



Hinc patriam sustinet

**Instituto Superior de Agronomia**  
**Universidade Técnica de Lisboa**

# **Estudo da organização genómica de sequências repetitivas em várias espécies de *Fagaceae***

**Sofia Graça Barbosa Alves**

Dissertação para obtenção do Grau de Mestre em  
**Biologia Funcional**

Orientador: Professora Leonor Morais-Cecílio

Co-Orientador: Doutora Margarida Rocheta

## **Júri:**

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

Vogais: Doutor José Eduardo Lima Brito, Professor Associado da Universidade de Trás-os-Montes e Alto Douro;

Doutora Maria Leonor Mota Morais Cecílio, Professora Auxiliar do Instituto Superior de Agronomia da Universidade Técnica de Lisboa;

Doutora Margarida Maria Pedro Rocheta.

Lisboa, 2010

# Agradecimentos

Em primeiro lugar quero agradecer às minhas orientadoras, Prof. Leonor Morais-Cecílio e Dr.<sup>a</sup> Margarida Rocheta. Obrigada Prof. Leonor por me ter aceitado mais uma vez. Obrigada às duas, por tudo o que me ensinaram, por se terem mostrado sempre disponíveis e compreensivas e me terem apoiado nos momentos em que duvidava das minhas capacidades, espero ter estado à vossa altura e não vos ter desiludido.

Um agradecimento especial à Dr.<sup>a</sup> Teresa Ribeiro por me ter deixado continuar o seu trabalho, pelos conselhos e pela ajuda nas escolhas, em que a sua experiência com o material por vezes me fez muita falta.

Agradeço a Prof. Wanda Viegas me ter deixado fazer o meu trabalho de mestrado no Laboratório de Genética.

Agradeço a todos os meus colegas de laboratório a ajuda que me prestaram. Agradeço especialmente ao Miguel Bento, pelos conselhos, ajuda, pelas brincadeiras que me faziam descontraír e pela companhia das manhãs. Agradeço à Diana Tomás pela amizade e apoio, dentro e fora do laboratório, Sofia Róis parece que a conheço há tanto tempo quanto aos outros, trazendo sempre uma “boa onda” mesmo com sono. Obrigada Augusta Barão, por ser uma amiga de todos nós e estar sempre de olho, espero ter dado a ajuda que precisava nas arrumações. Agradeço à Dr.<sup>a</sup> Manuela Silva pelo apoio. Agradeço à Vera Inácio que apesar de ter chegado no fim, me apoiou e ajudou muito, obrigada! Agradecimento aos restantes pela companhia e sorrisos, Edna Ribeiro, Margarida Carvalho, Ana Delaunay Caperta, Nuno Neves, Sofia Pereira, Patrícia Vidigal (Fisiologia), João Rato e Heike.

Quero agradecer também a todo o pessoal do laboratório de Microbiologia, em especial à Prof. Luísa Brito e à Eng. Ana Carla Silva.

Por fim mas não menos importantes, agradeço aos meus Pais por acreditarem sempre em mim e me ajudarem em todo o meu percurso, sem eles não seria possível. Obrigada Gonçalo, Aninhas e Hugo pelo vosso apoio nos momentos mais críticos.

# Resumo

A família *Fagaceae* é uma importante família de espécies florestais à qual pertencem os géneros *Fagus*, *Quercus* e *Castanea*. Algumas espécies desta família com importância nos ecossistemas portugueses incluem *Quercus ilex* subsp. *rotundifolia* (azinheira) e *Quercus suber* (sobreiro) nos montados e *Castanea sativa* (castanheiro) e *Fagus sylvatica* (faia) nas florestas. Os genomas eucariotas são particularmente ricos em sequências repetitivas, de que são exemplo os retroelementos, que assumem um papel central na sua estrutura e evolução. Neste trabalho foram caracterizados onze fragmentos AFLP do genoma de azinheira. A maioria dos fragmentos revelou serem homólogos de retroelementos, enquanto os restantes não possuem homologia conhecida. Para comparar a organização genómica de algumas destas sequências nas *Fagaceae* foi realizada a amplificação das regiões inter-sequências AFLP (ISA) em azinheira, sobreiro, castanheiro e faia. Estas amplificações revelam polimorfismos entre os diferentes genomas que são tanto maiores quanto mais afastadas se encontram as espécies. Quando estas regiões genómicas são mapeadas em núcleos e cromossomas das quatro espécies estudadas, surge, para além de padrões dispersos pelo núcleo, uma notável marcação em todas as regiões do organizador nucleolar (NOR), comprovando a contribuição dos retroelementos e outras sequências repetitivas para um padrão de organização genómico pouco frequente nos genomas vegetais.

Palavras-chave: *Quercus*; *Castanea*; *Fagus*; Retrotransposões; NORs hibridação *in situ* fluorescente

## Abstract

*Fagaceae* is an important forest trees family to which belong the genus *Fagus*, *Quercus* and *Castanea*. In this family, the species with great importance in the Portuguese ecosystems include *Quercus ilex* subsp. *rotundifolia* (Holm oak) and *Quercus suber* (cork oak) in *montado*, and *Castanea sativa* (chestnut) and *Fagus sylvatica* (beech) in the forests. Eukaryote genomes are particularly rich in repetitive sequences, such as retroelements that assume a major role in its structure and evolution. In this work it was characterized eleven AFLP fragments of *Quercus ilex* subsp. *rotundifolia* genome. The sequences studied were mostly homologous to retroelements while others don't show any homology. To compare the genomic organization of some of these sequences in *Fagaceae*, the amplification of inter-sequence AFLP regions, was held in Holm and cork oaks, chestnut and beech DNAs. Different patterns were detected between these genomes, reflecting the evolutionary differences between species. When the genomic regions were mapped in chromosomes and nuclei on the four species, besides their occurrence throughout the nucleus, there was also a notable presence in all nucleolar organizing regions (NOR) enlightening the great contribution of retroelements and other repetitive sequences for a genomic organization not commonly seen in the plant genomes.

Keywords: *Quercus*; *Castanea*; *Fagus*; Retrotransposons; NORs; fluorescence *in situ* hybridization

# Dissertation Summary

Eukaryote genomes have two distinct chromatin fractions: an euchromatic, mostly formed by unique copy sequences associated with gene activity, and a heterochromatic tightly condensed associated with gene silencing. Heterochromatin is composed of repeated sequences, mostly of transposable elements, which, in plant genomes can reach a higher proportion. Retrotransposons with long terminal repeats (LTR-retrotransposons) are the predominant retroelements in plant genomes. These elements comprise two main superfamilies *gypsy-like* and *copia-like* that differ in the order of reverse transcriptase, integrase and RNase H in their ORFs.

The transposition mechanism of these elements involving *copy and paste* steps seems to be responsible for genome size differences detected in evolutionary close eukaryotic organisms. Retroelements are preferentially inserted in certain regions of the eukaryotic genomes leading to the formation of highly repeated regions and thus, to the formation of heterochromatin.

45S ribosomal genes consist of tandem arrays of repeated units clustered at particular chromosomal *loci* the nucleolar organizing regions – NORs. NORs chromatin is organized in decondensed domains, where the active genes are located and heterochromatic domains correspond to the silenced ribosomal genes. Due to their repetitive and highly conserved nature in eukaryotes, these regions are easily detected in chromosomes and interphase nuclei, being commonly used in fluorescent *in situ* hybridization (FISH) for karyotype and evolutionary studies.

The *Fagaceae* family has special importance in forest European ecosystems. In Portugal this family is represented by genus *Fagus*, *Castanea* and *Quercus*, throughout the territory, being of particular relevance the species *Q. suber* (cork oak) and *Q. ilex* subsp. *rotundifolia* (Holm oak) which integrate the Mediterranean ecosystem – montado. The heterochromatic genome fraction of *Fagaceae* is poorly known being its study limited to a reduced number of cytogenetic and molecular reports.

The objective of this work is to study the genomic organization of the *Fagaceae* species with value for Portuguese forest and economy, through the isolation and characterization of repetitive sequences in *Quercus ilex* subsp. *rotundifolia* genome. For this purpose, several genomic fragments previously obtained through AFLP with primer combinations *EcoRI-AA/MseI-CAA*, *MseI-CAG/EcoRI-AC* and *MseI-CA/EcoRI-AG*, were isolated and cloned. The length of the obtained fragments range from 83 bp up to 525 bp and they all revealed to be AT rich. The homology searched unraveled two groups: one, enclosing the majority of sequences, with high homology with transposable elements and a smaller one, with no homology to known sequences. These results show that retroelements

are probably an important fraction of the Holm oak genome. The sequences homologous to retroelements include RNase H and reverse transcriptase (RT) proteins. To determine the relationship between these sequences and plant retrotransposons, we carried out phylogenetic analysis based on sequences homologous segments. This analysis revealed different phylogenetic associations, to each sequence, leading to the thought that they belong to distinct elements with different evolutionary history. The potential polymorphisms that resulted from distinct insertion sites in other *Fagaceae* (*Q. suber*, *C. sativa* and *F. sylvatica*) genomes were studied through PCR amplification using specific primers. Amplicons were obtained in all species producing distinct patterns which might indicate a variability of insertion sites in all species analyzed. The same methodology was used to characterize the new sequences without known homology. The results showed different amplicon patterns on the four species.

With the aim of mapping these AFLP inter-sequence genomic regions in the four *Fagaceae* genomes, the products of amplification of three fragments (one encoding an RT and two fragments without homology), obtained from Holm oak genome, were used as FISH probe. *In situ* hybridization revealed a dispersed pattern throughout the nucleus, as previously seen in other plant species retroelements. Moreover a clear accumulation of these sequences in the nucleolar organizer region was detected. Together these two distributions lead to different patterns in the four genomes, which were similar between the two *Quercus* species and more distinct in chestnut and beech. The presence of retroelements in NORs has already been described in different *Musa* species, which could be considered a consequence of a particular genome organization. Furthermore sequences belonging to the IGS region are known to spread to other chromosomes regions of several species, like in *Nicotiana*. Among the four species of *Fagaceae* studied, beech has an interphase chromatin organization distinct from the others, with clearly visible chromocenters. One of the FISH probes used, hybridized intensely with the chromocenter regions, as well as, with NORs. In order to obtain a specific probe for the heterochromatic pericentromeric regions of *Fagus sylvatica*, a probe was produced with beech DNA and the same primers. FISH in beech nuclei and this probe, revealed hybridization signal only in the NOR region, pointing to the failing of amplification of those sequence(s) which were presumably present in Holm oak FISH probe. In conclusion, the sequences isolated from *Q. ilex subsp. rotundifolia* revealed themselves as potential components of the heterochromatic fraction of *Fagaceae* genomes. Moreover, the sequences without database homology are strong candidates to be specific to this family, and can be therefore used to better understand its evolution.

## Lista de Figuras

<b>Figura 1.</b> Estrutura dos retrotransposões LTR “ <i>Gypsy</i> - e <i>Copia-like</i> ”.....	3
<b>Figura 2.</b> Filogramas construídos pelo método da máxima verosimilhança.....	24
<b>Figura 3.</b> Electroforese de agarose a 1 % dos fragmentos amplificados por PCR a partir de DNA genómico de quatro espécies de <i>Fagaceas</i> .....	27
<b>Figura 4.</b> Detecção simultânea com FISH das sondas <i>Rot8.24</i> e rDNA 45S.....	30
<b>Figura 5.</b> Detecção simultânea com FISH das sondas <i>Rot23</i> e rDNA 45S.....	32
<b>Figura 6.</b> Detecção simultânea com FISH das sondas <i>Rot20</i> e rDNA 45S.....	33
<b>Figura 7.</b> Detecção simultânea com FISH das sondas <i>Sy/20</i> e rDNA 45S.....	34

## Lista de Tabelas

<b>Tabela 1.</b> Fragmentos de DNA obtidos ( <i>Rot</i> ) pela técnica de AFLP e respectivos <i>primers</i> utilizados.....	8
<b>Tabela 2.</b> Condições utilizadas na reacção de PCR para confirmação da inserção dos fragmentos no vector.....	9
<b>Tabela 3.</b> Fragmentos de DNA e respectivos <i>primers</i> utilizados na análise comparativa em <i>Fagaceas</i> .....	11
<b>Tabela 4.</b> Mistura de hibridação.....	13
<b>Tabela 5.</b> Fragmentos AFLP obtidos com diversas combinações de enzimas/ <i>primers</i> .....	17
<b>Tabela 6.</b> Sequências proteicas obtidas pela tradução dos clones, respectiva <i>open reading frame</i> e sequências homólogas em diferentes espécies.....	22
<b>Tabela 7.</b> <i>Primer(s)</i> desenhados para cada uma das sequências e a sua orientação no genoma.....	26

# Abreviaturas

µg - micrograma

µl – microlitro

AFLP – *Amplified Fragment Length Polymorphism*

Anti-Dig-FITC - anti-digoxigenina conjugada com o fluorocromo isotiocianato de fluoresceína.

BLAST – Basic Local Alignment Search Tool

BSA – albumina de soro de bovino

DAPI - 4',6-diamidino-2-fenilindol

dATP – desoxiadenosina trifosfato

dCTP – desoxicitosina trifosfato

dGTP – desoxiguanina trifosfato

DNA – ácido desoxirribonucleico

dNTP – mistura de nucleótidos

dTTP – desoxitimina trifosfato

dUTP – desoxiuridina trifosfato

*EcoRI* – enzima de restrição isolada de estirpes de *Escherichia coli*.

FISH – fluorescent in situ hybridization

*fow* – *forward*

*g* – aceleração gravitacional

g – grama

h – hora(s)

IPTG - isopropil β-D-1-tiogalactopiranósido

kb – kilo-pares de bases

l – litro

LB – meio Luria Broth

M – molar

mg - miligrama

min - minuto(s)

ml - mililitro

mM – milimolar

Mpb – Mega pares de bases (1 milhão de pares de bases)

*MseI* – enzima de restrição isolada de *Micrococcus spp.*

NOR(s) – Região(ões) organizadora(s) do nucléolo (*Nucleolar organizer region(s)*)

ORF – open reading frame

pb - pares de bases

PCR – polymerase chain reaction

p/v – peso/volume

*rev* – *reverse*

RNA – ácido ribonucleico

RNase – ribonuclease, degrada RNA

s - segundo(s)

SDS – dodecil sulfato de sódio

*spp.* - espécies

SSC – citrato de sódio salino

subsp. – sub-espécie

Tris = Trizma – tris[hidroximetil] amino-metano

v/v – volume/volume

X-Gal - bromo-cloro-indolil-galactopiranósido

# Índice

Agradecimentos .....	i
Resumo .....	ii
Abstract .....	iii
Dissertation Summary.....	iv
Lista de Figuras .....	vi
Lista de Tabelas.....	vi
1. Introdução .....	1
1.1. Organização de genomas eucariotas. ....	1
1.2. Elementos Transponíveis.....	1
1.3. Genes ribossomais.....	4
1.4. Fagaceae.....	5
1.5. Objectivos.....	6
2. Materiais e Métodos .....	7
2.1. Material Vegetal.....	7
2.1.1. Tratamento c-mitótico e fixação de raízes.....	7
2.2. Análise de DNA.....	7
2.2.1. Extração de DNA genómico.....	7
2.2.2. Clonagem e sequenciação dos fragmentos de AFLP.....	8
2.2.3. Análise Bioinformática das sequências.....	10
2.2.4. Análise comparativa por PCR.....	10
2.3. Hibridação in situ Fluorescente.....	11
2.3.1. Realização de esfregaços .....	11
2.3.2. Pré-tratamento das lâminas .....	12
2.3.3. Preparação das sondas.....	13
2.3.4. Preparação da mistura de hibridação.....	13
2.3.5. Desnaturação e hibridação .....	13
2.3.6. Lavagens pós-hibridação .....	14
2.3.7. Detecção dos locais de hibridação .....	14
2.3.8. Coloração do DNA com DAPI.....	14
2.3.9. Visualização e microfotografia.....	15
3. Resultados .....	16
3.1. Caracterização dos fragmentos de AFLP .....	16
3.2. Análise filogenética de sequências.....	21

3.3. Caracterização das regiões genómicas inter-sequências AFLP em diferentes espécies de Fagaceae.....	25
3.3.1. Caracterização molecular.....	25
3.3.2. Caracterização citogenética - Hibridação <i>in situ</i> fluorescente.....	28
4. Discussão .....	35
5. Conclusões e Perspectivas Futuras .....	38
6. Referências.....	39

# 1. Introdução

## 1.1. Organização de genomas eucariotas.

O genoma das plantas é constituído por várias moléculas de DNA lineares situadas no núcleo e por DNA extranuclear circular nas mitocôndrias e plastídeos.

