

Universidade do Porto



Botânica Molecular Forense:  
O DNA na identificação de espécies vegetais

Susana Maria Martins de Barros Lopes

Dissertação para a obtenção do grau de Mestre em Ciências Forenses

2011

Universidade do Porto

Botânica Molecular Forense:  
O DNA na identificação de espécies vegetais

Susana Maria Martins de Barros Lopes

Dissertação para a obtenção do grau de Mestre em Ciências Forenses

Orientadora: Prof. Doutora Laura Cainé

Co-Orientadora: Prof. Doutora Maria de Fátima Pinheiro

2011

## **AGRADECIMENTOS**

À Prof. Doutora Teresa Magalhães, na qualidade de diretora da Delegação do Norte do Instituto Nacional de Medicina Legal, presto o meu agradecimento pela oportunidade que me deu para realizar o presente trabalho nesta Instituição.

À Prof. Doutora Laura Cainé, minha orientadora, agradeço-lhe o apoio e orientação, mas principalmente por ter-me lançado este desafio numa área tão auspiciosa das ciências forenses.

À Prof. Doutora Maria de Fátima Pinheiro, minha co-orientadora, agradeço-lhe todo o apoio, opiniões e sugestões.

Aos membros e Doutoranda no Serviço de Genética e Biologia Forense: Ana, David, Joana, Gabriela, Lurdes, Maria João, Paula, Raquel, Dona Amélia e Dona Maria José, agradeço-lhes pela forma como me acolheram e ajudaram ao longo deste trabalho.

Aos meus amigos e colegas, que me ajudaram e apoiaram desde o início deste trabalho.

À minha Família, por todo o apoio e paciência, o meu mais profundo agradecimento.

## RESUMO

A Botânica Forense é a aplicação do estudo das plantas ou de vestígios de origem vegetal na resolução de crimes ou outras questões legais. Embora bastante útil, continua a ser uma ciência pouco utilizada na área forense, tendo como maior aplicação a identificação de espécies e de plantas ilegais. No entanto, a identificação de espécies de plantas através da Botânica Clássica não é por vezes possível pelo facto do material botânico se encontrar bastante degradado, impedindo a sua identificação pelos métodos tradicionais. Por sua vez, a Botânica Molecular, a partir de técnicas baseadas no DNA, permite identificar quantidades vestigiais de plantas.

Neste estudo foram analisadas 38 espécies de plantas, nomeadamente plantas não vasculares (criptógamas não vasculares), plantas vasculares sem sementes (criptógamas vasculares) e com sementes (angiospérmicas e gimnospérmicas). Seguindo as recomendações do CBOL (*Consortium for the Barcode of Life*), utilizou-se um sistema "código de barras" constituído por porções de dois genes cloroplastidiais, o *matK* (maturase K) e o *rbcL* (ribulose-1,5-bifosfato carboxilase/oxigenase), para identificação das plantas terrestres.

Para além da identificação de espécies de plantas, outros objetivos deste trabalho consistiram na otimização dos métodos de extração e de amplificação por PCR (*Polymerase Chain Reaction*) do DNA das plantas, usando dois pares de *primers* igualmente recomendados pelo CBOL: o par (*3F\_KIM/1R\_KIM*), para o gene *matK* e o par (*rbcLa\_f/rbcLa\_r*), para o gene *rbcL*. Estes dois pares de *primers* em conjunto, amplificaram e sequenciaram 73,68% das plantas em estudo. Por terem sido inicialmente desenhados para as angiospérmicas, sendo este o maior grupo de plantas terrestres, este dois pares de *primers* amplificaram e sequenciaram maioritariamente estas plantas. Contudo, neste trabalho, obteve-se igualmente sucesso na amplificação e sequenciação de criptógamas, não vasculares e vascular (fetos e musgo), e em 2 gimnospérmicas com o par de *primers* (*rbcLa\_f/rbcLa\_r*). Em contraposição, o par de *primers* (*3F\_KIM/1R\_KIM*) para o gene *matK* não proporcionou resultados na amplificação e sequenciação destas plantas.

Este estudo pretende ser um contributo não só com a sua aplicação em laboratórios forenses em Portugal, mas também para colaborar na base de dados de identificação de espécies botânicas.

## ABSTRACT

Forensic Botany is the application of plant studies or plant material in solving crimes or other legal issues. Although very useful, Botanical resources remain underutilized in forensic investigations with its most common application limited to identifying specific as well as suspected illegal plants. Nevertheless, plant species identification through Classic Botany is not always possible due to the often degradation of plant material preventing its identification by traditional methods. In these cases, Molecular Botany with DNA-based techniques can identify even small traces of plants.

In this study, 38 plant species were analyzed including non-vascular plants (non-vascular cryptogams), seedless vascular plants (vascular cryptogams) and plants with seeds (gymnosperms and angiosperms). Following the recommendations of CBOL (Consortium for the Barcode of Life), a core-barcode system consisting of portions of two plastid coding regions, *matK* (maturase K) and *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase), was used for plant species identification.

As well as plant species identification, other objectives of this work were optimization of extraction methods and PCR (*Polymerase Chain Reaction*) amplification of plant DNA, using two sets of *primers* also recommended by CBOL: the *primers* (*3F\_KIM/1R\_KIM*) for the gene *matK* and the *primers* (*rbcLa\_f/rbcLa\_r*) for the gene *rbcL*. These two sets of *primers* together, amplified and sequenced 73,68% of plant species in this study. Initially designed for angiosperms, the largest group of land plants, these sets of *primers* amplified e sequenced most successfully these plants. Nevertheless, it was also amplified and sequenced non-vascular and vascular cryptogams (ferns and moss) and two gymnosperms with the *primers* (*rbcLa\_f/rbcLa\_r*). In contrast, the *primers* (*3F\_KIM/1R\_KIM*) did not provided any kind of results in amplifying and sequencing these later plants.

This study aims not only its application in forensic laboratories in Portugal, but also to collaborate in genomic databases for plant species identification.

# ÍNDICE

LISTA DE ABREVIATURAS E SÍMBOLOS .....	viii
2. INTRODUÇÃO .....	1
2.1. BOTÂNICA FORENSE MOLECULAR – BREVE RESENHA HISTÓRICA .....	2
2.2. AS PLANTAS COM INTERESSE FORENSE .....	4
2.3. OS GENES CANDIDATOS .....	7
2.4. OS GENES ESCOLHIDOS: <i>MATK</i> E <i>RBCL</i> .....	11
2.4.1. <i>MATK</i> .....	11
2.4.2. <i>RBCL</i> .....	13
3. MATERIAL E MÉTODOS .....	14
3.1. AMOSTRAS .....	14
3.2. EXTRAÇÃO DO DNA.....	14
3.2.1. <i>DNEASY® PLANT KIT</i> .....	15
3.2.2. <i>BIOROBOT® EZ1</i> .....	16
3.3. AMPLIFICAÇÃO DO DNA .....	16
3.4. SEQUENCIAÇÃO DO DNA.....	17
3.5. SEPARAÇÃO E DETECÇÃO DOS PRODUTOS AMPLIFICADOS.....	18
3.6. ANÁLISE DAS SEQUÊNCIAS DE DNA E DETERMINAÇÃO DA ESPÉCIE .....	19
4. RESULTADOS.....	20
4.1. EXTRAÇÃO E AMPLIFICAÇÃO DO DNA.....	20
4.2. IDENTIFICAÇÃO DA ESPÉCIE .....	22
4.2.1. <i>MATK</i> .....	23
4.2.2. <i>RBCL</i> .....	27
5. DISCUSSÃO .....	30

5.1.	EXTRAÇÃO DO DNA.....	30
5.2.	AMPLIFICAÇÃO E SEQUENCIAÇÃO DO DNA.....	31
5.3.	IDENTIFICAÇÃO DA ESPÉCIE .....	31
5.3.1.	<i>MATK</i> .....	32
5.3.1.1.	ANGIOSPÉRMICAS .....	32
5.3.1.2.	GIMNOSPÉRMICAS.....	34
5.3.1.3.	CRIPTÓGAMAS .....	35
5.3.2.	<i>RBCL</i> .....	35
5.3.2.1.	ANGIOSPÉRMICAS .....	35
5.3.2.2.	GIMNOSPÉRMICAS.....	36
5.3.2.3.	CRIPTÓGAMAS .....	37
5.4.	CONSIDERAÇÕES FINAIS.....	37
6.	CONCLUSÕES.....	39
7.	REFERÊNCIAS BIBLIOGRÁFICAS.....	40

## LISTA DE ABREVIATURAS E SÍMBOLOS

### LISTA DE ABREVIATURAS

*atpF* - Gene que codifica as subunidade CFO I da ATP sintase

*atpH* - Gene que codifica a subunidade CFO III da ATP sintase

BLAST - *Basic Local Alignment Search Tool*

CBOL - *Consortium for the Barcode of Life*

COI - Citocromo oxidase I

ddNTP - Didesoxinucleotídeo trifosfato

dH<sub>2</sub>O - Água desionizada

DNA - Ácido desoxirribonucleico

dNTP - Desoxirribonucleotídeo trifosfato

*et al* - E outros.

IR (a ou b) - *Large Inverted Repeats*

ITSs - Espaçadores internos transcritos

kpb - Kilo pares de bases

LSC - *Large Single Copy Region*

*matK* - Gene que codifica a proteína maturase K

mtDNA - DNA mitocondrial

NSF - *National Science Foundation*

pb - Pares de bases

PCR - *Polymerase Chain Reaction*

PlantGDB - *Plant Genomic Database*

*psbA* - Gene que codifica a proteína D1 do fotossistema II do cloroplasto

*psbI* - Gene que codifica o polipeptídeo I de massa molecular baixa do sistema de fotossistema II do cloroplasto

*psbK* - Gene que codifica o polipeptídeo K de massa molecular baixa, do sistema de fotossistema II

RAPD - *Random Amplification of Polymorphic DNA*

*rbcL* - Gene que codifica a grande subunidade da ribulose-1,5-bifosfato carboxilase/oxigenase (RuBisCo)

rDNA - DNA ribossomal

RFLPs - *Restriction Fragment Length Polymorfisms*

RNA – Ácido ribonucleico

*rpoB* - Gene que codifica a subunidade  $\beta$  da RNA polimerase

*rpoC1* - Gene que codifica a subunidade  $\beta'$  da RNA polimerase

sp - Espécie

spp - Espécies

SSC - *Small Single Copy Region*

tRNA – Ácido ribonucleico transportador

*trnH* - Gene de tRNA

*trnK* - Gene de tRNA

## LISTA DE SÍMBOLOS

$\lambda$  - Comprimento de onda

°C - Graus Celcius

% - Percentagem

h - Hora(s)

$\mu\text{L}$  - Microlitro(s)

$\mu\text{M}$  - Micromolar

mg - Miligrama(s)

mL - Mililitro(s)

min - Minuto(s)

s - Segundo(s)

## **1. JUSTIFICAÇÃO E OBJETIVOS DESTE ESTUDO**

## 1. JUSTIFICAÇÃO E OBJETIVOS DESTE ESTUDO

A Botânica Forense tem como maior aplicação a identificação de espécies e plantas ilegais. Quando o material botânico se encontra bastante degradado e sem características morfológicas, este método de identificação clássico não proporciona resultados. Nestes casos, a Botânica Molecular permite identificar quantidades vestigiais de plantas, a partir de técnicas de tipagem de DNA.

A ubiquidade das plantas torna este tipo de provas bastante útil na prática forense, mas a ausência de um sistema de identificação através do DNA torna-se um obstáculo à identificação de material vegetal. Em Portugal, a nível forense, tanto quanto se conhece, não se procede à identificação de espécies de plantas através da Botânica Molecular. Deste modo, este estudo justifica-se pela necessidade de implementação de técnicas laboratoriais e colaboração em bases de dados de espécies botânicas.

Os objetivos deste trabalho foram:

- Comparar dois métodos de extração de DNA, utilizados em Botânica Molecular;
- Analisar e avaliar um sistema "código de barras" do DNA das plantas, constituído por porções de dois genes cloroplastidiais (*matK* e *rbcL*) e respetivos pares de *primers*, recomendados pelo CBOL (*Consortium for the Barcode of Life*), utilizados para amplificar e sequenciar vários grupos taxonómicos de plantas;
- Identificar plantas, indígenas, naturalizadas ou comuns, existentes na flora portuguesa;
- Colaborar no enriquecimento de bases de dados genómicas, com introdução de sequências homólogas de genes de espécies de plantas indígenas de Portugal, nestes repositórios.

## **2. INTRODUÇÃO**

## 2. INTRODUÇÃO

A Botânica Forense é a aplicação do estudo das plantas ou de vestígios de origem vegetal na resolução de crimes ou outras questões legais, encontrando-se subdividida em várias especialidades, incluindo a anatomia de plantas, a sistemática vegetal, a palinologia, ecologia de plantas e limnologia. Na última década, a Biologia Molecular surge como uma área fundamental, ampliando e complementando a investigação nestas disciplinas (Coyle *et al.*, 2001). Vários métodos de tipagem de DNA não humano têm sido utilizados para resolver casos penais e civis. A maioria dos cientistas forenses está familiarizada com os métodos de identificação humana. Em contraposição, a utilização de plantas, animais e insetos como prova, ainda não é uma prática corrente. Isto acontece em parte por falta de conhecimento das equipas de recolha de provas, que não valorizam suficientemente os vestígios botânicos (Coyle *et al.*, 2001). No entanto, a ubiquidade das plantas torna estas provas muito úteis nas investigações criminais (Ferri *et al.*, 2009).

### 2.1. BOTÂNICA FORENSE MOLECULAR – BREVE RESENHA HISTÓRICA

Ao longo da história, as plantas foram sendo utilizadas para fins medicinais, mas igualmente em suicídios ou crimes por envenenamento. O julgamento de Sócrates e a subsequente condenação ao suicídio por *Conium* fermentado permanecem na história da Botânica, mas apenas no século XX outros tipos de provas botânicas começaram a ser aceites para efeitos legais (Bock e Norris, 1997). O primeiro caso da era moderna do uso de vestígios botânicos em tribunal foi o sequestro e morte do filho de Charles Lindbergh, em 1932. Foi utilizada uma escada de madeira para aceder ao berçário, no segundo andar, para raptar o filho de Lindbergh. Arthur Koehler, especialista em identificação de madeiras do Laboratório de Produtos Florestais dos Serviços Florestais dos Estados Unidos, em Wisconsin, recolheu provas contra Bruno Richard Hauptman, que foi mais tarde condenado pelo crime (Coyle *et al.*, 2001).

