosum L .) tuber tissue. Proteomics. 2009;4154-4175.

Chen M, Presting G, Barbazuk W, Goicoechea L, Blackmon B et al. An integrated physical and genetic map of the rice genome. Plant Cell. 2002; 14:537-545.

Cheng Z, Presting G, Buel C, Wing R, Jiang J. High-resolution pachytene chromosome mapping of bacterial artificial chromosomes by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. Genetics. 2001*b*; 157: 1749-1757

Chokchaichmnankit P, Anamthawat-Jonsson K, Chulalaksananukul W. Chromosomal Mapping of 18S-25S and 5S Ribosomal Genes on 15 Species of Fagaceae from Northern Thailand. Silvae Genetica. 2008; 57: 5-13.

Coe E, Cone K, McMullen M, Chen S-S, Davis G, Gardiner J, Liscum E, Polacco M, Paterson AH, Sanchez-Villeda H, Soderlund C, Wing RA. Access to the maize genome: An integrated physical and genetic map. Plant Physiol. 2002; 128:9-12.

Collins T, Sander T. The Superfamily of SCAN Domain Containing Zinc Finger Transcription Factors. Madame Curie Bioscience Database. Landes Bioscience. 2000.

Costa F, Pereira T, Hodnett G, Pereira M, Stelly DCoquet C, Bauza E, Oberto G, Berghi A, Farnet AM, Ferre E, Peyronel D, Dal Farra C, Domloge N. Quercus suber cork extract displays a tensor and smoothing effect on human skin: An in vivo study. Drugs under experimental and clinical research. 2005;31(3):89-99.

Costa A, Pereira H, Oliveira A. A dendroclimatological approach to diameter growth in adult cork-oak trees under production. Trees-Structure and Function. 2001;15(7):438-443.

Fluorescent in situ hybridization of 18S and 5S rDNA in papaya (Carica papaya l.) and wild relatives. Caryologia. 2008; *61*(4), 411–416.

D'Auria J. Acyltransferases in plants: a good time to be BAHD. Current opinion in plant biology. 2006;9(3):331-340.

Darlington C, Wylie A. Chromosome atlas of flowering plants. George Allen and Unwin Ltd.,London. 1995; 519.

Datema E, Mueller LA, Buels R, Giovannoni JJ, Visser RGF, Stiekema WJ, van Ham RCGJ. Comparative BAC end sequence analysis of tomato and potato reveals overrepresentation of specific gene families in potato. BMC Plant Biol. 2008; 8:34.

Devi J, Ko J, Seo B. FISH and GISH : Modern cytogenetic techniques.Indian Journal of Biotechnology. 2005; 4:307–315.

Dietrich C, Perera M, Meeley R, Nikolau B, Schnable P. Characterization of two GL8 paralogs reveals that the 3-ketoacyl reductase component of fatty acid elongase is essential for maize (*Zea mays* L.) development. Plant Journal. 2005;42:844-861.

Doyle J, Doyle L. Isolation of plant DNA from fresh tissue. Focus. 1990;12:13-15. Eckerman C. Ekman R.Pap. Puu 1985. 3:100

Duminil J, Pemonge M, Petit R. A set of 35 consensus primer pairs amplifying genes and introns of plant mitochondrial DNA. Molecular Ecology Notes. 2002; 2(4):428-430.

Eickbush T, Jamburuthugoda V. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res. 2008; 134:221-234.

Eulgem T, Rushton P, Robatzek S. The WRKY superfamily of plant transcription factors. Trends in Plant Science. 2000;5:199-206.

Eustice M, Yu Q, Lai C, Hou S, Thimmapuram J, Liu L, Alam M, Moore P, Presting G, Ming R. Development and application of microsatellite markers for genomic analysis of papaya. Tree Genet Genomes. 2008; 4: 333-341.

Fabiane, R., Pereira, T. N. S., Hodnett, G. L., Pereira, M. G., and Stelly, D. M. (2008). Fluorescent in situ hybridization of 18S and 5S rDNA in papaya (Carica papaya l .) and wild relatives, *61*(4): 411–416.

Favre J, Brown S. A flow cytometric evaluation of the nuclear DNA content and GC percent in genomes of European oak species. Ann. Sci. For. 1996; 53: 915-917.

Fialho C, Lopes F, Pereira H. The effect of cork removal on the radial growth and phenology of young cork oak trees. Forest Ecology and Management. 2001;141(3): 251-258.