O núcleo dos eucariotas é envolvido pelo envelope nuclear que delimita uma matriz onde se encontra a cromatina constituída por DNA e proteínas que se organizam em cromossomas. A cromatina pode existir em dois tipos de organização diferentes: a eucromatina e a heterocromatina. A eucromatina é a porção mais descondensada correspondendo provavelmente a fibras de 30 nm e com grande potencial de expressão génica. Esta fracção é maioritariamente constituída por sequências em cópia única enquanto a heterocromatina é constituída por sequências repetidas que adoptam uma estrutura muito condensada estando associada ao silenciamento génico (BELMONT *et al.* 1999; WOODCOCK and DIMITROV 2001). A heterocromatina pode ser considerada constitutiva quando se apresenta condensada durante todo o ciclo celular, ou pode ser facultativa quando se apresenta associada apenas a um determinado estado do ciclo celular (AVRAMOVA 2002). A distribuição da componente heterocromática não é uniforme nos cromossomas de uma mesma espécie variando entre e ao longo destes. Numa análise citológica as regiões heterocromáticas podem ser claramente distinguidas sob a forma de blocos cuja distribuição varia de espécie para espécie: em *Arabidopsis thaliana* localizam-se nas regiões centroméricas e pericentroméricas (FRANSZ *et al.* 2000), em *Nicotiana tabacum* nas zonas teloméricas e sub-teloméricas (FAJKUS *et al.* 1995). Outra fracção heterocromática presente em todos os genomas eucariotas é a que diz respeito às regiões do organizador nucleolar (NORs) (NEVES *et al.* 2005). A heterocromatina, excepto a que se encontra no NOR é constituída por sequências não codificantes muito repetidas, onde se incluem numerosos elementos transponíveis (FRANSZ *et al.* 2000), que mostram uma preferência de inserção nestes domínios heterocromáticos (GAO *et al.* 2008).

## 1.2. Elementos Transponíveis

Os elementos transponíveis foram descobertos inicialmente em plantas por Barbara McClintock, em 1949, tendo sido reconhecidos como agentes mutagénicos que actuavam em genes individuais ou na estrutura geral do genoma (BENNETZEN 2000). Existem vários tipos de elementos transponíveis, no entanto, todos possuem duas propriedades, nomeadamente, moverem-se por diferentes locais do genoma e terem a capacidade de amplificar o seu

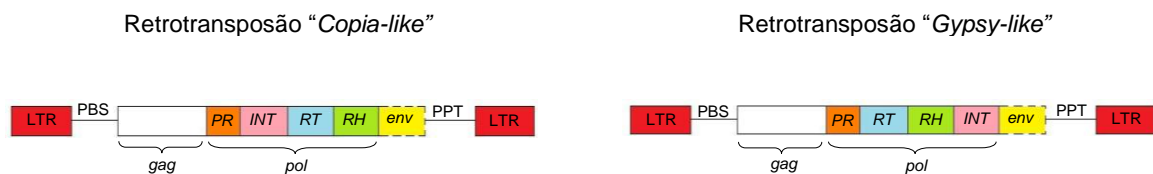
número de cópias dentro do genoma. As diferentes formas de executar estas duas actividades, permitem-nos distinguir as duas grandes classes e tipos de elementos transponíveis.

A primeira grande classe inclui os elementos transponíveis de DNA, que foram os primeiros a serem identificados devido a causarem alterações fenotípicas tanto nos tecidos germinais como nos somáticos. Estes elementos movem-se no genoma através de um mecanismo de excisão e consequente reintegração noutra local pela acção da enzima transposase, por eles codificada (CRAIG *et al.* 2002).

A segunda grande classe de elementos transponíveis compreende os retroelementos, ou elementos transponíveis de RNA. Esta classe é muito abundante nos eucariotas e em plantas constituem o grupo mais representado, chegando a contribuir com mais de 50% do genoma. Esta classe tem como característica principal a transposição por transcrição reversa. Assim, o DNA do retroelemento é directamente transcrito em RNA, que depois é transcrito reversamente para DNA, sendo esta nova cópia inserida no genoma. Os retroelementos podem ser classificados em quatro tipos, agrupados em elementos sem repetições terminais directas longas (LTRs - *Long Terminal Repeats*) e com repetições terminais directas longas. De entre os elementos sem LTR, a classe considerada mais antiga inclui os LINEs (*Long Interspersed Nuclear Elements*). Os LINEs codificam as proteínas *gag* (envolvidas no empacotamento intracelular do transcrito de RNA) e as proteínas que constituem o domínio *pol* (com função polimerásica) que inclui a enzima transcriptase reversa. As proteínas *pol* têm a capacidade de transcrever reversamente o RNA LINE em DNA, enquanto uma endonuclease (*EN*), também codificada pelo elemento, está associada provavelmente com a integração do fragmento de DNA no genoma. As inserções dos LINEs são flanqueadas por pequenas duplicações directas do DNA alvo (BENNETZEN 2000). No grupo de retroelementos sem LTR encontram-se também os SINEs (*Small Interspersed Nuclear Elements*) que não se encontram muito representados nas plantas mas podem ser uma grande fracção em alguns genomas animais (BATZER and DEININGER 2002).

A classe mais numerosa de retroelementos nas plantas inclui os retrotransposões com LTRs. O tamanho dos retrotransposões LTR pode variar entre várias centenas de bases a 10 kb, com LTRs que variam entre centenas a milhares de bases de comprimento. Os retrotransposões LTR possuem um local PBS (*Primer Binding Site*) a seguir ao LTR mais próximo da extremidade 5', assim como um local PPT (*PolyPurine Tract*) antes do LTR da posição 3'. Estas regiões são sítios de ligação do tRNA para a transcrição e transcrição-reversa, respectivamente (HAVECKER *et al.* 2004). A região central é constituída pelas ORFs (*Open Reading Frames*) das poliproteínas *gag* e *pol*. Tal como nos LINEs, a região *gag* codifica, para além das proteínas envolvidas na maturação do RNA do retrotransposão,

proteínas responsáveis pela montagem de partículas semelhantes a vírus (VLPs). A região *pol* codifica uma protease (PR), uma transcriptase reversa (RT), uma RNase H (RH) e uma integrase (INT) que permite a inserção do cDNA no cromossoma. De todos os retrotransposições LTR, os mais importantes pela sua grande contribuição genômica são, os tipo *Gypsy* (*Ty3/Gypsy-like*) e os tipo *Copia* (*Ty1/Copia-like*). A diferença entre estas duas famílias está na posição que a integrase ocupa na ORF do gene *pol* (Figura 1). Nos *Ty1/Copia-like* a integrase encontra-se antes da RT e RH enquanto nos *Ty3/Gypsy-like* localiza-se depois da RH (BENNETZEN 2000). Em alguns casos estes retrotransposições possuem uma ORF entre o gene *pol* e o LTR da extremidade 3' que codifica uma proteína *env* semelhante à que forma o envelope dos vírus, como se observa no retrotransposição “*copia-like*” *SIRE-1* de *Glycine max* (LATEN et al. 1998).



**Figura 1.** Estrutura dos retrotransposições LTR “*Gypsy-* e *Copia-like*”. LTR - extremidades terminais longas (*long terminal repeats*); PBS e PPT - sítios de ligação de primers (*primer binding site* e *polypurine tract*); regiões codificantes *gag-pol*; PR - protease; RT - transcriptase reversa; RH - RNase H; INT - integrase; *env* - alguns retrotransposições possuem sequências semelhantes às que codificam proteínas do *envelope* dos vírus.

É globalmente aceite que os retroelementos contribuem determinantemente para o tamanho e organização dos genomas eucariotas, nomeadamente dos genomas vegetais, sendo apontado o seu mecanismo de transposição como o responsável pela diferença de tamanho do genoma em espécies próximas. Por exemplo, o arroz (490 Mpb/1C) e a cevada (5439Mpb/1C) têm aproximadamente o mesmo número de genes, no entanto apresentam genomas de dimensões muito contrastantes sendo esta diferença atribuída, em parte, à amplificação destes elementos (VICIENT *et al.* 1999).

O local ou locais onde os elementos transponíveis se inserem num genoma pode originar tipos diferentes de organização genómica, uma vez que os retroelementos podem estar dispersos por todo o genoma ou associar-se a certas regiões cromossómicas. Por exemplo, em *Pisum sativum* L. sequências repetitivas correspondentes a partes de retrotransposições LTR foram identificadas ao longo dos cromossomas metafásicos, sem estarem associados a nenhuma região em particular (NEUMANN *et al.* 2001). Já nas espécies de *Helianthus* e nos seus híbridos, a análise de retrotransposições *Ty3/Gypsy-like* revelou estarem mais representados nas regiões centroméricas e pericentroméricas (Staton *et al.*

2009), tal como em *Arabidopsis* e milho (ANANIEV *et al.* 1998; PETERSON-BURCH *et al.* 2004). No caso do retrotransposição *Retand* de *Silene latifolia* verifica-se a sua presença predominantemente nas regiões sub-teloméricas (KEJNOVSKY *et al.* 2006), tal como acontece também com alguns retroelementos em cereais (ZHANG *et al.* 2004). Para além destas localizações foi recentemente descrita a presença do retrotransposição *monkey* na região do rDNA 45S de várias espécies do género *Musa* (BALINT-KURTI *et al.* 2000; HRIBOVA *et al.* 2007).

### 1.3. Genes ribossomais

Os genes ribossomais codificam para o RNA ribossomal e compreendem as unidades de transcrição 45S e a unidade 5S em *loci* distintos. Os genes ribossomais 45S encontram-se repetidos em tandem numa zona cromossómica particular, designada por constrição secundária ou NOR (*Nucleolar Organiser Region*). Cada unidade de repetição 45S compreende as sequências que codificam respectivamente para os genes rRNA 18S, 5.8S e 25S, que são transcritos por inteiro sendo posteriormente processados nos respectivos rRNAs e que juntamente com o rRNA 5S e proteínas irão formar as diferentes subunidades ribossomais. Os genes componentes de cada unidade 45S encontram-se separados por duas regiões intergénicas internas que são transcritas (ITS1 e ITS2 – *Internal Transcribed Spacer*), e cada uma destas unidades encontra-se separada da seguinte por uma região intergénica (IGS – *InterGenic Spacer*) que compreende duas regiões; uma região que não é transcrita - NTS (*Non-Transcribed Spacer*) e uma região transcrita (ETS – *External Transcribed Spacer*) que se localiza na região mais próxima do gene 18S.

Na região IGS, a zona NTS possui locais iniciadores e terminadores da transcrição e elementos repetitivos (sub-repetições), responsáveis pela regulação da transcrição dos genes rRNA a jusante (BAUER *et al.* 2009; LIM *et al.* 2004). Em *Quercus robur* e *Q. petraea* existem quatro sub-regiões dentro do IGS, uma região de sub-repetições contendo três elementos repetidos, uma região rica em AT contendo locais de ligação à matriz e uma origem de replicação putativa, uma região promotora contendo um local de iniciação de transcrição putativo, e por último a região ETS (BAUER *et al.* 2009). A região IGS apresenta grande variabilidade entre as espécies de vegetais, mas grande homogeneidade dentro do mesmo genoma, podendo ser usada como marcador na distinção de espécies e em estudos de evolução. Em *Q. robur* e *Q. petraea*, duas espécies estreitamente relacionadas, foram encontrados duas variantes IGS de dimensões diferentes com uma elevada homologia, indicando a separação recente destes dois *taxa* (BAUER *et al.* 2009). Esta região, ou parte dela, pode também estar presente noutros locais cromossómicos para além do NOR, contribuindo assim para a fracção repetitiva do genoma e conseqüentemente para a

formação da heterocromatina. No género *Nicotiana* uma repetição A1/A2 de cerca de 130 pb pertencente às sub-repetições da região IGS encontrando-se espalhada ao longo de todos os cromossomas em diferentes genomas deste género sob a forma de DNA satélite altamente repetitivo (LIM *et al.* 2004).

Devido ao excessivo número de genes ribossomais, nem todos se encontram activos, havendo uma zona de transcrição activa e uma zona onde os genes estão silenciados. Em consequência, a cromatina dos NORs pode organizar-se em dois tipos de domínios: um eucromático, com grande descondensação correspondente à constrição secundária, visível nalguns cromossomas mitóticos onde se encontram os genes ribossomais activos, e outro heterocromático correspondente aos genes ribossomais silenciados, ou seja, que não se encontram a transcrever e que se encontra adjacente ao(s) domínio(s) eucromático(s). O tipo de organização destes genes e a sua enorme homologia nos eucariotas fazem deles bons marcadores tanto moleculares como citogenéticos e torna-os uma ferramenta indispensável para a caracterização genómica (GILL *et al.* 2008; NEVES *et al.* 2005).

O número de NORs e a sua localização variam muito de espécie para espécie. Em Fagaceas, o número de NORs e a sua localização é conhecida nalgumas espécies de *Quercus* (ZOLDOS *et al.* 1999), e de outras espécies asiáticas. O seu número varia entre um locus em *Castanopsis argentea* até três loci em *Lythocarpus vestitus* (CHOKCHAICHAMNANKIT *et al.* 2008) e a sua localização pode variar entre terminal a intersticial.

#### **1.4. Fagaceae**

Todas as espécies de Fagaceas à excepção de duas (*Trignobalanum doichangensis*  $2n=14$  e *T. verticillata*  $2n=42$ ) possuem  $2n=24$  cromossomas (CHEN *et al.* 2007; KREMER *et al.* 2007). As espécies do género *Quercus* possuem um cariótipo muito semelhante, com cromossomas muito semelhantes e consequentemente, difíceis de distinguir, sendo por isso vantajoso, encontrar marcadores eficazes para diferenciar as várias espécies, tanto a nível molecular como citogenético.

Apesar das suas semelhanças os genomas destas espécies têm tamanhos diferentes sendo a azinheira a espécie com maior genoma publicado (ZOLDOS *et al.* 1998). A constituição e organização dos genomas das *Fagaceae* é pouco conhecida especialmente a contribuição de retrotransposões e outros elementos transponíveis nos genomas do género *Quercus* e outros géneros desta família. São do nosso conhecimento apenas dois estudos que referem a organização de sequências de retrotransposões em *Quercus* (RIBEIRO *et al.* 2008; ZOLDOS *et al.* 2001), e um outro que caracteriza alguns retroelementos com e sem extremidades LTR em *Fagus sylvatica* (EMILIANI *et al.* 2009).

## **1.5. Objectivos**

Este trabalho tem como objectivo estudar a organização genómica de espécies de Fagaceas importantes para as florestas e a economia portuguesa, através da análise de sequências repetitivas presentes no genoma de *Quercus ilex* subsp. *rotundifolia*. Este estudo baseou-se no isolamento e caracterização molecular e citogenética de fragmentos gerados por AFLP a partir do genoma de azinheira e no estudo comparativo da sua organização em *Quercus suber*, *Castanea sativa* e *Fagus sylvatica*.

## **2. Materiais e Métodos**

### **2.1. Material Vegetal**

Neste trabalho foram usadas sementes e folhas de *Quercus ilex* subsp. *rotundifolia*, *Quercus suber*, *Castanea sativa* e *Fagus sylvatica* para obtenção de raízes e DNA. As raízes de *Fagus sylvatica* foram colhidas de plantas envasadas com cerca de 3 anos.

As sementes foram germinadas a temperatura ambiente em areia húmida, e foram retiradas as raízes seminais com 5 cm de comprimento, tendo sido as sementes mantidas até ao desenvolvimento de novas raízes adventícias e posteriormente envasadas.

As folhas para posterior extracção de DNA foram colhidas e colocadas em azoto líquido, podendo, se não forem usadas de imediato, ser guardadas a -80°C.

#### **2.1.1. Tratamento c-mitótico e fixação de raízes**

Após recolha, as raízes foram lavadas e colocadas em tubos de microcentrífuga com 1000 µl de α-bromonaftaleno e mantidas durante 3 horas à temperatura ambiente no escuro. O α-bromonaftaleno funciona como agente c-mitótico, ou seja, induz a despolimerização dos microtúbulos e conseqüentemente a ausência do fuso acromático durante a mitose, conseguindo-se assim, um espalhamento dos cromossomas por todo o citoplasma.

As raízes foram fixadas em placas de Petri com etanol:ácido acético na proporção 3:1 com agitação durante 10 min. Seguidamente transferiram-se as raízes para tubos de ensaio pequenos com o mesmo fixador onde se mantiveram durante 3 minutos findos os quais o fixador foi substituído, repetindo-se este passo três vezes. No dia seguinte repetiu-se o último passo mais uma vez e colocaram-se os tubos de microcentrífuga a -20°C durante uma semana.

### **2.2. Análise de DNA**

#### **2.2.1. Extracção de DNA genómico**

A extracção de DNA genómico de *Q. ilex* subsp. *rotundifolia* e *C. sativa* foi feita manualmente fazendo uma moagem das folhas em azoto líquido em almofariz e seguindo o protocolo de isolamento de DNA a partir de pequenas quantidades de tecido de plantas, desenvolvido por Doyle & Doyle (DOYLE and DICKSON 1987). Para *Q. suber* e *F. sylvatica* após a moagem do material foi usado o kit de extracção da Qiagen, DNeasy® Plant Mini Kit.

## 2.2.2. Clonagem e sequenciação dos fragmentos de AFLP

Os fragmentos de DNA usados no início deste trabalho foram cedidos pela Dra. Teresa Ribeiro.

