Conservation genetics of threatened plants in NW Spain: **a practical approach**

PhD thesis by
Lúa López Pérez
UDC / 2014

Advisor: Rodolfo Barreiro Lozano
Department of Animal and Vegetal Biology and Ecology.

UNIVERSIDADE DA CORUÑA

UNIVERSIDADE DA CORUÑA

Conservation genetics in threatened plants in NW Spain: a practical approach.

PhD thesis by

Lúa López Pérez

UDC / 2014

A thesis submitted in fulfilment of the requirements of the Spanish Ministry of Education for the award of Doctor of Philosophy (Biological Sciences)

Advisor: Rodolfo Barreiro Lozano

PhD program: Environmental biology (RD 778/1998)

Department of Animal and Vegetal Biology, and Ecology

Dissertation submitted for the degree of Doctor of Philosophy to the University of A Coruña (Galicia, Spain)

UNIVERSIDADE DA CORUÑA

Declaration

I declare that this thesis, composed by myself and embodying work done by myself, has not been accepted in any previous application for a higher degree. All sources of references and quotation have been duly acknowledged.

Cover design by **Merienda diseño gráfico y fotografía** (tomalamerienda@gmail.com).

Contributions

Supervised by: Rodolfo Barreiro Lozano, Professor
University of A Coruña, Spain

Visit advisor: Marcus Koch, Professor
University of Heidelberg, Germany

Reviewed by: Jérôme Duminil, Associate Scientist
Université Libre de Bruxelles, Belgium.

Isabel Maneiro, Associate Scientist
Consiglio Nazionale delle Ricerche , Italy

UNIVERSIDADE DA CORUÑA

Rodolfo Barreiro Lozano, Professor at the Department of Animal Biology, Plant Biology, and Ecology (University of A Coruña) has supervised the thesis entitled "Conservation genetics of threatened plants in NW Spain: a practical approach" written by **Lúa López Pérez**. The thesis fulfills all the requirements of the Spanish Ministry of Education for the award of Doctor of Philosophy (Biological Sciences) and for the international mention.

In my opinion, the thesis defense can take place subject to the approval of the thesis examination board.

A Coruña, 28th of April, 2014.

Rodolfo Barreiro Lozano, PhD

"And then, I started to write…."

(Dick Powell, The Bad and the Beautiful/Cautivos del mal)

"All of the rocky and metallic material we stand on, the iron in our blood, the calcium in our teeth, the carbon in our genes, were produced billions of years ago in the interior of a red giant star. ***We are made of star-stuff. There are pieces of star within us all!"***

(The Cosmic Connection: An Extraterrestrial Perspective, Carl Sagan)

"Nothing shocks me. I'm a scientist. / Nada pode sorprenderme. Son un científico."

(Indiana Jones, Temple of Doom/O templo maldito)

CONTENTS

SUMMARIES

ABSTRACT

Appropriate management of plants of conservation concern requires reliable estimates of the magnitude and spatial distribution of genetic diversity as these species often combine features that make them potentially susceptible to genetic erosion. In this regard, the present thesis focuses on applying genetic markers to the conservation of rare and threatened plants.

In the first two chapters, genetic diversity and population structure of the clonal endemism *Centaurea borjae* is assessed using AFLPs and cpDNA sequences. *C. borjae* displayed intermediate-low genetic diversity compared to other plants with similar life-history traits. Gene flow seem to be restricted as populations separated by few hundred meters showed significant differentiation. Clonal frequency was lower than anticipated and might be related to soil type. Five Management Units were designated for conservation purposes and sampling for *ex situ* preservation should focus on individuals separated >80 m.

In the third chapter, the neutral and quantitative diversity of the endangered therophyte *Omphalodes littoralis* spp. *gallaecica* is investigated. The five extant populations displayed minimal to none neutral genetic diversity and a lack of gene flow between them. Reciprocal transplant experiments showed among-population differentiation in several quantitative traits but the pattern of differences did not fit the expectations of local adaptation. Instead, it seemed to be caused by genetic drift. Based on the genetic and phenotypic results, each population should be designated as an independent Evolutionary Significant Unit for conservation purposes.

The last chapter focuses on developing SSRs markers for threatened plants using EST sequences available in public databases. 257 genera were analyzed and 86% of them were successfully mined. As most of these genera lack an annotated genome, *Arabidopsis* and *Oryza* were used as controls for genome distribution analyses. Dimers

and trimmers were prevalent types of repeat. Control genomes revealed that trimmers were mostly located in coding regions while dimers were largely associated to untranslated regions. Finally, empirical trials showed that EST-SSRs had high amplification success and were 100% transferable between species in two tested genera.

RESUMEN

La adecuada gestión de plantas con especial interés para la conservación requiere conocer la magnitud y la distribución espacial de la diversidad genética, ya que estas especies a menudo presentan características que las hacen más susceptibles a la erosión genética. En este contexto, la presente tesis se centra en la aplicación de marcadores moleculares para la conservación de plantas raras y amenazadas.

En los dos primeros capítulos se investiga la diversidad genética y la estructura de población del endemismo clonal *Centaurea borjae* empleando AFLPs y secuencias del genoma del cloroplasto. *C. borjae* mostró una diversidad genética intermedia-baja en comparación con otras plantas con rasgos vitales similares. El flujo genético está restringido, ya que poblaciones distanciadas unos cientos de metros presentaron diferencias significativas. La frecuencia de clones fue inferior a la esperada y parece estar relacionada con el tipo de suelo. Finalmente, se recomienda establecer cinco Unidades de Gestión y mantener una distancia >80 m entre individuos recogidos para conservación *ex situ*.

A lo largo del tercer capítulo, se investiga la diversidad neutral y cuantitativa del terófito amenazado *Omphalodes littoralis* spp. *gallaecica*. Las cinco poblaciones existentes revelaron una diversidad genética neutral mínima o cero además de ausencia de flujo genético entre ellas. Mediante experiencias de trasplante recíproco, se encontraron diferencias entre poblaciones en varios caracteres cuantitativos pero

dicha diferenciación no se ajustó a un patrón de adaptación local. Por contra, la variación fenotípica parecía ser consecuencia de la deriva genética. En base a los resultados genéticos y fenotípicos, cada población debe considerarse como una Unidad Evolutivamente Significativa independiente a efectos de conservación.

El último capítulo se centra en desarrollar marcadores SSR para plantas amenazadas utilizando secuencias EST disponibles en bases de datos públicas. Se estudiaron 257 géneros y el 86% de ellos fueron analizados con éxito. Como la mayoría de estos géneros carecen de genomas anotados, *Arabidopsis* y *Oryza* se emplearon como controles para determinar la distribución de los EST-SSRs a lo largo del genoma. Dímeros y trímeros fueron los tipos de repeticiones más abundantes. Los genomas de control revelaron que los trímeros están distribuidos principalmente en regiones de codificantes, mientras que los dímeros se asocian mayoritariamente con regiones no codificantes. La tasa de amplificación fue buena. Además, fueron transferibles entre especies del mismo género.

RESUMO

Unha adecuada xestión en plantas con especial interese para a conservación require coñecer a magnitude e a distribución espacial da diversidade xenética, xa que estas especies a miúdo posúen características que as fan máis susceptibles á erosión xenética. Neste contexto, a presente tese centrase na aplicación de marcadores moleculares para a conservación de plantas raras e ameazadas.

Ó longo dos dous primeiros capítulos investigase a diversidade xenética e a estrutura poboacional do endemismo clonal *Centaurea borjae* empregando AFLPs e secuencias do xenoma do cloroplasto. *C. borjae* amosou una diversidade intermedia-baixa en comparación con outras plantas con rasgos vitáis similares. O fluxo xenético parece estar restrinxido, xa que poboacións distanciadas uns centos de metros presentaron diferencias significativas. A presencia de clons foi inferior á esperada e

parece estar relacionada co tipo de solo. Finalmente, recoméndase establecer cinco Unidades de Xestión e manter unha distancia >80 m entre individuos recollidos para conservación *ex situ*.

Ó longo do terceiro capítulo, investigase a diversidad neutral e cuantitativa do terófito ameazado *Omphalodes littoralis* spp. *gallaecica*. As cinco poboacións existentes revelaron unha diversidade xenética neutral mínima ou cero e ausencia de fluxo xenético entre elas. Os transplantes recíprocos amosaron diferencias entre poboacións para varios caracteres cuantitativos, non obstante dita diferenciación nos se axustou a un patrón de adaptación local. Pola contra, a variación fenotípica pareceu ser consecuencia da deriva xenética. En base ós resultados xenéticos e fenotípicos, cada poboación debe considerarse como unha Unidade Evolutivamente Significativa independente para fins da súa conservación.

O último capítulo centrase no desenvolvemento de marcadores SSR para plantas ameazadas empregando secuencias EST dispoñibles en bases de datos públicas. Estudiáronse 257 xéneros e o 86% dos mesmos foron analizados con éxito. Como a maioría de estes xéneros carecen de xenomas anotados, *Arabidopsis* e *Oryza* empregáronse como controles para determinar a distribución dos EST-SSRs ó longo do xenoma. Dímeros e trímeros foron os tipos de repeticións máis abundantes e os xenomas de control revelaron que os trímeros distribúense principalmente en rexións codificantes, mentres que os dímeros están maioritariamente asociados con rexións non codificantes. O éxito de amplificación dos EST-SSRs foi bo e ademais, foron transferibles entre especies do mesmo xénero.

INTRODUCTION

INTRODUCTION

Plant conservation genetics

Ecology is the science dealing with the interactions that determine the distribution and abundance of organism (Krebs, 1972). Thus, ecologists aim to understand the processes that influence biodiversity. In the modern world, a major concern is the loss of biodiversity that can be mostly attributed to human factors. Human influence has deeply altered the natural environment, modifying the territory, exploiting species directly, changing biochemical cycles and transferring species between areas. Main **threats to biodiversity** loss can be summarized as:

- Alteration and loss of habitats: the transformation of natural areas impacts the number and abundance of species.
- Introduction of alien species and genetically modified organisms: species introduced into a new environment can lead to disequilibrium in the ecosystem.
- Pollution: pollution alters the chemical and physical features of the environment, resulting in changes in the diversity and abundance of species.
- Climate change: Earth's surface warming affects biodiversity as it threatens species that are adapted to cold (i.e. polar species) or to high altitudes (i.e. alpine species).
- Overexploitation: excessive harvesting of natural resources may exhaust them.

In this scenario, **conservation biology** emerged with the aim to minimize the loss of biodiversity and to ensure the maintenance of threatened species. The publication in 1981 of "Conservation and Evolution" by Frankel and Soule pioneered the scientific framework for conservation biology by demonstrating how evolution and the dynamics of genetic diversity, within and among populations, are pivotal for

preserving endangered species. Since then, a growing body of literature has addressed conservation issues (Allendorf and Luikart, 2013; Hamrick and Godt, 1996; Frankham *et al.*, 2010; Mills, 2006).

The International Union for Conservation of Nature (IUCN) recommends preserving the biological diversity at three levels: genes, species, and ecosystem (McNeely *et al.*, 1990). In this context, conservation genetics arises as an applied science that uses molecular tools and evolutionary genetics for conservation purposes (Hamrick and Godt, 1996; Frankham *et al.*, 2010; Mills, 2006). Appropriate conservation strategies require reliable estimates of the magnitude and spatial distribution of genetic diversity within and among populations, as it is the raw material for species to evolve and adapt in response to changing environments (Frankham, 2005; Frankham *et al*., 2010; Hamrick and Godt, 1996). This knowledge is even more relevant in threatened and/or rare plants as they often combine several features that make them potentially susceptible to genetic erosion and lower adaptability: small population size, habitat specificity, and isolation (Ellstrand & Elam, 1993; Cole, 2003; Hamrick & Godt, 1996; Leimu *et al.*, 2006) (Fig. 1). From now on, and for a lighter reading, the term threatened and/or rare species will be referred only as rare species.

Species that have experienced a reduction in gene flow and/or population size have been found to be more sensible to genetic erosion due to small population size (Aguilar *et al.*, 2008; Honnay and Jacquemyn, 2007). In this context, many **rare species** occur in small isolated populations and usually display reduced levels of genetic diversity (Cole, 2003; Ellstrand and Elam, 1993). Nevertheless, the premise that rare plants have lower genetic diversity is far from universal and needs to be further examined (Gitzendanner and Soltis, 2000). Besides, low levels of neutral genetic diversity may not necessarily lead to a loss of adaptive variation (Bekessy *et al.*, 2003; Landguth and Balkenhol, 2012; Reed and Frankham, 2001; Reed and Frankham, 2003).

Still, it seems undeniable that many plant populations are currently experiencing severe reductions and a growing isolation that might compromise their evolutionary potential because of habitat fragmentation, habitat destruction and environmental stress. Under these circumstances, plant conservation genetics may play a pivotal role in the preservation of rare species.

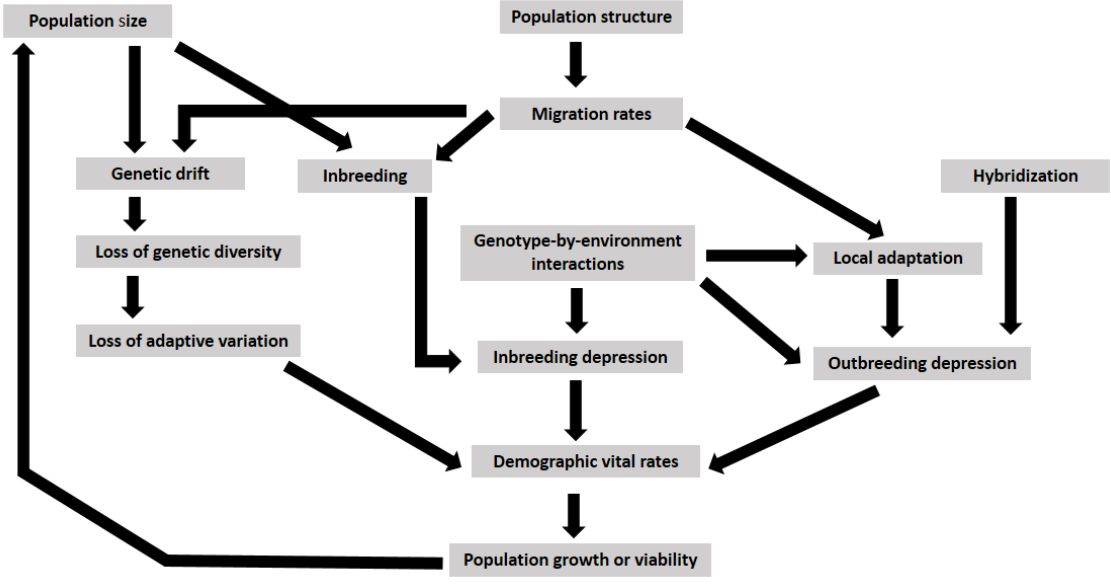


Fig. 1: Interacting factors in the conservation of natural populations (adapted from Allendorf *et al.*, 2010).

Most rare plants have **small population sizes** and their populations often experience a decreasing trend. In this regard, it is important to recall that census size (the number of individuals constituting a population) is usually larger than effective population size (N_e) (Wright, 1931). Species with small N_e are more prone to genetic bottlenecks and genetic drift (Hamrick *et al.*, 1991). Bottlenecks are sharp decreases in the number of individuals of a species that are highly likely to be accompanied by a significant loss in genetic diversity. Moreover, if the population undergoes several consecutive bottlenecks in time, the loss of genetic diversity will be exacerbated (Willi *et al.*, 2006). Isolated populations with reduced genetic diversity are also more sensitive to the effects of genetic drift (Ellstrand and Elam, 1993; Willi *et al.*, 2006).

When random genetic drift occurs, some alleles (specifically rare ones) may be lost just by chance and allele frequencies in subsequent generations probably differ from the parental ones causing the erosion of the genetic diversity of the population (Hamrick and Godt, 1996).

Severe reductions in population size are also likely to lead to **inbreeding** (the mating of relatives). Inbreeding occurs naturally in many plant species that reproduce by selfing (Huenneke, 1991). However, mating among relatives can have serious consequences for fitness in plants with mixed-mating and out-breeders (Angeloni *et al.*, 2011). Inbreeding can lead to the fixation of deleterious alleles, reducing reproductive output and survival (i.e. **inbreeding depression**) (Angeloni *et al.*, 2011). Despite earlier scepticism, there is now compelling evidence that inbreeding depression can have an impact on wild populations (Crnokrak and Roff, 1999; Keller and Waller, 2002), and that its negative effects increase in stressful habitats compared to benign ones (Armbruster and Reed, 2005). Nevertheless, the severity of inbreeding depression depends on several factors. Perennial species displayed significantly greater inbreeding depression than annual ones (Angeloni *et al.*, 2011). Likewise, outcrossing species usually displayed higher inbreeding depression than selfers (Angeloni *et al.*, 2011; Frankham *et al.*, 2010). Moreover, inbreeding depression was found to be positively correlated with increasing population size (Angeloni *et al.*, 2011). The latter may be a consequence of genetic purge as mating among relatives for long periods of time helps to remove deleterious alleles. Thus, genetic purge is more likely to occur in small rather that big populations (Crnokrak and Barret, 2002; Glémin, 2003; Goodwillie *et al.*, 2005).

The patterns of genetic diversity are shaped by multiple factors among which **life-history traits (LHTs)** are regarded as highly determinant (Hamrick *et al.*, 1991; Nybom, 2004). Genetic diversity can partitioned at species, within population and among population level. Life form, geographical range and breeding system are highly

influential at species level (Hamrick *et al.*, 1991, Nybom, 2004). Short-lived and annual plants usually display lower genetic diversity than long-lived ones (Nybom, 2004). Similarly, selfing, mixed-mating and animal-pollinated taxa commonly have less genetic diversity than their outcrossing counterparts (Hamrick *et al.*, 1991, Nybom, 2004). Plants with restricted geographical range commonly show less variation than widespread taxa. According to Hamrick *et al.* (1991), the patterns mentioned above is maintained when genetic variation is considered at within population level. However, the distribution of the genetic diversity among populations follows a different pattern. Annual and/or selfing species usually showed higher among-population differentiation than long-lived and/or outcrossed taxa; geographical range, however, seemingly had no effect on genetic diversity among populations (Gitzendanner and Soltis, 2000; Hamrick and Godt, 1990; Honnay and Jacquemyn, 2007). In general, species with limited potential to disperse display greater genetic differentiation among populations than those with efficient dispersal. In this regard, Loveless and Hamrick (1984) estimated that selfing species harbored 56% of their allelic diversity within populations. Despite the general assumption that LHTs correlate with the pattern of genetic diversity, recent studies have noted that this tenet must be further discussed (Duminil *et al.,* 2007; Duminil *et al.,* 2009). Most of the reviews about this topic did not consider the phylogenetic independency across the studied taxa in their analyses. When the latter is taken into account, genetic structure was shown to be influenced only by a few LHTs such as mating system for nuclear markers and dispersal mode or geographic range size for organelle markers (Duminil *et al.,* 2007). Besides, plant traits that correlate with generation time influence mating system and inbreeding depression affecting genetic drift and gene flow and eventually modifying the genetic structure of the population (Duminil *et al.,* 2009).

Dispersal is one of the core processes involved in the dynamics and evolution of plant populations (Ouborg *et al.*, 1999). Population spatial dynamics is determined

by seed and pollen movement, which often display different modes and distances of dispersal (Garcia *et al.*, 2007). Overall, restricted gene flow commonly results in spatial genetic structure while high levels of gene flow usually lead to a random distribution of genotypes (Turner *et al.*, 1982; Wright, 1943; Wright, 1978;). The extent of pollen dispersal is determined by the mediator vector. For example, wind-pollinated species usually have a wide-range dispersal while gene flow can be restricted in animal-pollinated plants depending on the behaviour of the disperser (Garcia *et al.*, 2007). Self-fertilizing and clonal species are expected to have very low dispersal (Hamrick and Godt, 1996). Likewise, seed movement is also shaped by the disperser vector. Dispersal is usually restricted to very short distances in plants that disseminate their seeds by gravity. In contrast, dispersal distance is notably longer in anemochorous or zoochorous species (Cain *et al.*, 2000). Species with very limited dispersal capabilities are expected to have a strong population structure due to the non-random spatial distribution of genotypes, where genetic similarity is higher among neighbouring than among more distant individuals (Wright, 1943).

Genetic differentiation among populations can also be consequence of **adaptation** rather than genetic-drift or restricted dispersal. In fact, plant populations are commonly assumed to be locally adapted (Leimu and Fischer, 2008). In the absence of other forces and constrains, resident genotypes in each population would have on average a higher relative fitness in their local habitats than genotypes arriving from other habitats (Kawecki and Ebert, 2004). However, when further examined, this premise does not seem to be a general trend. Only 43.5% of the species reviewed by Leimu and Fisher (2008) performed better in their local habitats than in foreign ones. Moreover, these authors noted that the ability of a plant to adapt seems to be independent of its life history, spatial and temporal heterogeneity, and geographic scale. Instead, they found that local adaptation was more commonly displayed by

large populations, supporting the idea that small populations may have significantly reduced their ability to cope with changing environments (Willi *et al.*, 2006).

Although it is widely acknowledged that many possible factors can determine the genetic variation and structure of a particular species, we often operate under the unproven assumption that rare plants may have their evolutionary potential diminished. This approach seems inappropriate. Instead, formulating scientifically rational **conservation actions** that may minimize the extinction risk of a particular plant requires the appropriate assessment of its genetic diversity and structure (Aguilar *et al.*, 2008; Frankham, 2010; Tallmon *et al.*, 2004). In this regard, the genetic information derived from neutral molecular markers seems a crucial element in the development of accurate conservation initiatives, both *in situ* and *ex situ*. *Ex situ* efforts in plants typically involve **germplasm** (mostly seeds) storage where a common issue is to attain a sampling regime that may encompass the full genetic diversity of the species and its local populations. For germplasm collection, a minimum sampling distance can be determined by fine-scale spatial genetic structure analysis (SGS) where a kinship coefficient quantifies the degree of relatedness between each pair of individuals (Vekemans and Hardy, 2004). SGS is then used to set the minimum distance between individuals that will guarantee a maximum coverage of the population genetic diversity. An example of this approach can be seen below in chapter 1 where SGS was used to recommend *ex situ* conservation actions.

The genetic management of endangered wild populations also involves delimiting **management units** (MUs) (Palsboll *et al.*, 2007) (see chapters 1 and 2 for further explanations). MUs are described as demographically independent units (Avise, 1995; Moritz, 1999) and they are diagnosed as populations displaying differences in allele frequencies at organelle DNA and/or nuclear loci (Avise, 1995; Moritz, 1994). When differentiation goes beyond divergences in allele frequencies and also involves differences in quantitative traits, the concept of MUs becomes

insufficient and **Evolutionary Significant Unit** (ESU) seem more appropriate (Crandall *et al.*, 2000; Moritz, 1999) (see chapter 3 for further information). The distinction between MUs and ESUs seems particularly relevant in cases where conservation strategies may involve an exchange of individuals between populations as translocations might be allowable between MUs but not between ESUs. The transfer of individuals adapted to local conditions might have negative consequences due to outbreeding depression (Mills, 2006).

Neutral markers are useful for determining genetic relationships among individuals, among populations (gene flow and population structure), or the demographic history, but they are considered to have no impact on phenotypes or fitness (Reed and Frankham, 2001). Interestingly, the characters of greatest concern in conservation biology are those associated with **quantitative variation** as it determines the ability of the species to cope with environmental changes and to evolve (Frankham *et al.*, 2010). Unfortunately, the relationship between neutral markers and adaptive variation has been found to be weak at best (Bekessy *et al.*, 2003; Reed and Frankham, 2001) and variation in quantitative traits is known to be due to both genetic and environmental factors. In chapter 3, there is an example where a plant with minimal to none neutral variation at deme scale still shows variability in a number of quantitative traits.

The recent increase of large, publicly available DNA sequence datasets generated by high-throughput techniques and the growing emphasis on functional genomics can greatly facilitate the use of molecular approaches in non-target species of conservation concern (Allendorf *et al.*, 2010; Luikart *et al.*, 2003). In chapter 4, we show a cost-effective procedure to develop molecular markers for population studies in endangered plants using DNA sequences generated by high-throughput. Is in this context where conservation genetics goes one step further evolving into **conservation genomics** (Ouborg *et al.*, 2010; Primmer, 2009). Even if conservation genomics is a

new disciple still in its infancy, it is quite promising. Genomics already has provided some interesting surprises, such as the discovery of adaptive loci that displayed high divergence between populations.

Methods in conservation genetics

There are several types of **molecular marker techniques** currently available but none of them can be regarded as universally "best". The most suitable technique to assess genetic variation depends upon both the question addressed and the type of genetic information available for the species (Allendorf and Luikart, 2013). In fact, the popularity of the major types of molecular markers has changed along the last two decades (Fig. 2). Here we provide a brief overview of the various markers used in conservation genetics with their respective applications (Table 1).

Genetic variation is most commonly inferred using markers that are expected to be neutral or nearly neutral, this is, that there is no evidence of selection involved in shaping their alleles frequencies (Höglund, 2009). Neutral markers have proved suitable for conservation studies interested in estimating population sizes, population structure, genetic variation, genetic drift and inbreeding (Allendorf and Luikart, 2013).

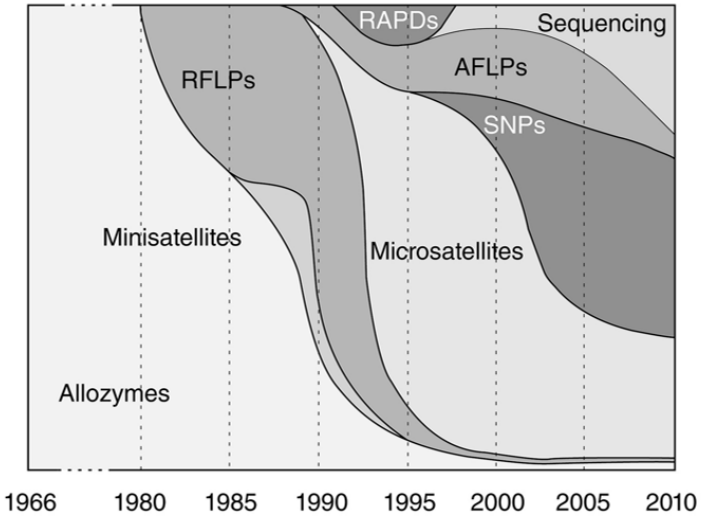


Fig. 2: Changes in the popularity of major molecular markers in conservation genetics. The horizontal axis indicates time and the vertical axis corresponds to the relative use of molecular markers at that time (extracted from Allendorf *et al.* 2013).

Among the most commonly used neutral markers, we have allozymes, Restriction Fragment Length Polymorphism (RFLPs), Microsatellites or Short Tandem Repeats (SSRs), Amplified Fragment Length Polymorphism (AFLPs) and DNA sequencing.

Table 1: Comparison of different molecular markers used in conservation genetics (adapted from Schötterer, 2004).

Markers	Advantages	Disadvantages
Allozymes	- Inexpensive - Universal protocols	- Require fresh or frozen material - Some loci show protein instability - Limited number of available markers - Can be a target of natural selection
RAPDs and AFLPs	- Inexpensive - Produces a large number of bands, which can then be further characterized individually	- Very sensitive to DNA quality, might lead to low reproducibility - Dominant - Difficult to analyse - Difficult to automate - Cross-study comparisons are difficult
Microsatellites	- Highly informative - Low ascertainment bias - Easy to isolate	- High mutation rate - Complex mutation behaviour - Not abundant enough - Difficult to automate - Cross-study comparisons require special preparation - Expensive development
DNA sequencing	- Highest possible level of resolution - Unbiased - Easy cross-study comparisons; data repositories already exist	- More expensive than the other techniques (but prices have experienced a continuous decrease)
SNPs	- Low mutation rate - High abundance - Easy to genotype - New analytical approaches in development - Easy cross-study comparisons; data repositories already exist	- Substantial rate heterogeneity among sites - Expensive development - Ascertainment bias - Low information content of a single SNP

Allozymes, also known as isozymes, are neutral, co-dominant markers described as alternative forms of a protein detected by electrophoresis that are the consequence of alternative alleles at a single locus (Allendorf and Luikart, 2013). Allozymes were the first molecular markers widely used in conservation genetics; they were very popular until the early nineties and there are many examples of their use at inferring genetic variation in rare plants. Two particularly relevant works are the

seminal paper by Hamrick and Godt in (1990) and the review by Hamrick (1983) published in the book "Genetics and Conservation". Today, the use of allozymes is mostly anecdotical and very few examples exist in the modern literature due to the low number of informative loci and doubts about their neutrality (Schlötterer, 2004).

The arrival of DNA-based markers revolutionized the field and promoted a shift from enzyme-based markers. DNA-based markers owe their popularity to the fact that they provided a direct survey of DNA variation rather than relying on variations in the electrophoretic mobility of proteins (Allendorf and Luikart, 2013). **Restriction Fragment Length Polymorphism (RFLP)** are dominant molecular markers generated by a single substitution in the restriction site recognized by an enzyme (e.g. from GAA**A**TTC to GA**T**TTC) that causes the absence of restriction in the individual. RFLP analyses of mitochondrial (mtDNA) and ribosomal (rDNA) DNA were largely used in the mid-1980s and early 1990s for investigating population structure and genetic variation (Avise, 1994) before being replaced by the more informative microsatellites.

Minisatellites are another marker of the past: tandem repeats that usually display length polymorphism as consequence of unequal crossing over or gene conversion. Like in RFLPs, the first step of minisatellites analysis involves the digestion of genomic DNA with restriction enzymes; however, they represent a different concept of molecular marker (Frankham *et al.*, 2010). Their extremely high variability revolutionized the genetic identification of individuals (i.e. DNA fingerprinting) in the late 1980s but they were very briefly used because they cannot be applied in standard population genetics given the high complexity of their banding patterns.

The main breakthrough in the history of the DNA markers was the invention of the **Polymerase Chain Reaction (PCR)** (Mullis *et al.* 1986; Mullis and Faloona, 1987). PCR allowed, for the first time, the amplification of a genomic region in many individuals without cloning or isolating large amounts of ultra-pure genomic DNA. The first widely used PCR-based markers were microsatellites or **Short Sequence Repeats**

(SSRs). These are short tandemly repeated sequences that have become the marker of choice in many population genetic analysis because of their co-dominance, high polymorphism and considerable abundance along the genome (Selkoe and Toonen, 2006). Nevertheless, SSRs also have disadvantages. Their development is a time-consuming and expensive task and they can suffer technical problems (e.g. PCR artefacts such as stutter peaks) that complicate their automatic scoring (Schötterer, 1998). Also, SSRs are species-specific, meaning that cross-amplification between relative species is very low and must be developed anew each time we move into a new species. However, see chapter 4 below for an example where a variant of SSRs (EST-SSRs) were highly transferable between species of the same genus.

Another class of PCR-based markers are **Randomly Amplified Polymorphic DNA (RAPD)** and **Amplified Fragment Length Polymorphism (AFLP)** (Schötterer, 1998), two types of marker that bind to multiple sites in the genome. Here, we restrict our comments to AFLP as the RAPD technique was soon avoided due to reproducibility problems and its presence in plant conservation studies is notably scarce. AFLPs are genome-wide markers that amplify restriction fragments by adding linkers. A main advantage of AFLPs is that they do not require previous knowledge of the genome (Allendorf and Luikart, 2013). This has been proved particularly useful in the study of population genetics of rare plant species (Mba and Tohme, 2005; Palacios *et al.*, 1999) and chapters 1 and 3 in this thesis provides other examples of the use of AFLPs in rare plants. AFLPs are dominant markers that do not allow detecting heterozygotes. Nevertheless, their dominant nature is offset by the high number of loci that can be detected. As in the case of SSRs, there are some technical problems that need to be considered when dealing with AFLPs. AFLPs require very high quality DNA that must be free of secondary metabolites such as polyphenols which can interfere with the restriction reaction eventually resulting in reproducibility issues (Bonin *et al.*, 2004).

Stutter peaks can be also common, hindering an automatic scoring (Schlötterer, 2004).

Finally, sequencing a particular region of the genome provides the most fine-grained information. Several regions of the **organelle DNA** have been widely used to investigate plants. Organelle DNA often displays uniparental inheritance with little or no crossing over compared to nuclear DNA (McCauley, 1995). In plant conservation genetics, organelle DNA has become a standard tool for assessing intraspecific population structure and gene flow. Chloroplast DNA is maternally inherited and it can only be dispersed by seeds but not by pollen (McCauley, 1995). Thus, contrasting patterns between organelle and nuclear markers can help to evaluate the relative influences of seed and pollen dispersal in the species genetic structure. Moreover, unlike SSRs or AFLPs, organelle-derived sequences can be historically ordered. As a result, they provide information on population histories (Avise, 2004) as shown in chapters 2 and 3 below. Chloroplast DNA and, to a lesser extent, mtDNA have been useful in plant conservation genetics interested in gene flow and phylogenetic histories reconstruction. A clear example of the latter is the use of the universal primers described by Taberlet *et al.* (1991) for the cpDNA region *trnT-L* (cited 2916 times, information from the ISI Web of Science). Chapters 2 and 3 used region *trnT-L* to ascertain the phylogeography of the two plants used in this thesis.

The recent explosion of **Next Generation Sequencing (NGS)** techniques have opened a new world of possibilities in conservation genetics. Large scale sequencing is becoming an accessible tool for studying natural populations. In this regard, Single Nucleotide Polymorphisms (SNPs) are the commonest type of polymorphism in the genome with a density of one every 200-500bp (Allendorf and Luikart, 2013). The most comprehensive way to identify SNPs towards the genome is through shotgun genome sequencing of a pool of individuals used as donors of genomic DNA. SNPs can be useful for describing genetic variation in natural populations; however, their

development is time- and cost-intensive (Schlötterer, 2004). Moreover, the position of the SNPs is impossible to know in non-model organism that lack an annotated genome. While SNPs located in intergenic regions or introns are consider to evolve neutrally, this premise does not hold for those located in exons (Allendorf and Luikart, 2013). Thus estimates of population structure can be biased due to selective pressures.

The marker types discussed above are selectively neutral, not affecting phenotypes or fitness (Reed and Frankham, 2001). So far, studies addressing adaptation were based in Quantitative Trait Loci (QTL) analysis and outlier loci analysis but none of them directly address variation in genes (Frankham *et al.*, 2010) (see chapter 1 for an example of outlier loci analysis). Molecular markers derived from genic regions are called **functional markers** (Andersen and Lübberstedt, 2003). Unlike QTLs and outlier loci analysis, functional markers target directly gene variation. Specific genes that are known to have an effect on relevant phenotypic traits (i.e. candidate genes) from which there is sequence information for PCR primer design are an example of functional markers (Allendorf and Luikart, 2013). However, this type of markers are scarce because there is no genome information for most of them. Nevertheless, since coding regions are highly conservative, annotated genomes from model plant species (e.g. *Arabidopsis* or *Oryza*) can be crossed with those from non-model species. In this regard, SNPs that are known to be located in coding regions are more likely to have a phenotypic effect that may affect fitness and might be used as functional marker (Allendorf and Luikart, 2013).

Expressed Sequence Tags (ESTs) can also be used as a source for functional marker development (Varshney *et al.*, 2005a) (see chapter 4 for further information on the use of ESTs as a source of funtional markers). In the absence of a complete genome, ESTs sequences remain a useful proxy to the genome because they derive from the transcript portion of the genome. SSRs derived from Expressed Sequence

Tags (**EST-SSRs**) have been widely used and proved very useful in model plants (i.e. crops) but their used in non-model organism is still on its infancy (Varshney *et al.*, 2005a). The growing availability of EST sequence data for a wide range of taxa makes this type of marker a promising option in future conservation genetics studies. Besides their linking to coding regions, a major advantage of EST-SSRs is their transferability (Varshney *et al.*, 2005b). Should EST sequences be available for a species closely related to our pet organism (e.g. congenerics), the set of EST-SSRs developed from these EST sequences will likely work in our organism. Moreover, compared to the time and money needed for conventional SSRs discovery, EST-SSRs can be produced in a very short time with no additional cost after accessing the EST database (Ellis and Burke, 2007).

Pet species

The work presented here focuses in two endemic plants of NW Spain: *Centaurea borjae* Valdés-Bermejo and Rivas Goday (1978) and *Omphalodes littoralis* spp. *gallaecica* M. Laínz (1971). Both species are catalogued as "endangered" by the IUCN and the Spanish Catalogue of Threatened Species (Serrano and Carbajal, 2011) (Ministerio de Medio Ambiente y Medio Rural y Marino, 2011), and listed as priority species in EU Habitats Directive (92/43/EEC, Annex II). Their total occupancy is estimated to be very small, which is one of the main reasons of why they are listed as endangered. Additionally, their habitats are considered Sites of Community Importance (SCI) within the Natura 2000 network.

Centaurea borjae is a relict paleopolyploid endemic to NW Spain (Garcia-Jacas and Susanna, 1992) (Fig. 1). It is found only in six enclaves, all of them cliffs spread along <40 km of the coastline (Valdes-Bermejo and Rivas Goday, 1978) (Fig. 1). It has been estimated that the total occupancy of the species does not exceed 5000 m^2 (Bañares *et al.,* 2004). *C. borjae* is a small (up to 6 cm tall), entomophilous outcrossing

plant with hermaphroditic flowers (Valdés-Bermejo and Agudo Mata 1983; Valdes-Bermejo and Rivas Goday, 1978). Its germination success seems to be very low (Gómez-Orellana Rodríguez, 2004; R. Retuerto pers. comm.; but see Izco *et al.,* 2003 for other estimates) and insect larvae can be easily found feeding on ripe fruits within mature flower heads (Fernández Casas and Susanna, 1986). The fruit lacks a pappus and presents an elaiosome. The latter suggests that ants may play a role in seed dispersal. *C. borjae* also produces rhizomes up to several meters long that can give rise to new rosette leaves.

Fig. 1: *Centaurea borjae* Basal rosette with flower (left) and typical habitat of *C. borjae* (right).

Despite its status as priority species, there are no data on the magnitude and structure of the genetic diversity of *C. borjae*. Its LHTs lead to conflicting hypothesis about its genetic variation. On one hand, the occurrence of clonal propagation together with the low germination success suggest that populations might be formed by ramets originating from a few genets with negative consequences for the genetic diversity of populations (Izco *et al.,* 2003). However, self-incompatible outcrossers often display considerable levels of genetic variation (Cole, 2003; Hamrick and Godt, 1996; Nybom, 2004) and polyploids generally maintain higher levels of genetic diversity in small populations than diploids with comparable population sizes (Soltis and Soltis, 2000). On the other hand, the occurrence of fruits without a pappus and the probable myrmecochory could be regarded as indicators of restricted seed dispersal (Cousens *et al.,* 2008; Gomez and Espadaler, 1998) that might result in

significant genetic differentiation at small spatial scales. Given this lack of empirical data, the genetic structure and diversity of *C. borjae* was investigated in the first two chapter of the present thesis in an effort to formulate informed and effective management guidelines for its conservation, both *in situ* and *ex situ*.

Omphalodes littoralis spp. *gallaecica* is a rare herb (total occupancy <100000 m^2) restricted to coastal dune systems in NW Spain (Romero Buján, 2005, Serrano and Carbajal, 2011; Gómez-Orellana Rodríguez, 2011) (Fig. 2). Due to threats faced by its sensitive habitat, its populations have undergone a continuous decline in the last decades (Bañares *et al.,* 2004). Hence, its actual distribution is extremely fragmented and the plant is known to be present only in five dune systems. *O. littoralis* spp. *gallaecica* is a self-compatible plant and autogamy has been suggested as the most probable mechanism for reproduction (Bañares *et al.,* 2004). Flowering period is very short and the ephemeral flowers last less than three days (Romero Buján, 2005). Seed are thought to be dispersed by animals through the adhesiveness of the fruit to their hair (Bañares *et al.,* 2004). Population size fluctuates greatly between years, multiplying or dividing by ten the number of individuals found any given year (Bañares *et al.,* 2004).

Fig. 2. Detail of *Omphalodes littoralis* spp. *Gallaecica.* Habit of a plant with flowers (left) and typical habitat (right).

As in the case of *C. borjae* and despite the conservation concern of *O. littoralis* spp. *gallaecica*, its population genetics and the variation of its ecophysiological traits have never been addressed. Since autogamy is speculated as the most probable mechanism of reproduction in this small therophyte, genetic diversity within populations might be low (Hamrick *et al.*, 1991; Nybom, 2004). Likewise, the considerable fluctuations in population sizes between years might have led to the genetic erosion of the populations due to consecutive bottlenecks (Willi *et al.*, 2006). However, the latter might be buffered in presence of a stable seed bank (McCue and Holtsford, 1998; Nunney, 2002). Finally, high rates of selfing are known to be related with high levels of differentiation among populations (Nybom, 2004; Hamrick and Godt, 1996). If high levels of differentiation among populations are mantained through time, population might even evolve independiently resulting in procesess of local adaptation (Leimu and Fischer, 2008). Thus, it migth be expected that *O. littoralis* spp. *gallaecica* will displayed high differentiation among populations that may eventually lead to local adaptation of its populations. In this regard, chapter 3 provides an exhaustive molecular and phenotypic study of the five extant populations of this rare herb. Molecular and phenotypic information was combined to propose guidelines for the conservation of this endangered plant.

OBJECTIVES

OBJECTIVES

- General objective:

- The main objective of this thesis was employing molecular markers to investigate the genetic variation in rare and threatened plant species. Results were interpreted from an applied point of view and specific management guidelines were proposed for the conservation of these organism.

- Specific objectives:

- **Chapter 1**: AFLP phenotypes were used to investigate the genetic variation and population structure of *Centaurea borjae.* AFLP-derived information was used to (1) infer the contribution of clonal reproduction, (2) determine if populations show signs of diminished genetic variation, (3) infer minimum inter-plant distance for appropriate germplasm collection, (4) determine whether populations are significantly differentiated from each other and, if so, whether it is possible to delineate management units.
- **Chapter 2**: The genetic structure of *Centaurea borjae* along its range and the historical processes behind it were investigated using sequences of the non-coding cpDNA region *trnT-F* (Taberlet *et al.*, 1991). cpDNA information was used to estimate the genetic diversity of *C. borjae*, investigate its demographic past, evaluate its population structure, identify populations of greater conservation concern and, finally, compare the pattern obtained with cpDNA sequences with the results of the AFLP shown in chapter 1.
- **Chapter 3**: An exhaustive molecular and phenotypic study of the five extant populations of the rare herb *Omphalodes littoralis* spp. *gallaecica* was carried out in this chapter. Chloroplast sequences form the trnT-F region and AFLP-genotypes were used to (1) ascertain whether *O. littoralis* spp. *gallaecica* is genetically impoverished as suggested by its life history traits, (2) whether its populations are significantly differentiated from each other, and (3), given that

O. littoralis spp. *gallaecica* is a therophyte, whether there are significant between-year differences in its genetic structure. On the other hand, a series of reciprocal transplant experiments were performed to investigate the adaptive component of several quantitative traits related to fitness. Phenotypic variation was examined to reveal whether there are there any phenotypic differences between populations. These differences were further investigated to assess whether they result from phenotypic plasticity or have a genetic basis and if they might be adaptive. Finally, molecular and phenotypic information were combined to propose specific guidelines for the conservation of this endangered plant.

- **Chapter 4**: This chapter explores a rather underexploited yet clearly promising application of EST-SSRs: the development of markers from public EST databases for use in evolutionary and conservation genetic studies of non-model plant species (with emphasis on threatened ones). All plant genera included in the International Union for Conservation of Nature and Natural Resources (IUCN) Plant Red List with EST sequences available in the GenBank EST database were searched for SSRs. Since most of these plant genera do not include model organisms, there are no available annotated reference genomes for comparison, hampering the location of the EST-SSRs within the genome (i.e. intergenic regions, introns, UTRs or exons). To minimize this obstacle, the EST sequences of two model genera with well-known annotated genomes were in-depth analyzed and used as a proxy: *Arabidopsis* was selected as a control for eudicots while *Oryza* was used as a guide for monocots. Finally, twenty-four of the developed SSR were tested for amplification, cross-amplification, and polymorphism in four species of conservation interest from two genera (*Trifolium fragiferum, Trifolium saxatile, Centaurea valesiaca* and *Centaurea borjae)*.

CHAPTER I

"Genetic guidelines for the conservation of the endangered polyploid *Centaurea borjae* (Asteraceae)"

Published as: Lopez L. & Barreiro R. (2013). Genetic guidelines for the conservation of the endangered polyploid Centaurea borjae (Asteracea). Journal of Plant Research. 126 (1): 81-93. doi: 10.1007/s10265-012-0497-3. Epub 2012 Jun 8.

