TECHNICAL UNIVERSITY OF CARTAGENA

DEPARTMENT OF AGRICULTURAL SCIENCE AND TECHNOLOGY

INSTITUTE OF PLANT BIOTECHNOLOGY

Ph.D. Thesis

Role of plastid markers in environmental studies on the example of the endangered species *Cistus heterophyllus*

Marta Pawluczyk

Supervised by

Dr. Marcos Egea Gutiérrez-Cortines

Dr. Julia Rosl Weiss

Cartagena 2017

Universidad Politécnica de Cartagena

DT-16

CONFORMIDAD DE SOLICITUD DEAUTORIZACIÓN DE DEPÓSITO DE TESIS DOCTORAL POR EL/LA DIRECTOR/A DE LA TESIS

Dr./a Marcos Egea Gutiérrez-Cortines y codirigida por el/la Dr./a Julia Weiss.Director/a de la Tesis doctoral **Role of molecular markers in environmental studies on the example of the endangered species *Cistus heterophyllus***

INFORMA:

Que la referida Tesis Doctoral, ha sido realizada por D/Dª.Marta Pawluczyk, dentro del programa de doctorado Técnicas Avanzadas en Investigación y Desarrollo Agrario y Alimentario, dando mi conformidad para que sea presentada ante la Comisión de Doctorado para ser autorizado su depósito.

La rama de conocimiento en la que esta tesis ha sido desarrollada es:

Ciencias
Ciencias Sociales y Jurídicas
Ingeniería y Arquitectura

En Cartagena, a 10 de enero de 2017

EL/LA DIRECTOR/A DE LA TESIS

MARCOS| EGEA| GUTIERREZ CORTINES

Firmado digitalmente por MARCOS| EGEA|GUTIERREZ CORTINES
Nombre de reconocimiento (DN): cn=MARCOS|EGEA|GUTIERREZ CORTINES, serialNumber=[redacted]
givenName=MARCOS, sn=EGEA GUTIERREZ CORTINES, ou=Ciudadanos, o=ACCV, c=ES
Fecha: 2017.01.18 16:33:22 +01'00'

Fdo.:

LA CODIRECTORA DE LA TESIS

Julia Rosl| Weiss

Firmado digitalmente por Julia Rosl|Weiss
Nombre de reconocimiento (DN): cn=Julia Rosl[redacted]
givenName=JULIA ROSL, sn=WEISS, ou=Ciudadanos, o=ACCV, c=ES
Fecha: 2017.01.19 10:34:11 +01'00'

Fdo.:

COMISIÓN DE DOCTORADO

Universidad Politécnica de Cartagena

DT-17

CONFORMIDAD DE DEPÓSITO DE TESIS DOCTORAL POR LA COMISIÓN ACADÉMICA DEL PROGRAMA

D/Dª. Francisco Artés Hernández Presidente/a de la Comisión Académica del Programa Técnicas Avanzadas en Investigación y Desarrollo Agrario y Alimentario.

INFORMA:

Que la Tesis Doctoral titulada, "**Role of molecular markers in environmental studies on the example of the endangered species Cistus heterophyllus**
", ha sido realizada, dentro del mencionado programa de doctorado, por D/Dª.-**Marta Pawluczyk**, bajo la dirección y supervisión del dirigida por el Dr./a Marcos Egea Gutiérrez-Cortines y codirigida por el/la Dr./a Julia Weiss.

En reunión de la Comisión Académica de fecha 22 de diciembre de 2016, visto que en la misma se acreditan los indicios de calidad correspondientes y la autorización del Director de la misma, se acordó dar la conformidad, con la finalidad de que sea autorizado su depósito por la Comisión de Doctorado.

La Rama de conocimiento por la que esta tesis ha sido desarrollada es:

- **Ciencias**
- Ciencias Sociales y Jurídicas
- Ingeniería y Arquitectura

En Cartagena, a 19 de enero de 2017

EL PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA

FRANCISCO DE ASIS|ARTES| HERNANDEZ

Fdo:----------

COMISIÓN DE DOCTORADO

To my M&M

Acknowledgments

I would like to thank those people without whom this PhD thesis could not have been completed.

First of all, I would like to express my gratitude to my thesis directors Julia Weiss and Marcos Egea-Cortines. With them I took my first steps in the lab and everything I know about lab work I owe to them. Thank you for your guidance and patience during the realization of this thesis. Marcos, thank you for your faith in me becoming a PhD! Julia, thank you for your invaluable help during the work on this manuscript.

I would also like to thank María José Vicente Colomer and Juan José Martínez Sánchez from the Plant Production Department for giving me the opportunity to participate in their project, for their trust, kindness, for sharing with me their great knowledge in botany and valuable lessons of field work.

I wish to acknowledge my colleagues from the Genetics group, PhD students and Doctors, now living all around the World: Izaskun Mallona Gonzalez – for your valuable comments during the editing of this manuscript, for all the field trips to get plant material and for "una caña en el puerto" when experiments didn´t work out, Luciana Delgado and Almu Bayo – for your assistance in my first PCRs, María Manchado, Pablo Galindo and María José Nicolas - thank you for helping me get plant material for my experiments and sharing with me your experience and long days spent in the lab. Thank you, Marta Terry, Victoria, Marina, Raquel, Claudio and Fernándo for nice company during my last seminars in Cartagena before the thesis defence. You´re a great team!

I appreciate Perla Gómez and Mariano Otón from Institute of Plant Biotechnology for your technical support

I am in debt of my co-authors Matthew G. Links, Mikel Egaña Aranguren and Mark D. Wilkinson for their contribution to this project.

Thanks to the members of Plant Production investigation group: Maira, Fran and Eli for receiving me so kindly and for your assistance while I was visiting your lab.

Thanks to Krzysztof Spalik for giving me the opportunity to stay in his research group at Warsaw University (Poland), and to Lukasz Banasiak for his essential help and the practical introduction in phylogenetic analysis.

I appreciate Jorge, Pedro and ANSE group for African samples of *Cistus heterophyllus* samples and for letting me use your greenhouses as the experimental area.

I am in debt with Dirección General de Universidades y Política Científica de la Consejería de Universidades, Empresa e Investigación and Dirección General de Patrimonio Natural y Biodiversidad de la Consejería de Agricultura y Agua for funding the research project “Molecular markers in conservation and management of the flora in the Murcia region”and making my work possible.

I'm most grateful to meet on my way Agnieszka Pietras-Lebioda – for sharing our first steps in research projects.

My special thanks to my Mum and Dad, my brother, the rest of my family and friends in Poland for your support and understanding my absence during my stay in Spain. Even when I was far away I could always count on you.

Finally, to Marinaldo and Marianna for being my greatest motivation and distraction at the same time during my work on this thesis.

Abstract

Molecular markers are a very powerful tool in many fields such as phylogenetics, evolutionary or conservation biology. However, it is not an easy task to find proper markers for rare species. The perfect marker depends on the biological question: for differentiation among closely related species we need a sensitive marker for highly differentiated region, whereas differentiation among organisms belonging to distant families requires markers for conserved regions.

Important is also the type of the DNA as source of our marker. Plastid DNA is preferred in plant phylogenetic projects whereas the analysis of hybridization events requires markers proceeding from nuclear DNA.

However, in case of rare species, scientist encounter a lack of sequence information about the genomes studied. In this case the only solution are universal markers already described for other organisms.

This project aims to analyse a set of molecular markers for tracing hybridization events in the population of an endangered species from the Cistaceae family, *Cistus heterophyllus* subsp. *carthaginensis*. The distribution of this subspecies is limited to only one natural population in the south-eastern Spain where individuals with wild type and hybrid phenotypes co-occure, suggesting hybridization events between the endangered population and the locally abundant *Cistus albidus*. These hybrids have been described in Africa as *C.* × *clausonis*.

We searched for DNA regions that allow discrimination between the wild type individuals and putative hybrids. The generated data could improve the species conservation strategy in order to avoid its extinction.

In chapter 1, we describe the possible application of plastid markers, regions known as ¨DNA barcodes¨ as markers for the aforementioned population. Noncoding DNA regions (*rbcL, trnK-matK)* were found as not

variable enough to be informative in closely related individuals. Intraspecific regions (*trnL-F, trnH-psbA*) presented a high rate of the evolutionary changes as indicated by their high variability. However, we found these markers as not sufficiently stable to give reliable information for the identification of wild type and hybrid individuals. Surprisingly, we observed heteroplasmy for *rpoB* and *rpoC1* genes in *C. heterophyllus* and the local *C.* × *clausonis*, but not in *C. albidus* or another species common to this region, *C. monspeliensis*.

We found two distinct alleles of *rpoB*, one present in all species and a second present only in *C. heterophyllus* and the local *C.* × *clausonis*. We also detected two alleles of *rpoC1*, one common to all species analyzed and a second present only in the local *C.* × *clausonis*. Our results show that there is a distinctive *rpoB* allele common to *C. heterophyllus* and *C.* × *clausonis* from Africa and Europe. The unique *rpoC1* allele found in the local *C.* × *clausonis* directs to a different origin of this small population, indicating that it is not a hybrid originating from *C. albidus* or C. *heterophyllus* currently present in this location.

Chapter 2 describes the application of the highly polymorphic internal transcribed spacer (ITS) region of the ribosomal DNA in the construction of a molecular tree in order to unravel the relationship among geographically isolated populations of *Cistus heterophyllus, Cistus albidus* and possible hybrids of these two species, *C.* × *clausonis* from Africa and Europe. Our data indicate that, depending on the individual and population, *C.* × *clausonis* phylogenetically resembles more either *Cistus heterophyllus* or *Cistus albidus* what might be related to the homogenization of variation between repeat types through concerted evolution.

In chapter 3, we present an issue that arose during the analysis of quantitative PCR data of the barcode markers. As we realized that there were significant differences between species in PCR efficiency of the same marker, we

decided to investigate if the observed bias may disturb species identification during metabarcoding of samples.

We used six universal *loci* and 48 plant species and quantified the bias at each step of the identification process from end point PCR to next-generation sequencing. End point amplification was significantly different for single *loci* and between species. Quantitative PCR revealed that the Cq threshold for various *loci*, even within a single DNA extraction, showed 2,000- fold differences in DNA quantity after amplification. Next generation sequencing (NGS) experiments in nine species showed significant biases towards species and specific *loci* using adaptor-specific primers. NGS sequencing bias may be predicted to some extent by the Cq values of qPCR amplification.

Resumen

Los marcadores moleculares son una herramienta muy poderosa en muchos campos como la filogenia, la biología evolutiva o la conservación. Sin embargo, no es una tarea fácil encontrar marcadores adecuados para especies raras. Los marcadores ideales tienen que ser de carácter informativo dependiendo de la cuestión biológica: la diferenciación entre especies estrechamente relacionadas requiere un marcador para regiones altamente diferenciados, mientras marcadores para la diferenciación entre organismos pertenecientes a familias distantes están seleccionados para la detección de regiones conservadas.

Importante es también el tipo de ADN como fuente de nuestro marcador. El ADN de plástidos se prefiere en proyectos sobre filogenia, mientras que la detección de eventos de hibridación demanda el análisis del ADN nuclear.

Sin embargo, en el caso de especies raras, los científicos se encuentran con una falta de información sobre secuencias de los genomas bajo estudio. En este caso la única solución es la aplicación de marcadores universales, ya descritos para otros organismos.

Este proyecto tiene como objetivo analizar un conjunto de marcadores moleculares para el rastreo de eventos de hibridación en la población de una especie en peligro de extinción de la familia Cistaceae, *Cistus heterophyllus* subsp. *carthaginensis*. La distribución de esta subespecie se limita a una sola población natural en el sureste de España, donde co-ocurren individuos con fenotipo silvestre y fenotipos híbridos, lo que sugiere eventos de hibridación entre esta población en peligro de extinción y una especie localmente abundante, *Cistus albidus*. Estos híbridos se han descrito en África como *C.* × *clausonis*.

Se realizaron búsquedas de regiones de ADN que permiten la discriminación de entre los individuos de tipo silvestre e híbridos supuestos. Los datos generados

podrían mejorar la estrategia de conservación de las especies con el fin de evitar su extinción.

En el capítulo 1, se describe la posible aplicación de marcadores moleculares plastídicos, regiones de marcadores conocidas como “códigos de barras”, para su aplicación de la población mencionado anteriormente. Regiones no codificantes de ADN (*rbcL, trnK-matK*) no resultaron lo suficientemente variables para ser informativas en individuos estrechamente relacionados. Regiones intra-específicas (*trnL-F, trnH-psbA*) presentan una alta tasa de cambios evolutivos, indicado por su alto grado de variabilidad. Sin embargo, encontramos que estos marcadores no son suficientemente estables como para proporcionar información fiable para la diferenciación entre individuos silvestres e híbridos. Sorprendentemente, se observó para los genes *rpoB* y *rpoC1* una heteroplasmia en C. *heterophyllus* y C. × *clausonis* local, pero no en C. *albidus* u otra especie comun a esta región, *C. monspeliensis*. Encontramos dos alelos distintos de *rpoB*, uno presente en todas las especies y un segundo presente sólo en *C. heterophyllus* y *C.* × *clausonis* local. También se detectaron dos alelos de *rpoC1*, uno común a todas las especies analizadas y un segundo presente sólo en *C.* × *clausonis* local. Nuestros resultados muestran que hay un alelo *rpoB* distintivo y común a *C. heterophyllus* y *C.* × *clausonis* de África y Europa. El alelo *rpoC1* unicamente encontrado en *C.*× *clausonis* local indíca un origen de esta pequeña población diferente que no resulta de una hibridación entre los *C. albidus* o *C. heterophyllus* actualmente presentes en esta ubicación.

El capítulo 2 describe la aplicación de regiones internas inter-espaciadas (ITS, internal transcribed spacer) ribosomales. Estos marcadores altamente polimórficos permiten la construcción de árboles filogenéticos moleculares con el objetivo de analizar las relaciones entre poblaciones geograficamente aislados de *Cistus heterophyllus, Cistus albidus* y posibles híbridos entre estos dos especies, *C.* × *clausonis* de África y de Europa. Nuestros datos indican que, depnediendo de individuo o población, *C.* × *clausonis* filogenéticamente parece

más a *Cistus heterophyllus* o *Cistus albidus* y problamente está relacionado a la homogenización de variación por evolution concertada.

En el capítulo 3, se presenta un problema que surgió durante el análisis de los datos de PCR cuantitativa de los marcadores de código de barras. Como resultaron diferencias significativas entre especies en la eficiencia de la PCR aplicando el mismo marcador molecular, decidimos investigar si el sesgo observado podría perturbar la identificación de especies durante el metabarcoding de muestras.

Utilizamos seis *loci* universales y 48 especies de plantas y cuantificamos el posible sesgo en cada paso del proceso de identificación desde PCR apunto final hasta la secuenciación. La amplificación a punto final fue significativamente diferente para un solo *loci* y entre las especies. Análisis por PCR cuantitativa reveló que el umbral Cq para diversos *loci*, incluso dentro de una sola extracción de ADN, mostró una diferencia de 2000 veces en la cantidad de ADN obtenida después de la amplificación. Experimentos de secuenciación de próxima generación (NGS) en nueve especies mostraron sesgos significativos hacia especies y *loci* específicos utilizando cebadores específicos del adaptador. El sesgo durante la secuenciación NGS se puede predecir en cierta medida por los valores Cq de amplificación en qPCR y depende de la secuencia primaria de ADN.

Original Publications of the Thesis

- Chapter 1 – Marta Pawluczyk, Julia Weiss, María José Vicente-Colomer, Marcos Egea-Cortines (2011) Two alleles of *rpoB* and *rpoC1* distinguish an endemic European population from *Cistus heterophyllus* and its putative hybrid (*C.* × *clausonis*) with *C. albidus*. *Plant Systematic and Evolution*. 298(2): 409-419.

- Chapter 3 – Marta Pawluczyk, Julia Weiss, Matthew G. Links, Mikel Egaña Aranguren, Mark D. Wilkinson, Marcos Egea-Cortines (2015) Quantitative evaluation of bias in PCR amplification and Next Generation Sequencing derived from metabarcoding samples. *Analytical and Bioanalytical Chemistry* 407(7): 1841-8.

- International Congress

  Chapter 1 - Marta Pawluczyk, Julia Weiss, María José Vicente-Colomer, Marcos Egea-Cortines. Use of DNA barcoding genes in genetic analysis of the *Cistus heterophyllus* subsp.*carthaginensis* unique population – XVIII Congreso de Federación de Sociedades Europeas de Biología de Plantas (FESPB) Valencia, 4-9 June, 2010

Table of Contents

List of Figures

List of Tables

Introduction

1. Molecular methods in ecological studies

In the past, environmental studies relied only on morphological data. This changed together with the technical development in biology at the end of the XXth century. Nowadays, botanical or phylogenetic studies combine phenotypic data with information generated in the laboratory via molecular techniques including the analysis of molecular markers applying PCR amplification and Next Generation Sequencing.

Molecular ecology is defined as "the application of molecular techniques to answer ecological questions" (Beebee & Rowe 2008). This interdisciplinary approach that applies an array of molecular tools in ecological studies became a milestone in modern ecological research and is responsible for the great advance in evolutionary biology research. Ecology not only is concerned with current state of populations and relation between organisms but also is inseparably related to the evolutionary history of organisms. Molecular biology helps to understand the origin of species and the ecological basis of their existence.

The great advance has been possible due to the development of at the times critical and now basic molecular techniques.

1.1 Traditional PCR and quantitative PCR

One of the most widely-spread and useful technique in molecular ecology is the polymerase chain reaction (PCR) described by Mullis and Faloona in 1987. The amplification of particular segments of isolated genomic DNA using oligonucleotide primers is repeated during various cycles. The primers, short sequences complementary to DNA stretches flanking the target sequence, are necessary starting points for DNA synthesis. Primers can be universal or specific. Universal primers amplify a specific DNA region in a range of species while specific primers are designed based on a specific and unique sequence for a species of interest. These primers usually amplify target sequences from only one species or few closely related species.

The invent of the PCR reaction allowed to isolate and amplify specific fragments of DNA from the background of large genomes. Moreover, it gave the possibility of obtaining billions of copies of a specific piece of DNA from the genome using very few starting copies of only a few nanograms. This is especially important if it is difficult to collect large amounts of tissue, for example in case of rare or endangered species. PCR opened up new possibilities of non-invasive sampling methods without the need to harm or destroy the organism. This is an important aspect when numerous samples are needed, as in the case of studies of population genetic. DNA amplification also allows working with old and/or degraded DNA, such as dry plant material from herbarium or samples from fossils.

Traditional PCR provides valuable genotypic information based on sizes or sequences of amplified products but cannot supply us with accurate estimates on the amount of DNA present in particular samples. Quantitative PCR (Q-PCR), also called real-time PCR, developed in 1990s (Higuchi *et al.* 1992, Higuchi *et al.* 1993), opened up this possibility. This method permits registration and quantification of the amplified product during each cycle by measuring fluorescence that is emitted when the fluorophore combines the DNA double strand. The two parameters that are widely used in the Q-PCR analysis are amplification efficiency and Cq (quantification cycle).

The optimal amplification efficiency is when from one copy of the template are generated two copies of the product. Then the efficiency is 100%. As the result of disturbances during the amplification process the efficiency can be lower (Mallona *et al.* 2011). The Cq parameter is the cycle in which the fluorescence reaches the established threshold (Luu-The *et al.* 2005, Bustin *et al.* 2009).

There are various ways that Q-PCR data can be useful in ecological studies. Firstly, it can give an important information about the influence of gene expression levels on the development of some phenotypic features for example salt tolerance (Gu *et al.* 2004). Q-PCR technique is often applied in species identification from complex samples in combination with melting peak analysis (Manter & Vivanco 2007, Derycke *et al.*2012). Finally, it can be used in more

technical approaches like the evaluation of applicability of specific molecular markers as described in Chapter 3.

1.2 Molecular markers

A molecular marker is a DNA sequence of the genome that permits differentiation among individuals of a population. DNA sequences may vary among different organisms but this variation usually is not displayed by any phenotypic features so it can be detected only by molecular analysis. Molecular markers are an important tool in species identification, phylogenetic reconstructions or revision of existing taxonomy.

There are different types of molecular markers depending on the technique that is used and the initial information about the organism that is going to be studied.

One of the PCR-based markers used by molecular ecologists are allozymes. Allozymes are variant forms of an enzyme used in important metabolic processes, that are coded by different alleles at the same locus. These variants do not have the same structure so they can be visualized by capillary electrophoresis. These markers exhibit high levels of functional evolutionary conservation throughout specific phyla (Hamrick & Godt 1990). They are often used to study evolutionary histories and relationships between different species. However, some organisms are monomorphic for their allozymes, which prevents their application and requires an alternative method to determine the evolutionary history of a taxa (Parker *et al.* 1998).

In case there are no previous studies on the organism of interest, as it can occur specially for endemic and rare species, no previous data are available that permit to design specific primers. In such a situation, the only solution are markers with unknown target regions in the genome. This type of markers can provide information about the genetic variability and allow differentiation among organisms. Examples for these markers are:

- Restriction Fragment Length Polymorphisms (RFLP) – in this method genomic DNA is digested by endonuclease enzymes, giving different-sized fragments of DNA. Restriction fragments are separated on an agarose gel, transferred to a

nylon membrane and visualized by a hybridizing radioactively- or fluorescently-labelled DNA probe (in Southern-blotting) (Botstein *et.al* 1980).

Some types of markers may not be applied in phylogenetic studies or reconstructions of species evolutive history because this type of markers is usually dominant and it is not possible to discriminate between homozygous or heterozygous individuals. Amongst these are:

- Amplified Fragment Length Polymorphism (AFLP) - a technique based on PCR amplification of digested fragments of genomic DNA. The initial step in AFLP is digestion of genomic DNA with two restriction enzymes, generating fragments with sticky ends. After ligation of adaptors to these ends, subsets of the digested fragments are selectively amplified (Vos *et al.* 1995).

- Randomly Amplified Polymorphic DNA (RAPD)– this method applies amplification of genomic DNA sequences with 10 base pair long primers of random nucleotide sequences (Williams *et al.* 1990).

For organisms with sequenced libraries published in public databases, specific markers can be applied. One of them is known as sequence-tagged site (STS) markers- a short (200-500 bp) DNA sequence with determined location in the genome. The STS concept was introduced by Olson *et al.* (1989). Some of STS types are:

- Single Sequence Repeat (SSR), also named microsatellite DNAs (MSATs) or Simple Sequence Repeat Polymorphisms (SSRP) – short, tandemly repeated, highly repetitive sequences of two, three or four nucleotides, located throughout the genome. They are present in nuclear and organellar DNA and usually appear in non-coding regions of the DNA. SSR markers were developed for usage in genetic mapping in humans (Litt & Luty 1989, Weber & May 1989).

- Inter-Simple Sequence Repeats (ISSR) – these markers involve PCR amplification using a single primer containing microsatellite repeated sequences. This primer amplifies the region between closely located and oppositely oriented SSRs. Primers can be designed for a microsatellite repeat only or it can be extended outside or inside the ISSR (Moreno *et al.* 1998, Fang & Roose 1997).

- Sequence Characterized Amplified Regions (SCARs)–this technique consists in PCR amplification using specific primers (15-30 bp length). Primers are designed from nucleotide sequences established in cloned and sequenced RAPDs (Hernández *et al.* 1995, ChungSun *et al.* 2000). SCARs markers are also successfully developed from AFLP (Xu *et al.* 2001) or ISSR primers (Ye *et al.* 2006).

- Cleaved Amplified Polymorphic Sequence (CAPS) - this method is a variation of RFLP but with previous PCR amplification of the fragment containing the variation. It is based on polymorphisms in the length of restriction fragment that create or eliminate restriction recognition sites in PCR regions amplified by specific oligonucleotide primers (Konieczny & Ausubel 1993).

The aforementioned markers are codominant. The other type of markers is based on direct DNA sequencing of targeted regions within the genome. This method determines the exact order of nucleotides within the DNA strand. It is the most powerful tool currently available to molecular ecologists.

The first widely used sequencing method was Sanger sequencing. The method, also called chain termination method, is based on repeated amplification of the DNA strand in presence of modified nucleotides, the dideoxynucleotides (ddNTPs) that prevent the addition of further nucleotides and stop the amplification.

While the original Sanger method used radioactively labelled primer or ddNTPs in four separate reaction tubes and in combination with four lanes on polyacrylamide gels for size separation (Sanger *et al.* 1977), this method has been automated in order to sequence more DNA in short time. First radioactive labelling was replaced by fluorescent ones (Martin *et al.* 1985) and the sequencing reaction was performed in one tube with each of the four ddNTPs labelled with a different fluorescent dye. A fluorescence sensor identifies each nucleotide based on the fluorescence emitted at a different wavelength (Smith *et al.* 1986, Ansorge *et al.* 1987).

Although this method of sequencing was the standard for several years, nowadays it has been largely replaced by automated Next- Generation Sequencing (NGS) methods. NGS methods permit parallel sequencing of even millions of DNA fragments which lowers the cost and increases the throughput of the process. First NGS methods described in 2005 by 454 Life Sciences enabled the simultaneous analysis of almost a million sequencing templates perched in the picotiter plate. It is based on sequencing by synthesis, also called pyrosequencing, method, published by Ronaghi *et al.* 1996. This method uses the luciferase enzyme which produces a light signal when activated by ATP, which is generated during incorporation of nucleotides that are added sequentially in a fixed order to the newly synthesized DNA strand. This approach results in a 100-fold increase in throughput and 6-fold cost reduction over the current Sanger sequencing technology (Margulies *et al.* 2005)

In the past decades, more NGS platforms were developed and made accessible for numerous laboratories. Nowadays the most common method used in research and clinical labs are the Illumina MiSeq System and the Ion Torrent Personal Genome Machine (PGM) System from Life Technologies. Both methods made sequencing more cost effective, faster and increased the throughput to 1Gb of sequence yield per run for Ion Torrent PGM and even 1.5-2Gb per run for MiSeq Illumina, at the same time producing high quality data. The accuracy for Ion Torrent reaches 98% and for Illumina 99.9%(Quail *et al.* 2012).

In the Illumina Mi Seq System, single stranded DNA fragments are ligated to generic adaptors that bind to a flat surface, followed by amplification *in situ* and formation of dense amplicon clusters. These clusters then form the templates for sequencing by synthesis with reversible terminator dNTP labelled with four different fluorescent dye. Each dNTPs added to the DNA strand of a cluster emits fluorescence which is imaged on the slide (Bentley *et al.* 2008)

The Ion Torrent technology instead of optical signals is based on pH changes which occur when dNTP is added to a DNA colony and an H^+ ion is released as a by-product. conductive to prepare the colonies, individual DNA strands with ligated adaptors are bound by hybridization to beads covered with adaptor complementary sequences. The DNA on the bead is amplified by

emulsion PCR, creating individual colonies from each DNA fragment. The beads are then applied to separate wells on a slide. The slide is cyclically flooded with a solution containing a single dNTP, buffers and polymerase. If the dNTP is incorporated into the strand, the pH changes, which is detected by the ion sensor. If two or more dNTPs are incorporated in the single cycle, the voltage is multiplied so the chip records two or more identical bases (Rothberg *et al.* 2011).

1.3 DNA barcoding and metabarcoding

When Sanger sequencing became an easy-access technique, researchers started to search a standardized protocol for species genotyping based on the use of a short DNA sequence from a particular region of the genome, thus providing a 'barcode' for species identification. In order to find universal markers that could be valid for all species, various markers were tested in a The International Barcode of Life project (iBOL). The core barcode for animals is the mitochondrial gene *cytochrome oxidase I* (*COI*) (Hebert *et al.* 2003, Waugh 2007). However, *COI* is not an appropriate barcode for most of the plant species (Kress *et al.* 2005). Therefore, various DNA regions where tested including the chloroplast genes *rbc-L*, *matK*, *rpoB*, *rpoC1* and the intergenic spacers *trnL-F*, *trnH-psbA, atpF–atpH, psbK–psbI* and ITS (Kress *et al.* 2005, Cowan *et al.* 2006, Fazekas *et al.* 2008, Hollingsworth *et al.* 2009 and 2011). Nevertheless, no 'golden mean' has been found. The barcode recommended for fungi is ITS (Schoch *et al.* 2012).

The Consortium for the Barcode of Life (CBOL), established in 2004, is an initiative dedicated to the development of DNA barcodes (http://www.barcodeoflife.org/). As a part of its activity scientist progressively define the barcode markers for each taxonomic group and standardize DNA barcoding protocols. Markers standardized by CBOL are accessible via the barcode of life data system (iBOLD) on http://www.boldsystems.org (Ratnasingham & Hebert 2007). The newly developed laboratory techniques gave researchers also the opportunity to analyse complex environmental samples containing a mixture of unknown species. This type of samples, so called metabarcoding samples, are used in a variety of fields, including microbial ecology, food safety, aero- and soil biology or ecosystem monitoring (Quéméré *et al.* 2013, De Barba *et al.* 2014). DNA metabarcoding combines two technologies:

DNA barcoding and NGS. The laboratory protocols of extracting, amplification and sequencing are followed by bioinformatic analysis of the generated sequence reads (Coissac *et al.* 2012).