A técnica de AFLP (*Amplified Fragments Length Polymorphism*) foi utilizada com o objectivo de obter fragmentos de sequências muito repetitivas do genoma de *Q. ilex* subsp. *rotundifolia* que pudessem ser usadas como sonda de FISH em diferentes genomas de Fagaceas (REAMON-BUTTNER *et al.* 1999). A técnica de AFLP (VOS *et al.* 1995) baseia-se na digestão completa de DNA genómico por endonucleases de restrição, seguida da amplificação selectiva por PCR e electroforese dum conjunto de fragmentos (Tabela 1).

**Tabela 1.** Fragmentos de DNA obtidos (*Rot*) pela técnica de AFLP e respectivos *primers* utilizados.

Fragmentos de DNA ( <i>Rot</i> )	Combinação usada para gerar os fragmentos AFLP	<i>Primers</i> de AFLPs <i>Forward / Reverse</i>
2*; 4; 6; 8.11; 8.24; 9; 10*; 10.1*	EcoRI-AA/MseI-CAA	5'-gactgcgtaccaattcaa-3'/ 5'-gatgagtctctgagtaacaa-3'
20*; 23	MseI-CAG /EcoRI-AC	5'-gatgagtctctgagtaacag-3'/ 5'-gactgcgtaccaattcac-3'
27	MseI-CAT/EcoRI-AG	5'-gatgagtctctgagtaacat-3'/ 5'-gactgcgtaccaattcag-3'

\* - Fragmentos caracterizados mas que não foram clonados neste trabalho

A designação dos fragmentos obtidos *Rot*, foi atribuída de acordo com a sua posição no gel de AFLP (de menor dimensão para o maior) e da numeração atribuída às colónias bacterianas.

Os fragmentos foram retirados do gel de AFLP, colocados em tubos de microcentrifuga e guardados a 4°C. A cada tubo adicionaram-se 30 µl de água destilada sendo posteriormente colocados durante 5 min em água a ferver e centrifugados a 16 000 g durante 10 min. O sobrenadante resultante serviu de molde para uma prévia amplificação do DNA realizada antes da clonagem.

Para a clonagem dos fragmentos usou-se o NZY-A PCR cloning kit da NZYTech, segundo o protocolo fornecido. Os fragmentos clonados foram *Rot4*, *Rot6*, *Rot8.11*, *Rot8.24*, *Rot9*, *Rot23* e *Rot27*.

As células transformadas foram plaqueadas em placas contendo meio LB com 16 g/l de Agar, 1 µl/ml de Ampicilina (100 mg/ml), 1 µl/ml de X-GAL (80 mg/ml) e 2.5 µl/ml IPTG (100 mM). A incubação decorreu a 37°C, durante a noite.

No dia seguinte as placas foram mudadas para 4°C durante 2 h a 3 h de modo a que a cor induzida pelo X-GAL se observasse bem. A partir de colónias brancas e azuis foi feito riscado. Posteriormente foi feito um PCR destas colónias para confirmação da inserção do fragmento seleccionado utilizando os *primers* M13 (Tabela 2).

**Tabela 2.** Condições utilizadas na reacção de PCR para confirmação da inserção dos fragmentos no vector.

Reacção de PCR		Programa PCR	
Colheita das bactérias da colónia		3 min a 96°C	
10x PCR buffer	2.5 µl	30 ciclos :	
MgCl <sub>2</sub> (50 mM)	0.75 µl	45 s a 94°C	
dNTPs (5 mM)	0.5 µl	45 s a 56°C	
<i>Primer</i> M13 <i>for</i> (50 pM)	0.5 µl	1:30 min a 72°C	
<i>Primer</i> M13 <i>rev</i> (50 pM)	0.5 µl	3 min a 72°C	
Água destilada desionizada	19.25 µl	Fim	
Taq polimerase (5 U/ µl)	0.5 µl		

Ao fazer a confirmação da inserção dos fragmentos observámos, contrariamente ao que era de esperar, que algumas colónias apresentando coloração azul possuíam fragmentos clonados, sendo assim consideradas falsos negativos. Esta situação ocorre somente com fragmentos de pequenas dimensões (<300 pb) e deve-se ao facto de que pequenas inserções não são capazes de impedir o funcionamento do gene *lacZ*. Durante o processo de detecção de colónias transformadas foram também encontrados casos de falsos positivos em que colónias brancas não possuíam o fragmento clonado. Neste caso, presume-se que o gene *lacZ* não foi activado, possivelmente devido às extremidades do plasmídeo se encontrarem danificadas.

Após confirmação da inserção do fragmento de DNA no plasmídeo, em gel de agarose 1 %, procedeu-se à sequenciação dos sete fragmentos.

### 2.2.3. Análise Bioinformática das sequências

Após sequenciação, todos os fragmentos de DNA foram analisados, inclusive os fragmentos *Rot20*, *Rot10* que já se encontravam publicados no GenBank e o fragmento *Rot10.1* cuja sequência estava disponível mas não publicada.

O primeiro passo da análise envolveu a detecção e remoção da zona das sequências pertencentes ao vector e à sequência do adaptador. Seguidamente, efectuou-se uma pesquisa de homologia utilizando as ferramentas Blastn (algoritmo de BLAST que procura homologias em bases de dados de nucleótidos usando uma sequência de nucleótidos) e Blastx (algoritmo de BLAST que procura homologias em bases de dados de proteínas usando uma sequência de nucleótidos traduzida) disponíveis na base de dados do NCBI. Foram utilizadas somente sete sequências de aminoácidos correspondentes a sete fragmentos AFLP (*Rot2*, *Rot6*, *Rot9*, *Rot10*, *Rot10.1*, *Rot20* e *Rot27*) para análise de filogenia (Tabela 1; ver Resultados-Tabela 6). Esta análise utilizou a ferramenta que se encontra disponível no site <http://www.phylogeny.fr/>. O algoritmo MUSCLE foi escolhido para efectuar o alinhamento das sete sequências com sequências de espécies escolhidas (ANISIMOVA and GASCUEL 2006; CHEVENET *et al.* 2006; DEREPPER *et al.* 2008; EDGAR 2004; GUINDON and GASCUEL 2003).

### 2.2.4. Análise comparativa por PCR

O conhecimento da sequência dos fragmentos AFLP permitiu o desenho de *primers* para as regiões ISA, específicos para os sete (*Rot4*, *Rot6*, *Rot8.11*, *Rot8.24*, *Rot9*, *Rot23* e *Rot27*) fragmentos clonados, assim como para a sequência *Rot20*, já publicada (Tabela 3). Supondo que a(s) sequência(s) amplificada(s) se encontra presente noutras espécies de Fagaceas, estes *primers*, desenhados para a sequência genómica de *Q. ilex* subsp. *rotundifolia*, foram usados para uma análise comparativa em outras espécies de Fagaceas: *Q. suber*, *Castanea sativa* e *Fagus sylvatica*.

**Tabela 3.** Fragmentos de DNA e respectivos *primers* utilizados na análise comparativa em Fagaceas.

Designação	<i>Primer Forward</i>	<i>Primer Reverse</i>
<i>Rot8.24</i>	-	5' ctccaactccgctccctcc 3'
<i>Rot9</i>	5' attctatgggtcctcagtgtttgg 3'	5' tctacatcctcagagagttcccat 3'
<i>Rot20</i>	5' aggcagccgggaactcatgc 3'	5' tccaaggctggtggtccatgc 3'
<i>Rot8.11</i>	5' gttgtgtggtttgtcacggatt 3'	-
<i>Rot4</i>	5' aacaagtttagcatccaagtcc 3'	-
<i>Rot23</i>	5' acaggaggcagagtaggtaagg 3'	-
<i>Rot6</i>	5' ggatgagatggattcctgttgg 3'	-
<i>Rot27</i>	5' agattctgcctactcttaccac 3'	-

## 2.3. Hibridação *in situ* Fluorescente

### 2.3.1. Realização de esfregaços

Os esfregaços foram feitos pela técnica de "drop" seguindo a metodologia de Zoldos *et al.* (1999) com pequenas modificações. Para tal, os ápices foram sujeitos a uma digestão enzimática com uma mistura de enzimas que digerem a parede celular<sup>1</sup>. Retiraram-se as raízes e lavaram-se duas vezes, durante 10 min em caixas de Petri contendo 1xEB<sup>2</sup>. Prosseguiu-se então com a limpeza das raízes separando o ápice da coifa e da zona meristemática. Seguidamente, as raízes foram digeridas em tubos de microcentrifuga, contendo a solução enzimática, a 37°C, durante aproximadamente três horas, com atenção para que ao fim da segunda hora os ápices fossem desfeitos com ajuda de uma micropipeta, porções maiores com dificuldade de digestão foram removidas.

As raízes digeridas e desfeitas foram centrifugadas na solução de digestão a 990 g, durante 4 min, e retirou-se o sobrenadante.

Conforme a quantidade do sedimento obtido colocaram-se entre 300 a 450 µl de 1xEB de maneira a dispersar as células, procedendo-se a nova centrifugação a 720 g durante 3 min. Removeu-se o sobrenadante e repetiu-se o último passo mais uma vez.

<sup>1</sup> Solução enzimática: 2% (p/v) celulase 'Onozuka' R10 (de *Trichoderma viride*, Serva), 3% (v/v) pectinase (de *Aspergillus niger*, solução em glicerol, Sigma), 0.3% pectoliase Y-23 (de *Aspergillus japonicus*, Sigma), 0.03% EDTA em 2 x SSC, pH 4.2

<sup>2</sup> 1xEB: A – Ácido cítrico monohidratado 0,1 M; B – Citrato trissódico desidratado 0,1 M; 10xEB: adicionar na proporção 2:3 (A:B); diluir 1:10 em água destilada para obter 1xEB.

Depois de removido o sobrenadante adicionaram-se cerca de 300 µl de fixador etanol-acético na proporção de 3:1 a -20°C. Centrifugou-se a 500 g, durante 4 min, e após se ter retirado o sobrenadante, repetiu-se o passo mais uma vez.

Conforme a quantidade de sedimento adicionaram-se entre 70 µl a 200 µl de fixador. Ressuspenderam-se bem as células e fez-se cair uma gota dessa suspensão sobre uma lâmina silanizada<sup>3</sup> soprando em seguida de maneira a espalhar e secar a gota.

Após a verificação da qualidade da preparação obtida em contraste de fase (microscópio Leitz DIALUX-22), repetiu-se o procedimento anterior para obter as restantes lâminas. No entanto, para material com citoplasma persistente, centrifugou-se de novo o tubo, retirando-se todo o sobrenadante e substituiu-se por ácido acético a 60%, deixando repousar até 2 min conforme a densidade do citoplasma. Centrifugou-se novamente durante 3 min e substituiu-se por novo fixador, prosseguindo com o *drop*.

O método de hibridação *in situ* fluorescente usado foi adaptado do procedimento usado por Schwarzacher and Heslop-Harrison (SCHWARZACHER and HESLOP-HARRISON 2000)

### **2.3.2. Pré-tratamento das lâminas**

Os pré-tratamentos foram iniciados com a digestão do citoplasma e fracção proteica com pepsina<sup>4</sup>. Colocou-se 200 µl de pepsina (por lâmina) e cobriu-se a região do esfregaço com lamela de plástico, mantendo as lâminas 10 min em câmara húmida a 37°C, lavando seguidamente em 2xSSC<sup>5</sup> (retirando-se as lamelas dentro da tina), 2x5 min com agitação.

Para se proceder à hibridação DNA/DNA, o material foi previamente tratado com RNase<sup>6</sup>, para degradar o RNA celular, impedindo a sua possível hibridação com a sonda. Para tal, colocou-se 200 µl de RNase (por lâmina) e cobriu-se com lamela de plástico, 1 h em câmara húmida a 37°C, lavando-se em seguida em 2xSSC (retirando-se as lamelas dentro da tina), 3x5 min com agitação.

Seguidamente procedeu-se à desidratação das lâminas, por imersão em etanol a 70 % (v/v) durante 3 min com agitação, e em etanol a 100% por mais 3 min deixando por fim secar ao ar.

---

<sup>3</sup> Silanização das lâminas: Dispõe-se 3 tinas com 200 ml de solução de silano (4 ml de silano + 196 ml de acetona), 200 ml de acetona e água destilada. Colocam-se as lâminas num suporte, passando 10s na solução de silano, depois pela acetona e pela água destilada. Seguidamente colocam-se as lâminas a secar na estufa a 37°C.

<sup>4</sup> Pepsina (actividade 3.200-4.500 unidades/mg de proteína) – Solução 2stock2 de 500 µg/ml em 0,01 M HCl. Usar uma diluição 1:75 em 0,01 M HCl.

<sup>5</sup> Saline Sodium Citrate: 0.3 M NaCl, 0.03 M citrato de sódio em água destilada, ajustar a pH 7,0.

<sup>6</sup> RNase – Solução “stock” 10 mg/ml em 10 mM Tris.HCl, pH 8, preparada a partir de ribonuclease A, solução em 10 mM Tris.HCl, pH 8 e 50% glicerol. Usar 100mg/ml diluindo 1/100 em 2xSSC.

### 2.3.3. Preparação das sondas

As regiões ISA *Rot8.24*, *Rot20* e *Rot23* amplificadas de DNA genómico de *Q. ilex* subsp. *rotundifolia* foram marcados através da técnica de PCR por incorporação dos nucleótidos marcados digoxigenina-11-dUTP ou biotina-11-dUTP tendo assim gerado as sondas *Rot8.24*, *Rot20*, *Rot23*. Os *primers Rot20* foram usados também para produzir uma sonda em genoma de faia que se designou *Syl20*. Simultaneamente foi também usada a sonda pTa71 como controlo por ter a sua localização conhecida nestes genomas. Esta sonda que permite identificar os genes ribossomais 45S, foi marcada com digoxigenina por incorporação de digoxigenina-11-dUTP (Roche) ou biotina-11-dUTP (Roche) pela técnica de “Nick Translation” (Roche), usando o protocolo fornecido com o produto.

### 2.3.4. Preparação da mistura de hibridação

A mistura de hibridação foi preparada segundo a Tabela 4.

**Tabela 4.** Mistura de hibridação.

Soluções	µl/lâmina
Formamida 100%	20
20xSSC	2
SDS 10% (p/v)	0,5
DNA esperma de salmão	1
Sonda	2 pTa71 5 <i>Rotx</i>
Sulfato de Dextrano 50%	8
Água esterilizada	Até 40

A formamida 100% é um destabilizador da molécula de DNA, o SDS 10 % é um detergente utilizado para quebrar as ligações entre os diferentes aminoácidos e aumentar a viscosidade. O sulfato de dextrano aumenta também a viscosidade e a concentração do DNA na amostra.

### 2.3.5. Desnaturação e hibridação

A mistura de hibridação foi primeiro desnaturada a 80°C (bloco térmico) durante 10 min, seguindo-se uma incubação em gelo durante 5 min.

A mistura de hibridação foi aplicada directamente nas lâminas (50 µl/lâmina) e coberta com uma lamela de plástico. Em seguida as lâminas foram colocadas no termo-ciclador tendo sido usado o programa que começa com uma desnaturação 78°C durante

10 min, seguida de uma diminuição gradual da temperatura até aos 37°C. Uma vez concluído o programa de desnaturação, as lâminas foram colocadas a incubar em câmara húmida a 37°C, onde ficaram até ao dia seguinte.

### **2.3.6. Lavagens pós-hibridação**

Estas lavagens permitem a remoção da sonda de locais onde se ligou com menor especificidade, mantendo-se apenas as heterohélices de DNA mais específicas.

As lâminas foram lavadas em 2xSSC, 3 min com agitação a 42°C, retirando-se as lamelas dentro da tina e seguidamente lavadas em formamida 20% em 0,1xSSC, durante 10 min com agitação a 42°C. Seguidamente, procedeu-se a duas lavagens em 2xSSC, durante 5 min com agitação a 42°C e a mais duas lavagens em 2xSSC à temperatura ambiente, durante 5 min com agitação. Por fim, foram lavadas duas vezes em 4xSSC/Tween<sup>7</sup> durante 5 min com agitação, à temperatura ambiente.

### **2.3.7. Detecção dos locais de hibridação**

De início procedeu-se ao bloqueamento de potenciais locais inespecíficos de ligação dos anticorpos tendo sido colocados 200 µl de solução de BSA 5 %<sup>8</sup> em cada lâmina cobrindo-se com lamelas de plástico, durante 5 min. Seguidamente adicionou-se 50 µl de solução de detecção a cada lâmina e cobriu-se com lamela de plástico, durante 1 h a 37°C em câmara húmida. A solução de detecção foi sempre feita com anti-Dig marcado com fluoresceína (FITC) (8 µg/ml) como anti-corpo primário e streptavidin-cy3 (5 µg/ml) em solução de BSA 5 %. Seguidamente, as lâminas foram lavadas três vezes com 4xSSC/Tween durante 5 min com agitação à temperatura ambiente.

### **2.3.8. Coloração do DNA com DAPI**

Logo após serem retiradas e escurridas da última lavagem, as lâminas foram montadas com 8 µl de DAPI (DAPI<sup>9</sup> em Citifluor AF1<sup>10</sup>), utilizando lamelas 24 mm x 32 mm e removendo o líquido em excesso com papel de filtro.