Filomena M, Bento S, Pereira H. A study of variability of suberin composition in cork from *Quercus suber L* , using thermally assisted transmethylation GC – MS. Journal of Analytical and Applied Pyrolysis. 2001;57: 45 - 55.

Fladung M, Kaufmann H, Markussen T, Hoenicka H. Construction of a Populus tremuloides Michx. BAC library. Silvae Genetica. 2008; 57(2):65-69.

Flavell R, Bennett M, Smith J, Smith D. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochemical Genetics. 1974; 4: 257-269,

Franke R, Briesen I, Wojciechowski T, et al. Apoplastic polyesters in Arabidopsis surface tissues - A typical suberin and a particular cutin. Phytochemistry. 2005;66(22):2643-2658.

Franke R, Schreiber L. Suberin — a biopolyester forming apoplastic plant interfaces. Current Opinion in Plant Biology. 2007;10:252-259.

Frelichowski J, Palmer M, Main D, Tomkins J, Cantrell R, Stelly D, Yu J, Kohel R, Ulloa M. Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Mol. Gen. Genomics. 2006; 479–491.

Gall J, Pardue M. Formation and detection RNA-DNA hybrid molecules in cytological preparations. Proc. Natl. Acad. Sci. U.S.A. 1969; 63: 378-383.

Gallois A, Burrus M, Brown S. Evaluation of the nuclear DNA content and GC percent in four varieties of Fagus sylvatica L. Ann Forest Sci. 1999; 56: 615- 618.

Gandini A, Neto CP, Silvestre AJD. Suberin: A promising renewable resource for novel macromolecular materials. Progress in Polymer Science. 2006;31(10):878-892.

Gill B, Friebe B. Plant cytogenetics at the dawn of the 21st century. Current opinion in plant biology. 1998; 2:109–15.

Glockner G, Szafranski K, Winckler T, Dingermann T, Quail M, Cox E, Eichinger L, Noegel A, Rosenthal A. The complex repeats of Dictyostelium discoideum. Genome Res. 2001; 11:585-594

Gonzalez V, Garcia-Mas J, Arus P, Puigdomenech P. Generation of a BAC based physical map of the melon genome. BMC Genomics. 2010; 11:339.

Graça J. Hydroxycinnamates in suberin formation. Protoplasma. 2010;85-91.

Graca J, Pereira H. Suberin Structure in Potato Periderm: Glycerol. Long-Chain Monomers. and Glyceryl and Feruloyl Dimers. Journal of Agricultural and Food Chemistry. 2000;48(11):5476-5483.

Graça J, Pereira H. The periderm development in. 2004;25(3):325 -335.

Graça J, Santos S. Linear Aliphatic Dimeric Esters from Cork Suberin. Biomacromolecules. 2006;7(6):2003-2010.

Graça J, Santos S. Suberin: a biopolyester of plants' skin. Macromolecular bioscience. 2007;7(2):128-135.

Green E, Riethman H, Dutchik J, Olson M. Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics.* 1991; 11:658 - 669.

Gu Y, Ma Y, Huo N, Vogel J, You F, Lazo G, Nelson W, Soderlund C, Dvorak J, Anderson O, Luo M. A BAC based physical map of Brachypodium distachyon and its comparative analysis with rice and wheat. BMC Genomics. 2009; 10:496.

Hamilton C, Frary A, Xu Y, Tanksley S, Zhang H. Construction of tomato genomic DNA libraries in a binary-BAC (BIBAC) vector. Plant Journal. 1999; 18(2):223-229.

Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. 1999; 41:95-98.

Harris P, Ferguson L. Dietary fibres may protect or enhance carcinogenesis. Mutation Research/Genetic Toxicology and Environmental Mutagenesis. 1999;443(1-2):95-110.

Hasterok R, Marasek A, Donnison I, Armstead I, Thomas A, King I, Wolny E, et al. Alignment of the genomes of Brachypodium distachyon and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence in situ hybridization. Genetics. 2006; 173(1), 349–62.

Havecker E, Gao X, Voytas D. The diversity of LTR retrotransposons. Genome Biol. 2004; 5:225.

Heslop-Harrison J. Genes in evolution: the control of diversity and speciation. Annals of Botany 2010; 106:437-438.

Heslop-Harrison J. Comparative genome organization in plants: From sequence and markers to chromatin and chromosomes. Plant Cell. 2000; 12: 617–635.

