

Universidade do Algarve

FCT – Faculdade de Ciências e Tecnologia

Biotechnologia (1º Ciclo)



## Relatório de estágio



### **Identificação de marcadores SSR e sua utilização na caracterização molecular de germoplasma de rúcula selvagem (*Diplotaxis tenuifolia*)**

**Estagiário:** João Pedro Martins dos Reis (nº 61179)

**Orientador:** Prof. Doutor José Manuel P. T. Leitão

(31 de Julho de 2020, Faro)

## Agradecimentos

Gostaria de expressar o meu agradecimento a todas as pessoas que contribuíram direta e indiretamente para a realização deste trabalho.

- Ao meu orientador, Professor Doutor José Leitão, pelo apoio, disponibilidade e todos os contributos na aquisição de conhecimentos essenciais para a realização deste trabalho;

- Aos meus pais, que me apoiaram e incentivaram diariamente a dar o melhor de mim, dizendo sempre para seguir o caminho que me trouxesse mais felicidade;

- Aos meus amigos, em especial ao Xavier Mestre, à Ana Sofia, à Ana Marques e ao João Inácio por todo o apoio e carinho manifestado ao longo dos últimos três anos;

A todos, o meu mais sincero obrigado!

---

(João Reis, 31 de Julho de 2020, Faro)

Este trabalho foi financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia. I.P., no âmbito do projeto Ref<sup>a</sup> PTDC/ASP-PLA/28963/2017.



## Resumo

A rúcula selvagem (*Diplotaxis tenuifolia* (L.) DC), é uma espécie herbácea, natural da região Mediterrânica e da Ásia Ocidental. A caracterização genotípica desta espécie é um ponto de partida para o estabelecimento de uma coleção de germoplasma e na identificação inequívoca da diversidade dos acessos em coleção. Este trabalho, teve como objetivo, a análise genotípica de quarenta acessos da coleção de *D. tenuifolia*, no quadro do projeto de investigação REMIRúcula. Devido às limitações impostas pelo COVID-19 à realização da parte prática deste trabalho, esta ficou limitada à extração, análise qualitativa, quantitativa e de amplificabilidade do DNA genómico de 40 acessos. A pesquisa dos dados genómicos lançados recentemente pela Universidade do Algarve nas bases de dados genómicos (NCBI), permitiu selecionar 20 *loci* microssatélite de *D. tenuifolia* e desenhar e obter primers para a sua utilização como marcadores SSR (Single Sequence Repeats). A análise qualitativa (degradação e contaminação com RNA) do DNA extraído foi feita por eletroforese em gel de agarose. A análise quantitativa por espectrometria-UV (Nanodrop) e a sua amplificabilidade testada por análise RAPD-PCR. Os primers desenhados para os marcadores SSR serão, a breve trecho, testados para a sua capacidade de revelar polimorfismos. Os marcadores selecionados serão utilizados para a análise transversal de toda a coleção de germoplasma (150 acessos) permitindo revelar as relações genéticas entre os acessos de *D. tenuifolia*, e entre estes e os acessos de outras espécies que também integram a coleção (*Eruca sativa*, *Diplotaxis catholica* e *Diplotaxis muralis*). Estes marcadores serão também utilizados na localização, identificação e clonagem de genes maiores (ou de QTL) envolvidos na resistência ao míldio, causado pelo oomiceta *Hyaloperonospora sp.*, objetivo central do supracitado projeto de investigação.

**Palavras-chave:** *Diplotaxis tenuifolia* · marcadores moleculares · marcadores SSR · recursos genéticos vegetais

# Índice

Agradecimentos .....	ii
Resumo .....	iii
Índice .....	iv
Lista das abreviaturas .....	v
Introdução .....	1
Revisão Bibliográfica .....	2
Materiais e Métodos .....	13
Resultados e discussão .....	16
Conclusão .....	19
Bibliografia .....	22
Anexos .....	27

## **Lista de abreviaturas**

**DNA** – “Desoxirribonucleic Acid”

**dNTPs** – “Deoxynucleotide Triphosphates”

**EDTA** – “Ethylenediamine tetraacetic acid”

**NCBI** – “National Center for Biotechnology Information”

**NGS** – “Next Generation Sequence”

**PCR** – “Polymerase Chain Reaction”

**QTL** – “Quantitative trait *loci*”

**RAPD** – “Random Amplified Polymorphic DNA”

**RNA** – “Ribonucleic Acid”

**RNase** – “Ribonuclease”

**rpm** – “Rotações por minuto”

**SDS** – “Sodium docedyl sulfate”

**SSR** – “Simple Sequence Repeats”

**SNP** – “Single Nucleotide Polymorphism”

**Tris** – “Tris(hydroxymethyl)aminomethane”

**UV** – “Ultravioleta”

## Introdução

A família Brassicaceae, engloba múltiplos géneros de plantas de elevado valor económico. Nesta família podemos encontrar espécies como a *Brassica oleracea* (L.) que pertencem todas as variedades de couve, e a *Brassica napus* (L.) comumente conhecida como colza ou canola, uma das maiores fontes mundiais de óleo vegetal.

Nesta família encontramos também a *Diplotaxis tenuifolia*, conhecida como rúcula selvagem, uma espécie herbácea, natural da região Mediterrânica e da Ásia Ocidental, amplamente utilizada atualmente na produção de saladas prontas para consumo cuja procura tem vindo a aumentar exponencialmente. O estabelecimento no país de uma coleção de germoplasma desta espécie, a identificação de acessos com forte resistência ao míldio (*Hyaloperonospora sp.*), doença que tem vindo a causar intensas perdas aos produtores nacionais e produtores por todo o mundo, e a identificação inequívoca dos diferentes acessos em coleção, são alguns dos objetivos centrais do projeto REMIRúcula, que congrega equipas do INIAV (Oeiras), da Universidade do Algarve, ITQB (Universidade Nova de Lisboa) e do Banco Português de Germoplasma Vegetal (Braga).

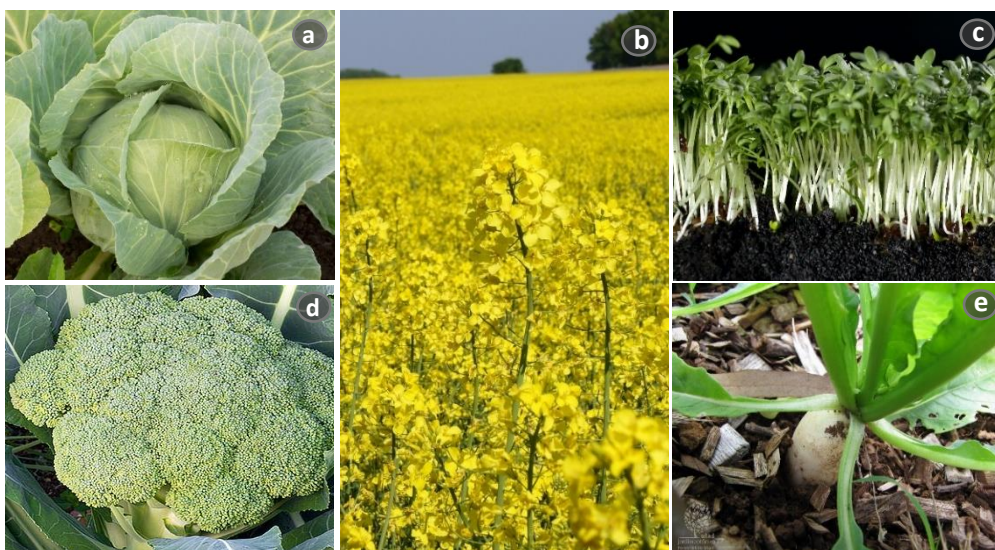
Este trabalho, interrompido pelo período de confinamento e outros períodos de contenção do COVID-19 tinha como objetivo a identificação inequívoca de 40 acessos da coleção de germoplasma de *Diplotaxis tenuifolia*, utilizando marcadores microssatélite (ou SSR).

O trabalho laboratorial, ficou limitado à extração do DNA dos referidos acessos e à sua avaliação qualitativa (eletroforese em gel de agarose) e quantitativa (espectrofotometria UV), e à sua amplificabilidade (análise RAPD-PCR).

Entretanto, a obtenção pelo Laboratório de Genómica e Melhoramento Genético de ampla informação genómica sobre a *D. tenuifolia* via next generation sequencing (Illumina), permitiu a identificação de centenas de loci microssatélite que foram pesquisados para desenhar primers para um primeiro conjunto de 20 marcadores SSR, que já foram obtidos e que urge testar.

## Revisão Bibliográfica

A família *Brassicaceae* (Fig. 1), é uma das principais e mais importantes famílias de plantas, englobando aproximadamente, cerca de 3700 espécies de 338 gêneros, entre as quais o gênero *Brassica* no qual estão incluídas espécies comuns na alimentação humana e, em particular, na dieta Mediterrânea. Neste gênero estão englobados todos os tipos de couve, nabos e outras espécies, como a rúcula (*Eruca sativa*) e a rúcula selvagem (*Diplotaxis tenuifolia*) atualmente muito populares para consumo em saladas, e espécies como a colza (*Brassica napus*), uma das principais fontes de óleo vegetal a nível mundial (EIP-AGRI Service Point, 2017; Petruzzello, 2020)



**Figura 1.** Imagens de várias espécies pertencentes à família Brassicaceae: a) Repolho - *Brassica oleracea* (L.) var. capitata L.; b) Colza - *Brassica napus* (L.) var. napus; c) Agrião - *Nasturtium officinale* (R.) B.; d) Brócolo - *Brassica oleracea* (L.) var. italica Plenck; e) Rabanete - *Raphanus sativus* (L.).

Neste gênero estão incluídos também vários genótipos de *Brassica oleracea*, utilizados como ornamentais e a *Arabidopsis thaliana* (Fig. 2), a espécie modelo mais utilizada em estudos sobre plantas (Couvreur, *et al.*, 2009) e o primeiro organismo eucariota a ter o seu genoma completamente sequenciado (The *Arabidopsis* Genome Initiative, 2000).

## *Diplotaxis tenuifolia*

A *Diplotaxis tenuifolia* (L.) DC (Fig 3), uma espécie herbácea muito rica em compostos nutricionais, popularmente conhecida como rúcula selvagem (wild rocket) originária da região Mediterrânica e da Ásia Ocidental, é o objeto central deste trabalho.

Um aspeto particular no cultivo desta espécie é a frequente confusão com a rúcula comum (*Eruca sativa*) (Fig. 4), com a qual a *D. tenuifolia* apresenta apenas ténues diferenças morfológicas ao nível das folhas, que são em geral, dependendo das variedades, mais estreitas e mais pontiagudas (Fig. 3).

Uma simples degustação das folhas das plantas pode fornecer alguma indicação sobre a espécie, pelo sabor mais intenso e adstringente das folhas da



**Figura 2.** *Arabidopsis thaliana* L.

*Diplotaxis*, característica mais ou menos pronunciada em diferentes variedades. No entanto, na altura da floração, a diferença torna-se evidente por apresentar a *D. tenuifolia* flores amarelas, característica claramente contrastante com as flores brancas da *E. sativa*.