ABSTRACT

Appropriate management of species of conservation concern requires designing strategies that should include genetic information as small population size and restricted geographic range can reduce genetic variation. We used AFLPs to investigate genetic variation within and among populations of the endangered narrow endemic *Centaurea borjae,* and found no evidence for genetic impoverishment despite its < 40 km range and potential for vegetative propagation. Genetic variation was comparable to other plants with similar life history (88% occurring within populations) and potential clone mates were less frequent than expected. Nonetheless, populations separated by few hundred meters showed signs of significant genetic differentiation suggesting low gene flow between them. Our results suggested that the three geographically closer populations located at the center of the range might be treated as a single management unit, while the remaining ones could be considered independent units. We found evidence of fine-scale spatial genetic structure up to 80 m indicating that the collection of germplasm for *ex-situ* conservation should focus on individuals separated >80 m to maximize genetic variation.

Keywords: *Centaurea borjae*, conservation, endangered species, genetic diversity, polyploidy.

CHAPTER 1

INTRODUCTION

Narrow endemics, i.e. taxa that occur in one or a few small populations confined to a single domain or a few localities (Kruckeberg and Rabinowitz, 1985), are interesting cases of naturally rare species. Small population sizes, habitat specificity, and isolation often account for their status as taxa of conservation concern which can also increase their sensitivity to demographic and environmental stochasticity (Frankham, 2005; Kruckeberg and Rabinowitz, 1985). These features also anticipate that narrow endemics may harbor low genetic variation. Genetic drift and inbreeding can lead to a loss of genetic diversity in isolated and small populations (Frankham *et al.,* 2002) with negative consequences for the evolutionary potential and which can also enhance the extinction risk (Frankham, 2005; Willi *et al.,* 2006). In this regard, a number of neutral marker studies have found that rare and/or endemic plants often show less genetic variability than widespread taxa (Cole, 2003; Ellstrand and Elam, 1993; Gitzendanner and Soltis, 2000; Hamrick and Godt, 1996). Nonetheless, the association between genetic diversity and range size is far from universal. Various comparative studies also revealed that endemic and rare taxa can maintain levels of diversity equal to or exceeding that of widespread congeners (Cole, 2003; Gitzendanner and Soltis, 2000). In fact, other factors besides range size can be influential for the genetic variability of a plant species as well. Outcrossing species commonly have higher levels of genetic diversity, and lower differentiation between populations, than selfing and clonal plants (Cole, 2003; Chung and Epperson, 1999; Hamrick and Godt, 1996; Nybom, 2004; Palacios *et al.,* 1999; Stehlik and Holderegger, 2000). Also, polyploids may harbor more genetic diversity when compared to diploid species (Soltis and Soltis, 2000). Predicting the actual genetic variation and structure of a particular narrow endemic is difficult and, instead, it must be investigated on a case by case basis.

Most members of the genus *Centaurea* (*Asteraceae*) are common and widespread. However, a few of them are endemics with a narrow distribution. An interesting example of this is *Centaurea borjae* Valdés-Bermejo and Rivas Goday (1978), a relict paleopolyploid, member of section Acrocentrum endemic to the Iberian Peninsula (Garcia-Jacas and Susanna, 1992) (Fig. 1). The origin of this hexaploid (2n=66, x=11) plant is somewhat obscure and the parental species are unknown. However, hexaploids in section Acrocentrum are commonly considered allopolyploids (Font, 2007; Font *et al.,* 2009). Habitat type is likely to play a determinant role in the existence of this perennial herb as it is found only along < 40 km of the marine coastline of NW Spain where it occurs in a few enclaves on the mid-upper slopes of very tall coastal cliffs (Valdes-Bermejo and Rivas Goday, 1978) (Fig. 1). Most enclaves are characterized by thin soils developed on a range of metamorphic substrata (serpentinites, amphibolites, gneisses). Recently, a new site was discovered on igneous soil (granitoid) in a relatively isolated isthmus (approximately, 25 km away from the other sites) (Soñora, 1994). It has been estimated that the total occupancy of the species does not exceed 5000 m^2 (Bañares *et al.* 2004). *C. borjae* is a small (up to 6 cm tall), entomophilous outcrossing plant with hermaphroditic flowers (Valdés-Bermejo and Agudo Mata 1983; Valdes-Bermejo and Rivas Goday 1978). Although not specifically tested in *C. borjae*, self-incompatibility is known to be common in *Asteraceae*, particularly among the members of the genus *Centaurea* (Colas *et al.,* 1997; Pisanu *et al.,* 2009). Flowering period ranges from June to August (Izco *et al.,* 2003). Besides, germination success seems to be very low (Gómez-Orellana Rodríguez, 2004; R. Retuerto pers. comm.; but see Izco *et al.,* 2003 for other estimates) and insect larvae are commonly found feeding on ripe fruits within mature flower heads (Fernández Casas and Susanna, 1986). The fruit lacks a pappus and, as in many *Centaurea* species, the presence of an elaiosome suggests that ants may play a role in seed dispersal. *C. borjae* produces rhizomes up to several meters long that can give rise to new rosette leaves. Rhizomes also serve as a belowground bud bank:

the plant is a poor competitor that gradually disappears as the surrounding plant community matures but rosette leaves readily resprout from dormant rhizomes if a disturbance destroys the surrounding community (Izco *et al.,* 2003).

Fig. 1: *Centaurea borjae* Basal rosette with flower (left) and typical habitat of *C. borjae* (right).

Centaurea borjae is catalogued by the IUCN as "endangered" (Gómez-Orellana Rodríguez, 2011) and listed as priority species by the "Habitats" Directive (92/43/EEC, Annex II). Additionally, the habitat occupied by this species is considered as a Site of Community Importance (SCI) within the Natura 2000 network of protected sites. Yet, and despite its status as priority species, there are no data on the magnitude and structure of the genetic diversity of *C. borjae*. Its life-history traits may lead to contradictory hypothesis about its genetic variation. Thus, the occurrence of clonal propagation together with the low germination success has led to the hypothesis that populations are made up by ramets originating from a few genets, with a negative impact on the magnitude of population-level genetic diversity (Izco *et al.,* 2003). Alternatively, self-incompatible outcrossers often display considerable levels of genetic variation (Cole, 2003; Hamrick and Godt, 1996; Nybom, 2004) and polyploids generally maintain higher levels of genetic diversity in small populations than do diploids with comparable population sizes (Soltis and Soltis, 2000). On the other hand the occurrence of fruits without a pappus and the probable myrmecochory indicate that seed dispersal could be restricted to relatively short distances (Cousens *et al.,* 2008; Gomez and Espadaler, 1998). Likewise, animal-pollinated plants can experience

limited gene flow depending on the behavior of the animal disperser (Ghazoul, 2005), leading to significant genetic differentiation at smaller spatial scales.

Knowledge of the genetic diversity and structure of endemic species is a prerequisite to formulate scientifically rational conservation programs, both *in situ* and *ex situ* (Frankham *et al.,* 2002). The genetic management of endangered wild populations often involves defining management units (Crandall *et al.,* 2000; Moritz, 1994) as well as actions intended to minimize the risk of extinction, e.g. rescue of small inbreed populations, management of fragmented populations (Aguilar *et al.,* 2008; Frankham, 2010; Tallmon *et al.,* 2004). The patterns of genetic diversity between populations can also be used to detect loci under selection, improving our knowledge of the species biology (Excoffier *et al.,* 2009; Frankham, 2010). Likewise, *ex situ* efforts in plants typically involve germplasm (mostly seeds) storage where a common issue is to attain a sampling regime that may encompass the full genetic diversity of the species and its local populations (Frankel *et al.,* 1995). However, an important limitation when studying rare and/or endemic plants is the need to obtain molecular markers for an organism with none or very scarce previous sequence information. In this regard, amplified fragment length polymorphisms (AFLP) are among the molecular markers most commonly used in plants (Mba and Tohme, 2005; Palacios *et al.,* 1999) and they have proven particularly useful in the study of rare and/or threatened species (e.g. Barnaud and Houliston, 2010; Kim *et al.,* 2005; Li *et al.,* 2008; Peters *et al.,* 2009; Stefenon *et al.,* 2008; Winfield *et al.,* 1998; Yan *et al.,* 2009). Compared to co-dominant markers (e.g. SSRs), AFLP do not allow detecting heterozygotes. However, the same limitation affects to co-dominant markers when dealing with polyploids (Bruvo *et al.,* 2004; Obbard *et al.,* 2006). In fact, banding patterns of polyploid organisms, whether obtained with co-dominant or with dominant markers, may not express individuals' genotypes and should be considered only as phenotypes (Kosman and Leonard, 2005).

In the present study, we used AFLP phenotypes to investigate the genetic variation and population structure of *Centaurea borjae* to obtain information that may contribute to a better management and conservation of this protected narrow endemic. We focused in the following questions: 1) how does clonal reproduction contribute to population sizes?; 2) do populations show signs of diminished genetic variation?; 3) what is the minimum inter-plant distance for appropriate germplasm collection?; 4) are populations significantly differentiated from each other and, if so, is it possible to delineate management units?

MATERIALS AND METHODS

Sample collection and DNA extraction

Our sampling scheme covered the entire distribution range of the species and included the only six known sites of *Centaurea borjae* (Izco *et al.,* 2003). Three sites were located on serpentine substrata, one on gneiss substrata, one on amphibolites soil, and one on a relatively isolated site with granitoid soil (see Fig. 2 in results). Rosette leaves were taken as putative individuals. Sampling covered the whole area occupied by the species at each site (see Table 1 for maximum inter-rosette distances at each site). Since *Centaurea borjae* displays an aggregated distribution, we followed a stratified design with 2-4 rosettes sampled per aggregation. Leaves were dried in silica gel and stored at -20°C until DNA extraction. DNA was extracted using the Wizard Magnetic Kit (Promega) according to the manufacturer's instructions. The quality of extracted DNA and negative controls were checked on 1.5% agarose gels.

AFLP analyses

As AFLP performance can be sensitive to reaction conditions (Bonin *et al.,* 2004), we used several control measures to guarantee the reproducibility of our AFLP fingerprints. First, selective primer combinations were chosen after screening twenty-

four pairs of primers with three selective bases on 20 individuals (3-4 individuals per sampling site). The whole procedure was repeated with new, independent DNA extractions of the same individuals to check for reproducibility. Four primer combinations generating reproducible, easily scorable profiles were chosen (*EcoRI*/*TruI*: TAG/CAT, TAG/CAG, TAG/CAC, TAC/CAA). Second, replicate DNA extractions were obtained for a new set of approximately 10% of the total number of individuals (evenly distributed among the 6 sampling sites) and run in parallel with the other DNA samples to monitor reproducibility. Samples and replicates were run in a blind-manner to avoid any bias during scoring. Individuals from each sampling site were evenly partitioned between the various 96-well plates used for PCR; replicates and originals were always run in separate plates; samples and replicates were randomly distributed within plates. Third, each batch of DNA extractions (24 samples) included a negative control with no sample added that went through the entire genotyping procedure (DNA extraction included). The estimated genotyping error (1.5%) was consistent with results of reproducibility tests conducted for AFLP both in plants and animals (Bonin *et al.,* 2004); none of the individual loci exceeded the maximum acceptable error rate (10%) recommended by Bonin *et al.* (2007).

AFLP analyses were performed according to Vos *et al.* (1995) with minor modifications and using nonradioactive fluorescent dye-labelled primers. Approximately 250 ng of genomic DNA were digested at 37°C for 3 hours in a final volume of 20 µl with 1.25 units of *EcoRI* and *TruI* (Fermentas) and 2x Tango Buffer (Fermentas). Digested DNA was ligated for 3 hours at 37ºC to double-stranded adapters (50 pmols of adaptors E, 5'-CTCGTAGACTGCGTACC-3' and 5'-AATTGGTACGCAGTCTAC-3', and M, 5'-GACGATGAGTCCTGAG-3' and 5'-TACTCAGGACTCAT-3') using 0.5 units of T4 DNA ligase (Fermentas). Then, 2 µl of the ligation product was pre-amplified with 0.3 µM of each single selective primer (*EcoRI*-T and *TruI*-C), 2.5 mM $MgCl_2$, PCR buffer 1x (Applied Biosystems), 0.8 µM dNTPs, 0.04

µg/µl BSA, 1M Betaine and 0.4 units of Taq polymerase (Applied Biosystems) in a final volume of 20 µl. Amplification conditions were 2 min at 72°C; 2 min at 94°C; 20 cycles of 30 s at 94 °C, 30 s at 56°C, and 2 min at 72 °C; and a final extension of 30 min at 60°C. Pre-amplification fragments were diluted 1:5 with Milli-Q water; 2.5 µl of the resulting solution were selectively amplified using 0.6 µM of the selective primers, 0.8 µM dNTPS, 2.5 mM $MgCl_2$, 0.04 µg/µl BSA, PCR Buffer 1x (Applied Biosystems) and 0.4 units of AmpliTaq Gold polymerase (Applied Biosystems) in a final volume of 10 µl. Selective amplification was performed as follows: 4 min at 95°C; 12 of cycles of 30 s at 94°C, 30 s at 65ºC (first cycle, then decreasing 0.7°C for each of the last 11 cycles), and 2 min at 72°C; 29 cycles of 30 s at 94ºC, 30 s at 56ºC, and 2 min at 72ºC; and a final extension of 30 min at 72°C. Digestion, ligation, and PCR reactions were performed in a PxE thermal cycler (Thermo Fisher Scientific Inc., Waltham, MA, USA). Selective amplification products were electrophoresed on an ABI 3130xl automated DNA (Applied Biosystems) sequencer with HD-500 as size standard (Applied Biosystems). Fragments from 70 to 400 bp were manually scored for presence/absence at each selected locus with the help of GeneMarker v.1.70 (SoftGenetics LLC, State College, PA, USA) following common recommendations (Bonin *et al.,* 2005). Scores of the 4 primer combinations were assembled into a single binary data matrix.

Data analysis

For the purposes of our data analyses, individuals collected from each sampling site were regarded as a putative population. Data analyses followed a phenotypic ("band-based") approach as it is often the case in studies that deal with polyploids or that combine various levels of ploidy (Abbott *et al.,* 2007; Andreakis *et al.,* 2009; Bonin *et al.,* 2007; Garcia-Verdugo *et al.,* 2009; Kosman and Leonard, 2005; Obbard *et al.,* 2006). Genetic diversity for each population as well as for the complete data set was estimated in GenAlex 6.41 (Peakall and Smouse, 2006) as the percentage

of polymorphic bands (5% criterion), the Shannon-Weaver Index of phenotypic diversity (H_{SW}), and the average dissimilarity (simple-matching coefficient) between pairs of individuals (H_{Phen}) (equivalent to Nei's gene diversity calculated from band frequencies, Kosman 2003). These estimates were supplemented with measurements of genotypic diversity based on the frequency of distinct multi-locus genotypes. To this aim, potential clones, i.e. individuals with identical banding pattern, were identified with the help of the program GenoType (Meirmans and Van Tienderen, 2004). As rates of somatic mutations are difficult to obtain for natural populations (Douhovnikoff and Dodd, 2003), the threshold value for genotype detection (i.e. maximum distance between two individuals at which they are still assigned to the same genotype) equaled the genotyping error rate estimated in our reproducibility tests (1.5%). Individuals with missing values for any loci were excluded from the genotype assignment. Genotypic diversity was estimated with the help of GenoDive (Meirmans and Van Tienderen, 2004) as number of genotypes (G), proportion of distinguishable genotypes, (G/N, where N is the number of individuals), effective number of genotypes ($G_{eff}=1/\sum p_i^2$, where p_i is the frequency of each i genotype), and evenness of genotypes (Eve = G_{eff}/G).

To detect possible loci under selection, and in order to minimize the possibility of false-positives, three different approaches were used. First, loci under selection were searched with the Bayesian method described in Beaumont and Balding (2004) and implemented in the software Bayescan (Foll and Gaggiotti, 2008). Bayescan estimates population-specific F_{ST} coefficients and uses a cut-off based on the mode of the posterior distribution to detect loci under selection (Foll and Gaggiotti, 2008). Bayescan was run by setting a sample size of 10000 and a thinning interval of 50 as suggested by Foll and Gaggiotti (2008), resulting in a total chain length of 550000 iterations. Loci with a posterior probability over 0.99 were retained as outliers, which corresponds to a Bayes Factor >2 (i.e. “decisive selection” (Foll and Gaggiotti, 2006))

and provides substantial support for accepting the model. Second, loci under selection were also identified using the approach of Beaumont and Nichols (1996) implemented in Mcheza (Antao and Beaumont 2011). Mcheza uses coalescent simulations to generate a null distribution of F_{ST} values based on an infinite island model for the populations; loci with an unusual high or low F_{ST} are regarded as under directional or stabilizing selection, respectively. Runs were performed with the infinite allele mutation model and the significance of the neutral distribution of F_{ST} was tested with 100000 simulations at a significance value P of 0.001. The multitest correction on false discovery rates (FDR) was set to 1% false positive to avoid overestimating the percentage of outliers. Finally, the Spatial Analysis Method (SAM) described by (Joost *et al.,* 2007) was used to investigate the relation between loci under selection and soil type. Unlike the previous procedures, SAM does not require defining the populations. It identifies alleles associated with environmental variables by calculating logistic regressions between all possible marker-environmental pairs and by comparing if a model including an environmental variable is more informative than a model including only the constant. In SAM, soil type was converted into a semi-quantitative scale following differences in the mineral composition (SiO_2 content) of parental rocks: granitic soil was scored as 1, gneisses and amphibolite soils as 2, and serpentine soil as 3. We followed a restrictive approach and a model was significant only if both G and Wald Beta 1 tests rejected the null hypothesis with a significance threshold set to 95% (P <0.00017 after Bonferroni correction). Bayescan, Mcheza and SAM were used under a conservative approach and the analyses were restricted to loci with a dominant allele frequency between 5% and 95%. This restriction decreases the probability that differentiation at a given locus would be incorrectly identified as a signature of selection just because it stood against low levels of background genetic variation resulting from the inclusion of low-polymorphism markers.

The presence of genetic structure was tested using a combination of individual-based and population-based approaches. First, pairwise simple-matching dissimilarities between individuals were visualized using Principal Coordinates Analysis (PCoA) as in Kloda *et al.* (2008). Second, the partitioning of the genetic diversity was evaluated by molecular variance analysis (AMOVA) (Excoffier *et al.,* 1992). Its significance was tested by 9999 random permutations of individuals among populations; the genetic variation apportioned to differences among populations was expressed as Φ_{PT}, an analogue of F_{ST}. Both AMOVA and PCoA were performed in GenAlex 6.41 (Peakall and Smouse, 2006). Third, the correlation between genetic and geographic distance between populations was tested for significance with a Mantel test as implemented in the Isolation by Distance Web Service 3.15 (Jensen *et al.* 2005) using 10 000 bootstrap randomizations. Finally, the network structure and genetic connectivity among populations was assessed with a network analysis based on graph theory that has proved useful in population genetics and landscape ecology (Dyer and Nason, 2004; Garroway *et al.,* 2008). The graph represents a landscape of discrete habitat patches as a set of nodes (populations) genetically interconnected by edges (gene flow) (Minor and Urban, 2007). The presence of an edge is determined by the genetic covariance of the connected populations; independent populations are shown unconnected. Networks were constructed with the online application Populations Graphs v2 (http://dyerlab.bio.vcu.edu/software/) and the analyses were carried out in the software Genetic Studio (Dyer, 2009). For graph construction, we retained the minimal edge set that sufficiently described the total among-population covariance structure; two populations shared an edge when there was significant covariance between them after removing the covariance that each population had with all the remaining populations. Significance was tested using edge exclusion deviance which identified the most important edges for each node in terms of genetic covariance. Extended and compressed edges were determined by regressing geographic and graph distances (Dyer, 2009). Graph distance was estimated as the minimal

topological distance connecting pairs of nodes. In a homogeneous IBD process, graph and geographical distances should be proportional. Alternatively, long distance migration can result in extended edges, i.e. relatively small graph distances between spatially distant populations, while high graph distances between spatially close populations are compressed edges revealing restricted migration (Dyer *et al.,* 2010).

The pattern of genetic differentiation was further investigated with individual-based Bayesian approaches. The option for spatial clustering of individuals implemented in BAPS 5.3 (Corander *et al.,* 2008) was run 3 times for each of K = 2–20 and the optimal partition determined by the program was used to estimate the levels of genetic admixture of individuals (with 200 reference individuals simulated for each genetic group and each original individual analyzed 20 times). The data was analyzed with an alternative Bayesian approach as implemented in Structure v.2.3.3 (Falush *et al.,* 2003; Hubisz *et al.,* 2009; Pritchard *et al.,* 2000). Structure was run assuming correlated allele frequencies. Ten runs with a burn-in period of 100 000 replications and a run length of 1 000 000 Markov chain Monte Carlo (MCMC) iterations were performed for a number of clusters ranging from K = 1 to 10. The value of K that captured most of the structure in our data was determined using the approach originally proposed by Pritchard *et al.* (2000) with further guidance derived from the procedure of Evanno *et al.* (2005) based on the rate of change of the estimated likelihood between successive K values. Runs of the selected K were averaged with the Clummp version 1.1.1 (Jakobsson and Rosenberg, 2007) using the LargeKGreedy algorithm and the G' pairwise matrix similarity statistics.

To investigate the fine-scale spatial genetic structure (fine scale SGS), the location of each individual sample was carefully recorded in three sites covering the whole range of the species (PR, PC, LI). The kinship coefficients between pairs of individuals (F_L) within each site were calculated following Loiselle *et al.* (1995). The hypothesis that there was significant SGS was tested by comparing the observed

regression slope of F_L on the logarithm of pairwise geographic distances, *b*, with those obtained after 10 000 random permutations of individuals among locations. Tests were conducted for each individual site as well as for the pooled data set. Standard errors for *b* were calculated by jackknifing over loci and used to test for significant differences among slopes. SGS was then quantified by an *Sp* statistic that represents the rate of decrease of F_L with distance (Vekemans and Hardy, 2004); *Sp* was calculated as $-b/[1-F_{(1)}]$, where $F_{(1)}$ is the average kinship coefficient between neighboring individuals. However, this approach assumes a linear relationship between F_L and ln of distance. Therefore, the SGS was visualized by plotting mean F_L estimates over pairs of individuals in a given distance interval against distance; the extent of the linear relationship was determined as the distance at which mean F_L showed no obvious trend. Estimates of *b* and *Sp* were restricted to these maximum distances and computed with the help of SPAGEDI (Hardy and Vekemans, 2002).

RESULTS

Genetic diversity measures

A total of 129 markers were scored in 180 individuals. Fifty-nine (45.7%) loci were segregating for the complete dataset and were retained for diversity estimates. Only one private band was detected in the geographically isolated PR. The estimates of total genetic diversity for the species (H_{Phen} = 0.258; H_{SW} = 0.413) were slighter above most of the values for single populations (Table 1). The three indices of genetic diversity were correlated across populations. OB exhibited the highest genetic diversity (86.4% polymorphic loci, H_{Phen} = 0.280; H_{SW} = 0.435) with values 20-25% higher than the estimates obtained at VH, the population with the lowest values for most indices (64.4% polymorphic loci, H_{Phen} = 0.192; H_{SW} = 0.309). The remaining four populations produced very similar estimates (69.5-74.6% polymorphic loci, H_{Phen} =

0.217-0.224; H_{SW} = 0.348-0.360), intermediate between OB and VH but slightly closer to the values observed in VH.

The 175 individuals used for genotype assignment (5 individuals were excluded due to the presence of missing values at some loci) produced 154 distinct genotypes (G_{eff} = 125, *G/N* = 0.880). Potential clone mates always occurred in the same population, often spatially close to each other. The presence and relative abundance of potential clone mates (i.e., genotypic diversity) depicted an arrangement of genetic diversity somewhat different from the image derived from non-genotypic indices. Again, OB produced the highest estimates (*G* = 29, G_{eff} = 28.1, *G/N* = 0.967) and VH produced the lowest (*G* = 21, G_{eff} = 15.0, *G/N* = 0.700). However, Table 1 shows the occurrence of two groups of populations with very different levels of diversity. Most of the individuals sampled in the three southernmost populations (OB, PC, and the geographically remote PR) had distinct genotypes, while 24-30% of the rosettes sampled in the three northernmost ones (OBB, VH, LI) were potential clone mates with identical AFLP banding patterns. As a results, the various estimates of genotypic diversity were clearly higher in southernmost populations (*G* = 25- 29, G_{eff} = 25.1-28.1, *G/N* = 0.961-0.967) than in northernmost ones (*G* = 21-26, G_{eff} = 15.0-20.5, *G/N* = 0.700-0.862). The index of evenness indicates that a few genotypes were repeatedly found in a considerable fraction of the individuals sampled in these northernmost sites.

Table 1. *Centaurea borjae.* Genetic characteristics of each sampling location based on 59 segregating loci.

		Band-based					Genotypic					
Pop	D_{max}	*N*	*PL*	*PB*	H_{Phen}	H_{SW} (±SE)	*N*	*G*	G_{eff}	*G/N*	*Eve*	*Sp*
LI	200	32	43 (72.9)	0	0.223	0.354 ±0.029	31	26	20.5	0.839	0.79	0.400
VH	320	30	38 (64.4)	0	0.192	0.309 ±0.030	30	21	15.0	0.700	0.71	N/A
OBB	240	29	44 (74.6)	0	0.224	0.360 ±0.027	29	25	17.2	0.862	0.69	N/A
OB	191	30	51 (86.4)	0	0.280	0.435 ±0.025	30	29	28.1	0.967	0.97	N/A
PC	600	30	41 (69.5)	0	0.217	0.348 ±0.028	29	28	27.1	0.965	0.97	0.132
PR	260	29	41 (69.5)	1	0.217	0.349 ±0.028	26	25	25.1	0.961	0.97	0.088
Total		180	59 (45.7)		0.258	0.413 ±0.022	175	154	125.0	0.880	0.81	0.185

LI, O Limo; VH, Vixia Herbeira; OBB, O Bico2; OB, O Bico; PC, Punta Candieira; Pr, Prior. D_{max} = maximum distance (in m) between samples, *N*, number of individuals; *PL*, number (and percentage) of polymorphic loci (5% criterion) (percentage for the total data set based on 129 scorable loci); *PB*, number of private bands; H_{Phen}, average simple-matching dissimilarity between pairs of individuals (equivalent to Nei's gene diversity for band frequencies); H_{SW}, Shannon-Weaver Index of phenotypic diversity; *G*, number of distinct genotypes; G_{eff}, effective number of genotypes; *Eve*, evenness; *Sp, Sp* statistic of autocorrelation (Vekemans and Hardy 2004).

Identification of possible loci under selection

Of the 129 reproducible AFLP loci, 59 had dominant allele frequencies ranging 5% to 95% and were included in outlier analyses (Table 2). Together, the three outlier detection approaches identified six loci as potentially under selection but only locus 31 was consistently detected as an outlier by the three procedures. In Bayescan, the six-population analysis identified two loci under selection: one under "very strong" selection $\log_{10}$BF>1.5 and another under "decisive" selection $\log_{10}$BF>2. Using the model of infinite alleles at a significance P value of 0.001, Mcheza only identified one locus under directional selection that coincided with the marker considered under "very strong" selection by Bayescan. After calculating logistic regressions between all possible marker-environmental pairs and with a significance threshold set to 95% after Bonferroni correction, SAM detected 5 loci associated with soil type. Again, this set of loci included locus 31 detected by both Mcheza and Bayescan.

Table 2. Detection of possible loci under selection.

	SAM		BAYESCAN	MCHEZA
	P value for G	*P* value for Wald Beta 1	log10(BF)	*P*(Simul F_{ST}<sample F_{ST})
Locus11	**2.98E-07**	**1.08E-06**	0.476	0.9852
Locus20	**1.37E-06**	**1.18E-05**	-0.104	0.8112
Locus23	**0.000117**	**0.000109**	-0.183	0.7520
Locus31	**5.55E-16**	**0.499992**	**1.8770**	**0.9992**
Locus38	**0.000167**	**0.000363**	-0.0885	0.6871
Locus41	0.196413	0.098697	**2.1280**	0.9621

Numbers in bold are loci detected as potentially under selection by SAM (*P* values for G and Wald Beta 1 with a significance threshold set to 95% corresponding to *P* <0.00017 after Bonferroni correction), BayeScan ($\log_{10}$(BF)>1.5 corresponding to "very strong selection"), and MCHEZA (*P* <0.001).

Since none of the six loci detected as outliers seemed linked to serpentine soil no obvious differences between serpentine LI, VH, OBB and non-serpentine sites OB, PC, PR were found. Instead, our results reveal that site PR had the largest influence on the detection of outlier loci. PR displayed a distinctive genetic composition for most of the loci detected by SAM (Table 3). Interestingly, locus 31 was private to PR. Similarly, PR also produced the highest (loci 11 and 38) or the lowest (loci 20 and 23) estimates for the frequency of the dominant allele.

Table 3. Population relative frequency of the dominant allele (as %) for six outlier loci.

	LI	VH	OBB	OB	PC	PR
Locus 11	**29.0**	**6.7**	**21.4**	40.0	10.3	78.3
Locus 20	**70.8**	**83.3**	**42.7**	50.0	44.8	13.0
Locus 23	**58.1**	**90.0**	**78.6**	66.7	62.1	30.4
Locus 31*	**0.0**	**0.0**	**0.0**	0.0	0.0	60.8
Locus 38	**51.6**	**50.0**	**28.6**	50.0	79.3	82.6
Locus 41	**83.9**	**36.7**	**35.7**	13.3	55.2	60.8

Numbers in bold are sites with serpentine soil. * indicates the locus detected as under selection by the three approaches

Population structure

AMOVA revealed that most 88% of the genetic variation occurred within populations (Table 4). Still, population differentiation was highly significant Φ_{PT}=0.119, $P < 0.0001$. The exclusion of PR from the dataset had minimal impact on the genetic differentiation, and Φ_{PT}=0.104 continued to be highly significant $P < 0.0001$.

Table 4. Analysis of molecular variance (AMOVA) based on 59 segregating markers in *C. borjae*.

Source of variation	d.f.	MSD	Variance components	*P*-value	Φ_{PT}
All (6) populations					
Among populations	5	34.86	0.933 (12%)	< 0.0001	0.119
Within populations	174	6.88	6.880 (88%)		
Main range (5) populations					
Among populations	4	31.18	0.803 (10%)	< 0.0001	0.104
Within populations	146	6.92	6.927 (90%)		

Separate analyses were carried out for the complete data set (6 populations) and for the subset of sites from the main range of the species (excluding the geographically isolated PR). *P*-values based on 9999 permutations. d.f. =degrees of freedom, MSD = mean squared deviations.

All pairwise Φ_{PT} were also significant $P < 0.05$ after Bonferroni correction for multiple testing. Even the comparison between the geographically close OB and OBB separated by 0.8 km was significant Φ_{PT}= 0.037. The highest level of differentiation occurred between VH and PR Φ_{PT} = 0.222. PR also yielded the highest Φ_{PT} values when compared to any of the other populations from Φ_{PT} = 0.114 for PR-PC to Φ_{PT} = 0.154 for PR-OBB. The Mantel test provided only weak evidence that genetic and geographic distances correlated along the species range. The moderately significant Mantel correlation was largely dependent on the inclusion of PR, the geographically isolated population, in the data set r = 0.1946, Mantel P = 0.036. Without PR, the correlation became non-significant.

The network generated by the 59 polymorphic loci only contained 10 out of the 15 possible edges indicating that the genetic covariance between populations was limited (Fig. 3). The network was largely consistent with an IBD pattern as 7 out of the 10 edges were proportional to geographical distance. PR, LI, and OBB produced the largest number of connections 4 edges each while OB, PC, and VH were less connected in genetic terms 3, 2, and 2 edges, respectively. Many edges involved geographically adjacent sampling sites; only PR, and to a lesser extent LI, showed connections with spatially distant populations but their edges were mostly proportional to geographical distance. VH was linked only by compressed edges highlighting its genetic isolation despite the geographical placement between OBB and LI.

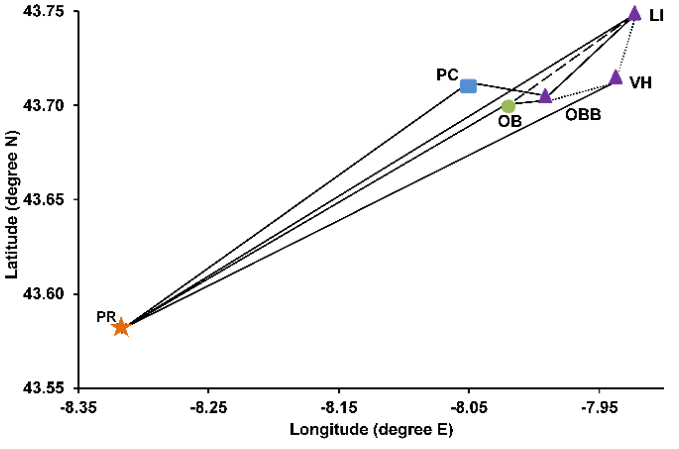


Fig.3. Genetic network of *C. borjae* created with 59 polymorphic loci. Site symbols indicate soil type: triangle, serpentine; circle, gneisses; solid square, amphibolites; star, granitoid. Populations connected by lines exhibit significant conditional genetic covariance. Solid lines indicate genetic distances proportional to spatial distances. Dotted lines ----- are compressed edges with relatively higher conditional genetic distance in respect to spatial separation, whereas dashed lines - - - - denote extended edges with small conditional genetic distance in respect to spatial separation. When necessary, coordinates for some populations have been slightly modified to avoid excessive line overlap.

Individual-based analyses produced results largely consistent with those obtained from population-based approaches. Confirming that most of the genetic variation occurs within populations, the PCoA plot 47% of the variation explained by the first two axes, Fig. 4 showed considerable overlap between the individuals collected at the 6 sites. However, the graph also revealed that the individuals from VH and PR formed two discrete groups with limited overlap.

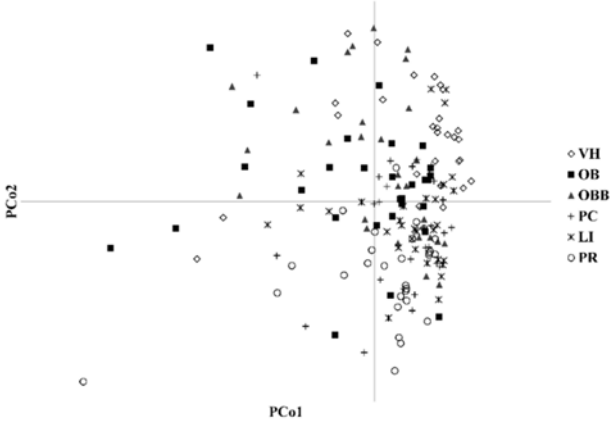


Fig. 4. Principal Coordinates Analysis PCoA of pairwise simple-matching dissimilarities between individuals of *C. borjae*. PCo1 and PCo2 explain 47% of total variation.

With AFLP markers treated as phenotypes, BAPS identified 9 genetic groups as the optimal partition log-likelihood value = -4332.5, probability for 9 clusters = 0.9996 although 2 out of the 9 genetic clusters consisted of one single individual each. Genetic admixture was generally low and most individuals 98% were assigned to a single cluster. The admixture clustering graph (Fig. 2) shows that the six populations can be divided into 4 groups according to their genetic lineage. Again, PR and VH consisted mainly of individuals assigned to one genetic group different in each sampling site while PC, OB, and OBB formed a larger group that was consistent with the overlap seen in the PCoA. One single genetic cluster dominated in these three populations 74% of the rosettes, although two other clusters also attained some

representation 15% and 7%, respectively. The plants collected in LI were evenly partitioned among 4 genetic clusters: two lineages 52% individuals were unique to LI while the other two were those also common in PR 23% and PC-OB-OBB 26%. Results from Structure corroborated the signal detected by BAPS. Log-likelihood values reached a plateau beyond *K* = 7, suggesting that a model with seven genetic clusters captured most of the structure in the data Pritchard *et al.* 2000. The method of Evanno *et al.* 2005 confirmed that the highest rate of change in the log probability of the data occurred both at *K* = 2 ΔK=108 and *K* = 7 ΔK=50. The partition for *K* = 2 seemed biologically meaningless. By contrast, clustering for *K* = 7 resembled the partition obtained with BAPS figure not shown but with a higher degree of admixture Dirichlet parameter α = 0.073.

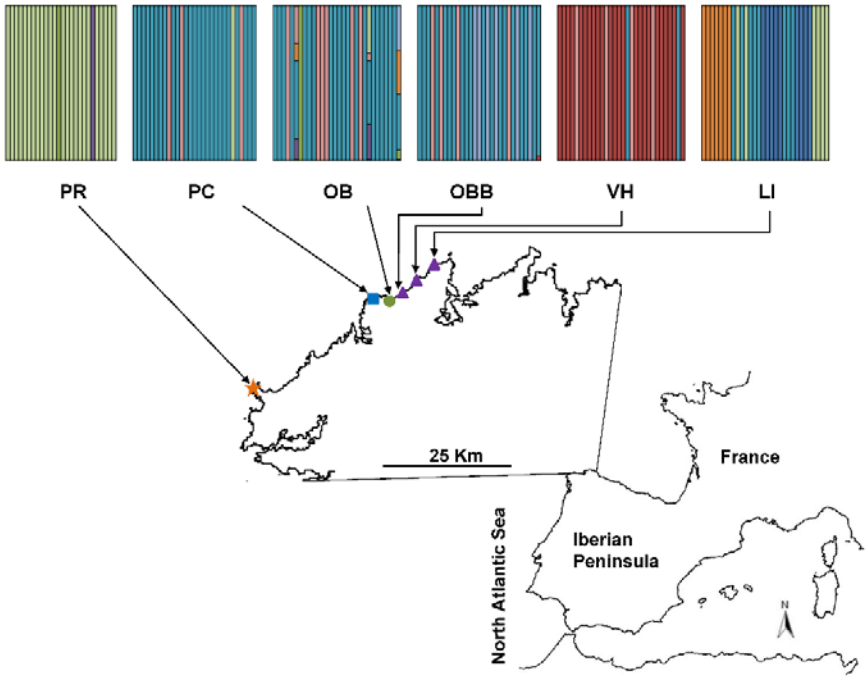


Fig. 2. Sites sampled in this study and population structure according to BAPS. Range occupancy is strongly fragmented into very small enclaves. Site symbols indicate soil type: triangle, serpentine; circle, gneisses; solid square, amphibolites; star, granitoid. The histogram shows the results of individual assignment by the admixture analysis performed for an optimal number of 9 genetic clusters *P* = 0.9996. Each vertical bar corresponds to one individual with patterns indicating the probability of assignment to each cluster.

Fine-scale spatial genetic structure

Average kinship coefficient decreased steadily until some distance in the three sites investigated for SGS (Fig. 5). Beyond that point, the relationship between the kinship coefficient and distance either experienced a rapid reduction in slope or disappeared. The distance for the change in slope varied among sites: 80 m in LI, 40 m in PC, and 35 m in PR. Calculations of *b* and the *Sp* statistic were restricted to these maximum distances to avoid any bias derived from this nonlinearity.

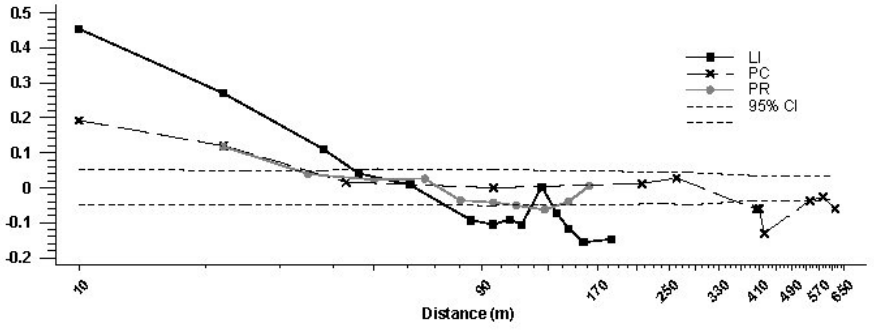


Fig. 5. Correlograms showing the mean kinship coefficient F_L as a function of distance for LI black solid squares, PC crosses, and PR grey solid circles clonal ramets included. Dotted lines are the 95% confidence belt for the null hypothesis of no spatial genetic structure determined by 10 000 permutations.

Slope *b* was always significant supporting the occurrence of SGS in the three sites and in the pooled data set P< 0.05. Slope comparison revealed significant differences among sites. The kinship coefficient fell more sharply with distance in LI *b*= -0.211 than in PC or PR -0.110 and -0.080, respectively; $P < 0.05$ for the comparison between LI and either PC or PR; the slopes of the latter two sites were statistically indistinguishable $P > 0.05$. One might suspect that the sharper slope of LI could be an artifact of a higher frequency of clonal ramets. In LI clone mates were detected separated as far as 20 m with an average clone distance of 8 m while in PC and PR distance among clone mates was 1 m one single pair per site. However, the exclusion of clonal replicates had a slight, non-significant impact in the estimate of *b* = -0.190; *P*

= 0.28 for the comparison of *b* estimated with and without clones. The variation in SGS among sites was further corroborated by the *Sp* statistic. Moreover, compared to *b, Sp* amplified the differences between sites as its value was three to four times higher in LI than in PC or PR Table 1. This change of magnitude resulted from the fact that LI simultaneously produced the lowest *b* and the highest F_1 = 0.473 estimates for the two values used to calculate *Sp* F_1 was 0.171 and 0.098 in PC and PR, respectively. Again, clone removal had little impact in *Sp* for LI *Sp* = 0.329, F_1 = 0.424.

DISCUSSION

Centaurea borjae shows a total occupancy typical of a narrow endemic (< 5 000 m2) arranged into a strongly fragmented distribution (Bañares *et al.*, 2004; Valdes-Bermejo and Rivas Goday, 1978). As a result, this plant is catalogued as endangered by national and supranational organisms (e.g. Gómez-Orellana Rodrígue,z 2011; Ministerio de Medio Ambiente y Medio Rural y Marino, 2011). According to the IUCN red list, major threats to its survival are a poor reproductive strategy together with the lack of appropriate habitat while other threats include livestock (trampling, predation) and tourism (trampling, anthropization) (Gómez-Orellana Rodríguez, 2011). Despite its conservation status, *C. borjae* has received little attention. In particular, its genetic variation has been totally overlooked. This gap in our knowledge can be filled using neutral markers such as AFLP. Although there is growing evidence that the correlation between neutral and adaptive variation might not be very high, a high neutral variation may indicate the potential for significant adaptive variation (Reed and Frankham, 2003).

How does clonal reproduction contribute to population sizes? A main concern for the long-term preservation of *Centaurea borjae* derived from the suspicion that its populations might be formed by a few genets with numerous ramets (Izco *et al.*, 2003). Clonal self-incompatible species have been reported to display lower genotypic

diversities than self-compatible ones (Honnay and Jacquemyn, 2008) and rare/narrow endemic plants with small populations seem to be more clonal than more widespread ones (Silvertown, 2008).

Our results confirm that potential clone mates do occur in every population and reveal a clumped clonal structure (i.e. clone mates were detected spatially close to each other) typical of plants that clone by organs that are not easily dispersed such as underground rhizomes (Vallejo-Marin *et al*., 2010). However, the high G/N estimates calculated for most populations (range 0.700-0.967) reveal a comparatively low extent of clonality since average G/N values in studies of clonal plants often are <0.65 (Vallejo-Marin *et al.,* 2010). While acknowledging that our estimates are likely to overestimate the clonal diversity of *C. borjae* since our ramet sampling was not exhaustive, as it is often the case in most studies (Vallejo-Marin *et al.,* 2010), they still suggest that clonal growth in *C. borjae* might not have the very large impact anticipated from direct observations of vegetative propagation in the field.

We found a lower clonal diversity in the three northernmost populations. Large differences in clonal diversity among populations of individual species seem common in plants (see Arnaud-Haond *et al.,* 2007 and references therein) and previous literature surveys have found that the frequency of clonality increases with population age or with increasing latitude (Silvertown, 2008). However, and to the best of our knowledge, geological substratum is the only consistent difference between our two sets of populations: serpentinites in the 3 northernmost sites; gneisses, amphibolites, and granitoids in the other 3 ones. Since serpentine soils are characterized by high levels of toxic heavy metals (Cr, Ni, Co) that may affect plant growth, it might be suggested that the conditions created by the serpentine soil may, at least partly, favor clonal propagation in *C. borjae*. In this regard, previous experimental studies have shown that clonal plants ameliorate the stressful effects of

serpentine soils through physiological integration among connected ramets (Roiloa and Retuerto, 2006).

None of the six loci detected as outliers in our analyses seemed linked to serpentine soil. Instead, the detection of outlier loci was largely influenced by the presence of one single population (PR). Given the peculiarities of outlier detection procedures (Excoffier *et al*., 2009; Foll and Gaggiotti, 2008), the influential role played by PR possibly derives from its geographic isolation. Moreover, even the locus that was simultaneously detected as an outlier by the three procedures must probably be regarded as an artifact of our sampling design (for further discussion on this topic see Supplementary Material S1).