---

<sup>7</sup> 4xSSC 0,2% (v/v) Tween 20 - Preparação de 500 ml: fazer uma diluição 1:5 de 20xSSC em água, adicionar 1 ml de Tween 20 e misturar bem devido à sua viscosidade.

<sup>8</sup> Solução de BSA 5% - Albumina de soro de bovino (BSA, B-8894, fracção V isenta de globulina, Sigma) em 4xSSC/Tween.

<sup>9</sup> DAPI (dihidroclorato de 4', 6-diamino-2-fenilindole), solução de 2 µg/ml.

<sup>10</sup> Citifluor AF1 – Solução glicerol/PBS redutora da dissipação de fluorescência

### **2.3.9. Visualização e microfotografia**

As preparações foram observadas usando um microscópio de epifluorescência Zeiss com objectiva e 100x com um conjunto de filtros apropriados para a detecção do FITC, Cy3 e DAPI. As imagens foram captadas utilizando uma câmara fotográfica digital AxioCam (com uma resolução de 2600dpi x 2060dpi) e com o software Axiovision. As imagens foram depois tratadas com o software Adobe Photoshop versão 7.0.

## 3. Resultados

### 3.1. Caracterização dos fragmentos de AFLP

Com o objectivo de caracterizar sequências repetitivas presentes no genoma de *Quercus ilex* subsp. *rotundifolia* foram isolados e clonados fragmentos genómicos previamente obtidos por AFLP. Neste trabalho foram caracterizados um total de 11 fragmentos. Dos sete fragmentos obtidos por clonagem neste trabalho, e quatro clonados pela Dr.<sup>a</sup> Teresa Ribeiro, oito foram obtidos com o par de *primers* *EcoRI-AA/Msel-CAA*, dois com o par *Msel-CAG/EcoRI-AC* e outro com o par *Msel-CA/EcoRI-AG* (Tabela 1), tal como descrito no ponto 2.2.2. Em todas as sequências correspondentes aos fragmentos caracterizados neste trabalho foram encontrados os *primers* de AFLP utilizados no seu isolamento, provando assim a origem dos mesmos. No entanto verificou-se que os fragmentos *Rot9* e *Rot10.1* possuíam o mesmo primer (*Msel-CAA*) nas duas extremidades indicando que esta sequência se encontra repetida e invertida no fragmento genómico isolado.

A caracterização dos fragmentos analisados encontra-se sumarizada na Tabela 5 e a combinação de primers AFLP usados encontra-se na Tabela 1. Os fragmentos isolados possuem tamanhos que variam entre 83 pb e 525 pb. Todos os fragmentos analisados correspondem a sequências ricas em AT, com percentagens variando entre 52 % e 71 %. Para investigar a origem destes fragmentos procedeu-se a uma pesquisa de homologias com o algoritmo BLASTN. Somente os fragmentos *Rot2*, *Rot6*, *Rot8.11*, *Rot9*, *Rot10*, *Rot20*, *Rot27* mostraram ter homologia com sequências conhecidas de nucleótidos com *e-values*, variando entre  $4e^{-68}$  e  $4e^{-3}$  (Tabela 5).

**Tabela 5.** Fragmentos AFLP obtidos com diversas combinações de enzimas/primers. Para cada clone indica-se o tamanho em pares de bases (pb) de cada fragmento assim como a percentagem de nucleótidos Adenina, Timina / Citosina, Guanina. As homologias obtidas são também assinaladas.

Clone	Sequência de DNA	Tamanho	AT/CG (%)	Homologia com nucleótidos	Homologia com proteínas	Descrição
Rot2	<u>AATTC</u> AACATAATCCCCATATCAGGGCTTCTATA GTAATTGCTGATGGGTTTTGCGGCCAAAAGTAA TTTGGATAGTACACCTTTTCATTATTGAGTTTTAT GAAGACCGTTGTTA	118 pb	64/36	<i>Q. ilex</i> subsp. <i>rotundifolia</i> gi 222353754(e-V 3e <sup>-21</sup> ) <i>P. trichocarpa</i> gi 224092102(e-V1e <sup>-20</sup> )	<i>R. communis</i> gi 255550139(e-V1e <sup>-11</sup> )	Mechano sensitive Channel
Rot4	<u>TAACA</u> AGCTTGTTAGACAAAaacaagtttagcatccaagtc cATATCAAACATTGAAACTCTTTTCAGCTTTTCAAG AGAATTTTCAGAAAGATTCTTATATTTCTCAACCA ATTTGTTGAATT	123 pb	71/29	SH	SH	
Rot6	<u>AATTC</u> AAGCAAATGGGAGTTAGCCATGAAggatgag atggattcctgttggGGAATCAGACATGGGAAGTACTG AATTACCAGTAGAAAAAAGACTTTGCACAATAAG TGAGTATACTGAATAAAAAATAAGCATGATGGTA GCAAACGTTACAAGGTCAGATTAGTTGTTA	175 pb	63/37	<i>V. vinifera</i> gi 123704217( e-V3e <sup>-68</sup> )	<i>V. vinifera</i> gi 225470676(e-V5e <sup>-22</sup> )	RT_2 "copia-like"
Rot8.11	<u>AATTC</u> AACAAGTGAATATGAGTTGAGTGATTTATG ACTTCTCATATGTTTCATTTgtttgtgtggtttgtcacggat tGCCAAAGGGGGAGATTGTTAGGACATATGTGTT TCACTTGTTA	124 pb	65/35	<i>V. vinifera</i> gi 147777233(e-V 1e <sup>-06</sup> )	SH	
Rot8.24	<u>TAACA</u> AGAAATACTAAATATCAAAGTCCAATCTTA TTTCAAAAGATTCCCTAAACTCCTATCATGCGCACA TCCTCACGAAACTCCCAATACTTCTTATTTCTCAG AATCTGAATATGACACTTGCTATTTCTCAGAATTT GAATATAAAAGGTGAAGAAAAAAGAGGGGAAggg agggagcggagttggagGGAGTGAGCTTGAATT	207 pb	63/37	SH	SH	
Rot9	<u>TAACA</u> ATAATGAAAAACCTAACGAAAAGTTTGTG AATCTTCACATAaattctatgggtcctcagttttggATGTCAA GGGTATGGTCACATGAAATCTGAACGCCCTACAT ACTTGAGGTCTAAGGATAAGGCTATGGCTGTAAC CCTTAGTGATAATGAAGTTTCTGATGATGAGTCT GGATGTGACAAGAATGGAAATTTTCATTGCTTTCA CTGCTACCGCTGTAGTCAATGAAAGTGTGTATGT TGAAGAGAACCCTTCTGatggggaactctcaggatgtag aTCTTCAAGAAGCCTATAATAAAATTTGCAAAGTT GCTGTATAGGATGCTATGAATGTTGAACTTGGCT TGAAGAAAATTGCATCTCTTGAGCTTGATAAGAAA AATTTGCTTGTTA	407 pb	62/38	<i>L. japonicus</i> gi 189163206(e-V3e <sup>-3</sup> )	<i>G. max</i> gi 29423276(e-V1e <sup>-9</sup> )	Gag-pol "copia-like"

Rot10	<p><u>AATTC</u>AACGTCCAATACCGCCCGCATACTACCAT  GAAGGGACAAGCAGTCGCTGACTTCATTGCGGA  ATTCACCAATATGGAAGGCCAGGGGGCAGAAGA  GCATCCTCAATGGAGTATCCACGCGGATGGATTG  TCTAACAAAGCAAGCTTGCGGAAAAGGTATAGTAC  TCCACTCACTAGAAGGGGATGAGATTGAGTGCAT  GGTTCATCTTGACTTCCCTACGACCAACAATGAA  GCGGAGTACGTGGCTCTAGTGGTAGGACTGGAT  CTCGCCAAAGCAGTAGGGGCCATATGTGTGGTT  GTGTATTACAACTTTTAGGTGGTCACAAGTCAGG  TGAACGGTGACTACGAGTGCAAAGGTGAAAGGA  TGAAGAAATACTTGGAGCAAGTAAGGAAGCGGG  TGGGTGACCTTCAGGTTCGGGTTTGTTCAAATCCC  AAGGGAAGAGAATAAGCAAGCTGACTGCCTTGC  CAAAGCTGCATCAGCCGAACATATGCTCATCCCC  AATAAGGTACTTTCTTTTGTTA</p>	525 pb	52/48	<p><i>Q. ilex</i> subsp. <i>rotundifolia</i>  gi 222353756(e-V0,0)  <i>P. vulgaris</i>  gi 84782338(e-V4e<sup>-3</sup>)</p>	<p><i>S. latifolia</i>  gi 68685649(e-V1e<sup>-26</sup>)</p>	<p>RNase H  “<i>gypsy-like</i>”</p>
Rot10.1	<p><u>TAACA</u>AGAAAGGTGGCAGAGAGTTTGTGTGAAAC  AATAGGGGATATTCAAAAATCTAATGGAGCAGTG  GACGAGGATGGGGGTAGTTTTTCCATGTAAGG  GTGGTTATCGATATGACTCTCCTGTTGTGCAGAG  GAAGGCTCATCACTCTACCTAATGGAGGCAAAG  ATGGATTGGTTCAAGTATGAACAGCTCCCAAGT  TTATGCTATTGGTGCAAATGTCTGAATCATGATGA  TAGGGACTGCGATCTATGGGTCCAAAGCAATGGT  ACCCTAATTGTGGACCAGTAGTAGTTTGGCCCGT  CTCCGAGAGCTCCACCATACAAAGCTGTGGGAA  AGGATGTGACTTATGTTCCAGGGTACTTTGAAAG  GAGAGTTGCAGAACTCGTGTGCAAAGCAGAGG  TGAAGATAATGTCCACACCACGATGCACGTAGAT  TCAACCACCATGACTCCGATGGAGGCGGTTCCA  AACGTGGAAGTGAAGGGTCGGGTGAAAGTTTG  AACATGGAAGTTGTTA</p>	522 pb	54/46	SH	<p><i>A. thaliana</i>  gi 297815890(e-V7e<sup>-7</sup>)</p>	<p>Zinc-Finger  “<i>RT non-LTR like</i>”</p>

<i>Rot20*</i>	AGGTGGTGCAATATTaggcagccgggaactcatgcTTTTG CGGACATTCCTTGTTTAGGTTGCAAAGTCCACA CAGCAACGGATTTGGCGGTTCTTTGTTCTTCATA GGTACTATGTTAGAAAAGCCATTTGGGATGTTGAA TGGACTTGATAAAAACCAGCTGCTAGCAACATCTT GACTTCTTGAGTTATCTGAGTTTCTACATCAATGT GAAAGACCTTGGCTGGTTGGATTACTGGCTTGAC TCCTGGGTCCACATTGAGAgcatggaccaccagccttgg TCCAGACCAAGCATCTCATCATAGGTCCATGGGA ACACATCCATGTATTCTTGAGCAAGGCCACTAA TTGCTCCTTTTCTTATGCTGTTACTCAGGAACTTC ATCA	392 pb	55/45	<i>Q. ilex</i> subsp. <i>rotundifolia</i> gi 89179403(e-V0,0) <i>V. vinifera</i> gi 147766829(e-V2e <sup>-5</sup> )	<i>A. officinalis</i> gi 89179403(e-V1e <sup>-24</sup> )	RT_1 "gypsy-like"
<i>Rot23</i>	<u>TAacaggaggcagagtaggtaaggGTGTGACATGCTGAC</u> <u>CCAGCTTCAACTTCAGACTTGACACATGAAAAAC</u> <u>ATGGTGAATT</u>	83 pb	55/45	SH	SH	
<i>Rot27</i>	<u>TAACATGCTCagattctgccttactctttaccacCATATCATCT</u> <u>ATATATATCTCAATGTTTTTCTAACTGTGTCTCA</u> <u>AACATCCTAGTCATCATTCTTTGATAGGTAGCCC</u> <u>CTGAATT</u>	121 pb	64/36	<i>L. japonicus</i> gi 189163049(e-V2e <sup>-10</sup> )	<i>V. vinifera</i> gi 225433516 (e-V1e <sup>-7</sup> ) <i>S. latifolia</i> gi 68685649 (e-V2e <sup>-6</sup> )	RT_1 "gypsy-like"

Nota: Na sequência de DNA, as letras sublinhadas indicam a região reconhecida pelas enzimas de restrição usadas na técnica de AFLP e as letras minúsculas, indicam a sequência dos *primers* usados para a análise de PCR. SH - Sem homologia. Quando o *e-value* apresentou valor superior a 10<sup>-3</sup>, não foi considerado significativo para análise e foi excluído. \*- A sequência do fragmento *Rot20* não possui os adaptadores, pois foi sequenciada directamente a partir do fragmento AFLP pela Dr.<sup>a</sup> Teresa Ribeiro.

Seguidamente procedeu-se a uma análise com o algoritmo BLASTX tendo assim sido analisadas as possíveis sequências proteicas que resultam da tradução das sequências de DNA. Os fragmentos *Rot2*, *Rot6*, *Rot9*, *Rot10*, *Rot10.1*, *Rot20* e *Rot27* mostraram homologia com sequências proteicas conhecidas, com *e-values*, variando entre  $1e^{-26}$  e  $1e^{-7}$  (Tabela 5). Nestas sequências foi também feita uma pesquisa de domínios conservados tendo os fragmentos *Rot6*, *Rot10*, *Rot20* e *Rot27* apresentado homologia com alguns domínios conservados.

A sequência *Rot2* possui 118 pb e é parcialmente homóloga de um transportador de membrana mecano-sensitivo apresentando maior homologia com uma sequência de nucleótidos de *Populus trichocarpa* (*e-value*  $1e^{-20}$ ) e com uma sequência proteica de *Ricinus communis* (*e-value*  $1e^{-11}$ ).

*Rot27* é constituída por 121 pb, e é homóloga de uma sequência genómica do cromossoma 5 de *Lotus japonicus* (*e-value*  $2e^{-10}$ ), de uma potencial proteína de *Vitis vinifera* (*e-value*  $1e^{-7}$ ) e do retroelemento *Retand* isolado em *Silene latifolia* (*e-value*  $2e^{-6}$ ), tendo a sua sequência completa de aminoácidos presente um domínio conservado correspondente a uma porção de uma transcriptase reversa tipo “*gypsy*”.

O fragmento *Rot6* possui 175 pb e em ambas as bases de dados, nucleótidos e proteínas, obteve maior homologia com uma sequência genómica de *V. vinifera* (*e-value*  $3e^{-68}$ ) e com uma proteína hipotética também de *V. vinifera* (*e-value*  $5e^{-22}$ ) que mostrou ser uma transcriptase reversa de um retrotransposição “*copia-like*” com domínio conservado a partir do aminoácido 15.

O fragmento *Rot20* com 392 pb, já publicado no GenBank (gi 89179403) é homólogo de uma transcriptase reversa “*gypsy-like*” com domínio conservado a partir do aminoácido 67 da sua *reading frame*. Possui maior homologia com uma sequência genómica de *V. vinifera* (*e-value*  $3e^{-5}$ ) e com uma proteína da família da transcriptase reversa de *Asparagus officinalis* (*e-value*  $1e^{-24}$ ).

A sequência *Rot9* é constituída por 407 pb e é homóloga da região *gag-pol* de um retrotransposição “*copia-like*” apresentando homologia com uma sequência genómica presente no cromossoma 5 de *L. japonicus* (*e-value*  $3e^{-3}$ ) e com uma sequência poliproteica *gag-pol* de *Glycine max* (*e-value*  $1e^{-9}$ ).

O fragmento *Rot10.1* que possui 522 pb, não obteve homologia na base de dados de nucleótidos, mas a homologia com *Arabidopsis thaliana* (*e-value*  $7e^{-7}$ ), na base de dados proteica revelou que este fragmento corresponde a uma região *Zync-finger* de uma RNase H de um retrotransposição não-LTR.

O fragmento *Rot10* com 525 pb, revelou que a sua sequência corresponde ao domínio conservado de uma RNase H completa, de um retrotransposição “*gypsy-like*”, apresentando, para além da sequência já publicada no GenBank (gi 222353756), maior

homologia com uma sequência genômica de *Phaseolus vulgaris* (*e-value*  $4e^{-4}$ ) e com uma RNase H do retroelemento *Retand* isolado em *Silene latifolia* (*e-value*  $1e^{-26}$ ).

A sequência *Rot8.11* com 124 pb, possui somente homologia com uma sequência genômica de *V. vinifera* (*e-value*  $1e^{-6}$ ).

Os fragmentos *Rot23*, *Rot4*, e *Rot8.24*, com 83, 123 e 207 pb respectivamente, não revelaram qualquer homologia com sequências existentes nas bases de dados.

### **3.2. Análise filogenética de sequências**

Para determinar o grau de homologia com outras espécies e inferir sobre a evolução das sequências isoladas efectuou-se uma análise filogenética com espécies que possuem o seu genoma parcial ou totalmente sequenciado pertencentes a vários grupos taxonômicos: duas *Fabaceae* (*Medicago truncatula* e *Glycine max*) família filogeneticamente próxima da família *Fagaceae*, duas lenhosas de diferentes famílias, *Salicaceae* e *Vitaceae* (*Populus trichocarpa* e *Vitis vinifera*), uma *Brassicaceae* (*Arabidopsis thaliana*) e duas monocotiledóneas (*Zea mays* e *Oryza sativa*).