Hong C, Lee S, Park J, Plaha P, Park Y, Lee Y, Choi J, Kim K, Lee J, Lee J, Jin H, Choi S, Lim Y. Construction of a BAC library of Korean ginseng and initial analysis of BACend sequences. Mol Genet Genomics. 2004; 271:709–716

Hribová E, Neumann P, Matsumoto T, Roux N, Macas J, Dolezel J. Repetitive part of the banana (Musa acuminata) genome investigated by low depth 454 sequencing. BMC Plant Biol. 2010; 10:204.

Islam-Faridi M, Childs K, Klein P, Hodnett G, Menz M, Klein R, et al. A molecular cytogenetic map of sorghum chromosome1: Fluorescence *in situ* hybridization analysis with mapped bacterial artificial chromosomes. Genetics. 2002; 161: 345-353.

Jackson S, Cheng Z, Wang M, Goodman H, Jiang J. Comparative fluorescence in situ hybridization mapping of a 431-kb Arabidopsis thaliana bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the Brassica rapa genome. Genetics. 2000; 156: 833–838.

Jaillon O, Aury J, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 2007; 449:463-467.

Jaynes R. 1962. Chesnut chromosomes. Forest Sci 8: 372-377.

Jenkins G, Hasterok R. BAC 'landing' on chromosomes of Brachypodium distachyon for comparative genome alignment. Nature Protocols. 2006; 1: 6

Jiang L, Dimitrov G, Tran K, Shetty J, Malek JA, Feldblyum T, Nierman WC, Fraser CM. Mouse BAC ends quality assessment and sequence analyses. Genome Res. 2001; 11:1736–1745

Jiang J, Gill B, Wang G, Ronald P, Ward D. Genetics Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc. Natl. Acad. Sci. USA. 1995; 92: 4487-4491

Jiang J, Gill B. Nonisotopic in situ hybridization and plant genome mapping: The first 10 years. Genome. 1994; 37: 717–725.

Jiang J, Gill B. Current status and potential of fluorescence in situ hybridization in plant genome mapping. Genome Mapping in Plants (A. H. Paterson,ed.). 1996; 127–135.

Jiang J, Gill B. Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome. 2006; 1057–1068.

Jong J, Paul Fransz P, Zabel P. High resolution FISH in plants –techniques and applications. Trends in Plant Science. 1999; 4 (7).

Kelley J, Rounsley S, Field C, Craven M, Bocskai D, Adams M, Kim U. High throughput direct end sequencing of BAC clones. Nucleic Acids Research. 1999; 7: 1539–1546.

Kelley J, Field C, Craven M, Bocskai D, Kim U, Rounsley S, Adams M. High throughput direct end sequencing of BAC clones. Nucleic acids research. 1999; *27*(6): 1539–46.

Kim H, San Miguel P, Nelson W, Collura K, Wissotski M, Walling J, Kim J, Jackson S, Soderlund C, Wing R. Comparative physical mapping between Oryza sativa (AA genome type) and O. punctata (BB genome type). Genetics. 2007; 176:379-390.

Kim J, Childs K, Islam-Faridi MN, Menz M, Klein R, Klein P, et al. 2002. Integrated karyotyping of sorghum by in situ hybridization of landed BACs. Genome. 2002; 45: 402-412

Klein R, Morishige D, Klein P, Dong J, Mullet J. High Throughput BAC DNA Isolation for Physical Map Construction of Sorghum (Sorghum bicolor). 1998; 16: 351–364.

Kolattukudy P. Polyesters in higher plants. Biopolyesters. 2001;1-49.

Kolattukudy P. Biopolyester Membranes of Plants: Cutin and Suberin. Science. 1980; 208(4447): 990-1000.

Kofler R, Schlötterer C, Lelley T. SciRoKo: A new tool for whole genome microsatellite search and investigation. Bioinformatics. 2007; 23(13): 1683-1685.

Kremer A, Casasoli M, Barreneche T, Bodénès C, Sisco P, Kubisiak T, Scalfi M, Leonardi S, Bakker E, Buiteveld J, Severson J, Arumuganathan K, Derory J, Saintagne C, Roussel G, Bertocchi M, Lexer C, Porth I, Hebard F, Clark C, Carlson J, Plomion C, Koelewijn H, Villani F. Fagaceae Trees. In Genome Mapping and Molecular Breeding in Plants. Volume 7. 2007;161-184.