A partir deste caso muitos outros se resolveram através desta ciência. Na literatura deparamo-nos com crimes solucionados com recurso à Botânica Clássica, mas na era da Genética, a Botânica Molecular Forense surge como um instrumento de

grande importância no apoio à investigação criminal, já que, por vezes o material botânico encontra-se de tal forma degradado que não é possível a sua identificação pelos métodos tradicionais. O primeiro caso solucionado com recurso à tipagem de plantas através do DNA ocorreu igualmente nos Estados Unidos, em 1992, e ficou conhecido pelo caso Maricopa, um condado no Arizona. O corpo de uma mulher foi encontrado debaixo de uma árvore de Palo Verde, no deserto do Arizona. Próximo do corpo estava um *beeper*, eventualmente do suspeito, Mark Bogan. Bogan foi investigado e na parte de trás do seu camião foram encontradas algumas cascas de sementes de uma árvore de Palo Verde. No seguimento desta descoberta, as entidades judiciais questionaram os cientistas se através da análise de DNA era possível relacionar as cascas de sementes encontradas no camião de Bogan com as sementes de frutos da árvore, junto da qual foi descoberto o corpo. A polícia recorreu a um investigador da Universidade do Arizona, T. Helentjaris, que utilizou a técnica de *Random Amplification of Polymorphic DNA* (RAPD) para gerar um padrão de bandas das provas em questão. Pesquisou também uma pequena população de outras árvores de Palo Verde para determinar se os padrões de bandas eram únicos para cada exemplar. O seu testemunho baseado nos vestígios botânicos ajudou a condenar Mark Bogan (Coyle *et al.*, 2001).

Uma outra aplicação da Botânica Molecular Forense, e a mais utilizada, para além da identificação, é a individualização (atribuição de origem) de plantas como a *Cannabis sativa L.* (marijuana). A *Cannabis* pode ser propagada por semente ou perpetuada por meio da clonagem. Plantas propagadas por semente têm genótipos únicos, análogos aos dos seres humanos selecionados a partir de uma população aleatória. De igual forma, as plantas propagadas através da clonagem devem ter genótipos idênticos tal como os gémeos homocigóticos. O rastreamento de marijuana clonada é relativamente simples, sendo que as amostras com perfis idênticos têm uma origem genética comum (Coyle *et al.*, 2003).

A Botânica Forense é desta forma útil na identificação de espécies, aplicada ao mapeamento de um crime ou *geo-sourcing*, na análise de conteúdos estomacais ou espécies protegidas, na atribuição da origem e localização de espécies vegetais, bem como em redes de distribuição de droga. Estas provas são sem dúvida uma mais-valia

na resolução de casos forenses, sendo essencial aprofundar técnicas e procedimentos, tendo como objetivo a sua aplicação na área forense e um enriquecimento da base de dados de espécies vegetais.

## **2.2. AS PLANTAS COM INTERESSE FORENSE**

Neste capítulo será apresentado uma breve resenha sobre o reino das plantas, baseado na 2ª edição do livro Botany (Mauseth, 1998).

A vida na terra começou há cerca de 3,5 bilhões de anos. Inicialmente os organismos eram simples, como as atuais bactérias, no entanto cem milhões de células aumentaram de complexidade através da evolução por selecção natural. À medida que os organismos se tornavam mais complexos, surgiram novos avanços como a fotossíntese, que ocorreu há cerca de 2,8 milhões de anos, hoje presente em todas as plantas verdes. Um outro passo importante aconteceu quando o DNA, a molécula que transporta a informação genética, ficou confinada a um organelo – o núcleo da célula. Surgem assim as células eucariotas em oposição às procariotas, sem núcleo definido. Algumas células adquiriram organelos que produzem energia – as mitocôndrias; outras, organelos que produzem a fotossíntese e transformam a energia do sol em energia química sob a forma de hidratos de carbono – os cloroplastos. Os cloroplastos desenvolveram-se em algas e plantas. As plantas terrestres primitivas assemelhavam-se a algas, nas quais foram ocorrendo inúmeras mutações. A selecção natural eliminou as espécies menos adaptáveis, as plantas terrestres perderam as características das algas e, há cerca de 400 milhões de anos atrás, adquiriram características mais adequadas para sobreviver na terra.

Milhares de espécies surgiram, a grande maioria foi extinta, as mais aptas cresceram rapidamente, sobreviveram mais tempo e geraram mais descendência. As espécies que não foram extintas evoluíram noutras espécies, e assim sucessivamente. Existem hoje em dia mais de 400000 espécies de plantas, que não evoluíram em simultâneo, continuando a diversificar, estabelecendo a linha de evolução de plantas, o reino *Plantae*. As principais linhas de evolução deste reino foram as plantas não vasculares (criptógamas não vasculares), as vasculares simples (criptógamas

vasculares) que não produzem sementes (fetos e plantas similares) e as plantas vasculares com sementes (gimnospérmicas e angiospérmicas). Ainda não foram identificados todos os passos do sistema de classificação das plantas, nem tão-pouco todas as relações são claramente compreendidas, mas as principais características parecem ser consistentes com as dos dados existentes e espera-se que apenas pequenas alterações sejam feitas há medida que o conhecimento aumenta.

Para uma definição científica mais rigorosa, as plantas não vasculares são embriófitas (plantas que não são algas) e não possuem tecido vascular. Sendo embriófitas, têm esporângios e gametângios multicelulares: células reprodutoras rodeadas por uma ou várias camadas de células estéreis. Estas plantas, quase exclusivamente terrestres, têm cutícula em grande parte dos seus corpos e muitas delas têm estomas. Estas características ocorrem em todas as embriófitas, mas as plantas não vasculares são distintas, não possuindo nem xilema nem floema.

Não se sabe quão relacionados são os musgos, as hepáticas e antóceros; estes têm muitas características em comum, mas também diferem em muitos aspetos significativos. São comumente distribuídos em três divisões distintas: musgos, divisão Bryophyta; hepáticas, divisão Hepatophyta; e antóceros, divisão Anthocerotophyta. Esta classificação enfatiza as suas diferenças e sublinha a necessidade de se tratar cada grupo individualmente.

As plantas vasculares são tradicionalmente divididas naquelas que não produzem sementes (criptógamas vasculares) e naquelas que têm semente (espermatófitas). As criptógamas vasculares surgiram em primeiro lugar, e alguns dos seus membros foram mais tarde os ancestrais das plantas com sementes. Estas plantas apareceram pela primeira vez há cerca de 420 milhões de anos atrás, tendo evoluído a partir de algas verdes ancestrais. A evolução do tecido vascular teve profundas consequências: plantas relativamente fortes, com hastes verticais que poderiam ser apoiadas, grandes corpos de plantas que poderiam ser integrados e algumas partes das plantas que poderiam diferenciar-se e especializar-se. Os primeiros gametófitos foram os mais elaborados e todas as subsequentes evoluções do gametófito desenvolveram-se no sentido da simplificação e redução.

As criptógamas vasculares são embriófitas, não algas, por causa dos seus esporângios e gametângios multicelulares. Como possuem tecido vascular, não são musgos, nem hepáticas ou antóceros, e possuem uma divisão extensa e complexa. Neste trabalho será dado ênfase a um grupo de criptógamas vasculares, os fetos, pertencentes à divisão Pteridophyta.

Nas plantas vasculares com sementes, as gimnospérmicas são as plantas com "óvulos nus", isto é, não estão inseridos em carpelos. As angiospérmicas são as plantas com flores, com carpelos, com esporófilo feminino, formando uma estrutura tubular fechada, sendo a evolução da flor o acontecimento mais significativo neste grupo.

Estudos em fósseis indicam que as plantas com sementes são um grupo muito mais antigo do que se pensava anteriormente e muitos desenvolvimentos evolutivos foram ocorrendo: a origem das sementes e cones, a origem de um câmbio vascular e crescimento secundário, a origem de vasos e a origem de muitos compostos químicos de defesa. Os diferentes tipos de gimnospérmicas são bem mais distintos do que se esperava e algumas devem ter evoluído a partir das plantas ancestrais com flores. Conseqüentemente, muitos botânicos acreditam que a velha divisão "Spermatophyta" e classes "Gimnospérmicas" e "Angiospérmicas" deveria ser abandonada e cada grupo deveria ser classificado como divisão. Outros botânicos, têm uma opinião contrária; como as várias linhas evolutivas são relativamente próximas, formando um grupo natural, devem manter-se apenas numa divisão. De qualquer maneira, o conceito de "gimnospérmica" ainda é muito útil e é constantemente utilizado, pelo menos informalmente, e a palavra "angiospérmica" ainda é usada para se referir a plantas com flor.

As divisões de plantas com semente comumente aceites e usadas são as seguintes: (1) Coniferophyta, (2) Cycadophyta, (3) Ginkgophyta, (4) Gnetophyta (estes quatro são gimnospérmicas) e (5) Magnoliophyta (as plantas com flores).

Como foi acima referido, a Biologia é atualmente constituída por cinco reinos, existindo cinco principais grupos de organismos: plantas, animais, fungos, protistas e procariontes. Entre os eucariontes, as plantas, os animais e os fungos são bastante

distintos e fáceis de reconhecer. Os vários organismos desenvolveram-se em domínios distintos. O nível fundamental da classificação é a espécie que, idealmente, e teoricamente, é um conjunto de indivíduos ligados por um ancestral comum. Espécies relacionadas são agrupadas em géneros. O nível acima do género é o da família. Os níveis acima da família são a classe, a ordem, a divisão, e o reino.

O objetivo deste estudo será identificar plantas preferencialmente ao nível da espécie, a partir de técnicas de sequenciação de DNA.

### **2.3. OS GENES CANDIDATOS**

Mendel nos seus estudos sobre plantas, nomeadamente o cruzamento entre espécies e, conseqüentemente, as leis da hereditariedade trouxe mudanças na Biologia. Mas, desde o longínquo século XIX um longo caminho foi percorrido, com inúmeros progressos neste novo ramo da Biologia, a Genética. Antes de iniciarmos o estudo e metodologias de identificação de espécies de plantas através da Genética, ou melhor dizendo, da Botânica Molecular, é importante explicar o porquê dos genes ou marcadores moleculares escolhidos para este propósito. Segue-se uma resenha histórica sobre os genes candidatos, escolhidos para a identificação das espécies.

Ao longo do século XX, vários genes foram estudados, na tentativa de identificar espécies; mas, para as plantas, a identificação de uma região consenso adequada tem sido problemática. O gene citocromo oxidase I (COI) do DNA mitocondrial (mtDNA) é excelente na identificação de várias espécies animais. No entanto, nas plantas tal não acontece, estudando-se outras regiões alternativas no genoma das plantas. De facto, o DNA mitocondrial evolui muito lentamente nas plantas para proporcionar uma região suficientemente variável, permitindo a discriminação entre as espécies (Cho *et al.*, 2004). Assim, a busca desta região ou “código de barras” de DNA está ainda em curso. O “código de barras” de DNA é uma técnica nova que usa sequências de DNA padrão, 400-800 pares de bases (pb), de um organismo, para determinar a sua identidade (Lahaye *et al.*, 2008a). Portanto, foi necessária a procura de sequências de genes adequados para a identificação de espécies de plantas. Com base em informação disponível a partir de estudos filogenéticos, foram estudados para a sua adequação

como "códigos de barras", sequências de genes, codificantes e não-codificantes, de DNA do cloroplasto juntamente com um gene do DNA nuclear (Vijayan e Tsou, 2010).

O "código de barras" das plantas, baseado num segmento de DNA nuclear herdado de forma bi-parental, que supostamente deveria fornecer mais informação sobre a identidade da espécie, até à data corresponde apenas à região dos espaçadores internos transcritos (ITSs) do DNA ribossomal (rDNA) (Kress *et al.*, 2005). A dificuldade na obtenção de universalidade na amplificação por PCR (*Polymerase Chain Reaction*) de genes, especialmente a partir de DNA degradado, com baixa qualidade e o baixo poder de discriminação de espécies, devido à conservação de genes funcionais em inúmeras linhagens, podem ser as razões principais do número limitado do estudo de tais genes (Vijayan e Tsou, 2010). No entanto, os ITSs do rDNA representam uma ferramenta suplementar no "código de barras" em grupos taxonómicos, em que a sequenciação deste *locus* é possível.

O genoma do cloroplasto partilha alguns atributos do genoma mitocondrial, tais como um grande número de cópias por célula, possibilidade de uso da PCR e a existência de *primers* universais. Daí, os genes cloroplastidiais serem considerados análogos aos genes mitocondriais e serem utilizados para identificação de espécies de plantas. No entanto, em comparação com genes do mtDNA em animais, os genes dos cloroplastos nas plantas tem um ritmo de evolução mais lento; portanto, encontrar sequências de genes adequados, com suficiente poder de discriminação constitui ainda um grande desafio (Wolfe *et al.*, 1987).

O genoma do cloroplasto das plantas superiores é uma estrutura circular com um tamanho de 120-160 kilo pares de bases (kpb). A arquitetura geral do genoma do cloroplasto é representada por duas regiões a *Large Single Copy Region* (LSC) e a *Small Single Copy Region* (SSC) interrompidas por duas cópias de *Large Inverted Repeats* (IRa e IRb) (Figura 1). O genoma do cloroplasto contém todos os genes de rRNA (quatro genes em plantas superiores), os genes de RNA transportador (tRNA) (35 genes) e outros genes para a síntese de proteínas no cloroplasto (~100 genes), essenciais à sua existência (Shinozaki *et al.*, 1986).



candidatas: *matK*, *rbcl*, *rpoB*, *rpoC1*, *atpF-atpH*, *trnH-psbA* e *psbI-psbK* (Figura 1). Os genes *matK*, *rbcl*, *rpoB*, *rpoC1* codificam para proteínas. O gene *matK* codifica a proteína maturase K; o gene *rbcl* codifica a subunidade grande da rubilose-1,5-bifosfato carboxilase/oxigenase; o gene *rpoB* codifica a subunidade  $\beta$  da enzima RNA polimerase do cloroplasto; e o gene *rpoC1* codifica a subunidade  $\beta'$  da enzima RNA polimerase do cloroplasto. As regiões *atpF-atpH*, *trnH-psbA* e *psbI-psbK* são espaçadores intergênicos. Os genes *atpF* e *atpH* codificam as subunidades CFO I e CFO III da ATP sintase, respectivamente. Os genes *psbK* e *psbI* codificam dois polipeptídeos de massa molecular baixa, K e I, respectivamente, do fotossistema II. Estas duas regiões são conservadas desde as algas até às plantas terrestres e parasitárias (Lahaye *et al.*, 2008a). O gene *trnH* é um gene de tRNA e *psbA* codifica a proteína D1 do fotossistema II. Todos estes dez genes são cloroplastidiais e localizam-se na região *Large Single Copy* (LSC) do genoma do cloroplasto.