Although DNA barcoding is very promising in biodiversity research, it presents a high dependency on PCR (Taberlet *et al.* 2012). Errors during amplification and sequencing, mispriming or degradation of the DNA template can significantly influence experiment results. This issue is more profoundly described in Chapter 3.

2. *Cistus heterophyllus* endangered species

2.1 *C.heterophyllus* taxonomy

C. heterophyllus was first described by Desfontaines in 1798.
Kingdom: Plantae
Clade: Angiosperms
Clade: Eudicots
Clade: Rosids
Order: Malvales
Family: Cistaceae
Genus: *Cistus*
Species: *Cistus hetrophyllus*

The taxonomy up to the family level (Cistaceae) - followed Angiosperm Phylogeny Group (2009). The taxonomy of Cistus genera described by Demoly & Montserrat (1995) has traditionally been based on vegetative (nerve number, shape, and hairiness of leaves) and reproductive characters (sepal number, petal colour, style length, and number of fruit valves).

Phylogenetic analysis performed by Guzmán & Vargas (2005) supported the existing taxonomy. Two major lineages within the Cistus genera, a purple-flowered clade (including *C. hetrophyllus*) and a white-flowered clade, were delineated (except *Cistus parviflorus* which has the purple flowers of the purple-

flowered clade, but the sessile stigmas of the white-flowered clade) (Fig. I.1). Latest studies also classified *C. heterophyllus* in the purple pink flowered clade of Cistus species, but in a subclade only with *C. albidus* and *C. creticus* (Civeyrel *et al.* 2011)

C. heterophyllus subsp. *carthaginensis* has two subspecies described by Crespo & Mateo (1988):

- *C. heterophyllus* subsp. *heterophyllus* – mainland Spain
- *C. heterophyllus* subsp. *carthaginensis* – north Africa

However, Jiménez *et al.* 2007 disputes the existence of these two subspecies based on genetic differentiation using RAPD markers.

2.2 *S*pecies description

C. heterophyllus is distributed over the Mediterranean coast, the south-eastern part of the Iberian Peninsula and the northern part of Africa (Morocco, Algeria). Together with others *Cistus* species (*C. albidus*, *C. ladanifer*, *C. salvifolius*, *C. laurifolius* and *C. monspeliensis*) it is the dominate Mediterranean evergreen scrub. Similar to other species, it shows an adaptation to Mediterranean environments through its ecological characteristics as fire dependent seed germination, insect-dependent pollination and spring dependent phenology (Navarro Cano *et al.* 2008).

C. heterophyllus is an erect, much-branched shrub, growing up to 80-90 cm tall. Branches are short and rigid, with a reddish brown, fibrous bark clothed with dense stellate-fasciculated and simple, long hairs.

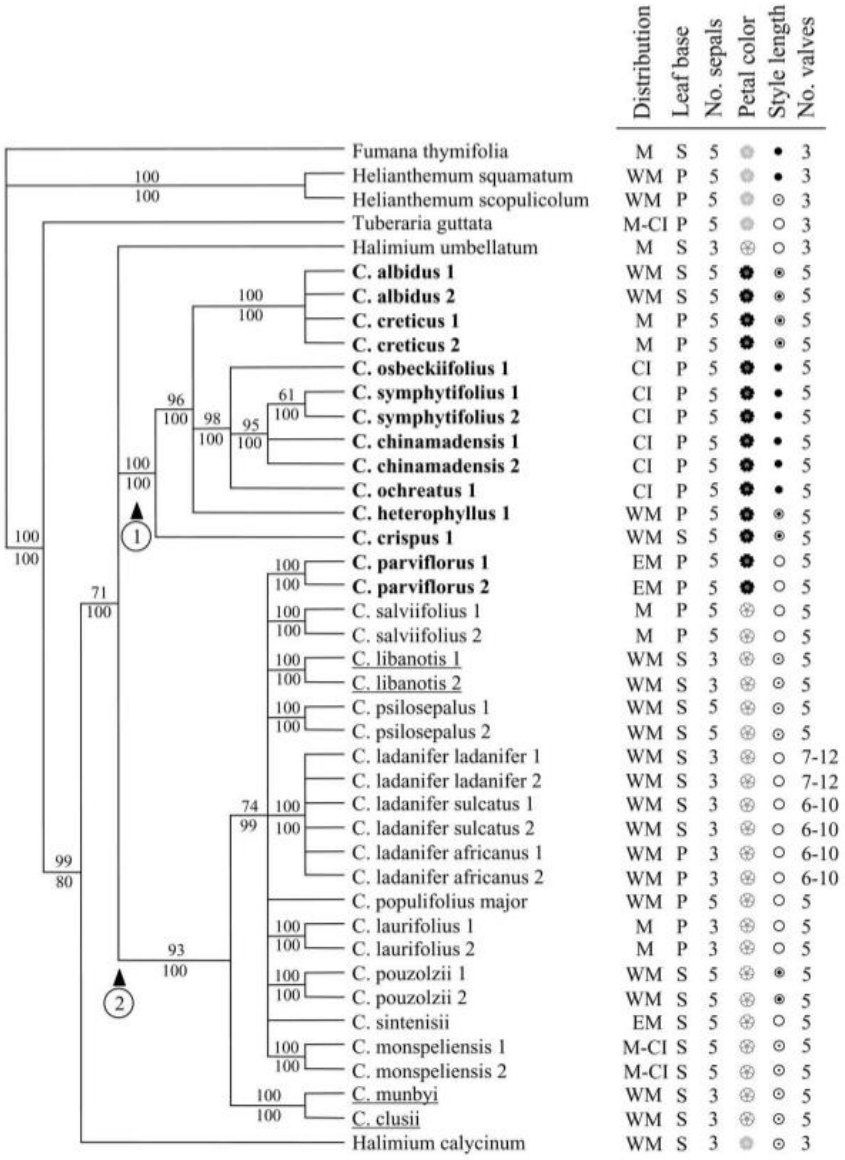


Figure I.1 Strict consensus tree from the combined analysis of *trnL-F, matK*, and ITS sequences. *F. thymifolia* - the outgroup taxon. Numbers above branches - bootstrap values. Numbers below branches - posterior probabilities. Species distribution (M, Mediterranean; WM, western Mediterranean; EM, eastern Mediterranean; and CI, Canary Islands) and five morphological characters: leaf base (P, petiolate; S, sessile), number of sepals (3, 5); petal colour (, yellow; , white; and , purple); style length (, sessile; , shorter than stamens; , as long as stamens; and , longer than stamens); and number of fruit valves (3, 5, 6 or more). Taxa circumscription in subgenera is coded as follows: Cistus (in bold); Leucocistus (in roman), and Halimioides (underlined) (Guzmán & Vargas 2005)

Leaves are sheating at the base, reticulately veined underneath, the upper surfaces are dark green with stellate and simple hairs, and the lower surfaces are whitish with a coating of short hairs, usually 5–20 mm (0.2–0.8 in) long. Upper leaves are attenuate, elliptical to lanceolate in shape, with short or without petiole sand leaf margins are turned under (revolute). The lower leaves are round or ovately rounded, with short stalks, the leaf margins are slightly turned under (revolute).

The flowers are large (up to 60 mm), terminating the branches, arranged in cymes of one to five individual flowers, each with five purplish-pink petals, usually with a yellow spot at the base. The five sepals have stellate hairs, plus some longer simple hairs. Stamens numerous (100-150), surrounding the style. Peduncles have brownish red color, hairy, one-flowered with two leafy, lanceolate, sessile brackets.

The fruit capsule is about 9 mm high, brown covered with simple hairs, containing angular brownish seeds (Warburg 1968, Navarro Cano *et al.* 2008).

The two subspecies described before have different distribution of hairs. In *C. h.* subsp. *heterophyllus*, the young stems have many stellate hairs and many longer simple hairs whereas the leaves have scattered long simple hairs. In *C. h.* subsp. *carthaginensis*, the young stems have many stellate hairs and fewer longer simple hairs whereas the leaves – scarce have simple long hairs. Moreover, the outer two sepals of *C. h.* subsp. *carthaginensis* are smaller (8.5-mm long by 6-mm wide) than in *C. h.* subsp. *heterophyllus* (average 10 mm long by 9-mm wide). Petals and flowers of *C. h.* subsp. *carthaginensis* are also smaller than petals and flowers of *C. h.* subsp. *heterophyllus* (Ferrer-Gallego & Ferrando 2013).

2.3 Ecological situation of *C. heterophyllus*

C. h. subsp. *heterophyllus* is distributed on the Mediterranean coast of north Africa in Morocco and Algeria (Fig. I.2).

The distribution of *C. h.* subsp. *carthaginensis* Pau (Crespo & Mateo 1988) is limited to only two populations located on the Spanish Mediterranean coast: only one individual in la Pobla de Vallbona (Valencia) and the unique natural population in Europe, discovered in 1994 in Sierra Minera (Murcia region) in the Regional Parque de Calblanque, Monte de las Cenizas y Peña de Águila (Navarro Cano *et al.* 2008). This population comprises 22 individuals, ten of them present a phenotype indicating possible hybridization events with *C. albidus* species.

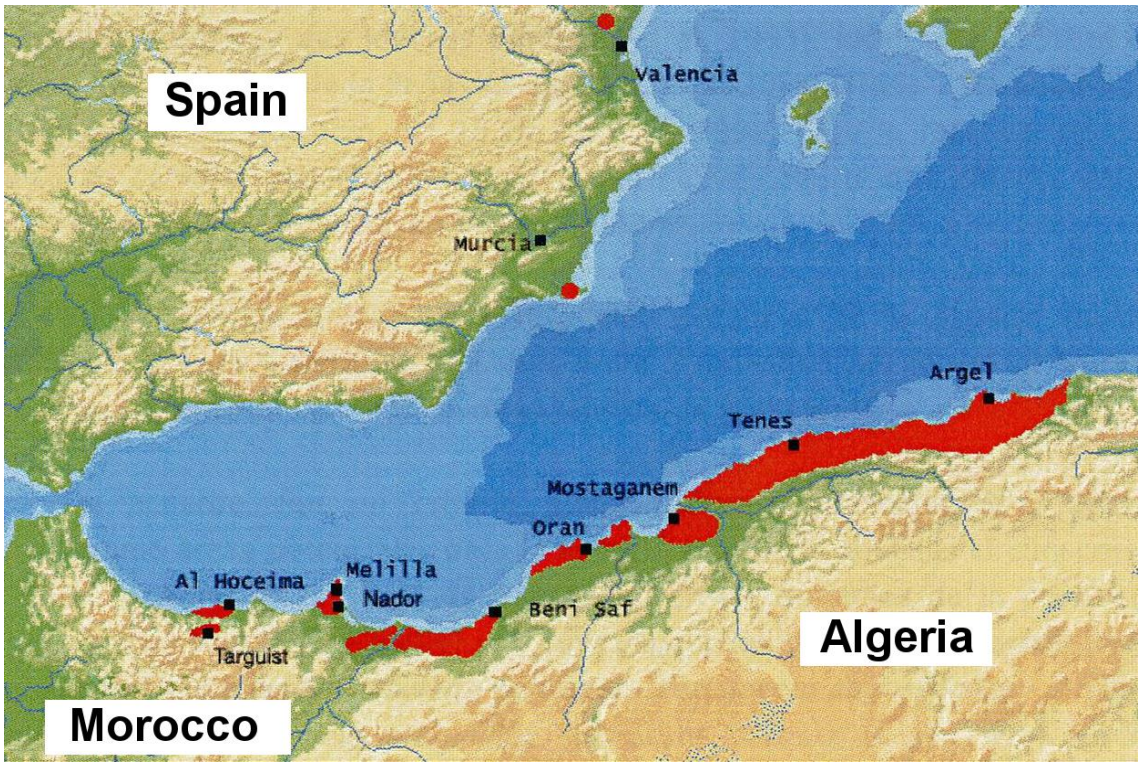


Figure. I.2 *C. heterophyllus* distribution (Navarro Cano *et al.* 2008)

In the beginnings of XX century, individuals of *C. heterophyllus* subsp. *carthaginensis* were abundant in Sierra Minera (Sancti Spiritu Mountain and Peña del Aguila) of Murcia region. These individuals were described last by Jimenéz in 1903 and 1908. In the 50s of the XXth century because of the surface mining activity and the degradation of natural habitat, the *C. heterophyllus* population in this location was considered to be extinct. Nevertheless, in 1993, a population of seven adult and two young individuals was rediscovered (Robledo

et al. 1995). This newly discovered population was destroyed by fire in 1998 but one year later it reactivated from bank seeds remaining in the soil and stimulated to germinate by the fire. This population exists up till now and was investigated in this project. The newly regenerated population comprises individuals showing a possible hybrid phenotype, similar to that from *C. heterophyllus* × *C. albidus* described by Font Quer & Maire in Cavanillesia III (1930) in northern Africa.

Sudden extinction of the numerous individuals indicates that the *C. heterophyllus* population in Murcia region is placed at a recent genetic bottleneck. The unexpected reduction of the size of this population due to fires and human disturbance led to the reduction of the genetic diversity that influences the robustness of the population and its ability to survive selecting environmental changes.

An endogamic depression is one of problems leading to species extinction, especially in endemic or isolated populations (Frankham 1998). To avoid endogamic depression, plants developed mechanisms that prevent autopolinization (Ivanov *et al.* 2010) and as its consequence reduction of the genetic variability (Spielman *et al.* 2004). These mechanisms include autoincompatibility, a common trait in the Cistaceae family (Herrera 1992, Talavera *et al.* 1993). Studies on the reproductive biology of *C. heterophyllus* (Boscaiu & Güemes 2001) confirmed gametophitic autoincompatibility also for this species.

As the *C. heterophyllus* population from Murcia region is the unique population of this subspecies, the number of individuals is decreasing and the survivability of individuals is low. It is presumed that hybridization events led to an increase in genetic variability. But at the same time, it is the major danger for the species existence (Navarro Cano *et al.* 2008). A spontaneous hybridization process between clones of the unique *C. heterophyllus* individual from Valencia was also described (Boscaiu & Güemes 2001).

2.4 Conservation strategy of the endangered species

Cistus heterophyllus subsp. *carthaginensis* is listed as critically endangered (CR) in the IUCN Red List (Güemes *et al.* 2006). The protection of this species includes both *ex situ* and *in situ* conservation strategies. The *ex situ* activities are: seed storage and reintroduction of seedlings in Murcia region. Now, there are 7 populations descendent of seeds from the Murcian population before the fire in 1998 (Navarro Cano & Rivera 2001) and from seeds from anterior populations (Navarro Cano *et al.* 2008). The attempts of the reintroduction of *in vitro* cultivated plant were not successful (Rosato *et al.* 2016).

3. State of art

Most studies concerning *C. heterophyllus* species are based on morphological analysis such as colour of flowers, leaves morphology etc., or on reproduction strategy of the species (Demoly & Montserrat, 1995)

During the last twenty years, thanks to the development of molecular techniques and the interest in molecular taxonomy or evolutive biology, we observe an increasing number of molecular studies on phylogeny and biogeography of the Cistaceae family.

First results in molecular diversity and taxonomy published for Cistus species were allozyme analysis (Batista *et al.* 2001; Farley & McNeilly 2000).

Amplified Length Polymorphism method (AFLP) was used to describe diversity and genetic structure of *C. ladanifer* (Carlier *et al.* 2008).

A complex phylogenetic and systematic analysis based on plastid DNA markers was presented by Guzmán & Vargas (2005) and Guzmán *et al.* (2009).

The unique molecular study on the population from Murcia (Llano del Beal) was performed by Jiménez *et al.* (2007) using Random Amplified Polymorphic DNA markers (RAPD). His results suggested that all individuals comprising this unique population were hybrids.

In our laboratory, we performed experiments implementing plastid DNA containing recommended DNA barcoding regions (Kress *et al.* 2005; Kress & Erickson 2007). Results show that two genes *rpoB* and *rpoC1* are useful for differentiation between *C. heterophyllus*, *C. albidus* and its possible hybrids. Other analyzed genes as *rbcL, trnK-matK* and intergenic spacer *trn L-F* were not sufficiently variable to be informative in case of closely related species.

Being aware that not only molecular data are a unique and reliable source of data, we combine DNA analysis with morphological analysis in our studies. During this experiments we observed that PCR efficiency for the same plastid marker was very different among species and consequently also the final amount of amplification product. This bias may be intensified during further amplification

steps during next generation sequencing. This problematic is relevant especially for the analysis of environmental samples, possible leading to underestimation or non-detection of species within the sample mixture.

4. Objectives

1. Development of plastid molecular markers for determination of the genetic structure of the unique population of *C. heterophyllus* subsp. *carthaginensis* located in Llano del Beal described by Navarro Cano (2008).

2. Phenotypic and molecular identification of wild type and hybrid individuals form the Llano del Beal population.

3. Application of chloroplast markers for the classic- and NGS barcoding strategies and assessment of their biases in metabarcoding studies.

Chapter 1 - Two alleles of *rpoB* and *rpoC1* distinguish an endemic European population from *Cistus heterophyllus* and its putative hybrid (*C.* × *clausonis*) with *C. albidus*

1.1. Introduction

Cistus is a genus of flowering plants from the rockrose family (Cistaceae), containing 21 species (Guzmán *et al.* 2009). These perennial, evergreen shrubs are characteristic of the Mediterranean region and are also found on the Canary Islands (Batista *et al.* 2001). Their big, visible flowers and resistance to harsh environmental conditions such as drought, poor and fire-degraded soils (Carlier *et al.* 2008; Ellul *et al.* 2002; Roy and Sonié 1992) have made these species common ornamental plants.

The most common species in Southern Spain are *C. albidus* and *C. monspeliensis* (Demoly & Montserrat, 1995). *C. heterophyllus* Desf. (1978) is an Ibero – African endemic species which is highly endangered in Europe. African individuals of *C. heterophyllus* are classified as *C. heterophyllus* subsp. *heterophyllus* whereas several individuals belonging to two populations in Spain (only locations in Europe) are classified as *C. heterophyllus* subsp. *carthaginensis* (Crespo and Mateo 1988). They can be found in two locations in the south-east on the Iberian Peninsula: one individual in La Pobla de Vallbona (Valencia) and the unique natural population in Llano del Beal in Parque Regional de Calblanque, Monte de las Cenizas y Peña del Águila (Murcia). In 1998 the Murcian population was destroyed by fire and it recovered itself one year later. Within this unique natural population of 22 plants, two distinct morphologies allow the separation into two subtypes. Twelve individuals resemble what would be a pure *C. heterophyllus* subsp. *carthaginensis* type whereas ten plants show an intermediate phenotype similar to hybrids described as *C.* × *clausonis* (*C. heterophyllus* × *C. albidus*) from northern Africa in Carvanillesia III by Font Quer and Maire (1930). These individuals throughout the text are called as *C.* × *clausonis* subsp. *carthaginensis*. Hybridization processes amongst *Cistus* species

have been described in the early XXth century (Card 1910; Simonet and Ansereau 1939). Furthermore hybridization between *C. albidus* and *C. heterophyllus* has been reported in the Microrreserva de flora de Tancat de Portaceli between clones of a valencian individual (Navarro Cano *et al.* 2008), suggesting that the genus *Cistus* can have some degree of cross hybridization.

There are no previous molecular studies to examine hybridization events in *Cistus*. Allozyme *loci* polymorphisms have been used to analyze populations of *C. salvifolius* (Farley & McNeilly 2000) but this type of markers requires genes encoding well-known proteins to be detectable. Very few of this type of markers are available and standardization of experimental procedure between laboratories is difficult (Lee *et al.* 2002). Application of RAPD markers suggested an introgression in the *C. heterophyllus* subsp. *carthaginensis* population from Murcia (Jiménez *et al.* 2007). A classification of the Cistaceae family based on plastid genes, including *C. heterophyllus* and *C. albidus*, confirms the existing taxonomy, but does not provide any information about intra-specific diversity of these species (Guzmán & Vargas 2005). In order to investigate the origin of *C.* × *clausonis* subsp. *carthaginensis*, we therefore consider it necessary to develop molecular markers at the population level for species of *C. heterophyllus* and *C. albidus*.

The barcoding DNA regions have been successfully adopted in many plant groups as a useful and informative method not only in species identification, but also in molecular phylogenetics and population genetics. We used selected plastid genome regions recommended by the 'Plant Working Group of the Consortium for the Barcode of Life (CBOL)': *rbcL, trnK-matK, rpoB, rpoC1, trnH-psbA* and *trnL-F* (Chase *et al.* 2007; Kress *et al.* 2005; Kress and Erickson 2007). It was suggested that these chloroplast markers could provide preliminary information of the extent and nature of population divergences and support comparative studies on population diversity (Hajibabaei *et al.* 2007). Molecular markers from the plastid genome are specially advantageous because they are considered as haploid, structurally stable, uniparentally inherited and non-recombinant (Guzmán *et al.* 2009). However chloroplast heteroplasmy (presence of more than one plastid genome within an individual) (Chat *et al.* 2002; Frey 1999) suggests

that sampling strategies could strongly influence phylogenetic outcomes (Wolfe and Randle 2004). In the present study, we performed plastid haplotype screening based on sequence comparison analysis in order to determine the origin of the local populations of *C. heterophyllus* and *C. × clausonis*.

1.2 Materials and methods

1.2.1 Sampling of plant material

Plant material (young leaves) of C. *heterophyllus* subsp. *carthaginensis* (12 individuals) and *C. × clausonis* subsp. *carthaginensis* (10 individuals) was collected from a population of Llano del Beal in Peña del Águila (Murcia). We also sampled one artificial population from Cartagena (Murcia) which originates from seeds of a Llano del Beal population recollected before a fire in 1998 and used by local authorities for *Cistus* reintroduction comprising three *C. albidus*, three *C. heterophyllus* and three *C. × clausonis*. Samples (dry material) of African C. *heterophyllus* subsp. *heterophyllus* (2 individuals) and *C. × clausonis* (1 individual) were obtained from two populations in Kebdana Mountains and Alhucemas. Additional samples of *C. albidus* were obtained from 5 populations (in total 10 individuals) from the south-east Spain. Seven *C. monspeliensis* samples were collected from the surroundings of the *C. × clausonis* subsp. *carthaginensis* population in Llano del Beal. Localities of sampled individuals were recorded using Garmin Colorado 300 GPS receiver (Fig. 1.1 and Supplementary Material - Table A.1.).

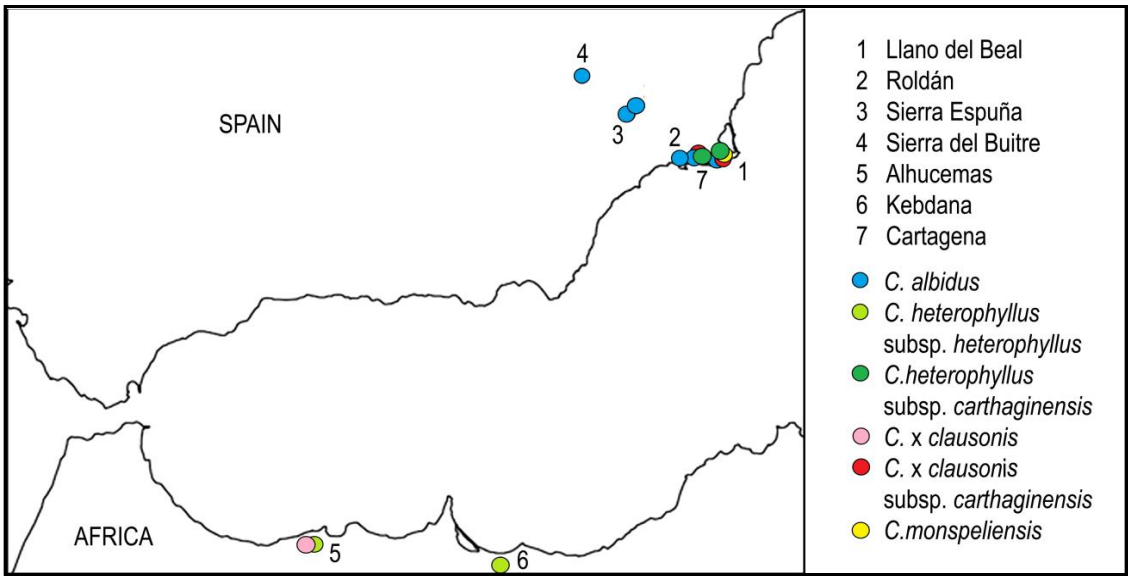


Figure 1.1 Localities of the samples examined in this study

1.2.2 Leaf and trichome analysis

Total leaf area was measured using ImageJ software available at (http://rsb.info.nih.gov/ij/) as described Delgado-Benarroch *et al.* (2009). We measured six leaves from five randomly chosen plants of each species (*C. heterophyllus* subsp. *carthaginensis, C.* × *clausonis* subsp. *carthaginensis* and *C. albidus*), n=30 for each species. Statistical analysis (Wilcoxon test) was performed with the R Stats Package (http://cran.r-project.org/). Photographs of the trichomes were taken using a Leica Stereomicroscope MZFLIII and Leica DC300F digital camera.

1.2.3 DNA extraction, cloning and sequencing

Fresh leaves were dried in silica gel and stored at -80 °C. Total genomic DNA was extracted using the commercial kit 'Plant NucleoSpin' (Machery and Nagel, Düren, Germany). Selected markers were amplified with GoTaq Polymerase (Promega, Madison, WI, USA) under the following PCR conditions: 95 °C for 2 min., 30-35 cycles: 95 °C for 30 s, 50-55 °C (Supplementary material - Table A.2) for 30 s and 72 °C for 1 min. The primers used in this experiment (*trnK-matK, rpoB, rpoC1, trnH-psbA*) have been described previously (Kress *et*

al. 2005; Kress and Erickson 2007) or were designed on the basis of sequences from GenBank (*rbcL* and *trnL-F*). PCR fragments were cloned using the *pGem T-Easy* kit (Promega, Madison, WI, USA). Five clones per single individual (*C. albidus*, *C. heterophyllus* subsp. *carthaginensis* and *C.* × *clausonis* subsp. *carthaginensis*) were amplified with gene-specific primers and were sequenced on an Abi Prism 3130XL Genetic Analyzer (Applied Biosystem, Foster City, CA, USA). Sequence accession numbers for the sequences are JF900405–JF900462.

1.2.4 Sequence analysis

Sequences were edited with the CodonCode Aligner V 3.5 software (CodonCode Co., Dedham, MA, USA) and manually corrected. Base quality assessment was performed with the same software according to Phred scores (Richterich 1998) and visual revision of chromatograms. Only bases with a quality value over 20 were accepted for further analysis. Sequence alignments were performed in CustalW2 with standard parameters (www.ebi.ac.uk/Tools/msa/clustalw2/).

Both intra- and inter-specific distances for *rbcL*, *trnK-matK* and *trnL-F* for *C. heterophyllus* and *C. albidus* species were computed using Kimura-2-Parameter (K2P) model in PAUP*4.0b10 (Swofford 2002). Average intra-specific distance and coalescent depth were used to characterize intra-specific variation (Meyer and Paulay 2005). Average inter-specific distance (Meyer and Paulay 2005) and the smallest inter-specific distance (Meier *et al.* 2008) represented the inter-specific divergence. For *rbcL* gene distance analysis we use also tree GenBank sequence accessions (FJ492042, FJ225860, FJ225868)

1.2.5 Real-time PCR, melting analysis for *rpoB* and *rpoC1* genes and identification of polymorphisms by restriction digestion

The *loci rpoB* and *rpoC1* of *C. heterophyllus* subsp. *heterophyllus* and subsp. *carthaginensis*, *C. albidus*, *C.* × *clausonis* subsp. *carthaginensis*, African *C.* × *clausonis* and *C. monspeliensis* were analyzed with the Mx3000P Q-PCR System using the SYBR Premix ExTaq™ (Takara Biotechnology, Dalian, China) with ROX as a reference dye. Primers used were the same as designed for *rpoB* and *rpoC1* genes sequencing (Supplementary Material – Table A.2). PCR

conditions were as following: 95°C for 5 min., 40 cycles of 95°C for 5s, 55°C for 20s and 72°C for 15s. Dissociation profiling was performed by applying a hold time of 1 min. at 95°C and increasing the temperature from 55 to 95° C rising by 0.5 °C per step.

We compared the melting curve of *rpoC1* gene amplification product for standard SYBR Green against the Type-it HRM kit (Qiagen, Valencia, CA, USA) containing Eva Green fluorescent dye. Representative groups of 5 samples for *C. heterophyllus* subsp. *carthaginensis*, 5 samples of *C. albidus* and 2 positive controls consisting in the cloned allele A and B were amplified under the following PCR conditions: 95°C for 5 min., 40 cycles of 95°C for 10s, 55°C for 30s and 72°C for 20s. For melting curve generation, a hold time of 1 min. at 95°C and a temperature range from 65 to 95°C rising by 0.5°C per step was applied.