Lander E, Linton L, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001: 860–921

Lai C, Yu Q, Hou S, Skelton R, Jones M, Lewis K, Murray J, Eustice M, Guan P, Agbayani R, Moore P, Ming R, Presting G. Analysis of papaya BAC end sequences reveals first insights into the organization of a fruit tree genome. Mol Gen Genomics. 2006; 276:1–12

Lapitan N, Brown S, Kennard W, Stephens J, Knudson D. FISH physical mapping with barley BAC clones. Plant J. 1997; 11: 149-156.

Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira LP, Jourdan I, Pot D, Andrade AC, et al. Construction and characterization of a Coffea canephora BAC library to study the organization of sucrose biosynthesis genes. Theoretical and Applied Genetics. 2005; 111(6):1032-1041

Liang H, Fang E, Tomkins J, Luo M, Kudrna D, Kim H, Arumuganathan K, Zhao S, Leebens-Mack J, Schlarbaum S, et al. Development of a BAC library for yellow-poplar (Liriodendron tulipifera) and the identification of genes associated with flower development and lignin biosynthesis. Tree Genetics and Genomes. 2007; 3(3):215-225.

Llaca V, Lou A, Young S, Messing J. Nested retrotransposons in Oriza sativa. 1998 *(unpublished)*.

Lopes M, Barros A. Pascoal Neto C. et al. Variability of cork from Portuguese Quercus suber studied by solid-state 13C-NMR and FTIR spectroscopies. Biopolymers. 2001;62(5):268-277.

Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu J, Tamarit D, Aguilar-Rodriguez J, Vicente-Ripolles M, Fuster G, Bernet G, Maumus F, Munoz-Pomer A, Sempere J, LaTorre A, Moya A. The Gypsy Database (GyDB) of Mobile Genetic Elements: Release 2.0 Nucleic Acids Research. 2011; 39.

Llorens C, Munoz-Pomer A, Bernad L, Botella H, Moya A. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biology Direct. 2009; 4:41

Livak J, Schmittgen T. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T) Method. Methods. 2001; *25*(4), 402-8.

Lopes M, Barros A, Pascoal C. et al. Variability of cork from Portuguese Quercus suber studied by solid-state 13C-NMR and FTIR spectroscopies. Biopolymers. 2001;62(5):268-277.

Lotfy S, Negrel J, Javelle F. Formation of ω-feruloyloxypalmitic acid by an enzyme from wound-healing potato tuber discs. Phytochemistry. 1994;35(6):1419-1424.

Lulai E, Suttle J. The involvement of ethylene in wound-induced suberization of potato tuber (*Solanum tuberosum L.*): a critical assessment. Post harvest Biology and Technology. 2004;34(1):105-112.

Lysak M, Fransz P, Ali H, Schubert I. Chromosome painting in Arabidopsis thaliana. Plant J. 2001; 28: 689–697.

Lysak M, Pecinka A, Schubert I. Recent progress in chromosome painting of Arabidopsis and related species. Chromosome Res. 2003; 11: 195–204.

Mahalingam R, Gomez-Buitrago A, Eckardt N, Shah N, Guevara-Garcia A DP, Raina R FN. Characterizing the stress/defense transcriptome of Arabidopsis. Genome Biol. 2003;4 (3):R20.

Mahairas G, Wallace J, Smith K, Swartzell S, Holzman T, Keller A, Shaker R, Furlong J, Young J, Zhao S, Adams M, Hood L. Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome. Proc. Natl. Acad. Sci. USA. 1999; 9739–9744.

Maluszynska J. In situ hybridisation in plants - methods and application. In Molecular Techniques in Crop Improvement (ed. Jain, S.M., Ahloowalia, B.S. and Brar,D.S.) 299–326 (Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002).

Marie D, Brown S. A cytometric exercise in plant DNA histograms with 2C values for 70 species. Biol Cell. 1993; 78:41–51.

Marín I. GIN transposons: genetic elements linking retrotransposons and genes. Mol Biol Evol. 2010 27:1903-11.