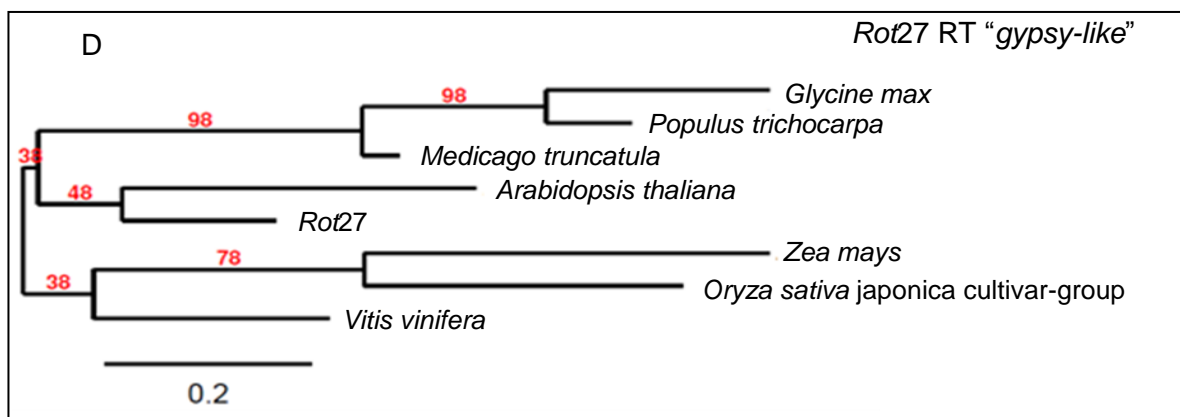
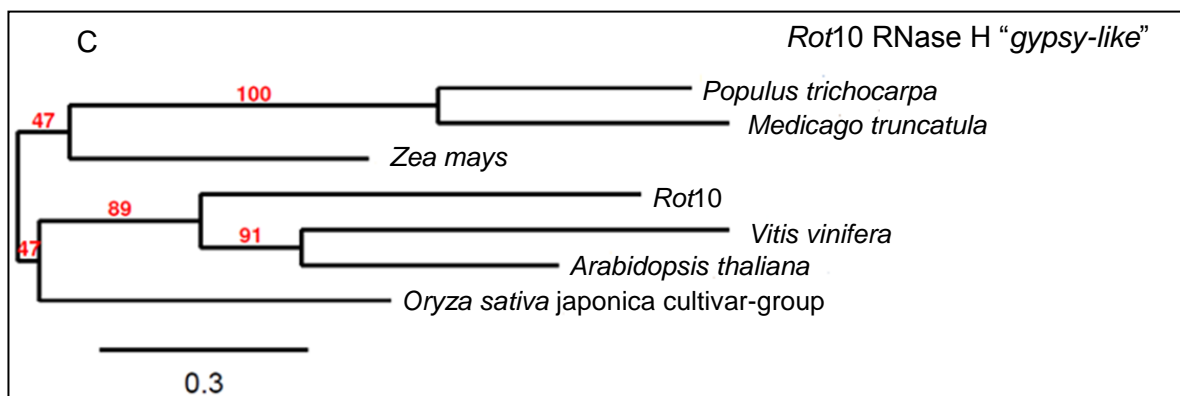
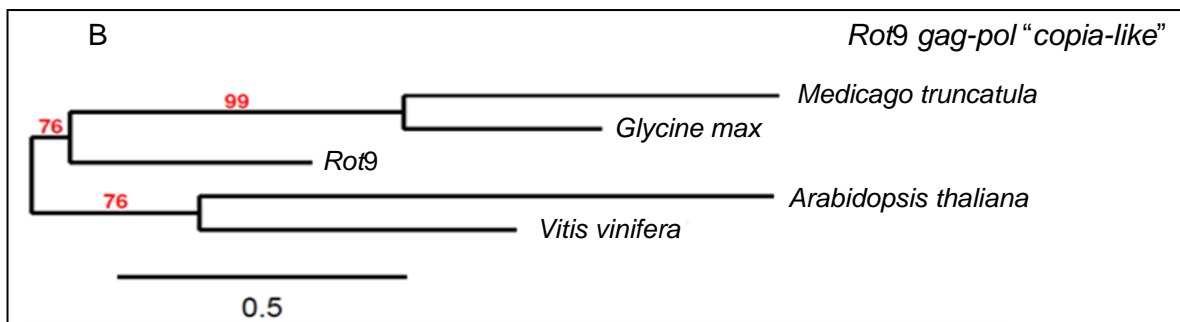
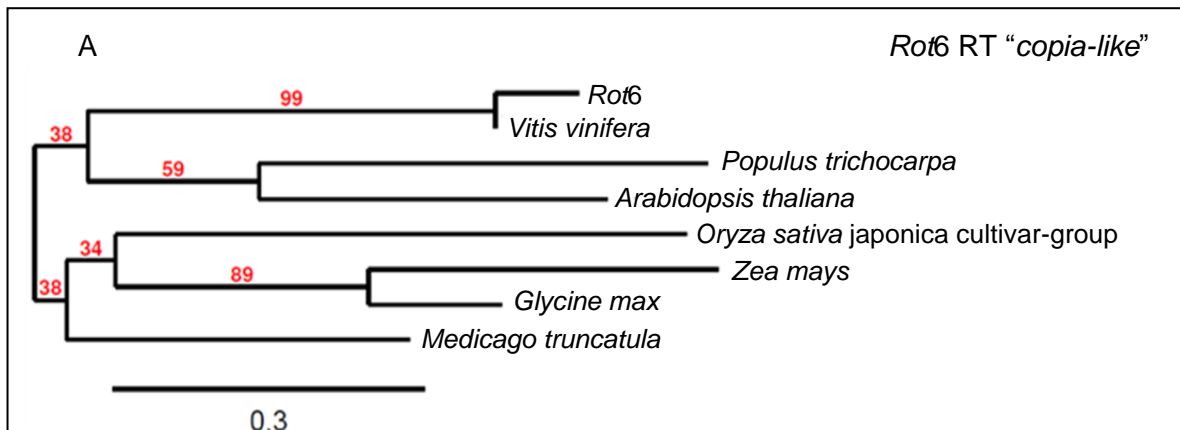
A construção das árvores filogenéticas foi feita a partir dos resultados obtidos com o Blastx, ou seja, foram usadas sequências proteicas. A escolha das sequências a utilizar na análise teve como base o valor de *e-value* igual ou menor a  $10^{-4}$  (Tabela 6) tendo sido apenas considerada a região homóloga em todas as espécies.

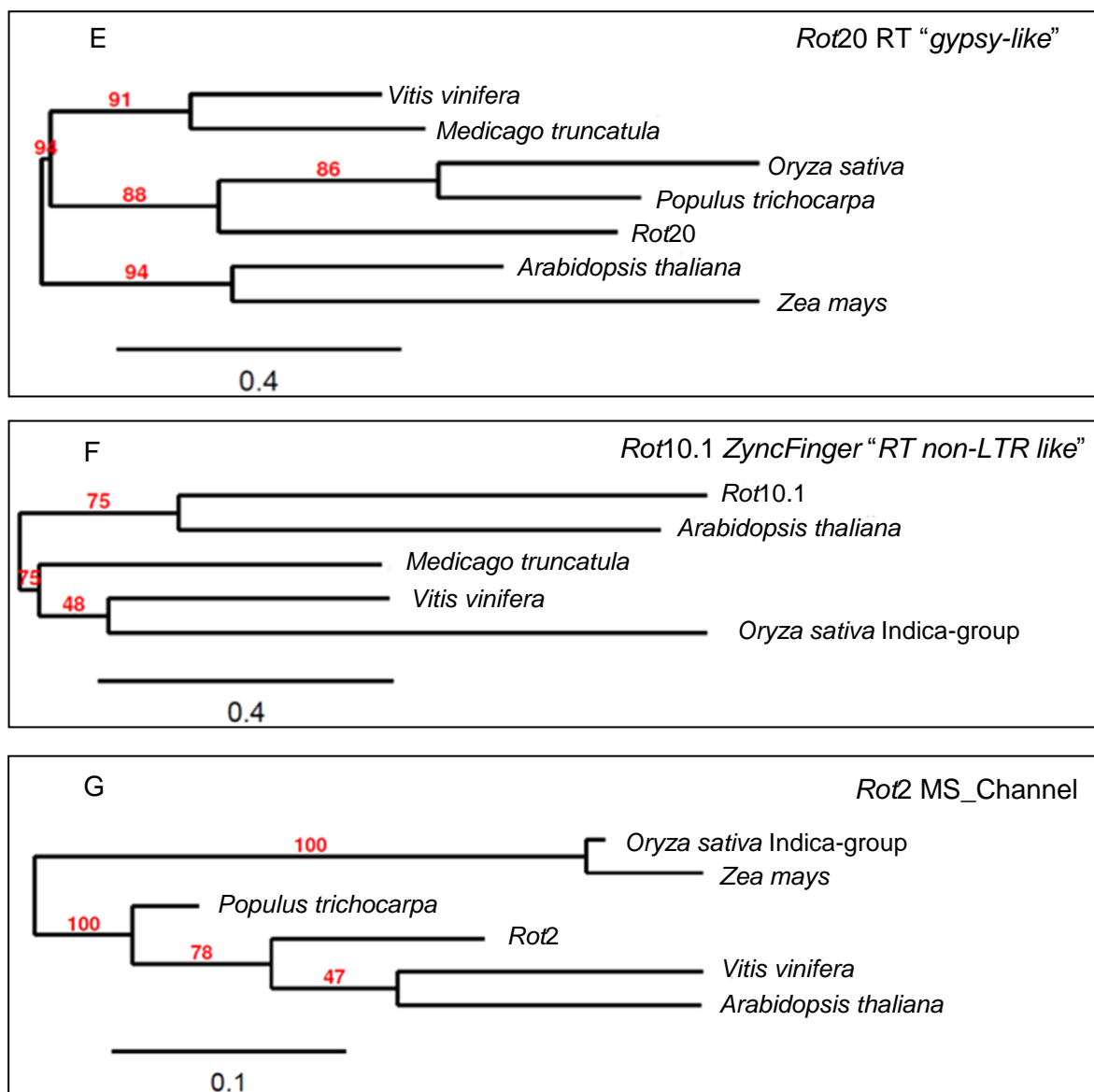
**Tabela 6.** Sequências proteicas obtidas pela tradução dos clones, respectiva open reading frame e sequências homólogas em diferentes espécies.

Nome	Sequência	Reading Frame	Homologia com proteínas
Rot2	TVFIKLNNEKVYYPNSVLAAPISNYRSPDMGDYVE	-3	<i>A.thaliana</i> gi 15239769; <i>V.vinifera</i> gi 296081700; <i>Z.mays</i> gi 259490196; <i>O.sativa</i> gi 125540611; <i>P.trichocarpa</i> gi 224092103
Rot6	WELAMKDEMDSLLGNQTWELTELPVEKKTLNKVVYIK NKHDGSKRYKVRV	+1	<i>V.vinifera</i> gi 225470676; <i>O.sativa</i> gi 77551524; <i>A.thaliana</i> gi 2642431; <i>Z.mays</i> gi 4234852; <i>P.trichocarpa</i> gi 10967636; <i>M.truncatula</i> gi 220966838; <i>G.max</i> gi 207367152
Rot9	QCFGCQGYGHMKSERPTYLRSKDKAMAVTSLSDNEVS DDESGCDKNGNFIAFTATAVVNESVYVEENPSDGELS EDVDL	+2	<i>G.max</i> gi 29423276; <i>M.truncatula</i> gi 124359853; <i>V.vinifera</i> gi 147801346; <i>A.thaliana</i> gi 9294242
Rot10	FNVQYRPHTTMKGQAVADFAEFTNMEGQGAEEHPQ WSIHADGLSNKQACGKGIVLHSLEGDEIECMVHLDFTP TNNEAEYVALVVGLDLAKAVGAICVVVYVNFVVTSQVN GDYECKGERMCKYLEQVRKRVDLQVGFVQIPREEN KQADCLAKAASAEHMLIPNKVLSFV	+3	<i>A.thaliana</i> gi 4417309; <i>V.vinifera</i> gi 147826766; <i>Z.mays</i> gi 33113963; <i>O.sativa</i> gi 110288766; <i>P.trichocarpa</i> gi 224123090; <i>M.truncatula</i> gi 87162496
Rot10.1	FFHVRVVIDMTLLLCRGLITLPNGGKRWIGFKYQLPS LCYWCKCLNHDDRDC	+3	<i>M.truncatula</i> gi 124359163; <i>O.sativa</i> gi 218196866; <i>V.vinifera</i> gi 147823314; <i>A.thaliana</i> gi 116830453
Rot20	QLVALLKEYMDVFPWYDEMGLDPRLLVHALNVDPG VKPVIQPAKVFHIDVETQITQEVKMLLAAGFIKSIQHPK WLSNIVPMKNKKNRQIRCCVDFCNLNKECPQKHEFP	-1 e -2	<i>O.sativa</i> gi 116310100; <i>V.vinifera</i> gi 147785235; <i>M.truncatula</i> gi 124359822; <i>A.thaliana</i> gi 8778610; <i>Z.mays</i> gi 4206306; <i>P.trichocarpa</i> gi 224107098
Rot27	NSGATYQRMTRMFETQLGKNIEIYIDDMVVKSKAESE HV	-1	<i>P.trichocarpa</i> gi 224147145; <i>Z.mays</i> gi 23928443; <i>G.max</i> gi 13559349; <i>O.sativa</i> gi 77555311; <i>M.truncatula</i> gi 87162496; <i>V.vinifera</i> gi 225433516; <i>A.thaliana</i> gi 4417309

Para o alinhamento das sequências proteicas e construção das árvores filogenéticas foi usado o software disponibilizado no site <http://www.phylogeny.fr/>. O alinhamento das sequências foi obtido com o algoritmo MUSCLE e para a construção das árvores filogenética foi utilizado o método da máxima verosimilhança (ANISIMOVA and GASCUEL 2006; CHEVENET *et al.* 2006; DEREPPER *et al.* 2008; EDGAR 2004; GUINDON and GASCUEL 2003).

Para o limite de *e-value* considerado ( $10^{-4}$ ), apenas Rot6 e Rot27 obtiveram homologia com as sete espécies escolhidas (Tabela 6; Figura 2:A,D).





**Figura 2.** Filogramas construídos pelo método da máxima verossimilhança. Os números nos ramos representam os valores de 100 *bootstrap*. O valor na barra horizontal indica o número de substituições.

Dois dos fragmentos estudados *Rot6* e *Rot9*, apresentam homologia com porções de retrotransposições “*copia-like*” (Tabela 5). No filograma de *Rot6* (Figura 2:A) os valores de *bootstrap* dos dois *clades* iniciais são baixos assim como o do valor do ramo que distingue *M. truncatula*, das outras espécies, no entanto *Rot6* aparece fortemente associada a *V. vinifera* com um *bootstrap* de 99 %. No filograma de *Rot9* (Figura 2:B) são nítidos dois *clades* onde esta sequência se encontra associada às duas espécies de *Fabaceae*.

Três dos clones estudados mostraram homologia, com retrotransposições “*gypsy-like*” (*Rot10*, *Rot27* e *Rot20*) (Tabela 5). Destes, *Rot10* e *Rot27* parecem pertencer ao mesmo

retrotransposição, devido à homologia com o mesmo retroelemento (*Retand* de *S. latifolia*) (Tabela 5), no entanto as suas árvores filogenéticas são distintas. Dos dois fragmentos apenas *Rot27* apresenta homologia com as sete espécies escolhidas enquanto *Rot10* apresenta homologia apenas com seis espécies (Tabela 6). Como os filogramas não têm raiz, não existe uma boa medida de fiabilidade dos *clades* iniciais, apenas se pode inferir sobre os *clades* seguintes. No filograma de *Rot10* (Figura 2:C) o clone encontra-se num *clade* com *V. vinifera* e *A. thaliana* suportado por um alto valor de *bootstrap* (89 %), ao contrário de *Rot27* (Figura 2:D), em que o *clade* onde se encontra tem um valor de *bootstrap* de 48 %. *Rot20* apresenta homologia, também, apenas com seis espécies (Tabela 6), encontrando-se num *clade* com choupo e arroz que é suportado por valores de *bootstrap* elevados (88 %) (Figura 2:E). O clone *Rot10.1* homólogo de um “zinc-finger” (Tabela 5) pertence possivelmente a um retrotransposição sem extremidades LTR, portanto diferente dos anteriores. No filograma (Figura 2:F) *Rot10.1* encontra-se no mesmo *clade* que *A. thaliana* suportado por um valor de *bootstrap* de 75 %. O filograma do fragmento *Rot2* (Figura 2:G), um transportador de membrana mecano-sensitivo (Tabela 5), encontra-se associado com *V. vinifera* e *A. thaliana*, o que é suportado por elevados valores de *bootstrap* (78 %). Os restantes *clades* são bem definidos com *bootstrap* de 100 %, separando assim monocotiledóneas de dicotiledóneas, e os posteriores agrupamentos que se observam no *clade* das dicotiledóneas.