To differentiate between *rpoC1* alleles A and B, PCR products of all sampled individuals were digested with *Cla*I restriction enzyme (Fermentas, Hanover, MD, USA) at 37°C overnight and separated by agarose gel electrophoresis.

Statistical analysis (Fligner test, Shapiro test, discriminant analysis) of RT-PCR melting data was performed with the stats package of the R environment (www.r-project.org). As the preliminary step for the discriminant analysis, melting data for *rpoC1* gene were modified according to data obtained from the restriction digestion analysis.

1. 3 Results

1.3.1 Phenotypic characteristics of individuals

The unique population of *C. heterophyllus* subsp. c*arthaginensis* from Llano del Beal is comprised of 22 individuals that display two distinct phenotypes, one group resembling what is described as pure *C. heterophyllus* subsp. *carthaginensis* (Fig. 1.2a) and a second group of plants that show intermediate vegetative phenotypes between *C. albidus* (Fig. 1.2b) and *C. heterophyllus*. This second group could be the result of hybridization between these two species and we refer to them as *C.* × *clausonis* subsp. *carthaginensis*

(Fig. 1.2c). Visual inspection of leaves from *C. albidus* obtained from the same hills at a distance of over three hundred meters, *C. heterophyllus* subsp. *carthaginensis* and *C.* × *clausonis* subsp. *carthaginensis* plants showed that the leaves of the *C.*× *clausonis* subsp. *carthaginensis* were intermediate in size between *C. albidus* and *C. heterophyllus*. Moreover, they resembled the African *C.* × *clausonis* (Fig. 1.2d).

Figure 1.2 Pictures of **a**. *C. heterophyllus* subsp. *carthaginensis*; **b.** *C. albidus*; **c.** *C.* × *clausonis* subsp. *carthaginensis* presenting intermediate phenotype; **d.** phenotypes of *Cistus* leaves: 1. *C. albidus*, 2. *C.* × *clausonis,* 3. *C.* × *clausonis* subsp. *carthaginensis*, 4. *C. heterophyllus* subsp. *heterophyllus*, 5. *C. heterophyllus* subsp. *carthaginenis*

Analysis of total leaf area by Kruskal-Wallis rank sum test showed significant differences between *C. albidus* vs. *C. heterophyllus* subsp. *carthaginensis* and *C.* × *clausonis* subsp. *carthaginensis* ($p < 0.05$). The post-hoc pairwise comparison using Wilcoxon test with the Bonferroni correction also gave significant differences between C. *heterophyllus* subsp. *carthaginensis* and the *C.* × *clausonis* subsp. *carthaginensis* ($p<0.05$) (Fig. 1.3). However, between these two groups of plants there is a considerable size overlap concerning leaf area suggesting that this parameter could be misleading in their discrimination.

Trichomes have been used as criteria for taxonomic assessment and are considered as reliable morphological features for *Cistus* species identification (Beilstein *et al.* 2006; Gulz *et al.* 1996; Hoot 1991; Khalik 2005).

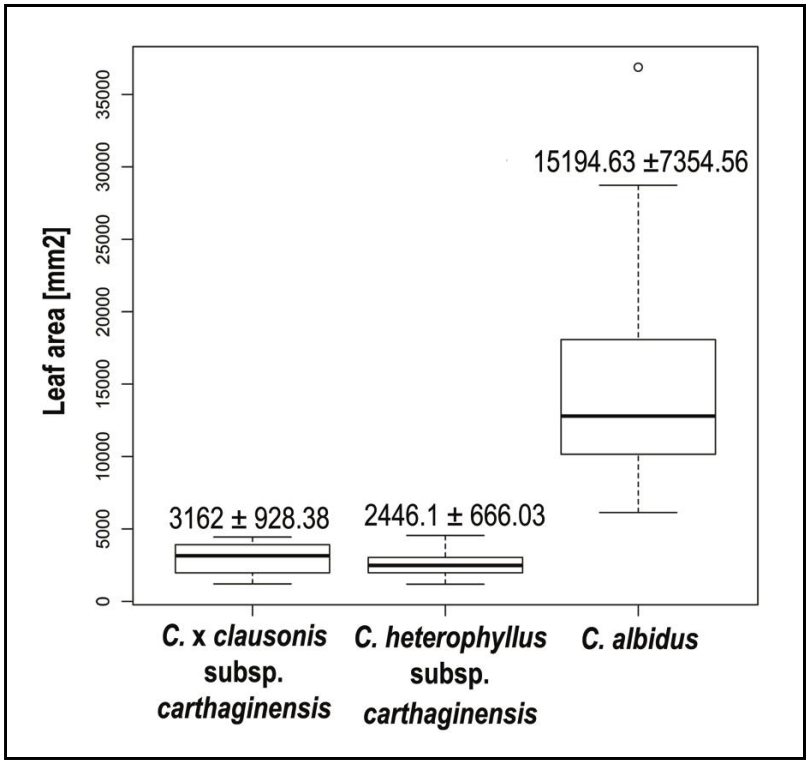


Figure 1.3 Total leaf area of *C.* × *clausonis* subsp. *carthaginensis*, *C. heterophyllus* subsp. *heterophyllus* and *C. albidus*

We analyzed their structure using stereomicroscopy. Long and thin star-shaped trichomes covered densely the adaxial surfaces of *C. albidus* leaves (Fig. 1.4 a,d). In *C. heterophyllus* this type of trichomes were short and thick, adhering to the lamina and less densely distributed than in *C. albidus* (Fig. 1.4 b,e). Trichomes of *C.* × *clausonis* subsp. *carthaginensis* showed intermediate phenotypes resembling *C. albidus* long and thick arms but more flattened as in *C. heterophyllus* (Fig. 1.4 c,f).

Leaf size and trichome shape allow easy determination between pure species of *C. albidus* and *C. heterophyllus*. However *C.* × *clausonis* subsp.

carthaginensis displayed an intermediate phenotype. Thus, it was not possible to constitute clear criteria and this kind of phenotypic classification can only be a preliminary step before molecular studies.

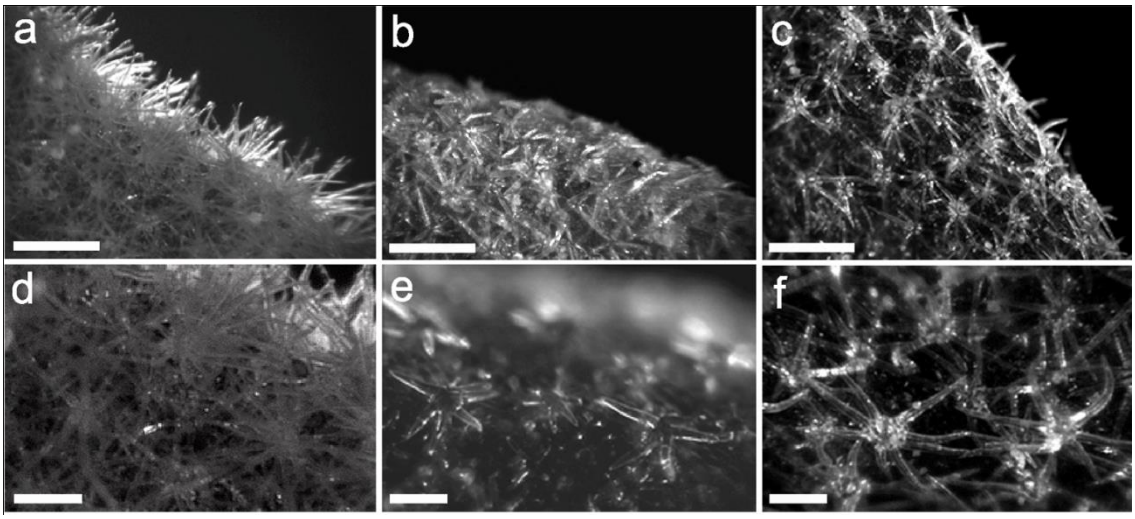


Figure 1.4 Trichomes of *C. albidus* (**a**,**d**), *C. heterophyllus* subsp. *carthaginensis* (**b**,**e**) and the *C.* × *clausonis* subsp. *carthaginensis*(**c**,**f**). White bars represent 15 μm

2.3.2 Molecular analysis

PCR amplification products of the genes *rbcL, rpoB* and *rpoC1* consistently produced high Phred quality scores sequences (Table 1.1). Sequences *trnK-matK, and trnL-F* intergenic spacer showed lower sequence quality (5 sequences of *trnL-F* had to be discarded) even so this information could be successfully analyzed after trimming and manual edition. The *trnH-psbA* region presented low quality (more than 60% of the base pairs were below 20 of quality score). Thus, this marker was excluded from further analysis.

Table 1.1 Comparison of analyzed chloroplast regions

Marker	Average length unedited/ analyzed (bp)	Number of alleles per plant	No. variable sites noninformative / informative sites	Quality value	GenBank accession no.
rbcL	1036 /629	1	4/0	51.96 ± 1.38	JF900445 - 47
trnK-matK	968/757	1	36/0	45.76 ± 3.28	JF900448 - 62
rpoB	508/468	2	2/10	55.58 ± 1.92	JF900405 - 19
rpoC1	518/403	2	68/3	55.37 ± 1.67	JF900420 - 34
trnL-F	379/343	-	93/127	45.78 ± 7.09	JF900435 - 44
trnH-psbA	308/ -	-	-	25.82 ± 2.08	-

2.3.3 Determination of intra- and inter-specific distances

In order to clarify possible relations between *C. albidus*, *C. heterophyllus* subsp. *carthaginensis* and the individuals of *C.* × *clausonis* subsp. *carthaginensis* we used the primer combinations described above. From the analyzed sequences *trnK-matK* and *rbcL* genes were highly conserved. *trnK-matK* presented only few uninformative variable sites. *rbcL* clone sequences were identical for each individual of *C. albidus*, *C. heterophyllus* subsp. *carthaginensis* and *C.* × *clausonis* subsp. *carthaginensis*). Therefore, three additional sequences (two of *C. albidus* and one of *C. heterophyllus*) from GenBank were included for the analysis. *trnL-F* intergenic spacer, as it was expected, showed high level of variability (Table 1.1).

Two parameters were used to describe intra-specific variation: average intra-specific distance (K2P) between all samples within species and the average coalescent depth – the maximum distance within each species. The lowest

average intra-specific distance was obtained for the *rbcL* gene as expected since it was highly conserved in all samples analyzed. The highest value for this parameter presented *trnL-F* intergenic spacer. The same tendency was observed for the coalescent depth (Table 1.2).

Table 1.2 Analysis of inter-specific divergences and intra-specific variation of analyzed barcode regions

Marker	***rbcL***	***trnK-matK***	***trnL-F***
Average intra-specific distances	0.0017 ± 0.0024	0.0043 ± 0.0012	0.0049 ± 0.0036
Coalescent depth	0.0026 ± 0.0036	0.0098 ± 0.0028	0.0150 ± 0.0128
Average inter-specific distances	0.0019 ± 0.0026	0.0053 ± 0.0029	0.0066 ± 0.0058
Minimum inter-specific	0.0000	0.0012	0.0000
'Barcoding gap' ratio	0.89	0.81	0.74

Inter-specific divergence was characterized by two parameters: average inter-specific distance (K2P) and the minimum inter-specific distance between individuals from all species. The highest average inter-specific distance was found using the *trnL-F* intergenic spacer, followed by *trnK-matK* gene, while *rbcL* presented the lowest value (Table 1.2). Both *trnL-F* and *rbcL* minimum inter-specific distance gave zero value indicating that at least two sequences from two analyzed species were almost identical. The minimum inter-specific distance calculated for the *trnK-matK* gene was also close to zero.

To differentiate species by barcodes the inter-specific variation or barcoding gap should be ten fold higher than the intra-specific variation (Hebert *et al.* 2004). The distance values for *rbcL*, *trnK-matK* and *trnL-F* genes were not big enough to show a barcoding gap (Table 1.2). Indeed, the inter-specific variation was less than one fold larger than the intra-specific variation. Furthermore, an overlap for both minimum inter-specific distance and coalescent depth values eliminated these regions as discrimination markers for our species of interest.

1.3.4 Heteroplasmy of *rpoB* and *rpoC1* genes

In contrast to the low information provided by *rbcL*, *trnK-matK* and *trnL-F*, we found two discriminative alleles of *rpoB* and *rpoC1*. We observed 12 bases of difference between allele A and B of *rpoB* dispersed over the 468 bp analyzed fragment (2.57% of divergence). Five clones of *C. albidus* individuals contained the same sequence (allele A). In contrast *C. heterophyllus* and *C.* × *clausonis* subsp. *carthaginensis* had three clones containing the A allele, and two clones for each individual were different (allele B). We also found two alleles of *rpoC1* with even higher divergence (11.41%, 46 bases in 403 bp). All sequences in *C. albidus* and *C. heterophyllus* were identical (allele A), but the individual *C.* × *clausonis* subsp. *carthaginensis* taken for sequencing had allele A sequence in four clones and a divergent one (allele B). Alignments (Fig. 2.5) of the corresponding translated ORFs coded by *rpoB* and *rpoC1* alleles show high variation at the amino acid level that nevertheless corresponds to amino acid changes found in *rpoB* and *rpoC1* genes of other plant species (Moo*re et a*l. 2010; L*ee et a*l. 2006; Sa*to et a*l. 1999).

```
allele A (rpoB)   QVALDSGVCVIAKHQGKIIYTDTEKIVLSGNGDTLRIPLVMYQGSNKNTCIHQT
allele B (rpoB)   QVALDSGVCMIAKHQGKIIYTDTEKLVLSGNMDTLRIPLVMYQGSKKNTCIHQT

allele A (rpoB)   PRVPRDKHIKKGQILADGAATIGGELALGKNVLVAYMPWEGYNFEDAVLISERL
allele B (rpoB)   PRVPRDKNIKKGQILSDSAATIGGELALGKNVLVAYMPWEGYNFEDAVLISERL

allele A (rpoB)   VYEDIYTSFHIRKYEIQTDVTSQGPEKITNEIPHLEAHLLRNLDKNGIVWWGIX
allele B (rpoB)   VYEDIYTSFHIRKYEIQTDVTSQGPEKITNEIPHLEAHLLRNLDKNGIVWLGIX


allele A (rpoC1)  FRETLLGKRVDYSGRSVIVVGPLLSLHRCGLPRELAIELFQTFVIRNLIRKNIA
allele B (rpoC1)  FRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHLA

allele A (rpoC1)  SNIGVAKRQIREKGQIVWQILEEVIQGHPVLLNRAPTLHRLGIQAFQPILVEGR
allele B (rpoC1)  SNIGVAKSKIREKEPIIWEILQEVMQGHPVLLNRAPTLHKLGIQAFQPVLVEGR

allele A (rpoC1)  AICLHPLVCKGFNADFDGDQMAVHVPLSLEAQAEARLLMFFSX
allele B (rpoC1)  AICLHPLVCKGFNADFDGDQMAVHVPLSLEAQAEARLLMFFSX
```

Figure 1.5 Alignment of the translated ORFs coding by: a. two alleles of *rpoB* gene. 162 amino acid fragment analyzed in this study contains 7 polymorphic positions; b. two alleles of *rpoC1* gene. 151 amino acid fragments show high number of polymorphism: 16

As clones corresponded to single individuals and there was a clear case of heteroplasmy for both *C. heterophyllus* and *C.* × *clausonis* subsp. *carthaginensis* individual, we tested the presence of the *rpoB* and *rpoC1* alleles in the rest of the population and on 12 individuals of *C. albidus* from other locations (see below, Fig. 1.1 and Supplementary Material – Table A.1).

2.3.5 *rpoB* discriminates between *C. albidus* and *C. heterophyllus* related individuals

As simple PCR amplification and sequencing without cloning was not feasible for haplotyping due to the heteroplasmy situation, we developed tools that allowed circumventing this problem. We used real-time PCR to obtain allelic discrimination by melting curve analysis. As expected we obtained differing melting temperatures for the *rpoB* A and B alleles. Melting temperature of allele A was 85.04 ± 0.20 °C and 83.38 ± 0.25°C for allele B. In a total of 12 *C. albidus* individuals from 5 different populations only allele A was present, whereas all *C. heterophyllus*, *C.* × *clausonis* (African origin) and *C.*× *clausonis* subsp. *carthaginensis* from Llano del Beal population had both alleles A, and B, absent in *C. albidus* (Fig. 1.6a and Supplementary Material – Figure S.1).

We tested the universality and consistence of the identified polymorphic molecular markers permitting identification of *rpoB* and *rpoC1* alleles, with two different dyes, one containing Eva Green fluorescent dye and a second one with SYBR Green. Both methods allowed correct discrimination. All amplicons produced with Eva Green dye had higher melting temperatures (1.6-1.7°C) than with SYBR Green. However for both genes *rpoB* and *rpoC1* differences of displacement of melting peaks of alleles A and B in PCR with Eva Green and SYBR Green according to Kruskal–Wallis analysis of variance were not significant ($p > 0.05$) (Table 3)(Akopov *et al.* 1988). Therefore we can conclude that independently from the type of fluorescent dye used, identification of *rpoB* and *rpoC1* alleles is consistent.

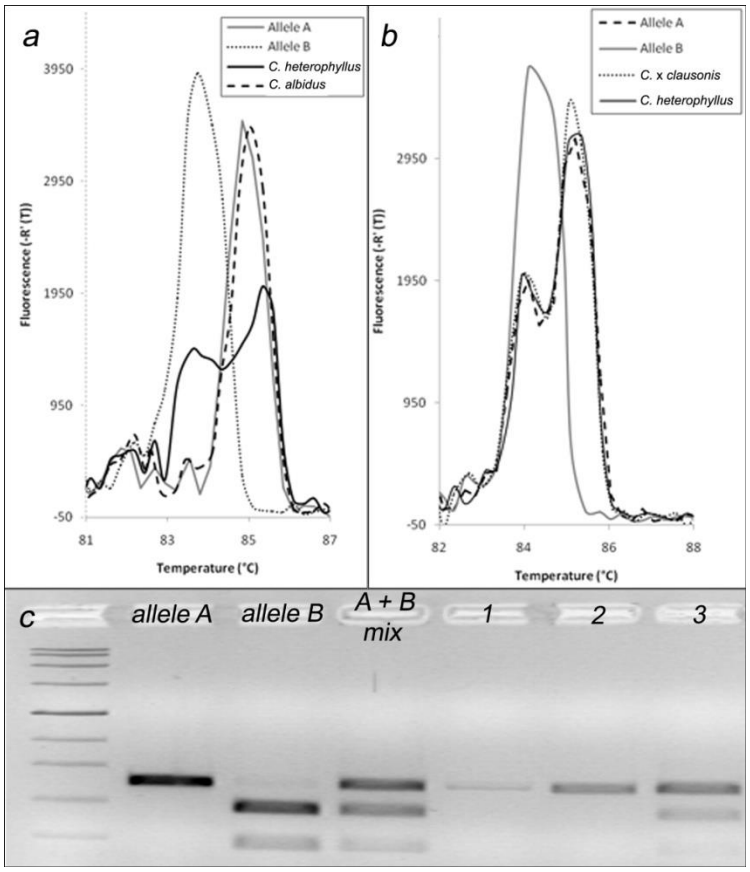


Figure 1.6a Melting curve qPCR analyses from of allele A (85.05 °C) and B (83.38 °C) of the *rpoB* gene present in *C. heterophyllus* individual. In *C. albidus* only allele A is present; **b**. Melting curve for *rpoC1* gene - allele A with double melting peak at 84.10 °C and 85.20 °C and allele B at 84.10 °C. Curve for *C. heterophyllus* without allele B and curve for *C.* × *clausonis* subsp. *carthaginensis* containing allele B; **c.** CAPS marker – *Cla*I enzyme digested only samples containing allele B or mix of two alleles. Samples containing only allele A remain undigested

Table 1.3 Comparison of melting data for two fluorescent dyes SYBR Green (Takara) and Eva Green (Qiagen) .[nd] P > 0.05 in ANOVA

Gene	Allele	Melting data for SYBR Green (Takara) (°C)	Melting data for Eva Green (°C)	Mean of difference in temperature between dyes
rpoB	Allele A	85.00 ± 0.23	83.26 ± 0.27	1.72 [nd]
	Allele B	83.44 ± 0.35	81.81 ± 0.10	1.69 [nd]
rpoC1	Allele A	85.13 ± 0.21	83.45 ± 0.14	1.6 [nd]
	Allele B	84.13 ± 0.18	82.31 ± 0.09	1.81 [nd]

1.3.6 *rpoC1* melting and restriction analysis discriminate between *C.* × *clausonis* subsp. *carthaginensis* and the rest of Cistus accessions

As expected for *rpoC1* we could detect two different melting peaks for allele A at 85.20 ± 0.20 °C and allele B at 84.10 ± 0.16 °C. Allele A was present in all 44 individuals analyzed (*C. albidus*, *C. heterophyllus* subsp. *heterophyllus* and subsp. *carthaginensis*, *C* × *clausonis* and *C.* × *clausonis* subsp. *carthaginensis*) (Supplementary Material – Figure S.1). There were numerous polymorphisms between the two alleles and the melting temperature should be distinct enough to differentiate them in qPCR. However the melting profile of allele A showed an additional peak at the same temperature as allele B (Fig. 1.6b), making it difficult to get clear cut decisions in populations. We found a *Cla*I restriction site between allele A and B by *in silico* restriction enzyme analysis, thus allowing direct amplification and digestion. We were able to confirm the *in silico* prediction as we found digestion of PCR products on allele B but not A. Furthermore, digested products were observed only in *C.* × *clausonis* subsp. *carthaginensis* from Llano del Beal population. We could not find this allele in other samples from Africa, neither in the samples of *C. heterophyllus* subsp. *carthaginensis* or any of the *C. albidus* analyzed (Fig. 1.6c) indicating that the B allele is specific of *C.* × *clausonis* subsp. *carthaginensis*.

As a continuation for a search of the origin of the *rpoC1* gene allele B found in *C.* × *clausonis* subsp. *carthaginensis* but absent in *C. albidus* and *C. heterophyllus subsp. carthaginensis*, we examined seven individuals of *C. monspeliensis* surrounding the population in distances of 300 to 500 meters. The melting profile obtained for *C. monspeliensis* was similar to the *C. albidus* and *C. heterophyllus* profiles, with exception of a second peak which was slightly displaced (Fig. 1.7). As this melting profile of *rpoC1* was not identical to the A allele found in the rest of the plants, we called it allele C. It showed a melting temperature of 84.62 ± 0.13 °C). Furthermore it showed a double peak highly similar to the allele A (Fig.1.7). Considering that *C. monspeliensis* is more distantly related than the rest of the species analyzed, the difference between allele A and C is probably justified. We further performed a restriction analysis and could not find a *ClaI* polymorphism in *C. monspeliensis*. Thus our data rules out *C. monspeliensis* as a possible ancestor of the *C.* × *clausonis* subsp. *carthaginensis*.

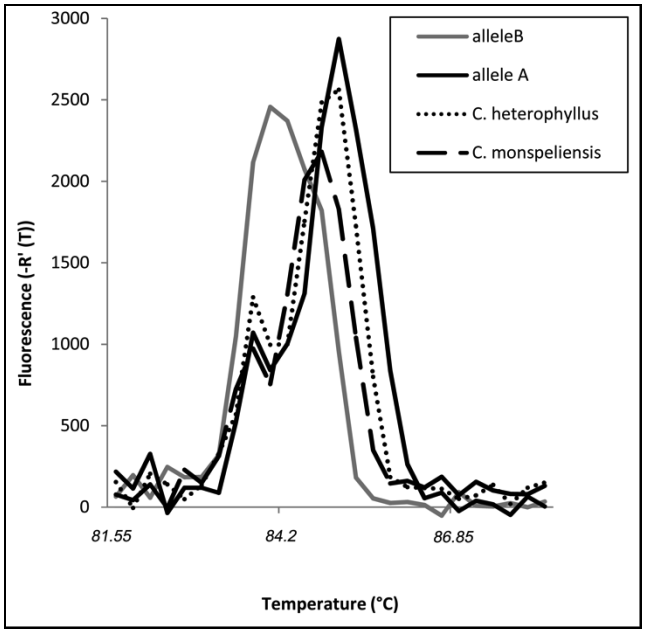


Figure 1.7 Melting curve qPCR analyses for rpoC1 gene in *C. heterophyllus* and *C. monspeliensis*. Allele A presents double peak at 84.10 °C and 85.20 °C, for *C. heterophyllus* an allele B at 84.13 ± 0.18 °C and for *C. monspeliensis* allele C at 84.62 ± 0.13 °C

1.3.7 Discriminant analysis of *rpoB* and *rpoC1* genes

We performed discriminant analysis of *rpoB* and *rpoC1* allelic data by regression model based on DNA melting data for both genes. Discriminant analysis separated individuals into three groups: one containing all *C. albidus* individuals, a second group containing the *C.* × *clausonis* subsp. *carthaginensis* and a third group containing all *C. heterophyllus* European and African and *C.* × *clausonis* from ex-situ material cultivated from seed material obtained from the old population in Llano del Beal (Fig. 1.8). The results suggest that *C.* × *clausonis* subsp. *carthaginensis* is clearly distinct from the rest of the accessions analyzed.

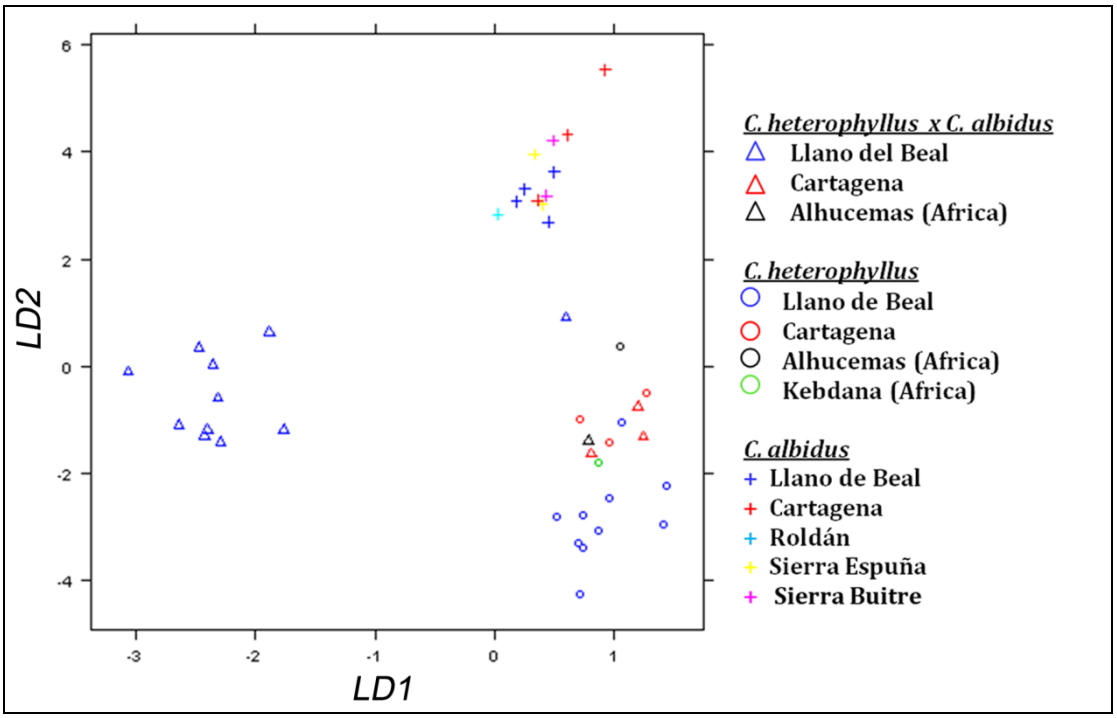


Figure 1.8 Discriminant analysis of *rpoB* and *rpoC1* genes shows three groups of individuals separate: 1st*C. albidus* from all populations, 2nd group containing all European and African *C. heterophyllus*, African and European individual with intermediate phenotype (*C.* × *clausonis*) excluding *C.* × *clausonis* subsp. *carthaginensis* from population in Llano del Beal that constitute 3rd group

1.4 Discussion

1.4.1 Phenotypic markers to study Cistus

Considerable advances in botanical phylogeny have been achieved with morphological characters. Here we show that leaf area can be successfully used to discriminate between *C. albidus* and *C. heterophyllus* but this parameter is less robust to discriminate between *C. heterophyllus* and *C.* × *clausonis* subsp. *carthaginensis*. Several reasons could account for the observed overlap. First a seasonal dimorphism has been described in the genus *Cistus*(De Micco and Aronne 2009). In other species growing in Spain like *Antirrhinum majus*, photoperiod causes important changes in leaf area (Bradley *et al.* 1996), indicating that determination of hybrids cannot always rely on macroscopic characters as they may show strong interactions with the time of the year and environmental conditions. Floral size in contrast has been found to be a stable phenotype in some species (Armbruste*r et al.* 1999; Weis*s et a*l. 2005) but it is more complex to analyze requiring *ex situ* studies to rule out environmental interactions that could affect the outcome of the experiment. Altogether we show that size markers, although easy to follow cannot always give clear-cut results.