Martel E, De Nay D, Siljak-Yakovlev S, Brown SC, Sarr A. Genome size and base composition in pearl millet and fourteen related species. J Hered. 1997; 88:139–143

Mao L, Wood T, Yu Y, Budiman MA, Tomkins J, Woo S, Sasinowski M, Presting G, Frisch D, GoV S, Dean RA, Wing RA. Rice transposable elements: a survey of 73,000 sequence tagged connectors. Genome Res. 2000; 10:982–990

Moire L, Schmutz A, Buchala A, et al. Glycerol Is a Suberin Monomer. New Experimental Evidence for an Old Hypothesis. Plant Physiology. 1999;119(3):1137-1146.

Moisy C, Garrison K, Meredith C, Pelsy F. Characterization of ten novel Ty1/copia-like retrotransposon families of the grapevine genome. BMC Genomics. 2008; 9:469.

Molinas M, Serra O, Hohn C, et al. A feruloyl transferase involved in the biosynthesis of suberin and suberin-associated wax is required for maturation and sealing properties of potato periderm. 2010;277-290.

Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martienssen R, Meier-Ewert S, Altmann T. A complete BAC-based physical map of the Arabidopsis thaliana genome. Nature Genet. 1999; 22:271-275.

Mun J, Kim D, Choi H, Gish J, Debelle F, Mudge J, Denny R, Endre G, Saurat O, Dudez A, et al. Distribution of microsatellites in the genome of Medicago truncatula: A resource of genetic markers that integrate genetic and physical maps. Genetics. 2006; 172:2541-2555.

Mun J, Kwon S, Yang T, Kim H, Choi B, Baek S, Kim J, Jin M, Kim J, Lim M, Lee S, Kim H, Kim H, Lim Y, Park B. The first generation of a BAC-based physical map of Brassica rapa. BMC Genomics. 2008; 9:280.

Narusaka Y, Narusaka M, Seki M, et al. Crosstalk in the responses to abiotic and biotic stresses in Arabidopsis: Analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray. Plant Molecular Biology. 2004;55(3):327-342.

Natividade J. Portugal. Subericultura. Ministério da Economia. Direcção Geral dos Serviços Florestais e Aquícolas; 1950.

Nederlof P, Flier S, Wiegant J, Raap A, Tanke H et al. Multiple fluorescence *in situ* hybridization. Cytometry. 1990; 11: 126-131.

Neil D, Villasante A, Fisher R, Vetrie D, Cox B, Tyler-Smith C. Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. Nucleic Acids Res. 1990; 18:1421-1428.

Nowotny M. Retroviral integrase superfamily: the structural perspective. EMBO Reports . 2009; 144-51.

O'Connor M, Peifer M, Bender W. Construction of large DNA segments in *Escherichia coli*. Science. 1989; 244: 1307-1312.

Ohri D, Ahuja M. Giemsa C-banded karyotype in Quercus L (oak). Silvae Genet. 1990; 39 (5-6), 216-219.

Paiva J, Prat E, Vautrin S, Santos M, San-Clemente H, Brommonschenkel S, Fonseca P, Grattapaglia D, Song X, Ammiraju J, Kudrna D, Freitas A, Bergès H, Grima-Pettenati J, Wing R. Advancing Eucalyptus genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. BMC Genomics. 2011a; 12:137.

Paiva J, Fevereiro P, Marques P, Rodrigues J, Provost G, Plomion C, Grima-Pettenati J, Bouchez O, Klopp C, Bergès H, Graça J. Deciphering cork formation in Quercus suber. BMC Proceedings. 2011b, 5: 172.

Pastuglia M, Ruffio-Chable V, Delorme V, Gaude T, Dumas C, Cock J. A functional S locus anther gene is not required for the self-incompatibility response in Brassica oleracea. Plant Cell. 1997.9:2065-2076.

Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C. Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J. 2006; 48(3):463–474.

Peleman J, Cottyn B, Van Camp W, Van Montagu M, Inzé D. Transient occurrence of extrachromosomal DNA of an Arabidopsis thaliana transposon-like element, Tat1. Proc. Natl. Acad. Sci. U S A. 1991; 88: 3618-22.

Pereira H. Chemical composition and variability of cork from Quercus suber L . Wood Science and Technology. 1988;218:211-218.

Pereira H. The thermochemical degradation of cork. Wood Science and Technology. 1992;26(4):259-269.

Pereira H, Lopes F, Graça J. The evaluation of the quality of cork planks by image analysis. Holzforschung-International Journal of the Biology. Chemistry. Physics and Technology of Wood. 1996;50(2):111-115.