Com um aumento na procura do mercado das saladas prontas para consumo, as hortaliças “baby leaf” desempenham um papel cada vez mais importante na dieta humana e, conseqüentemente, uma crescente importância na economia nacional, europeia e mundial (Nicoletti, *et al.*, 2007). Nestas circunstâncias, os ataques de pragas e doenças a esta cultura passaram a ter maior impacto na economia em geral e provocando repercussões em múltiplas empresas produtoras.

Este trabalho enquadra-se no projeto REMIRúcula que tem como promotor o INIAV (Instituto Nacional de Investigação Agrária e Veterinária), co-promotor a Universidade do Algarve e outros participantes, o ITQB e o Banco Português de Germoplasma Vegetal.

Este projeto tem como principal objetivo, “contribuir para uma maior sustentabilidade da produção de rúcula selvagem resistente ao míldio, garantindo alimentos mais seguros e saudáveis e a proteção ambiental” e, simultaneamente, impulsionar “o conhecimento, a prospeção e a valorização dos recursos genéticos



vegetais, com uma forte componente de investigação e inovação”. Através do estabelecimento de uma coleção de diferentes variedades de rúcula e a identificação de genótipos resistentes ao míldio (*Hyaloperonospora sp.*) este projeto promove a seleção de genótipos de interesse e aumentar a sustentabilidade da produção de rúcula selvagem (INIAV, 2019).



**Figura 3.** Plantas de *Diplotaxis tenuifolia* (L.) DC. Notem-se a as flores amarelas e as suas folhas, em geral, mais estreitas e pontiagudas .



**Figura 4.** Plantas de *Eruca sativa* Mill., A cor das folhas (brancas) é a característica pela qual, mais facilmente, esta espécie pode diferenciada da *D. tenuifolia*.

A preservação de uma agricultura sustentável e a conservação da diversidade vegetal em geral, bem como das espécies cultivadas em particular, são constantemente postas em causa por múltiplos fatores, como o crescimento contínuo da população mundial, a deflorestação, a urbanização, o uso de herbicidas, de múltiplos fitofármacos altamente poluentes e também pela utilização generalizada de um número reduzido de variedades de elite frequentemente geneticamente muito próximas entre si (Hoisington, et al., 1999; Nanduri, 2003; Liang, *et al.*, 2008; Boutin, *et al.*, 2014).

O fato da taxa de crescimento populacional ser superior à taxa de crescimento da área agrícola utilizada, fez decrescer, entre 1992 e 2016, a área de terreno cultivável per capita de 0,258 hectares para 0,192 hectares. No mesmo período, a população mundial cresceu 36% evoluindo de 5,45 mil milhões de pessoas para cerca de 7,42 mil milhões de pessoas, enquanto o terreno cultivável passou de 10,928% da área terrestre total, para apenas 11,06%. (The World Bank, (n.d); FAO, 2019).

Esta situação é particularmente preocupante em países subdesenvolvidos, cuja taxa de natalidade, normalmente muito elevada, requer o aumento significativo da produção agrícola, algo difícil de ser atingido nestes países (Bruinsma, 2003).

Além da introdução de novas tecnologias e de metodologias mais eficientes, a solução para o aumento da produção agrícola, passa pela utilização de variedades agrícolas altamente produtivas, frequentemente produzidas a partir de um “pool” genético de elite, normalmente muito reduzido no que se refere à sua variabilidade genética. Essa situação, leva a que uma larga diversidade de genótipos, em particular de variedades tradicionais locais, sejam negligenciadas, o que conduz à diminuição da diversidade genética das espécies cultivadas (National Research Council, 1993; Fu, 2006).

Por outro lado, a uniformidade genética das culturas constitui um problema que, em alguns casos, pode tomar proporções graves. Um exemplo clássico é o do ataque de *Phytophthora infestans* que se verificou na cultura da batata, em meados do século 19 na Europa, e que levou à morte por fome de aproximadamente 1 milhão de pessoas e à emigração de outros 2 milhões. Este problema foi particularmente sentido na Irlanda onde era predominantemente cultivada uma única variedade de batata (Irish Lumper), altamente suscetível a este patógeno (O’ Grada, 1995; Rao & Hodgkin, 2002).

## Conservação dos recursos genéticos.

A solução para o problema da erosão da variabilidade genética das espécies vegetais e em particular das espécies cultivadas, passa pela conservação de recursos genéticos que, aliada ao melhoramento genético de plantas, permite conservar a variabilidade genética existente, aumentar a produtividade das culturas e a segurança alimentar (Hoisington, et al., 1999; Nanduri, 2003).

A conservação de recursos genéticos vegetais é, em termos gerais, efetuada de dois modos diferentes: *in situ* ou *ex situ*.

Na conservação *in situ*, os recursos genéticos são mantidos no seu habitat de origem, por exemplo em parques naturais, uma forma de conservação eficiente em populações de espécies selvagens ou, quando se trata de espécies cultivadas, mantidas pelos agricultores que as cultivam tradicionalmente.

Esta última forma de conservação *in situ*, também designada como conservação “on-farm”, constitui um método dinâmico que permite a manutenção contínua das variedades tradicionais e a continuidade na evolução das culturas. No entanto, este tipo de conservação, apresenta alguns problemas, principalmente devido á complexidade das relações das instituições, que conservam recursos genéticos com agricultores locais, e a diversos fatores, socioculturais e económicos (Dullo *et al.*, 2010; Sthapit *et al.*, 2010). Em muitas situações, é necessário providenciar condições para que os agricultores possam ser ressarcidos das perdas provocadas pelo cultivo de variedades tradicionais, normalmente menos uniformes e produtivas que as variedades modernas. Isto é particularmente importante no caso de agricultores que pretendam vender a sua produção a grandes superfícies comerciais, que normalmente, exigem produtos agrícolas em grandes partidas e altamente uniformizados.

A conservação *ex-situ*, consiste na identificação e coleta de acessos em vários locais que são, em seguida, trasladados para bancos de germoplasma especializados onde, de acordo com as características das espécies e em particular a sua biologia reprodutiva, são mantidas sob a forma de sementes, coleções de campo e, em alguns casos, sob a forma de cultura *in vitro* (Astley, 1992; Nanduri, 2003).

Dada a quantidade de variedades agrícolas cujo produto final é a semente (grão) ou são multiplicadas por semente, a conservação de recursos genéticos mais comum é sob a forma de sementes.

Um dos obstáculos à conservação em bancos de sementes é a dificuldade na preservação de sementes recalcitrantes, que perdem a sua viabilidade no caso de passarem por processos intensos de congelação e/ou secagem. A preservação destes genótipos pode ser efetuada através da cultura de tecidos e de plantas *in vitro* (Tausch *et al.*, 2019; Vitis *et al.*, 2020) ou por criopreservação, que permite o armazenamento de germoplasma a longo prazo, estimulando a redução das atividades metabólicas do explante, simultaneamente diminuindo, todas as atividades celulares (Villalobos & Engelmann, 1995; Streczynski *et al.*, 2019).

Os acessos constantes das coleções de germoplasma podem ser intercambiados entre coleções e ser fornecidos para execução de vários estudos, em particular a empresas privadas ou a instituições públicas dedicadas ao melhoramento genético de plantas (Schreinemachers *et al.*, 2014).

Praticamente em todos os países existem instituições cujo objetivo principal é a recolha, preservação e utilização de recursos genéticos locais.

A nível mundial, existem várias organizações, cujo foco principal, é promover a acessibilidade e a qualidade dos alimentos.

A Food and Agriculture Organization of the United Nations (FAO), uma organização intergovernamental, apoia todos os estados membros e as suas instituições, no desenvolvimento de estratégias para a implementação de uma agricultura mais sustentável.

A FAO apoia diretamente institutos de investigação como o Consultative Group for International Agricultural Research (CGIAR), uma parceria entre 15 centros de investigação agrícola de excelência espalhados por todos os continentes, responsável por programas de investigação, financiados por vários países e organizações, com o objetivo principal de fazer avançar o conhecimento científico e inovação agrícola, para atingir as metas de desenvolvimento sustentável, estabelecidas pelas Nações Unidas. A esta organização estão associados 11 bancos de germoplasma internacionais, contendo mais de 700 000 acessos, distribuídos por todo o mundo (CGIAR, 2020).

Um dos principais centros de investigação membro do CGIAR é a Biodiversity International, sediada em Roma, Itália, cuja atividade se baseia no princípio de que a conservação e utilização de biodiversidade agrícola é globalmente importante para alcançar sustentabilidade do planeta e assegurar a nutrição da população.

Em Portugal, a coleção nacional de germoplasma é gerida pelo Banco Português de Germoplasma Vegetal (BPGV), abrigando cerca de 44 500 acessos de origem nacional e internacional (Barata, *et al.*, 2017). Na região do Algarve, o organismo responsável pela gestão regional de germoplasma é a DRAP Algarve, que mantém coleções de germoplasma de diversas espécies fruteiras (citrinos, alfarrobeiras, amendoeiras e figueiras, etc.) assim como, de castas de videira regionalmente importantes.

## **Importância dos recursos genéticos vegetais**

A preservação de recursos genéticos é determinante para o melhoramento de plantas, que partilha em pé de igualdade com o desenvolvimento tecnológico a responsabilidade do aumento vertiginoso da produção agrícola que se tem vindo a verificar no último século e meio e que constitui a base para a segurança alimentar mundial.

Baseado no conhecimento científico provido pela Genética e pelas novas ciências OMICs, o melhoramento de plantas cria novas variedades, em geral mais produtivas, mais bem adaptadas às condições específicas de cada região e às exigências das diferentes comunidades populacionais, e mais tolerantes a stresses bióticos e abióticos.

Na maioria dos casos, a solução para enfrentar novas situações de stresse abiótico e abiótico só pode ser encontrada nos recursos genéticos pré-existentes e mantidos em coleções de germoplasma (Singh, *et al.*, 2014).

A evolução da conservação de recursos genéticos vegetais, ditou ao longo dos anos, a necessidade de serem criados três principais tipos de coleções de germoplasma: coleções base, coleções ativas e coleções de trabalho.

Na coleção base, os acessos são guardados a longo prazo, a baixas temperaturas (-10° a -20°C) ou por vezes criopreservados (-150° a -196°C). A coleção base é, na maior parte do tempo, estática, com as sementes mantidas indefinidamente. No entanto, de tempos a tempos, de acordo com a espécie vegetal os acessos tem de ser avaliados para

verificar e refrescar a viabilidade do material conservado, e eliminar o material já não viável (National Research Council, 1993; Ruiz *et al.*, 1999).

Na coleção ativa, os acessos são conservados a temperaturas de -4°C. Tratam-se de coleções a curto prazo, onde o germoplasma se encontra em condições de ser distribuído para utilização em programas de melhoramento de plantas ou outros usos. (National Research Council, 1993; Ruiz *et al.*, 1999).

A coleção de trabalho, é constituído pelos acessos utilizados, principalmente, em programas de melhoramento de plantas. (National Research Council, 1993; Engels & Visser, 2003).

## **Marcadores moleculares**

Nas plantas, assim como em todos os organismos eucariotas, existem características fenotípicas que permitem a sua distinção. Algumas destas características podem permitir diferenciar plantas individuais no interior da mesma espécie. No entanto, quando se trata de identificar e discriminar plantas em circunstâncias particulares, em que essas características fenotípicas não estão presentes ou podem ser alteradas (por exemplo, pelas condições edafoclimáticas) ou queremos estabelecer as relações genéticas entre indivíduos, populações, espécies, ou outro nível taxonómico, as características fenotípicas passam a ser incapazes de fornecer resultados fiáveis.