1.4.2 Utility of barcode regions in closely related taxa analysis

Surprisingly we found that chloroplast DNA regions *rbcL*, *trnK-matK* and *trnL-F* recommended for plant species identification (Hollingsworth *et al.* 2009) were not sufficiently variable in order to discriminate between two closely related *Cistus* species. A good barcode marker should present high inter-specific and low intra-specific divergence (Lahaye *et al.* 2008). According to these criteria we found *rbcL* and *trnK-matK* as highly conserved among *C. heterophyllus, C. albidus* and *C.* × *clausonis* subsp. *carthaginensis*, and they cannot be used in population analysis due to a low substitution rate. Our data using *trnL-F* intergenic spacer confirmed, in agreement with other reports (Kress *et al.* 2005), that non-coding regions are more variable than plastid coding regions. This region has been used before in species identification (Ronning *et al.* 2005; Ward *et al.* 2005) and phylogenetic reconstructions (Chen *et al.* 2005; McDade *et al.* 2005). However our data show that *trnL-F* region is not useful for species

identification among closely related species, supporting previous findings (Taberlet *et al.* 2007).

1.4.3. Importance of sequence quality

The *trnH-psbA* intergenic spacer has been recommended as barcode marker (Chase *et al.* 2007),and has been confirmed as very informative and highly variable in taxa identification. However sequence quality is not always optimal (Kress *et al.* 2005). Indeed, sequence quality was consistently low in *trnH-psbA* so it had to be ruled out from our analysis. It is known that some DNA regions due to their genetic structure, as for example G and C repetitions, influence sequence quality and can inflate levels of heterozygosis (Lynch 2008, Mallona *et al.* 2011). Moreover numerous insertion/deletion and frequent inversions associated with palindromic motifs were referred as potential complication in further sequence alignment (Whitlock *et al.* 2010). Indeed we observed a high indel rate and AT repetitions in the *trnH-psbA* region. We conclude that although this region could be important to solve some phylogenetic disputes, it is also prone to errors that could obscure interpretation.

1.4.4 Real-time PCR melting profiles analysis as an efficient method for population studies

Melting analysis of markers amplified by PCR was already described as a very effective method for species identification (Winder *et al.* 2010). Detecting differences in allele melting temperatures for *rpoB* and *rpoC1* genes, we were able to develop a fast method for plastid haplotype determination by real-time PCR. Melting profiles provided consistent, easy to interpret data. Furthermore, it can help to avoid analysis errors in situations when plastid genome is not homoplasmic.

1.4.5 Chloroplast heteroplasmy

We found heteroplasmy in all the individuals analyzed of *C. heterphyllus* and *C.* × *clausonis* subsp. *carthaginensis*, comprising in both cases two alleles of genes *rpoB* and *rpoC1* indicating heteroplasmy in this genus. This phenomenon was already described in various plants as for example *Passiflora* (Hansen *et al.*

2007), *Medicago* (Matsushima *et al.* 2008) or *Senecio vulgaris* (Frey 1999). It can decrease utility or even declassify some types of markers for barcoding. There are three mechanisms leading to heteroplasmy: spontaneous mutations of the plant genome (Tilney-Bassett 1978), bipaternal inheritance (Metzlaff *et al.* 1981) or uniparental inheritance but with incomplete sorting-out for example during hybridization processes (Lax *et al.* 1987). In this respect, maternal inheritance had been already confirmed for the genus *Cistus* (Guzmán and Vargas 2009).

From our data, we conclude that the local *C. heterophyllus* subsp. *carthaginensis* and *C.* × *clausonis* subsp. *carthaginensis* are unique populations in Europe. The possibility that the rare *rpoC1* allele originates from hybridization with *C. albidus* or *C. monspeliensis* was excluded, indicating an ancient and unclear origin of this small endangered population.

1.5. Acknowledgements

Funding for these studies was provided by grant from Dirección General de Universidades y Política Científica de la Consejería de Universidades, Empresa e Investigación and Dirección General de Patrimonio Natural y Biodiversidad de la Consejería de Agricultura y Agua de la Comunidad Autónoma de la Región de Murcia**.** We especially thank ANSE (Asociación de Naturalistas del Sureste) for providing us with plant material, Dr. Juan José Martínez for help in fieldwork and Izaskun Mallona for help in statistical analysis and for comments on the manuscript.

Chapter 2 – Internal Transcribed Sequence (ITS) as a marker for the population structure of *Cistus heterophyllus* species

This chapter of the thesis is the result of a project at the Department of Plant Systematics and Geography, Institute of Botany, Warsaw University in cooperation with professor Krzysztof Spalik and PhD Łukasz Banasiak.

2.1 Introduction

The taxonomy of species of the genus *Cistus*, traditionally based on phenotypic markers for vegetative and reproductive characters, currently recognizes 21 species with the highest species diversity in the western Mediterranean (Guzmán & Vargas, 2009; Guzmán & Vargas, 2005). These perennial, evergreen and drought resistant shrubs bear attractive white (*Cistus monspeliensis*) or pink flowers (*Cistus albidus*) and are typical members of the Mediterranean vegetation (Batista *et al.* 2001) next to its usage as ornamental plant (Demoly & Montserrat 1995; Carlier *et al.* 2008; Ellul *et al.* 2002; Roy and Sonié 1992). Another pink flowered Ibero – African endemic species is *C. heterophyllus* Desf. (1978). The only two populations of this species in Spain are classified as *C. heterophyllus* subsp. *carthaginensis* (Crespo & Mateo 1988), one of them in Murcia, next to Llano del Beal village in Parque Regional de Calblanque, Monte de las Cenizas y Peña del Águila. While 12 of a total of 22 plants in this population resemble *C. heterophyllus* subsp. *carthaginensis*, the remaining plants resemble hybrids between *C. heterophyllus* and *C. albidus* described as *C.* × *clausonis* from northern Africa in Carvanillesia III by Font Quer and Maire (1930). Hybridization processes between individuals of different *Cistus* species are already described (Demoly, 1996). These processes are facilitated by the predominant self-incompatibility (Bosch, 1992) and equal chromosome number of all *Cistus* species (Ellul *et a*l. 2002).

Hybridization through interspecific matings with viable offspring occurs in 25% of plant species (Mallet, 2005) and may be followed by introgression or speciation (Baack & Rieseberg, 2007). Hybrid origin is often suggested by

morphological intermediacy, but additional tests are required to confirm hybridization and determine parentage. Polymorphic markers applied so far to examine hybridization events in the *Cistus* genera include allozymes (Farley &McNeilly 2000), RAPDs (Jiménez *et al.* 2007) and plastid genes (Guzmán and Vargas, 2005). A broad analysis of selected plastid genome regions for the species of *C. heterophyllus* and *C. albidus* compared to *C.* × *clausonis* subsp. *carthaginensis* on a population level revealed heteroplasmy in all the individuals analyzed of *C. heterphyllus* and *C.* × *clausonis* subsp. *carthaginensis,* but the alleles could not be traced back to *C. albidus* or *C. monspeliensis (*Pawluczyk *et al.* 2012*)*

Nuclear ribosomal DNA regions, in particular the internal transcribed spacer (ITS) region, proved very useful in detecting hybrid relationships of generic and interspecific nature in flowering plants and the reconstruction of reticulate evolutionary history. This DNA is distributed over several chromosomes, organized in hundreds to thousands of tandem repeats, each containing three ribosomal RNA genes, and intergenic (IGS) and external (ETS) and internal transcribed spacers. The latter are intercalated in the gene block 28S-5.8S-26S, while IGS separate consecutive gene blocks flanked by 5´and 3´ETS (Poczai & Hyvönen, 2010). Apart from the high variability of ITS regions due to quick evolution, which is superior over coding regions, additional advantages lie in the biparental inheritance, easy PCR amplification with universal primers available for various kind of organisms, multicopy structure and moderate size for easy sequencing. On the other hand, intra-individual ITS variability of an organism can be due to divergent paralogues, recombination events or be related to hybrid character and must be considered in phylogenetic analysis. (Buckler *et al.* 1997).

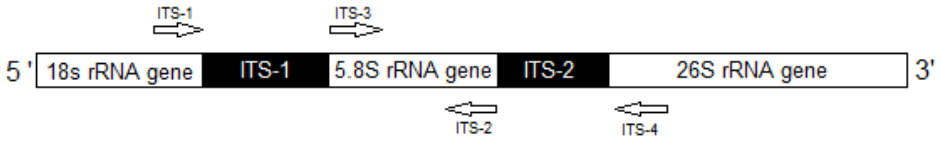


Figure. 2.1 Genes coding for ribosomal RNA and ITS primers used in this study

ITS polymorphisms shed light on hybrid evolution of individuals or species of *Ranuculus* (Hodač *et al. 2014)*, Glycine (Rausch*er et a*l.2002) and Hedera (Varg*as et a*l. 1999), among others. Nevertheless, one might encounter problems like loss of parental ribosomal DNA *loci* through recombination and segregation or the homogenization of variation between repeat types through concerted evolution. Furthermore, the ability to detect multiple ITS repeat types is highly dependent on their relative copy number in the genome of the hybrid (Rausch*er et a*l.2002; Poczai & Hyvönen, 2010).

In the present approach, we analysed ITS sequences for the geographically isolated populations of *Cistus heterophyllus, Cistus albidus* and possible hybrids of these two species *C.* × *clausonis* from Africa and Europe in order to construct a molecular tree and estimate the degree of relationship between them. We were also searching for the molecular data supporting existence of a subspecies of *Cistus heterophyllus*, which is *C. heterohyllus carthaginesis*.

2.2 Materials and Methods

2.2.1 Plant material

Plant material (young leaves) from 114 individuals was collected and analysed: 25 individuals of *C. albidus* from 10 different populations, 70 individuals of *C. heterophyllus* from 21 populations and 17 individuals of *C.* × *clausonis* from 4 different populations (Supplementary material - Table A.3). Samples were collected in Spain, Morocco and Algeria.

2.2.2 DNA extraction, amplification and sequencing

Total genomic DNA was extracted from leaves (dry or fresh) using the commercial kit 'Plant NucleoSpin' (Machery and Nagel, Düren, Germany) according to the instruction manual. ITS fragments were amplified with GoTaq Polymerase (Promega, Madison, WI, USA) under the following PCR conditions: 95 °C for 2 min., 35 cycles: 95 °C for 45 s, 56 - 57 °C for 45 s and 72 °C for 1 min. We used the primers (F: ITS-1, and R: ITS-4) described previously (White *et al.* 1990). When the amplification was not succeeded, to divide amplified fragment into two shorter, two pairs of primers F: ITS-1/R: ITS-2 and F: ITS-3/R: ITS-4 were used (Fig.2.1.). PCR products were sequenced on an Abi Prism 3130XL Genetic Analyzer (Applied Biosystem, Foster City, CA, USA).

2.2.3 Genetic variation and population analysis

To view, edit and align sequences we used MegAligner from DNASTAR Lasergene package (DNASTAR, Madison, WI) and Mesquite (Maddison & Maddison 2011). For phylogenetic analysis and trees construction we used Ape package (Paradis *et al.* 2004). Trees were edited with FigTree 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Haplotype network was constructed using Network 5 software(Bandel*t et al.* 1999) (www.fluxus-engineering.com). Network was constructed using median-joining (MJ) algorithm (weights= 10, epsilon parameter = 10). The star construction option was applied in order to delete non-MP links from the network.

2.3 Results and Discussion

2.3.1 Polymorphic sites in the ITS region

In this study, we found seven informative, polymorphic sites within the ITS region (Table 2.1). The ITS sequence was identical for all analysed populations of *C. albidus*. Most *C. heterophyllus* populations presented an ITS sequence characteristic for this species but surprisingly, within four populations of *C. heterophyllus*, one from Spain and three from Algeria (Cap Carbon 1, Cap Carbon 2 and El Afroun), individuals amplified alternative haplotypes for most ITS sites, either resembling *C. albidus* or *C. heterophyllus*. Three Algerian *C. heterophyllus* populations (from Sidi-Ferruch, Oued Nessarah and St. Claud) contained nearly exclusively ITS polymorphic sites resembling *C. albidus*. Nevertheless, as we analysed only one individual from each of these populations, we cannot exclude the possibility that individuals comprising these populations contained not only ITS haplotypes corresponding to *C. albidus* but also those from other Algerian populations.

The ITS region of the hybrid individuals *C.* × *clausonis* from Africa presented high similarity to *C. albidus* sequence. In contrast, within the Spanish *C.* × *clausonis* population from Llano de Beal and similar to *C. heterophyllus* from Llano de Beal, all polymorphic sites represented the alternative alleles typical for either *C. albidus* or *C. heterophyllus*.

Table 2.1 Polymorphic sites of the ITS region for different populations of *C. albidus*, *C. heterophyllus* and *C.* × *clausonis*. Colour codes indicate nucleotide characteristic for *C. albidus* (green), for *C. heterophyllus* (yellow) and polymorphic sites (white). IUPAC symbols are used to describe the polymorphic sites: Y = C + T, R = G + A, M = A + C.

Population location	Polimorphic sites						
	57	108	139	148	269	498	657
C. albidus							
Alcoy (Spain)	T	A	T	A	C	G	G
Llano del Beal (Spain)	T	A	T	A	C	G	G
Sierra Espuña (Spain)	T	A	T	A	C	G	G
Sierra del Buitre (Spain)	T	A	T	A	C	G	G
Roldán (Spain)	T	A	T	A	C	G	G
Aldea del Fresno (Spain)	T	A	T	A	C	G	G
Gurugu (Morocco)	T	A	T	A	C	G	G
Guro (Morocco)	T	A	T	A	C	G	G
Alhucemas (Morocco)	T	A	T	A	C	G	G
Tetuan (Morocco)	T	A	T	A	C	G	G
C. heterophyllus							
Llano del Beal (Spain)	Y	R	Y	M	Y	R	R
Béni Saf (Algeria)	C	G	C	C	T	A	A
Boutelis-Ain Turk (Algeria)	C	G	C	C	T	A	A
Cap Carbon 1 (Algeria)	Y	G	Y	Y	Y	A	R
Cap Carbon 2 (Algeria)	Y	G	Y	Y	T	A	R
El Afroun (Algeria)	C	R	Y	M	Y	R	R
Kristel (Algeria)	C	G	C	C	T	A	A
Monte Leon (Algeria)	C	G	C	C	T	A	A
Fort Santa Cruz (Algeria)	C	G	C	C	T	A	A
Saf Saf 1 (Algeria)	C	G	C	C	T	A	A
Saf Saf 2 (Algeria)	C	G	C	C	T	A	A
Oued Nessarah (Algeria)	C	G	T	A	C	G	G
Guyotville (Algeria)	C	G	C	C	T	A	A
St. Claud (Algeria)	C	G	T	A	C	G	G
Guro (Morocco)	C	G	C	C	T	A	A
Gurugu (Morocco)	C	G	C	C	T	A	A
Beni-Hadifa (Morocco)	C	G	C	C	T	A	A
Kebdana (Morocco)	C	G	C	C	T	A	R
Alhucemas 1 (Morocco)	C	G	C	C	T	A	A
Alhucemas 2 (Morocco)	C	G	C	C	T	A	A
C. x clausonis							
Llano del Beal (Spain)	Y	R	Y	M	Y	R	R
Alhucemas 1 (Morocco)	T	G	T	A	C	G	G
Alhucemas 2 (Morocco)	Y	G	T	A	C	G	G
Beni-Hadifa (Morocco)	T	R	T	A	C	G	G

2.3.2 Phylogenetic analysis

For the phylogenetic tree construction based on the ITS sequence analysis we used as outgroups *C. ladanifer*, *C. monspeliensis* and *C. laurifolius* species, belonging to the white flowered clade described before by Guzmán *et al.* (2009). *C. albidus* individuals are grouped together without regards for their provenance (Africa or Europe), which indicates high conservation of the ITS region within this species. The *C. creticus* individual is placed next to the *C. albidus* group, which supports the theory of Civeyrel *et al.* (2011) that *C. creticus* and *C. albidus* might derive from one ancestral species.

One of the *C.* × *clausonis* individuals from Llano del Beal (Spain) population branched together with *C. albidus* individuals.

C. heterophyllus presents higher variability in the ITS region. The phylogenetic analysis separates various groups within this taxon independently of their derivation: the first separate group comprises mainly African individuals and one individual from the European population, the second group branched on the tree in the direct neighbourhood with *C. albidus* species. This group is divided in two subgroups: one consists of three individuals from two different populations in Algeria and in the other subgroup, four individuals from African and European populations mix with hybrid individuals (*C.* × *clausonis*).

The intermediate position of *C.* × *clausonis* individuals on the tree confirms its hybrid character. However, the co-ocurence on the same branch of hybrid individuals and *C. heterophyllus* individuals is surprising. It could indicate that these individuals are already influenced by introgression processes or rather, as described Guzmán & Vargas (2010) based on the plastid *trnS–trnG* and *trnK–matK* sequences, because *C. heterophyllus* shared the same haplotype with closely related *C. albidus* and *C. creticus*.

Figure 2.2 Phylogenetic tree based on ITS region describing relation between *Cistus heterophyllus*, *Cistus albidus* and hybrids between these species.

2.3.3 Network analysis

The presented haplotype network separates the species *C. heterophyllus* (yellow) and *C. albidus* (blue). Whereas individuals of *C.* × *clausonis* are divided in two groups: some of them are closer related to C. *heterophyllus* and some to *C. albidus,* which supports the theory that *C.* × *clausonis* is a hybrid of these two species (Fig. 2.5). Within specific populations of *C.* × *clausonis,* we did not observe consistent relation to one of the parental species: in the Alhucemas (Morocco) population, two individuals are closely related to *C. albidus* and two other individuals to *C. heterophyllus*: in the Llano del Beal (Spain) population, five individuals are closely related to *C. albidus* and six individuals to *C. heterophyllus* (data not presented).

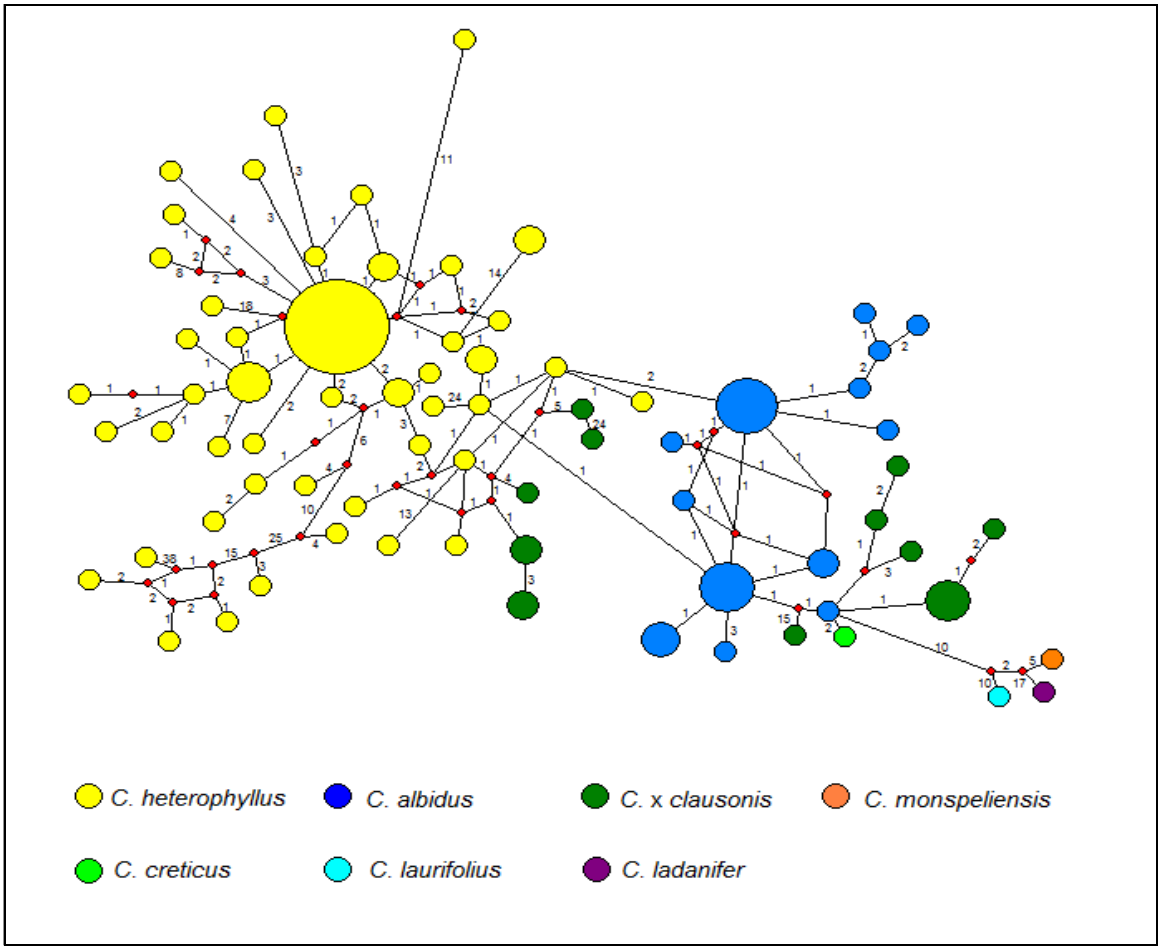


Figure 2.3 Haplotype network using ITS region sequences of selected Cistus species; the numbers in branches refer to mutational changes.

Only in case of the Beni-Hadifa (Algeria) population, both two analyzed individuals were closely related to *C. albidus*. We probably encounter here the phenomenon of homogenization of the copies of rDNA regions through recombination as proposed by Poczai and Hyvönen (2010), resulting in identical paralogous copies which segregated for either *C. albidus* or *C. heterophyllus*. Differing intraindividual ITS copies would be expected in the putative hybrid, as found for hybrids among species of Arabis (Koch, 2003).

2.4 Conclusions

As substitution occurring in non-coding spacer regions of the ribosomal DNA can be considered as neutral mutations without any constrains, ITS1 and 2 evolved much faster than coding regions. This fact make ITS regions considered as appropriate molecular makers especially for closely related taxa and population studies (Che*n et a*l. 2010; Guzmán & Vargas 2005; Guzmán & Vargas 2009). Amplified sequences of ITS regions from geographically isolated populations of *Cistus heterophyllus, Cistus albidus* and possible hybrids of these two species, *C.* × *clausonis* from Africa and Europe were used, together with sequence information collected for *C. ladanifer*, *C. monspeliensis, C. creticus* and *C. laurifolius,* to create a phylogenetic tree and haplotype network.

Our data indicate that, depending on the individual and population, *C.* × *clausonis* phylogenetically resembles more either *Cistus heterophyllus* or *Cistus albidus* and this might be based on the observation that the multiple copies of rDNA regions are homogenized through concerted evolution.

Surprisingly, in our studies *C. heterophyllus* shows high differentiation in ITS region. Some of its haplotypes were identical with *C. albidus* species. This result might lead to the conclusion that *C. albidus* species evolved from *C. heteropyllus* species or rather that the analysed *C. heterophyllus* individuals are already affected by the hybridization processes between these two species.

Chapter 3 - Quantitative evaluation of bias in PCR amplification and Next Generation Sequencing derived from metabarcoding samples

3.1 Introduction

Sequence analysis of complex DNA samples is an important approach to monitoring species distribution in biodiversity and population studies. Genetic material is assessed using universal genomic sequences "barcodes" that are informative regarding the species composition of the sample, as they contain sufficient polymorphisms between species that taxonomic discrimination becomes possible(Hajibabaei *et a*l. 2007). The barcoding approach has become a mainstream technique to identify species in insects(Hajibabaei *et a*l. 2006), very closely related plant species or hybrids(Pawluczyk *et a*l. 2012), or fungi (Kruger *et a*l. 2009) and bacteria (Links *et a*l. 2012).

In plants, seven chloroplast *loci* have been analysed as potential barcodes, the spacers *atpf-atph*, *trnH-psbA*, and *psbK-psbL*, and the genes *matK*, *rbcL*, *rpoB*, *rpoC1*, (Hollingsworth *et a*l. 2009; Kress and Erickson 2007). Metabarcoding involves DNA amplification of barcode *loci* from mixed population samples, followed by Next-Generation Sequencing (NGS). Sequenced fragments are then either assembled de novo and then aligned to known genome sequence (Links *et a*l. 2013), or are directly aligned to these genomic databases, thus becoming connected to specific taxa (Coissac *et a*l. 2012). Most often, the objective of these analyses is to arrive at a quantitative measure of the relative abundance of the various species in the sample.

Despite being a proven tool for taxonomic identification, the approach of PCR is subject to a wide variety of potential biases throughout the processes of amplification and sequence analysis, particularly when applied to mixed-population samples. These biases fall into three main categories. The first relates to differential barcode amplification success as a result of the barcode's universal primers. Depending on the marker/species combination, false-

negative results can occur when sequence variation at the universal priming sites in one of the species prevents efficient annealing of the universal barcode primer for that species. A second type of bias relates to the efficiency of the amplification reaction, which may differ from species to species based on the sequence composition of their specific variant of the barcode. As a result, the proportion of sequences representing each species in the original sample may bear little resemblance to the proportion of that species in that population. Finally, there may also be biases introduced during the preparation of DNA libraries for sequencing. For instance sample dilution has a strong effect on the correlation between biological and read quantities in bacterial samples(Ame*nd et a*l. 2010). A combination of barcoding and NGS have been in some cases confirmed by qPCR, showing that while the exact quantification is not precise, trends in the population structure are faithful (Lin*ks et a*l. 2014).

Despite knowing that these potential biases exist, the degree to which each source of bias affects the outcome of a metabarcoding experiment, and their relative importance, have not been well quantified. Moreover, by quantifying these biases and relating them to the specific sequences being studied, it may be possible to formulate approaches for post facto normalization of metabarcode data to better-reflect the population make-up. For example, PCR efficiency is an important parameter of Quantitative PCR analysis of gene expression(Platts *et a*l. 2008, Pfaffl *et a*l. 2002, Mallona *et al.* 2011), and while a variety of algorithms exist that predict the efficiency of PCR amplification, these are currently not considered in any of the normal barcoding or metabarcoding pipelines. Amplification efficiency for a given DNA sequence depends heavily on the G+C content of the amplicon(Mallona *et al.* 2011), DNA secondary structure (D'hae*ne et a*l. 2010), previous sample treatment (von Holst *et a*l. 2010). Under optimal PCR conditions with 100% amplification efficiency, two copies of DNA are generated from each template during exponential phase of amplification, and such a reaction is said to have an efficiency of 2. This efficiency can also affect another important statistic, namely Cq a relative measure of the predicted concentration of the target amplicon in a PCR reaction, and a measurement that is widely used in qPCR analysis(Bustin *et a*l. 2009; Schmittgen *et a*l. 2008). These kinds of statistics will be even more relevant to NGS technologies that

introduce additional PCR amplification steps, such as Ion Torrent or 454/Roche that utilize an emulsion PCR during library construction(Mardis 2008).

The present study, therefore, aims to first quantitatively analyze PCR success and evaluate amplification efficiency and Cq values as a tool for predicting amplification success. In this study, we undertake a survey of six well-known plant barcoding markers and apply them to 48 species from 34 different plant families. In addition, we apply the Ion Torrent sequencing method simultaneously for mixed species PCR products of three barcoding primers *rbcL*, *rpoB* and *rpoC1* starting with equal amounts of PCR products, to quantitatively measure the bias introduced by this step of the metabarcoding study.

Our results reveal that quantitative and even qualitative interpretation of metabarcoding data based on read-abundance is fraught with potential, serious biases. We present, in detail, a dissection of the degree of bias introduced at each step in the typical laboratory practice of barcode marker analysis from mixed DNA samples.

3.2 Materials and Methods

3.2.1 Plant material

Plant material 48 plant species belonging to 33 different families was gathered from the local fruit market, field sampling, botanical records and our own collections (Table 3.1).

3.2.2 DNA extraction and real-time PCR

Two independent genomic DNA samples were extracted from fresh leaf using the commercial kit 'Plant NucleoSpin' (Machery and Nagel, Düren, Germany). All extracted samples were quantified with a Nanodrop 2000 and, after isopropanol-ethanol precipitation, all samples were diluted to 50 ng/µl in order to have identical concentrations.