Do populations show signs of diminished genetic variation? No evidences of genetic impoverishment were detected in *Centaurea borjae*. Instead, our data revealed relatively high levels of genetic variation both at species and at population level. The percentage of polymorphic loci in *C. borjae* is comparable to estimates obtained in other outcrossing plants (Despres *et al.,* 2002; Kato *et al.,* 2011; Morden and Loeffler, 1999; Tero *et al.,* 2003; Vilatersana *et al.,* 2007). Genotypic diversity was likewise high and revealed a low percentage of clone mates in comparison with other clonal species (Arnaud-Haond *et al.,* 2007; Silvertown, 2008; Vallejo-Marin *et al.,* 2010). Also, our AFLP-derived estimates of HPhen compare well with values obtained using dominant markers in other perennial outcrossers with mid successional status (Nybom, 2004). Allogamous perennials, particularly when long-lived, often yield the highest mean levels of within-population diversity in plant studies (Nybom, 2004). In this regard, the diversity recorded within populations of *C. borjae* is in the mid to high end of the values typically found in plants studied with dominantly inherited markers.

Our estimates for *C. borjae* also fall within the range of values inferred for other endemic members of the genus *Centaurea* investigated with dominant markers:

Centaurea nivea (Sözen and Özaydin, 2009), *Centaurea wiedemanniana* (Sözen and Özaydin, 2010), or *Feminasia balearica* (formerly known as *Centaurea balearica*, Vilatersana *et al.,* 2007) (see Table S1). The latter are all diploids while polyploids like *C. borjae* are often expected to maintain higher levels of heterozygosity than their diploid counterparts (Soltis and Soltis, 2000). Still, Table S1 suggests that ploidy level exerts an uncertain influence on the estimates of genetic diversity obtained for other members of the genus. Table S1 also shows that while endemic *Centaurea* often display less genetic variation than their widespread counterparts, some endemic taxa reach levels of diversity equaling that of their widespread congeners as observed in other studies (Cole, 2003; Gitzendanner and Soltis, 2000). In fact, the differences between endemic and widespread *Centaurea* shown in Table S1 could be partially attributed to the different maker system used to investigate each type of taxa as many endemic *Centaurea* were studied with allozymes while most of the widespread taxa were investigated with microsatellites.

The retention of moderate-high levels of genetic diversity seems consistent with some features of *C. borjae*. Allogamous, insect-pollinated species like *C. borjae* often show higher genetic diversity than self-pollinated plants (Hamrick and Godt, 1996; Kim *et al.,* 2005; Takahashi *et al.,* 2011). Also, the presence of seed, bulb, or bud (*C. borjae*) banks is known to buffer plant populations against dramatic changes in genetic composition (see Ellstrand and Elam, 1993 and references therein). Likewise, endemic does not necessarily equate to rare. Some endemic/restricted species can be locally abundant and, consequently, less sensitive to the effects of genetic drift. In this regard, only rough estimates of local abundance are available for *C. borjae* (Bañares *et al*., 2004; Izco *et al*., 2003) but our observations suggest that local populations are made up of a few thousand rosette leaves that, given our G/N ratios, possibly represent comparably high numbers of genetically distinct individuals. Finally,

polyploids generally maintain higher levels of heterozygosity than their diploid progenitors (Soltis and Soltis, 2000).

What is the minimum inter-plant distance for appropriate germplasm collection? The fine-scale SGS found in *Centaurea borjae* indicates that rosette leaves at close distances can be more related than spatially random pairs. The values of the *Sp* statistic for *C. borjae* fit the higher end of the estimates compiled by Vekemans and Hardy (2004) for 47 plant species. Therefore, our results are in agreement with the strong SGS expected in species with low dispersal, clonal reproduction, and/or low density (Vekemans and Hardy, 2004). Albeit solid in statistical terms, *Sp* cannot be easily translated into guidelines for conservation. Likewise, the x-intercept in an autocorrelogram, another commonly used SGS parameter, has been severely criticized by its high sensitivity to sampling strategy (e.g. Zeng *et al*., 2010). In this regard, Vekemans and Hardy (2004) noted that there is one case where a critical distance, more useful for conservation purposes, can still be defined; if F_L decreases steadily until some distance x, showing no further trend, SGS can be said to occur until x. This seems to be the case in *C. borjae* where the extent of SGS deducted with this procedure would vary from 35-40 m in PR-PC to 80 m in LI. Therefore, as a general recommendation, efficient germplasm collection should avoid rosettes separated <80 m although distances as short as 35-40 m might be acceptable in southernmost sites. These distances will also prevent the collection of clone mates.

Are populations significantly differentiated from each other and, if so, is it possible to delineate management units? Several pieces of evidence suggest that dispersal and/or gene flow is restricted in *Centaurea borjae*. First, the moderate, but significant, among-population variability detected at population scale is consistent with a scenario of low gene flow, although any conclusion about gene flow based on Φ_{ST} estimates must be made with caution, particularly when dealing with wild populations that likely violate the model assumptions behind this statistic (Marko and

Hart, 2011; Whitlock and McCauley, 1999). Second, the fine-scale SGS detected in *C. borjae* is typical of plants with restricted dispersal and/or gene flow (Chung *et al*., 2005; Jump and Peñuelas, 2007; Sebbenn *et al.*, 2011). Finally, the network analysis also indicates restrictions to connectivity with only ten out of the fifteen possible edges present in the network and with the detection of some compressed edges connecting spatially close populations.

The trend for endemic species to be poor colonizers has received support in comparative studies with widespread congeners (Lavergne *et al.*, 2004) and seems consistent with unpublished evidence indicating that seed output and germination success is very low in *C. borjae* (R. Retuerto, pers. comm.) probably due to a high sterility of the achenes (Valdés-Bermejo and Agudo Mata, 1983). Limited dispersal also seems consistent with several life-history traits of *C. borjae*. Thus, although many pollinators can cross large distances in flight, animal-mediated pollen dispersal can be limited depending on the behavior of the animal disperser and/or the frequency and distribution of floral resources (Ghazoul, 2005). Likewise, the absence of a pappus and probable myrmecochory of *C. borjae* suggest that seed dispersal could be restricted to short distances (Cousens *et al*., 2008; Gomez and Espadaler, 1998). In this regard, evidence for low pollen flow rate among populations and very limited seed dispersal by ants has also been reported for *Centaurea corymbosa*¸ another endemic member of the genus *Centaurea* (Hardy *et al.,* 2004; Imbert, 2006). Likewise, heavy cypselas and restricted pollen dispersal were invoked as plausible causes for the very low levels of gene flow found in the related taxa *Feminiasia balearica* (Vilatersana *et al.*, 2007).

Our AFLP data consistently identified the set PC-OB-OBB as clearly differentiated from the other populations. Moreover, the individual-based analysis assigned most of the rosette leaves sampled in the PC-OB-OBB set to a genetic cluster that does not occur in PR, VH, or LI. Therefore, our data supports the designation of PR, VH, LI, and the PC-OB-OBB set as distinct MUs. Interestingly, genetic diversity and

differentiation in PR was comparable to the values estimated in other populations indicating that its geographical isolation did not have any obvious consequence on these genetic attributes. Still, most of the outlier loci detected in our analyses showed a different frequency of the dominant allele in this population. This included the only private marker found in our study, suggesting that PR may have separated long time ago (Vilatersana *et al.*, 2007). Alternatively, a portion of the rosette leaves sampled in PR share their genetic lineage with samples from LI, at the other end of the distribution range of the species, suggesting that both populations were connected in the past or episodes of long-distance dispersal.

It has been claimed that the very specific habitat of *Centaurea borjae* (thin, often ultrabasic, soils on sea cliffs) is in continuing decline due to human pressure and grazing (Gómez-Orellana Rodríguez, 2011). Yet, this claim is debatable. Excessive grazing and trampling, for example, are expected to have a negative impact on populations but moderate grazing of potential competitors possibly facilitates the persistence of *C. borjae* since this plant avoids areas with dense overlying vegetation. As for human pressure, the complete range of *C. borjae* falls within the Natura 2000 network (SCI ES1110002) implying that significant human developments require approval from environmental authorities. Moreover, the steep slope and harsh environmental conditions typical of the areas occupied by *C. borjae* provide an innate protection by rendering these sites unattractive and/or unsuitable to human activities. Alternatively, modeling efforts predict that the habitat suitable for *C. borjae* could disappear in the next 30 years due to global warming (Project PNACC; http://secad.unex.es/wiki/oeccpr). If so, *ex situ* conservation could be imperative and our results recommend that seed collection should avoid rosette leaves separated <80 m. Actually, no matter the immediate threats, *ex situ* conservation may seem

unavoidable if we recall that polyploids are regarded as evolutionary dead ends that experience higher extinction rates than diploids (Mayrose *et al.,* 2011).

In conclusion, *Centaurea borjae* showed no signs of decreased genetic variation. Even the frequency of potential clone mates was lower than anticipated, although we found some evidence that they might be more frequent in northernmost populations linked to serpentine soil. As in other outcrossing perennials, most of the genetic variation occurred within populations. Nonetheless, the significant genetic differentiation detected in our study suggests that population connectivity could be low while the fine-scale SGS reinforces the image of a plant with limited dispersal. The moderate genetic differentiation and similar genetic lineage deducted for three geographically close populations located at the center of the range suggests that they might be more closely related that the remaining populations. In situ conservation measures should consider these groups of populations as separate management units.

ACKNOWLEDGEMENTS

This research was supported by the project 07MDS031103PR Xunta de Galicia. We deeply appreciate the help of Maria Quintela with network, the outlier loci detection and the STRUCTURE analysis. We also thank three anonymous reviewers for insightful comments on an earlier version of the manuscript.

Table S1. Genetic diversity in species of the genus *Centaurea*.

Species	Range size	Pop.	Habitat and biological traits	Ploidy	Marker	Species/Pop diversity	References
C. borjae	Endemic	6	Perennial herb, no pappus, enthomoplilous outcrosser, insect pollinated, sea cliffs, asexual reproduction.	6x	AFLP	0.258/0.192-0.258 I=0.413/ I= 0.309-0.435	This study
C. corymbosa	Endemic	6	Perennial, fruit with pappus, mostly self-incompatible, entomophilous outcrosser, limestone cliff, insect pollinated.	?	Allozymes	0.074/0.03-0.074	(Colas et al. 1997)
					Allozymes	0.20/0.11-0.26	(Fréville et al. 2001)
					SSR	0.50/0.36-0.62	(Fréville et al. 2001)
C. horrida	Endemic	7	Dwarf, long-living, sea cliffs, outcrossing, insect pollinated.	2x	SSR	No data/0.603-0.854	(Mameli et al. 2008)
C. nivea	Endemic	5	Perennial, rhizomatous plant, calcareous soils.	2x	RAPD	0.296/0.244-0.258 I=0.451/ I=0.372-0.389	(Sözen and Özaydin 2009)
C. wiedemanniana	Endemic	6	Perennial.	2x	RAPD	0.278/0.183-0.211 I=0.429/ I= 0.283-0.324	(Sözen and Özaydin 2010)
Feminasia balearica[a]	Endemic	7	Shrub, enthomophilous outcrosser, siliceous costal cliff, deciduous pappus.	2x	AFLP	0.237/0.157-0.190	(Vilatersana et al. 2007)
C. cineraria	Endemic	2	Perennial, limestone cliff.	2x	Allozymes	No data/0.126-0.186	(Bancheva 2006)
C. ucriae	Endemic	3	Limestone cliff.	2x	Allozymes	No data/0.130-0.205	(Bancheva 2006)
C. todari	Endemic	2	Limestone cliff.	2x	Allozymes	No data/0.226-0.276	(Bancheva 2006)
C. tenorei	Endemic	3	Perennial herb.	2x-4x	Allozymes	0.08/No data	(Palermo 2002)
C. parlatoris	Endemic	3	Perennial herb.	2x	Allozymes	0.34/No data	(Palermo 2002)
C. maculosa spp. *maculosa*	Widespread	5	Perennial, self-incompatible, enthomophilous, monocarpic, calcareous rocky places.	2x	Allozymes	No data/0.044-0.170	(Fréville et al. 1998)
C. solstitialis	Widespread	22	Annual. Alien range (North America, since 1800).	2x	Allozymes	No data/0.257-0.417	(Sun 1997)
C. jacea	Widespread	5	Perennial, enthomophilous, self-incompatible, fruit with pappus.	2x	Allozymes	No data/0.27-0.45	(Hardy and Vekemans 2001)
		5		4x		No data/0.36-0.41	
C. diffusa	Widespread	8	Outcrosser. Alien range (North America, since 1907).	2x	SSR	No data/0.436-0.692	(Marrs et al. 2008b)
		5	Outcrosser. Native range (Eurasia).	2x-4x	SSR	No data/0.311-0.592	
C. stoebe spp. *micranthos*	Widespread	11	Alien range (North America).	4x	SSR	No data/0.616-0.809	(Marrs et al. 2008a)
		15	Native range (Eurasia).	4x	SSR	No data/0.521-0.856	
C. africana	Widespread	1	Perennial,	2n=30[b]	Allozymes	0.35/0.35	(Garnatje et al. 1998)

Range size, Ploidy, and *Biological traits* as indicated by the authors (? = ploidy not available in the reference). *Pop* is the number of local populations used for genetic diversity estimates. Diversity values are Nei's gene diversity unless otherwise indicated (*I* = Shannon information index).[a] Formerly known as *Centaurea balearica* J.J. Rodr. [b] Ploidy not indicated.

CHAPTER 2

"Patterns of chloroplast DNA polymorphism in the endangered polyploid *Centaurea borjae* (Asteraceae): implications for preserving genetic diversity."

Published as: Lopez L. & Barreiro R. (2013). Patterns of chloroplast DNA polymorphism in the endangered polyploid *Centaurea borjae* (Asteraceae): implication for preserving genetic diversity. Journal of Systematics and Evolution. 51 (4): 451-460. doi: 10.1111/jse.12012.

ABSTRACT

A previous study with AFLP fingerprints found no evidence of genetic impoverishment in the endangered *Centaurea borjae* and recommended that four management units (MUs) should be designated. Nevertheless, the high ploidy (6x) of this narrow endemic plant suggested that these conclusions should be validated by independent evidence derived from non-nuclear markers. Here, the variable *trnT-F* region of the plastid genome was sequenced to obtain this new evidence and to provide an historical background for the current genetic structure. Plastid sequences revealed little genetic variation; calling into question the previous conclusion that *C. borjae* does not undergo genetic impoverishment. By contrast, the conclusion that gene flow must be low was reinforced by the strong genetic differentiation detected among populations using plastid sequences (global F_{ST} = 0.419). The spatial arrangement of haplotypes and diversity indicate that the populations currently located at the center of the species range are probable sites of long-persistence whereas the remaining sites may have derived from a latter colonization. From a conservation perspective, four populations contributed most to the allelic richness of the plastid genome of the species and should be given priority. Combined with previous AFLP results, these new data recommended that five, instead of four, MUs should be established. Altogether, our study highlights the benefits of combining markers with different modes of inheritance to design accurate conservation guidelines and to obtain clues on the evolutionary processes behind the present-day genetic structures.

Key words: *Centaurea borjae*, conservation, cpDNA, genetic diversity, narrow endemic, *trnT-F*.

CHAPTER 2

INTRODUCTION

According to the International Union for Conservation of Nature (IUCN), genetic variation is a key component of biodiversity and must be preserved (www.iucn.org). Genetic variation is essential to facilitate evolutionary responses to environmental change (Lande, 1988; Reed and Frankham, 2003). Low levels of genetic diversity can reduce the evolutionary potential and increase the short-term extinction risk of a species (Frankham *et al.*, 2002; Willi *et al.*, 2006; Allendorf and Luikart, 2007). In this context, proper conservation of biodiversity requires reliable estimates of the magnitude and the spatial distribution of genetic variation within and among populations (Hamrick and Godt, 1996; Frankham *et al.*, 2002). This knowledge is even more relevant in narrowly occurring plants as they often combine a number of features that make them potentially susceptible to genetic risks: reduced population size, habitat specificity, and isolation (Ellstrand and Elam, 1993; Hamrick and Godt, 1996; Cole, 2003).

Centaurea borjae Valdés B. and Rivas G. (1978) is a good example of the latter. A narrow endemic in the otherwise widespread genus Centaurea (Asteraceae), this small perennial plant has a total occupancy below 5000 m2 (Bañares *et al.*, 2004). It occurs in a few enclaves concentrated in 16 km of costal cliffs in North West Iberian Peninsula, except for a geographically isolated population that was discovered 25 km apart from the other sites (Soñora, 1994). Given its extremely narrow range, *C. borjae* is listed as “endangered” by national (Spanish Catalogue of Threatened Species) and international (IUCN) organizations (Gómez-Orellana Rodríguez, 2011), and included among the “priority species” of the Habitats Directive (92/43/EEC, Annex II). In addition to its small range, this plant possibly has little potential for dispersal. Thus, several pieces of evidence suggest that seed production may be small. Rosette leaves develop a single capitulum (rarely 2) per year that, according to some estimates, produce a limited number of viable fruits (7 fruits per capitulum on average; Izco,

2003). Moreover, although *C. borjae* is an entomophilous outcrosser with hermaphroditic flowers, self-incompatibility is known to be common in other Centaurea (Colas *et al.*, 1997; Pisanu *et al.*, 2009; Sun and Ritland, 1998). Some estimates indicate that germination success is likewise low (Retuerto R, 2012, unpublished data; but see Izco, 2003 for other estimates; Gómez-Orellana Rodríguez, 2004), possibly aggravated by the fact that insect larvae are commonly found feeding on ripe fruits within mature flower heads (Fernández Casas and Sussana, 1986). Finally, seed dispersal is thought to be limited too, as the fruit lacks a pappus. Instead, the presence of an elaiosome suggests that ants may play a role in seed dispersal as they do in many other Centaurea (Imbert, 2006). In comparison, vegetative propagation can be considerable because *C. borjae* produces asexual rhizomes up to several meters long. Despite the above, a previous survey of *C. borjae* with highly polymorphic nuclear markers (amplified fragment length polymorphism, AFLP) failed to detect signs of genetic impoverishment (Lopez and Barreiro, 2012). Contrary to the expectations of a predominantly vegetative propagation, the AFLP fingerprints revealed that clone mates were rare. Still, *C. borjae* did show substantial differentiation among locations. Even sites separated by only a few hundred meters were significantly different. This strong genetic differentiation was consistent with the poor dispersal capacity anticipated by its biological traits and suggested that gene flow must be low among populations. Moreover, there was evidence that gene flow was likewise restricted within populations because small-scale spatial analyses revealed a significant autocorrelation for distances up to 35–80 m. This limited gene flow explains why, with the help of Bayesian assignment methods, we proposed to divide the range of *C. borjae* into four management units for conservation purposes.

Genome-wide markers such as AFLP and random amplified polymorphic DNA (RAPD) have been widely used in plant studies because of their easiness to produce large numbers of highly variable markers in species that lack prior sequence

information (Mba and Tohme, 2005; Schaal *et al.*, 2003). These fingerprinting techniques have been very fruitful in a wide range of applications (see Meudt and Clarke, 2007 and references therein) but they also have shortcomings. In this regard, our set of AFLP markers for *C. borjae* featured a very high resolving power as evidenced by the fact that most of the rosette leaves sampled in our study showed a distinct fingerprinting profile. As a result, our data seemed particularly well suited for analyses that involved an individual-based approach such as population assignment procedures, detection of small-scale spatial genetic structure, and identification of potential clone mates. However, their accuracy for analyses that required a population-based approach, e.g. estimates of genetic diversity and differentiation at the population level, was less clear. Due to the dominant mode of inheritance of AFLP/RAPD, allele frequencies are not directly accessible. Instead, data analysis relies on certain assumptions or resorts to alternative approaches (e.g. band-based analysis) which may raise concerns over bias in their estimates (Bonin *et al.*, 2007). The latter seems particularly worrisome in polyploids such as *C. borjae,* a hexaploid with 2n = 66 and x = 11. Moreover, it also implies that their genome offers more opportunities to hide part of the genetic diversity to the predominantly nuclear AFLP markers.

Another important limitation of AFLP/RAPD is that their data cannot be historically ordered. As a result, they provide only indirect information about population histories (Avise, 2004). However, the distribution of genetic variation in plant populations is strongly affected both by currents patterns of microevolutionary forces, such as gene flow and selection, and by the phylogenetic history of populations (Schaal *et al.*, 2003). The latter can only be inferred using markers with a different mode of inheritance, being chloroplast-DNA (cpDNA) variation a frequent choice in population-level studies of plants. Moreover, cpDNA is maternally inherited in most angiosperms (McCauley, 1995). Therefore, it generally informs of the genetic structure that results from seed flow, a variable that relates more easily to

demographic connectivity among populations, while the gene flow detected by nuclear markers is mostly due to pollen transfer (Ouborg *et al.*, 1999). Last but not least, the haploid nature of cpDNA obviates the hurdles encountered while working with polyploids. The merits of cpDNA markers for intraspecific studies have been demonstrated in applications that range from population structure, to phylogeography, or into the reconstruction of the evolutionary history of endemic and endangered species (Aizawa *et al.*, 2008; Ge *et al.*, 2011; Gong *et al.*, 2011; Liu *et al.*, 2010; Molins *et al.*, 2009; Zhou *et al.*, 2010). Similarly, the comparative analysis of chloroplast and nuclear DNA variation has become a widely used approach to get a more thorough view of the genetic structure in population-level studies of plants (e.g. Kato *et al.*, 2011).

Here, we employed sequences of the non-coding cpDNA region trnT-F (Taberlet *et al.*, 1991) to investigate the genetic structure of *C. borjae a*long its range and the historical processes behind it. Results were compared to those obtained previously with unordered AFLP markers; AFLP are widely acknowledged as predominantly nuclear in origin (Meudt and Clarke, 2007; Nybom, 2004). More specifically, in this study we aimed to: (1) estimate the genetic diversity of *C. borjae* using cpDNA sequences, (2) investigate its demographic past, (3) evaluate its population structure, (4) identify populations of greater conservation concern and, finally, (5) compare the pattern obtained with cpDNA sequences with the results of the AFLP study.

MATERIAL AND METHODS

Sample collection and storage

Our sampling scheme covered all known populations of *Centaurea borjae* (Izco, 2003) (see Fig. 1 in results). As this plant tends to display a clumped distribution, 2–4 rosette leaves per aggregation were sampled covering the whole area occupied

by the species at each site. Fresh leaves were dried in silica gel and stored at –20 °C prior processing. The samples used in the present study are the same as those analyzed for AFLP by Lopez and Barreiro (2012).

Sequencing

DNA was isolated using the Wizard Magnetic Kit (Promega, Madison, USA) according to the manufacturer's instructions. DNA integrity and negative controls were verified on 1.5% agarose gels. The *trnT-F* region encompasses three different fragments, two intergenic spacers (*trn*T-*trn*L and *trn*L-*trn*F) and the intron *trnL*. The three fragments were sequenced following Taberlet *et al.* (1991) with minor modifications. First, PCR reactions for the intergenic spacer *trnT-trnL* were performed in 25 µL using primers *a* and *b* (Taberlet *et al.*, 1991). Reactions contained 1x reaction buffer, 2 mmol/L MgCl2, 0.2 mmol/L of each dNTP, 0.5 µmol/L of each primer, 1 µL of genomic DNA, and 1.25 units of DNA polymerase (Applied Biosystems). The *trnL* intron and the intergenic spacer *trnL-trnF* were amplified using primers c and d, and e and f respectively. PCR mixes for these fragments included 1x reaction buffer, 1.5 mmol/L MgCl2, 0.2 mmol/L of each dNTP, 0.5 µmol/L of each primer, 1 µL of genomic DNA, and 0.35 units of DNA polymerase (Applied Biosystems). Regardless of the fragment, PCR profiles consisted of 2 min denaturation at 94°C followed by 35 cycles of 1 min denaturation at 94 °C, 1 min annealing at 50 °C and 90 s of extension at 72 °C, with a final elongation step of 3 min at 72 °C. PCR products were visualized on 1.5% agarose gels and purified with 1 µL of Exonuclease I (20 U/µL) and 2 µL of FastAP (10 U/µL) (Fermentas). Purified PCR products were bi-directionally sequenced under BigDye Terminator cycling conditions on an Automatic Sequencer 3730XL (Applied Biosystems, USA). Trace files were trimmed and assembled in CodonCode Aligner 3.7.1 (CodonCode Corporation, USA). Sequences were then aligned using ClustalW (Thompson *et al.*, 1994) as implemented in DnaSP v 5.0 (Librado and Rozas, 2009; Rozas *et al.*, 2003). Since the non-recombinant nature of cpDNA makes it equivalent

to a single-locus marker, sequences from the three fragments were combined into a single haplotype for each individual. Singleton polymorphisms and haplotypes occurring in on single individual were confirmed by reanalysis, starting from the sequencing reaction step, to discard PCR errors and/or sequencing artifacts.

Data analysis

The prior study with AFLP markers detected the occurrence of clones in some populations. Clone mates were also sequenced for cpDNA; however, only individuals with unique AFLP fingerprints were retained for statistical analyses unless otherwise stated. Distinct haplotypes were identified with the help of DnaSP v.5 (Rozas *et al.*, 2003) while their genealogy was inferred with the median-joining network algorithm implemented in Network 4.6 (Bandelt *et al.*, 1999). Genetic diversity was evaluated for each population as haplotype diversity (H_d; Nei, 1987) and nucleotide diversity (π) using Arlequin 3.5 (Excoffier *et al.*, 2005). Besides, the average intrapopulation diversity (h_s) and the total diversity (h_t) were estimated using Permut (Pons and Petit, 1996). The contribution of each population to total haplotypic diversity (C_T) and total haplotypic richness (C^r_T) was estimated using Contrib (Petit *et al.*, 1998; Pons and Petit, 1996) These contributions were partitioned into two components: one related to the level of diversity of the population (C_S and C^r_S) and the other to its divergence from the others populations (C_D and C^r_D).

Population structure was assessed by an analysis of molecular variance (AMOVA) based on haplotypes frequencies (Excoffier *et al.*, 1992); the significance of the FST statistic was tested by 1023 permutations calculated with Arlequin v3.5 (Excoffier *et al.*, 2005). A rough estimate of migration rate (N_m) among populations due to seed dispersal was estimated using the expression $F_{ST}=1/(1+2N_m)$ (Hudson *et al.*, 1992; Slatkin, 1993). Also, Permut was used to calculate and compare two measures of population differentiation, G_{ST} and N_{ST}, under the assumption that a significantly higher N_{ST} value suggests the existence of phylogeographic structure

(Pons and Petit, 1996). The correlation between geographic and genetic distances was investigated with the Mantel test implemented in the IBD Web Service (Bohonak, 2002), and its significance was determined after 1000 randomizations.

RESULTS

Phylogenetic relationships and geographical distribution of haplotypes

Among the three non-coding regions, only the intergenic spacers *trnT-L* and *trnL-F* showed polymorphism and were retained for statistical analyses. These two intergenic spacers were aligned with a total consensus length of 1003 bp: 577 bp for the *trnT-L* region and 426 bp for the *trnL-F* one. Sequences have been deposited in the GenBank database under Accession Nos. KC522681–KC522692. No intragenomic polymorphism (heteroplasmy) was detected in this study. Sequences were rich in A and T nucleotides (A/T content = 68%) in agreement with the nucleotide composition of non-coding chloroplast regions (Kelchner, 2000). Seven segregating sites were detected that included five point mutations and two indels. Three point mutations and the two indels occurred in the *trnT-L* region while only two point mutations were detected in the shorter *trnL-F* fragment. Altogether, these seven variable positions defined six haplotypes in the cpDNA. Three of them (H1, H2, and H6) were frequently sampled and comprised >95% of the individuals whereas the other three were very rare and only occurred in one or two individuals each.

The parsimony network yielded a neat genealogy free from ambiguities (Fig. 1). According to this network, the minimum number of mutations necessary to explain the data was seven. Its topology revealed the occurrence of two groups of haplotypes separated by three mutations. This partition in two groups largely resulted from the two point mutations detected in the *trnL-F* fragment. One group contained only two haplotypes and was dominated by H1, the most common haplotype in our data set that was also widely distributed along the species range. The other group consisted

of haplotypes H4–H6 arranged in a star-like phylogeny around H2. All the haplotypes at the tips of the genealogy were always separated by one single mutational step from the most widespread haplotype within each group.

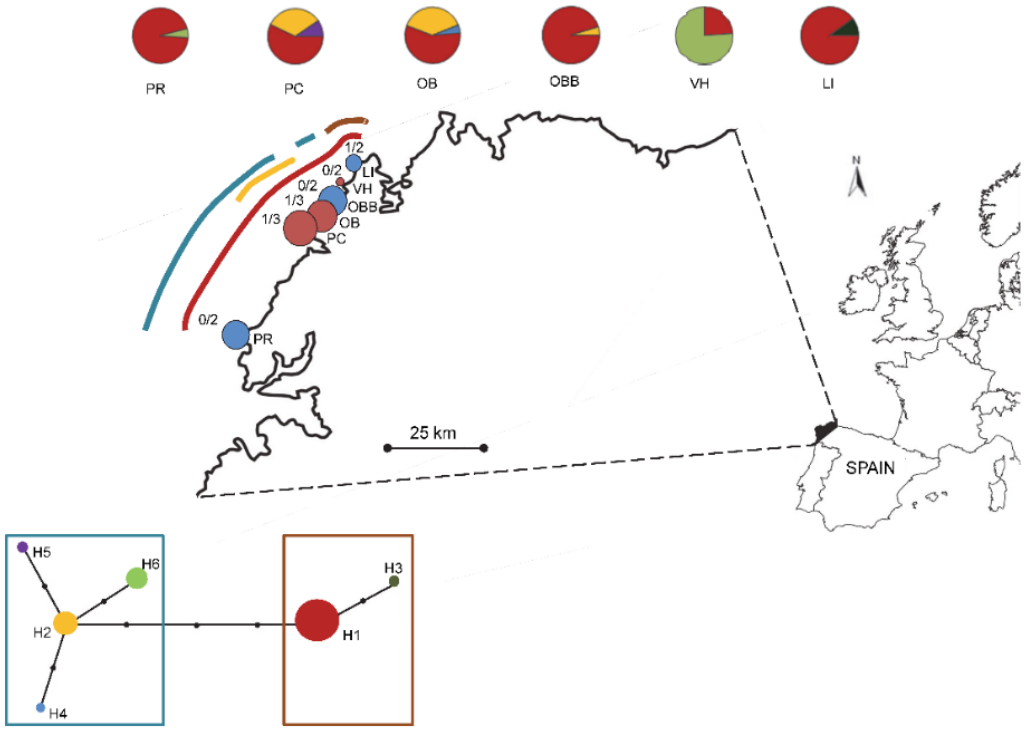


Fig. 1. Map showing the locations investigated for *Centaurea borjae*, the distribution of the chloroplast haplotypes, diversity plot, and haplotype network. Location codes: LI, VH, OBB, OB, PC, and PR. Pie charts show relative abundances of six haplotypes (codes H1-H6) in each population with colors matching the haplotype network. In the diversity plot, bubble sizes are proportional to deviation from the mean diversity for all populations; red fill indicates diversity above the mean whereas blue fill shows diversity below the mean. The proportion of private haplotypes for each population (number of private haplotypes/total number of haplotypes) is shown beside each bubble. Thick lines summarize the distribution of older haplotypes H1, H2 and of haplotypes derived from them, color-coded to the haplotype network. The median-joining network analysis is represented at the bottom. The size of each circle is proportional to haplotype frequency across populations. Each black-dot in the line between two adjacent haplotypes indicates a mutational step.

As shown in Fig. 1, none of the populations was monomorphic although one of the haplotypes always prevailed over the others comprising >50% of individuals. In most cases, the prevailing haplotype was the widely distributed H1. The only exception was population VH which was dominated by H6. This gave VH a peculiar character even though haplotype H1 was also found here in nearly 25% of the individuals. Haplotype H6 was also detected in PR; however, its presence in PR was residual as it was only detected in a single individual. On the other hand, haplotype H2, the second-most widespread haplotype in *C. borjae*, seemed restricted to the three populations at the center of the species range (PC, OB, OBB). Remarkably, H6 and H2 were never found in sympatry despite the fact that our genealogy indicated that H6 possibly derives from H2. Finally, low frequency haplotypes H3, H4, and H5 were unique to populations LI, OB, and PC, respectively.

Chloroplast haplotype diversity and population differentiation

Total haplotype diversity (H_d) for the species was 0.490 ± 0.048 and total nucleotide diversity (π x 102) was 0.157 ± 0.104 whereas total gene diversity (*h*t) was 0.515 ± 0.132 using the approach proposed by Pons and Petit (1995). On the other hand, average within-population gene diversity (*h*s) was 0.317 ± 0.089. Haplotype and nucleotide diversity were highly correlated across populations and ranged from 0.095 to 0.581 and from 0.019 to 0.172, respectively (Table 1). Their highest estimates were recorded at populations PC and OB (*H*d= 0.581 and 0.552, π x 102 = 0.172 and 0.164, respectively) at the center of the species range. PC and OB also contained two out of the three private haplotypes detected in *C. borjae* (Fig. 1). In comparison, the populations at each end of the distribution range (LI and PR) produced estimates below the mean for all populations but their values were similar to those recorded in OBB, a population that is very close to OB. The peculiar VH showed diversity values close to the mean for all populations. Since *C. borjae* reproduces asexually, diversity estimates were recalculated including the putative clones detected at each location.

This involved 10 individuals with an AFLP pattern identical to others already included in our data set. In the field, clone mates were found spatially clumped and they always had the same cpDNA haplotype. Overall, clone mate occurrence was low (18.2%) and had minimal impact on the estimates of diversity (results not shown). Actually, rather than decrease, the estimates of diversity increased slightly because many clone mates belonged to poorly representing haplotypes, resulting in a more balanced distribution of haplotypes within populations.

Table 1 Genetic diversity measures of *Centaurea borjae* at the six known locations.

Location	n	H	S	H_d (SD)	π x 10^2 (SD)
LI	20	2_+	4	0.190 (0.108)	0.019 (0.028)
VH	21	2	4	0.381 (0.101)	0.152 (0.106)
OBB	21	2	4	0.095 (0.084)	0.029 (0.035)
OB	21	3_+	3	0.552 (0.066)	0.164 (0.112)
PC	21	3_+	4	0.581 (0.075)	0.172 (0.116)
PR	19	2	1	0.105 (0.092)	0.042 (0.045)
Total	123	6	7	0.490 (0.048)	0.157 (0.104)

LI, VH, OBB, OB, PC, PR; n, number of sampled individuals; H, number of haplotypes ($_+$ denotes the occurrence of one private haplotype); S, number of segregating sites; H_d, haplotypic diversity; π x 10^2, nucleotide diversity; SD, standard error.

The populations of *C. borjae* contributed differently to total haplotype diversity (C_T) and richness (C^r_T) (Fig. 2). Population VH contributed much more to the total diversity than the others as shown by its C_T value (nearly 30%). This was mostly due to its strong divergence (C_D = 25.6%) because its diversity was essentially similar to the average (C_S = 2.5%). On the other hand, the two populations at the center of the distribution range (PC, OB) also had a positive total contribution to total diversity (C_T = 4.3% and 4.2%, respectively). However, in comparison with VH, their positive contribution was due to their diversity (C_S = 10.2% and 9.1%) whereas their divergence was below the average (C_D = –4.9% and –6.0%). The results based on the

contribution to total allelic richness were similar. Again, it was VH that contributed the most to total allelic richness (C^r_T = 22.9%) because it was enormously differentiated from the other populations (C^r_D = 25.2%). Likewise, OB and PC had positive total contribution; in OB, the positive contribution was due to its richness (C^r_S = 4.6%) whereas both richness and differentiation contributed the same in PC (C^r_S = 5.3%, C^r_D = 4.8%). Finally, the contribution to allelic richness showed an interesting difference: LI, at the northern end of the distribution range, also had a considerable net contribution to allelic richness (C^r_T = 5.9%) due to their important differentiation (C^r_D = 8.2%).

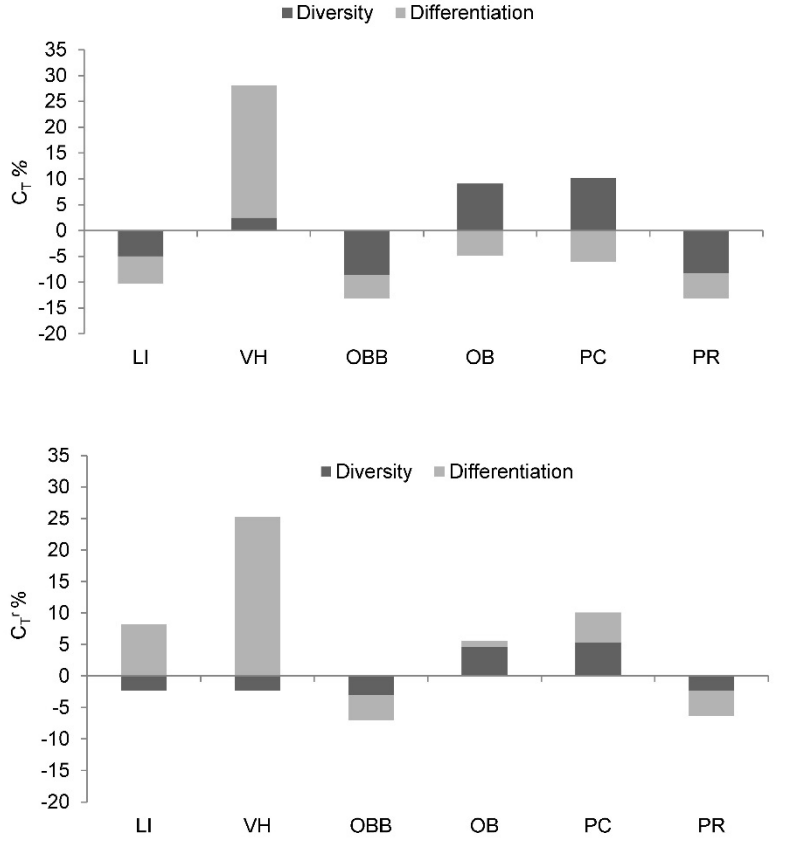


Fig. 2. Contribution to total haplotype diversity (C_T) and haplotypic richness (C^r_T) of each population of *Centaurea borjae* inferred with cpDNA haplotypes. The black and the grey bars represent the

contribution of diversity (C_S and C'_S) and differentiation (C_D and C'_D), respectively. Location codes: LI, VH, OBB, OB, PC and PR.

The AMOVA analysis revealed that 41.9% of the genetic variation was due to differences between populations. The resulting F_{ST} value was high and significant (0.419, $p < 0.001$) while the overall level of the inferred gene flow (N_m) was low (0.69 migrants per generation among populations). Despite this strong global differentiation, an examination of the pair-wise F_{ST} values provided statistic support to the occurrence of three sets of populations with similar haplotype composition. Populations PR, OBB, and LI were characterized by an overwhelming prevalence (>90%) of haplotype H1. Interestingly, this group does not consist of spatially contiguous populations; instead, its components are separated by other populations with totally different haplotype composition (Fig. 1). On the other hand, sites PC and OB were characterized by a more balanced partition between H1 (60%) and H2 (35%). Finally, VH displayed a clearly discordant composition, being the only population dominated by a haplotype other than H1. Despite the strong differentiation and low level of global gene flow, N_{ST} (0.380 ± 0.106) was not found to be significantly different from G_{ST} (0.383 ± 0.102) ($p > 0.05$ after 1000 permutations), indicating a lack of phylogeographic structure. Likewise, a Mantel test revealed no evidence of isolation by distance when testing for the correlation between genetic and geographic distances ($R^2 = 0.023$, $P = 0.357$) after 1000 randomizations.

DISCUSSION

As other endemics, *Centaurea borjae* may be prone to the consequences of genetic drift and inbreeding that, together with the fragmentation of its habitat, may threaten the long-term survival of its populations (Ellstrand and Elam, 1993). In this regard, a prior study with AFLP found no signs of genetic depletion in *C. borjae* (Lopez and Barreiro, 2012). However, the adequacy of the AFLP technique as a tool to obtain accurate estimates of diversity in *C. borjae* seemed debatable. One might suspect that

AFLP estimates may be biased by the interplay of the dominant mode of inheritance of the markers with the hexaploidy of *C. borjae*. In this context, sequencing non-coding regions of the cpDNA seemed a straightforward complement to obtain more comparable estimates (Kato *et al.*, 2011).

In comparison with our prior AFLP study (Lopez and Barreiro, 2012), the maternally inherited cpDNA provided some evidence of genetic depletion in *C. borjae*. First, the total number of haplotypes in *C. borjae* was typically lower than the values reported in widespread plants (Fang *et al.*, 2010; Su *et al.*, 2011) but similar to those of other narrow endemics (Artyukova *et al.*, 2011; Migliore *et al.*, 2011; Molins *et al.*, 2009). Likewise, nucleotide diversity was low and similar to estimates reported for other endemics (see Artyukova *et al.*, 2011 and references therein). Finally, species diversity, as h_t, was below the average computed for chloroplast regions in angiosperms (h_t = 0.712, range 0.375–0.993) using values compiled by Petit *et al.* (2005). Moreover, total diversity, as H_d or as h_t, was also well below the values for cpDNA in other endemic herbs (Artyukova *et al.*, 2011; Molins *et al.*, 2009; Zhou *et al.*, 2010).

A similar incongruence between nuclear and cpDNA markers has been reported elsewhere (e.g. Zhao and Gong, 2012). It has been often attributed to differences in mutation rate and effective population size (Schaal *et al.*, 1998); the latter effect might be aggravated in hexaploids such as *C. borjae.* A detailed examination of the results of *C. borjae* shows that the low haplotype diversity results from the predominance of a single widespread haplotype across most populations: haplotype H1 was detected in nearly 70% of individuals, prevailing in 5 out of the 6 populations. In comparison, other endemics such us *Senecio rodriguezii* also had populations largely dominated by one haplotype (Molins *et al.*, 2009) but the prevailing haplotype changed between populations resulting in high species diversity (h_t) but low average within-population diversity (h_s). On other occasions (e.g.

Oxytropis chankaensis, Artyukova *et al.*, 2011), populations were characterized by a more balanced partition of individuals among several (2–3) haplotypes that made both h_t and h_s high.

The structure of genetic variation across a species' range is typically interpreted in terms of contemporary (e.g. effective population size, gene flow) and historical (e.g. fragmentation, founder events) factors (Schaal *et al.*, 2003). Likewise, both contemporary and historical factors explain the present day population pattern detected in *C. borjae*. The predominance of a single, widespread haplotype in most populations cannot be attributed to intense current gene flow. Instead, both AFLP and cpDNA reveal a strong differentiation between populations. Moreover, prior studies at small scale indicate that gene flow is restricted even within populations (Lopez and Barreiro, 2012). Alternatively, the current arrangement of haplotypes may be a consequence of the demographic history of the plant. The coalescence theory predicts that older alleles will be more broadly distributed geographically; also, the tip nodes of a network will likely represent descendants derived from ancestral, interior nodes (Posada and Crandall, 2001). Accordingly, haplotypes H1 and H2 would represent some old polymorphism that had been long-maintained and their co-occurrence in PC and OB suggests that this area is a site of long persistence of the species. The same conclusion is reached by analyzing the spatial distribution of genetic diversity and private haplotypes. Long-maintained populations are known to be more diverse and to contain private haplotypes (Maggs *et al.*, 2008); two conditions met by PC and OB. In this scenario, the remaining sites may have derived from subsequent colonization from the central area and their lower genetic diversity would be the product of a founder effect.

On the other hand, *C. borjae* shows a decrease in genetic diversity towards its range limits that mimics a small-scale version of the pattern anticipated by the central-marginal hypothesis (Brussard, 1984). The latter is one of the hypotheses drawn from

the controversial abundant-center assumption (Sagarin *et al.*, 2006). According to the central-marginal hypothesis, geographically peripheral populations would experience stronger drift as a result of their smaller effective size and greater isolation. This will be further exacerbated if peripheral populations suffer founder events or more stressful environmental conditions (Eckert *et al.*, 2008). Regarding isolation, the southern range-edge fits the expectations of the central-marginal model as population PR is clearly separated from the others by a large stretch of unsuitable habitat. The same does not apply to the northern edge because its populations are not particularly isolated. Yet, the lack of isolation does not mean that other assumptions of the model are not applicable to this northern edge. Despite the small range occupied by the plant, the 3 northernmost populations show distinct environmental conditions due to the extremely intricate geology of the area: these 3 northernmost sites are located on serpentine substratum that contrasts with the ultra-basic (PC, OB) and granitoid (PR) soils found at the other locations. Serpentine soils are characterized by high levels of toxic heavy metals (Cr, Ni, Co) that are known to be stressful for plant growth. In fact, our previous study with AFLP revealed that serpentine soils had an impact on the genetic structure and variation of *C. borjae*. Serpentine populations had a larger occurrence of clones mates and a stronger small-scale spatial genetic structure than non-serpentine locations (Lopez and Barreiro, 2012). Therefore, the stressful ambient conditions generated by the serpentine soils may have led to smaller effective population sizes and more intense genetic drift.