Para ultrapassar essas dificuldades na identificação e criar uma inequívoca caracterização dos genótipos, recorre-se a métodos de caracterização genotípica, em que os mais utilizados são os marcadores-DNA.

Os marcadores-DNA, são bastante eficientes na caracterização da variabilidade genética de algumas espécies, como foi efetuado anteriormente em espécies de *B. juncea* (Vinu *et al.*, 2013), de *B. oleracea* var. *capitata* (Faltusová *et al.*, 2011) (Fig. 1., a)), ou em géneros como *Lupinus* spp. (Talhinhas, *et al.*, 2003).

São normalmente utilizados na localização e identificação de alguns genes de interesse, importantes em programas de melhoramento de plantas (Winter & Kahl, 1995; Mohan *et al.*, 1997; Rao & Hodgkin, 2002).

O uso de técnicas de genotipagem, é essencial na proteção de propriedade intelectual de genótipos melhorados. A utilização de perfis de DNA, com base no uso de marcadores, permite identificar inequivocamente um determinado genótipo e permite salvaguardar de forma eficaz a propriedade intelectual que sobre ele penda (Castellana *et al.*, 2020).

Para as organizações e empresas que se dedicam ao melhoramento de plantas ou grandes empresas de produção e comercialização de sementes, um aspecto muito significativo é o grau de pureza das mesmas.

A aplicação das análises genóticas na avaliação da pureza genética das variedades é um aspecto fundamental da utilização de marcadores moleculares, dado que estes (assim como o DNA genómico), ao contrário das características morfológicas não é influenciado por fatores ambientais (Krishna *et al.*, 2020; Zhang *et al.*, 2020).

Os marcadores moleculares são atualmente fundamentais na gestão dos bancos de germoplasma, ao permitir o registo e a identificação inequívoca dos acessos, avaliar a variabilidade genética existente nas coleções, estudar as relações genéticas entre os acessos e evitar a manutenção excessiva de genótipos iguais (Epperson *et al.*, 1997; Phippen *et al.*, 1997).

A utilização dos marcadores moleculares na identificação de acessos duplicados, permite rentabilizar o espaço e diminuir os custos de manutenção das coleções (Albuquerque *et al.*, 2019; Singh *et al.*, 2019).

A utilização de técnicas de análise genotípica é também usual na resolução de problemas de homonímia (uma denominação para vários genótipos distintos) e problemas de sinonímia (várias denominações para um único genótipo) (Muzzalupo *et al.*, 2014).

Uma ação importante para evitar o excesso de acessos conservados é a criação de coleções nucleares que abrigam a maior quantidade possível de diversidade genética de uma determinada espécie ou população, reduzindo ao mínimo a redundância genética no interior da coleção (Belaj *et al.*, 2012).

Após o desenvolvimento da técnica de Polymerase Chain Reaction (PCR), começaram imediatamente a ser desenvolvidos marcadores-DNA baseados nesta técnica, substituindo os até então utilizados marcadores-DNA baseados na técnica de hibridação

molecular: RFLPs (Restriction Fragment Length Polymorphisms) e VNTRs (Variable Number of Tandem Repeats).

Técnicas como a Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Inter-Simple Sequence Repeats (ISSR), e Simple Sequence Repeat (SSR) foram desenvolvidas nos últimos anos e são ainda amplamente utilizadas em análises genéticas.

Os marcadores RAPD foram os primeiros a ser desenvolvidos. (Welsh and McClelland, 1990; Williams *et al.*, 1990). Utilizando primers arbitrários que amplificam fragmentos aleatórios, estes marcadores são bastante populares devido à sua simplicidade e à sua capacidade de originar uma grande quantidade de produtos de amplificação, o que os torna bastante úteis em muitos tipos de análise, como a caracterização de diversidade genética (Talhinhas, *et al.*, 2003), a avaliação de similaridade genética de cultivares (Goulão *et al.*, 2001), na identificação de espécies e subespécies (Bardakci & Skibinski, 1994) e identificação de cultivares (Hu & Quiros, 1991). No laboratório de Genómica e Melhoramento Genético da FCT, UAlg, após muitos anos de utilização em larga panóplia de estudos, a utilização destes marcadores está praticamente reduzida à avaliação da amplificabilidade de novas amostras de DNA.

Outro tipo de marcadores amplamente utilizados em ensaios genotípicos, são os marcadores SSR, também conhecidos como microssatélites. Os microssatélites são constituídos por pequenas repetições em tandem de motivos com cerca de 2-6 bp, localizadas em locais específicos ao longo dos genomas, com um número de repetições extremamente variáveis de genoma para genoma, pelo que estes marcadores são normalmente muito polimórficos (Feng, *et al.*, 2016). No entanto, tratando-se de marcadores específicos (não-aleatórios) requerem o conhecimento prévio de sequências genómicas que contenham sequências microssatélite, algo atualmente de relativamente fácil execução com recurso à sequenciação massiva paralela (NGS).

Os marcadores SSR apresentam, normalmente, elevada reprodutibilidade e co-dominância, o que os torna muito utilizados nas mais variadas análises genómicas, como a análise de relações genéticas (Cabrita, *et al.*, 2014), a construção de mapas genéticos (Farinhó, *et al.*, 2004; Wang, *et al.*, 2017) e na identificação de cultivares (Ercisli, *et al.*, 2011) e discriminação de indivíduos (Powell, *et al.*, 1996).



## **NGS e genotyping by sequencing**

O desenvolvimento de técnicas de sequenciação de nova geração (NGS), possibilitou, em comparação aos métodos comuns, uma maior rapidez em análises genómicas. Apesar da rapidez de sequenciação ter aumentado, a inovação nas técnicas de sequenciação, geraram também uma redução nos custos gerais das análises genómicas (Poland & Rife, 2012).

A utilização de NGS possibilita a sequenciação total do genoma, no entanto, com o constante desenvolvimento de novas ferramentas bioinformáticas, é possível a utilização da NGS também para a identificação de Single Nucleotide Polymorphisms (SNP), também chamada de genotipagem por sequenciação. A identificação de SNP pode ser considerada vantajosa em relação a outros tipos de marcadores moleculares, pois estes aparecem em frequências relativamente elevadas, o que pode ser bastante positivo nas análises genóticas de populações, ou até na construção de mapas genéticos de espécies vegetais (Poland & Rife, 2012; Ting *et al.*, 2014; Chung *et al.*, 2017).

## **Materiais e Métodos**

### **Extração de DNA genômico**

Extraíu-se DNA de 40 diferentes acessos (**Anexo 1**) da coleção de germoplasma de rúcula selvagem que se encontra em construção no INIAV no âmbito do projeto REMI Rúcula.

As folhas de pelo menos 3 plantas de cada uma das variedades foram removidas, enxaguadas com água da torneira e secas com papel absorvente, para remover o excesso de água. A nervura central das folhas foi removida com o auxílio de um bisturi e a restante massa foliar foi macerada num almofariz contendo azoto líquido, até ser obtido um pó muito fino.

Parte do macerado foi colocado num tubo eppendorf, contendo 500 µL tampão de extração de DNA (250 mM Tris-HCL pH 8.0, 25 mM EDTA, 1% SDS) até que o volume total atingisse, aproximadamente, os 750 µL. Em seguida adicionou-se RNase A (20 µg/ml) e incubou-se em banho maria a 65°C durante 15 minutos.

Já à temperatura ambiente, adicionou-se 1 volume de fenol: clorofórmio: álcool isoamílico (25:24:1) e após agitação manual durante, pelo menos, 1 minuto a amostra foi centrifugada em microcentrifuga a 13 000 rpm durante 3 minutos.

A fase superior foi recolhida para novo tubo e após adição de 1 volume de clorofórmio: álcool isoamílico (24:1) e agitação manual durante 1 minuto a amostra foi novamente centrifugada a 13 000 rpm durante 3 minutos. A extração com 1 volume de clorofórmio: álcool isoamílico (24:1) foi repetida, pelo menos duas vezes, até se obter uma interfase completamente clara e limpa.

Após a última centrifugação, adicionaram-se 3 volumes de etanol absoluto à fase superior coletada para um novo tubo eppendorf e, após agitação para revelar o DNA precipitado, a amostra foi armazenada a -20°C.

No dia seguinte, a solução armazenada foi então centrifugada a 13 000 rpm durante 10 minutos. O sobrenadante foi decantado e pequenas gotículas residuais nas paredes do tubo removidas com micropipeta. O pellet contendo DNA foi deixado secar à temperatura ambiente durante 2 horas, sendo posteriormente solubilizado em 50 µL de tampão TE<sub>0.1</sub> (10 mM Tris; 0.1 mM EDTA).

### **Avaliação qualitativa do DNA genómico extraído.**

A integridade do DNA extraído e o nível de contaminação com RNA foram verificados, submetendo as amostras a eletroforese em gel de agarose (1,4%), flanqueadas pelo marcador de DNA “DNA Ladder III” (NZYTECH).

A avaliação qualitativa da amplificabilidade, das amostras de DNA foi efetuada através da análise RAPD com o primer OPE-03 da Operon Technologies Inc.

As reações RAPD-PCR foram realizadas num volume final de 15 µL, contendo 10 ng de DNA genómico, com 1x de Gel Load Reaction Buffer, 0,16 mM de dNTPs, 0,6 U de NZYTaq DNA Polimerase, 1,33 µM de primer OPE-03 e água Milli-Q até perfazer o volume. A reação PCR foi efetuada num Termociclador Biometra UNO II (Thermoblock, Biotron) como a seguir descrito: desnaturação inicial a 94°C durante 1 min. e 30 seg., seguido de 35 ciclos de 30seg. a 94°C, 30seg. a 36°C, 1 min a 72°C e um ciclo final de 10 min. a 72°C. Os produtos de PCR foram separados por eletroforese em gel de agarose (2%).

Os géis de agarose foram após término da eletroforese, submersos numa solução de brometo de etídio, analisados com um transiluminador UV e fotografados com uma câmara digital Canon EOS 1300D

### **Avaliação quantitativa do DNA genómico extraído.**

A avaliação quantitativa do DNA foi efetuada por espectrofotometria em NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific), utilizando o tampão Tris-EDTA 0.1 mM como branco. Após a quantificação, as amostras foram diluídas em água Milli-Q para a concentração apropriada (2.5 ng/ µL) para serem utilizadas como soluções de trabalho (4 µL em cada reação PCR).

### **Análises dos acessos da coleção por marcadores SSR.**

Recentemente, o Laboratório de Genómica e Melhoramento Genético adicionou um novo “Bioproject”: “*Diploaxis tenuifolia* isolate: DT550 Genome sequencing

(PRJNA624903)” e sequências de 500 loci de microssatélites (acessões MT317453 a MT317952) às bases de dados genômicos (NCBI).

Entre as 500 sequências de loci de microssatélites, foram selecionados 20 loci para desenvolvimento de marcadores microssatélite ou SSR (**Anexo 2**).