Table 3.1 List of plant species analysed

Plant species	Family	Location/Donor population
Spinacia oleracea	Amaranthaceae	Murcia, Spain/ commercial
Pistacia lentiscus	Anacardiaceae	Murcia, Spain/ natural
Daucus carota	Apiaceae	Murcia, Spain/ commercial
Nerium oleander	Apocynaceae	Murcia, Spain/ artificial
Arisarum vulgare	Araceae	Murcia, Spain/ natural
Phoenix dactylifera	Arecaceae	Murcia, Spain/ commercial
Aloe vera	Asphodelaceae	Murcia, Spain/ artificial
Lactuca sativa	Asteraceae	Murcia, Spain/ commercial
Cynara scolymus	Asteraceae	Murcia, Spain/ commercial
Brassica oleracea botrytis	Brassicaceae	Murcia, Spain/ commercial
Brassica oleracea italica	Brassicaceae	Murcia, Spain/ commercial
Diplotaxis erucoides	Brassicaceae	Murcia, Spain/ natural
Lobularia maritima	Brassicaceae	Murcia, Spain/ natural
Arabidopsis thaliana	Brassicaceae	Murcia, Spain/ artificial
Silene vulgaris	Caryophyllaceae	Murcia, Spain/ natural
Cistus albidus	Cistaceae	Murcia, Spain/ natural
Cistus heterophyllus	Cistaceae	Murcia, Spain/ natural
Aeonium arboreum	Crassulaceae	Murcia, Spain/ natural
Cucumis sativus	Cucurbitaceae	Biala Podlaska, Poland/ commercial
Ecballium elaterium	Cucurbitaceae	Murcia, Spain/ natural
Chamaecyparis sp.	Cupressaceae	Murcia, Spain/ artificial
Arbutus unedo	Ericaceae	Murcia, Spain/ artificial
Ricinus communis	Euphorbiaceae	Murcia, Spain/ artificial

Plant species	**Family**	**Location/Donor population**
Ceratonia siliqua	Fabaceae	Murcia, Spain/ natural
Pisum sativum	Fabaceae	Murcia, Spain/ artificial
Vicia faba	Fabaceae	Murcia, Spain/ artificial
Quercus coccifera	Fagaceae	Murcia, Spain/ natural
Pelargonium x hortorum	Geraniaceae	Murcia, Spain/ artificial
Leucobryum glaucum	Leucobryaceae	Biala Podlaska, Poland/ natural
Anagallis arvensis	Myrsinaceae	Murcia, Spain/ natural
Callistemos sp.	Myrtaceae	Murcia, Spain/ artificial
Olea europaea	Oleaceae	Murcia, Spain/ artificial
Oxalis pes-caprae	Oxalidaceae	Murcia, Spain/ natural
Pinus silvestres	Pinaceae	Biala Podlaska, Poland/ natural
Antirrhinum majus	Plantaginaceae	Murcia, Spain/ artificial
Zea mays	Poaceae	Murcia, Spain/ commercial
Oryza sativa	Poaceae	Murcia, Spain/ artificial
Hordeum vulgare	Poaceae	Murcia, Spain/ commercial
Piptatherum miliaceum	Poaceae	Murcia, Spain/ natural
Portulacaria afra	Portulacaceae	Murcia, Spain/ artificial
Galium verrucosum	Rubiaceae	Murcia, Spain/ natural
Populus alba	Salicaceae	Murcia, Spain/ artificial
Petunia hybrid	Solanaceae	Murcia, Spain/ artificial
Solanum tuberosum	Solenaceae	Murcia, Spain/ commercial
Solanum lycopersicum	Solenaceae	Murcia, Spain/ commercial
Thymelaea hirsuta	Thymelaeaceae	Murcia, Spain/ natural
Vitis vinifera	Vitaceae	Murcia, Spain/ commercial
Asphodelus fistulosus	Xanthorrhoeaceae	Murcia, Spain/ natural

Single species reactions were performed from the two independent DNA extractions with three technical replicas for a total of six PCR reactions per species using 100 ng DNA/reaction. Real-time PCR reactions were performed as described previously (Mallona *et al.* 2011). The primers used in this experiment (*rbcL-a*, *matK*, *rpoB*, *rpoC1,trnL-F, trnH-psbA*) have been described previously (Hollingsworth *et al.* 2009).

Equal amounts of genomic DNA from three species were used to create the mixed-species metabarcoding templates. Amplifications were performed using an initial DNA quantity of 150 ng corresponding to 50ng of each of the three genomes. Sequencing reactions comprised nine species.

3.2.3 qPCR efficiency and Cq calculation

qPCR efficiency and Cq were computed using qpcR, R package(Ritz *et a*l. 2008). Efficiency value (E) was calculated as EcpD2=F(cpD2)/F(cpD2)-1, in which F is raw fluorescence at cycle x, and cpD2 is cycle number at second derivative maximum of the curve(Spies*s et a*l. 2008).

3.2.4 Determination of relative abundance of sequences from PCR products of mixed genomic DNA by semiconductor sequencing

PCR products generated by amplifying, separately, the chloroplast barcoding sequences *rbcL-a*, *rpoC1* and *rpoB* from mixed genomic DNAs (100 ng each) were pooled equivalently to yield a final amount of 100ng. Initial time of digestion was adjusted to yield 300 bp fragments. Preparation of samples for library construction and sequencing were performed using the Ion Torrent Next generation sequencing Kits (Life Technologies, CA, USA) according to the manufacturer´s instructions. Briefly PCR products were fragmented using the Ion Shear Plus reagent to a fragment size of 200 bp. The corresponding fragments were ligated to adaptors and size fractionated using E-Gel electrophoresis, obtaining fragments of average 330bp. Emulsion PCR was performed using One-touch system according to the manufacturers protocol and sequencing was performed using 314 Ion Torrent chips. A total of 333,274 reads with a mean read length of 159bp were computationally analyzed in order to identify species origin of each fragment by aligning the reads with a library of

known Chloroplast sequences using Bowtie2 (Po*lz et a*l. 1998). We extracted from the resulting SAM file a map of reads to the known chloroplast sequences using a Perl script from the mPuma pipeline(Links *et al.* 2013). The analysis can be reproduced, with the same parameters and data, at the following Galaxy installation. page: http://biordf.org:8983/u/mikel-egana-aranguren/p/sources-of-bias-in-applying-barcoding-markers-for-sequence-analysis-of-environmental-samples.

3.3 Results

This work aimed to reveal and quantify the biases that can occur during metabarcoding analyses. We executed our analyses using the most widely-accepted plant barcodes, quantitated our results using widely-accepted practices such as qPCR, and followed normal protocols for library construction and NGS. At each stage, we re-normalized the samples such that we knew the precise quantities and relative abundances of the input DNA. In addition, although it is known that the size of the PCR amplification product plays a major role in bias within bacterial community pyrosequencing projects (Suzuki & Giovannoni 1996), the size of the amplicons analysed here is below the 1Kb threshold identified in those studies. Thus we should be able to safely exclude that as a possible cause of bias in this study.

3.3.1 Suitability of barcodes depending on plant species

The worst possible outcome of a metabarcode analysis is false-negative, i.e. lack of amplification of a species barcode despite presence of that taxon in the population. As such, our first analysis assessed PCR success. As expected, it varied both between barcode markers, and between the 48 plant species tested. Barcode primers for the *matK* gene were the least successful, giving positive results in only 50% of the tested species, followed by *rbcL* which amplified in 82% of species. The *rpoB* and *rpoC1* genes as well as the short intergenic spacers *trnL –F* and *trnH - psbA* proved to be the most universally successful barcoding markers, amplifying in close to 90% of the investigated species. Our data however, gives a within species assessment of PCR success based on six independent amplifications. As none of the samples had a complete failure of

amplification with all primer combinations we can conclude that DNA quality was not a limiting factor for amplification.

3.3.2 qPCR parameters for specific barcodes depending on plant species

The second phase of the analysis addressed whether end point PCR results are the outcome of PCR efficiency. As shown in Fig. 3.1, amplification efficiency during qPCR varied between barcode markers. The highest average efficiency, based on amplification from all species, corresponded to the markers *trnL–F* and *trnH - psbA* followed by *rpoB*, *rpoC1* and *rbcL*. The *matK* barcode showed the lowest average efficiency among all species. The efficiencies of *matK, rbcL* and *rpoC1*, but not *rpoB* and *trnH – psbA*, were significantly different from high-efficiency marker *trnL-F* ($p<0.0001$ for *matK* and *rbcL* and $p=0.0013$ for *rpoC1*). PCR efficiencies considering all barcode markers for selected species are summarized in Table 3.2 showing that both the barcode target and the species it is amplified from govern efficiency.

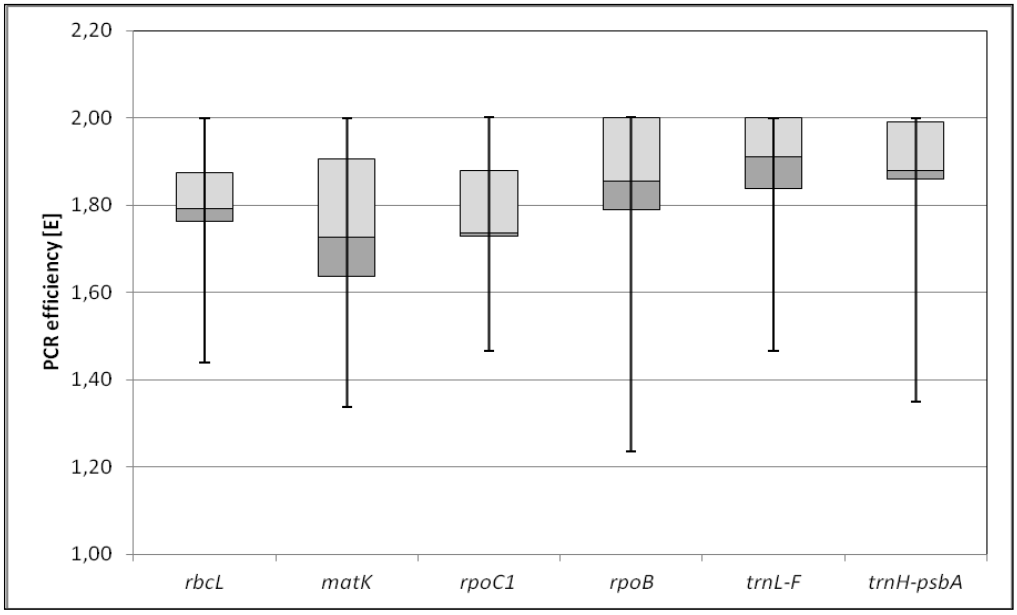


Figure 3.1 Boxplot of PCR efficiency data for six barcoding markers derived from qPCRs of 48 plant species. The graphic shows only successful amplification data with an efficiency >1

Table 3.2 PCR efficiency evaluated in a selection of plant species. Samples with NA were non-successful PCR amplifications

Plant family	***rbcL-a***	***matK***	***rpoC1***	***rpoB***	***trnL-F***	***trnH-psbA***	**Average ± SD**
Oxalidaceae (Oxalis pes-caprae)	1.89	1.83	1.70	1.78	1.91	1.90	1.84 ± 0.08
Cistaceae (Cistus heterophyllus)	1.83	1.80	1.66	1.71	1.90	1.95	1.81 ± 0.11
Poaceae (Zea mays)	1.85	NA	1.72	1.97	1.80	1.91	1.85 ± 0.10
Oleaceae (Olea europaea)	1.76	1.51	1.79	1.88	1.93	1.95	1.80 ± 0.16
Salicaceae (Populus alba)	1.78	1.78	1.78	1.89	1.98	1.98	1.87 ± 0,10
Poaceae (Oryza sativa)	NA	1.82	1.79	1.72	1.98	1.81	1.82± 0,10
Apiaceae (Daucus carota)	1.94	NA	1.85	2.00	1.98	2.00	1.95 ± 0.06
Solananceae (Solanum tuberosum)	1.70	1.70	1.85	1.84	1.95	2.00	1.80 ± 0.12
Scrophulariaceae (Antirrhinum majus)	1.79	1.82	1.98	1.99	2.00	2.00	1.93 ± 0.1
Arecaceae (Phoenix dactylifera)	1.87	1.90	1.97	1.97	2.00	1.84	1.92 ± 0.06
Cucurbitaceae (Cucumis sativus)	1.84	1.80	1.91	1.99	1.98	1.91	1.9 ±0.07
Amaranthaceae (Spinacia oleracea)	1.90	1.42	1.99	2.00	2.00	1.99	1.88 ± 0.23

Plant family	***rbcL-a***	***matK***	***rpoC1***	***rpoB***	***trnL-F***	***trnH-psbA***	**Average ± SD**
Vitales (Vitis vinifera)	1.82	1.85	1.75	1.94	1.89	1.95	1.87 ± 0.08
Solanaceae (Petunia hybrida)	1.73	1.73	1.86	1.85	1.93	1.94	1.84 ± 0.09
Fabaceae (Ceratonia silique)	1.83	1.70	1.84	1.79	1.91	1.91	1.83 ± 0.08
Fagaceae (Quercus coccifera)	NA	NA	1.68	1.72	1.90	1.86	1.79 ± 0.11
Thymelaeaceae (Thymelea hirsuta)	1.88	NA	1.73	1.78	1.81	1.75	1.79 ± 0.06
Xanthorrhoeaceae (Asphodelus fistulosus)	1.81	NA	1.73	1.76	1.78	1.84	1.78 ± 0.04
Brasicaceae (Brassica oleracea)	1.70	NA	1.76	1.82	1.76	1.67	1.74 ± 0.06
Asteraceae (Cynara Scolymus)	1.49	1.62	1.50	1.49	1.49	1.40	1.5 ± 0.07
Average	1.80	1.73	1.79	1.84	1.89	1.88	
Standard deviation	0.10	0.14	0.12	0.13	0.12	0.14	

As PCR success could be the result of initial priming and some samples gave no amplification we compared the priming site for the worst performing pair of primers (2.1.f matK and 5r matk) with their corresponding priming sites of negative performers *Zea mays*, *Quercus coccifera* and *Brassica oleracea*, *Oryza sativa* as middle quality, and *Vitis vinifera* that had the best overall amplification with this marker (Fig. 3.2). Indeed mispriming may explain the lack of amplification in the case of *Zea mays* but it is not obvious the differences in the other samples. Furthermore amplification efficiency may be affected by other parameters beyond priming (see below).

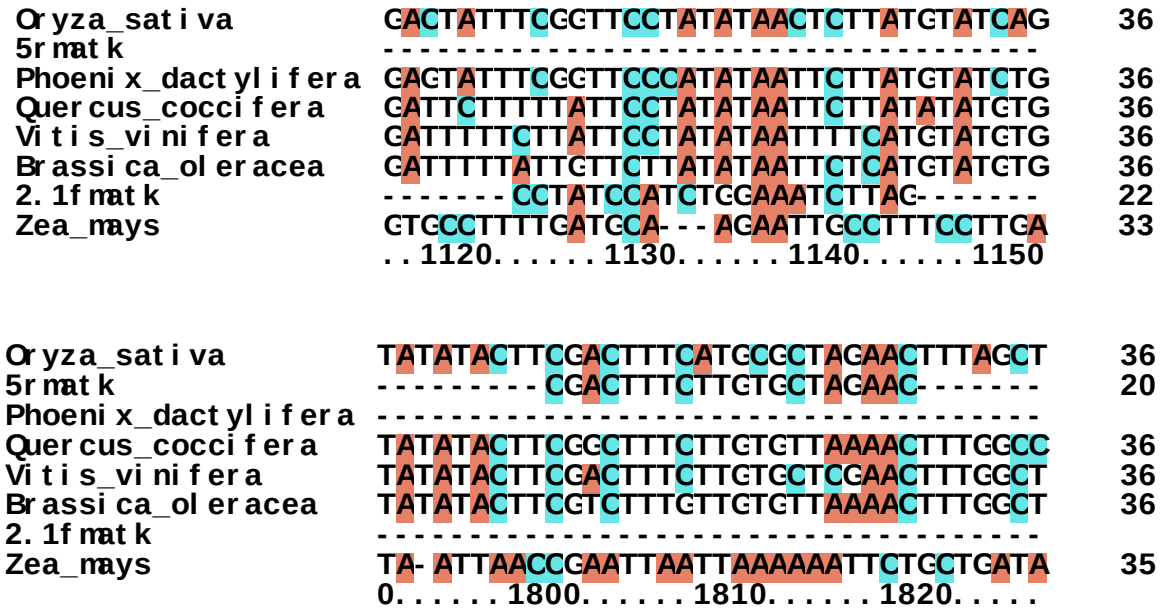


Figure 3.2 Annealing of primers 2.1f-matk and 5rmatk to sequences rendering negative amplification (*Quercus coccifera, Brassica oleracea* and *Zea mays)* and positive amplification (*Oryza sativa*, *Vitis vinifera* and *Phoenix dactylifera)*

Looking at intra-species variation for all barcodes, Cq values varied widely in this case also (Fig. 3.3, Table 3.3). Some extreme cases of intraspecific variation were found in *Oryza sativa* where *rbcL* showed no amplification whereas *trnL-F* had a Cq of 11.93 (Table 3.3).

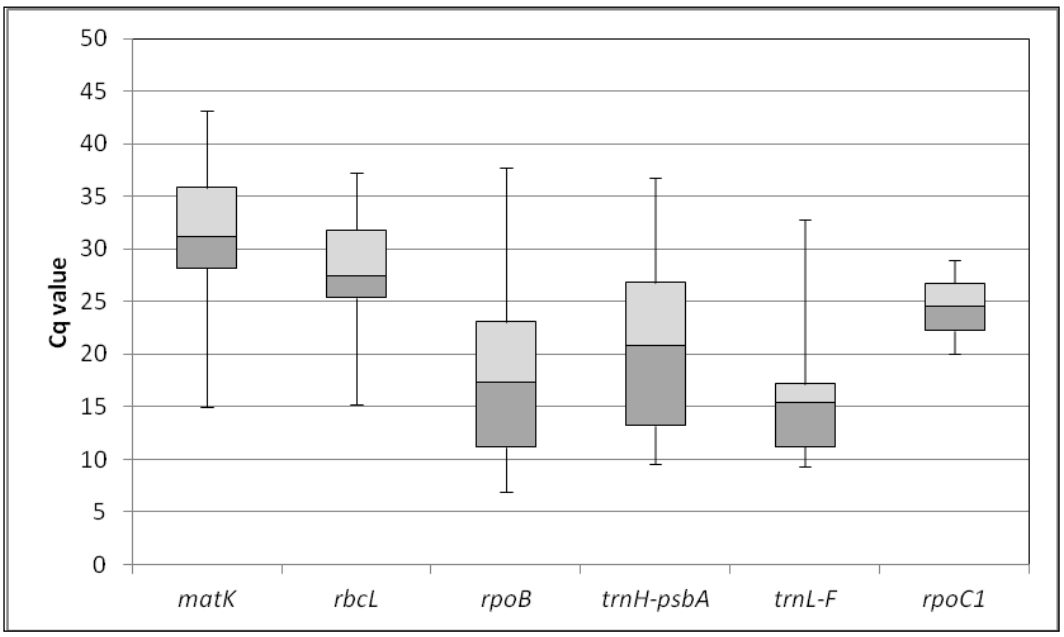


Figure 3.3 Boxplot of Cq values for six barcoding markers derived from qPCRs of 48 plant species

Beyond the false-negatives, other important differences in Cq were observed for the various markers. In *O. sativa*, the difference in Cq between *matK* (28.55) and *trnL-F* (11.93) is extremely large. If one were to apply the delta-CT formula (Schmittgen *et al.*2008), and assumed an average efficiency for both markers (efficiency = 1.9), the predicted differences in starting DNA level would be 2116-fold based on the estimates from these two barcodes. This was not an isolated case as we found negative amplification of *rbcL* or *matK* and positive albeit differing Cq values in 20% of the species tested for this parameter (*Zea mays*, *Daucus carota*, *Quercus coccifera* and *Asphodelus fistulosa*).

Cq values also varied significantly among species considering all six markers together and these differences did not correlate with the average efficiency of the PCR amplification. For example, *Z. mays* exhibited an average efficiency over all barcodes of 1.88±0.08 and an average Cq of 30.76±4.67, while *Solanum tuberosum* exhibited a similar average efficiency of 1.86±0.15, yet had a Cq of 15.98±5.30. Moreover, for any given barcode, PCR efficiency and Cq values also proved to be independent variables, based on regression analysis (R2 between 0.37 and 0.003).

Table 3.3 Cq qPRC values obtained in a selection of plant species. Samples with NA correspond to unsuccessful amplifications.

Plant family	***rbcL-a***	***matK***	***rpoC1***	***rpoB***	***trnL-F***	***trnH-psbA***	**Average ± SD**
Oxalidaceae (Oxalis pes-caprae)	30.99	36.24	22.63	23.44	19.41	27.76	26.75 ± 6.18
Cistaceae (Cistus heterophyllus)	25.83	28.80	24.85	25.01	16.74	18.86	23.35 ± 4.58
Poaceae (Zea mays)	34.74	NA	22.35	25.17	20.15	26.06	25.69 ± 5.57
Oleaceae (Olea europaea)	26.05	23.86	17.82	15.18	16.74	17.52	19.53 ± 4.36
Salicaceae (Populus alba)	24.13	29.89	15.29	13.82	13.25	13.90	18.38 ± 6.96
Poaceae (Oryza sativa)	NA	28.55	14.52	22.77	11.93	25.02	20,56 ± 7.06
Apiaceae (Daucus carota)	15.82	NA	13.06	9.77	20.15	25.95	26.95 ± 6.31
Solananceae (Solanum tuberosum)	16.77	20.55	10.16	8.65	10.53	10.90	12.93 ± 4.66
Scrophulariaceae (Antirrhinum majus)	27.81	33.83	13.06	12.72	12.06	15.08	19.09 ± 9.34
Arecaceae (Phoenix dactylifera)	31.39	16.06	10.81	15.32	10.12	19.95	17.28 ± 7.81
Cucurbitaceae (Cucumis sativus)	27.17	29.71	9.89	9.13	9.02	23.57	18.08 ± 9.77
Amaranthaceae (Spinacia oleracea)	29.66	19.59	8.94	25.32	9.40	10.40	17.22 ± 8.97
Vitales (Vitis vinifera)	33.15	18.17	17.65	13.66	13.88	15.48	18.67 ± 7.34
Solanaceae (Petunia hybrida)	28.38	19.47	11.02	10.28	10.42	11.03	15.10 ± 7.40

Plant family	***rbcL-a***	***matK***	***rpoC1***	***rpoB***	***trnL-F***	***trnH-psbA***	**Average ± SD**
Fabaceae (Ceratonia silique)	32.84	23.26	16.13	18.73	14.99	20.09	21.01 ± 6.50
Fagaceae (Quercus coccifera)	NA	NA	23.39	18.43	17.06	25.14	21.01 ± 3.87
Thymelaeaceae (Thymelea hirsuta)	29.52	NA	14.70	24.30	16.52	27.4	22.49 ± 6.58
Xanthorrhoeaceae (Asphodelus fistulosus)	26.73	NA	19.38	18.13	18.91	22.84	21.20 ± 3.58
Brasicaceae (Brassica oleracea)	24.55	NA	14.76	13.57	14.35	21.83	17.81 ± 5.02
Asteraceae (Cynara scolymus)	34.47	32.27	23.89	23.45	23.27	22.94	26.72 ± 5.21
Average	27.78	25.73	16.22	17.34	14.95	20.09	
Standard deviation	5.28	6.41	5.09	5.90	4.13	5.69	

Differences in efficiency or Cq may be related to amplification bias among template DNAs in environmental samples. We analyzed abundance of reads after sequencing in order to address this question.

3.3.3 Biases during pre-amplification and during emulsion PCR

The identification of genomic DNAs corresponding to different organisms in environmental samples requires sequencing of barcode-PCR products. Not all barcodes successfully amplify in each species. Table 3.4 shows the result of simultaneous sequencing of equal amounts of PCR products from mixed species templates amplified with barcode markers, *rbcL*, *rpoB* and *rpoC1*. The results reveal a strong bias in the number of reads corresponding each species contained in the equimolar starting sample. In the case of marker *rpoB*, most reads (95%) corresponded to *Solanum tuberosum* and only 0.02% to *Zea mays*. The number of reads was not related to the PCR efficiencies of the species, but was related to their Cq values when amplified separately (Table3.4).

Analysis of read numbers also showed a strong bias in the number of total reads corresponding to each of the barcodes (Table 3.4). Although equal amounts of PCR product from pre-amplification were used to create the amplicon library, only 11.2% of all reads were identified as *rbcL* fragments, 36.5% as *rpoB* fragments and 52.3% as *rpoC1* fragments. These results are significantly different from an expected 33.3% per reaction (Chi-square test $p < 2.2 e\text{-}16$). The relative percentages in read number proved independent of PCR efficiencies of the specific markers but correlated with average Cq values of the marker for three species amplified.

Table 3.4 Average PCR efficiencies, Cq values and sequence reads derived from PCR products of barcodes *rbcL*, *rpoB* and *rpoC1* using ion semiconductor sequencing

	Barcoding locus				
	rbcL		% of reads	PCR_{eff} of the species	Cq of the species
Average PCR_{eff} for the amplified species (together)	1.81±0.09	*Oxalis pes-caprae*	0.87	1.89±0.04	30.99±0.82
Average Cq for the amplified species (together)	26.97±7.52	*Vitis vinifera*	4.21	1.82±0.02	33.15±0.78
Total reads	34239	*Solanum tuberosum*	94.92	1.69±0.04	16.77±0.88
% of total reads	11.2				
	rpoB				
Average PCR_{eff} for the amplified species (together)	1.85±0.14	*Zea mays*	0.02	1.71±0.13	25.01±0.7
Average Cq for the amplified species (together)	21.79±5.00	*Cistus heterophyllus*	1.13	1.97±0.06	25.17±0.27
Total reads	111407	*Olea europea*	98.85	1.86±0.01	16.28±0.26
% of total reads	36.5				

	Barcoding locus *rpoC1*		% of reads	PCR_{eff} of the species	Cq of the species
Average PCR_{eff} for the amplified species (together)	1.74±0.06	*Cistus heterophyllus*	0.34	1.66±0.04	24.85±1.24
Average Cq for the amplified species (together)	18.22±4.96	*Oryza sativa*	36.57	1.79±0.02	14.52±0.54
Total reads	159923	*Populus alba*	63.09	1.78±0.03	15.29±1.51
% of total reads	52.3				

As emulsion PCR for NGS sequencing is performed with primers that correspond to ligated adaptors, and nevertheless a relationship between Cq values and final number of reads is maintained, we can conclude that the main bias that can be encountered in metabarcoding projects is related to the specific sequence of the barcode fragment. This seems to be independent of any primer-specific effect such as internal priming, etc., as it is consistent over two different primer pairs. Library construction can produce at least 4.6 fold differences when comparing *rbcL* against *rpoC1*.

3.4 Discussion

Similarity between primer and template, as well as the regional G+C content of a template, are factors that influence PCR efficiency (Benita *et.al.* 2003, Polz *et.al.* 1998). The low PCR success, particularly in case of *matK* with 50% PCR failure in a screening of 48 species, is probably due to lack of similarity between primer and template, since no highly-conserved sites flanking the most variable parts of this barcoding marker exist (Kress & Erickson 2007). Indeed, indels and mispriming may account for lack of success in PCR amplification (see Fig. 3.2). However it is not a straightforward assessment to understand the lack of amplification that may be also the result of specific features of the DNA strand amplified.

The Cq parameter is widely used in qPCR analysis (Bustin *et.al.* 2009, Schmittgen *et.al.* 2008) and we applied this to assess intraspecific and interspecific variability in both PCR success and as a possible parameter to estimate final read numbers in NGS experiments. Surprisingly, there was a wide range of Cq values identified within a single species, and even within a single DNA extraction, something completely unexpected as Cq values are thought to relate to DNA/cDNA quantities. These ranges were far beyond the 1-2 cycles that might arise from sampling and manipulation errors.

Our results show that PCR efficiency varies among barcoding markers and species, but that these differences in efficiency does not relate to the corresponding Cq values as measure of PCR success. The Cq values in contrast,

proved to be a valuable parameter for the estimation of PCR success as *matK* and *rbcL* showed the highest Cq values during qPCR. The late take-off in the qPCR assay for *rbcL* and *matK* probably reflect an excess of mismatches between primers and templates as Cq values also varied significantly among species over the whole range of markers that may be related to DNA quality and/or PCR inhibiting substances contained in the sample.

One of the most common aims in analysing environmental samples is to estimate the relative abundance of species based on determining the quantity of their template DNAs. In principle, equal amounts of template DNA from different species should lead to 1:1 amplicon numbers. However, Suzuki and Giovannoni (1996) observed preferential amplification of certain bacterial fragments in mixed templates with lower G+C content. Our results show the situation is similar in plants, with a strong bias in relative read number among three species after Ion Torrent sequencing. Low read numbers corresponded to species with high Cq values for a given marker, whereas PCR efficiency seemed unrelated, indicating that species with lower Cq's for a given marker are preferentially amplified.

As such, further improving the reliability of amplification, and utilization of sequence content features to derive and apply quantitative data-normalization algorithms, are certainly areas of significant interest for future development in metabarcoding and NGS analysis.