Em 2005, em Londres, na primeira conferência internacional do CBOL, Kress e colaboradores sugeriram a utilização do espaçador intergênico *trnH-psbA* para identificar as espécies vegetais de modo rápido e preciso, justificando a escolha por este ser mais variável que os genes e, conseqüentemente, com melhores resultados para a identificação. Esta proposta foi contestada por outros grupos de trabalho, exigindo um estudo mais sistemático sobre o assunto. Nesse mesmo ano, novamente Kress e colaboradores, publicaram um artigo onde referiam que idealmente seria preferível uma região "código barras" com apenas um *locus*; no entanto, desde plantas com flores a musgos, bem como hepáticas, existe muita variação. A solução adequada seria uma região "código de barras" com dois *loci*: o espaçador intergênico *trnH-psbA* e parte de um gene chamado *rbcl* (Pennisi, 2007). Nos meses que se seguiram surgiram outras propostas que incluíam três *loci* (Chase *et al.*, 2007; K. J. Kim *et al.*, não publicado). Apesar das inúmeras investigações em curso não foi possível chegar a um consenso, sendo "este propósito considerado mais longo que o previsto e certamente mais longo que o esperado" (Pennisi, 2007).

Na segunda conferência internacional do CBOL, em Tapei, em setembro 2007, foram propostos pelo menos cinco "códigos de barras" de DNA diferentes, mas os

investigadores, novamente, não conseguiram chegar a um consenso (Lahaye *et al.*, 2008a).

Recentemente, em 2009, o CBOL analisou estudos de vários grupos de trabalho e decidiu optar por duas propostas de regiões "código de barras". A primeira proposta contemplava uma região "código de barras" com dois *loci* (*matK* e *rbcl*) e uma segunda proposta constituída por uma região "código de barras" com três *loci* (*rbcl*, *matK*, e *trnH-psbA*). O comité executivo do CBOL, após ponderar as vantagens e desvantagens das duas propostas, recomendou dois genes cloroplastidiais, o gene *matK* e o gene *rbcl*, como sistema "código de barras", que identificam cerca de 70-75% das espécies. No entanto, o comité decidiu avaliar novamente esta proposta num prazo de 18 meses, pois considerou ser necessário a introdução de melhoramentos nesta estratégia, em conjunto com os novos dados para a caracterização do *matK* e do *rbcl*, bem como com os progressos obtidos com *primers* propostos do gene *matK* e técnicas de sequenciação para regiões não-codificantes, como a *trnH-psbA*.

## 2.4. OS GENES ESCOLHIDOS: *MATK* E *RBCL*

Neste estudo de identificação de espécies de plantas foram escolhidos os genes recomendados pelo CBOL. Segue uma breve análise destes genes.

### 2.4.1. *MATK*

Entre os genes cloroplastidiais, o *matK* é um dos genes candidatos mais promissores a "código de barras" nas plantas (Yu *et al.*, 2011). Tem um tamanho de cerca de 1550 pb e codifica a enzima maturase K, que está envolvida no *splicing* de intrões tipo-II de RNA transcritos (Neuhaus e Link, 1987). O *matK* tem sido utilizado como um marcador para a construção de filogenias de plantas, devido à sua rápida evolução e à sua presença ubíqua nas plantas (Hilu e Liang, 1997). Sendo uma região codificante, o *matK* é bastante útil em reconstruções filogenéticas a níveis taxonómicos elevados, como a ordem ou a família, mas igualmente a níveis taxonómicos mais baixos, como o género ou a espécie (Chase *et al.*, 2007; Lahaye *et al.*, 2008b). No entanto, a falha de amplificação por PCR do *matK* em alguns grupos

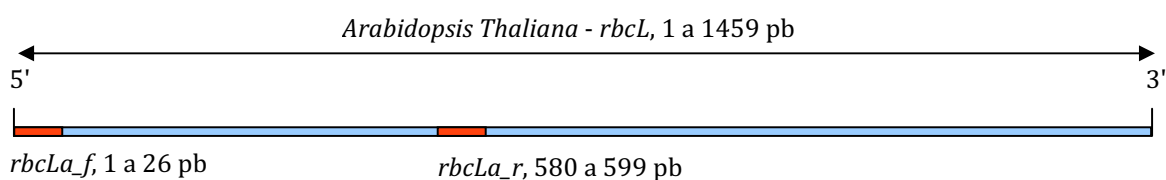


### 2.4.2. *RBCL*

Entre os genes dos plastídeos, o gene *rbcl* é o mais estudado e, por isso, a maioria dos grupos de investigação testou a sua adequação ao "código de barras" das plantas. Este gene está localizado na *Large Single Copy Region* do cloroplasto e codifica a subunidade grande da rubilose-1,5-bifosfato carboxilase/oxigenase (RuBisCo). Como a RuBisCo é uma enzima fotossintética muito importante, o *rbcl* foi o primeiro gene das plantas a ser sequenciado (Zurawski *et al.*, 1981). Este gene tem sido usado amplamente em estudos filogenéticos de plantas com mais de 10000 sequências do *rbcl*, já disponíveis no *GenBank* (Chase *et al.*, 2007).

Ao nível da família e nos níveis acima, o gene *rbcl* é de longe o preferido para inferir relações de filogenia, tanto em angiospérmicas como em fetos. De facto, a maioria dos estudos filogenéticos sugerem que o gene *rbcl* é o mais adequado para reconstruir as relações até ao nível de género, mas não é útil para níveis específicos (Soltis e Soltis, 1998). A fim de se obter a discriminação de espécies, é necessário sequenciar os cerca de 1430 pb, o que é um fator limitante do seu uso, porque uma região "código de barras" de DNA ideal deve ser suficientemente curta para amplificar DNA degradado, analisado através de uma sequenciação única (Chase *et al.*, 2007).

A secção "código de barras" proposta para o *rbcl* (Chase *et al.*, 2007) corresponde à região 5' final do gene indicado na Figura 3. Na sequência da *Arabidopsis thaliana*, a região está compreendida entre 1-599 pb (27-579 excluindo a sequência do *primer*). Neste estudo, o conjunto de *primers* utilizados para amplificar o gene da RuBisCo é o *rbcl*<sub>a</sub> (*forward* e *reverse*), proposto por Kress e Erickson (2007) e igualmente recomendado pelo CBOL.



**Figura 3** – Região "código de barras" do gene *rbcl* na sequência da *Arabidopsis thaliana*.

### **3. MATERIAIS E MÉTODOS**

### 3. MATERIAL E MÉTODOS

#### 3.1. AMOSTRAS

Foi colhido material botânico (folhas frescas, maioritariamente) de 38 plantas, nomeadamente, de criptógamas não vasculares, plantas vasculares sem semente (criptógamas vasculares) e com semente (gimnospérmicas e angiospérmicas), na zona norte de Portugal, entre os meses de maio e outubro. As plantas escolhidas para este estudo foram algumas consideradas indígenas, naturalizadas ou comuns na flora portuguesa (Decreto-Lei 565/99; <http://www.afn.min-agricultura.pt/portal/gestao-florestal/gfs/eafu-pt>), como o carvalho (*Quercus robur*), o sobreiro (*Quercus suber*), o castanheiro (*Castanea sativa*), o medronheiro (*Arbutus unedo*), o pinheiro (*Pinus pinaster*), o teixo (*Taxus baccata*), a oliveira (*Olea europaea*), o eucalipto (*Eucalyptus globulus*), o loureiro (*Laurus nobilis*), o alecrim (*Rosmarinus officinalis*) e a hera (*Hedera helix*). Por se tratar de uma época do ano com grande floração, frutificação e maturação dos frutos, também se escolheram várias folhas de árvores de fruto, arbustivas ou trepadeiras como a macieira (*Malus domestica*), a cerejeira (*Prunus avium*), o damasqueiro (*Prunus armeniaca*), o morangueiro (*Fragaria vesca*), a videira (*Vitis vinifera*), a cameleira (*Camellia japonica*), a roseira (*Rosa* sp), entre outras. Também houve o cuidado de colher criptógamas vasculares como os fetos (*Nephrolepis* spp) e uma criptógama não vascular, um musgo (*Hypnum cupressiforme*).

O material botânico, após a colheita, foi devidamente identificado, etiquetado e guardado em envelopes de papel a -20° C, até à extração do DNA.

#### 3.2. EXTRAÇÃO DO DNA

Cada amostra de material botânico foi retirada dos envelopes e seccionada com uma lâmina de bisturi descartável. De seguida, procedeu-se à sua trituração com a ajuda de um pilão e almofariz. Por fim, pesou-se 5-10 mg de planta triturada e transferiu-se para um tubo *Eppendorf* de 1,5 mL.

O DNA foi extraído através da utilização de diferentes processos, de modo a possibilitar a comparação entre métodos.

A identificação da espécie botânica foi realizada em 50 amostras, que foram submetidas aos métodos de extração com o *DNeasy® Plant Kit* (Qiagen) e usando o *BioRobot® EZ1* (Qiagen).

### 3.2.1. *DNEASY® PLANT KIT*

A extração utilizando o *DNeasy® Plant Kit* proporciona uma extração purificada e total de DNA (genômico, mitocondrial e cloroplastidial), sem recorrer à extração orgânica. O material botânico é inicialmente triturado e lisado, pela adição de um tampão de lise. Após a lise celular, as proteínas e polissacarídeos são precipitados e removidos por centrifugação. Ao lisado limpo é adicionado um tampão com etanol, permitindo a ligação do DNA a uma membrana de sílica-gel em coluna e a remoção de contaminantes, proteínas e polissacarídeos através de dois processos de lavagem. Por fim, o DNA purificado é eluído.

O protocolo de extração foi seguido de acordo com o manual *DNeasy® Plant Kit* (Qiagen, 2006), com apenas uma alteração inicial. Após o tecido ter sido triturado com a ajuda de um pilão e almofariz, foi incubado durante 12-18 h, a 56°C, em 400 µL do tampão de lise fornecido e 30 µL de proteinase K (10 mg/mL) (Stoeckle *et al.*, 2011). Os restantes passos foram efetuados de acordo com o protocolo do fabricante (Figura 4).



**Figura 4-** Amostras de plantas (cerca de 5 mg) maceradas com pilão e almofariz, em *Eppendorfs* de 1,5 mL, após digestão no tampão de lise e proteinase K, durante 12-18 h, a 56°C.

### 3.2.2. *BIOROBOT*<sup>®</sup> *EZ1*

Este aparelho efectua de modo automático a extração e purificação de ácidos nucleicos. A extração dos ácidos nucleicos é conseguida através de partículas magnéticas de sílica que se ligam ao DNA, em amostras devidamente preparadas e lisadas. Estas partículas, são depois separadas por uma fonte magnética, sofrendo posteriormente um processo de lavagem. Na fase final, o DNA é eluído numa solução de baixa força iónica (Cainé, 2010).

A extração foi realizada, em geral de acordo com as indicações do manual *EZ1*<sup>®</sup> *DNA Tissue Handbook* (Qiagen, 2010a), referentes ao protocolo de purificação do DNA a partir de tecidos, utilizando o cartão pré-programado *EZ1 DNA Forensic Card* e seleccionando o volume final de eluição de 50 µL.

O protocolo sofreu apenas uma pequena alteração inicial: após o tecido ter sido triturado com a ajuda de um pilão e almofariz, foi incubado durante 12-18 h a 56°C, em 190 µL do tampão de lise fornecido e 10 µL de proteinase K (10 mg/mL).

### 3.3. AMPLIFICAÇÃO DO DNA

A amplificação de sequências específicas de DNA correspondentes ao genes mencionados, foi efectuada pelo método de reação em cadeia da polimerase, PCR (Mullis, 1990).

Todas as reações de amplificação decorreram num termociclador *Veriti*<sup>®</sup> (Applied Biosystems) e foram realizadas na sala de pré-PCR.

Para as amplificações individuais de cada gene cloroplastidial, utilizou-se na reação de PCR um volume final de 20 µL, com 10 µL *HotStar Taq Plus Master Mix Kit* (Qiagen, 2010b), 1 µL DNA extraído e 1 µM (para o *matK*) ou 0,5 µM (para o *rbcl*) de cada *primer* (*forward* e *reverse*).

Na amplificação da região "código de barras" do gene *matK*, utilizaram-se os *primers* *3F\_KIM* (5'-CGT-ACA-GTA-CTT-TTG-TGT-TTA-CGA-G-3') e *1R\_KIM* (5'-ACC-CAG-TCC-ATC-TGG-AAA-TCT-TGG-TTC-3') e para a região "código de barras" do gene

*rbcl*, utilizaram-se os *primers* *rbcla\_f* (5'-ATG-TCA-CCA-CAA-ACA-GAG-ACT-AAA-GC-3') e *rbcla\_r* (5'-GTA-AAA-TCA-AGT-CCA-CCR-CG-3').

As condições da PCR usadas tiveram como referência os vários protocolos do CBOL e consistiram numa desnaturação inicial durante 5 min a 95°C; 35 ciclos: desnaturação (30 s a 94°C), *annealing* (1 min a 52°C para o *matK* e a 55°C para o *rbcl*), extensão (1min a 72°C); e extensão final a 72°C, durante 10 min.

Para os fragmentos de DNA que não amplificaram, as temperaturas de *annealing* experimentais, para ambos os pares de *primers*, variaram entre 48°C e 60°C, e número de ciclos entre 35 e 40.

De seguida, os produtos da PCR foram visualizados em gel de agarose a 2% (SeaKem®GTG®, FMC BioProducts), ao qual foi adicionado 6 µL de brometo de etídio (Sigma). De seguida, a cada 3 µL de produto amplificado foi adicionado 3 µL de azul de bromofenol (Amresco). Paralelamente, foi utilizado um marcador de peso molecular de 123 pb (Invitrogen). O gel foi visualizado sobre luz ultra-violeta num transiluminador (LKB, Bromma).

Todos os produtos de amplificação com tamanho correto foram purificados com ExoSAP (*exonuclease I and shrimp-alkaline phosphatase*, USB Corporation) e sequenciados em ambas as direções (*forward* e *reverse*) com o BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems).

### 3.4. SEQUENCIAÇÃO DO DNA

A sequenciação de DNA é um processo que permite identificar a ordem dos nucleotídeos numa amostra. Um dos métodos mais utilizados consiste em identificar continua e sequencialmente o último nucleotídeo incorporado na extremidade de alongamento da cadeia. A reação de síntese processa-se em condições iónicas e de pH apropriadas, na presença da enzima DNA polimerase, de dNTPs, de um único *primer* (*forward* ou *reverse*) e de ddNTPs (os terminadores), marcados com fluorescência. Os ddNTPs são moléculas semelhantes aos dNTPs, aos quais falta o grupo OH no carbono 3' da desoxirribose, impedindo, deste modo, a ligação de outro dNTP e,

consequentemente, o alongamento da cadeia de DNA. No final são produzidas várias cadeias simples de DNA com terminadores incorporados, com diferentes tamanhos, que constituem a sequência alvo.