### **3.3. Caracterização das regiões genómicas inter-sequências AFLP em diferentes espécies de Fagaceae**

#### **3.3.1. Caracterização molecular**

De modo a estudar regiões adjacentes ou parcialmente homólogas aos fragmentos de AFLP previamente caracterizados e a determinar o modo como estas regiões se encontram organizadas no genoma de azinheira, procedeu-se à amplificação das regiões ISA através de PCR usando *primers* desenhados para os fragmentos previamente sequenciados (Tabela 7).

**Tabela 7.** *Primer(s)* desenhados para cada uma das sequências e a sua orientação no genoma.

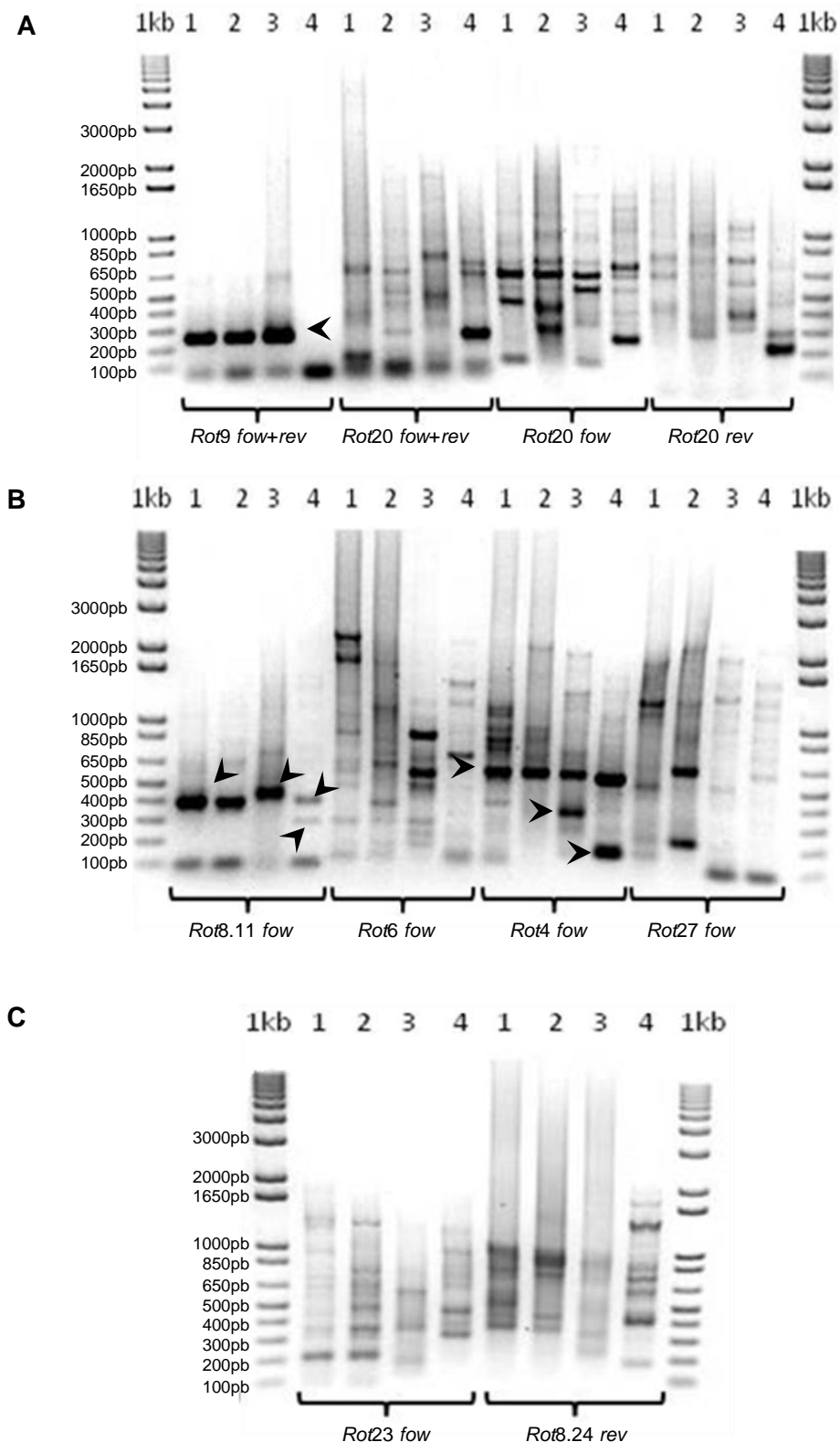
Nome	<i>Primer</i>	Posição do <i>primer</i> no genoma
<i>Rot4</i>	5' aacaagtttagcatccaagtcc 3'	
<i>Rot6</i>	5' ggatgagatggattccttgg 3'	
<i>Rot8.11</i>	5' gttgtgtggtttgtcacggatt 3'	
<i>Rot8.24</i>	5' gggaggagcggagttggag 3'	
<i>Rot9</i>	5' ttctatgggtcctcagtgtttgg 3' 5' atggggaactctctgaggatgtaga 3'	
<i>Rot20</i>	5' ggcagccgggaactcatgc 3' 5' gcatggaccaccagccttga 3'	
<i>Rot23</i>	5' acaggaggcagagtaggtaagg 3'	
<i>Rot27</i>	5' agattctgccttactcttaccac 3'	

(F), *primer forward*; (R), *primer reverse*.

Simultaneamente, com o objectivo de comparar a organização genómica destas sequências noutras espécies da mesma família, estudaram-se os padrões de amplificação obtidos com os mesmos *primers* em *Q. suber*, *C. sativa* e *F. sylvatica*.

Com o objectivo de conhecer melhor as sequências para as quais não foi detectada homologia usaram-se nesta análise os clones *Rot4*, *Rot8.11*, *Rot8.24*, *Rot23*. Para além destes fragmentos e devido à organização dos retroelementos nos genomas, foram também usadas as sequências homólogas de retroelementos *Rot6*, *Rot9*, *Rot20* e *Rot27* (Tabela 5). Devido ao tamanho reduzido de algumas sequências e assumindo que os retroelementos se podem encontrar muito repetidos no genoma e em orientações diferentes optou-se por usar, nalguns casos, um só *primer* (*fow* ou *rev*), o que nos permite detectar sequências que partilham o mesmo primer em orientação inversa, para além de um par de *primers* (*fow* e *rev*) nas sequências *Rot9* e *Rot20* (Tabela 3 e 7).

Todas as amplificações originaram padrões diferentes de bandas em todas as espécies estudadas, apesar das duas espécies de *Quercus* partilharem, algumas vezes padrões mais semelhantes com fragmentos do mesmo tamanho (Figura 3:A).



**Figura 3.** Electroforese de agarose a 1 % dos fragmentos amplificados por PCR a partir de DNA genômico de quatro espécies de *Fagaceas*. 1-*Q. ilex* subsp. *rotundifolia*; 2-*Q. suber*; 3-*C. sativa*; 4-*F. sylvatica*. Os *primers* foram desenhados a partir de seqüências de *Q. ilex* subsp. *rotundifolia*. *fow* – *primer forward*; *rev* – *primer reverse*; *fow+rev* – *primers forward+reverse*. As setas indicam as bandas referidas no texto. O marcador usado foi 1Kb<sup>+</sup>.

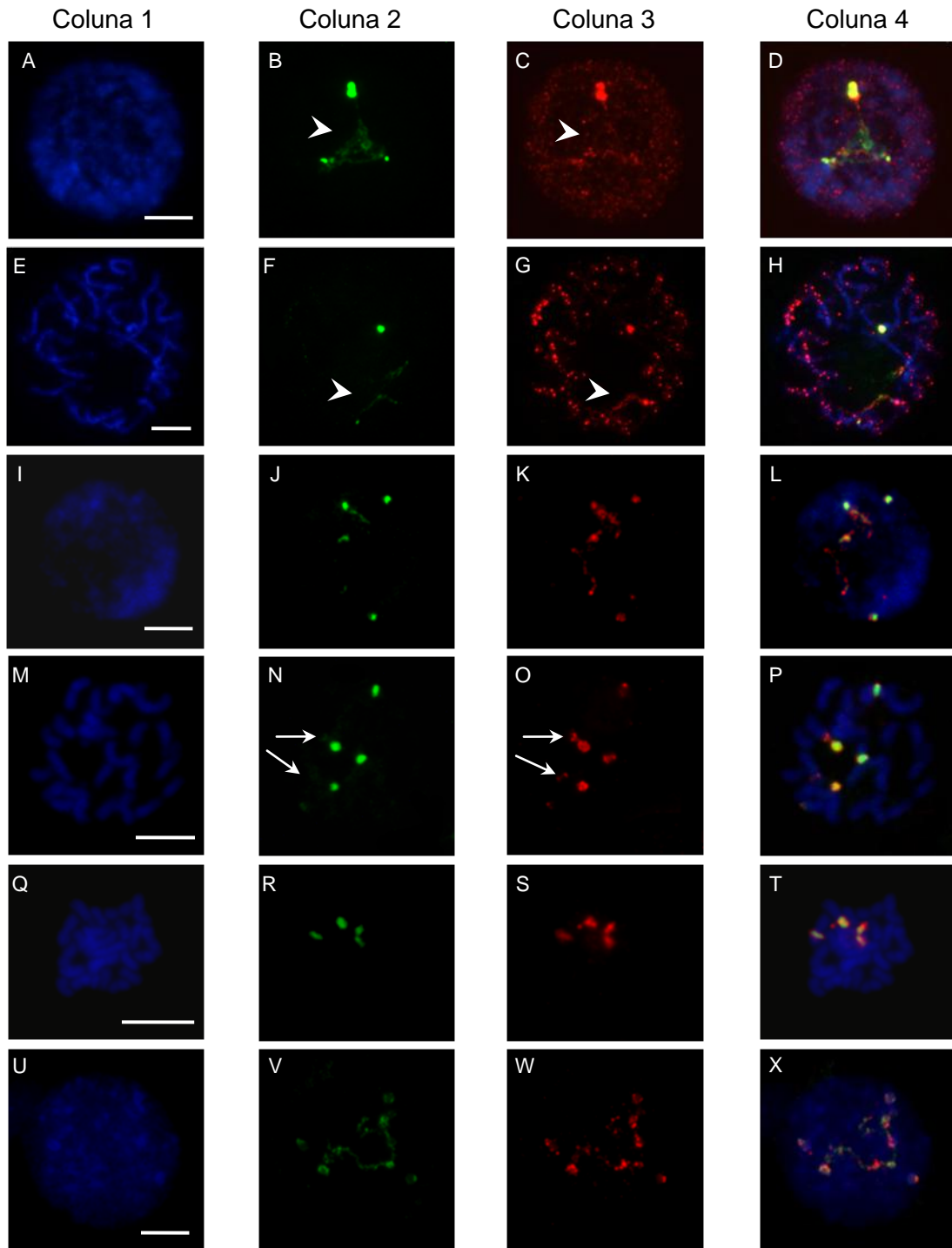
*Rot9* e *Rot8.11* apresentam o mesmo padrão em azinheira e em sobreiro. Assim, usando os primers desenhados para *Rot9* (Tabela 3) foi amplificada uma banda de ~250 pb em azinheira, sobreiro e castanheiro (seta), enquanto em faia não se obteve qualquer amplificação (Figura 3:A). Usando os primers para *Rot8.11* (Tabela 3) observou-se uma banda nas duas espécies de *Quercus* com ~400 pb (seta). Em *C. sativa* obteve-se também uma única banda embora com um tamanho um pouco maior (~450 pb, seta) e em *F. sylvatica* obtiveram-se duas bandas, de ~300 pb e 400 pb, embora menos nítidas (Figura 3:B). Tanto com os primers de *Rot8.24* como de *Rot23* (Tabela 3) os padrões são semelhantes nas duas espécies de *Quercus* e diferentes em castanheiro e faia, tendo sido obtido um máximo de sete fragmentos entre os ~200 pb e os 1.8 Kb (Figura 3:C). O padrão de amplificação usando os primers desenhados para *Rot4* (Tabela 3) apesar de ser diferente em todas as espécies, é amplificado em *Quercus* e em castanheiro um fragmento de ~650 pb (seta). Em castanheiro e faia amplificou-se mais um fragmento bem nítido em cada uma das espécies com ~350 pb (seta) e ~200 pb (seta), respectivamente (Figura 3:B). O padrão de amplificação dos primers desenhados para *Rot6* e *Rot27* (Tabela 3) é diferente em todas as espécies, com fragmentos variando entre os ~150 pb e ~2.5Kb (Figura 3:B). O padrão obtido com os primers desenhados para a sequência *Rot20* (Tabela 3) é distinto em todas as espécies, quer com o conjunto de *primers fow* e *rev* quer quando foi usado um *primer* individualmente variando o tamanho dos fragmentos entre 100 pb e ~1.5 Kb (Figura 3:A).

### 3.3.2. Caracterização citogenética - Hibridação *in situ* fluorescente

Com o objectivo de estudar a distribuição e organização *in situ* das regiões genómicas intersequências AFLP, obtidas no genoma de azinheira e comparar os padrões com os observados noutras espécies de *Fagaceae* procedeu-se a uma análise por FISH em núcleos interfásicos e em cromossomas de *Q. ilex* subsp. *rotundifolia*, *Q. suber*, *C. sativa* e *F. sylvatica*.

Para a hibridação *in situ* foram escolhidos os fragmentos resultantes das amplificações descritas anteriormente obtidas em genoma de azinheira a partir das sequências, *Rot8.24*, *Rot23* e *Rot20* (Figura 3:A,C). As sequências genómicas obtidas com *Rot8.24* e *Rot23* foram escolhidos para esta análise pois os respectivos fragmentos AFLP não apresentaram homologia com sequências conhecidas (Tabela 5), sendo que esta seria uma maneira de os caracterizar. Adicionalmente, a sequência de *Rot20* também foi escolhida pois é homóloga de um retroelemento (Tabela 5). Em todas as situações foi utilizada simultaneamente a sonda de rDNA 45S de trigo pTa71 de modo a permitir uma análise comparativa com os *loci* ribossomais.

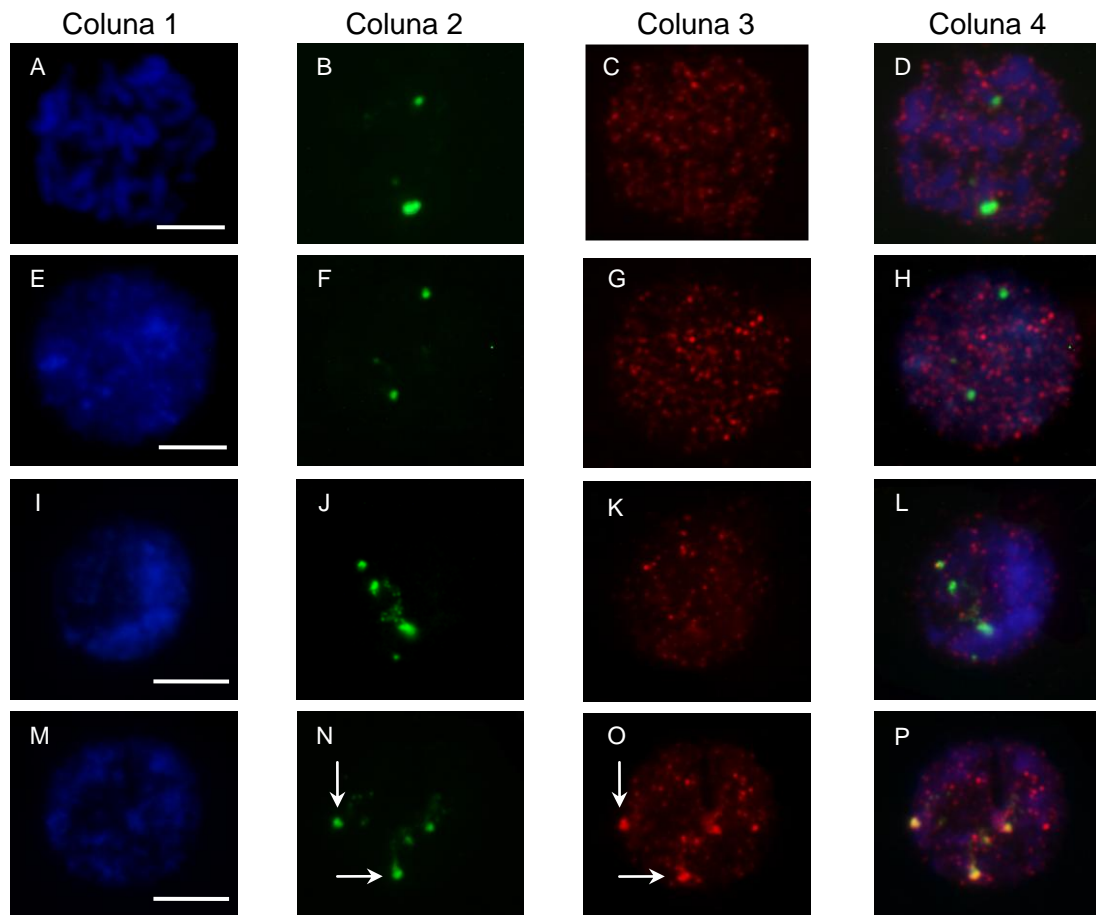
A sonda *Rot8.24* apresenta em *Q. ilex* subsp. *rotundifolia*, e em *Q. suber* (figura não apresentada) um padrão disperso por todo o genoma marcando fortemente a região do NOR em núcleos interfásicos, incluindo o domínio descondensado (Figura 4:A-D cabeça de seta), e a constrição secundária em cromossomas profásicos (Figura 4:E-H; cabeça de seta).