Gene flow was low in *C. borjae* (N_m=0.6930), resembling estimates for other endemics with similar biological traits (Liu *et al.*, 2010; Zhou *et al.*, 2010). Moreover, the significant F_{ST} obtained for the chloroplast genome was almost four times higher than the F_{ST} calculated with nuclear markers. Maternally inherited markers are expected to display larger differentiation than biparentally inherited nuclear ones as the former can be dispersed between populations only by seeds whereas the latter

can migrate by both pollen and seeds (Ouborg *et al.*, 1999; Ghazoul, 2005). Thus, the higher differentiation detected with cpDNA supports the conclusion that seeds in *C. borjae* disperse less than pollen (McCauley, 1995). Likewise, low dispersal seems consistent with several biological traits of *C. borjae*: lack of pappus, probable myrmecochory, and low germination success. Previous studies with another endemic Centaurea also indicated low dispersal and gene flow (Hardy *et al.*, 2004; Imbert, 2006). Finally, the lack of correlation between genetic and geographic patterns could be seeing as further evidence that the neighboring populations are not connected by gene flow.

According to Petit *et al.* (1998), the criterion to select priority populations for conservation should encompass the uniqueness of a population and its diversity level, with an emphasis on allelic richness. In this regard, the uneven distribution of cpDNA polymorphism among populations leads to prioritizing four enclaves in terms of their contribution to haplotype richness and diversity: LI, VH, OB and PC. By preserving these four populations, all known haplotypes will be maintained. In comparison, neither OBB nor PR provided any significant contribution and their conservation might be seen as less relevant. These results complement our prior findings with nuclear markers. The Bayesian analysis of AFLP led to the designation of four MUs (Management Units; *sensu* Moritz, 1994) that, remarkably, clustered OB, OBB, and PC into a single unit. Therefore, should we stick to the conservation guidelines derived from AFLP data, OB and PC would be considered genetically redundant. By contrast, the cpDNA data revealed that both PC and OB have private alleles and are not interchangeable in conservations terms. Likewise important, the four populations identified as priority by cpDNA only included three of the four MUs designated with nuclear markers. The excluded MU was the geographically isolated PR that, according to AFLP, has a certain level of uniqueness: a private band and noticeably different marker frequencies (Lopez and Barreiro, 2012). The disagreement between markers

with different mode of inheritance is well known and it possibly reflects differences in the time-span covered by each set of markers (deep, longer-term historical structure for cpDNA; shallow, contemporary one for AFLP) (Avise, 2004). In this regard, the fact that PR showed a singular composition in nuclear markers but not in its chloroplast genome suggests that its separation from the main range of the species is a relatively recent event. According to Avise (2004), combining markers with different mode of inheritance is important to design accurate management strategies. In our study, the combination of AFLP and cpDNA data suggests that five, instead of four, management units should be designated for *C. borjae*: LI, VH, OB-OBB, PC, and PR.

In summary, our study highlights the convenience of combining markers with a different mode of inheritance to obtain a more comprehensive image of the genetic structure. This knowledge is essential to design appropriate conservation strategies. Both sets of markers supported the idea of restricted gene flow between populations with seed dispersal more constrained than pollen migration. However, cpDNA data showed symptoms of genetic depletion that went unnoticed to the nuclear markers. Moreover, the plastid sequences provided insights into the demographic history of the plant. PC and OB appear as the probable sites of long-persistence of the species whereas other sites may have derived from a latter colonization. The cpDNA data also allowed us to select candidate populations that should be given priority for conservation. Combined with AFLP data, it is proposed that five MUs should be designated to ensure the maintenance of all the genetic polymorphism detected in *C. borjae*.

ACKNOWLEDGEMENTS

Research supported by grant 07MDS031103PR (Xunta de Galicia). Lua Lopez acknowledges support from Universidade da Coruña (contratos predoutorais UDC 2012).

CHAPTER 3

"A multi-faceted approach for the conservation of the endangered *Omphalodes littoralis* spp. *gallaecica*."

Published as: Lopez L., Retuerto R., Barreiro R. A comprehensive studio in the endangered *Omphalodes littoralis* subsp. *gallaecica*: genetic and phenotypic approach for its preservation. Perspectives in Plant Ecology and Evolution. (2nd revision)

ABSTRACT

Genetic diversity is now regarded as a key component of biodiversity and its assessment has become a frequent addition to conservation studies. However, due to practical limitations, most studies assess genetic variation using neutral markers while the variability of evolutionary relevant quantitative traits is typically overlooked. Here, we simultaneously assessed neutral and quantitative variation in an endangered plant to identify the mechanism behind their spatial arrangement and to propose conservation guidelines for maximizing mid- to long-term survival. *Omphalodes littoralis* spp. *gallaecica* is a self-fertilizing therophyte with an extremely narrow and fragmented distribution. Regardless of the marker set (non-coding sequences of cpDNA or Amplified Fragment Length Polymorphism loci), the five extant populations of O. littoralis showed minimal to no neutral genetic diversity and a lack of gene flow between them. Moreover, genetic structure was identical in samples collected on two consecutive years suggesting that the seed bank cannot buffer against genetic loss. High rates of self-fertilization together with a strongly fragmented distribution and recurrent bottlenecks seem the likely mechanisms that may have led to a dramatic loss of genetic variation in a classic scenario drawn by genetic drift. Despite the extremely narrow distribution range, reciprocal transplant experiments revealed that the populations differed in several quantitative traits and that these differences likely have a genetic basis. Nevertheless, the pattern of differences among populations did not fit the expectations of local adaptation. Instead, phenotypic variation seemed another outcome of genetic drift with important implications for conservation because each population should be designated as an independent Evolutionary Significant Unit (ESU). Our study evidences the benefits of combining neutral markers with appropriate assessments of phenotype variation, and shows that even endemics with extremely narrow ranges can contain multiple conservation units.

Keywords: conservation, genetic structure, *Omphalodes littoralis*, phenotypic traits, selfing, rare plant, reciprocal transplants, local adaptation.

INTRODUCTION

Designing and implementing appropriate measure that enhance the long-term survival of populations is a major challenge in conservation (Ellstrand and Elam, 1993). In this regard, the genetic structure of endangered populations has become a primary focus of research since theory predicts that intraspecific genetic variation is pivotal for the persistence of species (Ouborg *et al.,* 2006). Under the premise that populations may achieve their greatest evolutionary potential by maximizing their genetic diversity, conservation efforts often aim to preserve the most divergent populations and/or those displaying the largest levels of variation (Moritz, 1994).

Due to practical limitations, the genetic structure is usually assessed with neutral molecular markers even if their suitability for the purposes of conservations has been repeatedly questioned (Landguth and Balkenhol, 2012; Reed and Frankham, 2001). Instead, quantitative traits are those of most concern for conservation because they are directly related to the species' fitness (Frankham *et al.,* 2010). As natural selection act directly on phenotypes, not on genotypes, these traits reflect the species' ability to undergo adaptive evolution as well as the consequences of inbreeding and outbreeding on reproductive fitness (Allendorf and Luikart, 2012). Unfortunately, current evidence suggests that neutral variation may not be an accurate indicator of quantitative variation; consequently, making decisions based only on genetic differences detected by neutral markers is not without risk (Frankham *et al.,* 2010; Hedrick, 2001; Landguth and Balkenhol, 2012; Reed and Frankham, 2003). In this context, a multifaceted approach that combines neutral and phenotypic data should provide a more comprehensive picture of the genetic structure, eventually leading to better conservation management.

Phenotypic variation among individuals results from the interaction between genotype and environment (Kawecki and Ebert, 2004). In the absence of other forces, populations are expected to develop traits that provide an advantage under their local

environment resulting in a pattern where resident genotypes are better fitted to their local conditions than genotypes from other habitats (Williams, 1996). This pattern is known as local adaptation (Ashton and Mitchell, 1989). Nevertheless, local adaptation may be hindered by gene flow, confounded by genetic drift, and constrained by a lack of genetic variation (Lenormand, 2002). Disentangling whether the variation observed in quantitative traits is inheritable or results from phenotypic plasticity is challenging because genotypes cannot be directly inferred from observed phenotypes (Frankham *et al.,* 2010). Instead, reciprocal transplants are required to evaluate the relative contribution of phenotypic plasticity and genetics (Kawecki and Ebert, 2004).

From a conservation perspective, rare and/or endemic plants are of great concern because of their intrinsic characteristics: small population size, habitat specificity, and geographic isolation (Frankham *et al.,* 2010). These features can be detrimental for the evolutionary potential of the species due to low genetic diversity, strong genetic drift, and inbreeding depression (Cole, 2003; Frankham *et al.,* 2010; Höglund, 2009; Willi *et al.,* 2006). However, rarity is only one of several factors known to have an impact on the species' genetic structure. Life history traits, particularly life form and breeding system, have long been recognized as greatly influencing the distribution pattern of genetic diversity in plant populations (Hamrick and Godt, 1996). Namely, selfing species can maintain lower levels of genetic diversity and higher levels of differentiation among populations compared to outcrossers (Nybom, 2004; Hamrick and Godt, 1996).

The small annual *Omphalodes littoralis* spp. *gallaecica* M. Laínz (1971) is a rare herb (total occupancy <100000 m^2) restricted to coastal dune systems in northwest Iberian Peninsula (Romero Buján, 2005; Serrano and Carbajal, 2011) (Fig. 1). In the last decades, its populations experience continuous decline due to the threats faced by its sensitive habitat (Bañares *et al.,* 2004); as a result, its current distribution is extremely fragmented and today the plant is known to occur in just five dune systems. Because

of this rarity, *O. littoralis* spp. *gallaecica* is catalogued as "endangered" by both the IUCN and the Spanish Catalogue of Threatened Species (Serrano and Carbajal, 2011) (Ministerio de Medio Ambiente y Medio Rural y Marino, 2011), and listed as a priority species in the EU Habitats Directive (92/43/EEC, Annex II). Additionally, its habitat is considered as a Site of Community Importance (SCI) within the Natura 2000 network. *O. littoralis* spp. *gallaecica* is a self-compatible plant and autogamy has been suggested as the most probable mechanism of reproduction (Bañares *et al.,* 2004). Flowering period is very short, from March to April, and the ephemeral flowers last less than three days (Romero Buján, 2005). Seed are thought to be dispersed by animals through the adhesiveness of the fruit to their hair (Bañares *et al.,* 2004). Population size fluctuates greatly between years, multiplying or dividing by ten the number of individuals (Bañares *et al.,* 2004).

Fig. 1. Detail of *Omphalodes littoralis* spp. *gallaecica* with flower and its typical habitat. Picture belongs to Baldaios' dune system.

Despite the status of *O. littoralis* spp. *gallaecica* as a species of conservation concern, its population genetics and the variation of its ecophysiological traits between populations have never been addressed. Here, we aim to fill this gap with our knowledge with an exhaustive molecular and phenotypic study of the five extant populations of this rare herb. We used sequences of the chloroplast DNA *trnT-F* region and genotypes derived from mostly-nuclear Amplified Fragment Length

Polymorphism (AFLP) markers to address the following questions: a) is *O. littoralis* spp. *gallaecica* genetically impoverished as it might be suggested by its life history traits?; b) are its populations significantly differentiated from each other?; c) given that *O. littoralis* spp. *gallaecica* is a therophyte, are there significant between-year differences in its genetic structure? On the other hand, we performed a series of reciprocal transplant experiments to investigate the adaptive component of several quantitative traits related to fitness. Phenotypic variation was examined with an aim to answer: d) are there any phenotypic differences between populations?; if so, e) do these differences result from phenotypic plasticity or do they have a genetic basis?; f) are they adaptive?. Finally, we combined the molecular and phenotypic information to propose specific guidelines for the conservation of this endangered plant.

MATERIAL AND METHODS

Sample collection and DNA extraction

Samples for genetic analyses were collected on two consecutive years (2009 and 2010). In March 2009, plants (31-34 per site) were randomly sampled along the whole area occupied by the species at each of the five dune systems currently inhabited by *Omphalodes littoralis* spp. *gallaecica* (see Fig. 2 in results). One year later, sampling was repeated at three of the systems (DN, BD, and XN). Sampling was non-destructive to meet the requirements of regional conservation authorities; only two-three leaves per individual were collected, dried in silica gel, and stored at -20°C until DNA extraction. DNA was extracted using the Wizard Magnetic Kit (Promega) and DNA extracts were further purified with PowerClean DNA Clean-up Kit (Mobio, CA, USA) following manufacturers' protocols. The quality of the extracted DNA and negative controls were systematically checked on 1.5% agarose gels.

AFLP analyses and cpDNA sequencing

Since AFLP performance can be sensitive to reaction conditions (Bonin *et al.,* 2004), we used several control measures to guarantee the reproducibility of our AFLP fingerprints. First, selective primer combinations were chosen after screening twenty-four pairs of primers with three selective bases on 20 individuals (4 individuals per sampling site). The entire process was repeated with new, independent DNA extractions of the same individuals to assess reproducibility. Nine primer combinations were chosen due to their reproducible and easily scorable profiles (EcoRI/TruI: TCA/CAT, TAC/CAT, TAC/CAA, TCC/CTG, TAG/CTG, TCT/CTA, TCT/CAT, TGC/CAG and TGC/CAT). Second, replicate DNA extractions were obtained for 10% of the individuals used in the study (evenly distributed among the 5 sampling sites) and run in parallel with the other DNA samples to monitor reproducibility. Samples and replicates were run in a blind-manner to avoid any bias during scoring. Individuals from each sampling site were evenly partitioned between the various 96-well plates used for PCR while replicates and originals were always run in separate plates; both samples and replicates were randomly distributed within plates. Third, a negative control with no sampled tissue added was included in each set of DNA extractions (24 samples) and went through the entire genotyping procedure. The estimated genotyping error (0.5%) was consistent with results of reproducibility tests conducted for AFLP both in plants and animals (Bonin *et al.,* 2004); none of the individual loci exceeded the maximum acceptable error rate (10%) recommended by Bonin *et al.* (2007).

AFLP analyses were performed according to Vos *et al.* (1995) with minor modifications and using nonradioactive fluorescent dye-labeled primers. Approximately 250 ng of genomic DNA were digested at 37°C for 3 hours in a final volume of 20 µl with 1.25 units of EcoRI and TruI (Fermentas) and 2x Tango Buffer (Fermentas). Digested DNA was ligated for 3 hours at 37ºC to double-stranded

adapters (50 pmols of adaptors E, 5'-CTCGTAGACTGCGTACC-3' and 5'-AATTGGTACGCAGTCTAC-3', and M, 5'-GACGATGAGTCCTGAG-3' and 5'-TACTCAGGACTCAT-3') using 0.5 units of T4 DNA ligase (Fermentas). Then, 2 µl of the ligation product was pre-amplified with 0.3 µM of each single selective primer (EcoRI-T and TruI-C), 2.5 mM MgCl2, PCR buffer 1x (Applied Biosystems), 0.8 µM dNTPs, 0.04 µg/µl BSA, 1M Betaine and 0.4 units of Taq polymerase (Applied Biosystems) in a final volume of 20 µl. Amplification conditions were 2 min at 72°C; 2 min at 94°C; 20 cycles of 30 s at 94 °C, 30 s at 56°C, and 2 min at 72 °C; and a final extension of 30 min at 60°C. Pre-amplification fragments were diluted 1:5 with Milli-Q water; 2.5 µl of the resulting solution were selectively amplified using 0.6 µM of the selective primers, 0.8 µM dNTPS, 2.5 mM MgCl2, 0.04 µg/µl BSA, PCR Buffer 1x (Applied Biosystems) and 0.4 units of AmpliTaq Gold polymerase (Applied Biosystems) in a final volume of 10 µl. Selective amplification was performed as follows: 4 min at 95°C; 12 of cycles of 30 s at 94°C, 30 s at 65ºC (first cycle, then decreasing 0.7°C for each of the last 11 cycles), and 2 min at 72°C; 29 cycles of 30 s at 94ºC, 30 s at 56ºC, and 2 min at 72ºC; and a final extension of 30 min at 72°C. Digestion, ligation, and PCR reactions were performed in a PxE thermal cycler (Thermo Fisher Scientific Inc., Waltham, MA, USA). Selective amplification products were electrophoresed on an ABI 3130xl automated DNA (Applied Biosystems) sequencer with HD-500 as size standard (Applied Biosystems). Fragments from 70 to 400 bp were manually scored for presence/absence at each selected locus with the help of GeneMarker v.1.70 (SoftGenetics LLC, State College, PA, USA) following common recommendations (Bonin *et al.,* 2005). Scores of the nine primer combinations were assembled into a single binary data matrix.

The *trnT-F* region was sequenced according to Taberlet (1991) with minor modifications. PCR reactions for the intergenic spacer between *trnT-trnL* were performed in 25 µl using primers a and b (Taberlet *et al.,* 1991), containing 1x reaction

buffer, 2 mM $MgCl_2$, 0.2 of each dNTP, 0.5 µM of each primer, 1 µl of genomic DNA and 1.25 units of DNA polymerase (Applied Biosystems). The *trnL* intron and the intergenic spacer *trnL-trnF* were amplified using primers c-d and e-f, respectively. PCR mix incorporated 1x reaction buffer, 1.5 mM $MgCl_2$, 0.2 of each dNTP, 0.5 µM of each primer, 1 µl of genomic DNA and 0.35 units of DNA polymerase (Applied Biosystems). PCR profiles consisted of 2 min denaturation at 94°C followed by 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 50° C and 90 s of extension at 72°C with a final elongation of 3 min cycle at 72°C. PCR products were visualized on 1.5% agarose gels and purified with 1 µl of Exonuclease I (20 u/µl) and 2 µl of FastAP (10 u/µl) (Fermentas). Purified PCR products were bi-directionally sequenced on an Automatic Sequencer 3730XL (Applied Biosystems, USA) following manufacturer's recommendations. Sequences were trimmed with CodonCode Aligner (CodonCode Co., MA, USA) and aligned using Clustal-W (Thompson *et al.,* 1994) implemented in DnaSP v 5.0 (Librado and Rozas, 2009; Rozas *et al.,* 2003).

Reciprocal transplants and phenotypic measures

In May and June 2009, seeds were collected from at least 40 randomly selected native plants growing in each of the five dune systems (sites). Seeds from each site were bulked and stored at 8º C in a cool chamber until sowed in November 2009. At each and every site, reciprocal transplants were accomplished by sowing seeds from the five origins in 10 haphazardly selected small plots. Plots where arranged into three/four areas within each site, covering all the possible environmental variability. Before sowing, the first 10 cm of the top soil of each plot were carefully inspected and any native *Omphalodes littoralis* spp. *gallaecica* seed was removed. Sowing plots consisted of shallowly buried plastic trays with 60 alveoli filled with local soil; alveoli (2 cm x 2 cm x 2 cm) were tagged according to the provenance of their seeds. Twelve seeds per origin were randomly sowed per tray (60 seed per tray; 600 seeds per site, 120 from each origin). Considering that sand dune species are reported to achieve

maximum germination rates when buried 0.5-4 cm deep (Maun, 1994), seeds were sown two centimeters deep. The low depth of the alveoli allowed interactions among the root systems of the plants.

From the date of sowing until the end of the life cycle of *O. littoralis* spp. *gallaecica* in late May-early June (precise date varies with provenance), the experimental sites were visited at least once per month to record germination, establishment, and survival. Visit frequency increased as necessary at the time of fruiting to collect the new seeds before dispersal. At the end of the growing season, plants were individually harvested, transported to the laboratory, and separated into roots, shoots and reproductive mass. Roots were washed and all plant material was oven-dried at 35º C until constant weight to the nearest 0.0001 g (Mettler AJ100, Griefensee, Switzerland). Stem DW (dry weight) combined stems and leaves, reproductive DW included calyxes and seeds, while shoot DW included all above-ground biomass (i.e. stems, leaves, and reproductive biomass). Total DW encompassed root and shoot DW.

Data analysis

For data analyses, plants from each dune system were considered members of a putative population. With AFLP markers, genetic diversity was estimated with the help of GenAlex 6.41 (Peakall and Smouse, 2006) as the percentage of polymorphic bands (5% criterion), the expected heterozygosity (H_e) (equivalent to Nei's gene diversity), and the Shannon-Weaver Index (H_{SW}). Private bands unique to a single population were also detected with GenAlex 6.41 (Peakall and Smouse, 2006). Since autogamy has been suggested as the most probable mechanism of reproduction of O. littoralis subsp. gallaecica, the former estimates were complemented with measures of genotypic diversity based on the frequency of distinct multi-locus genotypes. Potential clone mates, i.e. individuals with identical banding pattern, were identified with the software GenoType (Meirmans and Van Tienderen, 2004). Since rates for

somatic mutations are difficult to determine for natural populations (Douhovnikoff and Dodd, 2003), the genotyping error rate estimated in our reproducibility tests was set as the threshold value for genotype detection (maximum distance between two individuals at which they are still assigned to the same genotype). Genotypic diversity was estimated with the help of GenoDive (Meirmans and Van Tienderen, 2004) as number of genotypes (G), effective number of genotypes ($G_{eff}=1/\sum p_i^2$, where pi is the frequency of each i genotype), proportion of distinguishable genotypes, (G/N, where N is the number of individuals), genotypic diversity ($G_d=(n/n-1).(1-\sum p_i^2)$, where n is the sample size), and evenness (Eve = G_{eff}/G).

The partitioning of the genetic diversity and the occurrence of differences between years were evaluated by the analysis of molecular variance (AMOVA) (Excoffier *et al.,* 1992) implemented in GenAlex 6.41 (Peakall and Smouse, 2006). Its significance was tested by 9999 random permutations of individuals among populations/generations and genetic variation was expressed as Φ_{PT}, an analogue of F_{ST}. Population structure was further investigated by calculating the pairwise simple-matching dissimilarities between populations and depicted in a Principal Coordinates Analysis (PCoA) as in Kloda *et al.,* (2008). Also, the correlation between pairwise Φ_{PT} statistics and log-transformed geographic distances was assessed with the Mantel test (10000 bootstrap randomizations) implemented in the Isolation by Distance Web Service (Jensen *et al.,* 2005). Finally, the arrangement of genetic differentiation was further investigated with the individual-based Bayesian approach implemented in BAPS 5.3 (Corander *et al.,* 2008). The option for clustering of individuals was run 3 times for each of K = 1–20. The optimal partition determined by the program was used to estimate the levels of genetic admixture of each individual (with 200 reference individuals simulated for each genetic group and each original individual analyzed 20 times).

The *trnT-trnF* region of the cpDNA amplified in this study contains two intergenic spacers, *trnT-trnL* and *trnL-trnF*, and the *trnL* intron (Taberlet *et al.,* 1991). Given the non-recombinant nature of cpDNA, the three fragments were combined into a single sequence for each individual. The various distinct haplotypes found in our data set were identified with the help of Geneious v.6.1.6 (Biomatters Ltd., Auckland, New Zealand). The phylogenetic relationships between haplotypes were inferred using the median-joining algorithm implemented in Network 4.6 (Bandelt *et al.,* 1999). For the phylogeographic reconstruction, indels were treated as a fifth state (Simmons *et al.,* 2007). Population diversity estimated as haplotype diversity (H_d) and nucleotide diversity (π) was calculated with Arlequin 3.5 (Excoffier *et al.,* 2005) while intra-population genetic diversity (h_s) and total genetic diversity (h_t) were estimated using Permut (Pons and Petit, 1996). The contribution of each population to the total haplotype diversity (C_T) and the total haplotypic richness (C_T^r) were estimated with Contrib (Petit *et al.,* 1998). C_T and C_T^r were partitioned into two components, the contribution due to a population's own level of diversity (C_S and C_S^r), and its differentiation from other populations (C_D and C_D^r), respectively.

Population structure was again estimated by an analysis of molecular variance (AMOVA) based on haplotype frequencies (Excoffier *et al.,* 1992) and its significance assessed by calculating the F_{ST} statistic (after 1023 permutations) (Excoffier *et al.,* 2005). Since N_{ST} estimates significantly higher than G_{ST} values suggests the presence of phylogeographic structure, the software Permut (Pons and Petit, 1996) was used to estimate the G_{ST} statistic based on haplotype frequencies and N_{ST} values based on both haplotype frequencies and distances between haplotypes (number of mutational steps). Finally, the correlation between geographic and genetic distance was inferred using a Mantel test implemented in the IBD web service (Bohonak, 2002) and its significance was determined after 10000 randomizations.

The various phenotypic traits measured in the reciprocal transplant experiments were analyzed with a split-plot mixed-model design to test for differences among populations. The linear model tested was $y_{ijklm} = \mu_{ijk} + \gamma_l + e(PC)_{kl} + e(SP)_{ijklm}$, where *i* indicates provenance (*i*=1,…, 5), *j* represents transplant site (*j* =1,…,5), *k* indicates the area within each site (*k* =1,…,4), ℓ indicates tray (ℓ = 1,…,10) and *m* is each individual observation (*m*=1,…,n). y_{ijklm} is the individual value for a variable, μ_{ijk} is the mean for the variable at the *ijk* treatment, γ_l indicates the effect of each ℓ tray where $\gamma_l \approx N(0, \sigma_\gamma)$, $e(PC)_{kl}$ is the random error due to the plot where $e(PC)_{kl} \approx N(0, \sigma_{PC})$, and the last term in the model refers to the random error caused by the split where $e(SP)_{ijklm} \approx N(0, \sigma_{SP})$. Given the controversy about the pattern of deme x habitat interaction that should be taken as diagnostic for local adaptation in reciprocal transplants, we followed the two criteria proposed by Kawecki and Ebert (2004). First, we tested the "*local vs. foreign*" hypothesis that compares demes within habitats: should local adaptation occur, "*local"* demes are expected to perform better than demes from other habitats ("*foreign*" demes). Second, we tested the "*home vs. away*" criterion that compares a deme's fitness across habitats: should local adaptation occur, demes should perform better when growing at their own habitat ("*home"*) than at others ("*away"*). Although both criteria were examined, the "*local vs. foreign*" test provides more convincing evidence of local adaptation because the "*home vs. away*" test may confound the effects of divergent selection with intrinsic differences in habitat quality (Kawecki and Ebert, 2004). In the "*local vs. foreign*" tests, we considered the error caused by origin, area, and tray while error in the "*home vs. away*" tests included sites, area, and tray. Significance of the interactions (p-value <0.05) was always tested with the Tukey's Studentized Range (HSD) (Montgomery, 2008) after Bonferroni correction (Wright, 1992). Analyses were conducted with the statistical software R v. 3.0.1. (R Development Core Team, 2013) using packages *nlme* and *lsmeans.*

RESULTS

Genetic diversity and structure

A total of 276 reproducible AFLP markers were scored in the 165 individuals sampled in 2009. Eighty-one (29.35%) loci were segregating for the whole data set and were retained for diversity estimates. Overall, 26 private bands were detected in all populations: one in population DN; two in BD, PC, and TC each; and 19 in XN (Table 1). Estimates of total genetic diversity for the species (H_e=0.356; H_{SW}=0.530) were one or two orders of magnitude higher than the values observed at individual populations where diversity was consistently low. The various indices of genetic diversity were correlated across populations: diversity was low at DN (20.99% polymorphic loci, H_e=0.069, H_{SW}=0.104), very low at PC and TC (1.23% polymorphic loci, H_e=0.006, H_{SW}=0.008 and 3.70% polymorphic loci, H_e=0.011, H_{SW}=0.016, respectively), and zero at BD and XN.

Table 1: Genetic diversity in *Omphalodes littoralis* subsp. *gallaecica* based on AFLP data.

Pop	N	PLP	P_b	H_e (±SE)	H_{SW}	*G*	G_{eff}	*G/N*	G_d	*Eve*
DN	34	17 (20.99)	1	0.069 (±0.017)	0.104 (±0.025)	33	32.11	0.97	0.99	0.97
BD	34	0 (0.00)	2	0.000	0.000	1	1.00	0.03	0.00	0
PC	34	1 (1.23)	2	0.006 (±0.006)	0.008 (±0.008)	2	1.84	0.06	0.47	0.92
TC	30	3 (3.70)	2	0.011 (±0.008)	0.016 (±0.011)	3	2.76	0.10	0.66	0.92
XN	33	0 (0.00)	19	0.000	0.000	1	1.00	0.03	0.00	0
Total	165	81 (29.35)	26	0.356 (±0.016)	0.530 (±0.018)	40	8.42	0.24	0.89	0.21

N, number of individuals; PLP, percentage of polymorphic loci (under 5% criterion); P_b, number of private bands (percentage for the total data set based on 276 scorable loci); H_e; Expected Heterozygosity (± standard error); H_{WS} Shannon-Weaver Index (± standard error); *G*, number of genotypes; G_{eff}, number of effectives genotypes; G_d, genotypic diversity; *Eve*, evenness of the effective number of genotypes. Nei's gene diversity was calculated using segregating fragments only.

The 165 individuals only produced 40 distinct genotypes (G_{eff}=8.42, *G/N*=0.24); moreover, most individuals shared just seven genotypes, explaining the low evenness recorded at species level (*Eve*=0.21).Nevertheless, none of the genotypes detected in the study occurred in more than one population so that each

local deme had a distinctive set of AFLP genotypes. Genotypic diversity echoed the changes between populations seen above for genetic diversity. However, while genetic diversity was consistently low across populations, genotypic diversity in DN could be described as high as almost every individual sampled at this site exhibited a distinct genotype (G=33, G_{eff}=32.11, G/N=0.97, G_d=0.99). In contrast, most of the individuals sampled at the other four dune systems shared just one (BD, XN) or a very few (two in PC, three in TC) genotypes producing very low estimates of the G/N ratio at these sites (<0.10). Nonetheless, the high evenness recorded at PC and TC (0.92) indicates that the few haplotypes found on these sites were evenly partitioned among individuals.

Genetic differentiation was extremely high and almost reached the theoretical limit of one (Φ_{PT} = 0.963, P < 0.0001), indicating that nearly all the genetic variation (96%) was due to differences between populations. Pairwise comparisons were likewise high and significant (Φ_{PT} > 0.79 and P < 0.05 after Bonferroni correction for each and every pairwise comparison). The most diverse population, DN, displayed the lowest pairwise Φ_{PT} values while the southernmost and relatively isolated XN showed the highest differentiation (Φ_{PT} > 0.94). A PCoA plot based on genetic distances among individuals (95.50% of variation explained by the first two axes, Fig. 3) revealed the three well-resolved groups that seemed consistent with the geographical placement of their population of origin. Thus, the genotype found at the southernmost site XN (33 individuals with identical AFLP genotype) was clearly separated from those recorded at other sites, echoing the very high pairwise Φ_{PT} values estimated for this population. Likewise, the remaining four demes were arranged into two groups of geographically consecutive sites (BD-DN and PC-TC, respectively). Despite the apparent correlation between genetic distance and geographical position suggested by the PCoA, the Mantel test found no evidence of isolation-by-distance (r = 0.0462, Mantel P = 0.5323). As for changes over time, when the same set of AFLP markers was

scored in samples collected one year later at three of the sites (DN, BD and XN), the genetic structure and diversity were nearly identical to those obtained in 2009 to the point that AMOVA revealed non-significant differences between years (Φ_{PT} = -0.009, P = 0.931).

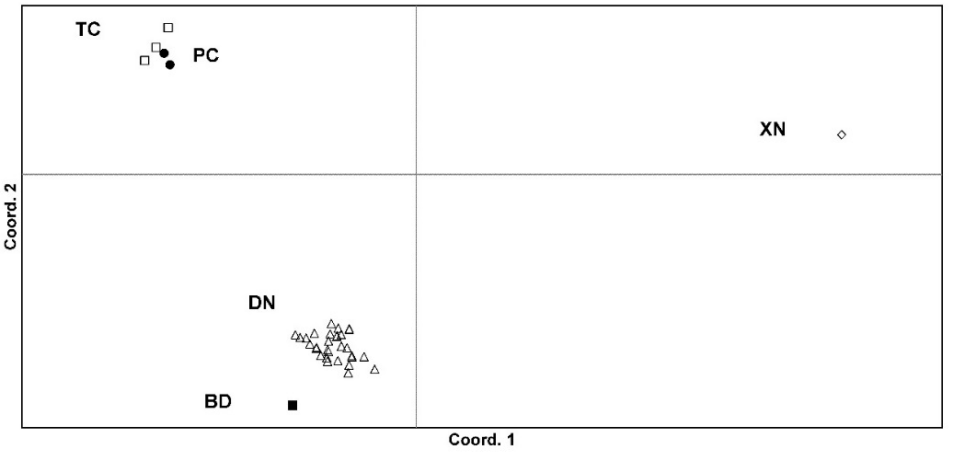


Fig. 3: Principal Coordinates Analysis calculated from simple-matching pairwise distances between individuals of *Omphalodes littoralis* spp *gallaecica* collected at five dune systems and scored with 81 segregating AFLP loci.Individuals coded by sampling site: TC, open squares; PC, filled circles; BD, filled squares; DN, open triangles; XN, open diamonds. Individuals with identical AFLP genotype appear as a single symbol. Together, coordinates 1 and 2 explain 95.50% of the total variation.

The individual-based Bayesian analysis corroborated the results obtained with the population-based approaches confirming that most of the genetic variation occurred among populations. In BAPS, the optimal partition identified five genetic groups that perfectly matched the five sampling populations (log-likelihood value = -1267.78, probability for 5 clusters = 0.9996). Moreover, no sing of genetic admixture was detected for any individual (Fig. 2).

Among the three non-coding fragments sequenced for the *trnT-trnF* region, only the intergenic spacer *trnT-trnL* was polymorphic. Therefore, the *trnL* intron and the intergenic spacer *trnL-trnF* were excluded from further analyses. The alignment of the *trnT-trnF* fragment resulted in a final consensus sequence of 762 pb. Sequences

were rich in A and T nucleotides, with A/T content of 73.80%, in accordance with the nucleotide composition of non-coding chloroplast regions (Kelchner, 2000). One point mutation and two indels of 11 pb and 22 pb, respectively, defined four haplotypes. The phylogenetic relationships among haplotypes shown by the parsimony network displayed a star-like shape with haplotype H1 in a central position (Fig. 2).

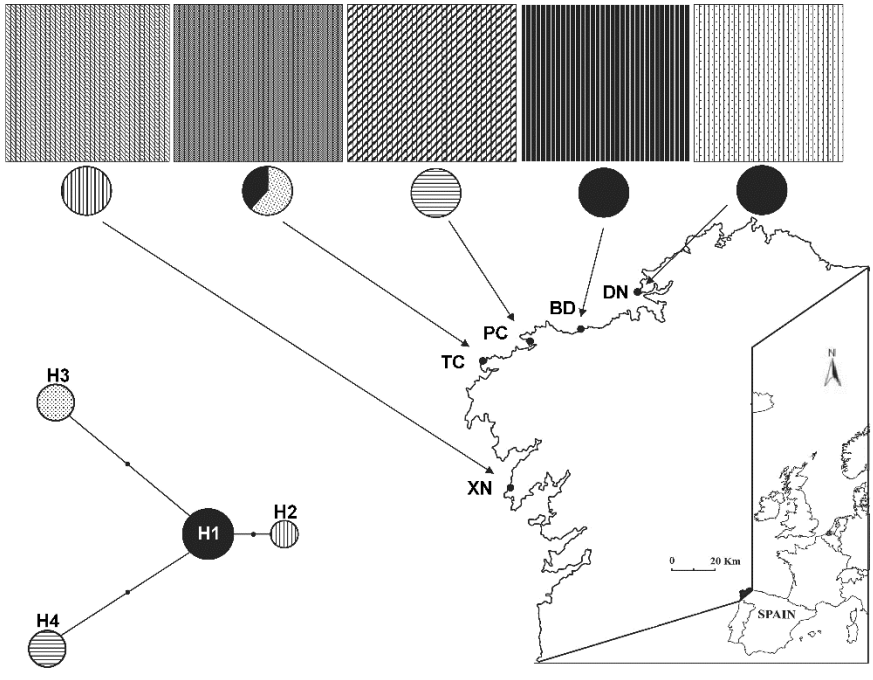


Fig. 2: Sampling sites, genetic structure based on AFLP genotypes, and cpDNA haplotypic network of *Omphalodes littoralis* subsp. *gallaecica*. Range occupancy is strongly fragmented into very small enclaves. Locations: DN, DB, PC, TC and XN. The histogram on the top depicts individual assignment by an admixture analysis performed for an optimal number of 5 genetic clusters (*P*=0.9996) using AFLP genotype data. Each vertical bar represents one individual with patterns indicating the probability of assignment to each cluster. Pie charts show the relative abundance of four cpDNA haplotypes (H1-H4) in each population; patterns match the haplotype median-joining network shown on the bottom-left. Circle size in the network is proportional to haplotype frequency across populations; black-dots indicate mutational steps.

Haplotype H2 was separated from the central H1 by one mutational step, while both H3 and H4 were separated from H1 by two relatively large indels each (11-bp

long in H3, 22-bp long in H4). Most populations showed a single cpDNA haplotype except TC where two were detected. The central haplotype H1 also was the most abundant (nearly 47% of the individuals) and the most widely distributed. Unlike the other haplotypes, H1 was detected at three sites while H2, H3 and H4 were restricted to XN, TC, and PC, respectively.

Estimates of total genetic diversity for the species based on cpDNA were H_d = 0.687, $\pi x10^2$= 1.154, h_s= 0.095 and h_t= 0.829 respectively (Table 2). Population diversity was even lower than that recorded with AFLP. Four out of five populations were dominated by a single haplotype and their within population diversity was zero. Interestingly, the set of demes with no cpDNA variation included DN, the only site where almost each individual displayed a distinctive AFLP genotype. On the other hand, the only location with two haplotypes (TC) exhibited intermediate to high values of haplotypic and nucleotide diversity (H_d = 0.473, $\pi x10^2$=1.386) because its two haplotypes were evenly partitioned among individuals.

Table 2: Genetic diversity measures of *Omphalodes littoralis* subsp. *gallaecica* based on cpDNA.

Population	N	S	H	H_d (SD)	$\pi x10^2$ (SD)
DN	32	0	1	0.000 (0.000)	0.000 (0.000)
BD	31	0	1	0.000 (0.000)	0.000 (0.000)
PC	32	0	1	0.000 (0.000)	0.000 (0.000)
TC	31	22	2	0.473 (0.054)	1.386 (0.723)
XN	32	0	1	0.000 (0.000)	0.000 (0.000)
Total	158	34	4	0.687 (0.023)	1.154 (0.593)

N, number of sampled individuals; *S*, number of segregating sites; *H*, number of haplotypes; H_d, haplotypic diversity; and $\pi x10^2$, nucleotide diversity.

As for the contribution to haplotypic diversity and richness, some populations clearly contributed more than others (Fig. 4). Three populations contained all the cpDNA haplotypes detected in the species and, consequently, they were the only ones with a positive total contribution to haplotypic diversity (PC, XN) and richness (PC, XN,

and TC). Their positive contribution was mostly due to their differentiation from other populations (components C_D and C_D^r) rather than to their own level of diversity (components C_S and C_S^r). The latter reflects the fact that each population was mostly (TC) or totally (PC, XN) dominated by a private cpDNA haplotype. In comparison, the contribution of the two northernmost populations (BD and DN) was from negative (diversity, C_T) to negligible (richness, C_T^r) because they only contained the widespread haplotype H1 that was occurred in TC.

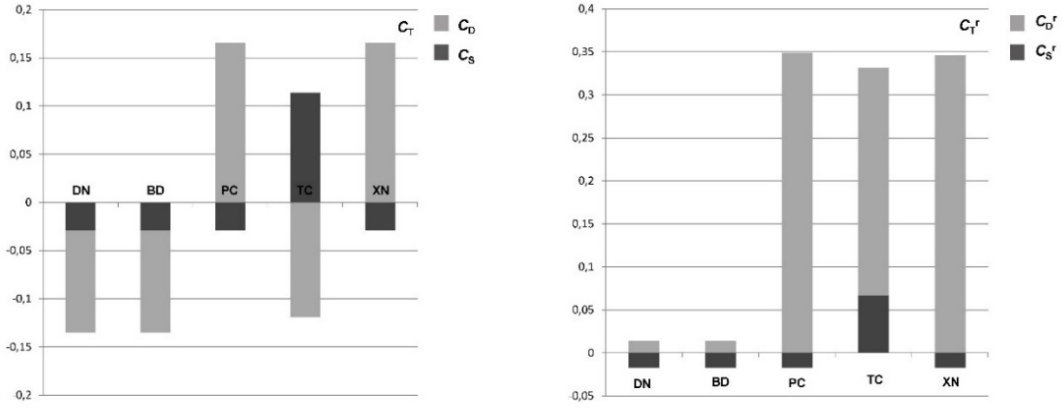


Fig. 4: Contribution to total cpDNA haplotype diversity (left, C_T) and haplotypic richness (right, C_T^r) of each population of *Omphalodes littoralis* spp *gallaecica*. Grey and black bars represent the contribution due to the diversity (C_S and C_S^r) and differentiation (C_D and C_D^r) of each population.

As seen with the AFLP genotypes, AMOVA revealed that most of the cpDNA variation (80.44%) was due to differences among populations, rendering a very high and significant F_{ST} estimate (0.804, P<0.001). Also, F_{ST} values were always high and significant except for the comparison DN-BD, two populations dominated by the same haplotype (H1). No evidence of phylogeographic structure was detected because the magnitude of population differentiation inferred from haplotype frequencies (G_{ST}=0.886) was not significantly different (P>0.05 after 1000 permutations) from the level inferred taking haplotype divergence into account (N_{ST}=0.873). Likewise, the

Mantel test found no support to an isolation by distance pattern (r=0.048; P=0.515 after 10000 randomizations).

Phenotypic analysis

Some trays were lost due to vandalism meaning that only 4 trays in XN, 7 in PC, and 8 in TC reached the end of the experiment. The GLM analysis showed that the partition of trays into several areas per site had no significant influence on the values of the various phenotypic traits with the only exception of mean seed DW (Table 3). Therefore, GLM analyses were repeated ignoring the arrangement into areas except for the latter variable. These analyses revealed significant differences between transplant sites for most variables suggesting that our plants performed better in some dune systems than in others. An examination of the mean values recorded at each transplant site revealed no obvious pattern (Fig. 5), although several variables (seed no., reproductive DW, total DW) seem to have reached higher values in the two southernmost sites.

Table 3: General linear model, "l*ocal vs. foreign*" and "*home vs. away*" tests for the quantitative traits investigated in the reciprocal transplants of *Omphalodes littoralis* subsp. *gallaecica*.

	GLM			*Local vs. Foreign*	*Home vs. Away*
	Area	Site	Origin	*Local vs. Foreign*‡	*Home vs. Away*‡
Seed number	NS	***	***	NS	NS
Mean seed weight (g DW)	***	***	**	NS	NS
Reproductive weight (g DW)	NS	***	***	NS	NS
Root weight (g DW)	NS	NS	***	NS	NS
Stem weight (g DW)	NS	**	***	NS	NS
Total weight (g DW)	NS	***	***	NS	NS

The effects of Area, Site and Origin are specified for the GLM. Significance is represented as NS (not significant), * ($0.05 \leq p \geq 0.001$), ** ($p<0.001$) and *** ($p<0.0001$). *Local vs. Foreign*‡ indicates that it has been corrected by the origin while *Home vs. Away*‡ represents that it has been corrected by the location of growth.

Provenance (origin) also had a significant influence on phenotype indicating that part of the variation seen at the various traits must have a genetic basis.

Regardless of the transplant site, the individuals from DN usually outperformed those from other origins producing more biomass and more seeds, even when the plants from other provenances were growing at their own site of origin (Fig. 5).

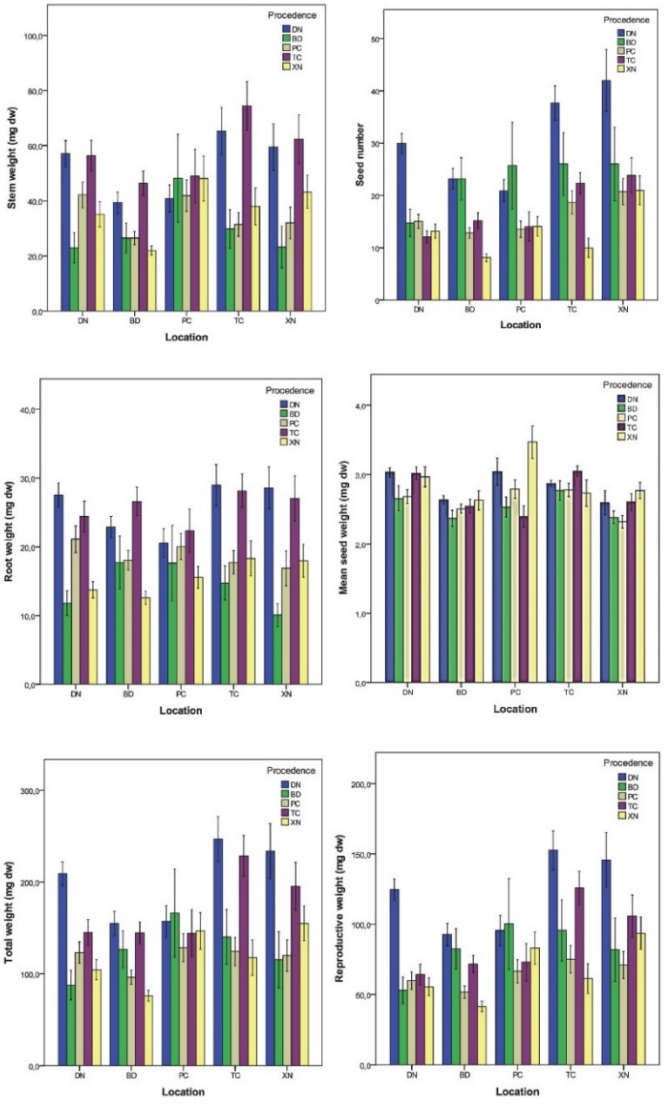


Fig. 5: Mean for the quantitative traits studied in *Omphalodes littoralis* subsp. *gallaecica.* Axis Y indicates the value of the studied phenotypic trait (from upper-left to the right-bottom: Seed number, Mean seed DW, Reproductive DW, Stem DW, Root DW and Total DW). Axis X represents the location of growth. For each location all possible origins are represented with colors (blue for DN, green for BD, grey for PC, purple for TC and yellow for XN). Each vertical bar represents the mean for a given phenotypic trait for a deme growing in a certain location and with a specific origin. The standard error is indicated in each vertical bar.