Deste modo, procedeu-se à sequenciação direta com ABI PRISM® BigDye® Terminator v1.1 Cycle Sequencing Kit, de acordo com o protocolo do fabricante (Applied Biosystems, 2002). O volume final de 10  $\mu\text{L}$  consistiu em: 5,75  $\mu\text{L}$  água desionizada ( $\text{dH}_2\text{O}$ ) autoclavada, 1,5  $\mu\text{L}$  *Reaction Mix*, 0,75  $\mu\text{L}$  tampão de sequenciação, 1  $\mu\text{L}$  de *primer*, à concentração de 3,2  $\mu\text{M}$  e 1  $\mu\text{L}$  produto amplificado purificado.

A reação de sequenciação foi efetuada para ambas as cadeias, *forward* e *reverse*, tendo sido utilizados, separadamente, os mesmos *primers* da PCR. Os terminadores não incorporados e os excessos de *primer* e tampão foram diretamente removidos da reação de sequenciação com DyeEx™ 2.0 Spin Kit (Qiagen). Esta remoção foi efetuada através de um gel de filtração, uma resina pré-hidratada que separa as moléculas de acordo com o seu peso molecular. O procedimento foi realizado de acordo com o manual de utilização DyeEx™ Handbook (Qiagen, 2002) e com o protocolo referente a DyeE 2.0 Spin Protocol.

### 3.5. SEPARAÇÃO E DETEÇÃO DOS PRODUTOS AMPLIFICADOS

Durante a eletroforese capilar com deteção automática por fluorescência, após uma excitação com um laser de árgon, cada um dos fluorocromos emite o seu máximo de fluorescência a um comprimento de onda ( $\lambda$ ) diferente, detetado depois pelo sequenciador automático. Durante a recolha de dados, os sinais fluorescentes são separados por difração de acordo com o comprimento de onda (Cainé, 2010).

A separação dos produtos amplificados e a sua deteção decorreu no sequenciador automático ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems) de acordo com os manuais do fabricante (Applied Biosystems, 2001). Os dados obtidos foram automaticamente recolhidos pelo ABI PRISM® Data Collection Software v.3.0.0.

### 3.6. ANÁLISE DAS SEQUÊNCIAS DE DNA E DETERMINAÇÃO DA ESPÉCIE

Após a análise automática das sequências, os eletroforetogramas obtidos devem ser confirmados, sendo por vezes necessário serem corrigidos e os respectivos ficheiros em formato FASTA (utilizado para representar sequências) alinhados manualmente, para permitir a posterior identificação da espécie. Para a identificação da espécie, o formato FASTA da sequência a analisar será submetido a uma base de dados de sequências genómicas de plantas disponível *online*, a partir do PlantGDB, pertencente ao NSF (*National Science Foundation*). O PlantGDB utiliza o motor de busca BLAST (*Basic Local Alignment Search Tool*), que estabelece homologias entre sequências (<http://www.plantgdb.org/cgi-bin/blast/PlantGDBblast>). No final, os resultados serão apresentados em valores percentuais, por ordem decrescente de grau de similaridade, ou seja, por ordem do valor máximo de identidade (*Maximum Identity Value*) da sequência ou sequências homólogas (sequências das diversas espécies já identificadas) com a sequência submetida.

## **4. RESULTADOS**

## 4. RESULTADOS

### 4.1. EXTRAÇÃO E AMPLIFICAÇÃO DO DNA

Foram estudadas 38 espécies de plantas, tendo sido selecionadas 50 amostras para extração com o *DNeasy® Plant Kit* e no *BioRobot® EZ1*. Destas amostras, 10 foram extraídas com o *DNeasy® Plant Kit* e 16 foram extraídas no *BioRobot® EZ1*. De 12 espécies de plantas foram ainda recolhidas 2 amostras e cada um dos grupos foi extraído por cada um dos métodos, de forma a poder compará-los. Nas 28 amostras extraídas no *BioRobot® EZ1* foram obtidas 21 sequências de DNA cloroplastidial (Tabelas I e II); enquanto, nas 22 amostras extraídas com o *DNeasy® Plant Kit* foram obtidas apenas 9 sequências de DNA cloroplastidial (Tabelas I e III).

**Tabela I** - Resultados obtidos (sequência de DNA cloroplastidial) segundo os diferentes métodos de extração efetuados em amostras de espécies de plantas.

Método de extração	Amostras estudadas	Obtenção de DNA
<i>BioRobot® EZ1</i>	28	21
<i>DNeasy® Plant Kit</i>	22	9

**Tabela II** - Resultados obtidos (sequência de DNA cloroplastidial) com o método de extração *BioRobot® EZ1* efetuado em 28 amostras de espécies de plantas.

Família	Espécie	Nome comum	Extração <i>BioRobot® EZ1</i>	DNA
<b>Angiospérmicas</b>				
<i>Ericaceae</i>	<i>Arbutus Unedo</i>	Medronheiro	Sim	Sim
<i>Myrtaceae</i>	<i>Eucalyptus globulus</i>	Eucalipto	Sim	Sim
<i>Fagaceae</i>	<i>Castanea sativa</i>	Castanheiro	Sim	Não
<i>Ebenaceae</i>	<i>Diospyros kaki</i>	Diospireiro	Sim	Sim
<i>Rosaceae</i>	<i>Fragaria vesca</i>	Morangeiro	Sim	Sim
<i>Araliaceae</i>	<i>Hedera helix</i>	Hera	Sim	Sim
<i>Asteraceae</i>	<i>Hypochaeris radicata</i>	Leituga	Sim	Sim
<i>Aquifoliaceae</i>	<i>Ilex aquifolium</i>	Azevinho	Sim	Não
<i>Asteraceae</i>	<i>Lactuca sativa</i>	Alface	Sim	Sim
<i>Lauraceae</i>	<i>Laurus Nobilis</i>	Loureiro	Sim	Sim
<i>Rosaceae</i>	<i>Malus domestica</i>	Macieira	Sim	Sim

**Tabela II (cont.)** - Resultados obtidos (sequência de DNA cloroplastidial) com o método de extração *BioRobot® EZ1* efetuado em 28 amostras de espécies de plantas.

Família	Espécie	Nome comum	Extração <i>BioRobot® EZ1</i>	DNA
<b>Angiospérmicas</b>				
<i>Lamiaceae</i>	<i>Mentha spicata</i>	Hortelã	Sim	Não
<i>Rosaceae</i>	<i>Prunus armeniaca</i>	Damasqueiro	Sim	Sim
<i>Rosaceae</i>	<i>Prunus avium</i>	Cerejeira	Sim	Sim
<i>Rosaceae</i>	<i>Prunus domestica</i>	Ameixeira	Sim	Não
<i>Rosaceae</i>	<i>Pyrus communis</i>	Pereira	Sim	Sim
<i>Lamiaceae</i>	<i>Rosmarinus officinalis</i>	Alecrim	Sim	Não
<i>Solanaceae</i>	<i>Solanum lycopersicum</i>	Tomateiro	Sim	Sim
<i>Amaranthaceae</i>	<i>Spinacia oleracea</i>	Espinafre	Sim	Não
<i>Fagaceae</i>	<i>Quercus robur</i>	Carvalho	Sim	Sim
<i>Fagaceae</i>	<i>Quercus suber</i>	Sobreiro	Sim	Sim
<i>Vitaceae</i>	<i>Vitis vinífera</i>	Videira	Sim	Sim
<b>Gimnospérmicas</b>				
<i>Cupressaceae</i>	<i>Chamaecyparis lawsoniana</i>	Camecíparris	Sim	Sim
<i>Pinaceae</i>	<i>Pinus pinaster</i>	Pinheiro	Sim	Não
<i>Taxaceae</i>	<i>Taxus baccata</i>	Teixo	Sim	Sim
<b>Criptógamas vasculares</b>				
<i>Lomariopsidaceae</i>	<i>Nephrolepis cordifolia</i>	Feto	Sim	Sim
<i>Lomariopsidaceae</i>	<i>Nephrolepis exaltata</i>	Feto de Boston	Sim	Sim
<b>Criptógamas não vasculares</b>				
<i>Hypnaceae</i>	<i>Hypnum cupressiforme</i>	Musgo	Sim	Sim

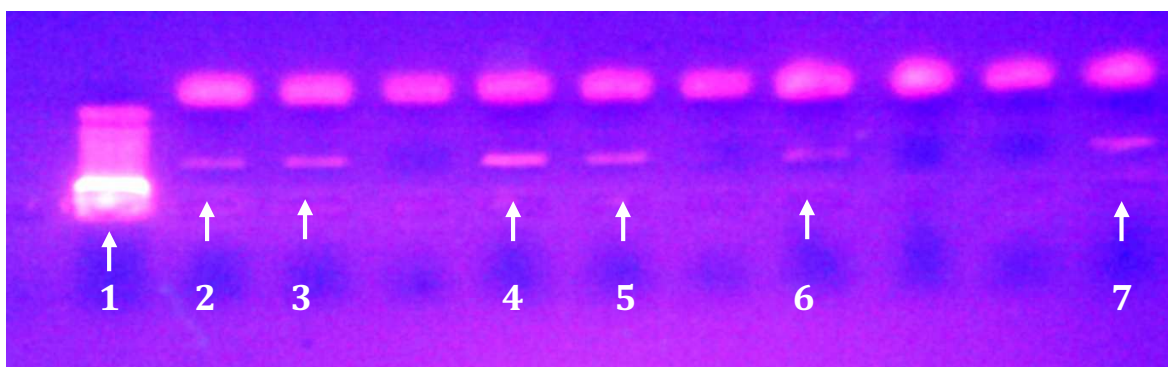
**Tabela III** - Resultados obtidos (sequência de DNA cloroplastidial) com o método de extração *DNeasy® Plant Kit* efetuado em 22 amostras de espécies de plantas.

Família	Espécie	Nome comum	Extração <i>DNeasy® Plant Kit</i>	DNA
<b>Angiospérmicas</b>				
<i>Xanthorrhoeaceae</i>	<i>Aloe Vera</i>	Aloe Vera	Sim	Sim
<i>Ericaceae</i>	<i>Arbutus Unedo</i>	Medronheiro	Sim	Não
<i>Brassicaceae</i>	<i>Brassica oleracea</i>	Couve	Sim	Não
<i>Theaceae</i>	<i>Camellia japonica</i>	Cameleira	Sim	Sim
<i>Fagaceae</i>	<i>Castanea sativa</i>	Castanheiro	Sim	Não
<i>Rutaceae</i>	<i>Citrus limon</i>	Limoeiro	Sim	Não
<i>Rutaceae</i>	<i>Citrus reticulata</i>	Tangerineira	Sim	Sim
<i>Rutaceae</i>	<i>Citrus sinensis</i>	Laranjeira	Sim	Sim
<i>Ebenaceae</i>	<i>Diospyros kaki</i>	Diospireiro	Sim	Não
<i>Rosaceae</i>	<i>Fragaria vesca</i>	Morangeiro	Sim	Não
<i>Asteraceae</i>	<i>Hypochaeris radicata</i>	Leituga	Sim	Não
<i>Lauraceae</i>	<i>Laurus Nobilis</i>	Loureiro	Sim	Sim
<i>Rosaceae</i>	<i>Malus domestica</i>	Macieira	Sim	Sim
<i>Lamiaceae</i>	<i>Mentha spicata</i>	Hortelã	Sim	Não

**Tabela III (cont.)** - Resultados obtidos (sequência de DNA cloroplastidial) com o método de extração *DNeasy® Plant Kit* efetuado em 22 amostras de espécies de plantas.

Família	Espécie	Nome comum	Extração <i>DNeasy® Plant Kit</i>	DNA
<b>Angiospérmicas</b>				
<i>Oleaceae</i>	<i>Olea Europaea</i>	Oliveira	Sim	Sim
<i>Passifloraceae</i>	<i>Passiflora edulis</i>	Maracujazeiro	Sim	Não
<i>Rosaceae</i>	<i>Prunus armeniaca</i>	Damasqueiro	Sim	Sim
<i>Rosaceae</i>	<i>Prunus avium</i>	Cerejeira	Sim	Não
<i>Rosaceae</i>	<i>Pyrus communis</i>	Pereira	Sim	Não
<i>Solanaceae</i>	<i>Solanum melongena</i>	Beringela	Sim	Não
<i>Amaranthaceae</i>	<i>Spinacia oleracea</i>	Espinafre	Sim	Não
<i>Rosaceae</i>	<i>Rosa sp</i>	Roseira	Sim	Sim

Após a amplificação, os produtos da PCR foram visualizados em gel de agarose a 2%, ao qual se adicionou brometo de etídio, de forma a confirmar se houve ou não sucesso na amplificação. A Figura 5 exemplifica este procedimento com uma fotografia de um gel de produtos amplificados, cujo DNA foi extraído no *BioRobot®EZ1* e com o *DNeasy® Plant Kit*.



**Figura 5** - Gel e produtos amplificados, cujo DNA foi extraído no *BioRobot®EZ1* (2; 3; 4 e 5) e com o *DNeasy® Plant Kit* (6 e 7): 1 - Ladder de 123 pb; 2 - *Malus domestica*; 3- *Arbutus unedo*; 4 - *Laurus nobilis*; 5 - *Prunus armeniaca*; 6 - *Malus domestica* e 7 - *Laurus Nobilis*.

#### 4.2. IDENTIFICAÇÃO DA ESPÉCIE

Na identificação da espécie foi utilizado o sistema "código de barras" das plantas recomendado pelo CBOL, constituído por uma porção da região codificante de dois genes cloroplastidias (*matK* e *rbcL*) e os respetivos pares de *primers*, (*3F\_KIM/1R\_KIM*), para o *matK* e (*rbcLa\_f/rbcLa\_r*), para o *rbcL*, na amplificação e

sequenciação destas regiões, obtendo-se 28 sequências de DNA cloroplastidial. Estas sequências foram posteriormente submetidas na base de dados genômica BLAST, obtendo-se uma ou mais sequências homólogas, com um valor máximo de identidade em percentagem, para cada espécie.