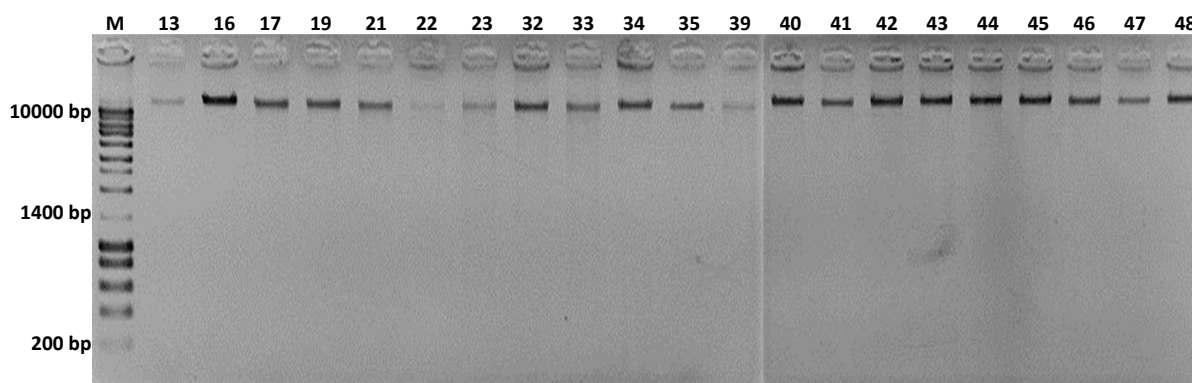
Os “primers” para amplificação dos 20 marcadores microssatélites foram desenhados, com apoio do programa Fast-PCR, de forma a possuírem temperaturas de “melting” muito próximas de 50°C e conteúdo em GC superiores a 40%. A sequência específica de cada primer foi adicionalmente validada contra os dados primitivos da sequenciação massiva paralela (NGS) utilizando o software Tablet 1.19.09.03.

A síntese dos 40 primers (**Anexo 3**), foi entretanto solicitada à empresa NZYTech Lda.

## Resultados e Discussão

### Análise Qualitativa do DNA

A análise por eletroforese em gel de agarose do DNA genômico extraído dos vários acessos, demonstrou que o DNA se encontrava em ótimas condições em termos de não-degradação e sem contaminação clara com RNA (Figura 5).

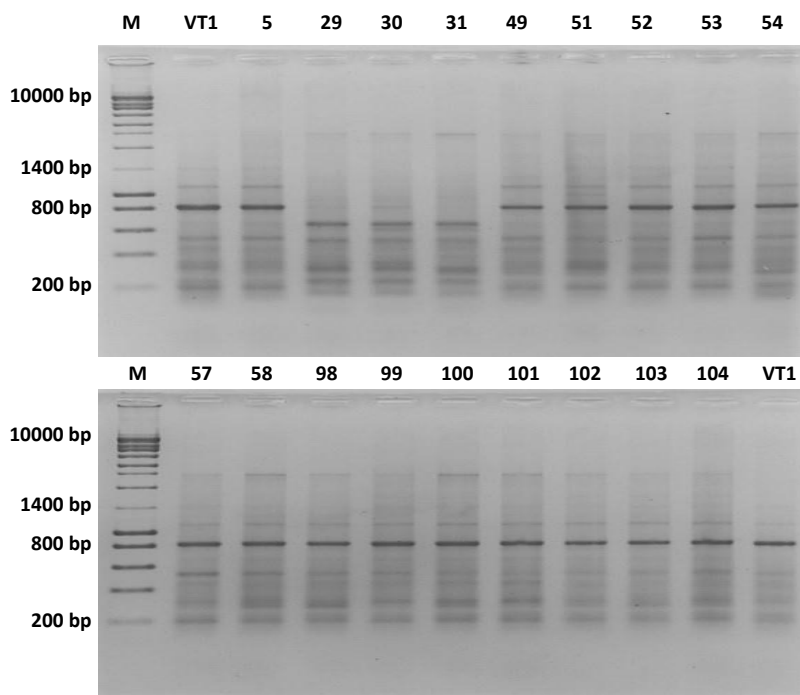


**Figura 5.** Análise qualitativa do DNA genômico extraído de *Diplotaxis tenuifolia*. A existência de uma única banda de DNA no gel de agarose e a ausência de arrastamentos pronunciados imediatamente abaixo, indicam que a DNA extraído dos vários acessos se encontra relativamente intacto. A ausência de manchas de arrastamentos de ácidos nucleicos na segunda metade do gel são um bom testemunho da boa atividade da RNase\_A adicionada durante o processo de extração do DNA e da praticamente ausente contaminação com RNA.

Na Figura 5, é possível também verificar que a quantidade de DNA extraído foi diferente de acesso para acesso, uma consequência direta das quantidades de massa foliar utilizada na extração de cada amostra.

### Análise RAPD

O acesso VT1 (**Anexo 1**), amavelmente cedido pela empresa Vitacress, foi usado como referência (controle) da espécie *D. tenuifolia*. A avaliação da amplificabilidade do DNA por análises RAPD permitiu identificar três acessos (29, 30 e 31) bastante distintos do acesso de referência VT1 e de todos os outros acessos que, entre si, apresentam ligeiras diferenças (Fig. 6). Estes 3 acessos, enviados de Itália, foram mais tarde confirmados, pela cor branca das flores, como sendo de *Eruca sativa*, totalmente resistentes aos isolados de *Hyaloperonospora sp.* (míldio) obtidos em plantas de *Diplotaxis* fortemente infetadas.



**Figura 6.** Gel de agarose (2%). Perfis de amplificação com o primer RAPD OPE-03. M - Marcador NZYTech Ladder III. Note-se a similaridade dos padrões de amplificação RAPD do acesso VT1 e da larga maioria dos acessos, exceto 3 (acessos 29,30,31) que foram posteriormente identificados como sendo de *Eruca sativa*.

## Análise Bioinformática

Um programa recente de sequenciação (NGS – Illumina) de DNA extraído de núcleos celulares de *Diplotaxis tenuifolia* parcialmente purificados, realizado pela Universidade do Algarve no âmbito do projeto REMIRúcula em colaboração com o INIAV, resultou em 105,7 M sequências que agregam 15.3 Gbp (Fig. 7). (<https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11802588>),

Runs: 1 run, 105.7M spots, 15.3G bases, 4.8Gb

Run	# of Spots	# of Bases	Size	Published
<a href="https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11802588">SRR11802588</a>	105,745,217	15.3G	4.8Gb	2020-05-16

**Figura 7.** Dados no NCBI da sequenciação de *Diplotaxis tenuifolia*.

Estas sequências estão agrupadas em 167 140 contigs (ainda não publicados) dos quais foram extraídos 500 loci microssatélite, publicados em separado nas bases de dados genômicos ([www.ncbi.nlm.nih.gov/nuccore/?term=diplotaxis%20microsatellite%20](http://www.ncbi.nlm.nih.gov/nuccore/?term=diplotaxis%20microsatellite%20))

Dos 500 loci microssatélite foram escolhidos 30 para re-seleção e transformação de 25 em marcadores microssatélites (SSR).

Foram desenhados primers com temperatura de “melting” próximas dos 50°C e analisados contra os contigs resultantes da sequenciação utilizando o software Tablet 1.19.09.03 e diretamente contra as “raw sequences” publicadas como um SRA (SRR11802588), para validar as sequências dos primers.

A necessidade de uma validação aturada das sequências dos primers advém de um problema recorrente às sequências dos loci de microssatélite. É muito comum, que após passagem sobre uma sequência microssatélite, várias DNA polimerases passem a cometer múltiplos erros de síntese, e conseqüentemente erros de sequenciação, antes de regressarem a um funcionamento normal.

A análise BLAST de sequências de 25bp, que incluem as sequências dos primers que se pretende validar, contra as sequências depositadas sob a forma de SRA no NCBI permitiu também validar as sequências dos primers que se pretende mandar sintetizar (Fig. 8).

### Reads (separated)

a)

```
>gnl|SRA|SRR11802588.100003134.2 100003134 (Biological)
TAGAAATGTCGGGCCATTGATTGTCCCTAAGCTCAATTGGTAAACAGAAAGAAACCAACAA
ATACACCAAACACACACACACACACACACACACACACACACACTTAAATGTGATA
AA
```

### Reads (separated)

b)

```
>gnl|SRA|SRR11802588.95883654.2 95883654 (Biological)
ATCAGAGGGTCAACCCGCGTGACTTTGACGAAACGAAGCCCGTGTGTGTGTGTGTGTGTG
TGTGTGAGTGAAGCTGGTTAGGGTTTTGTCAGAAGGCTCGACAGAGTGTATTAGTCGTGG
TCAAGTTAGAAGAGTGTTCCTGTT
```

**Figura 8.** Dois exemplos (muito parciais) da análise BLASTn de sequências de 25bp, que incluem sequências de primers a validar, contra a base de dados SRA: a) microssatélite DT14CA18; b) microssatélite DT41GT12.

## Conclusão

Apesar do trabalho ter sido interrompido pelo período de confinamento e outros períodos, que se seguiram, de contenção do COVID-19, foi possível realizar o trabalho acima exposto. Entretanto, já se encontram sintetizados os primers necessários para amplificação dos 20 marcadores SSR selecionados, podendo ser retomada a seleção dos marcadores mais polimórficos, que permitam identificar de forma inequívoca os vários acessos da coleção de germoplasma que se encontra a ser organizada no âmbito do projeto REMI Rúcula, estudar as relações genéticas entre os vários acessos, as relações genéticas com outras espécies do género *Diplotaxis* e com outros géneros da família Brassicaceae.

Estes marcadores SSR irão também ser utilizados para a localizar, identificar e clonar um gene maior (ou QTL) que confere forte resistência ao Míldio (*Hyaloperonospora sp.*) na rúcula selvagem (*Diplotaxis tenuifolia*).