The outperformance of DN was particularly pronounced when growing at their site of origin (at the north edge of the distribution range of the species) or when they had been transplanted to the two southernmost sites (TC, XN). In fact, DN plants produced more seeds and grew better (reproductive and total DW) at TC or XN that at home. TC plants were second to those from DN in terms of biomass production (stem, root, and total DW) but not in seed production. Despite the significant differences detected between sites and between origins, neither the "*local vs. foreign*" nor the "*home vs. away*" tests found significant differences for any quantitative trait, providing no support to the predictions of the hypothesis of local adaptation in *Omphalodes littoralis* spp*. gallaecica.*

DISCUSSION

Taxa listed as endangered by the IUCN Red List of Threatened Species are considered to face a very high risk of extinction in the wild (IUCN 2012). In the particular case of *Omphalodes littoralis* spp*. gallaecica,* its status as endangered was granted attending to criteria of area of occupancy only: the plant occupies 10 hectares (well below the threshold of 500 km^2 used by IUCN for endangered species), this area of occupancy is in continuing decline due to many threats, and populations undergo extreme fluctuations (Serrano and Carbajal, 2011). Leaving aside the fact that the plant possibly meets the IUCN's criteria for a higher level of risk (Critically Endangered), we have found new reasons for concern about the mid- to long-term survival of this dune dweller. Our results strongly suggest that effective population sizes must be much smaller that census estimates. In fact, we found only 40 distinct genotypes among 165 genotyped individuals; to make things worse, three quarters of them were concentrated in a single local deme so that most populations contained one or very few distinct genotypes. Moreover, even the population with the highest number of genotypes showed very low genetic diversity indicating that its various genotypes were closely related to each other. Therefore, we think that the effective

abundance of this endangered plant is much smaller than previously thought and should be considered a further reason for concern.

The low levels of within-population variation recorded in *Omphalodes littoralis* spp. *gallaecica* are consistent with its life history traits. Annual selfing taxa such as *Omphalodes littoralis* spp. *gallaecica* usually display the lowest levels of within-population variation (Nybom, 2004). Also, various comparative studies have found that narrow endemics are often less diverse than widespread taxa (Cole, 2003; Gitzendanner and Soltis, 2000; Hamrick and Godt, 1990). Despite the above, the diversity shown by most of the extant populations of *Omphalodes littoralis* spp. *gallaecica* still is remarkably low. The estimates of H_e obtained with AFLP markers in four out of the five sites (range: 0.000-0.011) are one or two orders of magnitude below the average H_{pop} estimated for annuals and/or selfing plants using markers with the same mode of inheritance (Nybom, 2004). And the situation is even worse if we consider the variation displayed by the cpDNA because most populations seemingly contained a single haplotype.

The spatial arrangement of genetic variation is typically explained by contemporary (e.g. effective population size, gene flow) and historical (e.g. fragmentation, founder events) factors (Schaal *et al.,* 2003). AFLP markers are typically associated with recent processes while cpDNA is more often related to ancient history (Avise, 2004). In the particular case of *Omphalodes littoralis* spp. *gallaecica,* both AFLP and cpDNA suggest that gene flow must be very restricted. While acknowledging that caution must be exerted when drawing conclusions about gene flow based on Φ_{ST} (Marko and Hart, 2011; Whitlock and McCauley, 1999), the fact that an overwhelmingly majority of genetic variation was due to differences between populations is consistent with a scenario of restricted gene flow. Also, the occurrence of private ALFP markers at each and every population together with the fact that each population had its own ALFP genetic lineage in the Bayesian analysis

suggest that they must have been separated for a long time. This conclusion is reinforced by the analysis of the cpDNA variability where most of the haplotypes detected in our study were private to a single population and each population showed a distinct cpDNA composition except for the two northernmost sites (BD and DN). According to coalescent theory, central and widespread haplotypes such as H1 may be regarded as ancestral (Posada and Crandall, 2001). Thus, the occurrence of H1 in three non-adjacent populations possibly suggests that the various local demes might have been connected in a distant past. From a conservation perspective, the extreme fragmentation and isolation revealed by the lack of gene flow among local demes suggests that the genetic rescue of one population by others seems highly unlikely without external help.

The strong among-populations differentiation detected using markers with different mode of inheritance is again consistent with the life history traits of *Omphalodes littoralis* spp. *gallaecica*. Selfing taxa are known to partition most of their genetic variation to differences between populations rather than to variability among individuals within populations (Duminil *et al.,* 2007). Together with the extremely low within-population diversity showed before, this high differentiation among-populations suggests that this small plant could be reflecting the effects of genetic drift. The latter would be exacerbated if we recall that this narrow endemic typically shows strong fluctuations in population size indicating that the plant could experience recurrent bottlenecks over the years. The very low within-population diversity shown by *Omphalodes littoralis* spp. *gallaecica* is a matter of concern. Populations with low genetic diversity can be threatened by stochasticity, even by relatively minor events, and are less capable to cope with environmental changes and/or stressful conditions (Frankham, 2005). Furthermore, small populations that fall below a certain effective size may enter an “extinction vortex” where reproductive dynamics favor inbreeding leading to lower reproduction, increased mortality, and smaller population sizes. In

this regard, high levels of self-fertilizing and small fragmented populations have been shown to be related to inbreeding depression (Angeloni *et al.,* 2011; Leimu *et al.,* 2010). As inbreeding depression can lead to a decrease in the number of populations, often in an irreversible fashion, that may result in the extinction of the species (Lande, 1993), there are reasons to worry that the long-term survival of this already endangered plant might be threatened. Nevertheless, while inbreeding depression has negative consequences for plant fitness, its impact is known to be smaller in self-compatible than in obligate outcrossing species (Leimu *et al.,* 2006).

While the large fluctuations in population size experienced by many annuals could compromise their genetic diversity, other attributes of their life cycle can act in the opposite direction. Some annual taxa have a large reservoir of viable seeds from which individuals may be drawn in the future (Levin, 1990). In these cases, a stable seed bank could have an important role buffering against the genetic loss (McCue and Holtsford, 1998; Nunney, 2002). However, this seems not be the case in *Omphalodes littoralis* spp. *gallaecica*. In agreement with previous observations in other taxa (Honnay *et al.,* 2008), our analysis revealed that the local demes of *Omphalodes littoralis* spp. *gallaecica* maintain a constant genetic composition between consecutive years. Thus, the inability of the seed bank to act as a reservoir of hidden genetic diversity adds further concern to the long-term persistence of the species.

An interesting result of our study is the finding that populations separated by just a few kilometers show statistically significant differences in their quantitative traits. While this variation could simply be a phenotypic response to subtle changes in the local environment of each site, our reciprocal transplant experiments indicate it actually involves a genetic component. Unlike what would be expected in a scenario of local adaptation, the individuals from one site (DN) commonly outperformed those from the others regardless of the transplant location. Initially, there is no clear explanation to the better fitness of the plants from DN. The only obvious difference

between DN and the other populations is that the former displays higher levels of within-population genetic diversity. Therefore, it seems tempting to speculate that the increased performance of its individuals could be related to the higher variation detected using neutral markers. While a correlation between neutral genetic diversity and fitness is far from universal, it is widely accepted that a lack of diversity can lead to the deleterious effects of inbreeding (Angeloni *et al.,* 2011; Landguth and Balkenhol, 2012; Reed and Frankham, 2003).

Conventional wisdom assumes that self-compatible species are expected to display a strong adaptation to local conditions given their usually high levels of genetic differentiation (Leimu and Fischer, 2008). However, while the populations of *Omphalodes littoralis* spp*. galaecica* are strongly isolated from each other, the patterns of quantitative differences detected in our reciprocal transplant experiments do not match the expectations under local adaptation. Instead, the inheritable differences in quantitative traits detected among populations must result from processes other than local adaptation. In the absence of gene flow, local adaptation can be confounded by genetic drift and/or constrained by a lack of genetic variation (Kawecki and Ebert, 2004). This might be the case of *Omphalodes littoralis* spp*. galaecica* where the lack of evidence in support of local adaptation suggests that genetic drift might be responsible for the differences among demes in their quantitative traits. Also, the higher performance of the plants from DN suggests that this population may be particularly relevant for the preservation of the species.

From a conservation perspective, the criterion to select priority populations should consider its uniqueness and variation level with an emphasis on allelic richness (Petit *et al.,* 1998). Our cpDNA analysis revealed that three out of five populations cover the complete genetic variation of the species (PC, TC and XN) and should be designated at least as MUs (management units *sensu* Moritz, 1994). However, our results also indicate that cpDNA contains only a portion of the genetic history of the

species. The more variable AFLP markers showed that each population belonged to a different genetic lineage. Moreover, the AFLP results also revealed that DN is the population with the largest genetic variation even though its cpDNA diversity is zero and totally redundant with other sites (the only haplotype detected in DN also occurs in BD and TC). Therefore, and unlike the cpDNA results, the AFLP markers indicate that each and every extant population of *Omphalodes littoralis* spp. *gallaecica* should receive equal attention given their unique genetic composition; consequently, five rather than three conservation units should be designated, one per population. In fact, by a simple simulation exercise we can estimate the genetic loss derived from the disappearance of one population. Total gene diversity (H_e) decreases from 11.2% to 27.5% depending on which population is simulated to disappear. Eventually, it seems likewise reasonable to suggest that the five MUs should be designated as ESUs (evolutionary significant unit *sensu* Moritz, 1994) given the significant differences in inheritable quantitative traits detected among these populations. The proposal of five ESU is done while noticing that the differences in the quantitative traits among these ESUs are non-adaptive but a result of genetic drift. However, we still think that the occurrence of these differences indicate that the various local demes are not interchangeable and may have a different potential to evolve. In this regard, practices involving the translocation of individuals between sites are strongly discouraged because of the strong genetic isolation between the populations of this endangered therophyte (Sletvold *et al.,* 2012).

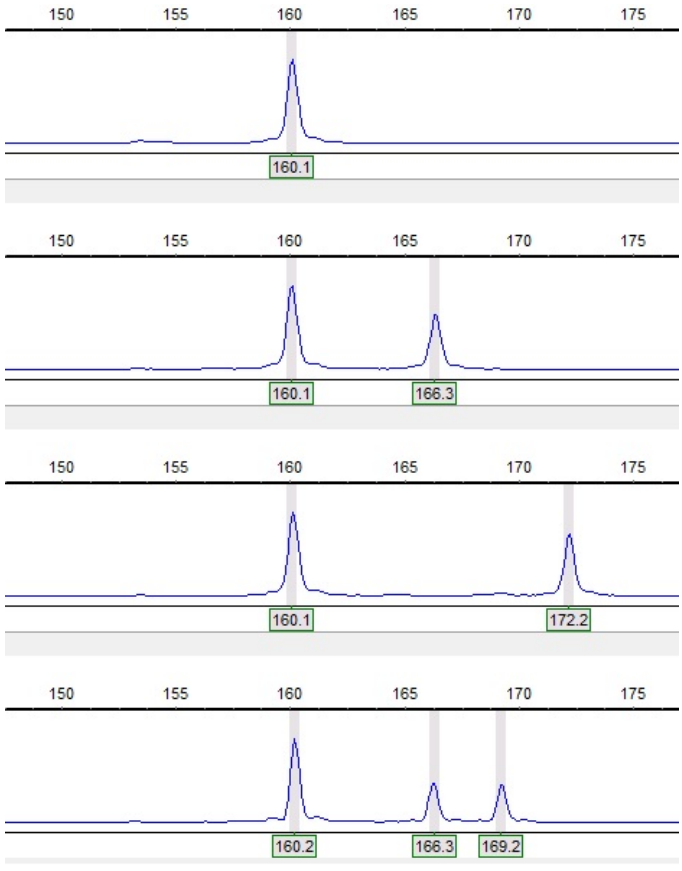
In summary, we have shown that by combining selfing with a strongly fragmented distribution, a narrow endemic plant can reach extremely low genetic variation within populations but high differentiation between local demes. Moreover, the various demes of *Omphalodes littoralis* spp. *gallaecica* also differ in their quantitative traits and these differences have a genetic basis, contradicting the initial assumption that populations living in a very narrow range under similar

environmental conditions should display a more homogeneous ecophysiology. Our reciprocal transplant experiments indicate that this variation in *O. littoralis* cannot be attributed to local adaptation. Instead, high rates of self-fertilization together with recurrent bottlenecks caused by dramatic interannual fluctuations in population size may have led to a decrease in genetic diversity in a classic scenario drawn by genetic drift. Regardless of the mechanism behind the pattern, the current arrangement of genetic diversity is of some concern from a conservation perspective. Effective population sizes are much smaller than previously thought while the lack of gene flow among local demes suggests that if the plant disappears from one dune system, recolonization without assistance is highly unlikely. The plants from the only deme with moderate genetic diversity consistently outperformed those from other populations with minimal to zero diversity, suggesting that the latter might have diminished their ability to cope with the environment. We recommend that each population should be designated as an independent ESU because of their distinctive genetic and phenotypic make-up. Eventually, our study highlights that range size, geographic distance, and homogeneous environment may not be accurate indicators to delineated conservation strategies.

ACKNOWLEDGEMENTS

Research supported by grant 07MDS031103PR (Xunta de Galicia). L. L. acknowledges support from Universidade da Coruña (contratos predoutorais UDC 2012).

CHAPTER 4



Mining molecular markers from public EST databases in the study of threatened plants

Published as: Lopez L., Koch M., Fisher M. and Barreiro R. Mining molecular markers from public EST databases in the study of threatened plants. Journal of American Botany. (submitted)

ABSTRACT

Simple Sequence Repeats (SSR) are widely used in population genetic studies but their *de novo* development is costly and time-consuming. The ever-increasing available DNA datasets generated by high-throughput techniques offer new and inexpensive alternatives for SSRs discovery. In particular, Expressed Sequence Tags (EST) have been used as a SSRs' source for plants of economic relevance but their application to non-model species has been overlooked. We explored SSRs discovery from publicly available EST databases (GenBank-NCBI) of non-model species, with special emphasis on threatened plants (all genera with available EST listed by the International Union for Conservation of Nature and Natural Resources). EST sequences of two model genera with fully annotated genomes, *Arabidopsis* and *Oryza*, served as controls for EST-SSRs genome distribution analysis. From a total of 14 498 726 EST sequences from 257 endangered genera, 17 076 SSRs from 222 genera had suitable primer information. Dimers and trimers were the prevalent repeats. Control genomes revealed that trimmers, together with hexamers, were mostly located in coding regions while dimers were largely associated with untranslated regions. Performance and transferability of EST-SSRs was tested in four species from two eudicot genera, *Trifolium* and *Centaurea*, finding considerable amplification success (41.67-66.67%) and very high (100%) transferability between congenerics. The high cross-species transferability suggests that the number of possible target species would potentially increase in a significant manner. Altogether, our study supports the use of EST databases as an extremely affordable and fast alternative for developing SSRs markers in threatened plants.

Keywords: conservation, EST-SSR, functional markers, population genetics, threatened plants.

CHAPTER 4

INTRODUCTION

Since the 1980s, the fast advent of molecular markers technology has revolutionized the field of genetics by changing the pace and accuracy of genetic analysis. Today, the analysis of DNA variation is a key component in plant genetics studies addressing relevant aspects such as evolution, phylogeny or conservation (Allendorf and Luikart, 2012; Frankham *et al.*, 2004; Höglund, 2009). Among the various types of molecular marker used for these purposes, Simple Sequence Repeats (SSRs) are often regarded as the markers of choice. Microsatellites or SSRs are short tandemly repeated DNA regions that are ubiquitous in pro- and eukaryote genomes (Morgante *et al.*, 2002; Tautz and Renz, 1984; Toth *et al.*, 2000). They are considered "ideal" markers because of their abundance, multiallelic behavior, high polymorphism and codominant inheritance (Ritland, 2000). Unfortunately, *de novo* development of SSRs is an expensive and time-consuming task (Squirrell *et al.*, 2003). Moreover, genomic SSR are usually species-specific, meaning that specific markers developed for one taxon cannot be directly transferred to another (Selkoe and Toonen, 2006).

With the recent and growing emphasis on structural functional genomics, the number of large datasets of DNA sequences generated by high-throughput technologies has greatly increased for a wide variety of taxa. In this context, Expressed Sequence Tags (ESTs) databases available for public use appear as an attractive alternative for SSRs mining and development (Ellis and Burke, 2007). Microsatellites generated from ESTs (EST-SSRs) display several advantages over those derived from genomic DNA. First, time and costs for SSRs development are considerably lower. Instead of the weeks required for SSRs development with conventional approaches, it takes 2-3 days to obtain a batch of EST-SSRs markers with primers from existing databases. Second, any type of SSR motif can be detected in EST-SSR mining while a subset of predefined motifs are favored in conventional approaches that involve an enrichment step. Third, SSRs have found to be moderately abundant (≈2-5%) in EST

sequences given their preferential association with the non-repetitive fraction of the plant genome (Morgante *et al.*, 2002; Kantety *et al.*, 2002). Finally, EST-SSRs located in conserved regions are highly transferable between related species, even across genera, because the conserved flanking sequences are ideally suited for primer design. Nevertheless, EST-SSRs also show some disadvantages. Their development is restricted to organisms with existing EST sequence data, although microsatellite mining from EST sequences of related species is a promising alternative. In addition, EST-SSRs are expected to display lower levels of polymorphism than anonymous SSRs as they are linked to conserved regions of the genome (Ellis and Burke, 2007; Varshney *et al.*, 2005a). Nonetheless, several studies with EST-SSRs found moderate to high levels of polymorphism (Aleksic and Geburek, 2014; Fraser *et al.*, 2004; Pashley, 2006). Finally, another possible concern is that EST-SSRs might bias the estimates of population divergence if one assumes a neutral model of drift, mutation and migration (Luikart *et al.*, 2003). However, Woodhead *et al.* (2005) reported that measures of population structure derived from ESR-SSRs were consistent with those from anonymous SSRs. In fact, several studies indicate that only a very small fraction of genes might have experienced recent positive selection (Tiffin and Hahn, 2002; Victoria *et al.*, 2011)

EST-SSRs can be considered “functional markers” because ESTs represent a portion of the transcribed region of the genome under certain conditions (Andersen and Lübberstedt, 2003; Varshney *et al.*, 2005a). For a majority of these markers, a “putative function” can be deduced by comparison against annotated reference genomes. EST-SSRs with dinucleotide motifs are known to be favored in Untranslated Regions (UTRs) and introns, while trinucleotides are frequent in coding regions (CDS) (Morgante *et al.*, 2002). Thus, compared with anonymous SSR, EST-SSRs offer the opportunity to detect variation in transcribed portions of the genome that could show a marker-trait association (Varshney *et al.*, 2005a). For example, contractions or

expansions in the 5' UTRs can alter the transcription or translation of their respective genes (Li *et al.*, 2004; Zhang *et al.*, 2006) while length variation in microsatellite loci located in 3' UTRs has been linked to gene silencing and expression levels of flanking genes, among others (Conne *et al.*, 2000; Thornton *et al.*, 1997). On the other hand, changes in coding regions may entail a change in function or, even, a loss of function (Li *et al.*, 2004).

To date, EST-SSRs markers have been successfully used for resolving phylogenies (Tabbasam *et al.*, 2013) and to increase resolution in comparative genetic mapping studies by cross-referencing genes between species (Varshney *et al.*, 2005b; Yu *et al.*, 2004). These studies have mostly focused on species of economic importance (i.e. crops) and model species (Aggarwal *et al.*, 2007; Blair and Hurtado, 2013; Fukuoka *et al.*, 2010; Gao *et al.*, 2003; Kantety *et al.*, 2002; Mishra *et al.*, 2011; Simko, 2009; Varshney *et al.*, 2005b). Surprisingly, there are very few examples in the literature on the use of EST-SSRs in threatened plants, despite the fact that they could be regarded as a potentially powerful tool for addressing conservation-related questions (Aleksic and Geburek, 2014; Liewlaksaneeyanawin *et al.*, 2004).

The present study explores a rather underexploited, yet clearly promising, application of EST-SSRs: developing markers from public EST databases for evolutionary and conservation genetic studies of non-model plant species, with a special emphasis in threatened ones. In particular, we searched all plant genera included in the International Union for Conservation of Nature and Natural Resources (IUCN) Plant Red List that had EST sequences available in the GenBank EST database (dbEST). Since most of these genera do not include model organisms, normally there are no available annotated reference genomes for comparison, thus hampering the location of the EST-SSRs within the genome (i.e. intergenic regions, introns, UTRs or exons). To minimize this obstacle, EST sequence data sets for two model genera with well-known annotated genomes were in-depth analyzed and used as a proxy.

Arabidopsis was selected as a control for eudicots while *Oryza* was used as a guide for monocots. Finally, a proof-of-concept study was undertaken by testing for amplification, cross-amplification and polymorphism twelve EST-SSRs in four species from two genera (*Trifolium fragiferum, Trifolium saxatile, Centaurea valesiaca* and *Centaurea borjae*). These four species are of conservation interest due to their threatened status: *Trifolium saxatile* and *Centaurea borjae* are listed by the IUCN while *Trifolium fragiferum* and *Centaurea valesiaca* are included in the Swiss Red List.

MATHERIAL AND METHODS

Sequence data sources

By September 2013, 16 031 555 EST sequences were downloaded from the dbEST database in GenBank at the NCBI website (http://www.ncbi.nlm.nih.gov/dbEST/). Batch files of EST sequences were downloaded in FASTA format. The dataset included 14 498 726 records for 257 genera (*Oryza* included) listed both in IUCN Red List and dbEST plus 1 532 829 records for *Arabidosis*. Whenever full-length cDNA sequences were available, they were included in the dataset along with the ESTs.

EST-SSR detection and primer design

SSRs were detected in the EST dataset with the help of QDD, an open access software which provides a user-friendly tool for microsatellite detection and primer design from large sets of DNA sequences using FASTA files as input (Meglecz *et al.*, 2010). The output file is a list with the ID of the EST sequence that contains the SSR, number and type of repeats, location, and primers information. Before EST-SSR searches, QDD assembled the ESTs of each genus into unigenes (contigs and singletons) to avoid redundancy. Non-redundant EST unigenes were then screened for perfect SSRs. Only Class I microsatellites were considered (Temnykh *et al.*, 2001),

defined as DNA sequences containing at least 20 bp, that is ten repeats for dinucleotides (DNRs), seven repeats for trinucleotides (TNRs), five repeats for tetranucleotides (TRNs) and four repeats for penta- (PNRs) and hexanucleotides respectively (HNRs). Mononucleotides were excluded from EST-SSR searches as their polymorphism is often difficult to interpret. To have enough flanking sequence of appropriate quality for primer design, only EST sequences larger than 100 bp were taken into account during EST-SSR searches. EST-SSRs primers were designed with the version of Primer3 embedded in QDD (Rozen and Skaletsky, 2000) under the following criteria: length ranging from 18-23 nucleotides (optimum 20 bp), annealing temperature 55-65 ºC (optimum 60ºC), GC content 30-70% (optimum 50%) and PCR product size from 90 to 320 bp.

Basal Local Alignment Search Tool (BLAST) searches in Oryza and Arabidopsis

EST sequences for control genera *Oryza* and *Arabidopsis* were run in QDD following the criteria specified above. QDD output files were then used as input for a BLASTn search against *Oryza sativa* and *Arabidopsis thaliana* reference genomes using default parameters specified on the NCBI website. Whenever a positive hit was found (i.e. >98% of coincidence), the matching gene sequence was downloaded and aligned in Geneious 6.1.6 (created by Biomatters, available from http://www.geneious.com/) and the distribution of the SSRs along the genome (UTRs, exons, non-coding regions) was inferred using the annotated gene information derived from the BLASTn search. As a double-check, a BLASTx search against *Oryza* and *Arabidopsis* reference protein databases was also conducted for EST-SSRs using default criteria.

DNA isolation, PCR conditions, and amplification of SSRs

Six individuals of *Trifolium fragiferum,* seven from *Centaurea valesiaca,* two of *Trifolium saxatile* and one from *Centaurea borjae* were used for testing amplification and polymorphism in twelve primer pairs of EST-SSRs. Fresh leaves were dried in silica

gel until DNA extraction. Leave tissue from each plant was collected in a 2.0 ml Eppendorf tube, frozen with liquid nitrogen and ground to fine powder with a Mini-BeadBeater (Glen Mills Inc, NJ, US). DNA was extracted using the Wizard Magnetic Kit (Promega, US) according to the manufacturer's instructions. The quality of the extracted DNA and negative controls were checked in 1.5% agarose gels. Amplification was tested with regular PCR reactions performed in 25 µl containing 1x reaction buffer, 2 mM $MgCl_2$, 0.2 of each dNTP, 0.16 of each primer, 1 µl of genomic DNA and 0.5 units of DNA polymerase (NZyTech, Portugal). PCR profiles consisted of 5 min denaturation at 94°C followed by 35 cycles of 30 s denaturation at 94°C, 50 s annealing at 59° C, and 45 s of extension at 72°C, with a final elongation step of 35 min at 72°C. PCR products were screened on 2% agarose gels. Primer pairs that had successfully amplified in the first round where re-tested with the M13 tail method of Schuelke (2000). PCR reactions were performed in 25 µl containing 1x reaction buffer, 2 mM $MgCl_2$, 0.2 of each dNTP, 0.04 µM of the forward primer with the M13 tail, 0.16 of the reverse and the M13-FAM primer respectively, 1 µl of genomic DNA and 0.5 units of DNA polymerase (NZyTech, Portugal). PCR profiles included 5 min denaturation at 94°C followed by 35 cycles of 30 s denaturation at 94°C, 50 s annealing at 59°C, and 45 s of extension at 72°C, followed by eight additional cycles of 30 s denaturation at 94°C, 45 s annealing at 53° C, and 45 s of extension at 72°C, and a final elongation step of 35 min at 72°C. PCR products were screened on 2% agarose gels and sized on an ABI-3730XL DNA analyzer (Applied Biosystems, US) using a 500HD size ladder. PCR reactions from one primer pair that produced PCR amplicons larger than expected were purified with 1 µl of Exonuclease I (20 u/µl) and 2 µl of FastAP (10 u/µl) and bi-directionally sequenced (BigDye Terminator cycling conditions) in an Automatic Sequencer 3730XL (Applied Biosystems, US).

Compositional analysis of SSR mining

Occurrence and frequency of SSR motifs in the IUCN genera were analyzed after importing QDD output files into MATLAB and Statistics Toolbox 2013a (MathWorks Inc., MA, US). Repeat types, number of repeats, and frequency were calculated for each genus using a combination of sorting and counting functions. Results were displayed in tabular and graphical representations. To provide a broader view, results from IUCN genera were grouped into eight taxonomic groups following Ruhfel *et al.* (2014): Florideophyceae, Charophyceae, Monilophyta, Lycopodiophyta, Acrogymnospermae, Magnoliidae, Monocotyledoneae and Eudicotyledoneae.

RESULTS

Frequency and distribution of SSRs in Arabidopsis and Oryza

The dbEST database contained 1 342 281 *Oryza* ESTs sequences. After filtering out redundant and short (<100bp) records, 2626 EST sequences (1912 singletons and 714 contigs) were left available for SSR search and produced 521 perfect EST-SSRs with primer pairs (19.19%). On the other hand, the *Arabidopsis* dataset contained 1 532 829 EST sequences that, after filtering, was reduced to 899 EST sequences (616 singletons and 283 contigs) that contained 151 perfect SSRs with primer pairs (16.80%). In both cases, filtering had a large impact on the number of EST records available for SSR search, suggesting a high rate of redundant and/or short records in the EST database.

Although only sequences assigned to *Oryza* were downloaded from the dbEST, just 23.80% of the sequences with EST-SSRs did not rendered a significant hit in the BLASTn search against the *O. sativa* reference genome. Similarly, the BLASTn comparison of *Arabidopsis* EST-SSRs sequences against the *A. thaliana* reference genome produced 7.95% of unsuccessful searches. The SSRs derived from these sequences were excluded from further analyses and distribution and position was

determined for 397 EST-SSR of *Oryza* and 139 of *Arabidopsis* (Table 1). Trinucleotide repeats were the commonest repeat size in both genera with very similar relative abudances: 61.96% in *Oryza* and 69.78% in *Arabidopsis*. Dimers were second in abundance, with a frequency of 23.29% in *Oryza* and 17.27% in *Arabidopsis,* while tetra- and pentanucleotides were scarce in both genera (<5%). Hexamers displayed intermediate frequencies in *Oryza* (11.59%) and *Arabidopsis* (8.63%).

Table 1: Number and distribution of the EST-SSRs found for the EST sequences of *Oryza* and *Arabidopsis*.

	Genomic		Intron		UTR		Exon		Total	
	Oryza	*Arabidopsis*	*Oryza*	*Arabidopsis*	*Oryza*	*Arabidopsis*	*Oryza*	*Arabidopsis*	*Oryza*	*Arabidopsis*
Dimers	17 (23.29)	2 (8.33)	26 (35.62)	5 (20.83)	29 (39.73)	16 (66.67)	1 (1.37)	1 (4.17)	73 (23.29)	24 (17.27)
Trimers	18 (7.32)	2 (2.06)	16 (6.50)	2 (2.06)	70 (28.46)	26 (26.80)	142 (57.72)	67 (69.07)	246 (61.96)	97 (69.78)
Tetramers	3 (17.65)	0 (0.00)	3 (17.65)	0 (0.00)	9 (52.94)	3 (100.00)	2 (11.77)	0 (0.00)	17 (4.28)	3 (2.16)
Pentamers	4 (2.67)	0 (0.00)	1 (.67)	0 (0.00)	10 (66.67)	3 (100.00)	0 (0.00)	0 (0.00)	15 (3.78)	3 (2.16)
Hexamers	6 (13.04)	1 (8.33)	3 (6.52)	0 (0.00)	13 (28.26)	1 (8.33)	24 (52.17)	10 (83.33)	46 (11.59)	12 (8.63)
Total	48 (12.09)	5 (3.60)	49 (12.34)	7 (5.04)	131 (33.00)	49 (35.25)	169 (42.57)	78 (56.12)	397 (100.00)	139 (100.00)

Included only EST sequences downloaded from the dbEST database (NCBI) that had a match in their respective reference genomes using BLASTn. SSRs search only consider EST sequences larger or equal to 100bp, and SSRs ≥20 bp. Numbers between parentheses correspond with the proportion for each class.

The various SSR motifs were grouped into classes according to base complementarity and depending on the reading frame (for groups see Fig. 1, from now on in the text will be identified with the first motif repeat). Dinucleotide motifs displayed similar patterns in both genera as the AG group was the most abundant, the AC group had an intermediate frequency, motifs from the AT group were rare and those from the GC group went undetected (Fig. 1). Despite that the AG group prevailed in both genera, it was clearly commoner in *Oryza* than in *Arabidopsis*. Unlike dimers, trimmers displayed different patterns in each genus. Various trimeric motifs that were common in *Oryza*, went unrecorded (GGC and ACG) or very rare (AGC, ACC and AGG) in *Arabidopsis*. GGC group dominated in *Oryza*, with a frequency of 19.51% while the motifs from the groups AAG, AGC and AGG had intermediate values, and the group AAT was clearly underrepresented with only a 1.15% (Fig. 1). In comparison,

trimmers in *Arabidopsis* were dominated by the AAG group with a 48.45% abundance, while two groups (AGC and AAT) were very scarce (i.e. only one and two SSR detected, respectively). No motif from the ATG group was found on either genera.

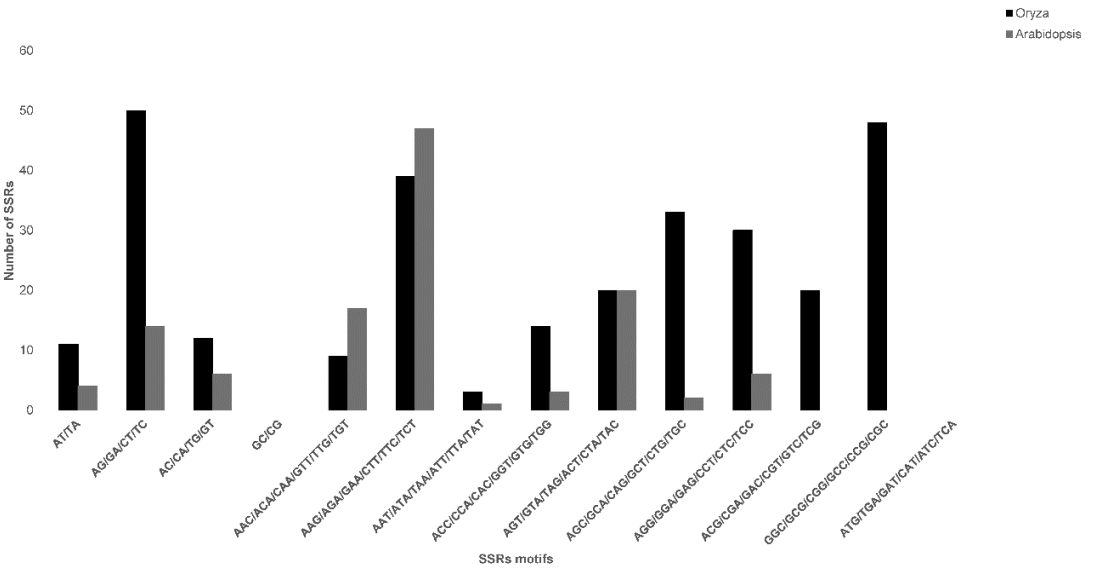


Fig 1: Di- and trinucleotide distribution obtained with iQDD software in *Oryza* and *Arabidopsis* EST sequences that had positive hits in *Oryza sativa* (japonica cultivar-group) and *Arabidopsis thaliana* reference genomes with BLASTn (NCBI).

Four categories were considered for the position of the EST-SSRs along the genome according to the alignments derived from BLASTn results: genomic, introns, untranslated regions (UTRs), and exons. The majority of EST-SSRs were located in exons (42.57% in *Oryza,* 56.12% in *Arabidopsis*) followed by UTRs (33.00% and 35.25% in *Oryza* and *Arabidopsis,* respectively) (Table 1) and only a small fraction was found in non-coding regions (i.e. intergenic regions and introns). The proportion of EST-SSRs found in non-coding regions greatly varied between genera, representing 24.43% in *Oryza* but only 8.64% in *Arabidopsis*. Repeats of different size showed characteristic locations along the genome. Thus, trimmers and hexamers were mostly concentrated in coding regions (exons) with frequencies 57.72 and 52.17% respectively in *Oryza*, and 69.07 and 83.33% in *Arabidopsis*. By contrast, dimers mostly occurred in UTRs (39.73 and 66.67% in *Oryza* and *Arabidopsis*, respectively) but they were also

relatively common in non-coding regions. Tetra- and pentanucleotide repeats were scarce but they occurred preferentially associated to UTRs and non-coding regions in both genera.

EST-SSRs analysis from the IUCN genera

Two hundred and fifty-seven genera from the IUCN plant red list were mined for SSR using EST sequences available in dbEST (NCBI). These 257 genera included two Floriedophyceae, one Cariophyceae, three Lycopodiophyta, five Monilophyta, 18 Acrogymnospermae, three Magnoliidae, 58 Monocotyledoneae, and 167 Eudicotyledoneae. Overall, 14 498 726 sequence were screened for SSR discovery (Table 2). In a few cases, SSR search and primer design were unsuccessful due to a very low number of EST sequences in the input file or sequences that did not fulfilled the predefined criteria (i.e. sequences under 100 bp or highly redundant sequences). As a result, 222 genera were successfully mined for SSR rendering 17 076 microsatellites with primers (see Table S1 in supplementary material). Like in the control genomes, dimers (30.73%) and trimers (39.03%) were the commonest type of SSR while tetramers and pentamers were very scarce (<10%), and hexamers displayed an intermediate position. Nonetheless, when the frequency of the various classes of SSR was analyzed in detail, there were differences among taxonomic groups (Fig. 2). Trimers were commoner than dimers in eudicots and monocots. In Acrogymnospermae, hexamers clearly dominated representing more than one third of the SSRs. Furthermore, trimers were overwhelmingly overrepresented in Lycopodiophyta (64.1%) while dimers were heavily abundant in Monilophyta (81.65%) Finally, tetramers and pentamers were consistently rare across genera except in Florideophyceae.

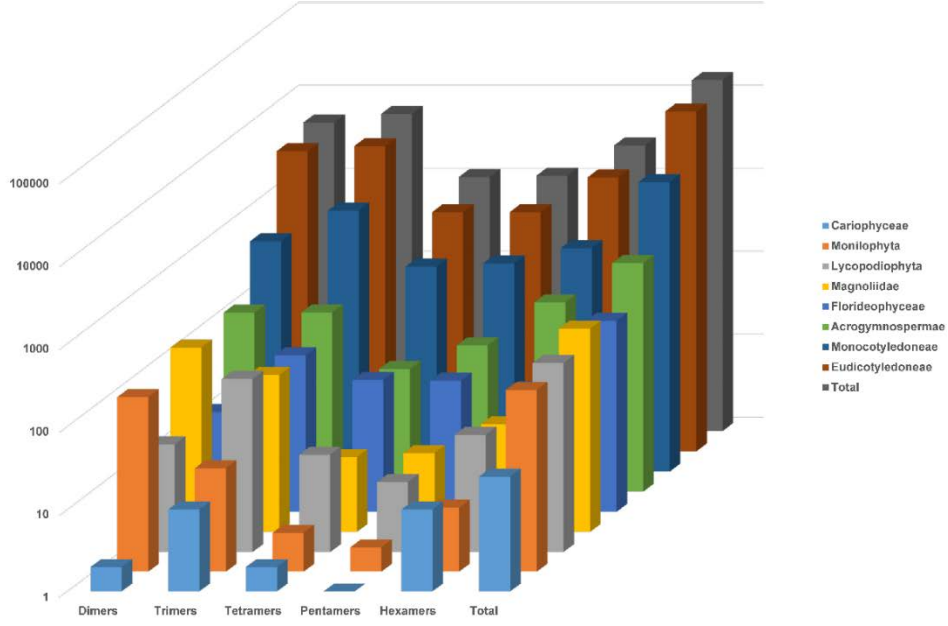
Table 2: Number of SSRs motifs found in 257 genera included in the IUCN red list with EST sequences in the dbEST.

Taxonomic groups	N_{genera}	$N_{Genera\ SSR}$	EST seqs.	Dimers	Trimers	Tetramers	Pentamers	Hexamers	Total	Common motifs
Florideophyceae	2	2	16645	2 (8.0)	10 (40.0)	2 (8.0)	1 (4.0)	10 (40.0)	25 (0.15)	ACG/GGC
Cariophyceae	1	1	88280	16 (8.0)	77 (38.5)	39 (19.5)	38 (19.0)	30 (15.0)	200 (1.17)	AG/TGA
Acrogymnospermae	18	15	1191184	144 (22.26)	145 (25.444)	30 (5.26)	58 (10.18)	193 (33.86)	570 (3.34)	AG/AT/CAG
Lycopodiophyta	3	3	101292	20 (10.53)	122 (64.21)	15 (7.89)	7 (3.68)	26 (13.68)	190 (1.51)	AG/CAG/TGA
Monilophyta	5	3	35665	129 (81.65)	18 (11.39)	3 (1.89)	2 (1.27)	6 (3.80)	158 (0.93)	AG/TGA
Magnoliidae	3	3	57672	167 (59.22)	78 (27.66)	8 (2.84)	9 (3.19)	20 (7.09)	282 (1.65)	AG/AT/CAG
Monocotyledoneae	58	37	3197142	598 (19.24)	1395 (44.88)	296 (9.52)	323 (10.39)	496 (15.96)	3108 (18.20)	AG/AT/AAG/CGG
Eudicotyledoneae	167	158	9810846	4172 (33.26)	4820 (28.23)	767 (6.11)	769 (6.13)	2015 (16.03)	12543 (73.45)	AG/AT/AAG/TGA
	257	222 (86.38)	14498726	5248 (30.73)	6665 (39.03)	1160 (6.79)	1207 (7.07)	2797 (16.37)	17076	

For the SSRs search only EST sequences larger or equal to 100bp, and SSRs motif with 20 or more pair of bases were considered. Numbers between parentheses correspond with percentages.

Overall, the most abundant dimeric motifs were from the AG group. For trimmers there was no consensus along all the groups studied but the AGT and AGC groups were the commonest. When each taxonomic group was considered separately, the AT group was also very common in Spermatophyta (Acrogymnospermae and Angioespermae), second only to the AG group. In red algae the ACG and GGC groups were the most frequent. Moreover, trimers rich in GC displayed high abundance in Monocotyledoneae while it was absent from the remaining groups of Streptophyta. Tetramers, pentamers and hexamers were too scarce in most taxa to allow an appropriate analysis of their distribution. Only in Acrogymnospermae, the distribution of hexanucleotides was examined in detail finding that ATCGGG and ATGGCG were the main motifs.

Fig 2: Distribution of SSRs motif in 222 IUCN red list genera grouped into eight large taxonomic gropus (Florideophyceae, Cariophyceae, Lycopodiophyta, Monilophyta, Acrogymnopermae, Magnoliidae, Monocotyledoneae and Eudicotiledonea). The axis Y (logarithmic scale) represents the number of SSR.



Amplification and transferability of the EST-SSRs

A subset of 24 pairs of EST-SSRs primers (12 pairs per genus) were chosen to test amplification performance in two genus of Eucotyledonae, *Trifolium* and *Centaurea* (Table 3). A total of 85 293 *Trifolium* EST sequences were run for SSR search rendering 130 EST-SSR with their primers. Likewise, the 53 422 EST sequences analyzed for *Centaurea* returned 306 EST-SSRs and their primers. Thirteen out of the 24 pairs of primers yielded a clear amplification product (amplification rate 54.2%). Nevertheless, the amplification success differed between genera and *Centaurea* displayed a higher amplification rate (66.7%) than *Trifolium* (41.7%). All loci produced amplification products of the expected size, except for locus C6 of *Centaurea* that generated an amplicon longer than expected, suggesting the presence of a non-transcribed intron inside; which was further confirmed by the sequencing of the PCR product. The protocol from Schuelke (2000) had mostly no impact on PCR performance since all the pair of primers that amplified in the first round with untransformed primers also did with the M13-tail ones. However, locus C7 produced an unspecific second band, larger than the one obtained in the first round, with method of Schuelke (2000).

The selected primers were also used to assess the cross-species transferability in two species, *C. borjae* and *T. saxatile*. Only two individuals of each species were used in this process as the aim was test the level of transferability among species of the same genus rather than polymorphism. Cross-species amplification was considered successful when an amplification band was observed in the electrophoresis gel. Under this criterion, the rate of successful transferability was 100%, since all the primers that worked on one species also did it on its counterpart.

Table 3: Characteristics of the EST-SSR loci tested for amplification in *Trifolium* and *Centaurea*. Loci with several GenBank EST gi correspond to consensus sequences generated by QDD.