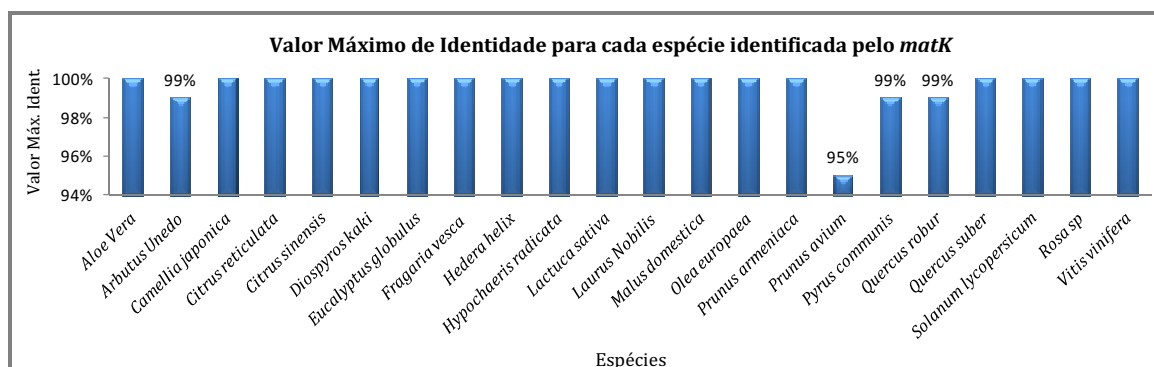
#### 4.2.1. *MATK*

Foi extraído com sucesso o DNA cloroplastidial de 22 espécies de plantas, usando dois métodos de extração (*DNeasy® Plant Kit* e *BioRobot® EZ1*). As sequências homólogas obtidas através do BLAST, após amplificação e sequenciação foram comparadas, escolhendo-se aquelas cuja extração proporcionou melhores resultados. Na Tabela IV é possível observar a correspondência a nível taxonómico de género ou espécie, bem como a extração considerada.

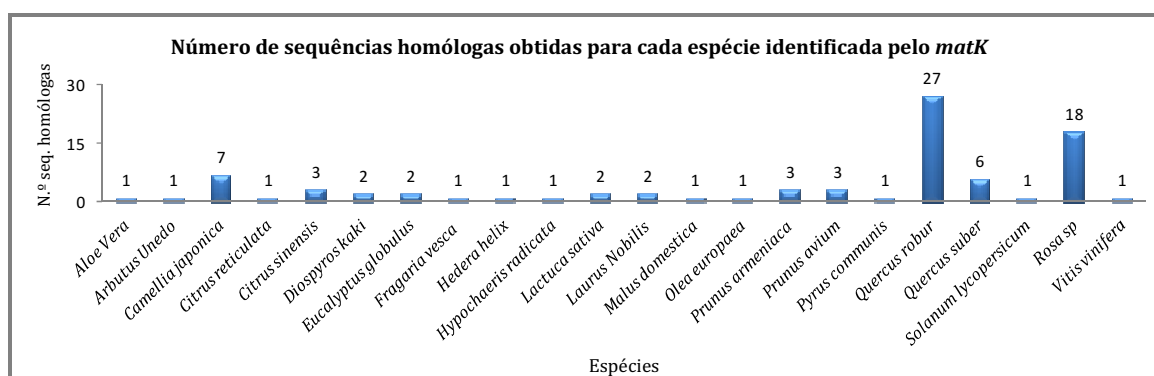
**Tabela IV** - Espécies de plantas, correspondência a nível taxonómico para a região "código de barras" do gene *matK* e método de extração de DNA considerado.

Espécies de plantas	Correspondência taxonómica	Extração utilizada
<b>Angiospérmicas</b>		
<i>Aloe Vera</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Arbutus Unedo</i>	Género	<i>BioRobot® EZ1</i>
<i>Camellia japonica</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Citrus sinensis</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Citrus reticulata</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Eucalyptus globulus</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Diospyros sp</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Fragaria vesca</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Hedera helix</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Hypochaeris radicata</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Lactuca sativa</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Laurus Nobilis</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Malus domestica</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Olea europaea</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Prunus armeniaca</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Prunus avium</i>	Género	<i>BioRobot® EZ1</i>
<i>Pyrus communis</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Quercus robur</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Quercus suber</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Rosa sp</i>	Género	<i>DNeasy® Plant Kit</i>
<i>Solanum lycopersicum</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Vitis vinifera</i>	Espécie	<i>BioRobot® EZ1</i>

Nas Figuras 6 e 7, é possível visualizar o valor máximo de identidade e o número das sequências homólogas obtidas, para cada espécie.

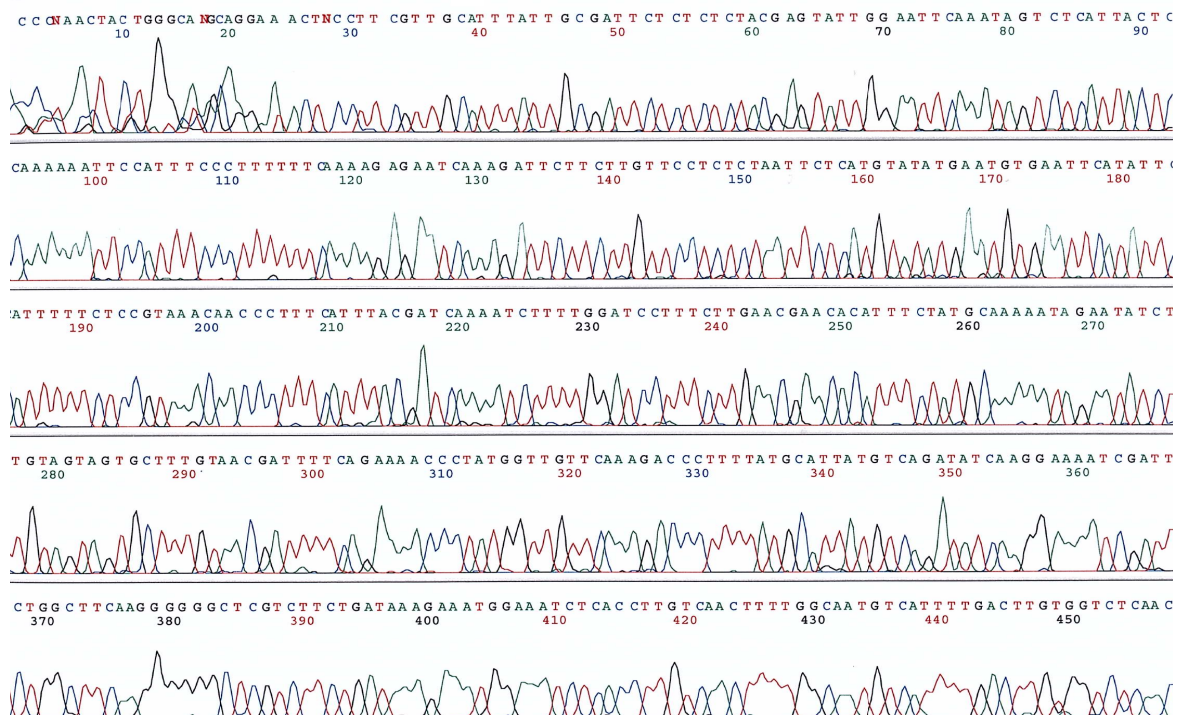


**Figura 6** - Espécies de plantas angiospérmicas e o respetivo valor máximo de identidade, por sequência submetida no BLAST, para uma região do gene *matK*.

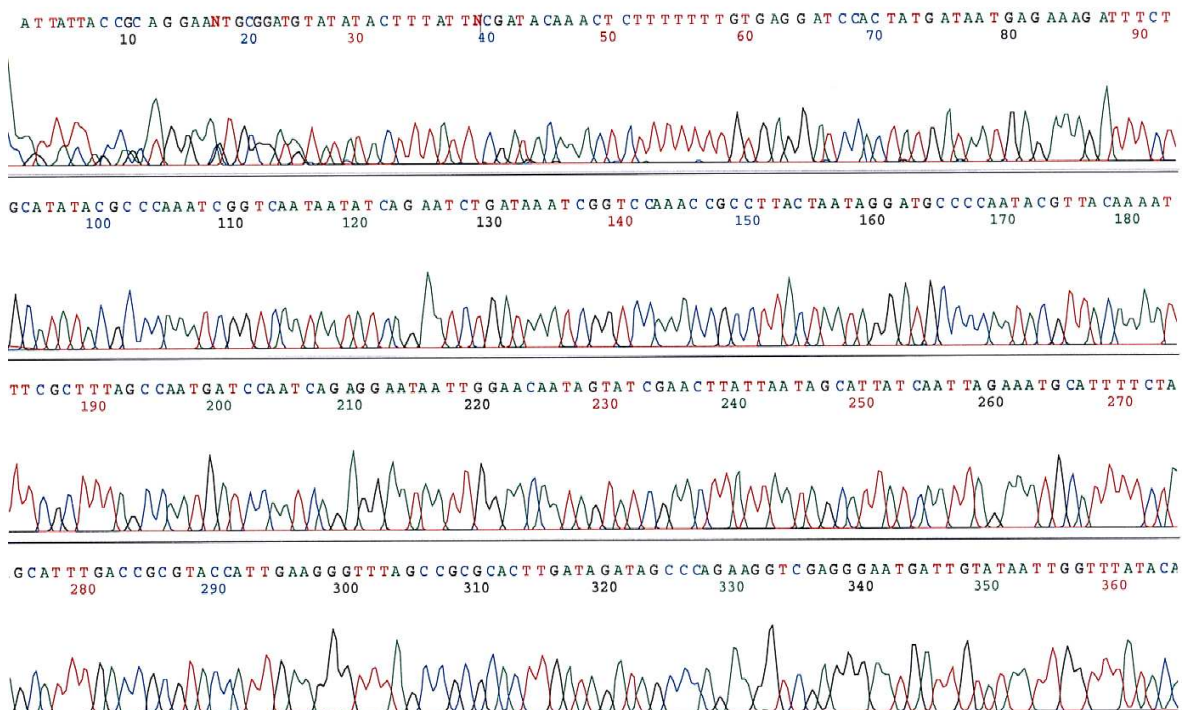


**Figura 7** - Espécies de plantas angiospérmicas e o respetivo número de sequências homólogas obtidas, por sequência submetida no BLAST, para uma região do gene *matK*.

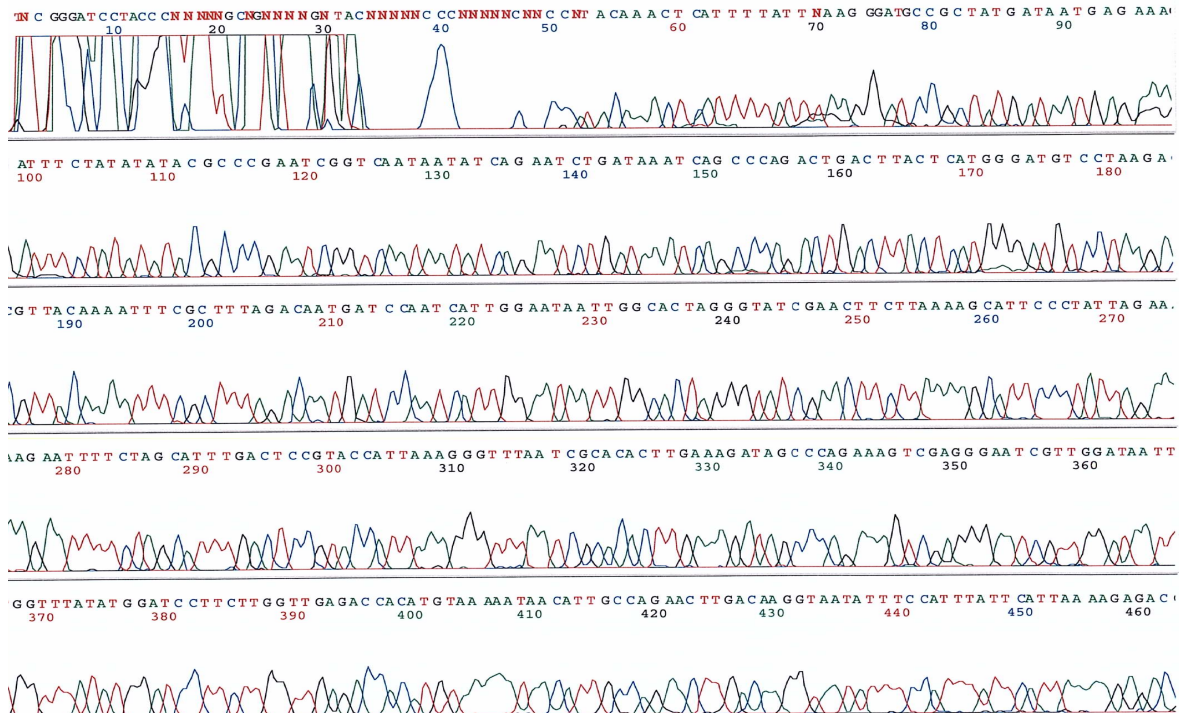
As angiospérmicas foram o grupo de plantas com a maior percentagem de sucesso na amplificação e sequenciação com o par de *primers* da região "código de barras" do gene *matK*, proporcionando uma sequência homóloga em 50% das espécies de plantas amplificadas e sequenciadas, quando a respetiva sequência obtida foi submetida no BLAST. Nas Figuras 8, 9, 10 e 11 é possível visualizar os eletroforetogramas de sequências obtidas após a amplificação e sequenciação de uma região do gene da *matK*, de 3 plantas indígenas da flora portuguesa, a *Laurus nobilis*, a *Hedera helix* e a *Arbutus Unedo*, extraídas pelo *BioRobot® EZ1*, bem como de uma espécie de rosa, a *Rosa sp*, extraída com o *DNeasy® Plant Kit*.



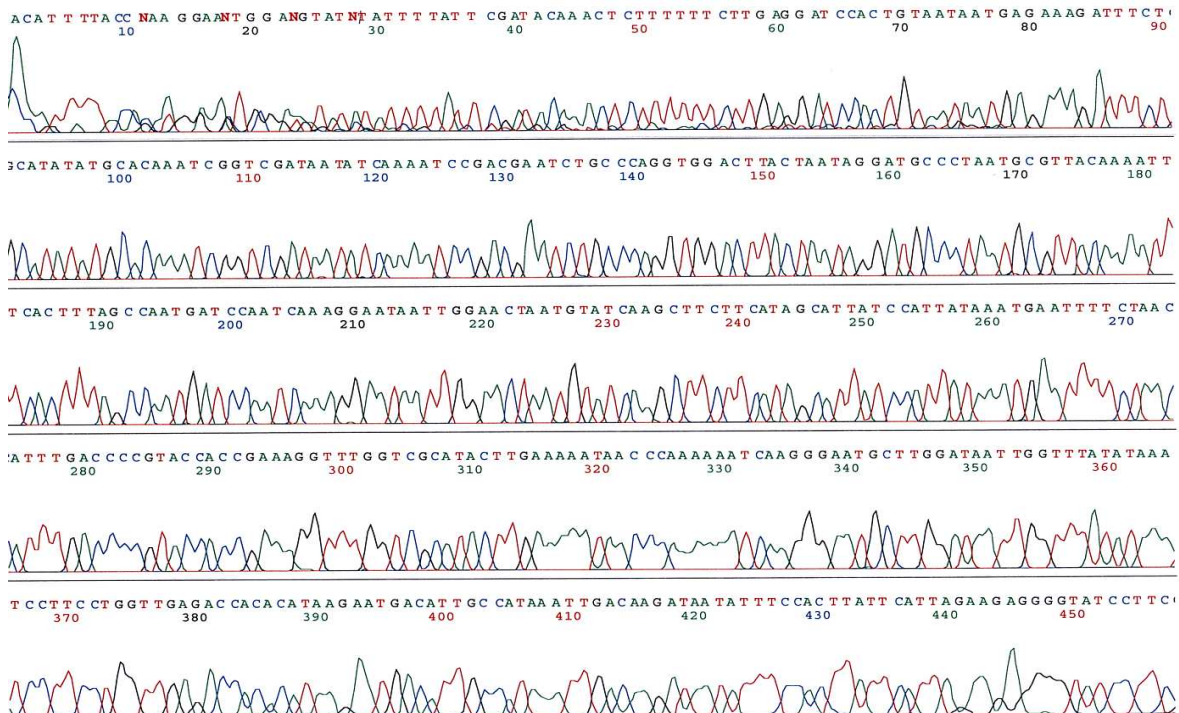
**Figura 8** - Sequência obtida para uma angiospérmica (*Laurus nobilis*) após extração do DNA no BioRobot® EZ1, amplificação e sequenciação de uma região do gene *matK*.