3.5 Acknowledgments

This work was performed as partial fulfilment of the PhD of Marta Pawluczyk. This work was funded by the Comunidad Autónoma de la Región de Murcia Project "Molecular markers in conservation and management of the flora of Murcia Region" ("Marcadores moleculares en conservación y gestión de la flora murciana").

3.6 Data availability

Raw and processed data will be made publicly available via entries in Data Dryad, and a formal Data Descriptor will be published detailing the methodologies and workflows used, as well as rich descriptions of the data elements themselves. The analytical workflow for sequence processing and mapping are already publicly available as a Galaxy workflow, as described in the manuscript, and can be freely re-run at any time. The analysis can be reproduced, with the same parameters and data, at the following Galaxy installation. page: http://biordf.org:8983/u/mikel-egana-aranguren/p/sources-of-bias-in-applying-barcoding-markers-for-sequence-analysis-of-environmental-samples.

3.7 Authors contributions

MP, MEC and JW designed experiments, MP and JW performed experiments; MP, JW, MEC, MEA and MDW analyzed data; MP, JW, MEC, MGL and MDW wrote the manuscript. All authors corrected the first draft and approved the manuscript.

General conclusions

This chapter summarizes conclusions arising from this dissertation and presents possible future investigation as a continuation of this study.

1. General conclusions

Chapter 1

- Morphological analysis of leaves and trichomes supports the theory that *C.* × *clausonis* is hybrid between *C. heterophyllus* and *C. albidus*.
- The plastid genes *rbcL, trnK-matK* and the intergenic spacer *trn L-F* are not sufficiently variable to be informative in case of such closely related species as *C. heterophyllus* and *C. albidus*.
- Heteroplasmy was found in *C. heterophyllus* and *C.* × *clausonis* individuals for *rpoB* and *rpoC1,* genes. were found to be useful for differentiation between *C. heterophyllus*, *C. albidus* and its hybrids *C.* × *clausonis*.
- *rpoB* gene discriminates *C. albidus* from *C. heterophyllus* and *C.* × *clausonis*.
- *rpoC1* gene differentiate between *C.* × *clausonis* subsp. *carthagenensis* and the rest of analysed *Cistus* individuals.

Chapter 2

- Phylogenetic analysis based on ITS region separates *C. albidus* and *C. heterophyllus* on two different branches
- Hybrid provenance of *C.* × *clausonis* was supported by intermediate position on the phylogenetic tree and the haplotype network.
- *C. heterophyllus* individuals presenting haplotype similar to hybrid individuals might suggest that: 1) these individuals are already affected

by the introgression processes or 2) *C. heterophyllus* is an ancestral taxon for *C. albidus*.

Chapter 3

- PCR success is the first indicator of the similarity between primer and template.
- PCR efficiency is the parameter that permits effective evaluation of the utility of universal markers in studies on particular organisms/ taxa.
- Bias existing in PCR amplification and NGS can interfere with correct, especially quantitative, analysis of matabarcoding samples

2. Future investigations

Since our studies present contrary results to Jiménez *et al.* (2007) concerning the application of molecular data as part of the conservation strategy for the endangered species *C. heterophyllus*, further analysis is required. Application of different types of makers as microsatellites could be advantageous.

Supplementary material

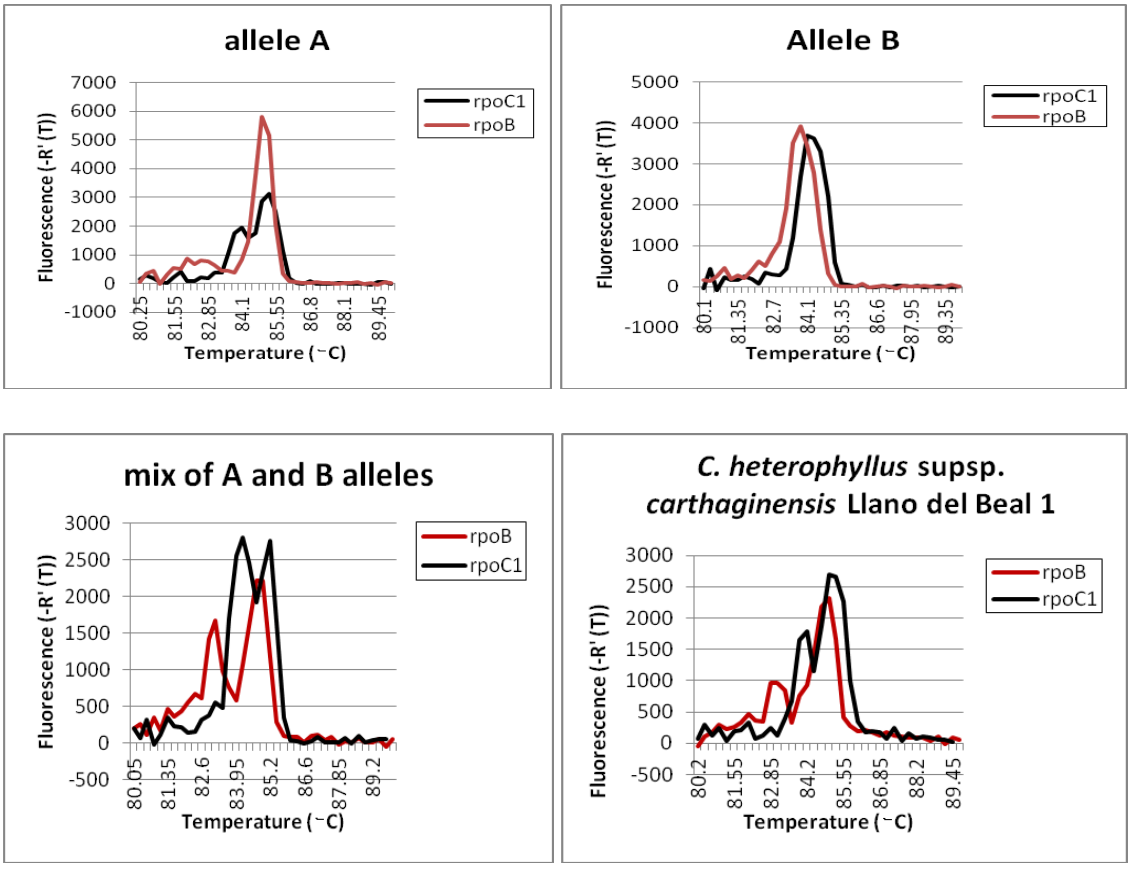
Table A.1 Sampled populations

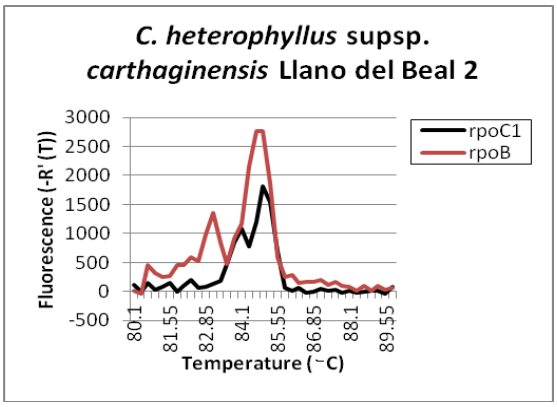
Simple number on the map (Fig.2.1)	X	Y	Geographic data	Species or plants group	Nº of individuals
1	691063	4165107	Llano del Beal	*C. albidus*	4
				C. heterophyllus subsp. *carthaginensis*	12
				C. × *clausonis* subsp. *carthaginensis*	10
				C. monspeliensis	7
2	673136	4162724	Roldán	*C. albidus*	1
3	632351	4199420	Sierra Espuña	*C. albidus*	2
4	596108	4223734	Sierra del Buitre	*C. albidus*	3
5	405637	3894315	Alhucemas	*C. heterophyllus* subsp. *hetrophyllus*	1
				C. × *clausonis*	1
6	540534	3879344	Kebdana	*C. heterophyllus* subsp *hetrophyllus*	1
7	676976	4164020	Cartagena	*C. heterophyllus* subsp. *carthaginensis*	3
				C. × *clausonis* subsp. *carthaginensis*	3
				C. albidus	3

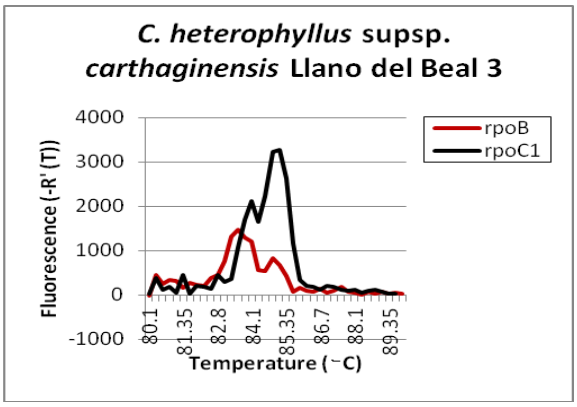
Table A.2 Primers sequences used in this study

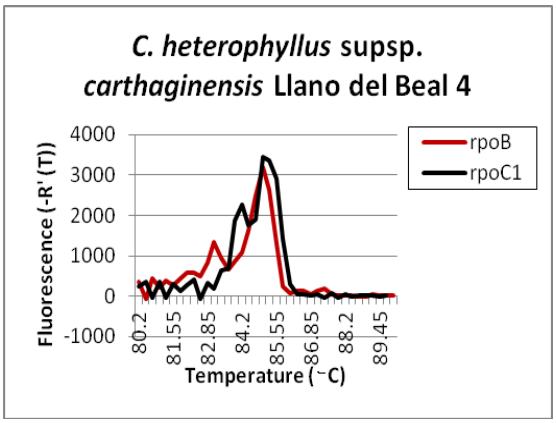
Name	Code	Primer sequences	Annealing temp. (°C)
trnK-matK (Kress & Erikson 2007)	2.1f	CCTATCCATCTGGAAATCTTAG	50
	5r	GTTCTAGCACAAGAAAGTCG	50
rpoB (Kress & Erikson 2007)	2f	ATGCAACGTCAAGCAGTTCC	55
	4r	GATCCCAGCATCACAATTCC	55
rpoC1 (Kress & Erikson 2007)	1f	GTGGATACACTTCTTGATAATGG	55
	3r	TGAGAAAACATAAGTAAACGGGC	55
trnH – psbA (Kress & Erikson 2007)	F	ACTGCCTTGATCCACTTGGC	55
	R	CGAAGCTCCATCTACAAATGG	55
trnL-F	F	TTACTATTTTTTTTGCCTACCCTCTC	55
	R	TTCAGTCCTCTGCTCTACCG	55
rbcL	F	TCCTGAATATGAAACCAAAGATACTG	50
	R	GTATCCATTGCTTCAAATTCGAA	50

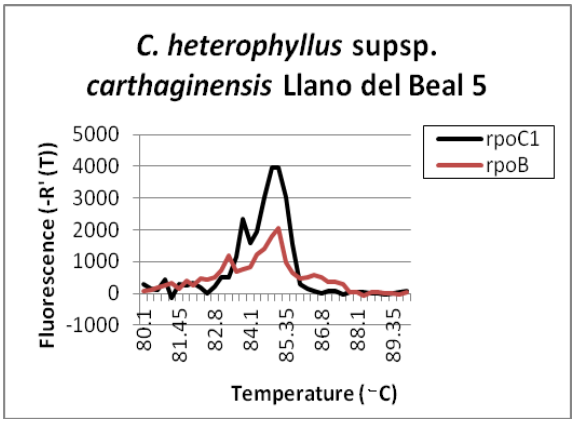
Figure S.1 Melting profiles of analyzed individuals

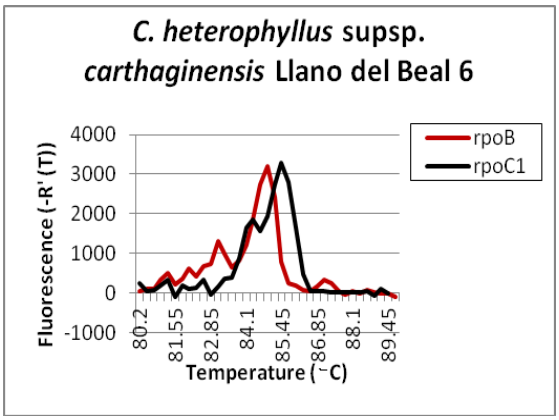


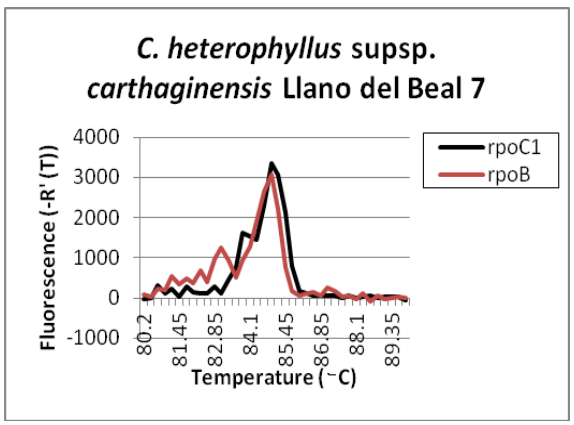

C. heterophyllus supsp. carthaginensis Llano del Beal 2
Fluorescence (-R' (T))
Temperature (°C)
rpoC1
rpoB

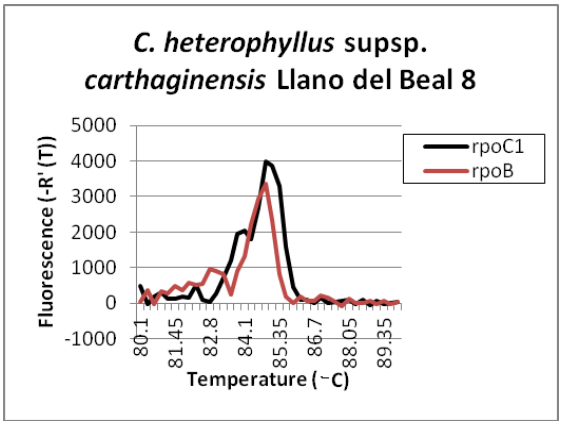

C. heterophyllus supsp. carthaginensis Llano del Beal 3
Fluorescence (-R' (T))
Temperature (°C)
rpoB
rpoC1

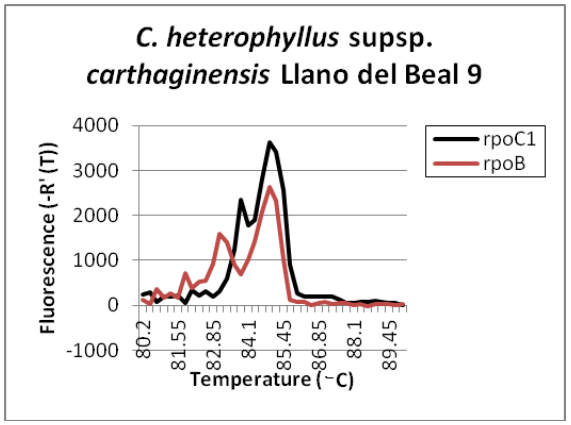

C. heterophyllus supsp. carthaginensis Llano del Beal 4
Fluorescence (-R' (T))
Temperature (°C)
rpoB
rpoC1

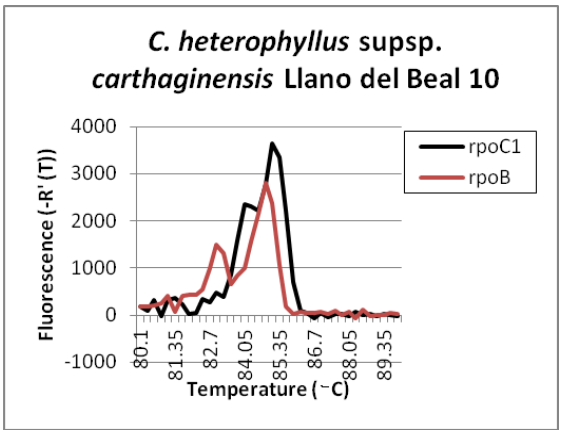

C. heterophyllus supsp. carthaginensis Llano del Beal 5
Fluorescence (-R' (T))
Temperature (°C)
rpoC1
rpoB

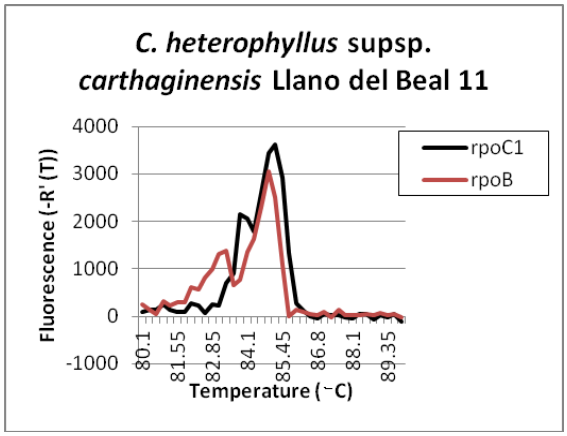

C. heterophyllus supsp. carthaginensis Llano del Beal 6
Fluorescence (-R' (T))
Temperature (°C)
rpoB
rpoC1

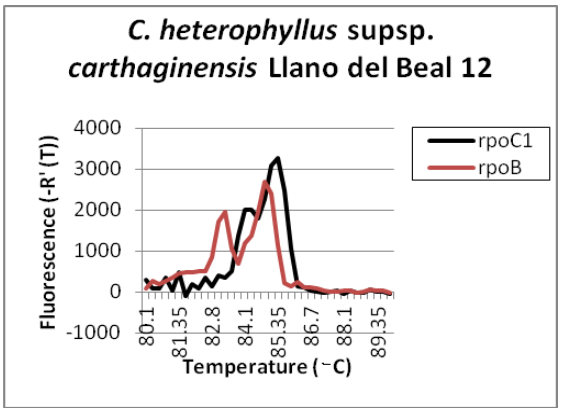

C. heterophyllus supsp. carthaginensis Llano del Beal 7
Fluorescence (-R' (T))
Temperature (°C)
rpoC1
rpoB

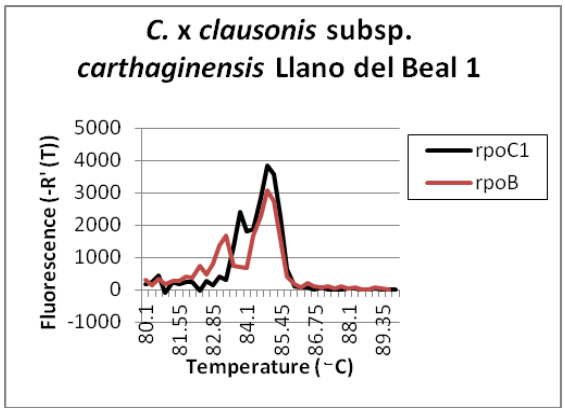

C. heterophyllus supsp. carthaginensis Llano del Beal 8
Fluorescence (-R' (T))
Temperature (°C)
rpoC1
rpoB

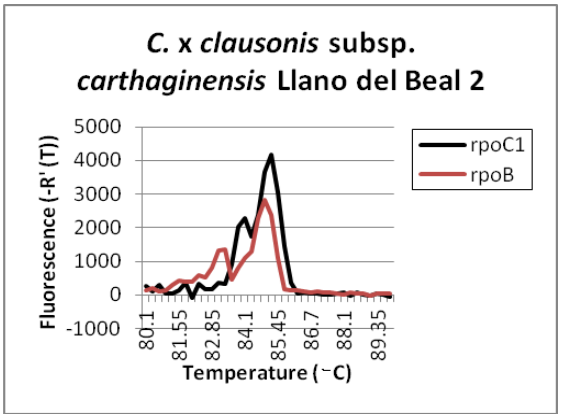

C. heterophyllus supsp. carthaginensis Llano del Beal 9
Fluorescence (-R' (T))
Temperature (°C)
rpoC1
rpoB

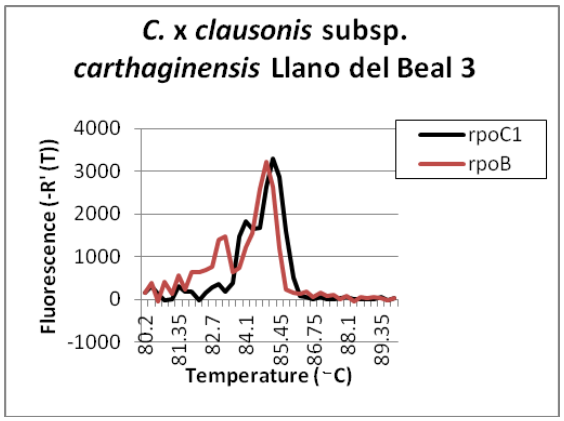

C. heterophyllus supsp. carthaginensis Llano del Beal 10
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.7
84.05
85.35
86.7
88.05
89.35
Temperature (°C)
rpoC1
rpoB

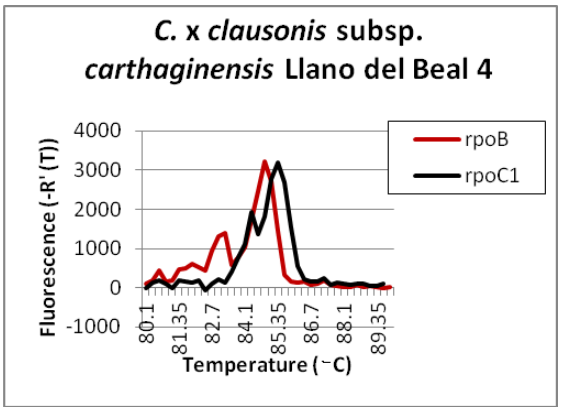

C. heterophyllus supsp. carthaginensis Llano del Beal 11
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.55
82.85
84.1
85.45
86.8
88.1
89.35
Temperature (°C)
rpoC1
rpoB

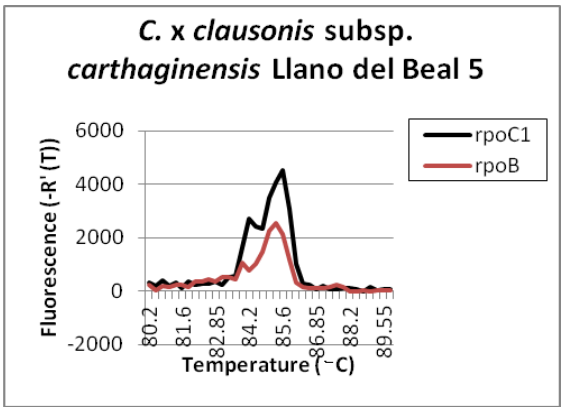

C. heterophyllus supsp. carthaginensis Llano del Beal 12
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.8
84.1
85.35
86.7
88.1
89.35
Temperature (°C)
rpoC1
rpoB

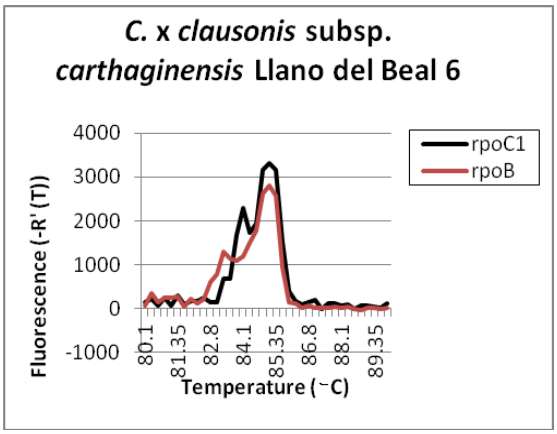

C. x clausonis subsp. carthaginensis Llano del Beal 1
Fluorescence (-R' (T))
5000
4000
3000
2000
1000
0
-1000
80.1
81.55
82.85
84.1
85.45
86.75
88.1
89.35
Temperature (°C)
rpoC1
rpoB

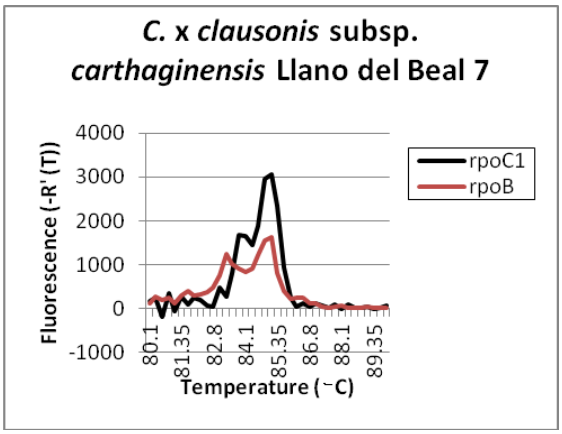

C. x clausonis subsp. carthaginensis Llano del Beal 2
Fluorescence (-R' (T))
5000
4000
3000
2000
1000
0
-1000
80.1
81.55
82.85
84.1
85.45
86.7
88.1
89.35
Temperature (°C)
rpoC1
rpoB

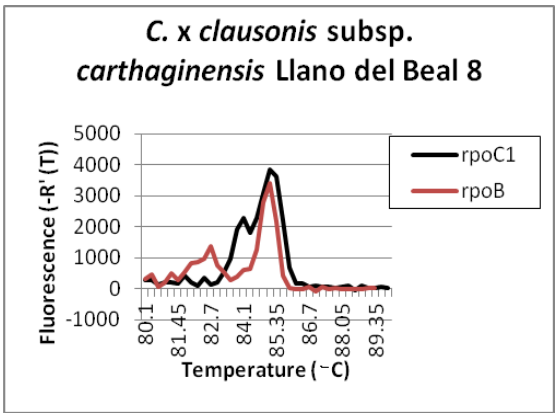

C. x clausonis subsp. carthaginensis Llano del Beal 3
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.2
81.35
82.7
84.1
85.45
86.75
88.1
89.35
Temperature (°C)
rpoC1
rpoB

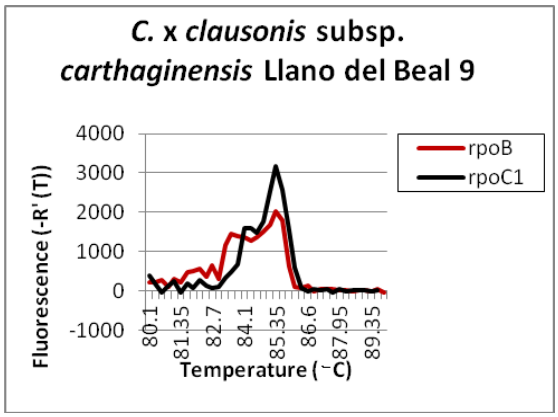

C. x clausonis subsp. carthaginensis Llano del Beal 4
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.7
84.1
85.35
86.7
88.1
89.35
Temperature (°C)
rpoB
rpoC1

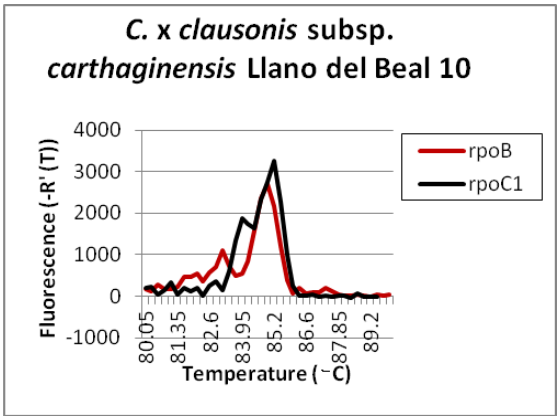

C. x clausonis subsp. carthaginensis Llano del Beal 5
Fluorescence (-R' (T))
6000
4000
2000
0
-2000
80.2
81.6
82.85
84.2
85.6
86.85
88.2
89.55
Temperature (°C)
rpoC1
rpoB

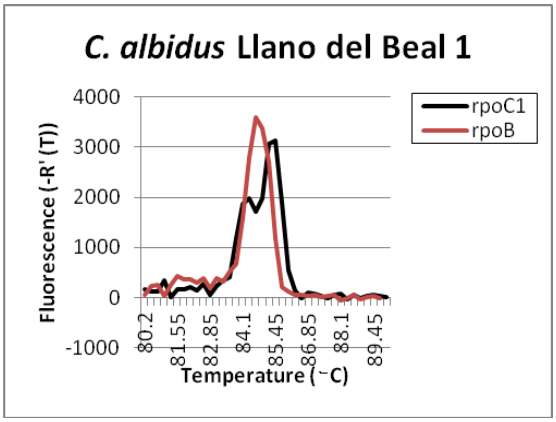

C. x clausonis subsp. carthaginensis Llano del Beal 6
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.8
84.1
85.35
86.8
88.1
89.35
Temperature (°C)
rpoC1
rpoB