Genus	GenBank EST gi	Repeat motif	Primer sequence	PCR product
T6-*Trifolium*	gi86106666 gi86105378	$(AG)_{11}$	T6_F: CAACCAGTGGTGTGAGTAGGAG T6_R: ACGTTGGTGGAGAGGTTGAG	113-115bp
T7-*Trifolium*	gi428283538	$(AG)_{13}$	T7_F: ATCACGCTTCACTCCTCCAC T7_R: CAACTCCAAGCTTAAGATCGTGTA	no PCR product
T1-*Trifolium*	gi428292074	$(AG)_{11}$	T1_F: AGATTCCCACCAATCTCCCT T1_R: CAATACGCGGGTCTTGATCT	257-261bp
T2-*Trifolium*	gi86106666 gi86105378	$(AAT)_{7}$	T2_F: TTCCGGTTAGGTTAGGGTTT T2_R: TTTTCACATCTTCCGAAGCC	no PCR product
T3-*Trifolium*	gi428285635	$(AGT)_{8}$	T3_F: CACCACATATGCAACCACAA T3_R: GTCGACGACGGTTGTTACCT	no PCR product
T8-*Trifolium*	gi428291122	$(ACC)_{7}$	T8_F: GCAAAACTCAAGAGAACGGC T8_R: GGATGTCTTCGGAGGTGAGA	no PCR product
T9-*Trifolium*	gi428292435	$(ACC)_{7}$	T9_F: ACAACCCATTTGCCTCAAAG T9_R: TTTTCACTTCCACCACCTCC	124-127bp
T10-*Trifolium*	gi86119186	$(ACC)_{9}$	T10_F: TCCACTAGTTCTAGAGCGGC T10_R: TCCTGTAAACTGGAGGAGCC	no PCR product
T11-*Trifolium*	gi86124411	$(AGG)_{8}$	T11_F: TGGCGGTGGTGACTTATACA T11_R: TGTTTGGCAGTGGTGATGTT	no PCR product
T4-*Trifolium*	gi86125686	$(ACC)_{8}$	T4_F: GCTGCCACAGCACTACCAG T4_R: AATATTACCGTGAATGAAGCTCAG	110bp
T5-*Trifolium*	gi86097190	$(ACCT)_{5}$	T5_F: TGAGTTCCGAGTTAAGGCTCA T5_R: TTCGGTAACTCCGAGGATTG	227-231bp
T12-*Trifolium*	gi428282514	$(AATCC)_{20}$	T12_F: GATTATTCAACCAAACGCCG T12_R: TAGAAAGCCACGCCAAGACT	no PCR product
C6-*Centaurea*	gi124618051	$(AC)_{11}$	C6_F: TGGGATGCAGTCCAGTCATA C6_R: TTGCAACTTGCCTGTACCAC	256bp
C1-*Centaurea*	gi148298213	$(AC)_{10}$	C1_F: GGGAACCACACCTTTCATCT C1_R: GATCTGGCTTGACCCAAGAA	133-135bp
C7-*Centaurea*	gi124669731 gi124688599	$(AC)_{12}$	C7_F: TCGTTTTCCGATCACAAACTC C7_R: CAATTTGGCGACATCTCCTT	141-143bp
C2-*Centaurea*	gi124680442	$(AAG)_{7}$	C2_F: CGCATTATGGAATAAACCCG C2_R: GCTTTCGACTTCATAAGCGG	305bp
C8-*Centaurea*	gi148296795	$(ACC)_{7}$	C8_F: CGATGTATACAGGTGGTGCG C8_R: GGAGAAGGGGAGACGTAAGG	141-144bp
C9-*Centaurea*	gi124675484	$(ACC)_{9}$	C9_F: AACGGTAGGAACCAGCATTG C9_R: GATCCTCTGGCAGGGTCATA	no PCR product
C10-*Centaurea*	gi124661102	$(AGC)_{7}$	C10_F: AGTTGCCAGAAAGGAGCAAG C10_R: TCGAGAACAATGGCCTATCC	no PCR product
C11-*Centaurea*	gi148292432	$(AGG)_{7}$	C11_F: TCCATGGATACAACCACCAA C11_R: GCGATATTCGGATGCAAAGT	160-172bp
C3-*Centaurea*	gi124632630	$(AGT)_{7}$	C3_F: GCCATCCCCTTCTCTACTCC C3_R: GTTACAGGTGACGATGGGGT	no PCR product
C4-*Centaurea*	gi124691992	$(AGGT)_{5}$	C4_F: CTGCACCTACCCAGAGAAGC C4_R: CGGGAGAGGGTAAATTGTGA	103-107bp
C12-*Centaurea*	gi124632477	$(AATCGG)_{4}$	C12_F: ATGCATTGAGAAGGCCAATC C12_R: AACTCGCAAGCCTTTTCAAG	no PCR product
C5-Centaurea	gi124673348 gi124676118 gi124669484	$(AAGCAG)_{5}$	C5_F: TTAAGCATTCTTCGAGGCGT C5_R: TCTATGCCTACGCCGATCTC	no PCR product

Despite the small number of individuals used in the polymorphism tests, two out of the seven EST-SSRs (28.75%) that yielded a PCR product of the expected size in *Centaurea* displayed polymorphism within species (Table 3): loci C1 and C11 in produced two and three genotypes, respectively. On the other hand, one of the dimeric loci of *Trifolium* (T1) displayed a stutter-peak profile and was discarded from further analysis. Among the four remaining loci, T5 and T9 were polymorphic revealing three and two genotypes, respectively (50% polymorphism). Finally, six out of the seven loci of *Centaurea* produced different genotypes in the two species used in our tests (87.77%) while three out of the four loci of *Trifolium* were polymorphic between species (75%).

DISCUSSION

Computational approaches allow the fast discovery of molecular markers from the ever-increasing publicly available genomic resources. Thus, SSRs derived from EST sequences arise as an excellent alternative to the classical techniques of anonymous microsatellites because of their fast and inexpensive discovery (Ellis and Burke, 2007). Besides, unlike anonymous SSRs, EST-SSRs markers have proven of great value in cross-species studies, linkage maps, and in discovering markers linked to genes rather than only in traditional population structure studies (Varshney *et al.*, 2005b). Thus far, EST-SSR development have almost exclusively targeted crop and model species, ignoring non-model ones (Aggarwal *et al.*, 2007; Blair and Hurtado, 2013; Fukuoka *et al.*, 2010; Gao *et al.*, 2003; Kantety *et al.*, 2002; Mishra *et al.*, 2011; Simko, 2009; Varshney *et al.*, 2005b). In this context, the present study has tried to fill this gap by focusing on developing EST-SSRs for evolutionary and conservation studies in non-model species, with a special emphasis on threatened plants.

Frequency and distribution of SSRs in Arabidopsis and Oryza

The frequency and distribution of short tandem repeats in EST sequences is highly variable among studies, in part because the efficiency of SSR discovery relies on several factors such as the mining tool used, the mining criteria, or the size of the EST sequences dataset (Aggarwal *et al.*, 2007; Blair and Hurtado, 2013). Differences in mining criteria such as searching for perfect and/or imperfect repeats, minimum numbers of repeats, or length of spacer in compound repeats usually lead to significant deviations in the number of microsatellites identified in a given species using the same dataset (Aggarwal *et al.*, 2007). Here, we opted for highly conservative criteria and only perfect repeats with a length equal or larger than 20 bp were considered (Blair and Hurtado, 2013). We did so in an effort to increase the polymorphism of the detected SSRs but, as a consequence, we probably obtained a lower number of EST-SSRs than would have been found if more relaxed parameters were set for the searching.

The in-depth analysis of EST-SSR frequency and distribution in *Arabidopsis* and *Oryza* revealed that trimmers and dimers contained more than 85% of the SSRs found. Furthermore, trinucleotide repeats comprehended the vast majority of SSRs, accounting for more than 60% of the detected loci. High frequencies of trimmers are known to be favored in higher plants in comparison with algae or mosses and have been invariably reported in most studies (Kantety *et al.*, 2002; Varshney *et al.*, 2005b; Victoria *et al.*, 2011). As expected in vascular plants, the AG group was the most abundant dinucleotide motif and low frequencies of the AT group were recorded in both genera (Kantety *et al.*, 2002; Morgante *et al.*, 2002; Temnykh *et al.*, 2001; Victoria *et al.*, 2011). In agreement with previous studies of monocots and eudicots, we found differences in the trinucleotide repeats of *Oryza* and *Arabidopsis*. GC-rich motifs, commonly dominant in monocots, were the most frequent trimmers in *Oryza* as the group GGC (Gao *et al.*, 2003; Temnykh *et al.*, 2001; Kantety *et al.*, 2002; Victoria

et al., 2011) while the AAG group prevailed in *Arabidopsis* where GC-rich motifs were scarce (Victoria *et al.*, 2011).

Overall, a major fraction of EST-SSRs were located in CDS regions, an observation that seems consistent with the fact that EST-SSR derive from transcribed regions. Nevertheless, not every type of nucleotide repeat appeared in CDS regions with equal probability. Di, tetra and pentamers mostly concentrated in UTRs and, to a lesser extent, in other non-coding regions. However, trimmers and hexamers regularly occurred in CDS regions. Since the frequency and distribution of the various SSR repeats and motifs are a function of the dynamics and history of genome evolution, the predominance of trimeric repeats, especially trinucleotides, in ESTs has been attributed to selection against frameshift mutations caused by length variation in non-trimeric motifs (Morgante *et al.*, 2002). Large frequencies of dimers in UTRs and a prevalence of trimmers in CDS regions have been consistently reported in other plant studies (Gao *et al.*, 2003; Wang *et al.*, 1994). Since EST sequences derive from mRNA, the frequency of EST-SSRs located in non-coding regions might seem unexpectedly high. However, transcripts of unknown function with apparently little protein coding capacity are now known to overlap with protein-coding regions and they are often distributed in intergenic regions (Gingeras, 2007).

Interestingly, trinucleotides in *Oryza* were rich in GC motifs and more than 70% of these GC-rich trimmers were linked to CDS regions. CCG repeats have been found to be involved in many gene functions such as stress resistance, transcription regulation, or metabolic enzyme biosynthesis (Gao *et al.*, 2003). As trinucleotide repeats are usually related to coding regions, they usually involve a moderate number of repeats based on the limitation to non-perturbation of the triplet codon, which may result in low levels of polymorphism (Cho *et al.*, 2000). In contrast, dimers tend to display higher levels of variation as consequence of their association with UTRs and non-coding regions (Liewlaksaneeyanawin *et al.*, 2004; Yu *et al.*, 2004).

EST-SSRs analysis from the IUCN genera

The frequencies of the various nucleotide repeats and motifs in IUCN genera were highly consistent with the results obtained in the control genomes of *Oryza* and *Arabidopsis*. Trimers and dimers accounted for >60% of the EST-SSRs, while tetramers, pentamers and hexamers displayed lower frequencies. However, the abundance of the various types of nucleotide repeat differed between groups. Results for mocots and eudicots were highly consistent with those obtained in the two control genomes. They were likewise in agreement with previous findings in flowering plants where trimmers were the most abundant motifs followed by dimers (Victoria *et al.*, 2011). Similarly, AG was the commonest dimer, as it seems typically the case in angiosperms (Kantety *et al.*, 2002; Morgante *et al.*, 2002; Temnykh *et al.*, 2001; Victoria *et al.*, 2011). The pattern seen in the trimeric motifs of IUCN genera agreed with what we found in *Oryza* and *Arabidopsis*, corroborating the high abundance of CG-rich motifs in monocots and the AAG group in eudicots (Gao *et al.*, 2003; Kantety *et al.*, 2002; Temnykh *et al.*, 2001; Victoria *et al.*, 2011). Differences in the frequency of the various types of repeat and motif between taxonomic groups were expected because the SSR distribution is affected by the dynamics and history of genome evolution (Morgante *et al.*, 2002). Thus, Acrogymnospermae revealed a higher proportion of hexamers than mono and eudicots, and the leading motif in Acrogymnospermae, the AT group, was very scarce in angiosperms. Similar results have been reported for this group of plants in previous studies (Pinosio *et al.* 2014; Victoria *et al.* 2011). Unfortunately, the four groups of non-vascular plants were represented by too few genera to allow generalizations.

Amplification and transferability of the EST-SSRs

Amplification success in this study was similar to values reported in some studies of EST-SSRs (Cordeiro *et al.*, 2000; Rungis *et al.*, 2004) but lower than others

(Eujayl *et al.*, 2004; Wöhrmann and Weising, 2011). Unsuccessful primer amplification can be a consequence of non-transcribed introns located in the annealing primer region (Ellis and Burke, 2007). Also, some of the EST-SSRs detected in our searches could actually belong to a different species because, as revealed by our analysis of control genomes, a portion of EST sequences do not find a match in control genomes and might be a result of RNA contamination (Varshney *et al.*, 2005a).

Given their association with conserved regions of the genome, EST-SSRs are often assumed to be less polymorphic than their genomic counterparts (Ellis and Burke, 2007; Russell *et al.*, 2004; Varshney *et al.*, 2005a). However, studies comparing both types of marker showed that this premise does not always hold true and similar levels of polymorphism have been found in anonymous versus EST-SSRs (Fraser *et al.*, 2004; Pashley, 2006). In our study, polymorphism ranged from 25 to 28.57% within species and from 75 to 87.77% between species. Since only eight individuals of each species/genus were used to assess polymorphism, the levels estimated here must be taken with caution and cannot be consider a general attribute of EST-SSRs. The quality of the banding patterns was high, with clear peaks (except for locus T1), a flat baseline, and no null allele was detected. Cleaner profiles and lower frequencies of null alleles than those found in anonymous SSRs appears to be a general property of EST-SSRs (Pashley, 2006; Woodhead *et al.*, 2005; Wöhrmann and Weising, 2011). The lower levels of polymorphism usually attributed to EST-SSRs compared with anonymous SSRs may be compensated by their high rate of cross-species transferability (Aggarwal *et al.*, 2007; Pashley, 2006; Wöhrmann and Weising, 2011), which has been reported not only among congenerics but also across different genera (Varshney *et al.*, 2005b). Our results are highly congruent with the premise of high-transferability in EST-SSRs. All of the tested loci that successfully amplified in one species did the same in its counterpart, supporting that EST-SSRs are markers with a great potential for comparative studies among species.

Use of EST-SSR as molecular markers for studying threatened species

Whereas EST-SSRs can be essentially used for the same purposes of the genomic SSRs, their association with translated regions offers a range of possibilities not usually available in anonymous SSRs. Since microsatellites derived from EST sequences are associated with CDS regions, the function of these genes can often be identified by aligning the ESTs of interest against genomic sequence of a model organism such as *Arabidopsis* for eudicots and *Oryza* for monocots. Therefore, these markers could be useful in quantitative trait locus mapping within species and in comparative genomics studies among species due to their high cross-species transferability (Varshney *et al.*, 2005b). Likewise, EST-SSRs have also been considered a better option than anonymous SSRs for resolving phylogenetic studies (Tabbasam *et al.*, 2013).

Even if genomic SSRs seem a more suitable option for studies detecting intraspecific variation because they tend to display higher levels of polymorphism, this can be compensated combining both types of markers (Aleksic and Geburek, 2014; Wöhrmann *et al.*, 2011). A possible concern when dealing with EST-SSRs is that, as consequence of their association with genic regions, selection may influence the estimates of population genetic parameters (Pashley, 2006). However, several studies suggest that this may not be an issue as estimates of population differentiation were largely consistent with those derived from anonymous SSRs (Woodhead *et al.*, 2005). Of course, not every EST-SSR will behave as a neutral marker and loci linked to genes involving relevant traits may display a signature of selection. However, the latter may offer the chance to target “adaptive variation”, an issue of high relevance in studies addressing conservation issues (Frankham *et al.*, 2010). Our results suggest that conservation studies with adaptive variation in mind should focus on trimmers. Trinucleotide repeats are very likely to be located within exons, they are commoner and more polymorphic than hexamers. Besides, and as noted before, EST sequences

with SSRs can be cross-referenced with annotated genomes for sequence similarity and gene discovery. Dinucleotide repeats could be another good choice because they are known to be very polymorphic and our results show that they are mainly linked to UTRs, which are known to be involved in gene expression and other control functions (Conne *et al.*, 2000).

In summary, this study represents the first attempt to test the potential of publicly accessible EST databases as a source of SSRs discovery for threatened plant species at a broad scale. We successfully detected SSRs with primers for more than 87% of the 257 IUCN plant genera analyzed, thus providing EST-SSRs ready to test for 222 genera. Since EST-SSRs have proved to be highly transferable among species, the number of species that could be potentially targeted in studies using the set of loci presented here could eventually be quite large. A common limitation for many population genetics studies with non-model organism is the development of the set of molecular markers. Our study shows that EST databases are a valuable and suitable source for SSRs discovery. Once accessed the EST database, a set of EST-SSRs with primers can be produced in a couple of days with no further cost. In conclusion, our results highly support the use of existing EST databases for SSRs discovery in non-model plants as a bench tool for evolutionary and/or conservation studies of population geneticists and molecular ecologists.

ACKNOWLEDGEMENTS

The authors would like to thank Tina Wöhrmam for her helpful advice in an early stage of the study, Anne Kempel for providing the tissue samples and Lisa Kretz, Maria Rodriguez Lojo, Eva Wolf, Florian Michling and Nora Hohmann for their technical help during the experimental phase. This research was supported by the European Science Foundation (ConGenOmics Network) and University of A Coruña (contratos predoutorais UDC 2012).

SUPLEMENTARY MATERIAL

Table S1: List of IUCN plant genera mined for EST-SSRs with raw results. EX = extinct, EW = extinct in the wild, CR =critically endangered, EN = endangered, VU = vulnerable, NT =near threatened, LC =least concern, DD =data deficient.

Taxonomy				IUCN records								EST-SSRs						
Group	Class	Family	Genus	EX	EW	CR	EN	VU	NT	LC	DD	tot	n EST	di	tri	tetra	penta	hexa	Tot
Rodophyta	Florideophyceae	Gigartinaceae	Chondrus								2	2	4114					3	3
		Gracilariaceae	Gracilaria			1					1	2	12531	2	10	2	1	7	22
Charophyceae	Charophyceae	Characeae	Nitella							7		7	88280	16	77	39	38	30	200
Monilophyta	Osmundiopsida	Osmundaceae	Osmunda							1		1	2684	1	3	1			5
	Polypodiopsida	Adiantaceae	Adiantum	1			2			1		4	30540	122	16	3	1	4	146
		Azollaceae	Azolla							2		2	6						0
		Marsileaceae	Marsilea				1	2		13		16	61						0
		Pteridaceae	Ceratopteris							2		2	5125	7	2		1	2	12
Licopodiophyta	Isoetopsida	Isotaceae	Isoetes			8	4	2	3	5	6	28	338					1	1
	Lycopodiopsida	Lycopodiaceae	Huperzia				2	8	1	1		12	3451	1	9	2	1	10	23
	Sellaginellopsida	Selaginellaceae	Sellaginella				1		1	2		4	97503	19	113	13	6	15	166
Acrogymnospermae	Coniferopsida	Araucariacea	Araucaria			2	5	6				13	10						0
		Cupressaceae	Chamaecyparis				1	1				2	6702	0	2	1	1	2	6
			Taiwania					1				1	2624						0
			Cunninghamia					1		1		2	444		1				1
			Cryptomeria						1			1	57653	1	4	1	2	4	12
			Picea			2	5	5	2	22		36	546864	55	44	12	23	51	185
			Pinus			1	2	7	15	85		110	476951	38	29	6	15	64	152
			Pseudostuga					2	1	1		4	18142	1	1		4	8	14
			Abies			4	4	4	3	30	1	46	2806					2	2
			Larix					1		9		10	965		2			2	4
		Podocarpaceae	Podocarpus			3	14	8	11	51	11	98	677	3			3	1	7
		Sciadopytaceae	Sciadopitis					1				1	11	1	2				3
	Cycadiopsida	Taxaceae	Taxus			1	2	1	2	4		10	492						0
		Cycadaceae	Cycas			11	16	30	24	15	2	98	21997	10	1	2	2	1	16
		Zamiaceae	Zamia			15	13	11	20	1		60	20677	13	22	1	3	7	46
	Gnetopsida	Ephedraceae	Ephedra						1	51	4	56	4981		4	1	1	13	19
		Gnetaceae	Gnetum				1	3	6	25	6	41	10724	6	26	5	3	32	72
	Ginkgoopsida	Ginkgoaceae	Ginkgo				1					1	21590	16	7	1	1	6	31
Magnoliidae	Magnoliidae	Magnoliaceae	Lirodendron						1			1	24663	84	32	4	1	8	129
		Lauraceae	Persea				1	11	3	1		16	16558	56	14	2	6	8	86
		Aristolochiaceae	Aristolochia			2	3	6				11	16451	27	32	2	2	4	67

Monocotyledoneae	Liliopsida	Acoraceae	Acorus							2		2	9733	26	19	3	1	3	52
		Alliaceae	Allium			1	1	4	2	11	37	56	41796	17	16	2	3	13	51
		Amaryllidaceae	Narcissus				5			6	1	12	1						0
		Araceae	Zantedeschia							2		2	4283	9	2			1	12
			Scidapsus							1		1	18						0
			Colocasia							2		2	9						0
			Landoltia							1		1	154						0
		Asparagaceae	Asparagus				3	2		2	3	10	8519	16	20	1	2	4	43
		Agavaceae	Agave							1		1	4						0
		Bromeliaceae	Guzmania			4	10	19	5	6	1	45	1						0
		Costaceae	Costus					2		1		3	1221	1				1	2
		Dioscoreaceae	Dioscorea			1	2	2	1	8	2	16	44169	14	40	6	6	16	82
		Gramineae	Oryza				2			2		4	1342281	101	315	43	60	62	581
			Panicum			2	1			6		9	720785	63	205	39	56	96	459
			Saccharum							3		3	284272	64	178	33	50	92	417
			Sorghum							3		3	232681	59	174	45	57	61	396
			Festuca				1	4	2	5	1	13	139903	24	74	27	16	24	165
			Avena							1	1	2	79586	26	81	28	17	29	181
			Setaria							2		2	66027	7	17	4	11	3	42
			Elymus							1		1	27891	13	26	4	8	10	61
			Cenchrus			1						1	21708	7	16	5	3	3	34
			Agrostis				1		2	4	1	8	21655	4	5	2	2	3	16
			Eragrostis							11		11	15898	6	5	7	2	0	20
			Puccinellia			1		1				2	15037	6	18	3	5	2	34
			Urochloa							1		1	2207						0
			Eleusine							1		1	1934	1	3			1	5
			Bambusa							1		1	514	4					4
			Brachiaria							6		6	41	1					1
			Vetiveria							1		1	32	1					1
			Poa				1			8	1	10	30						0
			Agropyron				2		1			3	16						0
			Paspalum				2	1	6			9	7						0
			Digitaria							4	3	7	1						0
			Phalaris					1		1		2	1						0
		Hydrocharitaceae	Hydrilla							1		1	70						0
		Iridaceae	Crocus					2	1			3	6767	3	5	3	1	3	15

			Iris			1					2	3	6745	31	41	6	2	8	88
		Lemnaceae	Wolffia							3		3	1827		1				1
			Lenma				1			6		7	16						0
		Liliaceae	Lilium					1		1	1	3	3906					2	2
			Fritillaria				4	2			3	9	2158	2	2			2	6
			Tulipa				4	2			3	9	76						0
		Orchidaceae	Dendrobium			3	8	1		7	3	22	16183	31	18	3	2	9	63
			Phalaenopsis			2	1	1				4	8066	3	5	1		2	11
			Vanda				2			1		3	2271	5	5	1		1	12
			Oncidium							1		1	3138	6	5		1	2	14
			Cymbidium				2				1	3	78						0
			Calanthe				1	3			1	5	66						0
			Vanilla				1			1		2	31						0
			Anoectochilus				1	1			1	3	7						0
		Palmae	Calamus			4		1	1	10	2	18	2528	4	6		2	3	15
			Metroxylon					1		1		2	412						0
			Phoenix						3			3	411	4	5	2	1	1	13
		Pandaceae	Pandamus			5	1	12	5	3	2	28	977	4	1			1	6
		Posidoniaceae	Posidonia					1	1	6		8	5668	8	16	1		2	27
		Typhaceae	Typha							5		5	133						0
		Zingiberaceae	Zingiber				1	1		5	1	8	38186	21	48	13	7	22	111
		Zosteraceae	Zostera				2	2	2	8		14	18345	6	23	14	8	14	65
Eudicotyledoneaceae	Magnoliopsida	Acanthaceae	Acanthus					1		4		5	628		1				1
		Actinidiaceae	Actinidia				1	7				8	132593	278	54	10	13	55	410
		Aizoaceae	Sesuvium							1		1	38						0
		Anacardiaceae	Mangifera		2	1	7	15		6	8	39	1665	0	1	0	0	0	1
			Pistacia					2	2	1		5	1299	0	1	0	0	3	4
		Apocymaceae	Wrightia					3		1		4	16431	25	26	3	2	7	63
		Araliaceae	Panax				1					1	22225	31	32	11	14	28	116
			Eleutherococcus				5					5	464						0
			Saruma				1					1	10274	26	11	2	0	10	49
			Asarum					2				2	623			1			1
		Avicenniaceae	Avicennia					4		5		9	1893	1	2	0	0	0	3
		Berberidaceae	Berberis			2	2	9	2	2	14	31	10120	9	27	5	12	18	71
			Epimedium							6		6	4983	6	16	7	2	9	40
		Betulaceae	Alnus						1	2		3	32544	63	33	0	2	16	114

			Betula		1	3	2	2		3		11	5910	9	4	0	2	0	15
			Corylus			1				1		2	3						0
		Boraginaceae	Onosma			1					1	2	78						0
		Burseraceae	Commiphora					3	16		1	20	81	1	0	0	0	0	1
		Cactaceae	Opuntia							8		8	451	1	1	0	0	1	3
		Capparaceae	Cleome							1		1	4677	20	15	0	1	6	42
		Caricaceae	Carica				2	2	2			6	77393	57	36	1	4	15	113
		Caryocaraceae	Caryocar				2	1				3	958	0	1	0	0	0	1
		Caryophyllaceae	Silene			2	6	2	3	1		14	3849	4	10	3	1	5	23
			Dianthus			1	1	1	1	1	3	8	23276	4	86	8	13	27	138
			Gypsophila				1					1	148	0	0	0	0	1	1
		Celastraceae	Euonymus			1	4	4	2	5		16	51380	17	15	3	1	13	49
			Catha							1		1	445	23	7	2	0	3	35
		Chenopodiaceae	Beta			1		1			1	3	29830	6	24	1	1	33	65
			Suaeda							1		1	1703	0	4	1	0	3	8
			Salicornia				1					1	1432	0	1	0	0	1	2
			Atriplex			1						1	519	0	1	1	0	1	3
		Cistaceae	Cistus				1				1	2	2048	0	1	1	0	0	2
		Combretaceae	Terminalia		1		6	12	1		1	21	9						0
		Compositae	Lactuca				1	2	2	3	3	11	226709	84	213	23	32	138	490
			Artemisia			1	1	1	1	2	1	7	95480	12	40	6	9	28	95
			Centaurea	1		3	4	9	3		3	23	85293	24	146	33	26	77	306
			Carthamus							1		1	41584	30	96	21	9	29	185
			Barnadesia				1			1		2	28483	38	34	2	2	14	90
			Senecio		1	2	1	1		4	2	11	10219	2	9	3	2	13	29
			Stevia				1	2			2	5	5548	0	6	0	0	6	12
			Vernonia	1			1	3			1	6	1325	2	1	1	0	2	6
			Mikania			6	2		2		2	12	379	0	2	1	0	0	3
			Cirsium					1	1	2		4	173	0	1	0	0	1	2
			Hieracium			2		1	1		3	7	132						0
			Tanacetum			1	1					2	58						0
			Ageratum					1				1	20						0
		Convolvulaceae	Ipomoea					1		2		3	87095	73	72	23	23	34	225
		Crassulaceae	Kalanchoe					1		1		2	670	0	2	0	0	1	3
			Sedum					1			1	2	145						0
		Cruciferae	Brassica			1	1	1	2	3	3	11	1047449	238	299	24	36	86	683

			Thaspi						1			1	647						0
			Lepidium			1		3		3	3	10	322						0
			Erysimum			1		1				2	206						0
			Barbarea				1			2	7	10	31	0	6	0	0	0	6
			Rorippa			1		1		2	1	5	30						0
			Isatis					1			6	7	27						0
		Cucurbitaceae	Cucurbita							1		1	880	1	8	0	1	2	12
		Dipterocarpaceae	Shorea	1		102	34	3		6	2	148	39	2	3	0	0	0	5
		Ebenaceae	Diospyros			14	14	40	5	19	4	96	14187	17	11	4	3	8	43
		Ericaceae	Vaccinium				1	1				2	22401	110	22	4	8	6	150
			Rhododendron		1	1		6	2	1		11	1293	11	1	0	0	0	12
		Eucommiaceae	Eucommia						1			1	28456	84	28	1	5	17	135
		Euphorbiaceae	Euphorbia		1	23	30	76	6	33		169	59055	36	177	22	23	35	293
			Jatropha				1	2		1		4	46865	42	36	2	6	19	105
		Fagaceae	Quercus			4	8	23	2	16	18	71	150124	207	163	24	21	73	488
			Fagus					2				2	36740	50	47	9	4	28	138
			Castanopis			1		4	2			7	3354	23	9	1	4	9	46
		Geraniaceae	Geranium			2	1	6				9	27						0
		Gesneriaceae	Haberlea							1		1	373	0	0	0	0	2	2
			Paraboea			3	8	3		4	5	23	2						0
		Grossulariaceae	Ribes			2		2		1		5	8490	15	27	2	3	10	57
		Guttiferae	Garcinia			1	7	16	3	12		39	149	0	0	0	1	0	1
			Hypericum			1	2	6	1	7		17	18						0
		Illiciaceae	Illicium					2		2		4	233	1	0	0	0	0	1
		Junglandaceae	Juglans				2	4	2			8	18850	45	22	8	8	11	94
		Labiate	Salvia			2	4	5	1	1	2	15	11766	12	11	3	4	5	35
			Mentha						2	2		4	1663	1	4	0	0	0	5
			Lavandula							1		1	22						0
		Leguminosae	Glycine							2		2	1481192	143	208	21	32	79	483
			Phaseolus			1	1			2	1	5	581936	58	85	20	11	42	216
			Medicago			1	1	1	3	7	2	15	281872	55	117	26	28	45	271
			Arachis					1	1	1		3	254065	102	184	38	24	77	425
			Lotus			5	2	1		7	1	16	242485	61	149	20	21	74	325
			Vigna						1	2	1	4	189876	39	69	6	16	33	163
			Trifolium						1	19		20	53422	23	77	11	9	10	130
			Cicer				2			2		4	44766	13	18	2	8	9	50

			Cajanus						1	1		2	25639	3	3	4	3	3	16
			Acacia			4	9	16	19	32	4	84	11569						0
			Lupinus				1	1	4	14	5	25	9944	7	13	1	2	6	29
			Lathyrus			1			1	8		10	8880	3	4	0	3	3	13
			Pueraria							1		1	6365	7	10	1	6	5	29
			Ammopiptanthus			1						1	5755	2	14	1	0	9	26
			Vicia			4	1		1	9		15	5613	0	3	1	0	1	5
			Robinia							3		3	3095	4	4	0	0	0	8
			Caragana							3	1	4	1905	0	2	1	0	1	4
			Prosopis					1	2	2	3	8	1599	3	0	0	1	1	5
			Oxytropis							6	2	8	1245	0	1	0	0	0	1
			Macrotyloma							1		1	1025	0	0	0	1	0	1
			Copaifera					3		1		4	716	2	2	0	0	1	5
			Sesbania					1		8	1	10	374	0	1	0	0	1	2
			Stylosanthes							1		1	350	0	0	0	0	0	0
			Hymenaea					1		3		4	299	0	1	0	0	0	1
			Leucaena				3	4	1			8	150	0	0	1	0	0	1
			Astragalus	1		6	7	4	4	54		76	125						0
			Hedysarum						1	3	1	5	6						0
		Linaceae	Linum			1		2			1	4	286856	33	128	20	6	8	195
		Lythraceae	Punica					1		1		2	2383	6	3	1	0	3	13
			Sonneratia			2			1	4		7	249	2	0	0	0	0	2
		Malvaceae	Hibicus			2	2	2	1	3	2	12	47						0
			Abulition			3		1				4	5						0
		Moraceae	Ficus			3	10	11	1	23		48	9296	21	37	0	5	17	80
			Artocarpus					5				5	3						0
		Myrtaceae	Eucalyptus			1	1					2	167436	228	197	24	19	65	533
			Myrciaria					3				3	8						0
		Oleaceae	Fraxinus			1			1			2	12083	9	2	1	1	4	17
			Olea					1			1	2	9959	0	2	0	1	4	7
		Paeoniaceae	Paeonia				1					1	2204	2	2	0	0	1	5
		Papaveraceae	Papaver					1				1	26523	21	28	4	5	6	64
		Plantaginaceae	Plantago			2	2	1		1	1	7	3						0
		Platanaceae	Platanus					1		1		2	7						0
		Plumbaginaceae	Limonium			5	5	4	5	3	1	23	5534	0	2	1	1	0	4
			Plumbago							1		1	1842	2	1	0	0	6	9

		Podostemaceae	Polypleurum					1	3	6	4	14	9679	0	0	1	0	0	1
		Polygonaceae	Persicaria							10		10	3538	5	4	2	0	2	13
		Primulaceae	Primula				1	1	1	3		6	5672	6	4	0	3	5	18
			Aegiceras						1	1		2	68						0
		Ranunculaceae	Aquilegia			2				2	1	5	115666	73	217	35	14	46	385
			Adonis								1	1	4189	0	2	0	0	2	4
		Rhamnaceae	Ziziphus				1	3		1	1	6	3401	5	7	0	0	0	12
		Rhizophoraceae	Bruguiera			1				5		6	22543	48	49	6	8	7	118
			Rizophora						1	5		6	331						0
			Ceriops						1	3		4	101						0
			Kandelia							2		2	1						0
		Rosaceae	Malus				1	2			4	7	336190	183	67	17	24	31	322
			Prunus			2	7	7	3	7	4	30	111788	217	63	17	15	32	344
			Rubus					2				2	3061	9	10	1	0	6	26
			Pyrus			3	1	1	3		5	13	1942	11	2	1	0	1	15
			Photonia					2				2	44						0
			Potentilla					1	1		2	4	1						0
		Rubiaceae	Coffea				2	8				10	254474	102	94	19	16	27	258
			Oldenlandia	1				1		4	2	8	1117	0	1	1	0	0	2
			Psychotria			13	17	29	4	2	10	75	10						0
		Rutaceae	Citrus			1						1	568700	126	123	26	21	46	342
		Salicaceae	Populus			1	1	1	1	1		5	422517	176	207	51	45	88	567
			Salix			1	1	2	1	3		8	707						0
		Santalaceae	Santalum	1			1	2				4	95						0
		Sapindaceae	Paullinia					1				1	14888	5	3	0	1	2	11
			Dimocarpus						1			1	66	1	0	0	0	0	1
		Scrophulariaceae	Antirrhinum			1	1					2	25310	11	30	5	7	12	65
			Mimulus							2		2	279620	225	253	37	23	52	590
			Torenia							1		1	256	0	0	1	2	0	3
		Solanaceae	Solanum			6	6	11	14	2	8	47	725742	71	126	17	25	85	324
			Nicotiana							1		1	413102	111	148	16	25	65	365
			Withania							2		2	741	0	0	0	1	0	1
		Tamaricaceae	Tamarix							1		1	22712	8	6	5	4	4	27
		Theaceae	Camellia				1	10				11	49902	109	43	9	11	26	198
		Tiliaceae	Corchorus							1		1	840	0	1	0	0	0	1
		Tropaeolaceae	Tropaeolum			1	1	6			1	9	10507	0	3	0	0	1	4

		Ulmaceae	Ulmus			1	1	2				4	1277	0	1	0	0	0	1
		Umbelliferace	Daucus								1	1	18137	3	1	0	0	3	7
			Centella							1		1	443	2	3	0	0	1	6
			Apium			1			1	3		5	2336	1	3	0	0	0	4
			Blupleurum			3	1	1				5	237	5	2	0	0	0	7
			Angelica							1		1	63						0
		Urticaceae	Boehmeria						1			1	421	0	1	0	0	1	2
		Violaceae	Viola	1		2	1	1	1	1		7	49						0
		Vitaceae	Vitis							1		1	162674	69	73	18	23	47	230

CONCLUSIONS

CONCLUSIONS

General conclusions

Throughout the chapters of this thesis, various molecular tools were used to study the genetic variation and population structure of rare and/or threatened species. Results derived from this thesis support the use of molecular markers for conservation purposes. Conservation actions such as defining management units or establishing minimum inter-plant distance for seed collection for *ex situ* germplasm collection require population genetic information. Results also highlight the importance of combining molecular markers with different modes of inheritance for designing accurate management strategies. Management measures based in one type of molecular marker only can sometimes overlook populations of conservation concern.

Specific conclusions

Chapter 1:

- Clonal growth seemed relatively restricted in *C. borjae* although clonal diversity differed among populations and the northernmost ones have a higher abundance of clones. The only consistent difference between populations with higher and lower clonal incidence was the geological substratum. Northernmost populations occur on serpentine soils and it is speculated that these soils may affect plant growth by favoring clonal propagation.
- No evidences of genetic impoverishment were detected in *Centaurea borjae*. Instead, our data revealed relatively high levels of genetic variation at species and at population level. Diversity levels detected in *C. boraje* were comparable to those obtained in plants with similar life-history traits and fell within the range of values inferred for other endemic members of the genus *Centaurea* investigated with dominant markers.

CONCLUSIONS

- We found evidence of restricted gene flow among populations, in agreement with the poor dispersal abilities attributed to *C. borjae*. Likewise, the fine-scale SGS found in *Centaurea borjae* indicates that rosette leaves at close distances can be more related than spatially random pairs. The results fitted again with the expectations for a plant with low dispersal capabilities, clonal reproduction, and/or low density. For germplasm collection, rosettes separated <80 m should be generally avoided tp prevent collecting genetically close plants and/or clone mates.
- AFLP data consistently identified four genetic clusters that were designated as an independent management unit based on the restricted gene flow among populations detected and their genetic uniqueness. One MU was formed by the three central populations PC-OC-OBB, while the remaining three MUs encompassed one population each.

Chapter 2:

- Unlike AFLPs, chloroplast sequence data provided some evidence of genetic depletion in *C. borjae*. The incongruence between AFLP and cpDNA data was attributed to differences in mutation rate and effective population size.
- Like in the AFLP study, gene flow was low. In fact, estimates with cpDNA were lower than with AFLPs and seem consistent with several biological traits of *C. borjae*: lack of pappus, probable myrmecochory, and low germination success.
- The current arrangement of haplotypes suggest that the species might have persisted for a longer period of time at the center of its current distribution range.
- The uneven distribution of cpDNA polymorphism among populations leads to prioritizing four enclaves in terms of their contribution to haplotype richness and diversity: LI, VH, OB and PC. By preserving these four populations, all known haplotypes will be maintained. These results complement prior findings with nuclear markers because cpDNA data reveal that PC and OB have private alleles and are not interchangeable in conservations terms. Likewise important, the four populations

identified as priority by cpDNA only included three of the four MUs designated with nuclear markers. The excluded MU was the geographically isolated PR that, according to AFLP, has a certain level of uniqueness (a private band and noticeably different marker frequencies).

Chapter 3:

- Both AFLP and cpDNA recorded an extremely low genetic diversity in *Omphalodes littoralis* spp. *gallaecica* and minimal gene flow among populations. It is speculated that this pattern may be a consequence of strong genetic drift within populations.
- Still, cpDNA data suggests that the various local demes might have been connected in a distant past.
- The pattern of low genetic diversity and strong differentiation seems stable on consecutive years, suggesting the inability of the seed bank to act as a reservoir of hidden genetic diversity.
- The various populations differed in a number of quantitative traits and reciprocal transplant experiments indicated that these differences had a genetic component. However, the variation in quantitative traits could not be attributed to local adaptation.
- From a conservation perspective, the combination of genetic and quantitative trait analysis led to the designation of five Evolutionary Management Units (ESUs) and each population is recommended to be considered as a single ESUs given its molecular and phenotypic uniqueness.

Chapter 4:

- Trimers, followed by dimers, were the commonest SSR motifs in EST sequences of the control genomes of *Arabidopsis* and *Oryza.* We found differences in the type of motif between monocots and dicots: monocots were abundant in GC-rich motif.
- In general, EST-SSRs derived from control genomes were mostly located in coding regions. However, trimmers and hexamers were commonly found in CDS

regions while other motifs were mainly located in UTRs and, to a lesser extent, other non-coding regions.

- EST-SSRs with primers were found for 222 out of 257 genera of threatened plants.
- Trimers were also the commonest nucleotide repeats in IUCN genera but the frequency of the various types of SSR repeat differed among the studied taxonomic groups. Results for Angyospermae were consistent with those found in the control genomes where trimmers and dimers were the most abundant but the Acrogymnospermae revealed a high proportion of hexamers.
- Empirical tests indicate that our EST-SSRs have notable amplification success and very high transferability between congenerics, supporting the use of existing EST databases for developing SSRs in non-model plants as bench tool for evolutionary and/or conservation studies of population geneticists and molecular ecologists.

BIBLIOGRAPHY

BIBLIOGRAPHY

Introduction

- Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y and Lobo J (2008). Genetic consequences of habitat fragmentation in plant populations: susceptibles signals in plant traits and methodological approaches. Molecular Ecology 17:5177-5188.
- Allendorf FW, Hohenlohe PA and Luikart G (2010). Genomics and the future of conservation genetics. Nature Reviews Genetics 11:697-709.
- Allendorf FW and Luikart G (2013). Conservation and the Genetics of Populations. Blackwell Pub., Malden, MA.
- Andersen JR and Lübberstedt T (2003). Functional markers in plants. Trends in Plant Science 8:554-560.
- Angeloni F, Ouborg NJ and Leimu R (2011). Meta-analysis on the association of population size and life history with inbreeding depression in plants. Biological Conservation 144:35-43.
- Armbruster P and Reed DH (2005). Inbreeding depression in benign and stressful environments. Heredity 95:235-242.
- Avise JC (1994). Molecular Markers, Natural History and Evolution. Chapman & Hall, New York.
- Avise JC (1995). MItochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. Conservation Biology 9:686-690.
- Avise JC (2004). Molecular Markers, Natural History, and Evolution. Sinauer Associates, Sunderland, Mass.
- Bañares A, Blanca G, Güemes J, Moreno JC, Ortiz S (2004). Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Dirección General de Conservación de la Naturaleza, Madrid
- Bekessy SA, Ennos RA, Burgman MA, Newton AC and Ades PK (2003). Neutral DNA markers fail to detect genetic divergence in an ecologically important trait. Biological Conservation 110:267-275.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C and Taberlet P (2004). How to track and assess genotyping errors in population genetics studies. Molecular Ecology 13:3261-3273.
- Cain ML, Milligan BG and Strand AE (2000). Long-distance seed dispersal in plant populations. American Journal of Botany 87:1217-1227.
- Cole CT (2003). Genetic variation in rare and common plants. Annual Review of Ecology Evolution and Systematics 34:213-237.
- Cousens R, Dytham C, Law R (2008). Dispersal in Plants A Population Perspective. Oxford University Press, Oxford, UK.
- Crandall KA, Bininda-Emonds ORP, Mace GM and Wayne RK (2000). Considering evolutionary processes in conservation biology. Trends in Ecology & Evolution 15:290-295.
- Crnokrak P and Barret SCH (2002). Purging the genetic load: a review of the experimental evidence. Evolution 56:2347-2358.
- Crnokrak P and Roff DA, (1999). Inbreeding depression in the wild. Heredity 83:260-270.
- Duminil J, Fineschi S, Hampe A, Jordano P, Salvini D, Vendramin GG and Petit RJ (2007). Can population genetic structure be predicted from life-history traits? American Naturalist 169: 662-672.
- Duminil J, Hardy OJ and Petit RJ (2009). Plant traits correlated with generation time direclty affect inbreeding depression and mating system and indirectly genetic structure. BCM Evolutionary Biology 9:177-191.
- Ellis JR and Burke JM (2007). EST-SSRs as a resource for population genetic analyses. Heredity 99:125-132.
- Ellstrand NC and Elam DR (1993). Population genetic consequences of small population size: implications for plant

conservation. Annual Review of Ecology and Systematics 24:217-242.