**Figura 9** - Sequência obtida para uma angiospérmica (*Hedere helix*) após extração do DNA no BioRobot® EZ1, amplificação e sequenciação de uma região do gene *matK*.



**Figura 10** - Sequência obtida para uma angiospérmica (*Arbutus unedo*) após extração do DNA no *BioRobot® EZ1*, amplificação e sequenciação de uma região do gene *matK*.



**Figura 11** - Sequência obtida para uma angiospérmica (*Rosa sp*) após extração do DNA com o *DNeasy® Plant Kit*, amplificação e sequenciação de uma região do gene *matK*.

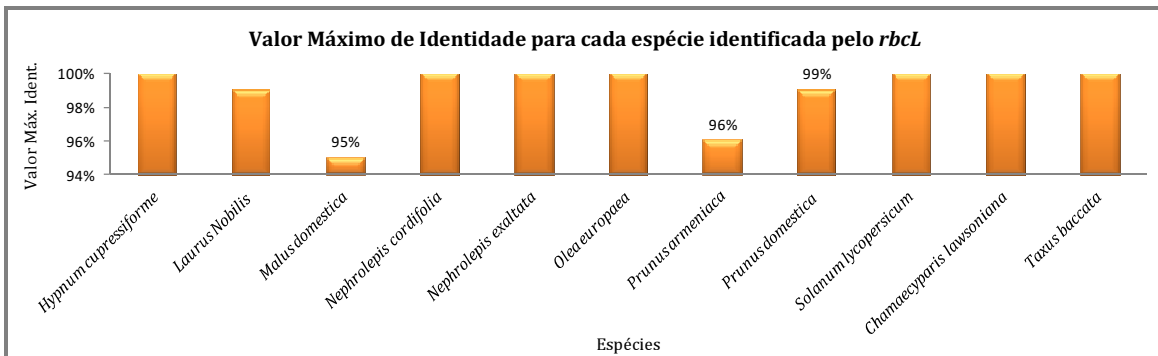
#### 4.2.2. *RBCL*

Foi extraído com sucesso o DNA cloroplastidial de 11 plantas com os dois métodos de extração (*DNeasy® Plant Kit* e *BioRobot® EZ1*). As sequências homólogas obtidas através do BLAST, após amplificação e sequenciação foram comparadas, tendo-se escolhido aquelas cuja extração proporcionou os melhores resultados. Na Tabela V é possível observar a correspondência a nível taxonómico de género ou de espécie, bem como a extração considerada. Nas Figuras 12 e 13 é possível visualizar o valor máximo de identidade e o número de sequências homólogas obtidas, para cada espécie.

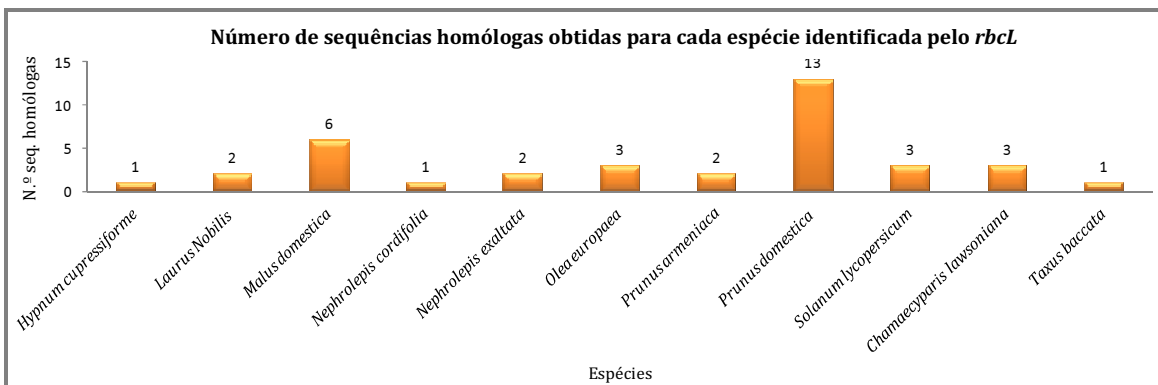
Para a região "código de barras" do gene *rbcl* obteve-se resultados em 6 angiospérmicas, 2 gimnospérmicas, 2 criptógamas vasculares e 1 não vascular. Nas Figuras 14, 15 e 16 é possível visualizar os eletroforetogramas de sequências obtidas após a amplificação e sequenciação de uma região do gene *rbcl*, de uma angiospérmica, a *Olea Europeae*, cujo DNA foi extraído com o *DNeasy® Plant Kit*. Mostra-se, igualmente os eletroforetogramas de uma criptógama vascular, a *Nephrolepis cordifolia* e de uma criptógama não vascular, a *Hypnum cupressiforme*, cujo DNA foi extraído no *BioRobot® EZ1*.

**Tabela V** - Espécies de plantas, correspondência a nível taxonómico para a região "código de barras" do gene *rbcl* e método de extração de DNA considerado.

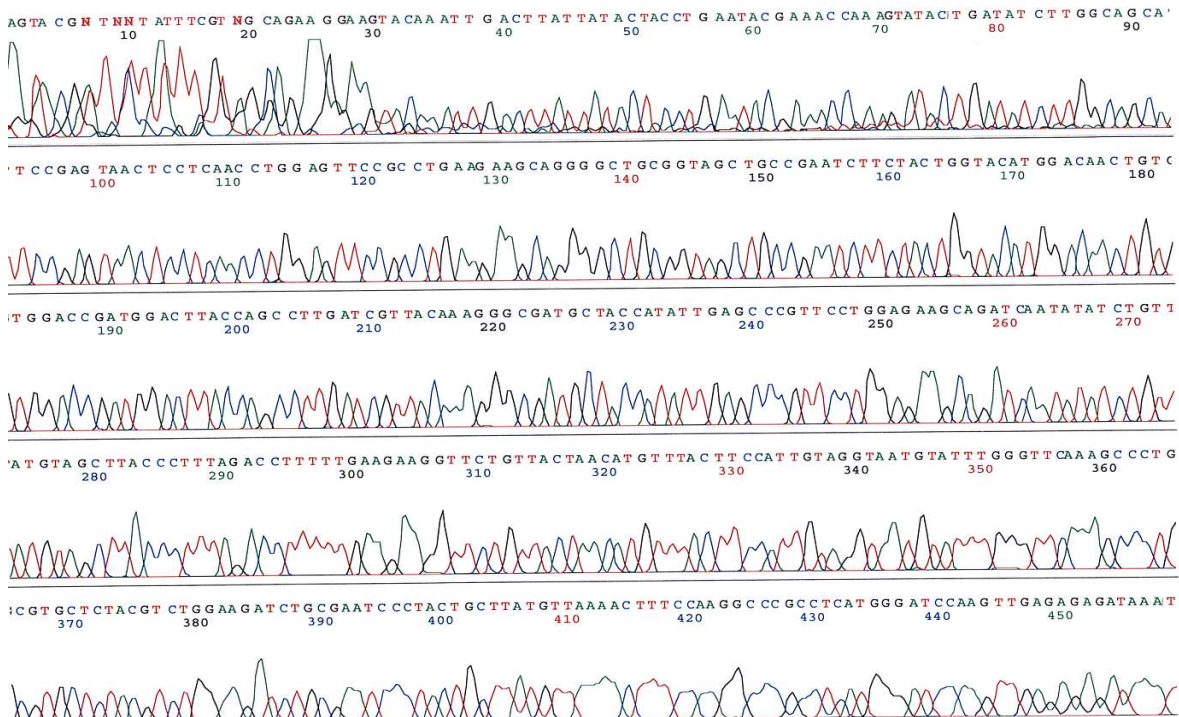
Espécies de plantas	Correspondência taxonómica	Extração utilizada
<b>Angiospérmicas</b>		
<i>Laurus Nobilis</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Malus domestica</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Olea europaea</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Prunus armeniaca</i>	Espécie	<i>DNeasy® Plant Kit</i>
<i>Prunus domestica</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Solanum lycopersicum</i>	Espécie	<i>BioRobot® EZ1</i>
<b>Gimnospérmicas</b>		
<i>Chamaecyparis lawsoniana</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Taxus baccata</i>	Espécie	<i>BioRobot® EZ1</i>
<b>Criptógamas vasculares</b>		
<i>Nephrolepis cordifolia</i>	Espécie	<i>BioRobot® EZ1</i>
<i>Nephrolepis exaltata</i>	Espécie	<i>BioRobot® EZ1</i>
<b>Criptógamas não vasculares</b>		
<i>Hypnum cupressiforme</i>	Espécie	<i>BioRobot® EZ1</i>



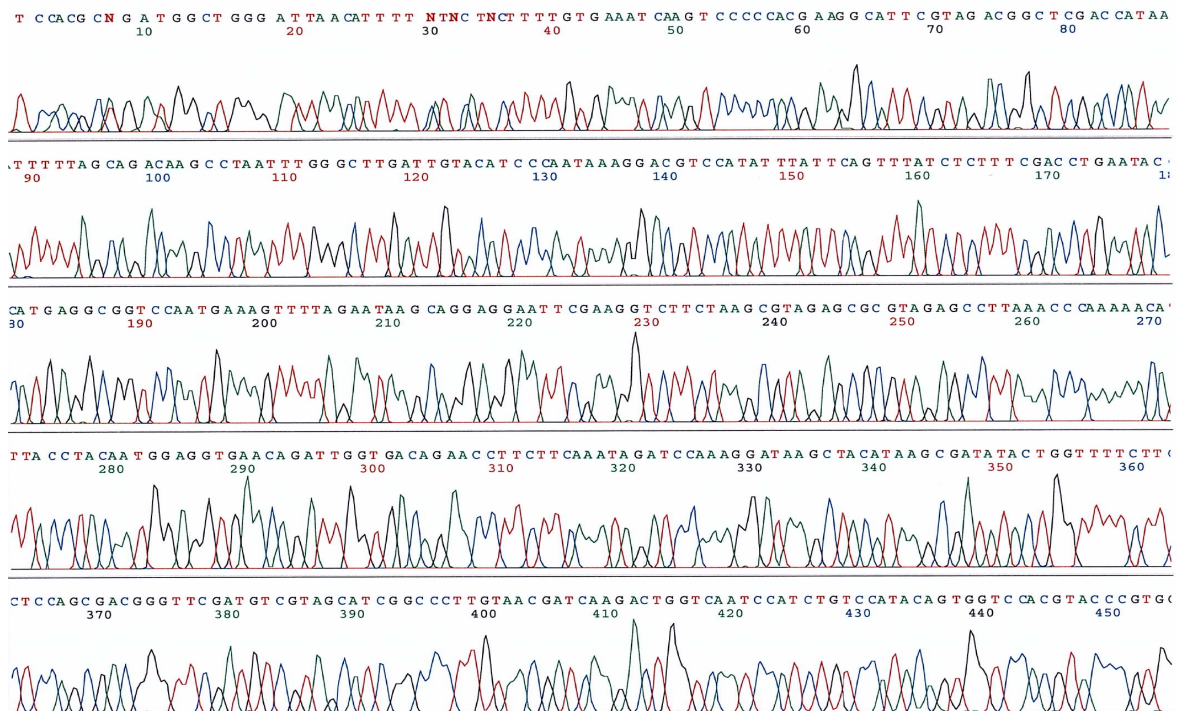
**Figura 12** - Espécies de plantas e o respetivo valor máximo de identidade, por sequência submetida no BLAST, para uma região do gene *rbCL*



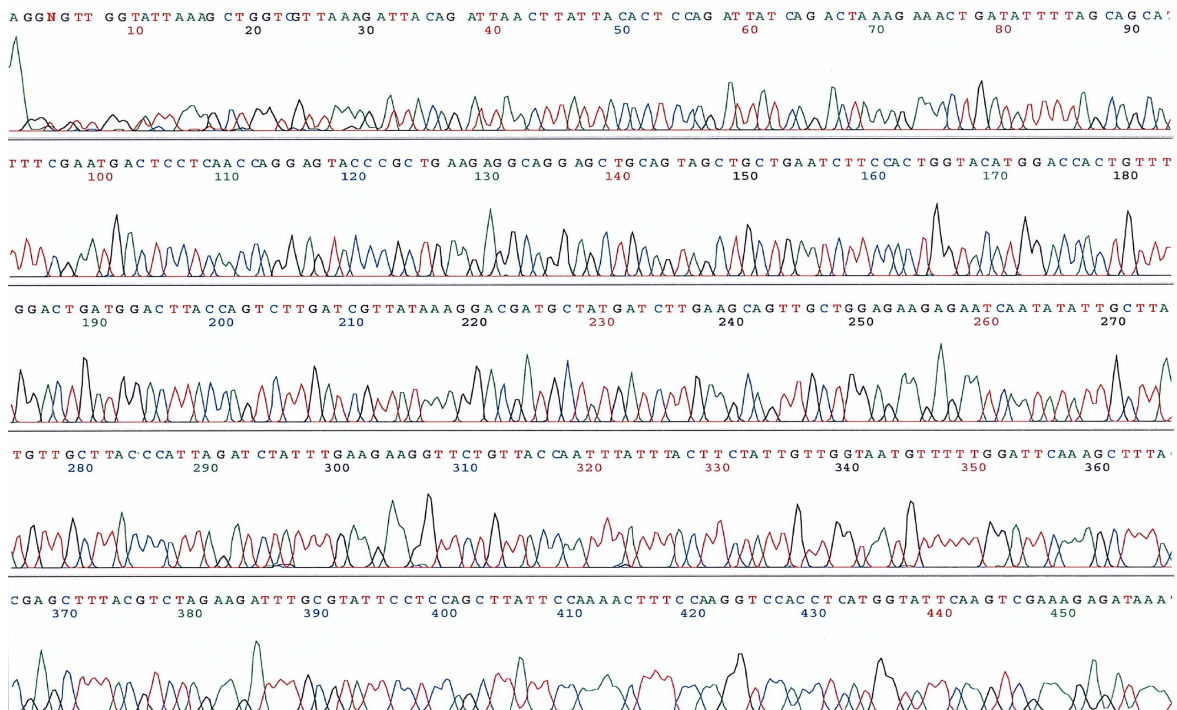
**Figura 13** - Espécies de plantas e o respetivo número de sequências homólogas obtidas, por sequência submetida no BLAST, para uma região do gene *rbCL*.