## Referências bibliográficas

- Albuquerque, H. Y., Oliveira, E. J., Brito, A. C., Andrade, L. R., Carmo, C. D., Morgante, C. V., . . . Faleiro, F. G. (2019). Identification of duplicates in cassava germplasm banks based on single-nucleotide polymorphisms (SNPs). *Scientia Agricola*, 76(4), 328-336. doi:10.1590/1678-992x-2017-0389
- Astley, D. (1992). Preservation of genetic diversity and accession integrity. *Field Crops Research*, 29(3), 205–224. doi: 10.1016/0378-4290(92)90026-6
- Barata, A. M., Gaspar, C., Rocha, F., & Lopes, V. (2017, June 8). Banco Português de Germoplasma Vegetal – 40 anos de conservação dos recursos genéticos em Portugal. *CULTIVAR*, 8, 85–90.
- Bardakci, F., & Skibinski, D. O. F. (1994). Application of the RAPD technique in tilapia fish: species and subspecies identification. *Heredity*, 73(2), 117–123. doi: 10.1038/hdy.1994.110
- Bélanger, J., & Pilling, D. (Eds.). (2019). The state of the worlds biodiversity for food and agriculture. Rome: FAO Commission on Genetic Resources for Food and Agriculture Assessments. doi: <http://www.fao.org/3/CA3129EN/CA3129EN.pdf>
- Belaj, A., Dominguez-García, M. D., Atienza, S. G., Urdíroz, N. M., Rosa, R. D., Satovic, Z., . . . Río, C. D. (2012). Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. *Tree Genetics & Genomes*, 8(2), 365-378. doi:10.1007/s11295-011-0447-6
- Boutin, C., Strandberg, B., Carpenter, D., Mathiassen, S., & Thomas, P. (2014). Herbicide impact on non-target plant reproduction: What are the toxicological and ecological implications? *Environmental Pollution*, 185, 295–306. doi: 10.1016/j.envpol.2013.10.009
- Bruinsma, J. (Ed.). (2003). World agriculture: towards 2015 - 2030. Rome: FAO.
- Cabrita, L., Apostolova, E., Neves, A., Marreiros, A., & Leitão, J. (2014). Genetic diversity assessment of the almond (*Prunus dulcis* (Mill.) D.A. Webb) traditional germplasm of

- Algarve, Portugal, using molecular markers. *Plant Genetic Resources*, 12(S1), S164–S167. doi: 10.1017/s1479262114000471
- Castellana, S., Ranzino, L., Beritognolo, I., Cherubini, M., Luneia, R., Villani, F., & Mattioni, C. (2020). Genetic characterization and molecular fingerprint of traditional Umbrian tomato (*Solanum lycopersicum* L.) landraces through SSR markers and application for varietal identification. *Genetic Resources and Crop Evolution*. doi:10.1007/s10722-020-00942-3
- CGIAR Genebank Platform. (2020, July 10). Retrieved July 20, 2020, from <https://www.cgiar.org/research/program-platform/genebank-platform/>
- Chung, Y. S., Choi, S. C., Jun, T., & Kim, C. (2017). Genotyping-by-sequencing: A promising tool for plant genetics research and breeding [Abstract]. *Horticulture, Environment, and Biotechnology*, 58(5), 425-431. doi:10.1007/s13580-017-0297-8
- Couvreur, T. L. P., Franzke, A., Al-Shehbaz, I. A., Bakker, F. T., Koch, M. A., & Mummenhoff, K. (2009). Molecular Phylogenetics, Temporal Diversification, and Principles of Evolution in the Mustard Family (Brassicaceae). *Molecular Biology and Evolution*, 27(1), 55–71. doi: 10.1093/molbev/msp202
- Dullo, M. E., Hunter, D., & Borelli, T. (2010). Ex Situ and In Situ Conservation of Agricultural Biodiversity: Major Advances and Research Needs. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 38(2), 123-135. <https://doi.org/10.15835/nbha3824878>
- EIP-AGRI Service Point. (2017). EIP-AGRI Brochure IPM for Brassica.
- Engels, J. M., & Visser, L. (2003). *A guide to effective management of germplasm collections*. Rome: International Plant Genetic Resources Institute.
- Epperson, J. E., Pachico, D. H., & Guevara, C. L. (1997). A Cost Analysis of Maintaining Cassava Plant Genetic Resources. *Crop Science*, 37(5), 1641-1649. doi:10.2135/cropsci1997.0011183x003700050039x
- Ercisli, S., Ipek, A., & Barut, E. (2011). SSR Marker-Based DNA Fingerprinting and Cultivar Identification of Olives (*Olea europaea*). *Biochemical Genetics*, 49(9-10), 555–561. doi: 10.1007/s10528-011-9430-z
- Faltusová, Z., Kučera, L., & Ovesná, J. (2011). Genetic diversity of *Brassica oleracea* var. capitata gene bank accessions assessed by AFLP. *Electronic Journal of Biotechnology*, 14(3). doi:10.2225/vol14-issue3-fulltext-4

- Farinhó, M., Coelho, P., Carlier, J., Svetleva, D., Monteiro, A., & Leitão, J. (2004). Mapping of a locus for adult plant resistance to downy mildew in broccoli (*Brassica oleracea* convar. *italica*). *Theoretical and Applied Genetics*, 109(7), 1392–1398. doi: 10.1007/s00122-004-1747-0
- Feng, S., He, R., Lu, J., Jiang, M., Shen, X., Jiang, Y., . . . Wang, H. (2016). Development of SSR Markers and Assessment of Genetic Diversity in Medicinal *Chrysanthemum morifolium* Cultivars. *Frontiers in Genetics*, 7. doi:10.3389/fgene.2016.00113
- FAO- Land Use. (2019, December 3). Retrieved July 20, 2020, from <http://www.fao.org/faostat/en/>
- Fu, Y.-B. (2006). Impact of plant breeding on genetic diversity of agricultural crops: searching for molecular evidence. *Plant Genetic Resources*, 4(1), 71–78. doi: 10.1079/pgr2006116
- Goulão, L., Cabrita, L., Oliveira, C. M., & Leitão, J. M. (2001). Comparing RAPD and AFLPTM analysis in discrimination and estimation of genetic similarities among apple (*Malus domestica* Borkh.) cultivars. *Euphytica*, 119(3), 259–270. doi: 10.1023/a:1017519920447
- He, W., Zhao, H., Yang, X., Zhang, R., & Wang, J. (2019). Patent analysis provides insights into the history of cotton molecular breeding worldwide over the last 50 years. *Journal of Integrative Agriculture*, 18(3), 539-552. doi:10.1016/s2095-3119(18)62012-x
- Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J., Skovmand, B., Taba, S., & Warburton, M. (1999). Plant genetic resources: What can they contribute toward increased crop productivity? *Proceedings of the National Academy of Sciences*, 96(11), 5937-5943. doi:10.1073/pnas.96.11.5937
- Hu, J., & Quiros, C. (1991). Identification of broccoli and cauliflower cultivars with RAPD markers. *Plant Cell Reports*, 10(10), 505–511. doi: 10.1007/bf00234583
- INIAV. (2019). Objetivos. Retrieved July 18, 2020, from <https://projects.inia.vt.pt/REMIRucula/pt/projeto/objetivos>
- Krishna, T. A., Maharajan, T., Roch, G. V., Ramakrishnan, M., Ceasar, S. A., & Ignacimuthu, S. (2020). Hybridization and hybrid detection through molecular markers in finger millet [*Eleusine coracana* (L.) Gaertn.]. *Journal of Crop Improvement*, 34(3), 335-355. doi:10.1080/15427528.2019.1709596

- Liang, Y.-Q., Li, J.-W., Li, J., & Valimaki, S. K. (2008). Impact of urbanization on plant diversity: A case study in built-up areas of Beijing. *Forestry Studies in China*, 10(3), 179–188. doi: 10.1007/s11632-008-0036-4
- Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C., & Sasaki, T. (1997). *Molecular Breeding*, 3(2), 87-103. doi:10.1023/a:1009651919792
- Muzzalupo, I., Vendramin, G. G., & Chiappetta, A. (2014). Genetic Biodiversity of Italian Olives (*Olea europaea*) Germplasm Analyzed by SSR Markers. *The Scientific World Journal*, 2014, 1-12. doi:10.1155/2014/296590
- Nanduri., K. R. (2003). Plant genetic resources: Advancing conservation and use through biotechnology. *African Journal of Biotechnology*, 3(2), 136-145. doi:10.5897/ajb2004.000-2025
- National Research Council. 1993. *Managing Global Genetic Resources: Agricultural Crop Issues and Policies*. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/2116>
- Nicoletti, Rosario & Raimo, Francesco & G, Miccio. (2007). *Diplotaxis tenuifolia*: biology, production and properties. *European Journal of Plant Science and Biotechnology*. 1. 36-43.
- O'Grada, C. (1995). Introduction. In *Great Irish famine* (p. 1). Cambridge: Cambridge Univ. Press.
- Petruzzello, M. (2020, April 09). List of plants in the family Brassicaceae. Retrieved June 23, 2020, from <https://www.britannica.com/topic/list-of-plants-in-the-family-Brassicaceae-2004620>
- Phippen, W. B., Kresovich, S., Candelas, F. G., & Mcferson, J. R. (1997). Molecular characterization can quantify and partition variation among genebank holdings: A case study with phenotypically similar accessions of *Brassica oleracea* var. *capitata* L. (cabbage) 'Golden Acre'. *Theoretical and Applied Genetics*, 94(2), 227-234. doi:10.1007/s001220050404
- Poland, J. A., & Rife, T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome*, 5(3), 92-102. doi:10.3835/plantgenome2012.05.0005
- Powell, W., Machray, G., & Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, 1(7), 215-222. doi:10.1016/s1360-1385(96)86898-0

- Rao, V. R., & Hodgkin, T. (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture*, 68(1), 1–19. doi: 10.1023/a:1013359015812
- Ruiz, M., Martín, I., & Cuadra, C. D. (1999). Cereal seed viability after 10 years of storage in active and base germplasm collections. *Field Crops Research*, 64(3), 229-236. doi:10.1016/s0378-4290(99)00044-1
- Schreinemachers, P., Ebert, A. W., & Wu, M. (2014). Costing the ex situ conservation of plant genetic resources at AVRDC – The World Vegetable Center. *Genetic Resources and Crop Evolution*, 61(4), 757-773. doi:10.1007/s10722-013-0070-5
- Singh, M., Bisht, I. S., Kumar, S., Dutta, M., Bansal, K. C., Karale, M., . . . Datta, S. K. (2014). Global Wild Annual Lens Collection: A Potential Resource for Lentil Genetic Base Broadening and Yield Enhancement. *PLoS ONE*, 9(9). doi:10.1371/journal.pone.0107781
- Singh, N., Wu, S., Raupp, W. J., Sehgal, S., Arora, S., Tiwari, V., . . . Poland, J. (2019). Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Scientific Reports*, 9(1). doi:10.1038/s41598-018-37269-0
- Sthapit, B.; Padulosi, S.; Mal, M. (2010) Role of on-farm/In situ conservation and underutilized crops in the wake of climate change. n. p. 145-156
- Streczynski, R., Clark, H., Whelehan, L. M., Ang, S., Hardstaff, L. K., Funnekotter, B., . . . Mancera, R. L. (2019). Current issues in plant cryopreservation and importance for ex situ conservation of threatened Australian native species. *Australian Journal of Botany*, 67(1), 1. doi:10.1071/bt18147
- Talhinhas, P., Neves-Martins, J., & Leitao, J. (2003). AFLP, ISSR and RAPD markers reveal high levels of genetic diversity among *Lupinus* spp. *Plant Breeding*, 122(6), 507–510. doi: 10.1111/j.1439-0523.2003.00892.x
- Tausch, S., Leipold, M., Reisch, C., & Poschlod, P. (2019). Dormancy and endosperm presence influence the ex situ conservation potential in central European calcareous grassland plants. *AoB PLANTS*, 11(4). doi:10.1093/aobpla/plz035
- The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. (2000). *Nature*, 408(6814), 796-815. doi:10.1038/35048692