Peterson D, Tomkins J, Frisch D, Wing R, Paterson A. Construction of plant bacterial artificial chromosome (BAC) libraries: An illustrated guide. Journal of Agricultural Genomics. 2002; 5.

Pine D. Brief Communications Inheritance and Subcellular Localization of Triose- Phosphate Isomerase in. Journal of Heredity. 1982;271-300.

Pinkel D, Straume T, Gray J. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences of the United States of America*. 1986; *83*(9): 2934–8.
Poulsen TS, Johnsen HE. BAC End Sequencing. Methods in Molecular Biology. 2004; 255: 157-161

Pollard M, Beisson F, Li Y, Ohlrogge J. Building lipid barriers: biosynthesis of cutin and suberin. Trends in Plant Science. 2008.

Provost G, Herrera R, Paiva J, et al. A micromethod for high throughput RNA extraction in forest trees. Biological research. 2007;40(3):291-7.

Rampant P, Lesur I, Boussardon C, Bitton F, Magniette ML, Bodénès C, et al.. Analysis of BAC end sequences in oak, a keystone forest tree species, providing insight into the composition of its genome. BMC Genomics. 2011; 12:292.

Ribeiro T, Barão A, Viegas W, Morais-Cecílio L. Molecular cytogenetics of forest trees. Cytogenet. Genome Res. 2008;120(3–4): 220–227.

Ribeiro T, Loureiro J, Santos C, Morais-Cecílio L. Evolution of rDNA FISH patterns in the Fagaceae. Tree Genetics and Genomes. 2011; *7*(6): 1113–1122.

Rocheta M, Carvalho L, Viegas W, Morais-Cecílio L. Corky, a gypsy-like retrotransposon is differentially transcribed in Quercus suber tissues. BMC research notes. 2012; *5*: 432.

Rosenblum B, Lee L, Spurgeon S, Khan S, Menchen S, Heiner C, Chen S. New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res. 1997; 25:4500-4504.

Safár J, Noa-Carrazana J, Vrána J, Barton J, Alkhimova O, Sabau X, Simková H, Lheureux F, Caruana M, Dolezel J, PiVanelli P. Creation of a BAC resource to study the structure and evolution of the banana (*Musa balbisiana*) genome. Genome. 2004; 47:1182–1191

Sambrook J, Fritsch EF, Maniatis T. Molecularc cloning: a laboratory manual, 2nd ed. 1989. Cold Spring Harbor Laboratory Press.

Sander T, Stringer K, Maki J, Szauter P, Stone J, Collins T. The SCAN domain defines a large family of zinc finger transcription factors. Gene. 2003; 310:29-38

SanMiguel P, Gaut B, Tiknonov A, Nakajima Y, Bennetzen J. The paleontology of intergene retrotransposons of maize. Nature Genetics. 1998; 20:43–45.

Schlueter J, Goicoechea J, Collura K, Gill N, Lin J-Y, Yu Y, Kudrna D, Zuccolo A, Vallejos C, Munoz-Torres M, Blair M, Tohme J, Tomkins J, McClean P, Wing R, Jackson S. BAC-end sequence analysis and a draft physical map of the common bean (Phaseolus vulgaris L.) genome. Tropical Plant Biol. 2008; 1:40-48.

Schmutz A, Buchala A, Ryser U, Changing the Dimensions of Suberin Lamellae of Green Cotton Fibers with a Specific Inhibitor of the Endoplasmic Reticulum-Associated Fatty Acid Elongases. Plant Physiology. 1996;110(2):403-411.

Schwarzacher T, Heslop-Harrison J. ''Practical In Situ Hybridization.'' Oxford Press, New York, NY. 2000 .

Sehgal S, Li W, Rabinowicz P, Chan A, Šimková H, Doležel J, Gill B. Chromosome arm-specific BAC end sequences permit comparative analysis of homoeologous chromosomes and genomes of polyploid wheat. BMC Plant Biology. 2012; 12:64

Serra O, Molinas M, Huguet G, Fluch S. A Genomic Approach to Suberin Biosynthesis. Society. 2007;144:419-431.

Shizuya H, Birren B, Kim U, Mancino V, Slepak T, et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc. Natl. Acad. Sci. USA. 1992; 89: 8794–8797.

Shultz J, Kazi S, Bashir R, Afzal J, Lightfoot D. The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. Theor. Appl. Genet. 2007; 1081–1090.