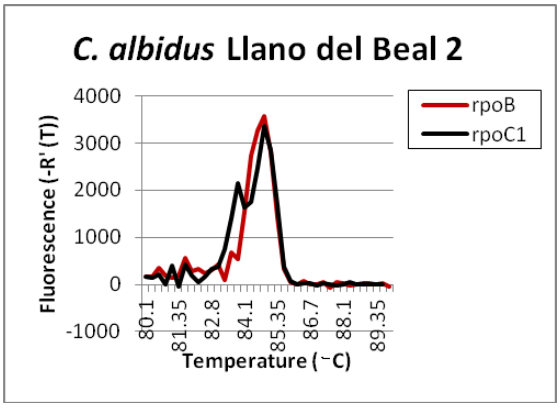

C. x clausonis subsp. carthaginensis Llano del Beal 7
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.8
84.1
85.35
86.8
88.1
89.35
Temperature (°C)
rpoC1
rpoB

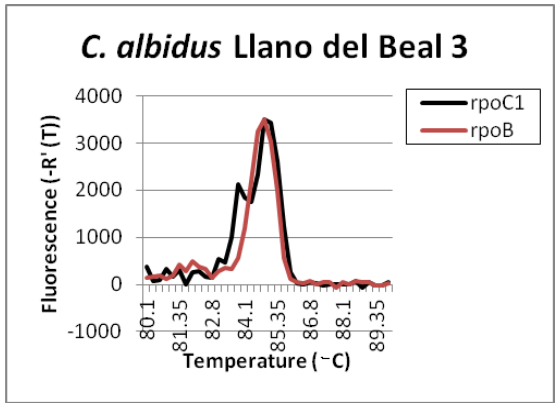

C. x clausonis subsp. carthaginensis Llano del Beal 8
Fluorescence (-R' (T))
5000
4000
3000
2000
1000
0
-1000
80.1
81.45
82.7
84.1
85.35
86.7
88.05
89.35
Temperature (°C)
rpoC1
rpoB

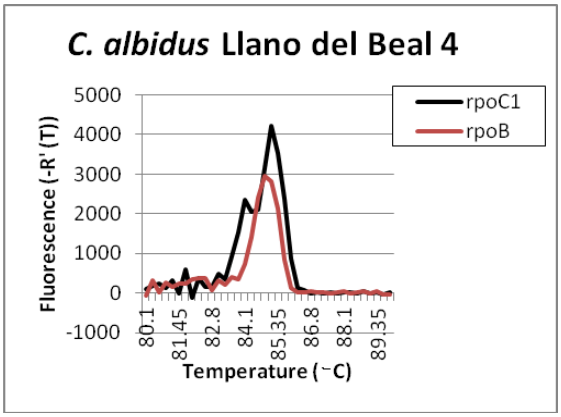

C. x clausonis subsp. carthaginensis Llano del Beal 9
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.7
84.1
85.35
86.6
87.95
89.35
Temperature (°C)
rpoB
rpoC1

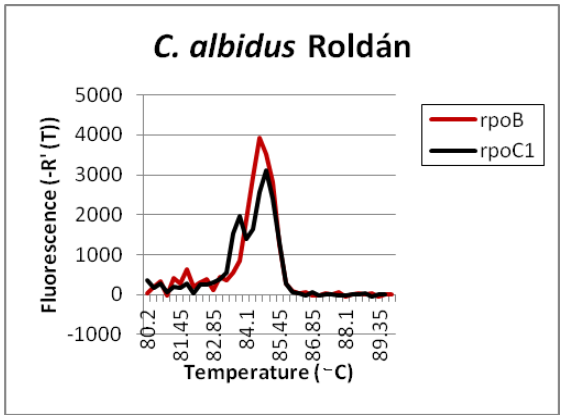

C. x clausonis subsp. carthaginensis Llano del Beal 10
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.05
81.35
82.6
83.95
85.2
86.6
87.85
89.2
Temperature (°C)
rpoB
rpoC1

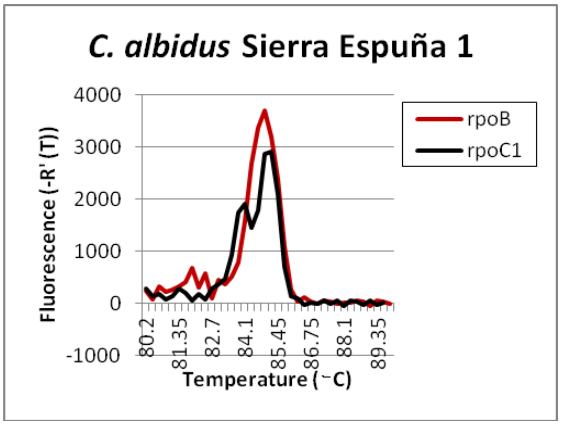

C. albidus Llano del Beal 1
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.2
81.55
82.85
84.1
85.45
86.85
88.1
89.45
Temperature (°C)
rpoC1
rpoB

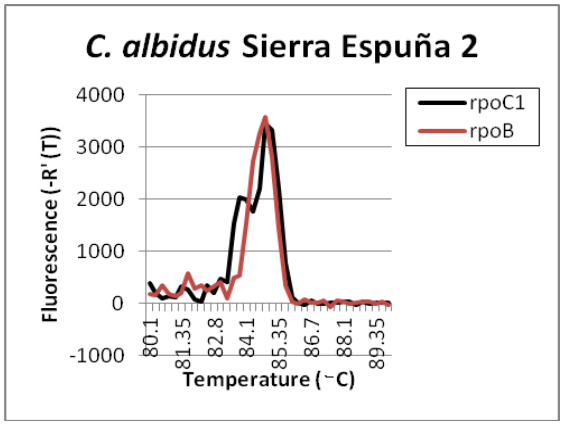

C. albidus Llano del Beal 2
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.8
84.1
85.35
86.7
88.1
89.35
Temperature (°C)
rpoB
rpoC1

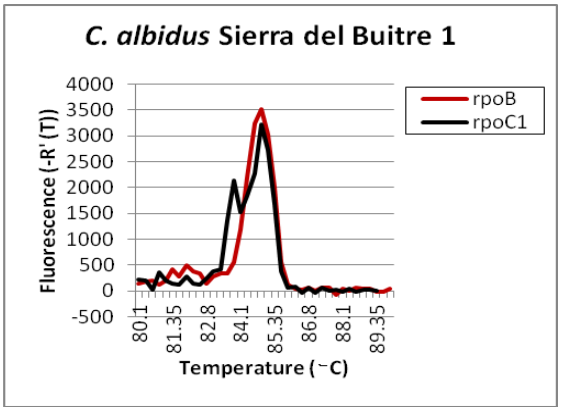

C. albidus Llano del Beal 3
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.8
84.1
85.35
86.8
88.1
89.35
Temperature (°C)
rpoC1
rpoB

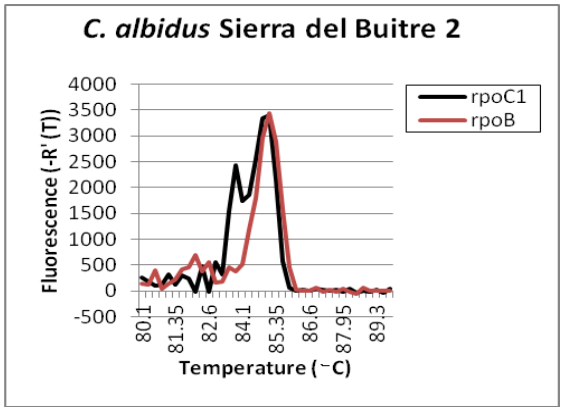

C. albidus Llano del Beal 4
Fluorescence (-R' (T))
5000
4000
3000
2000
1000
0
-1000
80.1
81.45
82.8
84.1
85.35
86.8
88.1
89.35
Temperature (°C)
rpoC1
rpoB

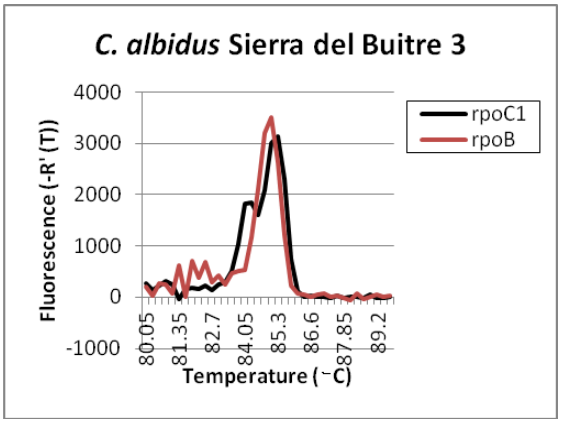

C. albidus Roldán
Fluorescence (-R' (T))
5000
4000
3000
2000
1000
0
-1000
80.2
81.45
82.85
84.1
85.45
86.85
88.1
89.35
Temperature (°C)
rpoB
rpoC1

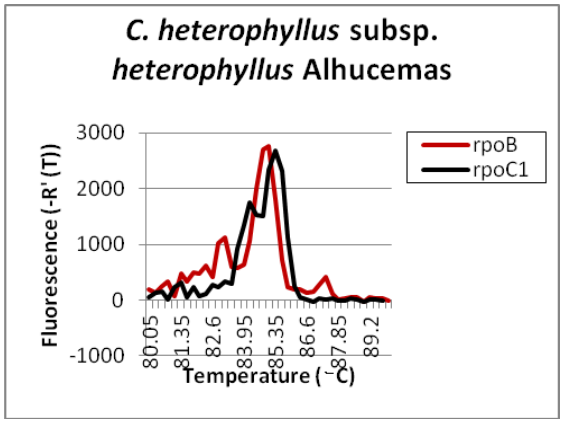

C. albidus Sierra Espuña 1
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.2
81.35
82.7
84.1
85.45
86.75
88.1
89.35
Temperature (°C)
rpoB
rpoC1

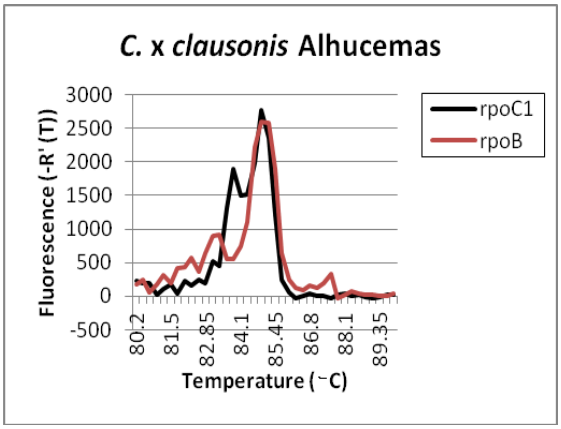

C. albidus Sierra Espuña 2
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.8
84.1
85.35
86.7
88.1
89.35
Temperature (°C)
rpoC1
rpoB

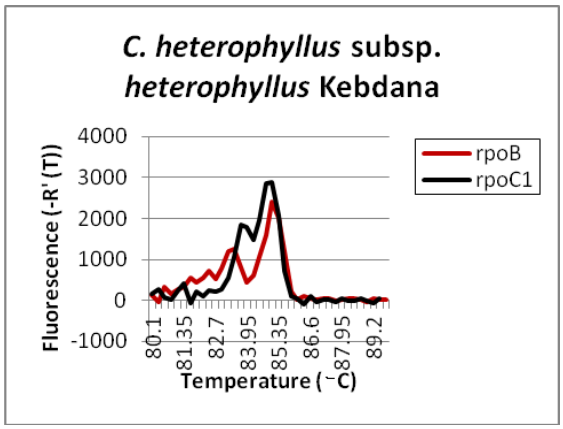

C. albidus Sierra del Buitre 1
Fluorescence (-R' (T))
4000
3500
3000
2500
2000
1500
1000
500
0
-500
80.1
81.35
82.8
84.1
85.35
86.8
88.1
89.35
Temperature (°C)
rpoB
rpoC1

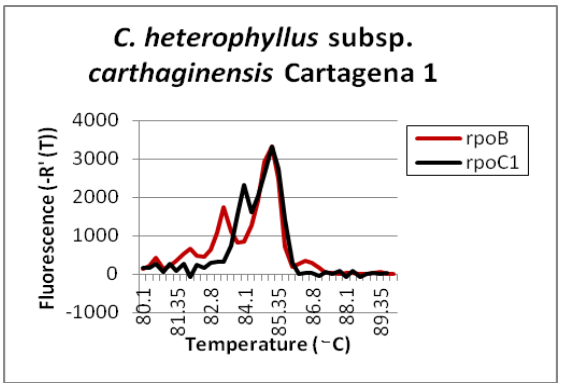

C. albidus Sierra del Buitre 2
Fluorescence (-R' (T))
4000
3500
3000
2500
2000
1500
1000
500
0
-500
80.1
81.35
82.6
84.1
85.35
86.6
87.95
89.3
Temperature (°C)
rpoC1
rpoB

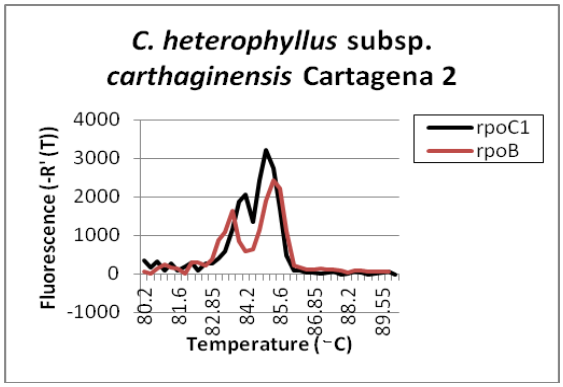

C. albidus Sierra del Buitre 3
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.05
81.35
82.7
84.05
85.3
86.6
87.85
89.2
Temperature (°C)
rpoC1
rpoB

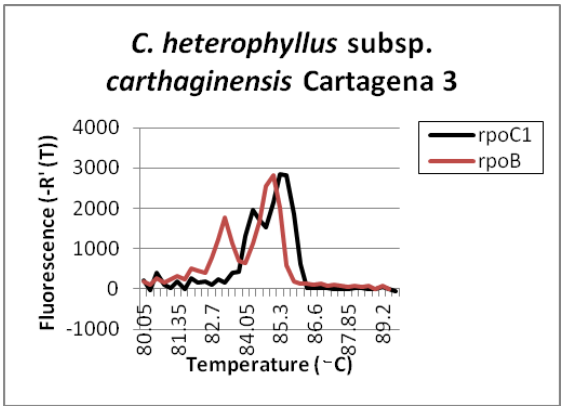

C. heterophyllus subsp. heterophyllus Alhucemas
Fluorescence (-R' (T))
3000
2000
1000
0
-1000
80.05
81.35
82.6
83.95
85.35
86.6
87.85
89.2
Temperature (°C)
rpoB
rpoC1

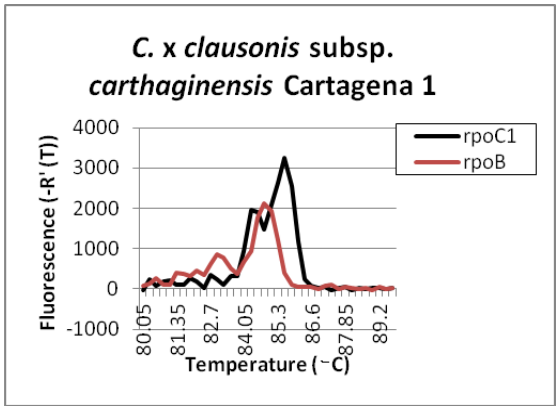

C. x clausonis Alhucemas
Fluorescence (-R' (T))
3000
2500
2000
1500
1000
500
0
-500
80.2
81.5
82.85
84.1
85.45
86.8
88.1
89.35
Temperature (°C)
rpoC1
rpoB

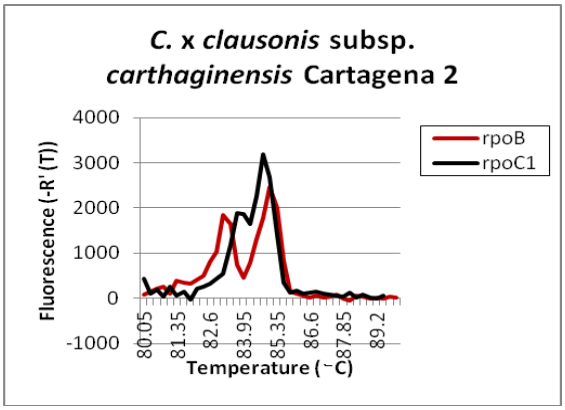

C. heterophyllus subsp. heterophyllus Kebdana
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.7
83.95
85.35
86.6
87.95
89.2
Temperature (°C)
rpoB
rpoC1

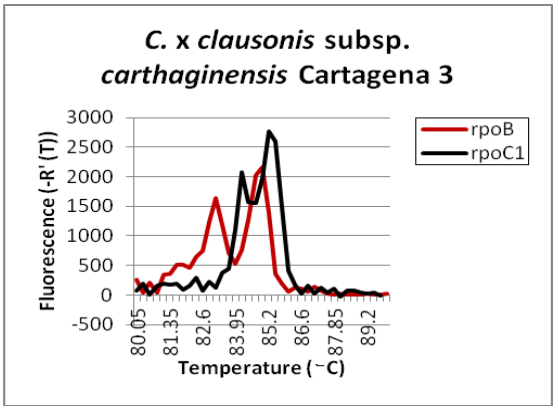

C. heterophyllus subsp. carthaginensis Cartagena 1
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.1
81.35
82.8
84.1
85.35
86.8
88.1
89.35
Temperature (°C)
rpoB
rpoC1

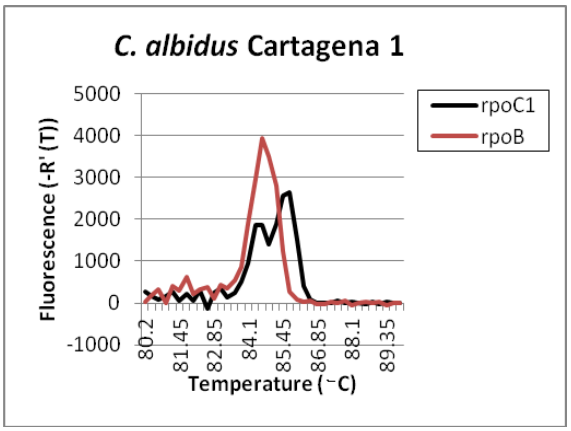

C. heterophyllus subsp. carthaginensis Cartagena 2
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.2
81.6
82.85
84.2
85.6
86.85
88.2
89.55
Temperature (°C)
rpoC1
rpoB

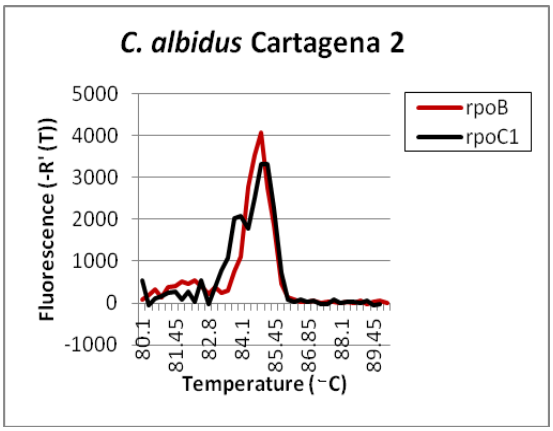

C. heterophyllus subsp. carthaginensis Cartagena 3
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.05
81.35
82.7
84.05
85.3
86.6
87.85
89.2
Temperature (°C)
rpoC1
rpoB


C. x clausonis subsp. carthaginensis Cartagena 1
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.05
81.35
82.7
84.05
85.3
86.6
87.85
89.2
Temperature (°C)
rpoC1
rpoB


C. x clausonis subsp. carthaginensis Cartagena 2
Fluorescence (-R' (T))
4000
3000
2000
1000
0
-1000
80.05
81.35
82.6
83.95
85.35
86.6
87.85
89.2
Temperature (°C)
rpoB
rpoC1


C. x clausonis subsp. carthaginensis Cartagena 3
Fluorescence (-R' (T))
3000
2500
2000
1500
1000
500
0
-500
80.05
81.35
82.6
83.95
85.2
86.6
87.85
89.2
Temperature (°C)
rpoB
rpoC1


C. albidus Cartagena 1
Fluorescence (-R' (T))
5000
4000
3000
2000
1000
0
-1000
80.2
81.45
82.85
84.1
85.45
86.85
88.1
89.35
Temperature (°C)
rpoC1
rpoB


C. albidus Cartagena 2
Fluorescence (-R' (T))
5000
4000
3000
2000
1000
0
-1000
80.1
81.45
82.8
84.1
85.45
86.85
88.1
89.45
Temperature (°C)
rpoB
rpoC1

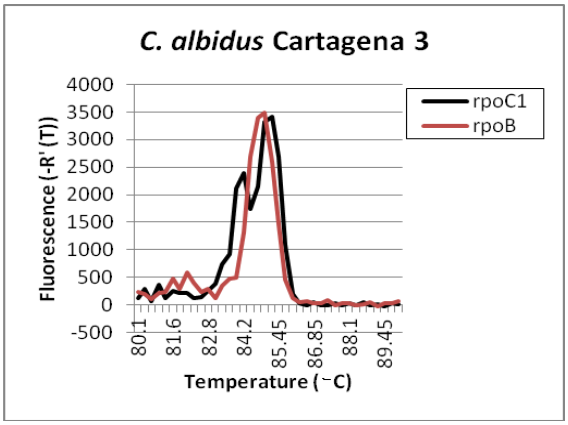

C. albidus Cartagena 3
Fluorescence (-R' (T))
4000
3500
3000
2500
2000
1500
1000
500
0
-500
80.1
81.6
82.8
84.2
85.45
86.85
88.1
89.45
Temperature (°C)
rpoC1
rpoB

Table A.3 List of individuals used in population analysis based on ITS fragment. Sample provider's abbreviation: PP Dept – Plant Production Department, UPCT; ANSE – Asociación de Naturalistas del Sureste; NCBI - NCBI GenBank; BG Geneve – Geneva Botanical Gardens

No.	Species	Sampling site	Country	Plant material	Accession no	Sample provider
1	*C. albidus*	Sierra Espuña	Spain	fresh	NA	PP Dept
2	*C. albidus*	Sierra Espuña	Spain	fresh	NA	PP Dept
3	*C. albidus*	Sierra Espuña	Spain	fresh	NA	PP Dept
4	*C. albidus*	Sierra Espuña	Spain	fresh	NA	PP Dept
5	*C. albidus*	Sierra del Buitre	Spain	fresh	NA	PP Dept
6	*C. albidus*	Sierra del Buitre	Spain	fresh	NA	PP Dept
7	*C. albidus*	Sierra del Buitre	Spain	fresh	NA	PP Dept
8	*C. albidus*	Sierra del Buitre	Spain	fresh	NA	PP Dept
9	*C. albidus*	Roldán	Spain	fresh	NA	Own collection
10	*C. albidus*	Roldán	Spain	fresh	NA	Own collection
11	*C. albidus*	Roldán	Spain	fresh	NA	Own collection
12	*C. albidus*	Gurugu	Morocco	dry	NA	ANSE
13	*C. albidus*	Gurugu	Morocco	dry	NA	ANSE
14.	*C. albidus*	Gurugu	Morocco	dry	NA	ANSE
15	*C. albidus*	Guro	Morocco	dry	NA	ANSE
16	*C. albidus*	Alhucemas	Morocco	dry	NA	ANSE
17	*C. albidus*	Alhucemas	Morocco	dry	NA	ANSE
18	*C. albidus*	Alcoy	Spain	fresh	NA	Own collection
19	*C. albidus*	Alcoy	Spain	fresh	NA	Own collection
20	*C. albidus*	Alcoy	Spain	fresh	NA	Own collection
21	*C. albidus*	Llano del Beal	Spain	fresh	NA	PP Dept
22	*C. albidus*	Llano del Beal	Spain	fresh	NA	PP Dept
23	*C. albidus*	Llano del Beal	Spain	fresh	NA	PP Dept
24	*C. albidus*	Llano del Beal	Spain	fresh	NA	PP Dept
25	*C. albidus*	Llano del Beal	Spain	fresh	NA	PP Dept
26	*C. heterophyllus*	Béni Saf	Algeria	dry	NA	ANSE
27	*C. heterophyllus*	Béni Saf	Algeria	dry	NA	ANSE
28	*C. heterophyllus*	Béni Saf	Algeria	dry	NA	ANSE
29	*C. heterophyllus*	Béni Saf	Algeria	dry	NA	ANSE
30	*C. heterophyllus*	Boutelis-Ain Turk	Algeria	dry	NA	ANSE
31	*C. heterophyllus*	Boutelis-Ain Turk	Algeria	dry	NA	ANSE
32	*C. heterophyllus*	Boutelis-Ain Turk	Algeria	dry	NA	ANSE
33	*C. heterophyllus*	Boutelis-Ain Turk	Algeria	dry	NA	ANSE
34	*C. heterophyllus*	Cap Carbon	Algeria	dry	NA	ANSE

No.	Species	Sampling site	Country	Plant material	Accesion no	Sample provider
35	*C. heterophyllus*	Cap Carbon	Algeria	dry	NA	ANSE
36	*C. heterophyllus*	Cap Carbon	Algeria	dry	NA	ANSE
37	*C. heterophyllus*	Cap Carbon	Algeria	dry	NA	ANSE
38	*C. heterophyllus*	Cap Carbon	Algeria	dry	NA	ANSE
39	*C. heterophyllus*	Cap Carbon	Algeria	dry	NA	ANSE
40	*C. heterophyllus*	Cap Carbon	Algeria	dry	NA	ANSE
41	*C. heterophyllus*	Cap Carbon	Algeria	dry	NA	ANSE
42	*C. heterophyllus*	El Afroun	Algeria	dry	NA	ANSE
43	*C. heterophyllus*	El Afroun	Algeria	dry	NA	ANSE
44	*C. heterophyllus*	El Afroun	Algeria	dry	NA	ANSE
45	*C. heterophyllus*	El Afroun	Algeria	dry	NA	ANSE
46	*C. heterophyllus*	Kristel	Algeria	dry	NA	ANSE
47	*C. heterophyllus*	Kristel	Algeria	dry	NA	ANSE
48	*C. heterophyllus*	Monte Leon	Algeria	dry	NA	ANSE
49	*C. heterophyllus*	Monte Leon	Algeria	dry	NA	ANSE
50	*C. heterophyllus*	Monte Leon	Algeria	dry	NA	ANSE
51	*C. heterophyllus*	Fort Santa Cruz	Algeria	dry	NA	ANSE
52	*C. heterophyllus*	Fort Santa Cruz	Algeria	dry	NA	ANSE
53	*C. heterophyllus*	Fort Santa Cruz	Algeria	dry	NA	ANSE
54	*C. heterophyllus*	Fort Santa Cruz	Algeria	dry	NA	ANSE
55	*C. heterophyllus*	Saf Saf	Algeria	dry	NA	ANSE
56	*C. heterophyllus*	Saf Saf	Algeria	dry	NA	ANSE
57	*C. heterophyllus*	Saf Saf	Algeria	dry	NA	ANSE
58	*C. heterophyllus*	Saf Saf	Algeria	dry	NA	ANSE
59	*C. heterophyllus*	Saf Saf	Algeria	dry	NA	ANSE
60	*C. heterophyllus*	Saf Saf	Algeria	dry	NA	ANSE
61	*C. heterophyllus*	Saf Saf	Algeria	dry	NA	ANSE
62	*C. heterophyllus*	Guro	Morocco	dry	NA	ANSE
63	*C. heterophyllus*	Guro	Morocco	dry	NA	ANSE
64	*C. heterophyllus*	Guro	Morocco	dry	NA	ANSE
65	*C. heterophyllus*	Guro	Morocco	dry	NA	ANSE
66	*C. heterophyllus*	Gurugu	Morocco	dry	NA	ANSE
67	*C. heterophyllus*	Gurugu	Morocco	dry	NA	ANSE
68	*C. heterophyllus*	Gurugu	Morocco	dry	NA	ANSE
69	*C. heterophyllus*	Beni-Hadifa	Morocco	dry	NA	ANSE
70	*C. heterophyllus*	Beni-Hadifa	Morocco	dry	NA	ANSE
71	*C. heterophyllus*	Kebdana	Morocco	dry	NA	ANSE
72	*C. heterophyllus*	Kebdana	Morocco	dry	NA	ANSE
73	*C. heterophyllus*	Kebdana	Morocco	dry	NA	ANSE
74	*C. heterophyllus*	Kebdana	Morocco	dry	NA	ANSE
75	*C. heterophyllus*	Alhucemas	Morocco	dry	NA	ANSE
76	*C. heterophyllus*	Alhucemas	Morocco	dry	NA	ANSE
77	*C. heterophyllus*	Alhucemas	Morocco	dry	NA	ANSE

No.	Species	Sampling site	Country	Plant material	Accesion no	Sample provider
78	*C. heterophyllus*	Alhucemas	Morocco	dry	NA	ANSE
79	*C. heterophyllus*	Alhucemas	Morocco	dry	NA	ANSE
80	*C. heterophyllus*	Alhucemas	Morocco	dry	NA	ANSE
81	*C. heterophyllus*	Sidi Ferruch	Algeria	dry	NA	BG Geneve
82	*C. heterophyllus*	Oued Nessarah	Algeria	dry	NA	BG Geneve
83	*C. heterophyllus*	Guyotville	Algeria	dry	NA	BG Geneve
84	*C. heterophyllus*	St. Claud	Algeria	dry	NA	BG Geneve
85	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
86	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
87	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
88	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
89	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
90	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
91	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
92	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
93	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
94	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
95	*C. heteropyllus*	Llano del Beal	Spain	fresh	NA	PP Dept
96	*C. ×clausonis*	Beni-Hadifa	Morocco	dry	NA	ANSE
97	*C. ×clausonis*	Beni-Hadifa	Morocco	dry	NA	ANSE
98	*C. ×clausonis*	Alhucemas	Morocco	dry	NA	ANSE
99	*C. ×clausonis*	Alhucemas	Morocco	dry	NA	ANSE
100	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
101	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
102	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
103	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
104	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
105	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
106	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
107	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
108	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
109	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
110	*C. ×clausonis*	Llano del Beal	Spain	fresh	NA	PP Dept
111	*C. ×clausonis*	Alhucemas	Morocco	dry	NA	ANSE
112	*C. ×clausonis*	Alhucemas	Morocco	dry	NA	ANSE
113	*C. heterophyllus*		Morocco		DQ092944	NCBI
114	*C. albidus*	Aldea del Fresno	Spain		DQ092932	NCBI
115	*C. albidus*	Tetuan	Morocco		DQ092933	NCBI
116	*C. creticus*	Kineta	Greece		DQ092937	NCBI
117	*C. ladanifer subsp. ladanifer*	Sierra de la Alhamilla	Spain		DQ092952	NCBI
118	*C. laurifolius*	Sierra de Segura	Spain		DQ092959	NCBI
119	*C. monspeliensis*	Sagres	Portugal		DQ092966	NCBI

References

Akopov SE, Orekhov AN, Tertov VV et al. (1988) Stable analogues of prostacyclin and thromboxane A2 display contradictory influences on atherosclerotic properties of cells cultured from human aorta. The effect of calcium antagonists. *Atherosclerosis*, 72:245-248. (Cited on page 31.)