- Fernández Casas J, Susanna A (1986). Monografía de la sección *Chamaecyanus* Willk. del género *Centaurea*. Treballs de l'Institut Botànic de Barcelona 10:1-174
- Frankel OH and Soule ME (1981). Conservation ans evolution. Cambridge University Press, Cambridge, England.
- Frankham R (2005). Genetics and extinction. Biological Conservation 126:131-140.
- Frankham R (2010). Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation 143:1919-1927.
- Frankham R, Briscoe DA and Ballou JD (2010). Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.
- Garcia C, Jordano and Godoy JA (2007). Contemporary pollen and seed dispersal in a *Prunus mahaleb* population: patterns in distance and direction. Molecular Ecology 16:1947-1955.
- Garcia-Jacas N, Susanna A (1992). Karyological Notes on *Centaurea* sect. *Acrocentron* (Asteraceae). Plant Systematics and Evolution 179:1-18
- Gitzendanner MA and Soltis PS (2000). Patterns of genetic variation in rare and widespread plant congeners. American Journal of Botany 87:783-792.
- Glémin S (2003). How are deleterious mutations purged? Drift versus nonrandom mating. Evolution 57:2678-2687.
- Gómez-Orellana Rodríguez L (2004). *Cenaturea borjae* Valdés Verm. & Rivas Goday. In: Bañares A, Blanca G, Güemes J, Moreno JC, Ortiz S eds Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Dirección General de Conservación de la Naturaleza, Madrid, 632-633
- Gómez-Orellana Rodríguez L (2011). *Centaurea borjae*. IUCN Red List of Threatened Species, Version 2011.1. Avaliable from www.iucnredlist.org
- Gomez C and Espadaler X (1998). *Myrmecochorous* dispersal distances: a world survey. Journal of Biogeography 25:573-580
- Goodwillie C, Kalisz S and Eckert GC (2005). The evolutionary enigma of mixed mating systems in plants: Occurrence, theoretical explanations, and empirical evidence. Annual Review of Ecology Evolution and Systematics 36:47-79
- Hamrick JL (1983). The distribution of Genetic Variation within and among Plant Populations. Pages 500-508 in Schonewald-Cox CD, Chambers SM, MacBryde B and Thomas L, editors. Genetics and Conservation. Benjamin/Cummings, London.
- Hamrick JL and Godt MJW (1990). Allozyme diversity in plant species. Pages 43-63 in Brown AHD, Clegg MT, Kahler AL, and Weir BS, editors. Plant populations genetics, breeding and genetic resources. Sinauer, Sunderland.
- Hamrick JL and Godt MJW (1996). Conservation Genetics of Endemic Plant Species. Pages 281-304 in Avise JC and Hamrick JL, editors. Conservation Genetics: Case Histories from Nature. Chapman & hall, New York.
- Hamrick JL, Godt MJW, Murawski DA and Loveless MD (1991). Correlation between species traits and allozyme diversity: implications for conservation biology. Pages 75-86 in Falk DAI and Holsinger KE, editors. Genetics and conservation of rare plants. Oxford University Press, US.
- Höglund J (2009). Evolutionary conservation genetics. Oxford University Press, Oxford.
- Honnay O and Jacquemyn H (2007). Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. Conservation Biology 21:823-831.
- Huenneke LF (1991). Ecological implications of genetic variation in plant populations. Pages 30-44 in Falk DAI and Holsinger KE, editors. Genetics and the conservation of rare plants. Oxford University Press, US.
- Izco J, Rodríguez Oubiña J, Romero MI, Amigo J, Pulgar I, Gomez M (2003). Flora endémica de A Coruña España: *Centaurea borjae* y *Centaurea ultreiae.* Diputación Provincial A Coruña, A Coruña

- Kawecki TJ and Ebert D (2004). Conceptual issues in local adaptation. Ecology Letters 7:1225-1241.
- Keller LF and Waller DM (2002). Inbreeding effects in wild populations. Trends in Ecology & Evolution 17:230-241.
- Krebs CJ 1972. Ecology. The Experimental Analysis of Distribution and Abundance. Harper and Row, New York.
- Landguth EL and Balkenhol N (2012). Relative sensitivity of neutral versus adaptive genetic data for assessing population differentiation. Conservation Genetics 13:1421-1426.
- Leimu R and Fischer M (2008). A Meta-Analysis of Local Addaptation in Plants. PLoS ONE 3:e4010.
- Leimu R, Mutikainen P, Koricheva J and. Fischer M (2006). How general are positive relationships between plant population size, fitness and genetic variation? Journal of Ecology 94:942-952.
- Loveless MD and Hamrick JL (1984). Ecological determinants of genetic structure in plant populations. Annual Review of Ecology and Systematics 15:65-95.
- Luikart G, England PR, Tallmon DA, Jordan S and Taberlet P (2003). The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics 4:981-994.
- Mba C and Tohme J (2005). Use of AFLP markers in surveys of plant diversity. Methods in Enzymology:177-201.
- McCauley DE (1995). The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends in Ecology and Evolution 10:198-202.
- McCue KA and Holtsford TP (1998). Seed bank influences on genetic diversity in the rare annual *Clarkia springvillensis* (Onagraceae). American Journal of Botany 85: 30-36.
- McNeely JA, Miller KR, Reid WV, Mittermeier RA, and Werner TB (1990). Strategies for Conserving Biodiversity. Environment 32:17-40.
- Mills LS (2006). Conservation of wildlife populations. Demography, Genetics, and Management. Wiley-Blackwell, Singapore.
- Ministerio de Medio Ambiente y Medio Rural y Marino (2011). Real Decreto 139/2011, de 4 de febrero, para el desarrollo del Listado de Especies Silvestres en Régimen de Protección Especial y del Catálogo Español de Especies Amenazadas. Boletin Oficial del Estado 46:20912-20951
- Moritz C (1994). Defining "Evolutionarily-Significant-Units" for conservation. Trends in Ecology & Evolution 9:373-375.
- Moritz C (1999). Conservation units and translocations: strategies for conserving evolutionary processes. Hereditas 130:217-228.
- Mullis K, Faloona FA, Scharf S, Saiki R, Horn G and Erlich H (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology 51:263-273.
- Mullis KB and Faloona FA (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology 155:335-350.
- Nunney L (2002). The Effective Size of Annual Plant Populations: The Interaction of a Seed Bank with Fluctuating Population Size in Maintaining Genetic Variation. The American Naturalist 160, 195-204.
- Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13:1143-1155.
- Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK and Hedrick PW (2010). Conservation genetics in transition to conservation genomics. Trends in Genetics 26:177-187.
- Ouborg NJ, Piquot and Van Groenendael JM (1999). Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology 87:551-568.
- Palacios C, Kresovich S and Gonzalez-Candelas F (1999). A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism (AFLP). Molecular Ecology 8:645-657.
- Palsboll PJ , Berube M, and Allendorf FW (2007). Identification of management units

using population genetic data. Trends in Ecology & Evolution 22:11-16.

- Primmer CR (2009). From Conservation Genetics to Conservation Genomics. Pages 357-368 in Ostfeld RS and Schlesinger WH, editors. Ecology and Conservation Biology.
- Reed DH and Frankham R (2001). How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55:1095-1103.
- Reed DH and Frankham R (2003). Correlation between fitness and genetic diversity. Conservation Biology 17:230-237.
- Romero Buján MI (2005). Flora endémica amenazada del litoral de Galicia: una visión actual. Recursos Rurais Series Cursos 2: 1-10.
- Schlötterer C (2004). The evolution of molecular markers - just a matter of fashion? Nature Reviews Genetics 5:63-69.
- Schötterer C (1998). Microsatellites. Pages 237-261 in Hoelzel AR, editor. Molecular Genetic Analysis of Populations. A Practical Approach. IRL Press at Oxford University Press, Oxford.
- Selkoe KA and Toonen RJ (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9:615-629.
- Serrano M and Carbajal R (2011). *Omphalodes littoralis* subsp. *gallaecica*. (2011, editor. IUCN Red List of Threatened Species. Version 2011.1.
- Soltis PS, Soltis DE (2000). The role of genetic and genomic attributes in the success of polyploids. PNAS 97:7051-7057
- Taberlet P, Gielly L, Pautou G and Bouvet J (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17:1105-1109.
- Tallmon DA, Luikart G and Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. Trends in ecology and evolution 19:489–496
- Turner ME, Stephens JC and Anderson WW (1982). Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. PNAS 79:203-207.
- Valdés-Bermejo E, Agudo Mata MP (1983). Estudios cariológicos de especies ibéricas del género *Centaurea* L. (Compositae). I. Anales del Jardín Botánico de Madrid 40:119-142.
- Valdés-Bermejo E, Rivas Goday S (1978). Estudios en el género *Centaurea* L. (Compositae): *C. borjae* sp. nov. (Sect. Borjae Sect. Nov). Anales del Instituto Botánico J. Cavanilles 85: 159–164.
- Varshney RK, Graner A and Sorrells ME (2005a). Genic microsatellite markers in plants: features and applications. Trends in Biotechnology 23:48-55.
- Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, and Graner A (2005b). Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Science 168:195-202.
- Vekemans X, and Hardy OJ (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. Molecular Ecology 13:921-935.
- Willi Y, Van Buskirk J and Hoffmann AA (2006). Limits to the adaptive potential of small populations. Annual Review of Ecology Evolution and Systematics 37:433-458.
- Wright S (1931). Evolution in Mendelian populations. Genetics 16:97-159.
- Wright S (1943). Isolation by distance. Genetics 28:114-138.
- Wright S (1978). Evolution and the Genetics of Populations. University of Chicago Press, Chicago.

Chapter 1

- Abbott RJ, Ireland HE and Rogers HJ (2007). Population decline despite high genetic diversity in the new allopolyploid species *Senecio cambrensis* (Asteraceae). Molecular Ecology 16:1023-1033.
- Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y and Lobo J (2008). Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Molecular Ecology 17:5177-5188.
- Andreakis N, Kooistra WHCF and Procaccini G (2009). High genetic diversity and connectivity in the polyploid invasive seaweed *Asparagopsis taxiformis* Bonnemaisoniales in the Mediterranean, explored with microsatellite alleles and multilocus genotypes. Molecular Ecology 18:212-226.
- Antao T and Beaumont MA (2011). Mcheza: a workbench to detect selection using dominant markers. Bioinformatics 27: 1717-1718.
- Arnaud-Haond S, Duarte CM, Alberto F and Serrao EA (2007). Standardizing methods to address clonality in population studies. Molecular Ecology 16:5115-5139.
- Bañares A, Blanca G, Güemes J, Moreno JC and Ortiz S (2004). Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Dirección General de Conservación de la Naturaleza, Madrid.
- Barnaud A and Houliston GJ (2010). Population genetics of the threatened tree daisy *Olearia gardneri* (Asteraceae), conservation of a critically endangered species. Conservation Genetics 11:1515-1522.
- Beaumont MA and Balding DJ (2004). Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology 13: 969-980.
- Beaumont MA and Nichols RA (1996). Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society B: Biological Sciences 263: 1619-1626.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon C, Brochmann C and Taberlet P (2004). How to track and assess genotyping errors in population genetics studies. Molecular Ecology 13: 3261-3273.
- Bonin A, Ehrich D and Manel S (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Molecular Ecology 16:3737-3758.
- Bonin A, Pompanon F and Taberlet P (2005). Use of amplified fragment length polymorphism AFLP markers in surveys of vertebrate diversity. Methods in Enzymology 395:145-161.
- Bruvo R, Michiels NK, D'Souza TG and Schulenburg H (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Molecular Ecology 13:2101-2106.
- Chung MG and Epperson BK (1999). Spatial genetic structure of clonal and sexual reproduction in populations of *Adenophora grandiflora* (Campanulaceae). Evolution 53:1068-1078.
- Chung MY, Suh Y, Lopez-Pujol J, Nason JD and Chung MG (2005). Clonal and fine-scale genetic structure in populations of a restricted Korean endemic, *Hosta jonesii* (Liliaceae) and the implications for conservation. Annals of Botany 96:279-288.
- Colas B, Olivieri I and Riba M (1997). *Centaurea corymbosa*, a cliff-dwelling species tottering on the brink of extinction: A demographic and genetic study. PNAS 94:3471-3476.
- Cole CT (2003). Genetic variation in rare and common plants. Annual Review of Ecology, Evolution, and Systematics 34:213-237.
- Corander J, Siren J and Arjas E (2008). Bayesian spatial modeling of genetic population structure. Computational Statistics 23:111-129.
- Cousens R, Dytham C and Law R (2008). Dispersal in Plants A Population Perspective. Oxford University Press, Oxford, UK.
- Crandall KA, Bininda-Emonds ORP, Mace GM and Wayne RK (2000). Considering evolutionary processes in conservation biology. Trends in Ecology and Evolution 15:290-295.
- Despres L, Loriot S and Gaudeul M (2002). Geographic pattern of genetic variation in the European globeflower *Trollius europaeus* L. (Ranunculaceae) inferred from amplified fragment length polymorphism markers. Molecular Ecology 11:2337-2347.

- Douhovnikoff V and Dodd RS (2003). Intra-clonal variation and a similarity threshold for identification of clones: application to *Salix exigua* using AFLP molecular markers. Theoretical and Applied Genetics 106:1307-1315.
- Dyer RJ (2009). GeneticStudio: a suite of programs for spatial analysis of genetic-marker data. Molecular Ecology Resources 9:110-113.
- Dyer RJ and Nason JD (2004). Population Graphs: the graph theoretic shape of genetic structure. Molecular Ecology 13:1713-1727.
- Dyer RJ, Nason JD and Garrick RC (2010). Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. Molecular Ecology 19:3746-3759.
- Ellstrand NC and Elam DR (1993). Population genetic consequences of small population size: implications for plant conservation. Annual Review of Ecology, Evolution, and Systematics 24:217-242.
- Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology 14:2611-2620.
- Excoffier L, Hofer T and Foll M (2009). Detecting loci under selection in a hierarchically structured population. Heredity 103: 285-298.
- Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 21:479-491.
- Falush D, Stephens M and Pritchard JK (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567-1587.
- Fernández Casas J and Susanna A (1986). Monografía de la sección *Chamaecyanus* Willk. del género *Centaurea*. Treballs de l'Institut Botànic de Barcelona 10:1-174.
- Foll M, Gaggiotti O (2006). Identifying the environmental factors that determine the genetic structure of populations. Genetics 174: 875-891.
- Foll M and Gaggiotti O (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. Genetics 180: 977-993.
- Font M (2007). Poliploïdia, filogènia i biogeografia en *Centaurea* L. secció Acrocentron Cass. DC., Universitat de Barcelona, Barcelona.
- Font M, Garcia-Jacas N, Vilatersana R, Roquet C and Susanna A (2009). Evolution and biogeography of *Centaurea* section Acrocentron inferred from nuclear and plastid DNA sequence analyses. Annals of Botany 103:985-997.
- Frankel OH, Brown AHD, Burdon JJ (1995). The Conservation of Pant Biodiversity. Cambridge University Press, Cambridge, UK.
- Frankham R (2005). Genetics and extinction. Biological Conservation 126:131-140.
- Frankham R (2010). Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation 143:1919-1927.
- Frankham R, Briscoe DA and Ballou JD (2002) Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.
- Garcia-Jacas N and Susanna A (1992). Karyological Notes on *Centaurea* sect. *Acrocentron* (Asteraceae). Plant Systematics and Evolution 179:1-18.
- Garcia-Verdugo C, Fay MF, Granado-Yela C, Rubio de Casas R, Balaguer L, Besnard G and Vargas P (2009). Genetic diversity and differentiation processes in the ploidy series of *Olea europaea* L.: a multiscale approach from subspecies to insular populations. Molecular Ecology 18:454-467.
- Garroway CJ, Bowman J, Carr D and Wilson PJ (2008). Applications of graph theory to landscape genetics. Evolutionary Applications 1:620-630.
- Ghazoul J (2005). Pollen and seed dispersal among dispersed plants. Biological Reviews 80:413-443.
- Gitzendanner MA and Soltis PS (2000). Patterns of genetic variation in rare and widespread plant congeners. American Journal of Botany 87:783-792.
- Gomez C and Espadaler X (1998). Myrmecochorous dispersal distances: a world survey. Journal of Biogeography 25:573-580.
- Gómez-Orellana Rodríguez L (2004). *Cenaturea borjae* Valdés Verm. & Rivas

Goday. In: Bañares A, Blanca G, Güemes J, Moreno JC, Ortiz S eds Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Dirección General de Conservación de la Naturaleza, Madrid, 632-633.

- Gómez-Orellana Rodríguez L (2011). *Centaurea borjae*. IUCN Red List of Threatened Species, Version 2011.1. Avaliable from www.iucnredlist.org.
- Hamrick JL and Godt MJW (1996). Conservation Genetics of Endemic Plant Species. Pages 281-304 in Avise JC and Hamrick JL, editors. Conservation Genetics: Case Histories from Nature. Chapman & hall, New York.
- Hardy OJ, González-Martinez SC, Fréville H, Boquien G, Mignot A, Colas B and Olivieri I (2004). Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae) I. Pattern of pollen dispersal. Journal of Evolutionary Biology 17:795-806.
- Hardy OJ and Vekemans X (2002). SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2:618–620.
- Honnay O and Jacquemyn H (2008). A meta-analysis of the relation between mating system, growth form and genotypic diversity in clonal plant species. Evolutionary Ecology 22:299-312.
- Hubisz MJ, Falush D, Stephens M and Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322-1332.
- Imbert E (2006). Dispersal by ants in *Centaurea corymbosa* (Asteraceae): What is the elaiosome for? Plant Species Biology 21:109-117.
- Izco J, Rodríguez Oubiña J, Romero MI, Amigo J, Pulgar I and Gomez M (2003). Flora endémica de A Coruña España: *Centaurea borjae* y *Centaurea ultreiae*. Diputación Provincial A Coruña, A Coruña.
- Jakobsson M and Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806.
- Jensen JL, Bohonak AJ and Kelley ST (2005). Isolation by distance, web service. BMC Genetics 6:13
- Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G and Taberlet P (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Molecular Ecology 16: 3955-3969.
- Jump AS, Peñuelas J (2007). Extensive spatial genetic structure revealed by AFLP but not SSR molecular markers in the wind-pollinated tree, *Fagus sylvatica*. Molecular Ecology 16:925-936.
- Kato S, Iwata H, Tsumura Y and Mukai Y (2011). Genetic structure of island populations of *Prunus lannesiana* var. *speciosa* revealed by chloroplast DNA, AFLP and nuclear SSR loci analyses. Journal of Plant Research 124:11-23.
- Kim SC, Lee C and Santos-Guerra A (2005). Genetic analysis and conservation of the endangered Canary Island woody sow-thistle, *Sonchus gandogeri* (Asteraceae). Journal of Plant Research 118:147-153.
- Kloda JM, Dean PDG, Maddren C, MacDonald DW and Mayes S (2008). Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: an example from British Restharrows *Ononis spinosa* and *Ononis repens*. Heredity 100:253-260.
- Kosman E (2003). Nei's gene diversity and the index of average differences are identical measures of diversity within populations. Plant Pathology 52:533-535.
- Kosman E and Leonard KJ (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. Molecular Ecology 14:415-424.
- Kruckeberg AR and Rabinowitz D (1985). Biological aspects of endemism in higher plants. Annual Review of Ecology, Evolution, and Systematic 16:447-479.
- Lavergne S, Thompson JD, Garnier E and Debussche M (2004). The biology and ecology of narrow endemic and widespread plants: a comparative study of trait variation in 20 congeneric pairs. Oikos 107:505-518.
- Li XX, Ding XY, Chu BH, Zhou Q, Ding G and Gu S (2008). Genetic diversity analysis and conservation of the endangered Chinese endemic herb *Dendrobium officinale* Kimura et Migo (Orchidaceae) based on AFLP. Genetica 133:159-166.

- Loiselle BA, Sork VL, Nason J and Graham C (1995). Spatial genetic-structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). American Journal of Botany 82:1420–1425.
- Marko PB, Hart MW (2011). The complex analytical landscape of gene flow inference. Trends in Ecology and Evolution 26:448-456.
- Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Bajer MS, Rieseberg LH and Otto SP (2011). Recently Formed Polyploid Plants Diversify at Lower Rates. Science 333:1257-1257.
- Mba C and Tohme J (2005). Use of AFLP markers in surveys of plant diversity. Methods in Enzymology 395:177-201.
- Meirmans PG and Van Tienderen PH (2004). Genotype and Genodive: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4:792-794.
- Ministerio de Medio Ambiente y Medio Rural y Marino (2011). Real Decreto 139/2011, de 4 de febrero, para el desarrollo del Listado de Especies Silvestres en Régimen de Protección Especial y del Catálogo Español de Especies Amenazadas. Boletin Oficial del Estado 46:20912-20951.
- Minor ES and Urban DL (2007). Graph theory as a proxy for spatially explicit population models in conservation planning. Journal of Applied Ecology 17:1771-1782.
- Morden CW and Loeffler WF (1999). Fragmentation and genetic differentiation among subpopulations of the endangered Hawaiian mint *Haplostachys haplostachya* (Lamiaceae). Molecular Ecology 8:617-625.
- Moritz C (1994). Defining "Evolutionarily-Significant-Units" for conservation. Trends in Ecology and Evolution 9:373-375.
- Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13:1143-1155.
- Obbard DJ, Harris SA and Pannell JR (2006). Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. Heredity 97:296-303.
- Palacios C, Kresovich S and Gonzalez-Candelas F (1999). A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism AFLP. Molecular Ecology 8:645-657.
- Peakall R and Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.
- Peters MD, Xiang QY, Thomas DT, Stucky J and Whiteman NK (2009). Genetic analyses of the federally endangered *Echinacea laevigata* using amplified fragment length polymorphisms AFLP- Inferences in population genetic structure and mating system. Conservation Genetics 10:1-4.
- Pisanu S, Filigheddu R and Farris E (2009). The conservation status of an endemic species of northern Sardinia: *Centaurea horrida* Badaro (Asteraceae). Plant Biosystems 143:275-282.
- Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155:945-959.
- Reed DH and Frankham R (2003). Correlation between fitness and genetic diversity. Conservation Biology 17:230-237.
- Roiloa SR and Retuerto R (2006). Physiological integration ameliorates effects of serpentine soils in the clonal herb *Fragaria vesca*. Physiologia Plantarum 128:662-676.
- Sebbenn AM, Carvalho ACM, Freitas MLM, Moraes SMB, Gaino APSC, da Silva JM, Jolivet C and Moraes MLT (2011). Low levels of realized seed and pollen gene flow and strong spatial genetic structure in a small, isolated and fragmented population of the tropical tree *Copaifera langsdorffii* Desf. Heredity 106:134-145.
- Silvertown J (2008). The evolutionary maintenance of sexual reproduction: Evidence from the ecological distribution of asexual reproduction in clonal plants. International Journal of Plant Science 169:157-168.
- Soltis PS and Soltis DE (2000). The role of genetic and genomic attributes in the success of polyploids. PNAS 97:7051-7057.
- Soñora X (1994). Nueva localidad de *Centaurea borjae* Valdés-Bermejo & Rivas Goday. Lazaroa 14:183-184.
- Sözen E and Özaydin B (2009). A preliminary study of the genetic diversity of the critically endangered *Centaurea nivea* (Asteraceae). Annals of Botany Fennici 46:541-548.
- Sözen E and Özaydin B (2010). A study of genetic variation in endemic plant

Centaurea wiedemanniana by using RAPD markers. Ekoloji 19:1-8.

- Stefenon VM, Gailing O and Finkeldey R (2008). Genetic structure of plantations and the conservation of genetic resources of Brazilian pine *Araucaria angustifolia*. Forest Ecology Management 255:2718-2725.
- Stehlik I and Holderegger R (2000). Spatial genetic structure and clonal diversity of *Anemone nemorosa* in late successional deciduous woodlands of Central Europe. Journal of Ecology 88:424-435.
- Takahashi Y, Takahashi H and Maki M (2011). Comparison of genetic variation and differentiation using microsatellite markers among three rare threatened and one widespread toad lily species of *Tricyrtis* section *Flavae* (Convallariaceae) in Japan. Plant Species Biology 26:13-23.
- Tallmon DA, Luikart G and Waples RS (2004). The alluring simplicity and complex reality of genetic rescue. Trends in Ecology and Evolution 19:489-496.
- Tero N, Aspi J, Siikamäki P, Jäkäläniemi A, Tuomi J (2003). Genetic structure and gene flow in a metapopulation of and endangered plant species, *Silene tatarica*. Molecular Ecology 12:2073-2085.
- Valdés-Bermejo E, Agudo Mata MP (1983). Estudios cariológicos de especies ibéricas del género *Centaurea* L. (Compositae). I. Anales del Jardín Botánico de Madrid 40:119-142.
- Valdes-Bermejo E, Rivas Goday S 1978 Estudios en el género *Centaurea* L. (Compositae): *C. borjae* sp. nov. (Sect. *Borjae Sect. Nov*.). Anales del Instituto Botánico AJ. Cavanilles 85:159-164.
- Vallejo-Marin M, Dorken ME, Barrett SCH (2010). The Ecological and Evolutionary Consequences of Clonality for Plant Mating. Annual Review of Ecology, Evolution, and Systematics 41:193-213.
- Vekemans X and Hardy OJ (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. Molecular Ecology 13:921–35.
- Vilatersana R, Susanna A and Brochmann C (2007). Genetic variation in *Femeniasia* (Compositae, Cardueae), an endemic and endangered monotypic genus from the Balearic Islands Spain. Botanical Journal of the Linnean Society 153:97-107.
- Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.
- Whitlock MC and McCauley DE (1999). Indirect measures of gene flow and migration: $F_{ST} \neq 1/4Nm+1$. Heredity 82:117-125.
- Willi Y, Van Buskirk J and Hoffmann AA (2006). Limits to the adaptive potential of small populations. Annual Review of Ecology, Evolution, and Systematics 37:433-458.
- Winfield MO, Arnold GM, Cooper F, Le Ray M, White J, Karp A and Edwards KJ (1998). A study of genetic diversity in *Populus nigra* subsp. *betulifolia* in the upper Severn area of the UK using AFLP markers. Molecular Ecology 7:3-10.
- Yan XB, Guo YX, Zhao C, Liu FY and Lu BR (2009). Intra-population genetic diversity of two wheatgrass species along altitude gradients on the Qinghai-Tibetan Plateau: its implication for conservation and utilization. Conservation Genetics 10:359-367.
- Zeng LY, Xu LL, Tang SQ, Tersing T, Geng YP and Zhong Y (2010). Effect of sampling strategy on estimation of fine-scale spatial genetic structure in *Androsace tapete* (Primulaceae), an alpine plant endemic to Qinghai-Tibetan Plateau. Journal of Systematics and Evolution 48:257-26.

Table S1

- Bancheva S, Geraci A and Raimondo FM (2006). Genetic diversity in the *Centaurea cineraria* group *(Compositae)* in Sicily using isozymes. Plant Biosystems- An International Journal Dealing with all Aspects of Plant Biology 140: 10-16.
- Colas B, Olivieri I and Riba M (1997). *Centaurea corymbosa*, a cliff-dwelling species tottering on the brink of extinction: A demographic and genetic study. PNAS 3471-3476.
- Fréville H, Colas B, Ronfort J, Riba M and Olivieri I (1998). Predicting endemism from population structure of a widespread species: Case study in *Centaurea maculosa* Lam. (Asteraceae). Conservation Biology 12: 1269-1278.
- Fréville H, Justy F and Olivieri I (2001). Comparative allozyme and microsatellite population structure in a narrow endemic plant species, *Centaurea corymbosa* Pourret (Asteraceae). Molecular Ecology 10: 879-889.
- Garnatje T, Susanna A and Messeguer R (1998). Isozyme studies in the genus *Cheirolophus* (Asteraceae: Cardueae-Centaureinae) in the Iberian Peninsula, North Africa and the Canary Islands. Plant Systematics and Evolution 213: 57-70.
- Hardy OJ and Vekemans X (2001). Patterns of allozyme variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. Evolution 55: 943-954.
- Mameli G, Filigheddu R, Binelli G and Meloni M (2008). The genetic structure of the remnant populations of *Centaurea horrida* in Sardinia and associated islands. Annals of Botany 101: 633-640.
- Marrs RA, Sforza R and Hufbauer RA (2008a). Evidence for multiple introductions of *Centaurea stoebe micranthos* (spotted knapweed, Asteraceae) to North America. Molecular Ecology 17: 4197-4208.
- Marrs RA, Sforza R and Hufbauer RA (2008b). When invasion increases population genetic structure: a study with *Centaurea diffusa*. Biol Invasions 10: 561-572.
- Palermo AM, Pelegrino G, Musacchio A and Menale B (2002). Allozymic variability in *Centaurea tenorei* Guss. ex Lacaita and in other species of *C. parlatoris* Heldr. group (Asteraceae). Plant Biosystems- An International Journal Dealing with all Aspects of Plant Biology 136: 331-337.
- Sözen E, Özaydin B (2009). A preliminary study of the genetic diversity of the critically endangered *Centaurea nivea* (Asteraceae). Annals of Botany Fenn 46: 541-548
- Sözen E, Özaydin B (2010). A Study of Genetic Variation in Endemic Plant *Centaurea wiedemanniana* by Using RAPD Markers. Ekoloji 19: 1-8
- Sun M (1997). Population genetic structure of yellow starthistle (*Centaurea solstitialis*), a colonizing weed in the western United States. Canadian Jornal of Botany 75: 1470-1478.
- Vilatersana R, Susanna A and Brochmann C (2007). Genetic variation in *Femeniasia* (Compositae, Cardueae), an endemic and endangered monotypic genus from the Balearic Islands (Spain). Botanical Journal of the Linnean Society 153: 97-107.

Chapter 2

- Aizawa M, Yoshimaru H, Katsuki T and Kaji M (2008). Imprint of post-glacial history in a narrowly distributed endemic spruce, *Picea alcoquiana*, in central Japan observed in nuclear microsatellites and organelle DNA markers. Journal of Biogeography 35: 1295–1307.
- Allendorf FW and Luikart G (2007). Conservation and the genetics of populations. Malden: Blackwell Publisher.
- Artyukova EV, Kozyrenko MM, Kholina AB and Zhuravlev YN (2011). High chloroplast haplotype diversity in the endemic legume *Oxytropis chankaensis* may result from independent polyploidization events. Genetica 139: 221–232.
- Avise JC (2004). Molecular markers, natural history and evolution. Chapman & Hall, New York.
- Bañares A, Blanca G, Güemes J, Moreno JC and Ortiz S (2004). Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Madrid: Dirección General de Conservación de la Naturaleza.
- Bandelt HJ, Forster P and Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37–48.
- Bohonak AJ (2000). IBD (Isolation by distance): A program for analysis of isolation by distance. Journal of Heredity 93: 153–154.
- Bonin A, Ehrich D, Manel S (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Molecular Ecology 16: 3737–3758.
- Brussard PF (1984). Geographic patterns and environmental gradients: the central-marginal model in *Drosophila* revisited. Annual Review of Ecology and Systematics 15: 25–64.
- Colas B, Olivieri I, Riba M (1997). *Centaurea corymbosa*, a cliff-dwelling species tottering on the brink of extinction: A demographic and genetic study. PNAS 94: 3471–3476.
- Cole CT (2003). Genetic variation in rare and common plants. Annual Review of Ecology and Systematics 34: 213–237.
- Eckert CG, Samis KE and Lougheed SC (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Molecular Ecology 17: 1170–1188.
- Ellstrand NC and Elam DR (1993). Population genetic consequences of small population size: implications for plant conservation. Annual Review of Ecology and Systematics 24: 217–242.
- Excoffier L, Laval G and Schneider S (2005). Arlequin (ver. 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47–50.
- Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 21: 479–491.
- Fang HL, Guo QS, Shen HJ, Shao QS (2010). Phylogeography of *Chrysanthemum indicum* L. (Compositae) in China based on *trnL-F* sequences. Biochemical Systematics and Ecology 38: 1204–1211.
- Fernández Casas J and Sussana A (1986). Monografía de la sección *Chamaecyanus* Willk. del género Centaurea. Treballs de l'Institut Bótanic de Barcelona 10: 5–174.
- Frankham R, Briscoe DA, Ballou JD (2002). Introduction to conservation genetics. Cambridge: Cambridge University Press.
- Ge XJ, Hwang CC, Liu ZH, Huang CC, Huang WH, Hung KH, Wang WK and Chiang TY (2011). Conservation genetics and phylogeography of endangered and endemic shrub *Tetraena mongolica* (Zygophyllaceae) in Inner Mongolia, China. BMC Genetics 12: 1–12.
- Ghazoul J (2005). Pollen and seed dispersal among dispersed plants. Biological Reviews 80: 413–443.
- Gómez-Orellana Rodríguez L (2011). *Centaurea borjae*. IUCN Red List of Threatened Species, Version 2011.1. Avaliable from www.iucnredlist.org.
- Gong X, Luan SS, Hung KH, Hwang CC, Lin CJ, Chiang YC and Chiang TY (2011). Population structure of *Nouelia insignis* (Asteraceae), an endangered species in southwestern China, based on chloroplast DNA sequences: recent demographic shrinking. Journal of Plant Research 124: 221–230.
- Hamrick JL and Godt MJW (1996). Conservation Genetics of Endemic Plant

Species. Pages 281-304 in Avise JC and Hamrick JL, editors. Conservation Genetics: Case Histories from Nature. Chapman & hall, New York.

- Hardy OJ, González-Martinez SC, Fréville H, Boquien G, Mignot A, Colas B and Olivieri I (2004). Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae) I. Pattern of pollen dispersal. Journal of Evolutionary Biology 17: 795–806.
- Hudson RR, Slatkin M, Maddison WP (1992). Estimation of levels of gene flow from DNA-sequence data. Genetics 132: 583–589.
- Imbert E (2006). Dispersal by ants in *Centaurea corymbosa* (Asteraceae): What is the elaiosome for? Plant Species Biology 21: 109–117.
- Izco J, Rodríguez Oubiña J, Romero MI, Amigo J, Pulgar I, Gomez M (2003). Flora endémica de A Coruña España: *Centaurea borjae* y *Centaurea ultreiae.* Diputación Provincial A Coruña, A Coruña.
- Kato S, Iwata H, Tsumura Y, Mukai Y (2011). Genetic structure of island populations of *Prunus lannesiana* var. *speciosa* revealed by chloroplast DNA, AFLP and nuclear SSR loci analyses. Journal of Plant Research 124: 11–23.
- Kelchner SA (2000). The evolution of non-coding chloroplast DNA and its application in plant systematics. Annals of the Missouri Botanical Garden 87: 482–498.
- Lande R (1988). Genetics and demography in biological conservation. Science 241: 1455–1460.
- Librado P and Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.
- Liu F, Zhao SY, Li W, Chen JM and Wang QF (2010). Population genetic structure and phylogeographic patterns in the Chinese endemic species *Sagittaria lichuanensis*, inferred from cpDNA *atpB-rbc*L intergenic spacers. Botany 88: 886–892.
- Lopez L and Barreiro R (2012). Genetic guidelines for the conservation of the endangered polyploid *Centaurea borjae* (Asteraceae). Journal of Plant Research. doi: 10.1007/s10265-012-0497-3.
- Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Vainola R, Viard F and Wares J (2008). Evaluating signatures of glacial refugia for north Atlantic benthic marine taxa. Ecology 89: 108–122.
- Mba C and Tohme J (2005). Use of AFLP markers in surveys of plant diversity. Methods in Enzymology 395: 177–201.
- McCauley DE (1995). The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends in Ecology and Evolution 10: 198–202.
- Meudt HM and Clarke AC (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends in Plant Science 12: 106–117.
- Migliore J, Baumel A, Juin M, Diadema K, Hugot L, Verlaque R and Medail F (2011). Genetic diversity and structure of a Mediterranean endemic plant in Corsica (*Mercurialis corsica*, Euphorbiaceae). Population Ecology 53: 573–586.
- Molins A, Mayol M and Rosselló JA (2009). Phylogeographical structure in the coastal species *Senecio rodriguezii* (Asteraceae), a narrowly distributed endemic Mediterranean plant. Journal of Biogeography 36: 1372–1383.
- Moritz C (1994). Defining "Evolutionarily-Significant-Units" for conservation. Trends in Ecology and Evolution 9: 373–375.
- Nei M (1987). Molecular evolutionary genetics. New York: Columbia University Press.
- Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13: 1143–1155.
- Ouborg NJ, Piquot Y and Van Groenendael JM (1999). Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology 87: 551–568.
- Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D and Vendramin GG (2005). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Molecular Ecology 14: 689–701.
- Petit RJ, El Mousadik A and Pons O (1998). Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12: 844–855.
- Pisanu S, Filigheddu R and Farris E (2009). The conservation status of an endemic species of northern Sardinia: *Centaurea horrida* Badaro (Asteraceae). Plant Biosystems 143: 275–282.

- Pons O and Petit RJ (1995). Estimation, variance and optimal sampling of gene diversity. I. haploid locus. Theoretical and Applied Genetics 90: 462–470.
- Pons O and Petit RJ (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144: 1237–1245.
- Posada D and Crandall KA (2001). Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology and Evolution 16: 37–45.
- Reed DH and Frankham R (2003). Correlation between fitness and genetic diversity. Conservation Biology 17: 230–237.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496–2497.
- Sagarin RD, Gaines SD and Gaylord B (2006). Moving beyond assumptions to understand abundance distributions across the ranges of species. Trends in Ecology and Evolution 21: 524–530.
- Schaal BA, Gaskin JF and Caicedo AL (2003). Phylogeography, haplotype trees, and invasive plant species. Journal of Heredity 94: 197–204.
- Schaal BA, Hayworth DA, Olsen KM, Rauscher JT and Smith WA (1998). Phylogeographic studies in plants: problems and prospects. Molecular Ecology 7: 465–474.
- Slatkin M (1993). Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47: 264–279.
- Soñora X (1994). Nueva localidad de *Centaurea borjae*, Valdes-Bermejo & Rivas Goday. Lazaroa 14: 183–184.
- Su Z, Zhang M and Sanderson SC (2011). Chloroplast phylogeography of *Helianthemum songaricum* (Cistaceae) from northwestern China: implications for preservation of genetic diversity. Conservation Genetics 12: 1525–1537.
- Sun M and Ritland K (1998). Mating system of yellow starthistle (*Centaurea solstitialis*), a successful colonizer in North America. Heredity 80: 225–232.
- Taberlet P, Gielly L, Pautou G and Bouvet J (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17: 1105–1109.
- Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673–4680.
- Valdés-Bermejo E and Rivas Goday S (1978). Estudios en el género *Centaurea* L. (Compositae): *C. borjae* sp. nov. (Sect. Borjae Sect. Nov). Anales del Instituto Botánico J. Cavanilles 85: 159–164.
- Willi Y, Van Buskirk J and Hoffmann AA (2006). Limits to the adaptive potential of small populations. Annual Review of Ecology and Systematics 37: 433–458.
- Zhao Y and Gong X (2012). Genetic structure of the endangered *Leucomeris decora* (Asteraceae) in China inferred from chloroplast and nuclear DNA markers. Conservation Genetics 13: 271–281.
- Zhou GY, Yang LC, Li CL, Xua WH and Chen GC (2010). Genetic diversity in endangered *Notopterygium forbesii* Boissieu based on intraspecies sequence variation of chloroplast DNA and implications for conservation. Biochemical Systematics and Ecology 38: 911–916.

Chapter 3

- Allendorf FW and Luikart G (2012). Conservation and the Genetics of Populations. Blackwell Pub., Malden, MA.
- Angeloni FN, Ouborg J and Leimu R (2011). Meta-analysis on the association of population size and life history with inbreeding depression in plants. Biological Conservation 144: 35-43.
- Ashton PJ and Mitchell DS (1989). Aquatic Plants: Patterns and Modes of Invasion, Attributes of Invading Species and Assessment of Control Programs. Pages 111-154 in Drake JA, Mooney HA, di Castri F, Groves RH, Kruger FJ, Rejmánek M, and Williamson M, editors. Biological Invasions: A Global Perspective. John Wiley & Sons, Chichester.
- Avise JC (2004). Molecular Markers, Natural History, and Evolution. Sinauer Associates, Sunderland, Mass.
- Bandelt HJ, Forster P and Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37-48.
- Bañares AG, Blanca J, Güemes JC, Moreno, and Ortiz S, editors (2004). Atlas y Libro Rojo de la Flora Vascular Amenazadas de España. Dirección General de Conservación de la Naturaleza, Madrid.
- Bohonak AJ (2002). IBD (Isolation by distance): A program for analysis of isolation by distance. Journal of Heredity 93: 153-154.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon C, Brochmann C, and Taberlet P (2004). How to track and assess genotyping errors in population genetics studies. Molecular Ecology 13: 3261-3273.
- Bonin A, Ehrich D and Manel S (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Molecular Ecology 16: 3737-3758.
- Bonin A, Pompanon F, and Taberlet P (2005). Use of amplified fragment length polymorphism (AFLP) markers in surveys of vertebrate diversity. Pages 145-161 in Zimmer E and Roalson E, editors. Molecular Evolution: Producing the Biochemical Data, Part B. Academic Press, San Diego, CA.
- Cole CT (2003). Genetic variation in rare and common plants. Annual Review of Ecology Evolution and Systematics 34: 213-237.
- Corander J, Siren J and Arjas E (2008). Bayesian spatial modeling of genetic population structure. Computational Statistics 23: 111-129.
- Douhovnikoff V and Dodd RS (2003). Intra-clonal variation and a similarity threshold for identification of clones: application to *Salix exigua* using AFLP molecular markers. Theoretical and Applied Genetics 106: 1307-1315.
- Duminil J, Fineschi S, Hampe A, Jordano P, Salvini D, Vendramin GG and Petit RJ (2007). Can population genetic structure be predicted from life-history traits? American Naturalist 169: 662-672.
- Ellstrand NC and Elam DR (1993). Population genetic consequences of small population size: implications for plant conservation. Annual Review of Ecology, Evolution and Systematics 24: 217-242.
- Excoffier L, Laval G and Schneider S (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.
- Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 21: 479-491.
- Frankham R (2005). Genetics and extinction. Biological Conservation 126, 131-140.
- Frankham R, Briscoe DA and Ballou JD (2010). Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.
- Gitzendanner MA and Soltis PS (2000). Patterns of genetic variation in rare and widespread plant congeners. American Journal of Botany 87: 783-792.
- Hamrick JL and Godt MJW (1990). Allozyme diversity in plant species. Pages 43-63 in Brown AHD, Clegg MT, Kahler AL, and Weir BS, editors. Plant populations genetics, breeding and genetic resources. Sinauer, Sunderland.
- Hamrick JL and Godt MJW (1996). Conservation Genetics of Endemic Plant Species. Pages 281-304 in Avise JC and Hamrick JL, editors. Conservation Genetics:

Case Histories from Nature. Chapman & hall, New York.

- Hedrick PW (2001). Conservation genetics: where are we now? Trends in Ecology & Evolution 16: 629-636.
- Höglund J (2009). Evolutionary conservation genetics. Oxford University Press, Oxford.
- Honnay O, Bossuyt B, Jacquemyn H, Shimono A and Uchiyama K (2008). Can seed bank maintain the genetic variation in the above ground plant population? Oikos 117: 1-5.
- IUCN (2012). Red List Categories and Criteria. Version 3.1. IUCN, Gland, Switzerland and Cambridge, UK.
- Jensen JL, Bohonak AJ and Kelley ST (2005). Isolation by distance, web service. BMC Genetics 6: 13.
- Kawecki TJ and Ebert D (2004). Conceptual issues in local adaptation. Ecology Letters 7: 1225-1241.
- Kelchner SA (2000). The evolution of non-coding chloroplast DNA and its application in plant systematics. Annals of the Missouri Botanical Garden 87, 482-498.
- Kloda JM, Dean PDG, Maddren C, MacDonald DW and Mayes S (2008). Using principle component analysis to compare genetic diversity across polyploidy levels within land complexes: an example from British Resthrrows (*Ononis spinosa* and *Ononis repens*). Heredity 100: 253-260.
- Lande R (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. American Naturalist 142: 911-927.
- Landguth EL and Balkenhol N (2012). Relative sensitivity of neutral versus adaptive genetic data for assessing population differentiation. Conservation Genetics 13: 1421-1426.
- Leimu R and Fischer M (2008). A Meta-Analysis of Local Adaptation in Plants. PLoS ONE 3:e4010.
- Leimu R, Mutikainen P, Koricheva J and Fischer M (2006). How general are positive relationships between plant population size, fitness and genetic variation? Journal of Ecology 94: 942-952.
- Leimu R, Vergeer P, Angeloni F, and Ouborg NJ (2010). Habitat fragmentation, climate change, and inbreeding in plants. Annals of the New York Academy of Sciences 1195: 84-98.
- Lenormand T (2002). Gene flow and the limits to natural selection. Trends in Ecology & Evolution 17: 183-189.
- Levin AD (1990. The Seed Bank as a Source of Genetic Novelty in Plants. The American Naturalist 135: 563-572.
- Librado P and Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.
- Marko PB and Hart MW (2011). The complex analytical landscape of gene flow inference. Trends in Ecology & Evolution 26: 448-456.
- Maun MA (1994). Adaptations enhancing survival and establishment of seedlings on coastal dune systems. Vegetatio 111: 59-70.
- McCue KA and Holtsford TP (1998). Seed bank influences on genetic diversity in the rare annual *Clarkia springvillensis* (Onagraceae). American Journal of Botany 85: 30-36.
- Meirmans PG and Van Tienderen PH (2004). Genotype and Genodive: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4: 792-794.
- Ministerio de Medio Ambiente y Medio Rural y Marino (2011). Real Decreto 139/2011, de 4 de febrero, para el desarrollo del Listado de Especies Silvestres en Régimen de Protección Especial y del Catálogo Español de Especies Amenazadas. Boletin Oficial del Estado 46: 20912-20951.
- Montgomery DC (2008). Design and Analysis of Experiments. John Wiley & Sons, New York.
- Moritz C (1994). Defining "Evolutionarily-Significant-Units" for conservation. Trends in Ecology & Evolution 9: 373-375.
- Nunney L (2002). The Effective Size of Annual Plant Populations: The Interaction of a Seed Bank with Fluctuating Population Size in Maintaining Genetic Variation. The American Naturalist 160, 195-204.
- Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13: 1143-1155.
- Ouborg NJ, Vergeer P and Mix C (2006). The rough edges of the conservation genetics paradigm for plants. Journal of Ecology 94: 1233-1248.