- The World Bank - Arable land (hectares per person). (n.d.). Retrieved July 20, 2020, from <https://data.worldbank.org/indicator/AG.LND.ARBL.HA.PC?end=2016>
- The World Bank - Arable land (% of land area). (n.d.). Retrieved July 20, 2020, from <https://data.worldbank.org/indicator/AG.LND.ARBL.ZS?end=2016>
- The World Bank - Population, total. (n.d.). Retrieved July 20, 2020, from <https://data.worldbank.org/indicator/SP.POP.TOTL?end=2016>
- Villalobos, V. M., & Engelmann, F. (1995). Ex situ conservation of plant germplasm using biotechnology. *World Journal of Microbiology & Biotechnology*, *11*(4), 375-382. doi:10.1007/bf00364612
- Vinu, V., Singh, N., Vasudev, S., Yadava, D. K., Kumar, S., Naresh, S., Bhat, S. R., & Prabhu, K. V. (2013). Assessment of genetic diversity in Brassica juncea (Brassicaceae) genotypes using phenotypic differences and SSR markers. *Revista de biologia tropical*, *61*(4), 1919–1934.
- Vitis, M. D., Hay, F. R., Dickie, J. B., Trivedi, C., Choi, J., & Fiegner, R. (2020). Seed storage: Maintaining seed viability and vigor for restoration use. *Restoration Ecology*. doi:10.1111/rec.13174
- Wang, L., Zhang, Y., Zhu, X., Zhu, X., Li, D., Zhang, X., ... Zhang, X. (2017). Development of an SSR-based genetic map in sesame and identification of quantitative trait loci associated with charcoal rot resistance. *Scientific Reports*, *7*(1). doi: 10.1038/s41598-017-08858-2
- Wang, X. W., Kaga, A., Tomooka, N., & Vaughan, D. A. (2004). The development of SSR markers by a new method in plants and their application to gene flow studies in azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. *Theoretical and Applied Genetics*, *109*(2), 352-360. doi:10.1007/s00122-004-1634-8
- Welsh, J., & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, *18*(24), 7213-7218. doi:10.1093/nar/18.24.7213
- Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J., & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, *18*(22), 6531–6535. doi: 10.1093/nar/18.22.6531
- Winter, P., & Kahl, G. (1995). Molecular marker technologies for plant improvement. *World Journal of Microbiology & Biotechnology*, *11*(4), 438-448. doi:10.1007/bf00364619

Zhang, S., Li, B., Chen, Y., Shaibu, A. S., Zheng, H., & Sun, J. (2020). Molecular-Assisted Distinctness and Uniformity Testing Using SLAF-Sequencing Approach in Soybean. *Genes*, *11*(2), 175. doi:10.3390/genes11020175

# Anexo 1

<b>Nº acesso</b>	<b>Nome do acesso</b>	<b>Nº acesso</b>	<b>Nome do acesso</b>
<b>VT1</b>	TZ 4194 (tratada)	<b>40</b>	Wild Rocket, CN WROC 2117 UNT
<b>1</b>	wild rocket, var. BELLEZIA UNT	<b>41</b>	Wild Rocket, CN WROC 2118
<b>3</b>	wild rocket, var. LETIZIA UNT	<b>42</b>	Wild Rocket, CN WROC 2119
<b>4</b>	wild rocket, var. PRONTO UNT	<b>43</b>	Wild Rocket, CN WROC 2407 UNT
<b>6</b>	wild rocket, var. SORRENTO (tratada verde)	<b>44</b>	Wild Rocket, CN WROC 2410 UNT
<b>13</b>	CN 922	<b>45</b>	Wild Rocket, CN WROC 2422
<b>16</b>	APOLLO	<b>46</b>	Wild Rocket, CN WROC 2425 UNT
<b>17</b>	ATHENA	<b>47</b>	Wild Rocket, CN WROC 2426 UNT
<b>19</b>	TZ1138 UNT	<b>48</b>	Wild Rocket, CN WROC 2433 UNT
<b>21</b>	SINOPE	<b>49</b>	Wild Rocket, CN WROC 2441 UNT
<b>22</b>	SINOPE UNT	<b>51</b>	Wild Rocket, CN WROC 2443 UNT
<b>23</b>	89-004 UNT	<b>57</b>	Wild Rocket, CN WROC 2450 UNT
<b>29</b>	Rúcula ( <i>E. vesicaria</i> sub. sp. sativa)	<b>58</b>	Wild Rocket, CN WROC 2451 UNT
<b>30</b>	Rúcula ( <i>E. sativa</i> )	<b>98</b>	<i>Diplotaxis tenuifolia</i> , var. BELLEZIA UNT
<b>31</b>	Rúcula ( <i>E. sativa</i> )	<b>99</b>	<i>Diplotaxis tenuifolia</i> , var. LETIZIA UNT
<b>32</b>	Rúcula selvatica ( <i>D. tenuifolia</i> )	<b>100</b>	<i>Diplotaxis tenuifolia</i> , var. JOLIZIA
<b>33</b>	Wild Rocket, CN WROC 2110	<b>101</b>	<i>Diplotaxis tenuifolia</i> , var. ANASTAZIA
<b>34</b>	Wild Rocket, CN WROC 2111	<b>102</b>	<i>Diplotaxis tenuifolia</i> , var. PRUDENZIA F1
<b>35</b>	Wild Rocket, CN WROC 2112 UNT	<b>103</b>	<i>Diplotaxis tenuifolia</i> , var. TRICIA
<b>39</b>	Wild Rocket, CN WROC 2116 UNT	<b>104</b>	Rúcula selvagem



## Anexo 2

>DT02CA16 (contig\_16076)

AGGCTAACATATCATATGAGGCTCGCGTTCAAGTTTAAAGAACTCTACTGAAGTTGGATTCCCTGTTGACAC  
AAAGAAGCAAGGAAAGCACCTAGTAAATTATAAGATTTGTGGAGAGAGCTAATTAACCATTTGCAAGTT  
ATTATAACGAACCTTGTATTGGTGAATCTTGTGACAACAACAACCCCTCCTGCTGGC **CCTTTGATCTCAC**  
**CAGTC**GTTTCACTCGGCTCAGACGTAGCCACTTGACATTGCACCTATTCAA **CACACACACACACACA**  
**CACACACACACACA**TCTAATCAAAATCTCTCTTAAAGTATATGTGAGAGTTAGCTAGAGCATAATTACG  
CCTTGTCAATTGCGTTTGTGTC **TCCACTATTACTGCTGCT**TGCCCTCTGCACCATTAGTAGTACTCCATTTT  
CTTAGTGTCTCGCCAGATGTTTCACTCGGCTAAGCCGTATCGTCCCCACCAGTTTACCCAACCTCCTCCA  
CTGGTATTATTCCCACACCTACAACCCACATATAGACACACTTTTTCACAACCTATACAAATCCAATTAGT

>DT12CA15 (contig\_58401)

GCTCCCCAAACGACGACGACGCTCAATGCGAAAGATAATATTTAATTTGGGGTTATCAAACAAACAAACA  
AAAAAGGACAGGCTACGTTGATTCATCACCATCTCAGCATTGTTTTTTCAACAGGTTTATCCAAGGGATA  
TACATGCATTCGGACTCATAGCCTGGACAGGGAACGACGTGTTGAACAGAAGGTATGAACATAATAATCC  
ATTAT **GCGTATGTCACCAAAGAT**TGTTTTTAAATTAAGAACAAA **CACACACACACACACACACACACA**  
**CACA**ATCCATATATATATTTGATAATGCAATGGAATTTGCAGGCTGAGATTGCTAGCAGAATCCAAGGGA  
TATAGACTTGATGACACTGGGCTG **TTCCCAGCTACTCACAG**CTCAAGTGGTAACCGGGTAAGGGCAACAA  
TGTGAATGATTAAGGGAATGAATGAATGATTTTATTTTATGTTTGAGCATGTGATCATGTTCTTTTT  
GTCAGGGTGGTAGAGCAAGTGCAGCATGAACTCTCCACAGAGCAAGAAGTTTTTCGATTTCT

>DT14CA18 (contig\_75179)

TAATCCTATCAACTCATATTATATAAATTATTAGAAACAAGCCATTAACTCTTCAGGTCTAGAAGGTTGAG  
ATTATTAGCCATTTTCTTATAAAGCATCTGCCACATAACGGCTCAGAGGAACACATTCTCTATAAATGT  
CTCAGTGATCAGCATAAAGTTACTGAAGACTAGAAAATGTCGGGC **CATTGATTGTCCCTAAGCT**CAATTGG  
TAACAGAAAGAAACCAACAAATACACCAAAA **CACACACACACACACACACACACACACA**CTT  
AAATGTGATAAAAAACACTATATCTAACCTTCAATTTACTCAAGACTGGCAAAA **ACTTATGAGCCCTAGC**  
**ATA**ATACGACAACCTTCTAGAGTAGATTGGAATTACCTTTCCCTTGGAACTCAACGAGCGCTGCATCGGCC  
ATTTTCTTGGTTTTGAAAGAAACAAACCCGTAACCACTCGACCTCCTTGGATTTTTCAGGAATATAA  
CAGTGGACACAACGTTCCAGTGTCTGCATCAAAGAACTCCTTAAGATGCTTGGCCCTTGCTTCAA

>DT35GT11 (contig\_46042)

TCAAGACGCTCTTCTAGAAAAAGTGCATGTTCTAATCCATGCAGAGCAATCCCAAGCTGAAAACACCA  
TGCAAGTGATCAAGAAATGTTTCATTAACATACCAAAGATTCATTACTATCATACTAGAGGACTTGTGT  
CGTCTTTGAAGTTAGATATACGTAGAATATGTCTCTGTGTGAATAATTGAGAAGCTCATCTTTGAATGT  
TGGATCACCTTCTCTTAGTGTCTGTGAATGACTATTA AAAA **CCTTAATAGCCACCTGCAT**GTTGATAAAA  
GCAATTTAATTAGAGCAGAGTTCATGTAACATAAAGAAA **GTGTGTGTGTGTGTGTGTGT**TTAAGTGACT  
CACAATCCAATTGTGAGTTCTAGATAATCTCTTGGCTAGTGCATGAATGCAGTAAGCAACATCTGCTCGT  
GGTGTGTC **CACAGATGTTGCAGAGAA**TATTTCTGTATTAAAGATTA AAAA ACTCTGACCGAATCAGAAAC  
TCTTCATAATCATCATAAAACCAATTCAGAAGAGAGAACTATTACTGCGAATATGTGCTTCTTTGGG

>DT25CA16 (contig\_116501)

ACGGTCAAAAATAAGAAAACCTAACATTATGGTGTGTTGTGTCCCTGGAACCAATGATACCGACTACTATGG  
AATTATTCAGGAGATCATGATGTTAGATATCATGGTGTGTTGGCTTGAAAATCATTGTTTTTCAAATAT  
TCGTGGTTGATACCACGAAAACCGCGTATGAGAAGACATGTGTCTGGTCTAGTTGATTTCT **CACCAG**  
**CAGACAATATG**TAAAAATATGATCCTTTTGTATTACCAAATAAGTAAAAA **CACACACACACACACACA**  
**CACACACACACA**TATATATATTATATATATATATAAATTTATAAAGATATAAAATTTATGATGTTATTGT  
ATAATAAACAGATAAAGTGTGATCAAGCATATTTTAT **TCCTTACCCACGGATATG**TCGACATTCAACCGAA  
GATTGGTAGGCAAGTGCAAAATTTTTTCCCGGGGAATCCGAGAAACAGCAAAAATGCTTTGACTGCAT  
GGCAAGATGATAGACGTGACCAAGTTGCTGAAAGTTTCAATGTTATGGGTGGAGACACATGTTGT

>DT04CA24 (contig\_17567)