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Mol.Ecol.*, 19:5555-5565. (Cited on page 50.)

Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.*, 161(2):105–121. (Cited on page 8.)

Ansorge W, Sproat B, Stegemann J, Schwager C, Zenke M (1987) Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.*, *15*(11): 4593-4602. (Cited on page 5.)

Armbruster WS, Di Stillo VS, Tuxill JD, Flores TC, Velasquez Runk JL (1999) Covariance and decoupling of floral and vegetative traits in nine neotropical plants: a reevaluation of Berg's correlation-pleiades concept. *Am. J. Bot.* 86:39–55. (Cited on page 36.)

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. *Curr. Opin. Genetics Dev.*, *17*(6):513–8. (Cited on page 39.)

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol*, 16(1):37. -48. (Cited on page 42.)

Batista F, Banares A, Caujape-Castells J *et al.* (2001) Allozyme diversity in three endemic species of Cistus (Cistaceae) from the Canary Islands: intraspecific and

interspecific comparisons and implications for genetic conservation. *Am. J. Bot.*, 88(9):1582-92. (Cited on pages 15, 19 and 39)

Beebee T, Rowe G (2008) *An Introduction to Molecular Ecology*, (2nd ed.) Oxford University Press, New York, USA (Cited on page 1.)

Beilstein MA, Al-Shehbaz IA, Kellogg EA (2006) Brassicaceae phylogeny and trichome evolution. *Am. J. Bot.*, *93*(4):607-619. (Cited on page 26.)

Benita Y, Oosting RS, Lok MC, et al. (2003) Regionalized GC content of template DNA as a predictor of PCR success. *Nucleic Acids Res.*, *31*(16):e99-e99. (Cited on page 66.)

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218):53-59. (Cited on page 6.)

Boscaiu M, Güemes J (2001) Breeding system and conservation strategy of the extremely endangered *Cistus carthaginensis* Pau (Cistaceae) of Spain. *Israel J. Plant Sci.,* 49:213-220. (Cited on page 13.)

Bosch J. (1992) Floral biology and pollinators of three co-occurring Cistus species (Cistaceae) *Bot. J. Linn. Soc.*, *109*(1):39–55. (Cited on page 39.)

Botstein D, Whiter R.L, Skolnick M, Davis R.W (1980) Construction of a genetic linkage map in man using Restriction Fragment Length Polymorphisms. *Am J. Hum Genet,* 32:314-331. (Cited on page 4.)

Bradley D, Vincent C, Carpenter R, Coen E (1996) Pathways for inflorescence and floral induction in Antirrhinum. *Development*, 122(5):1535-44. (Cited on page 36.)

Buckler ES, Ippolito A, Holtsford TP (1997) The evolution of ribosomal DNA divergent paralogues and phylogenetic implications. *Genetics*, *145*(3):821-832. (Cited on page 40.)

Bustin SA, Benes V, Garson JA *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin.Chem.*, 55:611-622. (Cited on pages 2, 50 and 66.)

Card M. (1910) Binary hybrids of the first generation in the Cistus genus and Mendelian characters. Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences 151:239-241. (Cited on page 20.)

Carlier J, Leit OJ, Fonseca F (2008) Population genetic structure of *Cistus ladanifer L.*(Cistaceae) and genetic differentiation from co-occurring *Cistus* species. *Plant Species Biol.*, 23:141-151. (Cited on pages 15, 19 and 39.)

Chase MW, Cowan RS, Hollingsworth PM *et al.* (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*, 56:295-299. (Cited on pages 20 and 37.)

Chat J, Decroocq S, Decroocq V, Petit RJ (2002) A case of chloroplast heteroplasmy in kiwifruit (Actinidia deliciosa) that is not transmitted during sexual reproduction. *J. Hered.*, 93(4):293-300. (Cited on page 20.)

Chen S, Xia T, Wang Y, Liu J, Chen S (2005) Molecular systematics and biogeography of Crawfurdia, Metagentiana and Tripterospermum (Gentianaceae) based on nuclear ribosomal and plastid DNA sequences. *Ann. Bot.*, 96(3):413-24. (Cited on page 36.)

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T., Pang X., Luo K., Li Y., Li X.,Jia X., Lin Y., Leone C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS One*, *5*(1):e8613. (Cited on page 48.)

ChungSun K, GungPyo L, DongHyeon H, KiHyun R, ChangHoo L. (2000) SCARs markers derived from RAPD for cultivar identification in *Pyrus pyrifolia*. *J. Korean Soc. Hort. Sci.,* 41(2):125-128. (Cited on page 5.)

Civeyrel L, Leclercq J, Demoly JP, Agnan Y, Quèbre N, Pélissier C, Otto T (2011) Molecular systematics, character evolution and pollen morphology of Cistus and Halimium (Cistaceae). *Plant Syst. Evol.*, 295(1): 23-54. (Cited on pages 9 and 45.)

Coissac E, Riaz T, Puillandre N (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol. Ecol.*, 21: 1834-1847. (Cited on pages 8 and 49.)

Cowan RS, Chase MW, Kress WJ, Savolainen V. (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon* 55:611–616. (Cited on page 7.)

Crespo MB, Mateo G (1988) Consideraciones acerca de la presencia de *Cistus heterophyllus* Desf. en la Península Iberíca. *Anales Jard. Bot. Madrid*, 45(1):165-171. (Cited on pages 9, 12, 19 and 39.)

D'haene B, Vandesompele J, Hellemans J (2010) Accurate and objective copy number profiling using real-time quantitative PCR. *Methods*, *50*(4):262-270. (Cited on page 50.)

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Mol. Ecol. Resour.*, *14*(2):306-323. (Cited on page 7.)

De Micco V, Aronne G (2009) Seasonal dimorphism in wood anatomy of the Mediterranean Cistus incanus L. subsp. incanus. *Trees Struct.Funct.*, 23:981-989. (Cited on page 36.)

Demoly JP (1996) Les hybrides binaires rares du genre Cistus L. (Cistaceae) = Rare binary hybrids of the genus Cistus L. (Cistaceae). *Anales Jard. Bot. Madrid.,* 54(1):6241-254 (Cited on page 39.)

Demoly JP, Montserrat P (1995) Cistus. In: Castroviejo S *et al.* (eds) Flora Iberica: plantas vasculares de la Península ibérica e Islas Baleares., 3:320–325. Real Jardín Botánico, CSIC, Madrid, Spain (Cited on pages 8, 15, 19 and 39.)

Delgado-Benarroch L, Weiss J, Egea-Cortines M (2009) The mutants compacta Shnlich, Nitida and Grandiflora define developmental compartments and a compensation mechanism in floral development in *Antirrhinum majus*. *J. Plant Res.*, 122:559-569. (Cited on page 22.)

Derycke S, Sheibani Tezerji R, Rigaux A, Moens T (2012) Investigating the ecologyand evolution of cryptic marine nematode species through quantitative real-timePCR of the ribosomal ITS region. *Mol Ecol Resour.*, 12(4):607-19. (Cited on page 2.)

Desfontaines RL (2013) Flora atlantica: Sive historia plantarum quae in Atlante, agro Tunetano et Algeriensi crescunt., 1:103-104 Cambridge University Press, Cambridge, UK. (Cited on page 8.)

Ellul P, Boscaiu M, Vicente O, Moreno V, Rossello JA (2002) Intra- and interspecific variation in DNA content in Cistus (Cistaceae). *Ann.Bot. (Lond)*, 90: 345-351. (Cited on pages 19 and 39.)

Fang DQ, Roose ML (1997) Identification of closely related citrus cultivars with inter-simple sequence repeat markers. *Theor. Appl. Genet.*, 95(3):408-417. (Cited on page 4.)

Farley RA, McNeilly T (2000) Diversity and divergence in *Cistus salvifolius* (L.) populations from contrasting habitats. *Hereditas*, 132:183-192. (Cited on pages 15, 20 and 40.)

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SC (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One, 3(7):e2802. (Cited on page 7.)

Ferrer-Gallego PP, Ferrando I. (2013) Cistus heterophyllus nothosubsp. marzoi, nsubsp. nova (Cistaceae). Bouteloua, 16: 27–33 (Cited on page 10.)

Font Quer P, Maire E (1930) *Cistus* × *clausonii*, hybrid = *C. albidus* × *C. heterophyllus*. Cavanillesia 3:59–60 (Cited on pages 13, 19 and 39.)

Frankham R(1998) Inbreeding and extinction: island populations. Conserv. Biol. 12:665-675. (Cited on page 13.)

Frey JE (1999) Genetic flexibility of plant chloroplasts. Nature, 398:115-116. (Cited on pages 20 and 38.)

Gu R, Fonseca S, Puskás LG, Hackler L, Zvara Á, Dudits D, Pais MS (2004) Transcript identification and profiling during salt stress and recovery of Populus euphratica. *Tree Physiol.*, *24*(3):265-276. (Cited on page 2.)

Güemes J, Francisco Jiménez J, Sánchez-Gómez P, Carrión Vilches MÁ (2006) *Cistus heterophyllus* ssp. *carthaginensis*. *The IUCN Red List of Threatened Species* 2006: e.T61679A12522799 (Cited on page 14.)

Gulz PG, Herrmann T, Hangst K (1996) Leaf trichomes in the genus Cistus. *Flora*, 191:85-104. (Cited on page 26.)

Guzmán B, Lledo MD, Vargas P (2009) Adaptive radiation in mediterranean *Cistus* (Cistaceae). *PLoS One.*, 4(7):e6362. (Cited on pages 15, 19, 20 and 45.)

Guzmán B, Vargas P (2005) Systematics, character evolution, and biogeography of Cistus L. (Cistaceae) based on ITS, *trnL-trnF*, and *matK* sequences. *Mol. Phylogenet. Evol.*, 37:644-660. (Cited on pages 8, 10, 15, 20, 39, 39 and 48.)

Guzmán B, Vargas P (2009) Long-distance colonization of the Western Mediterranean by *Cistus ladanifer* (Cistaceae) despite the absence of special dispersal mechanisms. *J. Biogeogr.*, 36:954-968. (Cited on pages 15, 38, 39 and 48.)

Guzmán B, Vargas P (2010) Unexpected synchronous differentiation in Mediterranean and Canarian *Cistus* (Cistaceae). *Perspect. Plant Ecol. Evol. Syst.* 12:163–174. (Cited on page 45.)

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA*, 103:968-971. (Cited on page 49.)

Hajibabaei M, Singer GA, Hebert PD, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.*, 23:167-172. (Cited on pages 20 and 49.)

Hamrick JL, Godt MJW(1990) Allozyme diversity in plant species. In: Brown AH *et al.* Plant population genetics, breeding, and genetic resources, 43-63. Sinauer Associates Inc., Sunderland, USA (Cited on page 3.)

Hansen AK, Escobar LK, Gilbert LE, Jansen RK (2007) Paternal, maternal, and biparental inheritance of the chloroplast genome in Passiflora (Passifloraceae): implications for phylogenetic studies. *Am. J. Bot.*, 94(1):42-6. (Cited on page 37.)

Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. *Proc Biol Sci., 270*(1512):313-321. (Cited on page 7.)

Hebert PD, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biol.* 2(10):e312 (Cited on page 29.)

Hernández P, Martin A, Dorado G. (1995) Development of SCARs by direct sequencing of RAPD products: a practical tool for the introgression and marker-assisted selection of wheat. *Mol Breed.,* 5(3):245–253. (Cited on page 5.)

Herrera J (1992) Flower variation and breeding systems in the Cistaceae. *Plant Syst. Evol.,* 179: 245-256. (Cited on page 13.)

Higuchi R, Dollinger G, Walsh PS, Griffith R (1992) Simultaneous amplification and detection of specific DNA sequences. *Biotechnology*, 10(4): 413-417. (Cited on page 2.)

Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*, 11(9):1026-30. (Cited on page 2.)

Hodač L, Scheben AP, Hojsgaard D, Paun O, Hörandl E (2014) ITS polymorphisms shed light on hybrid evolution in apomictic plants: a case study on the *Ranunculus auricomus* complex. *PloS One*, 9(7):e103003. (Cited on page 41.)

Hollingsworth PM, Forrest LL, Spouge JL et al. (2009) A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA*, 106: 2794. (Cited on pages 7, 36, 49 and 54.)

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. *PLoS One*, *6*(5):e19254. (Cited on page 7.)

Hoot SB (1991) Phylogeny of the Ranunculaceae based on epidermal microcharacters and macromorphology. *Syst. Bot.*, 16:741-755. (Cited on page 26.)

Ivanov R, Fobis-Loisy I, Gaude T (2010) When no means no: guide to Brassicaceae self-incompatibility. *Trends Plant Sci.*, *15*(7):387-394. (Cited on page 13.)

Jiménez JF, Sánchez-Gómez P, Rosselló A (2007) Evidencia de introgresión en *Cistus heterophyllus* subsp. *carthaginensis* (Cistaceae) a partir de marcadores moleculares RAPD. *Anal. Biol.*, 29:95-103. (Cited on pages 9, 15, 20, 40 and 70.)

Khalik KA (2005) Morphological studies on trichomes of Brassicaceae in Egypt and taxonomic significance. *Acta Bot. Croat.*, 64:57-73. (Cited on page 26.)

Koch M, Mummenhoff K, Al-Shehbaz IA (2003) Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae): a review of a decade of studies. *Ann. Missouri Bot. Garden,* 90:151–171. (Cited on page 48.)

Konieczny A, Ausubel FM (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* 4(2):403-410. (Cited on page 5.)

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One*, 2(6):e508. (Cited on pages 15, 20, 23, 49, 66 and 72.)

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA*, 102:8369-8374. (Cited on pages 7, 15, 20, 22, 36 and 37.)

Kruger M, Stockinger H, Kruger C, Schussler A (2009) DNA-based species level detection of *Glomeromycota*: one PCR primer set for all arbuscular mycorrhizal fungi. *New Phytol.*, 183:212-223. (Cited on page 49.)

Lahaye R, van der Bank M, Bogarin D *et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. USA*, 105:2923-2928. (Cited on page 36.)

Lax AR, Vaughn KC, Duke SO, Endrizzi JE (1987) Structural and physiological studies of a plastome cotton mutant with slow sorting out. *J. Hered.*, 78:147. (Cited on page 38.)

Lee SB, Kaittanis C, Jansen RK *et al.* (2006) The complete chloroplast genome sequence of Gossypium hirsutum: organization and phylogenetic relationships to other angiosperms. *BMC Genomics*, 7:61. (Cited on page 30.)

Lee SW, Ledig FT, Johnson DR (2002) Genetic variation at allozyme and RAPD markers in *Pinus longaeva* (Pinaceae) of the White Mountains, California. *Am. J. Bot.*, 89:566. (Cited on page 20.)

Links MG, Demeke T, Grafenhan T *et al.* (2014) Simultaneous profiling of seed-associated bacteria and fungi reveals antagonistic interactions between microorganisms within a shared epiphytic microbiome on *Triticum* and *Brassica* seeds. *New Phytol.*, 202:542-553. (Cited on page 50.)

Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE (2012) The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. *PLoS One*, 7(11):e49755. (Cited on page 49.)

Links MG, Chaban B, Hemmingsen SM, Muirhead K, Hill JE (2013) mPUMA: a computational approach to microbiota analysis by de novo assembly of operational taxonomic units based on protein-coding barcode sequences. *Microbiome*, 1:23. (Cited on pages 49 and 55.)

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.*, 44: 398-401. (Cited on page 4.)

Luu-The V, Paquet N, Calvo E, Cumps J (2005) Improved real-time RT-PCR method forhigh-throughput measurements using second derivative calculation and double correction. *Biotechniques*, 38(2):287-93. (Cited on page 2.)

Lynch M (2008) Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Mol. Biol. Evol.*, 25(11):2409-19. (Cited o page 37.)

Mallona I, Weiss J, Egea-Cortines M (2011) pcrEfficiency: a Web tool for PCR amplification efficiency prediction. *BMC.Bioinformatics.*, 12:404. (Cited on pages XXVII, 2, 37, 38, 50 and 54)

Manter DK, Vivanco JM (2007) Use of the ITS primers, ITS1F and ITS4, to characterize fungal abundance and diversity in mixed-template samples by qPCR and length heterogeneity analysis. *J. Microbiol. Methods*, 71(1):7-14. (Cited on page 2.)

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet.*, 24:133-141. (Cited on page 51.)

Martin WJ, Warmington JR, Galinski BR, Gallagher M, Davies RW, Beck MS, Oliver SG (1985) Automation of DNA sequencing: a system to perform the sanger dideoxysequencing reactions. *Nat. Biotechnol.*, *3*(10): 911-915. (Cited on page 5.)

Matsushima R, Hu Y, Toyoda K (2008) The model plant *Medicago truncatula* exhibits biparental plastid inheritance. *Plant Cell Physiol.*, 49(1):81-91. (Cited on page 38.)

McDade LA, Daniel TF, Kiel CA, Vollesen K, Lavin M (2005) Phylogenetic relationships among Acantheae (Acanthaceae): major lineages present

contrasting patterns of molecular evolution and morphological differentiation. *Syst. Bot.*, 30:834-862. (Cited o page 36.)

Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Syst.Biol.*, 57:809-813. (Cited on page 23.)

Mallet J. (2005) Hybridization as an invasion of the genome. *Trends Ecol. Evol.*, 20(5):229–37. (Cited on page 39.)

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, BravermanMS, ChenYJ, ChenZ, DewellS, DuL, Fierro JM, GomesXV, GodwinBC, HeW, HelgesenS, HoCH, IrzykGP, JandoSC, AlenquerMLI, JarvieTP, JirageKB, KimJB, KnightJR, LanzaJR, LeamonJH, LefkowitzSM, LeiM, LiJ, LohmanKL, LuH, MakhijaniVB, McDadeKE, McKennaMP, MyersEW, NickersonE, NobileJR, PlantR, PucBP, RonanMT, RothGT, SarkisGJ, Simons JF, SimpsonJW, SrinivasanM, TartaroKR, TomaszA, VogtKA, VolkmerGA, WangSH, WangY, WeinerMP, YuP, BegleyRF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057): 376-380. (Cited on page 6.)

Maddison WP, Maddison DR (2015). Mesquite: a modular system for evolutionary analysis. Version 2.75. 2011. *URL http://mesquiteproject. org*. (Cited on page 42.)

Metzlaff M, Börner T, Hagemann R (1981) Variations of chloroplast DNAs in the genus Pelargonium and their biparental inheritance. *Theor. Appl. Genet.,* 60:37-41. (Cited on page 38.)

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.*, 3(12):e422. (Cited on page 23.)

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. USA*, 107:4623-4628. (Cited on page 30.)

Moreno S, Martín JP, Ortiz JM (1998) Inter-simple sequence repeats PCR for characterization of closely related grapevine germplasm. *Euphytica*, 101(1), 117-125. (Cited on page 4.)

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via apolymerase-catalyzed chain reaction. *Methods Enzymol.*,155, 335-50. (Cited on page 1.)

Navarro Cano JA, Carrión Vilches MÁ, Robles Sánchez J, López Espinosa JA (2008). Plan de recuparación de *Cistus heterophyllus* subsp. *carthaginensis* (Jara de Cartagena) en la Región de Murcia. 1:4-13, 37, 46-67, BIOCYMA, Consejería de Agricultura y agua de la Región de Murcia, Murcia, Spain (Cited on pages 9, 11-14, 17 and 20.)

Navarro Cano JA, Rivera D (2001) Hacia la recuperación de la jara cartagenera en Murcia. *Quercus* 189:26-29. (Cited on page 14.)

Olson M, Hood L, Cantor C, Botstein D (1989) A common language for physical mapping of the human genome. *Science*, 245(4925): 1434-1435. (Cited on page 4.)

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289-290. (Cited on page 42.)

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA (1998) What molecules can tell us about populations: choosing andusing a molecular marker. *Ecology,* 79(2):361-382. (Cited on page 3.)

Pawluczyk M, Weiss J, Vicente-Colomer M, Egea-Cortines M (2012) Two alleles of *rpoB* and *rpoC1* distinguish an endemic European population from Cistus heterophyllus and its putative hybrid (*C.* × *clausonis*) with *C. albidus*. *Plant Sys. Evol.*, 298:409-419. (Cited on page 40 and 49.)

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.*, 30(9): e36. (Cited on page 50.)

Platts AE, Johnson GD, Linnemann AK, Krawetz SA (2008) Real-time PCR quantification using a variable reaction efficiency model. *Anal.Biochem.*, 380:315-322. (Cited on page 50.)

Poczai P, Hyvönen J (2010) Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Mol. Biol. Reports*, *37*(4):1897-1912. (Cited on pages 40, 41 and 48.)

Polz MF, Cavanaugh CM (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.*, *64*(10):3724-3730. (Cited on pages 55 and 66.)

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A**,** Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*(1):341. (Cited on page 6.)

Quéméré E, Hibert F, Miquel C, Lhuillier E, Rasolondraibe E, Champeau J, Champeau J, Rabarivola C. Nusbaumer L, Chatelain C, Gautier L, Ranirison P, Crouau-Roy B, Taberlet P, Chikhi L (2013). A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. *Plos One,* 8(3): e58971. (Cited on page 7.)

Ratnasingham S & Hebert PD (2007) BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Mol. Ecol. Notes*, *7*(3):355-364. (Cited on page 7.)

Rauscher JT, Doyle JJ, Brown AHD (2002) Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. *Molecular Ecology*, *11*(12):2691–2702, (Cited on page 41.)

Richterich P (1998) Estimation of errors in ôrawö DNA sequences: A validation study. *Genome Research*, *8*(3):251-259. (Cited on page 23.)

Ritz C, Spiess AN (2008) qpcR: a R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics.*, 24:1549-1551. (Cited on page 54.)

Robledo A, Navarro JA, Rivera D, Alcaraz F (1995) Los últimos ejemplares de jara de cartagena. *Quercus* 110:12-14 (Cited on page 13.)

Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.*, *242*(1): 84-89. (Cited on page 6.)

Ronning SB, Rudi K, Berdal KG, Holst-Jensen A (2005) Differentiation of important and closely related cereal plant species (Poaceae) in food by hybridization to an oligonucleotide array. *J. Agric. Food Chem.,* 53:8874-8880. (Cited on page 36.)

Rosato M, Ferrer-Gallego P, Totta C, Laguna E, Rosselló JA (2016). Nuclear rDNA instability in in vitro-generated plants is amplified after sexual reproduction with conspecific wild individuals. *Biol. J. Linn. Soc.*, 181(1):127–137, (Cited on page 14.)

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, *475*(7356):348-352. (Cited on page 7.)

Roy J, Sonié L (1992) Germination and population dynamics of Cistus species in relation to fire. *J. Appl. Ecol.*, 29:647-655. (Cited on pages 19 and 39.)

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.*, *74*(12):5463-5467. (Cited on page 5.)

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of Arabidopsis thaliana. *DNA Res.*, *6*(5):283-290. (Cited on page 30.)

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat.Protoc.*, 3:1101-1108. (Cited on pages 50, 60 and 66.)

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. USA*, 109(16): 6241-6246. (Cited on page 7.)

Simonet M, Ansereau P. (1939) The meoisis of two Cistus hybrids: C x hybridus Pouri and C x Rodier Verg var antipolintensis Dans. *Comptes Rendus*

Hebdomadaires Des Seances De L Academie Des Sciences 208:1526-1527. (Cited on page 20.)

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SBH, Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* 321: 674 – 679 (Cited on page 5.)

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. USA*, 101(42):15261–15264. (Cited on page 13.)

Spiess AN, Feig C, Ritz C (2008) Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry. *BMC Bioinformatics.*, 9:221. (Cited on page 54.)

Suzuki MT, Giovannoni SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.*, *62*(2):625-630. (Cited on pages 55 and 67.)

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts. (Cited on page 23.)

Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.*, 35:e14. (Cited on page 37.)

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.*, 21(8):2045-2050. (Cited on page 8.)

Talavera S, Gibbs PE, Herrera J (1993) Reproductive biology of *Cistus ladanifer* (Cistaceae). *Plant Syst. Evol.,* 186:123-134. (Cited on page 13.)

Tilney-Bassett RAE (1978) The inheritance and genetic behaviour of plastids. *The Plastids.Their Chemistry, Structure, Growth and Inheritance. Elsevier:* New York, USA, 251-524. (Cite on page 38.)

Vargas P, McAllister HA, Morton C, Jury SL, Wilkinson MJ (1999) Polyploid speciation in Hedera (Araliaceae): Phylogenetic and biogeographic insights based on chromosome counts and ITS sequences. *Plant Syst. Evol.*, *219*(3-4), 165–179. (Cited on page 41.)

von Holst C, Boix A, Marien A, Prado M (2010) Factors influencing the accuracy of measurements with real-time PCR: The example of the determination of processed animal proteins. *Food Control*, 24:142-147. (Cited on page 50.)

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.*, 23 (21):4407-4414. (Cited on page 4.)

Warburg EF (1968) *Cistus heterophyllus*, In Tutin TG, Heywood VH, Burges NA, Valentine DH, Walters SM, Webb DA, *Flora Europaea*: Rosaceae to Umbelliferae, 2:283, Cambridge University Press, Cambridge, UK (Cited on page 11.)

Ward J, Peakall R, Gilmore SR, Robertson J (2005) A molecular identification system for grasses: a novel technology for forensic botany. *Forensic Sci. Int.*, 152:121-131. (Cited on page 36.)

Waugh J. (2007) DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*, *29*(2):188-197. (Cited on pages 7.)

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can betyped using the polymerase chain reaction. *Am. J. Hum. Genet.*, 44:388–396. (Cited on page 4.)

Weiss J, Delgado-Benarroch L, Egea-Cortines M (2005) Genetic control of floral size and proportions. *Int. J. Dev. Biol.* 49:513-525. (Cited on page 36.)

White TJ, Bruns T, Lee SJWT, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications, 18:315-322. (Cited on page 42.)

Whitlock BA, Hale AM, Groff PA (2010) Intraspecific Inversions Pose a Challenge for the *trnH-psbA* Plant DNA Barcode. *PLoS One*, 5(7):e11533. (Cited on page 37.)

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.,* 18 (22),6531-6535. (Cited on page 4.)

Winder L, Phillips C, Richards N *et al.* (2010) Evaluation of DNA melting analysis as a tool for species identification. *Methods Ecol. Evol.*, *2*(3):312-320. (Cited on page 37.)

Wolfe AD, Randle CP (2004) Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematics. *Syst. Bot.*, 29:1011-1020. (Cited on page 21.)

Xu M, Huaracha E, Korban SS (2001) Development of sequence-characterized amplified regions (SCARs) from amplified fragment length polymorphism (AFLP) markers tightly linked to the Vf gene in apple. *Genome*, 44(1):63-70. (Cited on page 5.)

Ye Q, Qiu YX, Quo YQ, Chen JX, Yang SZ, Zhao MS, Fu CX (2006) Species-specific SCAR markers for authentication of *Sinocalycanthus chinensis*. *J. Zhejiang Uni. Sci.*, 7(11):868-872. (Cited on page 5.)