BIBLIOGRAPHY

- Peakall R and Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.
- Petit RJ, El Mousadik A and Pons O (1998). Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12: 844-855.
- Pons O and Petit RJ (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144: 1237-1245.
- Posada D and Crandall KA (2001). Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology & Evolution 16: 37-45.
- Reed DH and Frankham R (2001). How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55: 1095-1103.
- Reed DH and Frankham R (2003). Correlation between fitness and genetic diversity. Conservation Biology 17: 230-237.
- Romero Buján MI (2005). Flora endémica amenazada del litoral de Galicia: una visión actual. Recursos Rurais Series Cursos 2: 1-10.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.
- Schaal BA, Gaskin JF and Caicedo AL (2003). Phylogeography, haplotype trees, and invasive plant species. Journal of Heredity 94: 197-204.
- Serrano M and Carbajal R (2011). *Omphalodes littoralis* subsp. gallaecica.in I (2011, editor. IUCN Red List of Threatened Species. Version 2011.1.
- Simmons PM, Müller K and Norton PA (2007). The relative performance of indel-coding methods in simulations. Molecular Phylogenetics and Evolution 44: 724-740.
- Sletvold N, Grindeland JM, Zu P and Agren J (2012). Strong inbreeding depression and local outbreeding depression in the rewarding orchid *Gymnadenia conopsea*. Conservation Genetics 13: 1305-1315.
- Taberlet P, Gielly L, Pautou G and Bouvet J (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17: 1105-1109.
- Thompson JD, Higgins DG, and Gibson TJ (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673-4680.
- Vos P, Hogers R, Bleeker M, Reijans M, Lee vd T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23: 4407-4414.
- Whitlock MC and McCauley DE (1999). Indirect measures of gene flow and migration: FST≠1/(4Nm+1). Heredity 82: 117-125.
- Willi Y, Van Buskirk J and Hoffmann AA (2006). Limits to the adaptive potential of small populations. Annual Review of Ecology Evolution and Systematics 37: 433-458.
- Williams GC (1996). Adaptation and Natural Selection. Princeton University Press, New Jersey.
- Wright SP (1992). Adjust P-Values for Simultaneous Inference. Biometrics 48: 1005-1013.

Chapter 4

- Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, and Singh L (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theoretical and Applied Genetics 114:359-372.
- Aleksic MA and Geburek T (2014). Quaternary population dynamics of an endemic conifer, *Picea omorika*, and their conservation implications. Conservation Genetics 15:87-107.
- Allendorf FW and Luikart G (2013). Conservation and the Genetics of Populations. Blackwell Pub., Malden, MA.
- Andersen JR and Lübberstedt T (2003). Functional markers in plants. Trends in Plant Science 8:554-560.
- Blair MW and Hurtado N (2013). EST-SSR markers from five sequenced cDNA libraries of common bean (*Phaseolus vulgaris* L.) comparing three bioinformatic algorithms. Molecular Ecology Resources 13:688-695.
- Conne L , Stutz A and Vassalli JD (2000). The 3' untranslated region of messenger RNA: A molecular "hotspot" for pathology? Nature Medicine 6:637-641.
- Cordeiro GM, Casu R, McIntyre CL, Manners JM, and Henry RJ (2000). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *Erianthus* and *Sorghum*. Plant Science 160:1115-1123.
- Cho GY, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, and Cartinhour S (2000). Diversity of microsatellites derived from genomic libraries and GenBank sequence in rice (*Oryza sativa* L.). Theoretical and Applied Genetics 100:713-722.
- Ellis JR and Burke JM (2007). EST-SSRs as a resource for population genetic analyses. Heredity 99:125-132.
- Frankham R, Ballou JD, and Briscoe DA (2004). A Primer of Conservation Genetics. Cambridge University Press, Cambridge, UK.
- Frankham R, Briscoe DA, and Ballou JD (2010). Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.
- Fraser LG, Harvey CF, Crowhurst RN, and De Silva HN (2004). EST-derived microsatellites from *Actinidia* species and their potential for mapping. Theoretical and Applied Genetics 108:1010-1016.
- Fukuoka H, Yamaguchi H, Nunome T, Negoro S, Miyatake K and Ohyama A (2010). Accumulation, functional annotation, and comparative analysis of expressed sequence tags in eggplant (S*olanum melongena* L.), the third pole of the genus *Solanum* species after tomato and potato. Gene 450:76-84.
- Gao L, Tang J, Li H and Jia J (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. Molecular Breeding 12:245-261.
- Gingeras TR (2007). Origin of phenotypes: Genes and transcripts. Genome Research 17:682-690.
- Höglund J (2009). Evolutionary conservation genetics. Oxford University Press, Oxford.
- Kantety RV, La Rota M, Matthews DE and Sorrells ME (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology 48:501-510.
- Li B, Xia Q, Lu C, Zhou Z and Xiang Z(2004). Analysis on frequency and density of microsatellites in coding sequences of several eukaryotic genomes. Genomics Proteomics & Bioinformatics 2:24-31.
- Liewlaksaneeyanawin C, Ritland C, El-Kassaby Y and Ritland K (2004). Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theoretical and Applied Genetics 109: 361-369
- Luikart G, England PR, Tallmon DA, Jordan S and Taberlet P (2003). The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics 4:981-994.
- Meglecz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N and Martin JF (2010). QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26:403-404.
- Mishra RK, Gangadhar BH, Yu JW, Kim DH and Park SW (2011). Development and characterization of EST based SSR markers in Madagascar periwinkle (*Catharanthus roseus*) and their transferability in other medicinal plants. Plant Omics 4:154-162.

BIBLIOGRAPHY

- Morgante M, Hanafey M, and Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genetics 30:194-200.
- Pashley CH (2006). EST Databases as a Source for Molecular Markers: Lessons from *Helianthus*. Journal of Heredity 97:381-388.
- Ritland K (2000). Marker-inferred relatedness as a tool for detecting heritability in nature. Molecular Ecology 9:1195-1204.
- Rozen S and Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. Pages 365-386 *in* Misener S and Krawetz SA, editors. Methods in Molecular Biology, vol. 132: Bioinformatics Methods and Protocols. Humana Press, Totowa, NJ.
- Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE and Burleigh JG (2014). From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evolutionary Biology 14:23-49.
- Rungis D, Bérubé Y, Zhang J, Ralph S, Ritland EC, Ellis EB, Douglas C, Bohlmann J and Ritland K (2004). Robust simple sequence repeat markers for spruce (*Picea* spp.) from Expressed Sequence Twaags. Theoretical and Applied Genetics 109:1283-1294.
- Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G and Powell W (2004). A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47:389-398.
- Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18:233-234.
- Simko I (2009). Development of EST-SSR Markers for the Study of Population Structure in Lettuce (*Lactuca sativa* L.). Journal of Heredity 100:256-262.
- Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M and Powell W (2003). How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology 12:1339-1348.
- Tabbasam N, Zafar Y and Mehboob-ur-Rahman (2013). Pros and cons of using genomic SSRs and EST-SSRs for resolving phylogeny of the genus Gossypium. Plant Systematics and Evolution 300:559-575.
- Tautz D and Renz M (1984). Simple sequence are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research 12:4127-4138.
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, and McCouch SR (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Research 11:1441-1452.
- Thornton CA, Wyner JP, Simmons Z, McClain C and Moxley RT (1997). Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. Nature Genetics 16:407-409.
- Tiffin P and Hahn MW (2002). Coding sequence divergence between two closely related plant species: *Arabidopsis thaliana* and *Brassica rapa* spp. *pekinensis*. Journal of Molecular Evolution 54:746-753.
- Toth G, Gaspari Z and Jurka J (2000). Microsatellites in different eukariotic genomes: surveys and analysis. Genome Research 10:967-981.
- Varshney, R. K., A. Graner, and M. E. Sorrells (2005a). Genic microsatellite markers in plants: features and applications. Trends in Biotechnology 23:48-55.
- Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, and Graner A (2005b). Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Science 168:195-202.
- Victoria FC, da Maia LC and de Oliveira AC (2011). In silico comparative analysis of SSR markers in plants. BMC Plant Biology 11:15-30.
- Wang, Z, Weber JL, Zhong G and Tanksley SD (1994). Survey of plant short tandem DNA repeats. Theoretical and Applied Genetics 88:1-6.
- Wöhrmann T, Guicking D, Khoshbakht D and Weising K (2011). Genetic variability in wild populations of *Prunus divaricata* Ledeb. in northern Iran evaluated by EST-SSR and genomic SSR marker analysis. Genetic Resources and Crop Evolution 58:1157-1167.
- Wöhrmann T and K Weising (2011). In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae.

Theoretical and Applied Genetics 123:635-647.

- Woodhead M, Russel J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M and Powell W (2005). Comparative analysis of population genetic structure in *Anthyrium distentifolium* (Ptedidophyta) using AFLPs and SSRs from anonymous and trasncribed gene regions. Molecular Ecology 14:1681-1695.
- Yu, JK, Dake TM, Singh S, Benscher D, Li W, Gill B and Sorrells ME (2004). Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805-818.
- Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, Qian H and Tang K (2006). Conservation of noncoding microsatellites in plants: Implications for gene regulation. BCM Genomics 7:323-337.

ANNEX

EXTENDED SUMMARY

TÍTULO: Genética de la conservación de plantas amenazadas en el NW de la Península Ibérica: una aproximación práctica

Genética de la conservación

La genética de la conservación es una disciplina aplicada que se beneficia del uso de herramientas moleculares y evolutivas para conservar la biodiversidad (Avise and Hamrick, 1996; Frankham *et al.*, 2010; Mills, 2006). La diversidad de los genes constituye la materia prima de las especies para evolucionar y adaptarse en un ambiente en continuo cambio. Por lo tanto, para diseñar estrategias de conservación adecuadas es imprescindible conocer el nivel y la distribución de la diversidad genética dentro y entre poblaciones (Frankham, 2005; Frankham *et al*., 2002; Hamrick and Godt, 1996). Este conocimiento es aún más importante en especies raras y/o amenazadas.

Las especies raras y/o amenazadas a menudo poseen características tales como un pequeño tamaño de población, especificidad por un hábitat y/o aislamiento, que las hacen más susceptibles a sufrir procesos de erosión genética (Ellstrand and Elam, 1993; Cole, 2003; Hamrick and Godt, 1996; Leimu *et al.*, 2006). Las plantas con pequeños tamaños poblacionales son más suscceptibles a a sufrir cuellos de botella y deriva genética (Hamrick *et al.*, 1991). Los cuellos de botella conllevan una fuerte reducción en el número de individuos que habitualmente va acompañada de una disminución de la diversidad genética (Willi *et al.*, 2006). Del mismo modo, la deriva genética resulta en la pérdida de alelos por azar (Hamrick and Godt, 1996). Varias revisiones sugieren que las plantas raras y/o amenazadas tienden a poseer niveles de diversidad genética menores que los de especies más ampliamente distribuidas (Cole, 2003; Ellstrand and Elam, 1993). Sin embargo, esta afirmación está lejos de ser universal y necesita ser examinada con mayor detalle (Gitzendanner and Soltis, 2000).

Además, unos niveles bajos de diversidad genética neutral no necesariamente correlacionan con una pérdida de de variabilidad adaptativa (Bekessy *et al.*, 2003; Landguth and Balkenhol, 2012; Reed and Frankham, 2001; Reed and Frankham, 2003).

El patrón de diversidad genética en plantas está influenciado por múltiples factores entre los cuales cabe destacar el efecto de los rasgos vitales de la especie (Hamrick *et al.*, 1991; Nybom, 2004). La forma de vida, el rango de distribución y el tipo de reproducción afectan a la diversidad genética tanto a nivel de la especie como a nivel de la población. Las especies anuales, especies que se reproducen por autogamia y/o especies con rangos de distribución reducidos tienden a poseer menor diversidad genética que las perennes, de fecundación cruzada y/o ampliamente distribuidas (Hamrick *et al.*, 1991, Nybom, 2004). Por otra parte, las plantas anuales y/o autógamas acostumbran a manifestar mayor diferenciación entre poblaciones que las que tienen fecundación cruzada o son perennes (Gitzendanner and Soltis, 2000; Hamrick and Godt, 1990; Honnay and Jacquemyn, 2007). La dispersión es otro proceso determinante de la estructura genética (Garcia *et al.*, 2007). Especies con un movimiento restringido de polen y/o semillas suelen presentar fuerte estrucura genética mientras que las plantas con una elevada tasa de dispersión tienden a presentar una distribución aleatoria de genotipos (Turner *et al.*, 1982; Wright, 1943; Wright, 1978). Finalmente, la diferenciación genética entre poblaciones puede ser consecuencia de procesos de adaptación local en lugar de deriva genética o baja dispersión.

Para conocer el nivel y estructura genéticos de las poblaciones es necesario emplear marcadores moleculares. Actualmente, hay muchos tipos de marcador molecular pero ninguno es el marcador perfecto y la elección de cuál utilizar depende de la cuestión abordada. Entre los marcadores más utilizados en genética de conservación de plantas encontramos los AFLPs (*Amplified Fragment Length Polymotphism*), los microsatélites o SSRs (*Short Sequence Repeats*) y la secuenciación

de regiones del cloroplasto (Mba and Tohme, 2005; Selkoe and Toonen, 2006; Taberlet *et al.*, 1991). Los AFLP son marcadores que cubren todo el genoma amplificando fragmentos de restricción mediante la adición de ligandos. Una de sus principales ventajas es que no requieren conocimiento previo del genoma (Allendorf and Luikart, 2013) pero son marcadores dominantes que no permiten detectar heterocigotos. Sin embargo, su naturaleza dominante se ve compensada por el alto número de loci que pueden detectar. Los microsatélites son muy utilizados en genética de poblaciones por su naturaleza co-dominante, alto polimorfismo y considerable abundancia a lo largo del genoma (Selkoe and Toonen, 2006). Sin embargo, también tienen desventajas y su desarrollo es una tarea que consume tiempo y dinero. La secuenciación de fragmentos de ADN del cloroplasto es una información muy valiosa debido a que su modo de herencia difiere del de los marcadores moleculares neutrales como AFLPs y SSRs (McCauly, 1995). El ADN del cloroplasto se hereda principalmente de forma maternal en angiospermas y, por lo tanto, solo puede ser dispersado por semillas (McCauly, 1995). Además, sus secuencias puede ser ordenadas históricamente proporcionando información sobre la historia de las poblaciones (Avise, 2004).

La información derivada de marcadores neutrales como los citados arriba es un elemento crucial en el desarrollo de iniciativas de conservación efectivas, tanto *in situ* como *ex situ*. Por un lado, los esfuerzos de conservación *ex situ* consisten típicamente en el almacenar germoplasma (principalmente semillas). Para el muestreo de germoplasma es necesario mantener una distancia mínima de muestreo entre individuos que se determina mediante un análisis espacial de la estructura genética (Vekemans and Hardy, 2004). Por otra parte, la gestión *in situ* de poblaciones silvestres suele implicar el definir unidades de manejo (MUs) (Palsboll *et al*., 2007) que se diagnostican como poblaciones que presentan diferencias en las frecuencias alélicas de ADN de orgánulos y/o loci nucleares (Avise, 1995; Moritz, 1994). Cuando

la diferenciación va más allá de la simple divergencia en las frecuencias alélicas e implica también diferencias en rasgos cuantitativos se emplea el término unidad evolutivamente significativa (ESU) (Crandall *et al*, 2000; Moritz, 1999). Es importante saber con qué tipo o unidad se está tratando ya que intercambiar individuos entre Mus puede ser recomendable pero no lo es entre ESUs.

A pesar de que los marcadores neutrales son útiles para determinar las relaciones genéticas entre individuos, el flujo de genes, la estructura de la población, y la historia demográfica (Reed and Frankham, 2001) su uso como indicadores del potencial adaptativo de una especie es, en el mejor de lo casos, escaso (Bekessy *et al.*, 2003; Reed and Frankham, 2001). Con el reciente aumento de la disponibilidad de conjuntos de datos de ADN generados por NGS (Next Generation Sequencing) y el creciente énfasis en la genómica funcional, las nuevas técnicas y enfoques de datos ahora pueden ser aplicadas a las poblaciones naturales (Allendorf *et al*., 2010; Luikart *et al*., 2003). Es en este contexto donde la genética de la conservación va un paso más allá convirtiéndose en genómica de conservación, una disciplina todavía en su infancia resulta muy prometedora (Ouborg *et al*., 2010; Primmer, 2009).

Especies objeto de estudio

La presente tesis se centra en el estudio de la diversidad y estructura genética de dos endemismos del noroeste de España: *Centaurea borjae* Valdés- Bermejo y Rivas Goday (1978) y *Omphalodes littoralis* spp. *gallaecica* M. Lainz (1971). Ambas especies están catalogadas como "en peligro " por la IUCN y el Catálogo Español de Especies Amenazadas (Serrano y Carbajal, 2011; Ministerio de Medio Ambiente y Medio Rural y Marino, 2011), y catalogadas como especies prioritarias en la Directiva de Hábitats de la UE (92/43/CEE, Anexo II). Su ocupación total se estima que es muy reducida siendo una de las principales razones a las que deben su estatus de en peligro. Además, sus hábitats son considerados como lugares de importancia comunitaria (LIC) dentro de la red Natura 2000.

Centaurea borjae se encuentra sólo en seis localidades, todas ellas acantilados situados a lo largo de <40 km de la línea costera (Valdés- Bermejo and Rivas Goday, 1978). Se estima que la ocupación total de la especie no supera 5.000 m^2 (Bañares *et al.*, 2004). *C. borjae* es una pequeña planta (<6 cm de altura), con polinización cruzada entomófila y flores hermafroditas (Valdés-Bermejo and Agudo Mata, 1983; Valdés-Bermejo and Rivas Goday 1978). Su éxito de germinación parece ser muy bajo (Gómez-Orellana Rodríguez, 2004; Pers comm. R. Retuerto; pero ver Izco *et al.,* 2003 para otras estimas) y se pueden encontrar fácilmente larvas de insectos alimentándose dentro de los frutos (Fernández Casas and Susanna, 1986). El fruto carece de vilano y posee un elaiosoma que sugiere que las hormigas podrían desempeñar un papel en la dispersión de las semillas. *C. borjae* produce rizomas que pueden extenderse hasta varios metros y dar lugar a nuevas rosetas.

A pesar de su estatus como especie prioritaria para la conservación, no hay datos de la magnitud y estructura de su diversidad genética. Sus rasgos vitales pueden conducir a hipótesis contradictorias sobre su variación genética. Por un lado, la propagación clonal junto con la baja germinación llevan a pensar que las poblaciones podrían tener baja diversidad genética. Por otro lado, como especie de fecundación cruzada podría mostrar niveles considerables la diversidad genética (Cole, 2003; Hamrick and Godt, 1996; Nybom, 2004) y, además, los poliploides suelen mantener niveles más altos de diversidad genética en poblaciones pequeñas que los diploides (Soltis and Soltis, 2000). Finalmente, la presencia de frutos sin vilano y la probable mirmecocoria pueden considerarse indicadores de una dispersión restringida de semillas (Cousens *et al*., 2008; Gómez and Espadaler, 1998) que podría resultar en la diferenciación genética significativa a pequeñas escalas espaciales.

Omphalodes littoralis. spp. *gallaecica* es un pequeño terófito con una ocupación total <100.000 m^2 y cuya presencia está restringida a cinco sistemas de dunas costeras (Romero Buján, 2005; Serrano and Carbajal, 2011). Debido a las

amenazas que enfrenta su hábitat, las poblaciones de esta planta han sufrido una disminución continua en las últimas décadas (Bañares *et al.*, 2004). *O. littoralis* spp. *gallaecica* es una planta auto-compatible y la autogamia se ha sugerido como el mecanismo más probable de la reproducción (Bañares *et al.*, 2004). El período de floración es muy corto y las flores duran menos de tres días (Romero Buján, 2005). La semillas se cree que son dispersadas por animales a adheridas al pelo del animal (Bañares *et al.*, 2004). Su tamaño de población fluctúa mucho entre años, pudiendo multiplicar o dividir por diez el número de individuos (Bañares *et al.*, 2004).

Como en *C. borjae*, a pesar del interés para la conservación de *O. littoralis* spp. *gallaecica*, nunca se ha estudiado ni su diversidad y estructura genética, ni la variación de sus características ecofisiológicas. La probable autogamia sugiere que los niveles de diversidad dentro de poblaciones podrían ser bajos (Hamrick *et al.*, 1999; Nybom, 2004). Del mismo modo, las grandes fluctuaciones de tamaño de población entre años podrían conllevar una erosión genética por cuellos de botella consecutivos (Willi *et al.*, 2006). Por último, las altas tasas de autofecundación podrían resultar en una gran diferenciación entre poblaciones (Nybom, 2004; Hamrick and Godt, 1996). Si esos altos niveles de diferenciación se mantienen en el tiempo, es posible que las poblaciones evolucionen independientemente resultando en adaptación local (Leimu and Fischer, 2008). Por lo tanto, se esperaría que *O. littoralis* spp. *gallaecica* exhiba una gran diferenciación entre poblaciones que podría conducir a la adaptación local de éstas.

Objetivos

Objetivos generales:

• El objetivo principal de esta tesis es aplicar marcadores moleculares al estudio de la diversidad y estructura de población de plantas raras y/o amenazadas. Los resultados

se interpretan desde un punto de vista aplicado y se proponen medidas de conservación específicas.

Objetivos específicos:

• **Capítulo 1:** se utilizaron fenotipos AFLP para investigar la variación genética y la estructura poblacional de *Centaurea borjae*. La información derivada de los AFLPs se utilizó para (1) inferir la contribución de la reproducción clonal, (2) determinar si las poblaciones muestran signos de empobrecimiento genético; (3) inferir la distancia mínima entre plantas para la recolección de semillas para bancos de germoplasma; (4) determinar si las poblaciones se diferencian significativamente entre sí, y de ser así, si es posible delimitar unidades de gestión

• **Capítulo 2:** se estudia la estructura genética de *Centaurea borjae* a lo largo de su área de distribución y los procesos históricos detrás de ésta empleando secuencias de la región no codificante *trnT-F* del cloroplasto (cpDNA) (Taberlet *et al.*, 1991). Específicamente, en este capítulo se abordan los siguientes objetivos: (1) estimar la diversidad genética de *C. borjae* utilizando secuencias cpDNA, (2) investigar su pasado demográfico, (3) evaluar su estructura de la población, (4) identificar las poblaciones de mayor interés para la conservación y comparar el patrón obtenido con los resultados de los AFLP del capítulo 1.

• **Capítulo 3:** En este capítulo se lleva a cabo estudios moleculares y fenotípicos exhaustivos de las cinco poblaciones existentes de *Omphalodes littoralis* spp. *gallaecica*. Se utilizaron secuencias de la región *trnT-F* del cloroplasto y genotipos AFLP para determinar (1) si *O. littoralis* spp. *gallaecica* está empobrecida genéticamente como podrían indicar sus rasgos vitales; (2) comprobar si sus poblaciones están significativamente diferenciadas entre sí; (3) dado que *O. littoralis* spp. *gallaecica* es un terófito, determinar si hay diferencias significativas entre años consecutivos en su estructura genética. Además, se realizaron experimentos de

trasplante recíproco para investigar el componente adaptativo de varios rasgos cuantitativos relacionados con la *fitness*. Las informaciones molecular y fenotípica se combinaron para proponer directrices específicas para la conservación de esta especie en peligro de extinción.

• **Capítulo 4**: Este capítulo explora una aproximación todavía poco explotada, pero prometedora, de los EST-SSRs: el desarrollo de marcadores a partir de secuencias EST disponibles en bases de datos de públicas para utilizarlos en estudios de genética evolutiva y de conservación de plantas no-modelo, con énfasis en especies amenazadas. Se buscaron SSR en todos los géneros de planta de la Lista Roja de Plantas de la Unión Internacional para la Conservación de la Naturaleza y los Recursos Naturales (UICN) con secuencias EST disponibles en la base de datos GenBank EST (dbEST). Dado que la mayoría de estos géneros de plantas no incluyen organismos modelo, no hay genomas de referencia anotados disponibles, lo que dificulta la localización de los EST-SSRs dentro del genoma. Para minimizar este obstáculo, también se analizaron las secuencias EST de dos géneros modelo que sirvieron de especies sustitutas/representativas: *Arabidopsis* se seleccionó como control de eudicotiledóneas y *Oryza* como guía para monocotiledóneas. Por último, se testó la amplificación, polimorfismo y transferibilidad entre congéneres de doce loci SSR para cada genéro usando dos especies de cada género: *Trifolium fragiferum, Trifolium saxatile, Centaurea valesiaca* y *Centaurea borjae*.

Resultados y discusión

- *Centaurea borjae*

Una de las principales preocupaciones para la preservación a largo plazo de *Centaurea borjae* derivaba de la sospecha de que las poblaciones podrían estar formadas solo por unos pocos genetos con numerosos rametos. Los resultados mostraron que existen clones en todas las poblaciones pero su presencia no era tan

alta como se especulaba. Además, su abundancia variaba entre localidades, las localidades más al norte mostraron mayor abundancia de clones que las centrales y las de más al sur. Estas diferencias en diversidad clonal entre poblaciones parecen ser algo frecuente en plantas (ver Arnaud-Haond *et al.*, 2007 y sus referencias bibliográficas) y estudios anteriores han encontrado que la clonalidad aumenta con la edad de la población o la latitud (Silvertown, 2008). Sin embargo, la única diferencia consistente entre nuestros dos grupos de poblaciones es el sustrato geológico: serpentinitas en los 3 sitios más septentrionales; gneises, anfibolitas y granitos en los otros 3. Los suelos de serpentina se caracterizan por niveles altos de metales tóxicos que pueden afectar el crecimiento de la planta, lo que sugirie que las condiciones creadas por el suelo de serpentina podrían, al menos en parte, favorecer la propagación clonal en *C. borjae*. En este sentido, estudios experimentales anteriores con otras especies han demostrado que las plantas clonales mejorar los efectos estresantes de suelos a través de la integración fisiológica de sus rametos (Roiloa and Retuerto, 2006).

Las estimas de diversidad derivadas de los análisis AFLPs mostraron que *Centaurea borjae* no está genéticamente empobrecida y posee niveles de diversidad genética similares a otras especies con rasgo vitales similares (i.e. plantas perennes y/o con fecundación cruzada) (Nybom, 2004). Los valores encontrados caen dentro del rango de estimas obtenidas con marcadores dominantes en otros miembros de género C*entaurea*. Sin embargo, las estimas de diversidad obtenidas con cpDNA mostraron evidencias de empobrecimiento genético cuando se comparan con otras plantas raras.

Los análisis de estructura de población apuntaron a diferencias genéticas significativas entre poblaciones con ambos marcadores, lo que sería consistente con un escenario de flujo genético reducido. Ese flujo genético reducido entre poblaciones parece consistente con la capacidad de dispersión limitada que sugieren ciertas

características de *C. borjae.* La dispersión del polen mediada por animales puede ser limitada en función del comportamiento del animal dispersor y/o la frecuencia y distribución de los recursos florales (Ghazoul, 2005). Del mismo modo, la ausencia de vilano y la probable mimecocoria de *C. borjae* sugieren que la dispersión de semillas podría limitarse a distancias cortas (Cousens *et al.,* 2008; Gómez and Espadaler, 1998). La idea de flujo genético reducido se vio reforzada por los análisis AFLP de estructura genética espacial a pequeña escala que mostraron que plantas más próximas entre sí también estaban genéticamente más emparentadas. Por tanto, nuestros resultados mostraron una fuerte estructura espacial a pequeña escala típica de especies con baja dispersión, reproducción clonal, y/o de baja densidad poblacional (Vekemans and Hardy, 2004). Como el alcance de esa estructura a pequeña escala varía entre localidades (35-40 m a 80 m), se recomienda que las muestras para bancos de germoplasma estén separadas al menos 80 m.

La disposición actual de haplotipos de cpDNA puede ser una consecuencia de la historia demográfica de la planta. Basándonos en predicciones de la teoría de coalescencia (Posada and Crandall, 2001), los haplotipos H1 y H2 serían ancestrales y su co-ocurrencia en las localidades PC y OB sugiere que esta zona es un sitio de gran persistencia de la especie. La misma conclusión se alcanza con el análisis de la distribución espacial de la diversidad genética y haplotipos privados ya que las poblaciones más antiguas acostumbran ser más diversa y contenien haplotipos privados (Maggs *et al.*, 2008.), dos condiciones que se encuentran en PC y OB. En este escenario, los restantes sitios habrían derivado de la posterior colonización desde la zona central y su diversidad genética más baja sería producto de un efecto fundador.

Finalmente, se designaron 5 unidades de manejo en base a diferencias en las frecuencias de los loci AFLP y las frecuencias haplotípicas del cpDNA (LI, VH, OB-OBB, PC, and PR). Designar MUs en base a los de AFLPs o cpDNA por separado podría llevar a errores ya que con los AFLPs PC se consideraría parte de la MU OB-OBB mientras

que con cpDNA PR tampoco sería considerada una MU independiente. Esto pone de manifiesto la necesidad de combinar marcadores con distinto modo de herencia para formular medidas de conservación más precisas.

- *Omphalodes littoralis* spp. *gallaecica*

Los análisis genéticos de las poblaciones de *Omphalodes littoralis* spp. *gallaecica* revelaron niveles de diversidad muy bajos o nulos, en concordancia con sus rasgos de vida (especie anual que se reproduce por autogamia; Nybom, 2004). Así mismo, la estructura de población puso de manifiesto la ausencia de flujo genético entre poblaciones. El hecho de que todas las poblaciones poseyeran bandas AFLP privadas es indicativo de un fuerte aislamiento mantenido en el tiempo. Esto último fue confirmado por los resultados de las secuencias de cpDNA donde la casi todas las poblaciones mostraron una composición diferente y la mayoría de los haplotipos eran privados. De nuevo, esta enorme diferenciación fue consistente con los rasgos de vida de este pequeño terófito (Nybom, 2004). De acuerdo con la teoría coalescente, el haplotipo H1 podría ser considerado como ancestral y su aparición en tres poblaciones no adyacentes, sugiere que los diversos grupos locales podrían haber estado conectados en un pasado distante.

La extremadamente baja diversidad genética de las poblaciones, junto con su enorme diferenciación genética, sugiere que esta pequeña planta podría estar reflejando los efectos de la deriva genética. Este último podría estar agravado por cuellos de botella recurrentes como consecuencia de las fuertes fluctuaciones de tamaño poblacional típicas de este endemismo. La extremadamente baja diversidad de las poblaciones de *O. littoralis* spp. *gallaecica* es motivo de preocupación ya que pueden tener menor capacidad de respuestas frente a cambios ambientales y/o condiciones de estrés (Frankham, 2005). Las poblaciones pequeñas que caen por debajo de cierto tamaño efectivo pueden entrar en un "vórtice de extinción" donde la dinámica reproductiva favorecen la endogamia conduciendo a una disminución en

la reproducción, un aumento de la mortalidad y una reducción en el tamaño de las poblaciones más pequeñas. Por otra parte, la extrema fragmentación de la especie y el aislamiento entre sus poblaciones sugieren que es improbable un rescate genético de una población por otras.

Mientras que las grandes fluctuaciones de tamaño poblacional podrían comprometer la diversidad genética de *O. littoralis* spp. *gallaecica*, otros atributos de su ciclo de vida pueden actuar en dirección opuesta . Algunos taxa anuales tienen un gran banco de semillas viables de las que se pueden extraer individuos en el futuro que amortigüen la pérdida genética (McCue and Holtsford, 1998; Nunney, 2002). Sin embargo, este no parece ser el caso en *Omphalodes littoralis* spp. *gallaecica* ya que nuestros datos revelaron una composición genética constante en generaciones consecutivas. Por lo tanto, la incapacidad del banco de semillas para actuar como depósito de diversidad genética añade más preocupación sobre la persistencia a largo plazo de esta especie.

Los análisis de rasgos cuantitativos mostraron que poblaciones separadas por pocos kilómetros eran fenotípicamente diferentes. Si bien esta variación podría ser simplemente una respuesta fenotípica a sutiles cambios en el entorno local de cada lugar, nuestros experimentos de trasplantes recíprocos indican que en realidad poseen un componente genético. Sin embargo, a diferencia de lo que cabría esperar en un escenario de adaptación local, las plantas de un mismo sitio (DN) solían superar a las de los demás, independientemente de la ubicación del trasplante. Inicialmente, no hay una explicación clara para el mejor funcionamiento de las plantas de DN. La única diferencia evidente entre DN y las otras poblaciones es que DN muestra los niveles más altos de diversidad genética. Por lo tanto, parece tentador especular que el mayor rendimiento de sus individuos podría estar relacionado con la mayor variación genética neutral detectada por los marcadores moleculares.

Desde una perspectiva de conservación, al combinar los datos genéticos y fenotípicos, se recomienda establecer cinco ESUs. Es importante resaltar que la existencia de estas diferencias indican que los diversos grupos locales no son intercambiables entre si y pueden tener un potencial diferente para evolucionar. En este sentido, las prácticas de gestión que impliquen un desplazamiento de individuos entre sitios no parecen recomendables visto el fuerte aislamiento genético entre las poblaciones de este terófito en peligro de extinción (Sletvold *et al*., 2012).

- *EST-SSR para géneros de plantas amenazadas de la IUCN.*

Las aproximaciones computacionales permiten desarrollar, rápida y económicamente, marcadores moleculares a partir de recursos genómicos disponibles al público. En este contexto, el desarrollo de SSR derivados de secuencias EST surgen como una excelente alternativa a las técnicas clásicas de desarrollo de SSR anónimos (Ellis and Burke, 2007). El análisis de los genomas de control mostró que los trímeros y los dímeros constituyen más de 85% de los SSR encontrados, siendo trinucleótidos >60%. Estos resultados fueron consistentes con lo esperado para plantas superiores (Kantety *et al.*, 2002; Varshney *et al.*, 2005; Victoria *et al.*, 2011). Así mismo, AG fue el motivo más abundante en dinucleótidos mientras que AT mostró frecuencias bajas (Kantety *et al.*, 2002; Morgante *et al.*, 2002; Temnykh *et al.*, 2001; Victoria *et al.*, 2011). En lo que respecta a los trinucleótidos, los motivos ricos en GC fueron los más abundantes en *Oryza,* en concordancia con lo esperado en monocotiledóneas (Gao *et al.*, 2003; Temnykh *et al.*, 2001; Kantety *et al.*, 2002; Victoria *et al.*, 2011). En contraposición, los motivos ricos en GC eran escasos en *Arabidopsis*, lo que de nuevo coincide con resultados publicados en otros trabajos (Victoria *et al.*, 2011). El análisis de distribución a lo largo del genoma mostró que los EST-SSRs se localizan principalmente en regiones codificantes del genoma (CDSs), lo cual es consistente con el hecho de que estos marcadores están asociados con la porción que se transcribe. Sin embargo, la frecuencia de los distintos tipos de

repetición varía ampliamente a lo largo de las distintas regiones genómicas. Dímeros, tetrámeros y pentámeros se asociaron principalmente con UTRs y otras regiones no codificantes, mientras que trímeros y hexámeros se localizaron mayoritariamente en CDSs. Dado que la frecuencia y distribución de las diferentes repeticiones SSR y sus motivos son función de la dinámica y de la historia de la evolución del genoma, el predominio de repeticiones de triméricas en los ESTs se atribuye a la selección en contra de mutaciones que alteren el marco de lectura (Morgante *et al.*, 2002). La elevada frecuencia de dímeros en UTRs y la prevalencia de trímeros en CDS se han visto anteriormente en otros estudios de plantas (Gao *et al.*, 2003; Wang *et al.*, 1994).

El análisis de frecuencias de los diferentes tipos de repeticiones en los géneros de la UICN fue muy consistente con los resultados derivados de los genomas control de *Oryza* y *Arabidopsis*. Trímeros y dímeros representaron más del 60 % de los EST-SSRs, mientras que tetrámeros, pentámeros y hexámeros mostraron frecuencias más bajas. Sin embargo, la frecuencia de los diferentes tipos de repeticiones de nucleótidos divergió entre los grupos taxonómicos estudiados. Los resultados de angiospermas fueron consistentes con los obtenidos en los genomas control y con resultados anteriores en plantas con flores donde los trímeros eran los motivos más abundantes seguidos de dímeros (Victoria et al., 2011). Así mismo, el grupo más común de motivos era AG, como se ha visto en otras angiospermas (Kantety *et al.*, 2002; Morgante *et al.*, 2002; Temnykh *et al.*, 2001; Victoria *et al.*, 2011). El patrón de los motivos triméricos fue el mismo que para *Oryza* y *Arabidopsis*, corroborando la presencia de motivos ricos en GC en monocotiledóneas y el grupo AAG en las restantes angioespermas (Gao *et al.*, 2003; Kantety *et al.*, 2002; Temnykh *et al.*, 2001; Victoria *et al.*, 2011). Las diferencias de frecuencia de los diferentes tipos de SSR entre grupos taxonómicos es función de la dinámica y la historia evolutiva del genoma (Morgante *et al.*, 2002). De acuerdo con estudios previos, el grupo Acrogymnospermae reveló una alta proporción de hexámeros en comparación con

gimnospermas y el motivo más común fue de TA, que era muy escaso en angiosperma (Pinosio *et al.*, 2014; Victoria *et al.*, 2011). Los cuatro grupos que representan a las plantas no vasculares están representados por pocos géneros en nuestros análisis y no es posible hacer generalizaciones.

La tasa de éxito de amplificación fueron similares a las de algunos estudios anteriores con EST- SSR (Cordeiro *et al.*, 2000; Rungis *et al.*, 2004). Debido a la asociación de los EST con regiones conservadas del genoma, los EST-SSRs suelen mostrar menos polimórfismo que los SSRs clásicos (Ellis and Burke, 2007; Russell *et al.*, 2004; Varshney *et al.*, 2005). Sin embargo, esta premisa no es necesariamente cierta (Fraser *et al.*, 2004; Pashley, 2006) y, en nuestro estudio, el nivel de polimorfismo varió desde 25 hasta 28,57% dentro de las especies y de 75 a 87,77% entre especies. Dado que sólo se ensayaron ocho individuos de cada género, estos niveles de polimorfismo podrían estar subestimados y el polimorfismo real de nuestros EST-SSR podría ser mayor. Una de las mayores ventajas de los EST-SSRs es su alta tasa de transferibilidad entre especies (Aggarwal *et al*., 2007; Pashley, 2006; Wöhrmann and Weising, 2011) de un mismo género o, incluso, especies de diferentes géneros (Varshney *et al.*, 2005). Los resultados obtenidos en el presente estudio son congruentes con la premisa de alta transferibilidad en EST-SSRs ya que todos los cebadores que amplificaron con éxito en una especie también lo hicieron en su congénere.

En resumen, este trabajo pone de manifiesto el gran potencial del uso de secuencias EST disponibles en bases de datos públicas como fuente de SSR para plantas amenazadas. Se detectaron SSR con cebadores en 222 géneros de plantas. Teniendo en cuenta su elevada transferibilidad, el número de especies que se podrían favorecer de estos marcadores podría ser considerable. Además, como el desarrollo de marcadores es uno de los pasos donde se invierte más tiempo en los estudios de poblaciones, parece razonable sugerir que las bases de datos de EST son una valiosa

alternativa para el desarrollo de SSR ya que una vez que se accede a la base de datos de EST, solo se necesitan un par de días para tener una batería de SSR con cebadores listos para probar sin ningún coste.

AGRADECEMENTOS/ AGRADECIMIENTOS/ THANKS TO…

A mi director,

Por enseñarme, no solo ciencia, sino también literatura, cine e historia.

O meu pai,

Por inculcar en min un profundo respeto e admiración pola natureza; amo o que fago, son o que fago gracias a ti.

A mi madre,

Por la paciencia, apoyo y consuelo que solo ella puede dar.

A mi tita Geni y a mi Tito,

Por las aventuras en Londes, y en Amsterdan. Por ser mis padres adoptivos.

A mi primo Antón,

Porque él entiende que los peces son seres vivos y los humanos también, porque es mi primo favorito.

A Andrea,

Porque las charlas de política contigo son más divertidas.

A miña avoa,

Polos coellos de indias, os pitiños de cores, as culleres desaparecidas na leira.

A mis abuelos,

Por cuidar siempre de mi con todo vuestro corazón.

A Nana,

Por dejarme ejercer de hermana mayor, por ser la otra mitad en nuestro YingYang.

A Juanjo, Marian y Sara,

Por las comidas de los sábados en la Apillada.

A Fer,

For letting me be your "perfect woman... the Goddess. Goldie."

Os Moreno Leira,

Por facerme un oco na súa familia.

A mis compas del labo, las que ya se fueron, las que aún están y las que volverán,

Por los cafés y pitillos, por los viajes en coche compartidos, por los buceos en islas paradisíacas, por las horas puliendo "conejitos y corazones", por tantas cosas que aquí no caben, pero sobre todo por ser más que compañeras, por ser mis amigas.

A Xavi,

Polas palmeiras de chocolate.

A Sergio,

Por dejarme colonizar su despacho.

A Tania y Nati,

Por compartir conmigo las horas y horas de charla juntas mejorando el mundo. A Tania, por llevar a mi lado tantos años. Por estar siempre dispuesta a darlo todo en la pista de baile conmigo, por los veranos en Sta. Cruz, y aún más importante, por estar siempre ahí cuando necesito una amiga. A Nati, por seguir compartiendo conmigo esa enorme sonrisa que te caracteriza desde que te conozco.

A Lúa,

Por las cartas interminables, los disfraces, las horas al teléfono, porque aunque haga meses que no nos veamos ni hablamos es como si el tiempo nunca pasara.

A Aran,

Por ver siempre el lado bueno de la vida, tu alegría se contagia.

A los niños,

Por acogerme desde el primer día como si fuera del "barrio" y convertirme en Lubi, por llenar mi vida con partidas de LOL, la cascada, las invenciones culinarias, las ampliaciones de pantalla, las macacadas, los cacareos, etc. por eso y mucho más esta gatiña siempre tendrá un ronroneo para vosotros.

A mis compañeros de carrera,

A Mayte, Aida, Peib, Rober, Miguel, Rosa, y todos los que hicieron de los años de carrera uno de los mejores momentos de mi vida. En especial al primate albino,

Copito, con el que tantas, tantas, horas pasé, por los momentos teniente, la música ochentera, las anécdotas curiosas sin sentido.

Last but not least, To Eva, Nora and Flo,

For making the word "bügeln", my favorite German word! And because you are one of the main reasons that I want to come back.

Appropriate management of plants of conservation concern requires reliable estimates of the magnitude and spatial distribution of genetic diversity as these species often combine features that make them potentially susceptible to genetic erosion. In this regard, the present thesis focuses on applying genetic markers to the conservation of rare and threatened plants. In the first two chapters, genetic diversity and population structure of the clonal endemism Centaurea borjae is assessed using AFLPs and cpDNA sequences. C. borjae displayed intermediate-low genetic diversity compared to other plants with similar life-history traits. Gene flow seem to be restricted as populations separated by few hundred meters showed significant differentiation. Clonal frequency was lower than anticipated and might be related to soil type. Five Management Units were designated for conservation purposes and sampling for ex situ preservation should focus on individuals separated >80 m. In the third chapter, the neutral and quantitative diversity of the endangered therophyte Omphalodes littoralis spp. gallaecica is investigated. The five extant populations displayed minimal to none neutral genetic diversity and a lack of gene flow between them. Reciprocal transplant experiments showed among-population differentiation in several quantitative traits but the pattern of differences did not fit the expectations of local adaptation. Instead, it seemed to be caused by genetic drift. Based on the genetic and phenotypic results, each population should be designated as an independent Evolutionary Significant Unit for conservation purposes.The last chapter focuses on developing SSRs markers for threatened plants using EST sequences available in public databases. 257 genera were analyzed and 86% of them were successfully mined. As most of these genera lack an annotated genome, Arabidopsis and Oryza were used as controls for genome distribution analyses. Dimers and trimers were prevalent types of repeat. Control genomes revealed that trimmers were mostly located in coding regions while dimers were largely associated to untranslated regions. Finally, empirical trials showed that EST-SSRs had high amplification success and were 100% transferable between species in two tested genera.

Appropriate management of plants of conservation concern requires reliable estimates of the magnitude and spatial distribution of genetic diversity as these species often combine features that make them potentially susceptible to genetic erosion. In this regard, the present thesis focuses on applying genetic markers to the conservation of rare and threatened plants. In the first two chapters, genetic diversity and population structure of the clonal endemism Centaurea borjae is assessed using AFLPs and cpDNA sequences. C. borjae displayed intermediate-low genetic diversity compared to other plants with similar life-history traits. Gene flow seem to be restricted as populations separated by few hundred meters showed significant differentiation. Clonal frequency was lower than anticipated and might be related to soil type. Five Management Units were designated for conservation purposes and sampling for ex situ preservation should focus on individuals separated >80 m. In the third chapter, the neutral and quantitative diversity of the endangered therophyte Omphalodes littoralis spp. gallaecica is investigated. The five extant populations displayed minimal to none neutral genetic diversity and a lack of gene flow between them. Reciprocal transplant experiments showed among-population differentiation in several quantitative traits but the pattern of differences did not fit the expectations of local adaptation. Instead, it seemed to be caused by genetic drift. Based on the genetic and phenotypic results, each population should be designated as an independent Evolutionary Significant Unit for conservation purposes.The last chapter focuses on developing SSRs markers for threatened plants using EST sequences available in public databases. 257 genera were analyzed and 86% of them were successfully mined. As most of these genera lack an annotated genome, Arabidopsis and Oryza were used as controls for genome distribution analyses. Dimers and trimers were prevalent types of repeat. Control genomes revealed that trimmers were mostly located in coding regions while dimers were largely associated to untranslated regions. Finally, empirical trials showed that EST-SSRs had high amplification success and were 100% transferable between species in two tested genera.