TGTCTTGCAAAATAGGAGACTTCTAGGTAAATCTTGTAATTTTTGAAAATGATCATTAGAC4TTGTGGGT  
ATAAGAGCGATCGATCGATGCACAAGGTGAGACATCGATCGACGGTGTGTGTGAAATCCAACCTGACTT  
GAGTCTTCAGGATGAGAATTCACAATGCTCTAAAAATGGTCCAAAAGCTCTAATTGGCTCAAACAAGTACA  
**TGAACCTGTTAAGACTCCA**TAACAACCTCCAAAATGCATAGAGAATGTCATAAAA**CACACACACACACAC**  
**ACACACACACACACACACACACACACACACA**TATATATATATATATATATATATATATATATATATCATGGG  
TGAAAATTTGTAATTTCCATGACATATCATGTGTCTAAACATGTTTTAGAGTATTTATA**GGTCTAGGTA**  
**TGTTTTGGAG**TTATTTGGAACATTTAGAGCTTTTTGGATGACAAAACCTCATAGCTAACAAGAATCAGCTA  
AGTTAGAGTCGACATATGAGATTGGAATCGATCGACAATGCAACATTGGCGTCAATCAATGATT

>DT21CA16 (contig\_98741)

GAGTTTTGAGCATTTTTGCTCAGAGAAGCAAGAACAAGGAGGGAGAGAGCCTTACTCTTCCTCTGGGAGAAA  
CGGCCCCGCATCGAAAATAATCACCTAAGAACAAAAACAACGTCTGGTTTTGAAAGGCAAGATAGACTGAAAA  
AACGATCTCCGCATGTTTACATCAGTGTAGAAGTGTGCAACCTCCAAAGCAAGCGTCTTTGATGACAAAC  
GAAA**CGATGTCTTATCCATGAGC**TGCGACAACAACACATTCAGATGATTGGCTCTCTACTAA**CACACACA**  
**CACACACACACACACACACACACA**ACAGTACTGAAAAAGCATTCAACCTGTGGATCAGTGACAATGGCT  
ACTTTAGTAAAC**TTCCCATCACTCTGCAA**GACAAAAGAAGAAGCTAGCAGATATGACCAATCAGCTC  
TTCAAGAACAAAGACTCGTACTTTTCGGAACAGAACAAAGCAAACACAAAATTTGGAATGTCATTGTCTTAA  
AAAACATTTCTCATACTTTTCGGACAGAACAAGCAACTACAAGATCGAAGTGTCTCATTTGCAT

>DT25TA12 (contig\_212)

AAAAAAGTTATTCTAGTTTTGGTTTTGGTTCAGTTATTTCAAGTAATTTAGAAAAAAAATAACGATATTTT  
AGATAAAAATATTGAGTATTTTGAAAATATCAAATTTGTTAGAATGATTTAGATAAAAAGAGTTCAGATAG  
ATAGGATTTGTTTTGGATATTT**CGGATAAACATATCCGCTT**ACTTTTCAGGATTGATGATTGATAATTTTTT  
TAAAAATTTTTCGAAGACTTTTAAATAGATTTTTAAATTA**TATATATATATATATATATATA**TTTTGGCTAT  
GCTATGTATATAAATTAATTTTTGGGTTTTTAGTGTATTTAAACTCCCAACTATTTTTTAAATCGGAAGAAAA  
**AC (C) TCCA**ACTAAAATTTTATGTTTTTAAACCTTGAACATTTAA**ACCGTTAGTGATGTTAAC**TTCCGT  
TAAAAAATTCGTTTTTCTTAAACAGACATCCGTTAAATTTGAAACGCGGTGTTTCATATAATCACATAAAG  
GACACGCAATGTTTATAATCAACAAATATTTTAGTGATTTAAACCTCTAACTATTTTTTAAATAGGATGAA  
AAACC

>DT26TA17 (contig\_226)

CTTTTTGAATATTTTATAATATTTTTGAAATAAAGATTTCTTCAGTGCTAGGTAACATTATTTCTATATT  
TATTATTTGCAGCTCCAATGATAGTATACTTTGGGTCAACCATATAAAAGCCAAACCCGCTCAAGTGTATC  
CTTCATGCTATAATTTATTA**GCTGAACAACCATACGA**AATTTATTTTTAAAGAATTTGTAAATATCTACT  
TGTTTTTATTAACCTG**TATATATATATATATATATATATATATATATATA**TTGCAACATAATGGTTTTAGA  
GTATTTATCCAAAAAGTAGTAGTTAGTTAATCTGTTTTA**CTGTCCGTGATCGTAGTTAGA**AAAATAATTATA  
TCTGACGTTCTTAGAGATTCCCTAGTCGATTCTTATTATACTATAATCAAAGTTTTATTTGTCAAGCTACAT  
CGATTGAGTTTTCTTTTGTGTATAACCTACTAATAGGCAAATTCAAATTCATGATGAGCTTCGGATAAGTG  
CCTTTTGTATTGGTTATCGAATTCATGATAACTAATTGAAAAATAAATTTCTTCTTAACTAAATATC  
GATCTCA

>DT30TA11 (contig\_294)

TTAAAAATGAGCATACAAGAACCAAGAAAAAGTAGCTAAAATGACTCGTGTAACCTCAACGTACCAAA  
ACAAAAACCCTTTTGTTTATATTTCCCAAATCCTCCTCTCCTTTTTCTCATCTTCTCACCTCAAACCTCTCT  
CTATCTCTCTCTCTCGCCATATCAAGAGCTAAATATTTCTCTGAAAGCTGACTTCAAAGCTTTTTTCTA  
AA**GCAAAACAAAGCCAAATCC**TTCTCTAATTGAGATTTTTATT**TATATATATATATATATATATATA**AATTTAA  
AAAGATGGGGAGACATTCATGCTGTACAAACAGAAGCTGAGAAAAGGACTTTGGTCT**TCCTGAAGAAGAC**  
**GAGAAG**CTTCTTAGGTACATCACTAAGTACGGCCATGGTTGCTGGAGCTCTGTCCCTAAACAAGCTGGTA  
ATATAATTTCTTCTATTTATCCGTTTTATTATAGCTTTATTGTGTTAGATTAATCCTCCTTCTAATTC  
TGAATTATCATTAAAAATATTTGTAGGTTTGACAGAGGTGTGGAAAGAGCTGTAGATTAAGATGGATAAA  
CTA

>DT19TA12 (contig\_133)  
ATATGATTTATGTGGCTAGCTGGTTCATGCAGTTTTTAGCCAACATCTATTGATTATTGCATCAATCTT  
ATTGTTTAAAGCAAAGTCCCTTACAAATTGCAACAACAAAAAATAGTCATATTGCAAGAATGATATATTGTT  
GCAATATGCTCATATATTTTTTGCAAAATGAAACTGCAATAGAATTGTAACAAATTTAGTTT **CTTGCTG**  
**TTTGAATGTTGCAAAATTCATATTTGTTGCATAGACGTAGTAAAAATTAGCAACAAATTTGAAATGCTGT** **TAT**  
**ATATATATATATATATATATA**TCATATTTTTGCAACCATTTTTAATAGTGATTTTGTGTATATTTACAA  
CGTATTTTCTAACGATAAATGGT **GGCAAGTCATTGCTACG**TAAAAATATATAGTAAAACACATTGCAAAC  
GGTAATTATAGTTTTTTAATTAATAATTTTTATTGCAAAAAATATTTATATATAAATTTTTCCGTACGA  
TTGCAACGTATTGTTGCATTTTTGAGTTATTACAAAATTTGCTATATAAAAATATATAGCAAATGTTGTT  
GCAA

>DT10GA15 (contig\_10754)  
CATATTAGAATTTTTAAGAGGGAAAAACAGCAAACCATTTATTTTTAGTCAGTGAGCTTCCGATGTTATGTA  
TGATTTATGAGGCTATGATAACTGAAATATCAATTGATATGACTCGTTTTGCTACTGGCATCATAAATC  
TCTAGTTCTACTACAATAGAGAGAA **CACAAGAGGCAACAGAA**AAACTTAAAACCAACTCTATGCCAAGAG  
AAAACATAAAGGAAAGCCA **SAGAGAGAGAGAGAGAGAGAGAGAGAGA**AACCTTCTGGCTTCATCTTC  
TTTCAACTTAGCAGCATGGATCTTCTCAGTACAGACTTCAATTCTCTTTCTCAAG (T) GTTCCAGTGCTT  
ACAATATTTTCAACAAAAGAAGA **GAGGTAACATGAATGAGGC**GTTCCCTGAGTAAAAAAAAGTGAGACAC  
TTGCTTGCTAATACTGTACCTGCTTCTTAGGACCAGCAGTGAGTTTCATTTCCATGGTGAGGTTACCAA  
AACATGCTTTCTTAGGACCAGCAGTGAGTTTCATTTCCATGGTGAGGTTACCAAACACCCCTTA

DT55GT15 (contig\_107264)  
TAATCTTATCTGTAAGCTTTACGTATGCTCTATTGTTTTCCCTCTTTCTATCACGCATGCTATATTGTT  
TTCAATCTCATCTGTGAAGCTAATAATATCTTTTTTTACTCTTGCAGATGGCATCCCTCGCTTTCGATAA  
CGAACAAAACGATATTGATATGCATTTGTGAAGACCAAGAGTACCAAGAGACAAGTGCTAATGTGAATGCA  
ACATCCAGAGCATCCCTCAGGCGAGTAAGGTGAAGAAGTTGGTTGTGCCAAGGTTCTCGT **GTGTGGGAGC**  
**ATTTACACAAGAACCAAAG (C) GACCCGAGACAA** **GTGTGTGTGTGTGTGTGTGTGTGTGTGTGT**CACCATT  
GCCACAATATCTTTTCATGCGCATCTAAGTCAAGGACTTTCGAACCTGAAGCAGCATGCAAAAT **CTGCAAAG**  
**AACACCAGC**CGTTCTTAATTGGTCAGAAAACGGATCAACAACAGCTTGATAATGAAGGAAAGCTGAAGCC  
ATCTAGAGTTTCCGAGTCTCTTTTTCAAAGAAGCGTGTAATGAATTGATGGTGTTAGCAGAGTTGTC

>DT41GT12 (contig\_72063)  
TGATGAACAAAAAAATTCACCATAATTTTTTTCGTAAACATGGTGTAGCAGTCGTTGTTAATTGCCTTT  
ACGAAAGATACTTTATTTCCACGGTTTATGAAATGACCAGATAAGTCCGATATATATAGTCGGATCAAAC  
ATTTTAGAATATTTCTATTTACATGATCTGTACAACATAGTCAGGAATGGAGAAGGAAAATGAGAATAAA  
TCCCAACTGCTATTATCTGATATGAGATAATTAGTCACATAAGATCATATCAGAGGGTCAACCCGCGT **AC**  
**TTTGACGAAACGAAGCC** **GTGTGTGTGTGTGTGTGTGTGT**GAGTGAAGCTGGTTAGGGTTTTGTCAGA  
AGGCTCGACAGAGTGTATTAGTCGTGGTCAAGTTAGAAGAGTTGCTGTTCAATGGGGCTGCTGCTTGT  
GGGACTACAGATTCT **GCTTCTCTTTGGTTCTGAG**ATCCCTCACGTTGGAGATTGCAACCCTAGCACATCT  
ACCACACCGCTCTCATCTTCTTCTTCTTCTCTCTGTTGGAGAATGAATCAGATGTTGGTTTCT