**Figura 14** - Sequência obtida para uma angiospérmica (*Olea Europea*) após extração do DNA com o *DNeasy® Plant Kit*, amplificação e sequenciação de uma região do gene *rbCL*.



**Figura 15** - Sequência obtida para uma criptógama vascular (*Nephrolepis cordifolia*) após extração do DNA no BioRobot® EZ1, amplificação e sequenciação amplificação e sequenciação de uma região do gene *rbcl*.



**Figura 16** - Sequência obtida para uma criptógama não vascular (*Hypnum cupressiforme*) após extração do DNA no BioRobot® EZ1, amplificação e sequenciação amplificação e sequenciação de uma região do gene *rbcl*.

## **5. DISCUSSÃO**

## 5. DISCUSSÃO

### 5.1. EXTRAÇÃO DO DNA

Neste estudo utilizaram-se dois métodos de extração do DNA: o *DNeasy® Plant Kit* e o *BioRobot® EZ1*. Verificou-se que DNA extraído no *BioRobot® EZ1* proporcionou maior sucesso na obtenção de sequências de DNA cloroplastidial. Pois, das 28 amostras de plantas, obtiveram-se sequências de DNA cloroplastidial em 21. Quanto ao método de extração pelo *DNeasy® Plant Kit*, os resultados não foram animadores. Pois, das 22 amostras cujo DNA foi extraído com este *kit* obtiveram-se sequências de DNA cloroplastidial em apenas 9. Quanto aos dois grupos de 12 amostras extraídas com ambos os métodos, foram obtidas 9 sequências de DNA cloroplastidial no *BioRobot® EZ1* e apenas 3 sequências de DNA cloroplastidial com o *DNeasy® Plant Kit*.

Este menor número de resultados obtidos de amostras extraídas com o *DNeasy® Plant Kit* não se deve, provavelmente, à ineficiência do método, mas por ser necessário a utilização de amostras botânicas praticamente reduzidas a pó, com a ajuda do azoto líquido. Em sua substituição, utilizou-se o pilão e almofariz para macerar as plantas, sendo, de seguida o material botânico digerido no tampão de lise e proteinase K durante 12-18 h. Apesar dos bons resultados apresentados na literatura com esta alteração ao protocolo (Stoeckle *et al.*, 2011), o método *DNeasy® Plant Kit* não produziu os resultados desejados. Em consequência, o método escolhido, entre os dois disponíveis para este estudo, seria o *BioRobot® EZ1* devido à qualidade do DNA obtido, baixo risco de contaminação, não necessitando de azoto líquido, e bom desempenho demonstrado através dos resultados obtidos.

Quanto à quantidade de amostra necessária, houve alguns problemas na sua otimização. Os protocolos estabelecem até 100 mg de planta, mas este valor é manifestamente elevado. Após algumas experiências, conclui-se que grandes quantidades de material botânico inviabilizam a amplificação. Assim, a quantidade ideal está compreendida entre 5 mg e 10 mg; no entanto, foi possível amplificar o DNA extraído a partir de 1 mg de planta.

## 5.2. AMPLIFICAÇÃO E SEQUENCIAÇÃO DO DNA

Na amplificação por PCR foram escolhidos os pares de *primers* (*3F\_KIM/1R\_KIM*; K. J. Kim, não publicado), para gene o *matK* e o (*rbcLa\_f/rbcLa\_r*; Kress e Erickson, 2007), para o gene *rbcL*, com os respectivos protocolos recomendados pelo CBOL.

Para a região "código de barras" do gene *matK*, o protocolo de CBOL proporcionou bons resultados, para a maioria das plantas angiospérmicas, com temperatura de *annealing* de 52°C e 35 ciclos. Para amostras difíceis, isto é, as que não amplificaram, como as gimnospérmicas e criptógamas, a temperatura de *annealing* utilizada variou entre 48°C e 60°C e ciclos entre 35 e 40. Contudo, nenhuma das plantas deste grupo amplificou.

Para região "código de barras" do gene *rbcL*, obtiveram-se resultados mais modestos, tanto na amplificação como na sequenciação das plantas. Pois, foi apenas amplificado e sequenciado o DNA cloroplastidial de 11 espécies de plantas em 28, cujo DNA cloroplastidial foi extraído. Destas 11 plantas, apenas 4 amplificaram de acordo com o protocolo inicialmente estabelecido (*Hypnum cupressiforme*, *Olea europeae*, *Nephrolepis cordifolia* e *Nephrolepis exaltata*). Em relação às restantes, experimentaram-se várias temperaturas de *annealing*, conseguindo-se amplificar o DNA de 7 plantas (*Chamaecyparis lawsoniana*, *Laurus nobilis*, *Solanum lycopersicon*, *Malus domestica*, *Prunus armeniaca*, *Prunus domestica* e *Taxus baccata*) com uma temperatura de *annealing* de 50°C e igualmente com 35 ciclos.

## 5.3. IDENTIFICAÇÃO DA ESPÉCIE

Após a amplificação e sequenciação, obtiveram-se os respectivos eletroforetogramas e as sequências no formato FASTA foram submetidas na base de dados genómica BLAST. O número de pares de bases recomendados a submeter à base de dados genómica é de cerca de 500 pb, produzindo eletroforetogramas sem ambiguidades. No entanto, foi possível obter-se um valor máximo de identidade de 100% e apenas uma sequência homóloga na angiospérmica *Vitis vinifera*, com 331 pb.

Contudo, salvo algumas exceções, obtêm-se eletroforetogramas sem grandes ambiguidades, com identidade de 100% e uma sequência homóloga efetivamente a partir de 500 pb.

O sistema "código de barras" do DNA das plantas, constituído por porções de dois genes cloroplastidiais (*matK* e *rbcL*), em conjunto, identificaram 73,68% das plantas analisadas. Das 28 plantas identificadas, 25 tiveram correspondência ao nível da espécie e 3 ao nível do género (*Arbutus Unedo*, *Prunus avium* e *Rosa* sp). Quanto ao nível de discriminação, em 28 plantas identificadas, obteve-se uma sequência homóloga por sequência submetida no BLAST em 14 plantas, ou seja, um nível de discriminação de 50%.

### 5.3.1. *MATK*

#### 5.3.1.1. ANGIOSPÉRMICAS

As angiospérmicas constituem o maior grupo de plantas terrestres, tendo o este estudo incidido, em especial, sobre esse vasto grupo. A família com maior número de plantas estudadas foram as *Rosaceae* (Tabela II e III). Neste sentido, extraiu-se o DNA de 32 destas plantas e obtiveram-se 22 sequências de DNA cloroplastidial da região "código de barras" do gene *matK* (Figura 6). Contudo, não houve sucesso na amplificação e sequenciação das angiospérmicas *Brassica oleracea*, *Castanea sativa*, *Citrus limon*, *Ilex aquifolium*, *Mentha spicata*, *Passiflora edulis*, *Prunus domestica*, *Rosmarinus officinalis*, *Spinacia oleracea* e *Solanum melongena*. Das 22 sequências resultantes foi obtido um valor máximo de identidade de 100% para 18 plantas, ou seja, em 81,81%. Das restantes amostras de plantas, obteve-se um valor máximo de identidade de 99% para 3 plantas (*Arbutus Unedo*, *Quercus robur* e *Pyrus communis*) e de 95% para a angiospérmica *Prunus avium*.

O par de *primers* para o "código de barras" do gene *matK* (3F\_KIM/1R\_KIM) utilizado neste estudo é considerado o melhor par 'universal' disponível, amplificando e sequenciando cerca de 70% das angiospérmicas. Este par de *primers* é por vezes complementado com uma outra versão de par de *primers* (390F/1326R,

Cuénoud *et al.*, 2002) aumentando a amplificação e sequenciação em cerca de 10%. (Hollingsworth *et al.*, 2011). Neste estudo, o par de *primers* recomendado pelo CBOL, permitiu amplificar e sequenciar cerca de 68,75% de angiospérmicas. O valor determinado é similar ao de outros autores.

Quanto ao sucesso na discriminação de espécies nas plantas, este é menor do que em animais a partir do gene mitocondrial COI. Os níveis de discriminação variam entre *taxa*, mas uma discriminação de espécies inferior a 70% é bastante comum (Hollingsworth *et al.*, 2011). Idealmente, o número de sequências homólogas por sequência obtida, quando submetida na base de dados genómica BLAST, seria apenas uma. Neste estudo, foi obtida uma sequência homóloga em 11 plantas angiospérmicas das 22 amplificadas e sequenciadas, quando as respectivas sequências obtidas foram submetidas no BLAST, ou seja, obteve-se um nível de discriminação de 50% para este grupo de plantas (Figura 7).

Por serem plantas indígenas em Portugal, havia alguma expectativa quanto às plantas do género *Quercus* e um seu parente próximo, *Castanea sativa*, bem como as plantas angiospérmicas *Ilex aquifolium*, *Rosmarinus officinalis* e *Arbutus Unedo*, na identificação e discriminação destas espécies. O par de *primers* para a "código de barras" do gene *matK* não amplificou as angiospérmicas *Castanea sativa*, *Ilex aquifolium* e *Rosmarinus officinalis*. As plantas do género *Quercus* apresentaram um baixo nível de discriminação, com várias sequências homólogas por sequência submetida no BLAST: 6 para a *Quercus suber* (Sobreiro) e 27 para a *Quercus ruber* (Carvalho). Este facto é corroborado por estudos anteriores que demonstraram que o género *Quercus* é refratário ao nível da discriminação de espécies, tendo como provável explicação a hibridização e a baixa taxa de variação do genoma dos plastídeos (Piredda *et al.*, 2011). Do mesmo modo, espécies como as angiospérmicas *Rosa* sp e *Prunus avium* tiveram igualmente níveis de discriminação pouco satisfatórios e um valor máximo de identidade baixo, no caso da *Prunus avium*, sendo que estes valores refletem a dificuldade na amplificação e sequenciação de amostras de plantas da família das *Rosaceae* (Pang *et al.*, 2010). Contudo, neste estudo, a família das *Rosaceae* proporcionou bons resultados, com valores máximos de identidade de 100% para algumas espécies (*Fragaria vesca*, *Malus domestica* e *Prunus armeniaca*).

Alguns grupos de plantas são complexos de identificar ao nível da espécie, tendo em consideração a obtenção de apenas uma sequência homóloga para a mesma, resultando num bom poder de discriminação, para a região "código de barras" do gene *matK*. Por isso, não é possível identificar todas as plantas terrestres. Neste sentido, continuam a ser estudados *primers* específicos para determinados grupos taxonómicos, bem como *cocktails* de versões destes *primers* (Hollingsworth *et al.*, 2011).

Quanto à sequência obtida para a angiospérmica *Arbutus unedo* (Figura 10), obteve-se um valor máximo de identidade de 99% com apenas uma sequência homóloga. A sequência homóloga em causa, corresponde à *Arbutus canariensis* com um grau de similaridade apenas ao nível do género. Tal facto acontece porque a *Arbutus Unedo* tem apenas algumas sequências de genes na base de dados genómica BLAST, não incluindo as sequências do gene da *maturase K* nem do gene da RuBisCo. Perante esta circunstância, foi confirmado positivamente que as sequências dos genes da *maturase K* e da RuBisCo estavam no repositório da base de dados genómica BLAST para todas as plantas, exceto para as angiospérmicas *Passiflora edulis*, *Solanum melongena* e *Spinacia oleracea*, no caso do gene da *maturase K*. Efetivamente, o caso da angiospérmica *Arbutus Unedo* alertou para outras plantas cuja sequência dos genes da *maturase K* e RuBisCo ainda não estão nesses repositórios, sendo necessário enriquecer estas bases de dados para uma melhor e correta identificação das plantas.

#### 5.3.1.2. GIMNOSPÉRMICAS

No grupo das gimnospérmicas, extraiu-se o DNA cloroplastidial de 3 plantas (*Chamaecyparis lawsoniana*, *Pinus pinaster* e *Taxus baccata*); contudo, não foi possível amplificar as plantas deste grupo com este par de *primers*. Estudos anteriores confirmam que o género *Pinus* revela uma dificuldade acrescida na amplificação pois apresenta *loci* em duplicado (Hollingsworth *et al.*, 2011) e, relativamente ao género *Taxus*, este é difícil de amplificar com o par de *primers* (*3F\_KIM/1R\_KIM*), apresentando um baixo poder de discriminação para esta região, devido à falta de diferenças morfológicas entre espécies (Liu *et al.*, 2010). Consequentemente, neste

grupo de plantas, o CBOL recomenda uma outra versão deste par de *primers* para esta região "código de barras" (CBOL Plant Working Group, 2009).

Embora, a amostragem para as gimnospérmicas seja pequena, este facto reflete a realidade de outras investigações, que estudam maioritariamente plantas angiospérmicas, isto porque este par de *primers* foi desenhado para este último grupo de plantas, sendo difícil de amplificar por PCR noutros grupos taxonómicos.

### 5.3.1.3. CRIPTÓGAMAS

Nas criptógamas vasculares (*Nephrolepis* spp) e não vascular (*Hipnum cupressiforme*) estudadas não se obteve sucesso na amplificação e sequenciação com este par de *primers*.

O CBOL, através dos seus grupos de trabalho, verificou que o sucesso é limitado na amplificação e sequenciação de criptógamas (10%). Aliás, os fetos representam um desafio para a região "código de barras" do gene *matK*. O genoma de quase 90% de todas as espécies de fetos existentes não têm o arranjo *trnK/matK*. Ora, os pares de *primers* usados para amplificar a região "código de barras" do gene *matK* noutros grupos de plantas encontram-se entre o *trnK* ou os seus intrões, resultando assim numa sub-representação dos fetos em estudos anteriores. Além disso, a grande taxa de substituição de nucleotídeos do gene *matK* torna difícil desenhar pares de *primers* em grandes grupos taxonómicos, sendo necessários pares de *primers* específicos para estes taxa (Hilu *et al.*, 2003).