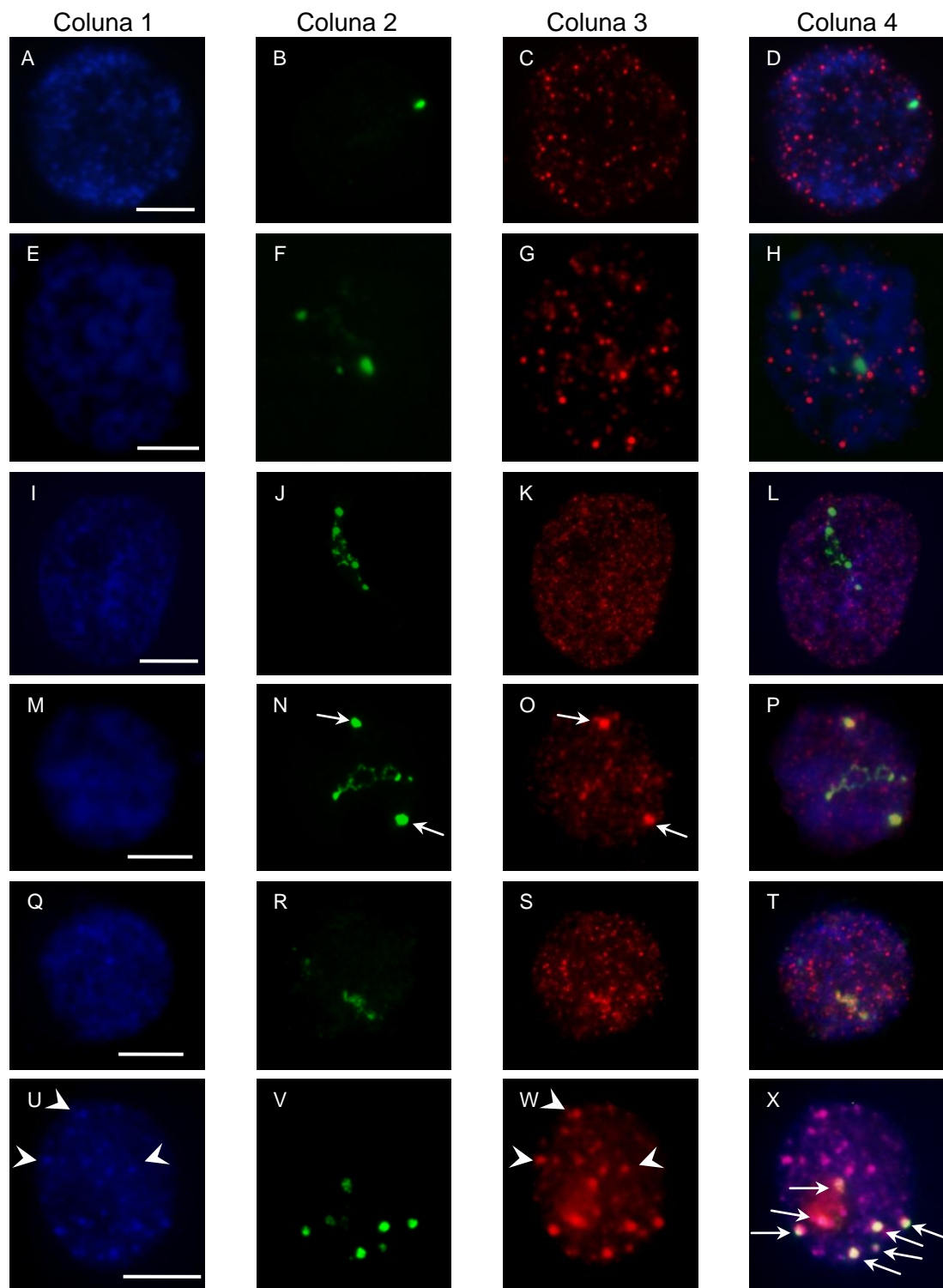
**Figura 4.** Detecção simultânea com FISH das sondas *Rot8.24* e rDNA 45S em núcleos interfásicos e cromossomas metafásicos de *Quercus ilex* subsp. *rotundifolia*: A-H; *Castanea sativa*: I-T e *Fagus sylvatica*: U-X. Os núcleos e cromossomas estão corados com DAPI (coluna 1) a sonda de rDNA 45S, pTa71 a verde (coluna 2), *Rot8.24* a vermelho (coluna 3) e sobreposição dos três fluorocromos na coluna 4. A sonda *Rot8.24* co-localiza-se parcialmente com o rDNA 45S (sinais amarelos) em todas as espécies. Marcação existe não só nos blocos condensados mas também nas zonas descondensadas (cabeças de seta). As zonas de não sobreposição são bem visíveis em *C. sativa* (setas brancas). Barra  $\cong$  10  $\mu$ m

No entanto, em *C. sativa* e *F. sylvatica* a marcação é restrita à região do NOR, onde se observa a co-localização parcial das duas sondas tanto na zona condensada como na região descondensada deste locus (Figura 4:I-X). A hibridação simultânea permite-nos detectar zonas sem sobreposição evidenciando uma zona próxima ao centrómero muito condensada (Figura 4:N,O; seta). Adicionalmente a intensidade da sonda *Rot8.24* é maior que a do rDNA 45S (Figura 4:N,O).

O mesmo padrão disperso com distribuição uniforme ao longo de todo o genoma foi observado com a sonda *Rot23* e *Rot20* em núcleos e cromossomas de *Q. ilex* subsp. *rotundifolia*, *Q. suber* e *C. sativa* (coluna 3, Figura 5:A-P e Figura 6:A-T). Curiosamente estas sondas não se co-localizam com o rDNA nos núcleos de *Quercus* contrariamente ao que acontece em núcleos de castanheiro e faia. Em núcleos de castanheiro onde a região 45S rDNA forma blocos bastante condensados, as duas sondas marcam estas zonas com intensidade (coluna 2 e 3, Figura 5:M-P e Figura 6:M-P, setas).



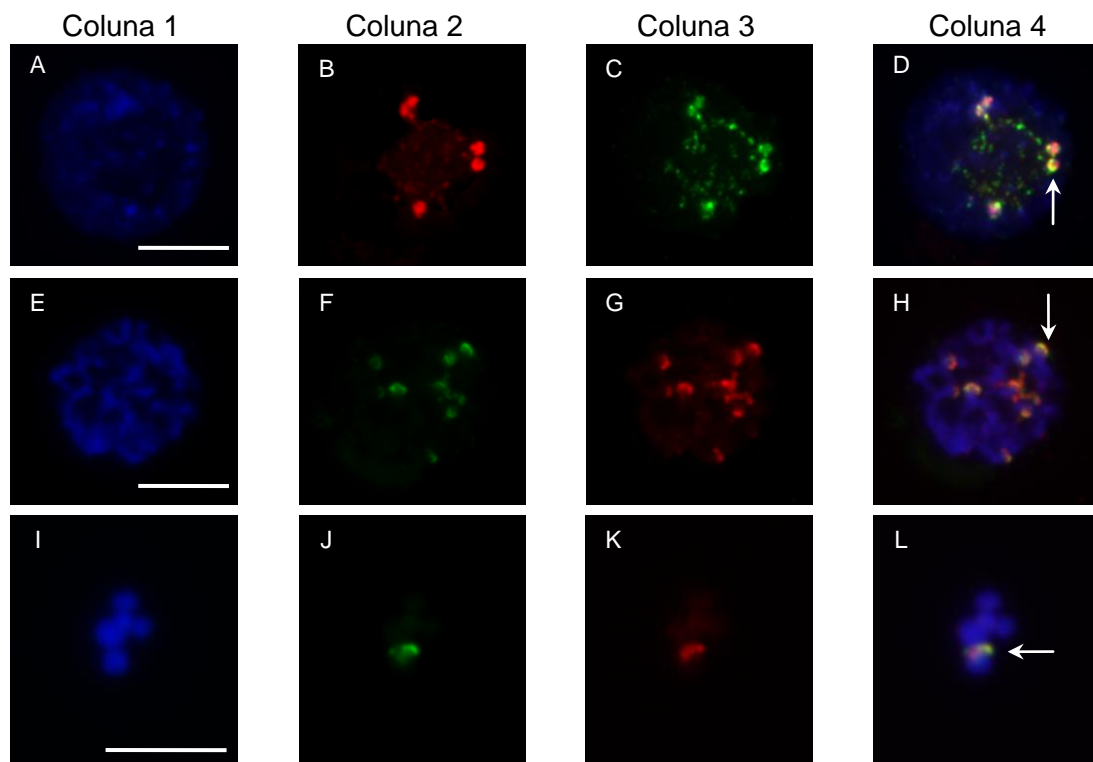
**Figura 5.** Detecção simultânea com FISH das sondas *Rot23* e rDNA 45S em núcleos interfásicos e cromossomas de *Quercus ilex* subsp. *rotundifolia*: A-H; *Quercus suber*: I-L; *Castanea sativa*: M-P. Os núcleos e cromossomas estão corados com DAPI (coluna 1). A sonda de rDNA 45S, pTa71 a verde (coluna 2), *Rot23* a vermelho (coluna 3) e sobreposição dos três fluorocromos na coluna 4. A sonda *Rot23* mostra um padrão disperso em todas as espécies e uma co-localização com os *loci* 45S em *C. sativa* (setas). Barra  $\cong$  10  $\mu$ m



**Figura 6.** Detecção simultânea com FISH das sondas *Rot20* e rDNA 45S em núcleos interfásicos e cromossomas de *Quercus ilex* subsp. *rotundifolia*: A-H; *Quercus suber*: I-L; *Castanea sativa*: M-T e *Fagus sylvatica*: U-X. Os núcleos e cromossomas estão corados com DAPI (coluna 1). A sonda de rDNA 45S, pTa71 a verde (coluna 2), *Rot20* a vermelho (coluna 3) e sobreposição dos três fluorocromos na coluna 4. A sonda *Rot20* mostra um padrão disperso em todas as espécies e uma co-localização com os *loci* 45S em *C. sativa* e *F. sylvatica* (setas). Em faia há uma marcada preferência de localização desta sonda pelos cromocentros (cabeças de seta). Barra  $\cong 10 \mu\text{m}$

Em faia, a sonda *Rot20* localiza-se também nas zonas heterocromáticas pericentroméricas (cabeça de seta) deste genoma para além da co-localização com os *loci* de rDNA (Figura 6:U-X, cabeças de seta).

Para tentar encontrar sequências específicas das zonas pericentroméricas de faia, produziu-se a sonda *Syl20* com DNA de faia utilizando os *primers* desenhados para a sequência *Rot20*. Esta sonda foi usada em núcleos e cromossomas de faia juntamente com a sonda de DNA ribossomal 45S pTA71. Contrariamente ao esperado, esta sonda apresenta exclusivamente uma co-localização com o rDNA 45S tanto nas regiões condensadas como na zona de descondensação marcando intensamente os NORs desta espécie em núcleos interfásicos (Figura 7:A-D) e em cromossomas profásicos (Figura 7:E-H) e metafásicos (Figura 7:I-L), não apresentando qualquer sinal de hibridação no resto do núcleo, incluindo nos cromocentros.



**Figura 7.** Detecção simultânea com FISH das sondas *Syl20* e rDNA 45S em núcleos interfásicos (A-D) e cromossomas (E-L) de *Fagus sylvatica*. Os núcleos e cromossomas estão corados com DAPI (coluna 1). A sonda de rDNA 45S pTa71 (coluna 2) a vermelho (B, D) e verde (F, H, J, L). A sonda *Syl20* (coluna 3) encontra-se a verde (C, D) e vermelho (G, H, K, L) e sobreposição dos três fluorocromos na coluna 4. Amarelo indica a co-localização das duas sondas (setas). Barra  $\cong 10 \mu\text{m}$

## 4. Discussão

A família *Fagaceae*, onde se incluem algumas espécies representadas em Portugal como *Quercus suber*, *Quercus ilex* subsp. *rotundifolia*, *Castanea sativa*, e *Fagus sylvatica*, é uma das famílias florestais mais importantes nos ecossistemas silvícolas do hemisfério Norte. O conhecimento dos seus genomas, tanto a nível molecular como citogenético, assume pois grande relevância. Neste trabalho foram caracterizadas várias sequências do genoma de *Quercus ilex* subsp. *rotundifolia* previamente obtidas por AFLP (Tabela 1) tendo como finalidade determinar a contribuição destas sequências na organização genómica das *Fagaceae*. Esta espécie foi a eleita por pertencer à mesma espécie de *Fagaceae* com o maior genoma publicado *Q. ilex* subsp. *ilex* (KREMER *et al.* 2007; ZOLDOS *et al.* 1998). A contribuição das sequências não génicas, nomeadamente das repetitivas para o tamanho dos genomas é amplamente aceite (BENNETZEN 2005; UNGERER *et al.* 2006), sendo assim de esperar maior facilidade na obtenção destas sequências em *Quercus ilex* subsp. *rotundifolia* do que noutras espécies com menor genoma.

Nesta caracterização foram estudados onze fragmentos (Tabela 1) dos quais seis pertencem a retrotransposições: dois “*copia-like*”, três “*gypsy-like*” e um retrotransposição sem extremidades LTR (LINE) (Tabela 5). Este resultado mostra que os retroelementos estão muito representados no genoma das *Fagaceae*, como aliás tem sido descrito para a maioria das espécies vegetais (BENNETZEN 2005; NEUMANN *et al.* 2001) e para outros *Quercus* (ZOLDOS *et al.* 2001). Estes seis fragmentos apresentam homologia com retroelementos existentes em espécies de diferentes famílias como *Fabaceae*, *Salicaceae*, *Vitaceae*, *Brassicaceae* e duas monocotiledóneas, indicando a sua natureza ubíqua nas plantas. No entanto, a análise filogenética feita a todas as sequências, assim como a análise de homologia com sequências do GeneBank revelou relações filogenéticas muito diversas (Tabela 5 e 6; Figura 2). Dos sete clones analisados somente *Rot9* produziu um filograma (Figura 2:B) que está de acordo com as relações filogenéticas actualmente aceites para as angiospérmicas (MOORE *et al.* 2010), agrupando as duas espécies de *Fabaceae* como sendo as mais homólogas, e colocando a sequência de *Quercus* (*Fagaceae*) com *V. vinifera* e *A. thaliana* num *clade* à parte. O alinhamento com as espécies de monocotiledóneas (*Z. mays* e *O. sativa*) não apresentou um *e-value* significativo. Este conjunto de resultados parece indicar que *Rot9* manteve um elevado grau de homologia ao longo da evolução das espécies de dicotiledóneas tendo-se distanciado das espécies de monocotiledóneas. Pelo contrário, todas as outras sequências de retrotransposições analisados não traduzem a evolução filogenética dos genomas em que se encontram, o que significa que de alguma forma estiveram sujeitas a uma pressão selectiva diferente. Este tipo de resultados poderia

indicar uma transmissão horizontal uma vez que dois dos critérios são satisfeitos: a) elevada conservação entre sequências de elementos transponíveis (TE) em espécies filogeneticamente distantes; b) incongruência entre elementos transponíveis e a filogenia do “hospedeiro”. Contudo, teria ainda que ser verificada uma distribuição descontínua dos TE num grupo de espécies e uma sobreposição geográfica, temporal e ecológica, para além de ter que se excluir a hipótese da existência de um polimorfismo ancestral que conduziu a uma fixação diferencial nas várias espécies. Adicionalmente, restrições selectivas a actuar sobre as sequências de TE necessitariam de ser consideradas (KUMEKAWA *et al.* 1999; STUART-ROGERS and FLAVELL 2001). Consideramos pois que vários factores, tais como, factores epigenéticos presentes no hospedeiro, plasticidade fenotípica e pressão selectiva diferencial actuaram sobre as sequências dos retroelementos conduzindo aos resultados obtidos.