>DT51GT11 (contig\_93338)  
GATGAAGGTTCCGAGAACTAACCAGCTTGGTGTCAAACCAAAGATGGGTTATACTATAAAATTCACTGGC  
TTTCGTGACCAGGTTTGTGTTTTGCAGACAAGATTCCTTTGTCTTTGCTTGAAGCTTATGTGTGTGTTCT  
TTTGGTGGTGTAAATAGGATGTTGCGAGCCTGACAAGTTTTTTCCAAAGTGCGTTTTGGGAAAACACCTGAA  
GAGAAA **CAGCTTCTGTTAGTGGTC**GGAAATGGGGAGAAGTGGATTTGAATG **GTGTGTGTGTGTGTGTGT**  
**GTGT**TGTCTTTCGCATTTAAGCTTCTTATGTGGATAAAGTCTTAGCTTTATTTCTTTTTTGTGTGTG  
TGTGTGAAGGGAATAA **TCTGACGTTTTTGGTTGG**GGGAAAGCAAGCTTTTGGAGTGTCTCTGGCTGATGT  
TTCACAGACTCAGCTTCAAGGAAAAATGATGTTTTTGTGGAGTTTCATGTTGATGACACTGCTGGTGCC  
AATGAGGTATGATATTGATGTAAGTGTGTGTTATGTTAGCTGAGTTGCTGATTCTGTTGTTTA

> DT58TA11 (contig\_653)  
 TTAGTGATTATGCTAGGTTGAAACGGTTTTAAAAGTAAATTGCAGGTTTGTGAGAAATGGTAATTTGCTC  
 GATTGAATCGGTTTTTGGTATTAACAATAAGGAAAGGTAGTTAGATTTAAGATTACTATTCAGATAATCA  
 GGATTATAATGCTATATATGCCTAATAAGTTGCATGCATGATATTA AAAAGATTGAAATATGCTAATATG  
 TCCTCGCGTTTTAGACTC GTCTATTGATCTGATGCCGAAAAATATCGATCGATGTTTTGACAGCGATATCG  
ATATATATATATATATATATATATTTGTGTCGTCGATCGGTTACTCTATACGAGTATCGATCGATATAGCTC  
 AGATCGAGTTGATGTGGATGCTCACCAAGA TCATAGATCGGTTCTCGCGTATCCCTAGAGGATCCTTAGCG  
 TATGCAAGATGATGTTAGGATGATGATTAACGAGTTAATTCATTCCAAGATCATATTCAGTTTGTCTACT  
 CTAGGATAAACATTAGAGTTCATCATGATAATGAATATCACAACTCAACAATCTATAGTTGGGGCT

> DT61TA11 (contig\_686)  
 AAAATAAAATCTAAACTCTTCACCAACAAAAGGTAACCTTCAGGTAATTTGATAAGCTAATGTGTAATCCA  
 CCATTAAGATTTTCATTTTTTATCTTTAGGTGAAAGAAATTACGAAAAAGCATGAGACCGTACATAACT  
 TATGGAGCTTCTATTTCAATTAATCTATATTAGTACCTTAGCAGTTGCTCAAACCAGGATTACATGACA  
 TGCAGTACATGTTA GTAATAAAGCGACCCTGCATGCACCAAGTATAATAATCTTAATTTTACA TATATAT  
ATATATATATATATAACTAATAAATTTGTGCAAAAAAACTTCTCAAATAAATAAATCTTTAAATTTGGAT  
 AGCAAAAACAA TGATACCTTCCCACCAGTTTCCTCTTGATGCGCTTTTGAATCGCCTTCCTCTCCATCTT  
 TCTTGTACTGCATAAGTTAAAGATAAAGAAGAAAAGCAGAATAATTTTGAATTAATAATACATATAGGG  
 ACTACACGTACATGTAAGTATATAAGTTTTTGTACGTTACTCACATAGGTGTTATGGGTATGATAGGTA  
 C

> DT182TA14 (contig\_17224)  
 TCTTGCTTCAGTTCATGAGATACAAAATGATTTGGGGAATCCATCATATATTATCAACGGTTTCTCTCAA  
 GTTCTCCAGATATGCGGATATGAGGTTGTTCTGTTGATAGTAAGGTCAAAGCTCACACAGAAGGTGCTCT  
 ACATGATTCGTAAATGTTGATGTGGTAGGGATTTAGAGCAAGAATT GACTAACTCTGTGTGTCACCTTC  
 TGTGCATATATATATATATATATATCTCAGCAAGAAGAAGA TATATATATATATATATATATATATATATC  
 TTGTTACTTCCGCTGATCATATTGCTCTGTTTCAGAGATTACAAGCAACAACCTCTTTTTAATTCCTCCAC  
GCCTGGAAC TGAAATTAAACTGAAGCATGCTGAGATGTTTTGTTCACCAAGCGGGGGCTTGTACCGACT  
 GTTTTCTGACTTTTTCTAATTATAAAAAGATGCGCATAAATGCATCGAACCTTGATGATACCTCTCAAACAT  
 AGTAAAAAGTTTTTACTATAAATAAAGTGAATAGTGAAGATGAAGCTATCACCAATCAATTTCTCAG

> DT189TA14 (contig\_19700)  
 AGCACCAAAGACTTCGAGGGGAGAGAGTTTCAGCCCAGTAGCACAAGCTTGCTGAATATTATACTGATGGT  
 GAGGAAAAAGCACTTGATTGTAAGCCTTAAGTGCATCCTTTAGAGAGTCATCACCTCTGTGAATCAT  
 ATTA AACACAGATTTAAACCCCGATAAAATTTACAACAAGCCACGGTCAAACCTTATA CAAGTAATTCAGTC  
TCGGGTATCTTGTATGGCTATATAA TATATATATATATATATATATATATATATCTCTGCTTGTGTGTGT  
 TAACATTTTTTCGAGAGAGTCTATATAAGAAACTTACTCAGTAGCCAATAACTCTTCGCACAGAACCTTG  
 ATCATTT CAAGACCTTGCTTTACCTTCAACAGGTTTCTTGTATGGCTACCAAATTTCTTTACACAGCCTA  
 CCTCAATGTCTTTGTCAATCATTTGCCCTCCAATGTGCGATATTGAACGCGATGCCACGGCCAGATCATTGAG  
 CTGGAAAAATAAGTTTTGTTTCTCAACGCTTGATATATAAACAAACACTCAATTTATAAATCGTCATGAT

> DT71GA16 (contig\_50994)  
 CTCTTATTGCACCTTTCACGCCCTGAATATTAACACAATGATGTTAAAGATCCTTCCACTTATCACTCAG  
 ATGAAAAACAAAATGCTAAAACATAACCTTAAGTGGAAATCCCAAGCATATTACAGCTGAGACTCTGATAT  
 CTTCACTCACGGCTACCATGCAACTTACTCTGCAGAAGCATAAGATCACAT CCAACATAGAAAGGTGCGT  
AACAGAACAACGTGC SAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATCTACCTGGAACCCATTGATTTG  
 CCAGCCAGTATCAAAGGATGACCAGGGAATCTAGCAGCAGTTTCTTTTACAAAATCCAGATGAAATTCAG  
 TAAGTTTTT CAGCTTTTGGAGGAACTCTTCTTTTCCCTCCAACAATATCTACCCACCCAGAAATTTCCAC  
 AAATTA AATGAAATCCCAATAAGTTTCTATCAAGAACACGAATATTATTATTA AAAAAAAAAATCTAGAAT  
 TTTACAGACATGGGTAATCAAATGTGATAACGTCAACAGCTGCTAAGCTCTTCTTTAGCCTATC

## Anexo 3

### PRIMERS

Nome:	Sequência (5'→3'):	Tm (°C):
DT02CA16_FW	<b>CCTTTGATCTCACCAGTC</b>	49,2
DT02CA16_RV	<b>AGCAGCAGTAATAGTGGA</b>	50,1
DT12CA15_FW	<b>GCGTATGTCACCAAAGAT</b>	49,6
DT12CA15_RV	<b>CTGTGAGTAGCTGGGAA</b>	50,3
DT14CA18_FW	<b>CATTGATTGTCCCTAAGCT</b>	49,4
DT14CA18_RV	<b>ATGCTAGGGCTCATAAGT</b>	50
DT35GT11_FW	<b>CCTTAATAGCCACCTGC</b>	49,3
DT35GT11_RV	<b>TTCTCTGCAACATCTGTG</b>	49,4
DT25CA16_FW	<b>CACCACGCAGACAATATG</b>	50,5
DT25CA16_RV	<b>CATATCCGTGGGTAAGGA</b>	50,2
DT04CA24_FW	<b>TGAACCTGTTAAGACTCCA</b>	50,3
DT04CA24_RV	<b>CTCCAAAACATACCTAGACC</b>	49,4
DT21CA16_FW	<b>CGATGTCTTATCCATGAGC</b>	49,7
DT21CA16_RV	<b>TTGCAGAGTGATGGGAA</b>	50,7
DT25TA12_FW	<b>CGGATAAACATATCCGCTT</b>	49,5
DT25TA12_RV	<b>GGTAAACATCACTAACGGT</b>	49,2
DT26TA17_FW	<b>GCTGAACAAACCATACGA</b>	49,8
DT26TA17_RV	<b>TCTAACTACGATCAGGACAG</b>	50
DT30TA11_FW	<b>GCAAACAAAGCCAATCC</b>	49,4
DT30TA11_RV	<b>CTTCTCGTCTTCTTCAGGA</b>	50,6
DT19TA12_FW	<b>CACAAGAGGCAACAGAA</b>	50,2
DT19TA12_RV	<b>CGTAGCAATGACTTGCC</b>	50,5
DT10GA15_FW	<b>CACAAGAGGCAACAGAA</b>	49,3
DT10GA15_RV	<b>GCCTCATTCATGTTACCTC</b>	50
DT55GT15_FW	<b>GTGTGGGAGCATTTCAC</b>	50,4
DT55GT15_RV	<b>CCTGGTGTTCCTTGCAG</b>	50,5
DT41GT12_FW	<b>ACTTTGACGAAACGAAGC</b>	50,6
DT41GT12_RV	<b>CTCAGAACCAAAGAGAAGC</b>	50,2
DT51GT11_FW	<b>CAGCTTTCTGTTAGTGGTC</b>	50,1
DT51GT11_RV	<b>CCAACCAAAAACGTCAGA</b>	50,2
DT58TA11_FW	<b>GTCTATTGATCTGATGCCG</b>	49,7
DT58TA11_RV	<b>GCGAGAACCGATCTATGA</b>	50,7
DT61TA11_FW	<b>GTAATAAAGCGACCCTGC</b>	50,3
DT61TA11_RV	<b>CTGGTGGGAAGGTATCA</b>	50,1
DT182TA14_FW	<b>GACTAACTCTGTGTGTGC</b>	49,6
DT182TA14_RV	<b>TTAATTCAGTTCCAGGCG</b>	49,5
DT189TA14_FW	<b>CAAGTAATTCAGTCTCGGG</b>	49,6
DT189TA14_RV	<b>GGTAAAGCAGGGTCTTG</b>	49,5
DT71GA16_FW	<b>CCAACATAGAAAGGTGCG</b>	50,5
DT71GA16_RV	<b>GAGTTCCTCCAAAAGCTG</b>	50,2