Silva S, Sabino M, Fernandes E, Correlo V, Boesel L, Reis R. Cork: properties. capabilities and applications. International Materials Reviews. 2005;50:345-365(21).

Silva J. Os Montados muito para além das árvores. Lisboa. Público. 2007; Comunicação Social SA; Fundação Luso-Americana para o Desenvolvimento

Sitte P. Zum feinbau der suberinschichten im flaschenkork. Protoplasma. 1962;54(4):555-559.

Soler M. Molecular genetics of cork formation. PhD Thesis. Univeritat de Girona. 2008.

Soler M, Serra O, Molinas M, et al. A Genomic Approach to Suberin Biosynthesis and Cork Differentiation. Plant Physiology. 2007;144(1):419-431.

Soler M, Serra O, Molinas M, et al. Seasonal variation in transcript abundance in cork tissue analyzed by real time RT-PCR. Tree Physiology. 2008;743-751.

Song S, Gerasimova T, Kurkulos M, Boeke JD, Corces VG. An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. Genes Dev. 1994; 8:2046-2057.

Stace C, Bailey J. The value of genomic in situ hybridization. (GISH) in plant taxonomic and evolutionary studies. In ''Molecular Systematics and Plant Evolution'' (P. M. Hollingsworth, R. M. Bateman, and R. J. Gornall, eds.). 1999; 199–210. London.

Stirling B, Newcombe G, Vrebalov J, Bosdet I, Bradshaw H. Suppressed recombination around the MXC3 locus, a major gene for resistance to poplar leaf rust. Theoretical and Applied Genetics. 2001; 103(8):1129-1137.

Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K. The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. Nature. 2000; 403:913-916.

Takayama S, Shimosato H, Shiba H, Funato M, Che F-S, Watanabe M, Iwano M, Isogai A. Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. Nature. 2001; 413:534 538.

Terol J, Naranjo MA, Ollitrault P, Talon M. Development of genomic resources for Citrus clementina: characterization of three deep coverage BAC libraries and analysis of 46,000 BAC-end sequences. BMC Genomics. 2008; 9:423.

Le Thierry d'Ennequin M, Panaud O, Siljak-Yakovlev S, Sarr A. First evaluation of nuclear DNA content in Setaria genus by flow cytometry. J Hered. 1998; 89:556–559.

Tomkins J, Fregene M, Main D, Kim H, Wing R, Tohme J. Bacterial artificial chromosome (BAC) library resource for positional cloning of pest and disease resistance genes in cassava (*Manihot esculenta* Crantz). Plant Mol Biol. 2004; 56: 555–561

Toribio M, Celestino C, Molinas M. Cork oak. Culture. 2005;(1):445-457.

Troggio M, Malacarne G, Coppola G, Segala C, Cartwright D, Pindo M, Stefanini M, Mank R, Moroldo M. A dense SNP-based genetic linkage map of grapevine (Vitis vinifera L.) anchoring Pinot Noir BAC contigs. Genetics. 2007; 176: 2637-2650.

Tuskan G, Gunter L, Yang Z, Yin T, Sewell M, DiFazio S. Characterization of microsatellites revealed by genomic sequencing of Populus trichocarpa. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere. 2004; 34(1):85-93.

VanGuilder H, Vrana K, Freeman W. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques. 2008;44(5), 619-26.

Venter J, Smith H, Hood L. A new strategy for genome sequencing. Nature. 1996; 381:364-366.

Walling J, Pires J, Jackson S. Preparation of samples for comparative studies of plant chromosomes using in situ hybridization methods. Methods in enzymology. 2005; 395: 443–60.

Walker J. A technique whose time has come. Science.2002; 64-65

Walker J. Structure and function of the receptor-like protein kinases of higher plants. Plant Molecular Biology. 1994; 26: 1599-1609.

Wells D. Tdd-4, a DNA transposon of Dictyostelium that encodes proteins similar to LTR retroelement integrases. Nucleic Acids Res. 1999; 27: 2408-2415.

Willetts N, Skurray R. Structure and function of the F factor and mechanism of conjugation. In Neidhardt, F.C. (ed.), *Escherichia coli and Salmonella typhimurium*: *Cellular and MolecularBiology*. American Society for Microbiology, Washington, D.C. 1987; 2:1110-1133.

Wills J, Craven R. Form, function, and use of retroviral gag proteins. AIDS. 1991; 5: 639-654.