Com o objectivo de conhecer a organização destas sequências em diferentes géneros de *Fagaceae*, foram amplificadas as regiões ISA de alguns clones homólogos de retroelementos em *Q. ilex* subsp. *rotundifolia*, *Q. suber*, *C. sativa* e *F. sylvatica* (Figura 3). A utilização destes três géneros de *Fagaceae* permite-nos comparar os padrões gerados e inferir a sua relação evolutiva, uma vez que *F. sylvatica* pertence ao género mais basal dentro da família (MANOS *et al.* 1999) e o género *Castanea* é considerado o mais recente, embora relativamente ainda próximo do género *Quercus* (MANOS *et al.* 2001). Os padrões gerados (Figura 3) são globalmente diferentes em todas as espécies, sendo possível detectar maiores semelhanças entre as espécies de *Quercus* reflectindo assim as relações filogenéticas mais próximas entre estas duas espécies. Alguns dos padrões, mostrando bandas múltiplas, foram obtidos com um único primer (Tabela 7), o que indica que este se encontra muito representado, presente em diferentes localizações e em direcções opostas no genoma das quatro espécies estudadas. Este tipo de organização é típica dos retroelementos que, para além de muito representados, podem encontrar-se inseridos em direcções opostas (BENNETZEN 2000; TAKAHASHI *et al.* 1997), inclusive dentro de outros retroelementos (KUYKENDALL *et al.* 2009).

Para além dos fragmentos com homologia no GeneBank, foram também isolados fragmentos que não apresentam qualquer homologia com sequências depositadas nas bases de dados (Tabela 5), o que sugere que se tratam de sequências genómicas específicas das *Fagaceae*. As sequências genómicas muito repetidas, também designadas DNA satélite são normalmente ricas em regiões AT, tal como as estudadas neste trabalho, e contribuem, assim como os retroelementos, para a heterocromatina que constitui uma grande fracção dos genomas vegetais (LAMB *et al.* 2007). De modo a caracterizar estas regiões, fomos estudar a localização e organização de uma sequência conhecida e duas sem homologia (*Rot20*, *Rot8.24* e *Rot23*, respectivamente) através da hibridação *in situ*

fluorescente (Figuras 4, 5 e 6) em núcleos interfásicos e cromossomas dos três géneros *Quercus*, *Fagus* e *Castanea*, usando como sonda as regiões ISA do genoma de *Q. ilex* subsp. *rotundifolia* (Figura 3). Os padrões de FISH são idênticos nas duas espécies de *Quercus* e diferentes dos de *Castanea* e *Fagus*, onde se observam padrões semelhantes entre elas. As três sondas mostram um padrão disperso (coluna 3 das Figuras 4:C,G; 5 e 6) como o normalmente exibido pelos retrotransposições que não têm inserção preferencial e por isso se encontram dispersos nos genomas (NEUMANN *et al.* 2001; WEBER *et al.* 2010). Para além deste tipo de organização, os retroelementos podem também acumular-se em regiões heterocromáticas específicas dos cromossomas como as regiões sub-teloméricas nos cereais (ANANIEV *et al.* 1998) e as pericentroméricas em *Arabidopsis* (PETERSON-BURCH *et al.* 2004). Nas espécies estudadas, as zonas heterocromáticas mais evidentes são os cromocentros de faia, onde se detecta uma acumulação da sonda produzida com os primers *Rot20* (Figura 6:W), evidenciando a potencial contribuição de retroelementos para a organização das regiões centroméricas ou pericentroméricas desta espécie. Porém, quando é usada a sonda *Syl20* que foi obtida a partir do genoma de faia o padrão é radicalmente alterado, observando-se hibridação somente nas regiões NOR. Esta situação indica que os fragmentos específicos da zona centromérica não foram amplificados no genoma de faia com estes primers (Figura 7, coluna 3). Surpreendentemente também as outras sondas usadas mostram uma co-localização com a região NOR, em todos os *loci* e ao longo de cada *locus*, tanto nos domínios condensados como nos descondensados. Nas regiões NOR as centenas ou milhares de unidades de rDNA 45S encontram-se repetidas em tandem (NEVES *et al.* 2005), de modo que o padrão FISH detectado mostra que estes elementos se encontram distribuídos ao longo de todo o *locus*, provavelmente intercalados com as cópias de rDNA 45S (Figura 4, 5 e 6, coluna 3). A existência de retrotransposições na região dos NORs está descrita unicamente para o genoma da bananeira (*Musa* sp.) onde o retroelemento *monkey* pertencente ao tipo “*gypsy*”, para além de se encontrar disperso ao longo de todos os cromossomas foi também observado no NOR (BALINT-KURTI *et al.* 2000). Tal como para o retrotransposição *monkey*, observa-se igualmente em Fagaceae uma correlação directa entre a intensidade de sinal 45S e a intensidade do sinal obtido com as sondas *Rot*. Verificou-se ainda que sequências repetidas em tandem pertencentes à região IGS do rDNA 45S se encontram também noutras regiões cromossómicas das bananeiras (Hribova *et al.* 2007), assim como já tinha sido observado em *Nicotiana* (LIM *et al.* 2004). Este tipo de localização indica um tipo de organização genómica em larga escala presente em genomas de dimensões médias, pois tal como as bananeiras que têm em média 611Mbp/1C (BENNETT and LEITCH 2005), os genomas das *Fagaceae* estudados neste trabalho possuem entre 544Mbp/1C em *Fagus sylvatica* (GALLOIS *et al.* 1999) e 970Mbp/1C em *Q. ilex* subsp. *rotundifolia* (Teresa Ribeiro, comunicação pessoal).

## 5. Conclusões e Perspectivas Futuras

A técnica de AFLP mostrou-se eficaz para o isolamento de sequências de natureza repetitiva pois os clones obtidos são majoritariamente homólogos de vários retroelementos e a sua distribuição nos núcleos das Fagaceas confirma a sua natureza altamente repetitiva.

Apesar de termos isolado e caracterizado apenas uma pequena porção de sequências repetitivas no genoma de *Q. ilex* subsp. *rotundifolia* estas permitem-nos começar a compreender a organização em larga-escala deste genoma e perceber como esta pode ter evoluído nos outros genomas de Fagaceas, desde o género mais ancestral até ao género *Castanea*.

As sequências que não apresentam homologia nas bases de dados são fortes candidatas a serem sequências específicas desta família que poderão ser utilizadas para melhor conhecer os seus membros. A clonagem e sequenciação das bandas pertencentes às regiões ISA que foram usados em FISH poderão revelar mais acerca destas sequências. A técnica de Southern Blot também trará mais respostas acerca da organização genómica destas sequências nomeadamente, a sua proporção no genoma, e potenciais polimorfismos.

Finalmente as regiões genómicas ISA parecem ter potencial na detecção de polimorfismos presentes nas espécies vegetais.

## 6. Referências

- ANANIEV, E. V., R. L. PHILLIPS and H. W. RINES, 1998 Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin. *Genetics* **149**: 2025-2037.
- ANISIMOVA, M., and O. GASCUEL, 2006 Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* **55**: 539-552.
- AVRAMOVA, Z. V., 2002 Heterochromatin in animals and plants. Similarities and differences. *Plant Physiol* **129**: 40-49.
- BALINT-KURTI, P. J., S. K. CLENDENNEN, M. DOLEZELOVA, M. VALARIK, J. DOLEZEL *et al.*, 2000 Identification and chromosomal localization of the *monkey* retrotransposon in *Musa* sp. *Mol Gen Genet* **263**: 908-915.
- BATZER, M. A., and P. L. DEININGER, 2002 Alu repeats and human genomic diversity. *Nat Rev Genet* **3**: 370-379.
- BAUER, N., T. HORVAT, I. BIRUS, V. VICIC and V. ZOLDOS, 2009 Nucleotide sequence, structural organization and length heterogeneity of ribosomal DNA intergenic spacer in *Quercus petraea* (Matt.) Liebl. and *Q. robur* L. *Mol Genet Genomics* **281**: 207-221.
- BELMONT, A. S., S. DIETZEL, A. C. NYE, Y. G. STRUKOV and T. TUMBAR, 1999 Large-scale chromatin structure and function. *Current Opinion in Cell Biology* **11**: 307-311.
- BENNETT, M. D., and I. J. LEITCH, 2005 Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Ann Bot* **95**: 45-90.
- BENNETZEN, J. L., 2000 Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology* **42**: 251-269.
- BENNETZEN, J. L., 2005 Transposable elements, gene creation and genome rearrangement in flowering plants. *Current Opinion in Genetics & Development* **15**: 621-627.
- CHEN, G., W.-B. SUN, C.-Y. HAN and A. COOMBES, 2007 Karyomorphology of the endangered *Trigonobalanus doichangensis* (A. Camus) Forman (*Fagaceae*) and its taxonomic and biogeographical implications. *Botanical Journal of the Linnean Society* **154**: 321-330.
- CHEVENET, F., C. BRUN, A. L. BANULS, B. JACQ and R. CHRISTEN, 2006 TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**: 439.
- CHOKCHAICHAMNANKIT, P., K. ANAMTHAWAT-JONSSON and W. CHULALAKSANANUKUL, 2008 Chromosomal mapping of 18S-25S and 5S ribosomal genes on 15 species of *Fagaceae* from Northern Thailand. *Silvae Genetica* **57**: 5-13.
- CRAIG, N. L., R. CRAIGIE, M. GELLERT and A. M. LAMBOWITZ, 2002 *Mobile DNA II*. American Society for Microbiology Press, Washington, DC.
- DEREEPER, A., V. GUIGNON, G. BLANC, S. AUDIC, S. BUFFET *et al.*, 2008 Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**: W465-469.
- DOYLE, J. J., and E. E. DICKSON, 1987 Preservation of plant-samples for DNA restriction endonuclease analysis. *Taxon* **36**: 715-722.
- EDGAR, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.
- EMILIANI, G., D. PAFFETI and R. GIANNINI, 2009 Identification and molecular characterization of LTR and LINE retrotransposable elements in *Fagus sylvatica* L. *iForest* **2**: 119-126.
- FAJKUS, J., A. KOVARIK, R. KRALOVICS and M. BEZDEK, 1995 Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. *Mol Gen Genet* **247**: 633-638.
- FRANSZ, P. F., S. ARMSTRONG, J. H. DE JONG, L. D. PARNELL, C. VAN DRUNEN *et al.*, 2000 Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* **100**: 367-376.

- GALLOIS, A., M. BURRUS and S. BROWN, 1999 Evaluation of the nuclear DNA content and GCpercent in four varieties of *Fagus sylvatica* L. *Annals of Forest Science* **56**: 615-618.
- GAO, X., Y. HOU, H. EBINA, H. L. LEVIN and D. F. VOYTAS, 2008 Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res* **18**: 359-369.
- GILL, N., C. S. HANS and S. JACKSON, 2008 An overview of plant chromosome structure. *Cytogenet Genome Res* **120**: 194–201.
- GUINDON, S., and O. GASCUEL, 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696-704.
- HAVECKER, E. R., X. GAO and D. F. VOYTAS, 2004 The diversity of LTR retrotransposons. *Genome Biology* **5**: -.
- HRIBOVA, E., M. DOLEZELOVA, C. D. TOWN, J. MACAS and J. DOLEZEL, 2007 Isolation and characterization of the highly repeated fraction of the banana genome. *Cytogenet Genome Res* **119**: 268-274.
- KEJNOVSKY, E., Z. KUBAT, J. MACAS, R. HOBZA, J. MRACEK *et al.*, 2006 *Retand*: a novel family of gypsy-like retrotransposons harboring an amplified tandem repeat. *Mol Genet Genomics* **276**: 254-263.
- KREMER, A., M. CASASOLI, T. BARRENECHE, C. BODÉNÈS, P. SISCO *et al.*, 2007 5 *Fagaceae* trees. *Genome Mapping and Molecular Breeding in Plants* **7 Forest Trees**: 161-187.
- KUMEKAWA, N., E. OHTSUBO and H. OHTSUBO, 1999 Identification and phylogenetic analysis of gypsy-type retrotransposons in the plant kingdom. *Genes Genet Syst* **74**: 299-307.
- KUYKENDALL, D., J. SHAO and K. TRIMMER, 2009 A Nest of LTR Retrotransposons Adjacent the Disease Resistance-Priming Gene NPR1 in *Beta vulgaris* L. U.S. Hybrid H20. *Int J Plant Genomics* **2009**: 576742.
- LAMB, J. C., W. YU, F. HAN and J. A. BIRCHLER, 2007 Plant chromosomes from end to end: telomeres, heterochromatin and centromeres. *Curr Opin Plant Biol* **10**: 116-122.
- LATEN, H. M., A. MAJUMDAR and E. A. GAUCHER, 1998 *SIRE-1*, a copyTy1-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc. Natl. Acad. Sci. USA* **95**: 6897–6902.
- LIM, K. Y., K. SKALICKA, B. KOUKALOVA, R. A. VOLKOV, R. MATYASEK *et al.*, 2004 Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. *Genetics* **166**: 1935-1946.
- MANOS, P. S., J. J. DOYLE and K. C. NIXON, 1999 Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (*Fagaceae*). *Mol Phylogenet Evol* **12**: 333-349.
- MANOS, P. S., Z.-K. ZHOU and C. H. CANNON, 2001 Systematics of *Fagaceae*: Phylogenetic tests of reproductive trait evolution. *International Journal of Plant Sciences* **162**: 1361–1379.
- MOORE, M. J., P. S. SOLTIS, C. D. BELL, J. G. BURLEIGH and D. E. SOLTIS, 2010 Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci U S A* **107**: 4623-4628.
- NEUMANN, P., M. NOUZOVÁ and J. MACAS, 2001 Molecular and cytogenetic analysis of repetitive DNA in pea (*Pisum sativum* L.). *Genome* **44**: 716–728.
- NEVES, N., M. DELGADO, M. SILVA, A. CAPERTA, L. MORAIS-CECÍLIO *et al.*, 2005 Ribosomal DNA heterochromatin in plants. *Cytogenet Genome Res* **109**: 104–111.
- PETERSON-BURCH, B. D., D. NETTLETON and D. F. VOYTAS, 2004 Genomic neighborhoods for *Arabidopsis* retrotransposons: a role for targeted integration in the distribution of the *Metaviridae*. *Genome Biol* **5**: R78.
- REAMON-BUTTNER, S. M., T. SCHMIDT and C. JUNG, 1999 AFLPs represent highly repetitive sequences in *Asparagus officinalis* L. *Chromosome Research* **7**: 297-304.
- RIBEIRO, T., A. BARAO, W. VIEGAS and L. MORAIS-CECILIO, 2008 Molecular cytogenetics of forest trees. *Cytogenetic and Genome Research* **120**: 220-227.
- SCHWARZACHER, T., and P. HESLOP-HARRISON, 2000 *Practical in Situ Hybridization*. Bios, Oxford.

- STATON, S. E., M. C. UNGERER and R. C. MOORE, 2009 The genomic organization of Ty3 / Gypsy- like retrotransposons in *Helianthus* (*Asteraceae*) homoploid hybrid species. *American Journal of Botany* **96**: 1646–1655.
- STUART-ROGERS, C., and A. J. FLAVELL, 2001 The evolution of Ty1-copia group retrotransposons in gymnosperms. *Mol Biol Evol* **18**: 155-163.
- TAKAHASHI, H., S. OKAZAKI and H. FUJIWARA, 1997 A new family of site-specific retrotransposons, SART1, is inserted into telomeric repeats of the silkworm, *Bombyx mori*. *Nucleic Acids Res* **25**: 1578-1584.
- UNGERER, M. C., S. C. STRAKOSH and Y. ZHEN, 2006 Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr Biol* **16**: R872-873.
- VICIENT, C. M., A. SUONIEMI, K. ANAMTHAWAT-JONSSON, J. TANSKANEN, A. BEHARAV *et al.*, 1999 Retrotransposon *BARE-1* and Its Role in Genome Evolution in the Genus *Hordeum*. *Plant Cell* **11**: 1769-1784.
- VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. VANDELEE *et al.*, 1995 Aflp - a New Technique for DNA-Fingerprinting. *Nucleic Acids Research* **23**: 4407-4414.
- WEBER, B., T. WENKE, U. FROMMEL, T. SCHMIDT and T. HEITKAM, 2010 The Ty1-copia families *SALIRE* and *Cotzilla* populating the *Beta vulgaris* genome show remarkable differences in abundance, chromosomal distribution, and age. *Chromosome Res* **18**: 247-263.
- WOODCOCK, C. L., and S. DIMITROV, 2001 Higher-order structure of chromatin and chromosomes. *Current Opinion in Genetics & Development* **11**: 130–135.
- ZHANG, P., W. LI, J. FELLERS, B. FRIEBE and B. S. GILL, 2004 BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements. *Chromosoma* **112**: 288-299.
- ZOLDOS, V., D. PAPES, S. C. BROWN, O. PANAUD and S. SILJAK-YAKOVLEV, 1998 Genome size and base composition of seven *Quercus* species: inter- and intra-population variation. *Genome* **4**: 162–168.
- ZOLDOS, V., D. PAPES, M. CERBAH, O. PANAUD, V. BESENDORFER *et al.*, 1999 Molecular-cytogenetic studies of ribosomal genes and heterochromatin reveal conserved genome organization among 11 *Quercus* species. *Theor Appl Genet* **99**: 969–977.
- ZOLDOS, V., S. SILJAK-YAKOVLEV, D. PAPES, A. SARR and O. PANAUD, 2001 Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus*. *Mol Genet Genomics* **265**: 234-241.