### 5.3.2. *RBCL*

#### 5.3.2.1. ANGIOSPÉRMICAS

O par de *primers* utilizado e recomendado pelo CBOL para a região "código de barras" do gene *rbcL*, obteve resultados insatisfatórios na amplificação e sequenciação das amostras de plantas angiospérmicas. Foram apenas amplificadas e sequenciadas 6 plantas angiospérmicas. Amplificou-se e sequenciou-se as

angiospérmicas *Laurus Nobilis*, *Solanum lycopersicum*, *Malus domestica*, *Olea europaea*, *Prunus armeniaca* e *Prunus domestica*. Para 3 destas plantas, cujo DNA foi amplificado e sequenciado, obteve-se um valor máximo de identidade de 100%, e uma sequência homóloga apenas numa planta, quando as sequências obtidas foram submetidas na base de dados genómica BLAST (Figuras 12 e 13). Nas espécies da família das *Rosaceae*, a região "código de barras" para o *rbcL* teve níveis de discriminação fracos e valor máximo de identidade baixos (*Malus domestica*, *Prunus armeniaca* e *Prunus domestica*).

Apesar da região "código de barras" do gene *rbcL* ser "fácil de amplificar, sequenciar e alinhar na maioria das plantas terrestres, embora com um poder discriminatório modesto" (Hollingsworth *et al.*, 2011), neste estudo não houve grandes sucessos na amplificação e sequenciação do DNA cloroplastidial das plantas angiospérmicas com este par *primers* recomendado para este região, apresentando valores máximos de identidade e poder de discriminação não tão animadores, quando comparados com a região "código de barras" do gene *matK*. Este resultado, não foi determinante para este estudo já que foi complementado com os bons resultados na amplificação e sequenciação, bem como com um razoável poder de discriminação para a região "código de barras" do gene *matK*.

#### 5.3.2.2. GIMNOSPÉRMICAS

Nas gimnospérmicas, das 3 espécies de plantas cujo DNA cloroplastidial foi extraído, 2 delas foram amplificadas e sequenciadas com sucesso: o *Chamaecyparis lawsoniana*, obtendo-se 3 sequências homólogas com um valor máximo de identidade de 100% e o *Taxus bacatta*, obtendo-se 1 sequência homóloga com um valor máximo de identidade de 100%, relativa às amostras sequenciadas (Figuras 12 e 13). De facto, estudos anteriores demonstram que a região "código de barras" do gene *rbcL* tem um bom poder de discriminação para o *Taxus baccata* (Liu *et al.*, 2011). Quanto à gimnospérmica *Pinus pinaster*, não se obteve sucesso na amplificação por PCR com este par de *primers*.

### 5.3.2.3. CRIPTÓGAMAS

Nas criptógamas vasculares (*Nephrolepis* spp), o protocolo seguido foi o inicialmente estabelecido; no entanto, estas plantas também amplificaram com sucesso com uma temperatura de *annealing* de 48°C. Extraíu-se o DNA de 2 espécies de plantas, amplificando-se e sequenciando-se com sucesso as duas. Das duas sequências resultantes obteve-se um valor máximo de identidade de 100% para ambas, com 1 sequência homóloga para a *Nephrolepis cordifolia* e 2 sequências homólogas para a *Nephrolepis exaltata*, quando submetidas na base de dados genómica BLAST (Figuras 12 e 13). Embora a amostragem deste grupo de plantas seja pequena, a região "código de barras" do gene *rbcl* obtém bons resultados na maioria dos fetos e um razoável poder de discriminação ao nível da espécie (Schneider e Schuettpelz, 2006).

Quanto à criptógama não vascular estudada (*Hypnum cupressiforme*), o protocolo seguido na amplificação foi o inicialmente estabelecido, utilizando este par de *primers*, obtendo-se um valor máximo de identidade de 100% e uma sequência homóloga, quando a sequência obtida foi submetida na base de dados genómica BLAST (Figuras 12 e 13). Embora tenha sido estudado apenas uma amostra neste grupo de plantas, estudos anteriores demonstraram que se obtém bons resultados na amplificação e sequenciação de musgos, utilizando este par de *primers*, para esta região "código de barras" do gene *rbcl* (Liu *et al.*, 2010).

## 5.4. CONSIDERAÇÕES FINAIS

Este estudo incluiu espécies de plantas indígenas, naturalizadas ou comuns à flora portuguesa, colhidas entre os meses de maio e outubro. Por se tratar de uma época do ano de grande floração, frutificação e maturação dos frutos, incluíram-se árvores de fruto na amostragem. Contudo, o estudo e recolha de plantas existentes na flora portuguesa devem continuar, de modo a aumentar conhecimentos na identificação de espécies de plantas desta área geográfica.

Os métodos de identificação de plantas através da tipagem do DNA, ou seja, a descoberta de uma região ou regiões "código de barras" de todas as plantas, ainda não é de todo possível. Deste modo, a colaboração entre vários grupos de trabalho é essencial, sendo necessário estabelecer protocolos de armazenamento de tecidos de plantas, bem como de amplificação por PCR e sequenciação do DNA. Idealmente, deveriam ser desenvolvidos protocolos de reações em *multiplex* para a PCR, bem como o desenho de pares de *primers* diferentes, de forma a que resultem nos mais variados grupos taxonómicos.

A diversidade das plantas terrestres não permite que as bases de dados genómicas possuam todas as sequências homólogas dos genes de espécies de plantas. Neste trabalho, verificou-se que uma planta indígena da flora portuguesa, a angiospérmica *Arbutus unedo*, não possui sequências homólogas para os genes *matK* e *rbcL*. Deste modo, este estudo poderá servir para acrescentar e enriquecer as bases de dados genómicas com sequências das plantas indígenas da flora portuguesa.

Este estudo sobre identificação de plantas com recurso à Botânica Molecular poderá servir igualmente de base em perícias forenses, uma vez que a Botânica Clássica requer conhecimentos específicos e aprofundados sobre plantas, não tendo por vezes aplicabilidade, pelo facto do tecido vegetal ser vestigial ou se encontrar degradado. Em condições ideais, devem ser usadas a Botânica Clássica e a Molecular na resolução de perícias forenses.

## **6. CONCLUSÕES**

## 6. CONCLUSÕES

- 1<sup>a</sup> O material botânico foi colhido na zona do norte de Portugal, tendo-se tido o cuidado de escolher algumas plantas indígenas, naturalizadas ou comuns, da flora portuguesa.
- 2<sup>a</sup> Os métodos de extração do DNA cloroplastidial utilizados foram o *kit DNeasy® Plant Kit* e o *BioRobot® EZ1*. O método mais eficaz foi o *BioRobot® EZ1*, pois proporcionou melhores resultados na obtenção de DNA para a amplificação e sequenciação das espécies de plantas.
- 3<sup>a</sup> O sistema "código de barras" do DNA das plantas, constituído por porções de dois genes cloroplastidiais (*matK* e *rbcl*), permitiu em conjunto identificar a maioria das plantas de todos os grupos taxonómicos em estudo.
- 4<sup>a</sup> A região "código de barras" do gene *matK* possibilitou a identificação de grande parte das plantas angiospérmicas, contrariamente ao gene *rbcl*, que não proporcionou resultados tão satisfatórios.
- 5<sup>a</sup> A região "código de barras" do gene *rbcl* facultou a identificação de duas gimnospérmicas e das criptógamas deste estudo, contrariamente ao gene *matK* que não proporcionou nenhum resultado para estes grupos taxonómicos.
- 6<sup>a</sup> A identificação da *Arbutus unedo*, planta indígena portuguesa, cuja sequência relativa ao *matK* não estava introduzida no BLAST, permitirá enriquecer a base de dados de sequências de DNA cloroplastidial. Por outro lado, obtiveram-se resultados promissores em plantas da família das *Rosaceae*.

Em face do exposto, este trabalho acrescentou conhecimentos na identificação de espécies de plantas, algumas das quais indígenas de Portugal, que poderão ser utilizados a nível forense no futuro.

## **7. REFERÊNCIAS BIBLIOGRÁFICAS**

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Applied Biosystems (2001) ABI PRISM® 3100 Genetic Analyzer User's Manual, Applied Biosystems, Inc.
- [2] Applied Biosystems (2002) BygDye® Terminator v1.1 Cycle Sequencing Kit Protocol, Applied Biosystems, Inc.
- [3] Bock JH e Norris DO (1997). Forensic Botany: an under-utilized resource. *J Forensic Sci* 42 (3): 364-367
- [4] Cainé L (2011) Entomologia Forense: Identificação Genética de Espécies em Portugal. Tese de Doutoramento em Ciências da Saúde, ramo Ciências Biomédicas - Faculdade de Medicina da Universidade de Coimbra. Coimbra. 98 pg.
- [5] CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proc Natl Acad Sci USA* 106 (31): 12794:12797.
- [6] Chase MW, Cowan RS, Hollingsworth PM, Berg C, Madriñán S, Petersen G, Seberg O, Jørgensen T, Camero, KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L e Wilkinson M (2007) A proposal for a standardized protocol to barcode all land plants. *Taxon* 56 (2): 295-299.
- [7] Cho Y, Mower JP, Qiu Y-L e Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc Natl Acad Sci USA* 101: 17741-17746.
- [8] Coyle HM, Ladd C, Palmbach T e Lee HC (2001) The Green Revolution: Botanical Contributions to Forensics and Drug Enforcement. *Croat Med J* 42 (3): 340-345.
- [9] Coyle HM, Palmbach T, Juliano N, Ladd C e Lee HC (2003) An Overview of DNA Methods for the Identification and Individualization of Marijuana. *Croat Med J* 44: 315-321.
- [10] Cuénoud P, Savolainen L, Chatrou LM, Powell M, Grayer RJ e Chase MW (2002) Molecular Phylogenetics of Caryophyllales based on Nuclear 18S rDNA and Plastid *rbcl*, *atpB*, and *matK* DNA Sequences. *Am J Bot* 89: 132-144.
- [11] Decreto-Lei n.º 565/99 de 21 de Dezembro. Diário da República n.º 295, I Série-A, Ministério do Ambiente, Lisboa.

- [12] Ferri G, Alù M, Corradini B e Beduschi G (2009) Forensic Botany: species identification of botanical trace evidence using multigene barcoding approach. *Int J Legal Med* 123: 395-401.
- [13] Hilu KW e Liang H (1997) The *matK* gene: sequence variation and application in plant systematics. *Am J Bot* 84 (6): 830– 839.
- [14] Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, AA Lawrence, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS e Chatrou LW (2003) Angiosperm phylogeny based on *matK* sequence information. *Am J Bot* 90: 1758–1776.
- [15] Hollingsworth PM, Graham SW e Little DP (2011) Choosing and Using a Plant DNA Barcode. *PLoS ONE* 6 (5): e19254.
- [16] Kress WJ, Zimmer EA, Weigt LA e Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* 102: 8369–8374.
- [17] Kress WJ e Erickson DL (2007) A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcl* Gene Complements the Non-Coding *trnHpsbA* Spacer Region. *PLoS ONE* 2: e508.
- [18] Lahaye R, Savolainen , Vincent , Duthoit , Sylvie , Maurin , Olivier , van der Bank, Michelle (2008a) A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. *Nature Preceedings*: hd1:10101/npre.2008.1896.1
- [19] Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG e Savolainen V (2008b) DNA barcoding the floras of biodiversity hotspots. *Proc Natl Acad Sci USA* 105: 2923–2928.
- [20] Liu JIE, Möller M, Gao LM, Zhang DQ e Li DZ (2011) DNA barcoding for the discrimination of Eurasian yews (*Taxus L.*, *Taxaceae*) and the discovery of cryptic species. *Molecular Ecology Resources* 11: 89–100.
- [21] Liu Y, Yan HFY, Cao T e Ge J (2010) Evaluation of 10 plant barcodes in Bryophyta (Mosses). *Journal of Systematics and Evolution* 48 (1): 36–46.
- [22] Neuhaus H e Link G (1987) The chloroplast tRNA(UUU) gene from mustard (*Sinapsis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. *Curr Genet* 11: 251–257.
- [23] Mauseth JD (1998) *Botany: An introduction to plant Biology*. 2<sup>th</sup> Ed. Jones and Bartlett Publishers Canada.

- [24] Ministério da Agricultura, Mar, Ambiente e Ordenamento do Território – Autoridade Florestal Nacional (2011), acessado em 22 de Novembro de 2011 no endereço eletrónico  
<http://www.afn.min-agricultura.pt/portal/gestao-florestal/gfs/eafu-pt>
- [25] Mullis KB (1990) The unusual origin of the Polymerase Chain Reaction. *Scientific American* 56-65.
- [26] Pang X, Song J, Zhu Y, Xu H, Huang I e Chen S (2010) Applying plant DNA barcodes for *Rosaceae* species identification. *Cladistics* 17: 165-170
- [27] Pennisi E (2007) Taxonomy. Wanted: a barcode for plants. *Science* 318: 190–191.
- [28] Piredda R, Simeone MC, Attimonelli M, Bellarosa R e Schirone B (2010). Prospects of barcoding the Italian wild dendroflora: oaks reveal severe limitations to tracking species identity. *Molecular Ecology Resources* 11: 72–83.
- [29] PlantGDB BLAST (2011), disponível no endereço eletrónico  
<http://www.plantgdb.org/cgi-bin/blast/PlantGDBblast>
- [30] Qiagen (2002) DyeEx™ Handbook, Qiagen Group.
- [31] Qiagen (2006) DNeasy® Plant Handbook, Qiagen Group.
- [32] Qiagen (2010a) EZ1® DNA Tissue Handbook, Qiagen Group.
- [33] Qiagen (2010b) HotStarTaq® Plus PCR Handbook, Qiagen Group.
- [34] Schneider H e Schuettpelz E (2006) Identifying fern gametophytes using DNA sequences. *Molecular Ecology Notes* 6: 989–991.
- [35] Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H e Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5: 2043–2049.
- [36] Soltis DE e Soltis PS (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. Em: Soltis DE, Soltis PM, Doyle JJ (eds) *Molecular systematics of plants. II. DNA sequencing*. Kluwer Academic Publishers, Dordrecht, pp 21–24.

- 
- [37] Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S e Little DP (2011) Commercial teas highlight plant DNA barcoding successes and challenges. *Sci Reports* 1: 42.
- [38] Vijayan K e Tsou CH (2010) DNA barcoding in plants: taxonomy in a new perspective. *Current Science*. 99 (11): 1530-1541.
- [39] Wolfe KH, Li WH e Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* 84: 9054–9058.
- [40] Yu J, Xue JH e Zhou SL (2011) New universal *matK* primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution* 49: 176–181.
- [41] Zurawski G, Perrot B, Bottomley W. e Whitefield PR (1981) The structure of the gene for the large subunit of ribulose-1,5 -bisphosphate carboxylase from spinach chloroplast DNA. *Nucleic Acids Res* 9: 3251–3270.