Woo S, Jiang J, Gill B, Paterson A, Wing R. Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*.Nucleic Acids Res. 1994; 22:4922-4931.

Wright D, Voytas, D. Potential retroviruses in plants: Tat1 is related to a group of Arabidopsis thaliana Ty3/gypsy retrotransposons that encode envelope-like proteins. Genetics. 1998; 149: 703–15.

Wu C, Sun S, Nimmakayala P, Santos FA, Meksem K, Springman R, Ding K, Lightfoot DA, Zhang HB. A BAC- and BIBAC-based physical map of the soybean genome. Genome Res. 2004; 14: 319-326.

Yilmaz A, Uslu E, Tekinbabaç M. Cytogenetic studies on Quercus L . (Fagaceae) species belonging to Ilex and Cerris section in TurkeY. Caryologia. 2011; *64*(3): 297–301.

Yu Q, Tong E, Skelton R, Bowers J, Jones M, Murray J, Hou S, Guan P, Acob R, Luo M, Moore P, Alam M, Paterson A, Ming R. A physical map of the papaya genome with integrated genetic map and genome sequence. BMC Genomics. 2009; 10: 371.

Yu Y, Tomkins J, Waugh R, Frisch D, Kudrna D et al. A bacterial artificial chromosome library for barley (Hordeum vulgare L.) and the identification of clones containing putative resistance genes. Theor. Appl. Genet. 2000; 101: 1093–1099.

Zhang HB, Wing RA. Physical mapping of the rice genome with BACs. Plant Mol Biol. 1997; 35: 115-127.

Zhang P, Li WL, Fellers J, Friebe B, Gill BS. BAC-FISH in wheat indentifies chromosome landmarks consisting of different types of transposable elements. Chromosoma. 2004a; 112: 288-299.

Zhao S, Shatsman S, Ayodeji B, Geer K, Tsegaye G, Krol M, Gebregeorgis E, Shvartsbeyn A, Russell D, Overton L, Jiang L, Overton L, Jiang L, Dimitrov G, Tran K, Shetty J, Malek JA, Feldblyum T, Nierman WC, Fraser CM. Mouse BAC ends quality assessment and sequence analyses. Genome Res. 2001; 11:1736-1745.

Ziolkowski PA, Sadowski J. FISH-mapping of rDNAs and Arabidopsis BACs on pachytene complements of selected Brassicas. Genome 2002; 45: 189–197.

Zoldos, V, Papes D, Brown SC, Panaud O, Siljak-Yakovlev S. Genome size and base composition of seven Quercus species: inter- and intra-population variation. Genome. 1998; 41(2): 162–168.

Zoldoš V, Papeš D, Cerbah M, Panaud O, Besendorfer V, Šiljakyakovlev S. Molecularcytogenetic studies of ribosomal genes and heterochromatin reveal conserved genome organization among 11 Quercus species. Theor Appl Genet 1999: 969–977.

Websites:

http://cnrgv. toulouse.inra.fr/
http://lipidlibrary.aocs.org/plantbio/polyesters/index.htm
http://science.jrank.org
http://www.abbottmolecular.com/us/technologies/real-time-pcr.html?icid=am-009-iva-04500900-001
http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/real-time- pcr.html
http://www.blast.ncbi.nlm.nih.gov/Blast.cgi (September 2012)
http://www.corkcomposites.amorim.com
http://www.expasy.org
http://www.gene-quantification.de
http://www.gydb.org/index.php/Main_Page (September 2012)
http://www.kofler.or.at/bioinformatics/SciRoKo/ (September 2012)
http://www.mbcf.dfci.harvard.edu/docs/oligocalc.html (January 2013)
http://www.ncbi.nlm.nih.gov/
http://www.ncbi.nlm.nih.gov/genbank/
http://www.ncbi.nlm.nih.gov/tools/primer-blast/
http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html (September 2012)
http://www.ncgr.org/research/jag
http://www.p2pays.org/ref/34/33161.pdf
http://www.realcork.org/artigo/84.htm.
http://www.realcork.org/artigo/292.htm
http://www.repeatmasker.org/(September 2012)
http://www.sigmaaldrich.com/etc/medialib/docs/Sigma/General_Information/qpcr_technical_guide.Par.0001.File.tmp/qpcr_technical_guide.pdf
Soper, D.S. (2011) "p-Value Calculator for Correlation Coefficients (Online Software)"