UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

SPECIATION IN SPATIALLY STRUCTURED POPULATIONS: IDENTIFYING GENES RESPONSIBLE FOR LOCAL ADAPTATION

Vera Lúcia Martins Nunes

DOUTORAMENTO EM BIOLOGIA

(BIOLOGIA EVOLUTIVA)

2011

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

SPECIATION IN SPATIALLY STRUCTURED POPULATIONS: IDENTIFYING GENES RESPONSIBLE FOR LOCAL ADAPTATION

Vera Lúcia Martins Nunes

Tese orientada por:

Professor Doutor Octávio S. Paulo

Professor Doutor Mark A. Beaumont

Professor Doutor Roger K. Butlin

DOUTORAMENTO EM BIOLOGIA

(BIOLOGIA EVOLUTIVA)

2011

This study was supported by Fundação para a Ciência e a Tecnologia (FCT) through a PhD scholarship (SFRH/BD/21306/2005) attributed to Vera L. Nunes and two research projects: POCI/BIA/59288/2004 -"Process of speciation: accessing the genic view" and POCTI/BSE/47999/2002 -"Coalescent methods applied to populations' analyses with microsatellites".

NOTA PRÉVIA

Para a elaboração da presente dissertação, e nos termos do Nº 1 do Artigo 40 do Regulamento de Estudos Pós-Graduados da Universidade de Lisboa, publicado no *Diário da República* nº 209 - II Série, de 30 de Outubro de 2006, foram usados integralmente artigos científicos publicados ou a submeter para publicação em revistas internacionais indexadas. Tendo os trabalhos referidos sido efectuados em colaboração, a autora da dissertação esclarece que participou integralmente na concepção e execução do trabalho experimental, na análise e interpretação dos resultados e na redacção de todos os manuscritos.

À memória da minha mãe

To the memory of my mother

TABLE OF CONTENTS

NOTE:

The varying format of some chapters in this thesis reflects the specific requirements of the scientific publications to which the presented manuscripts were submitted.

ACKNOWLEDGEMENTS | AGRADECIMENTOS

At the end of this long project, I would like to express my sincere gratitude to all those who have contributed to the fulfillment of this dissertation:

No final deste longo projecto, gostaria de expressar o meu sincero agradecimento a todos aqueles que contribuiram para a realização desta dissertação:

Ao Professor Octávio S. Paulo, por me ter proposto este projecto de doutoramento aliciante e me ter contagiado com o seu entusiasmo pela Biologia Evolutiva e pelos lagartos. Agradeço ainda a sua confiança nas minhas capacidades, as oportunidades de formação que me proporcionou ao envolver-me em vários projectos de investigação e a sua compreensão e o encorajamento perante os muitos desafios colocados por este projecto.

To Professor Roger K. Butlin and to Professor Mark A. Beaumont, for kindly accepted to co-supervise this thesis, for believing in the potential of this research project and for contributing with their expertise to the success of this project outcome. Thank you so much for your encouragement throughout these years and for your patience to attend all my doubts and to comment the many drafts of my manuscripts.

A todos os colegas com quem partilhei o laboratório, pela companhia, pela troca de experiências e pela entreajuda na resolução dos problemas recorrentes nas rotinas laboratoriais.

À Andreia Miraldo e à Inês Simões, por me transmitirem os seus conhecimentos nos meus primeiros meses no laboratório.

Um agradecimento muito especial a todos os que pertencem ou pertenceram ao grupo CoBiG^2, pelas amizade e boa disposição, pela partilha de conhecimentos e pelo encorajamento. Um grande obrigado aos "cobigos" que me auxiliaram na revisão dos

manuscritos e no domínio de ferramentas informáticas: Francisco Pina-Martins, Sofia Seabra, Eduardo Marabuto, Joana Costa, Diogo Silva, Sara Silva e Renata Martins.

Ao Pedro Moreira, à Paula Simões e à Dora Batista, por me proporcionarem a oportunidade de participar activamente nos seus projectos de investigação em paralelo com o meu projecto de doutoramento, permitindo alargar a minha experiência a outras técnicas e modelos de estudo no âmbito da Biologia Evolutiva.

À família e amigos, que acompanharam este processo muitas vezes à distância. A todos agradeço o apoio e a "torcida" para que este projecto chegasse a bom porto. Agradeço em especial ao meu pai e à minha irmã, por representarem um modelo de coragem e persistência na minha vida, fundamental para superar as provas mais duras deste projecto. Agradeço ao Hugo, que me acompanhou de perto em todos os altos e baixos desta caminhada. Agradeço ainda ao meu pequeno sobrinho, pela alegria e brincadeiras que preencheram as minhas visitas à "terrinha", e ao Quim e à Zé, sempre incansáveis, transportando "mimos" culinários entre Alcobaça e Lisboa.

RESUMO

A determinação da base genética de caracteres adaptativos em populações naturais é fundamental para melhor compreender a evolução da divergência adaptativa entre populações em ambientes heterogéneos e a forma como estas podem evoluir para formar novas espécies. O presente trabalho teve como principal objectivo a identificação de genes ou regiões genómicas envolvidas na adaptação local em populações espacialmente estruturadas, mas na ausência de barreiras físicas evidentes ao fluxo genético entre elas. O lagarto ocelado (ou sardão), *Lacerta lepida*, foi o modelo escolhido para este estudo, tendo sido analisadas populações ao longo de um gradiente ambiental na Península Ibérica, que é sobretudo condicionado pela variação climática. Duas subespécies de sardão são actualmente reconhecidas nos extremos opostos do gradiente ambiental, tendo por base a existência de diferenças morfológicas significativas, que sugerem a sua adaptação às condições ambientais locais. A subespécie *L. l. iberica* encontra-se restringida ao Noroeste da Península Ibérica, enquanto que a subespécie *L. l. nevadensis* substitui a subespécie nominal no Sudeste da área de distribuição da espécie.

Como primeira abordagem para a detecção de regiões do genoma sob acção da selecção, foi realizado um *genome scan* com AFLPs. Esta estratégia permite gerar centenas de marcadores genéticos em qualquer organismo, distribuídos pelo genoma, sem que para tal seja necessário ter conhecimento genético prévio da espécie, sendo por isso muito útil para espécies não-modelo, como o sardão, com poucos recursos genómicos disponíveis. A identificação de *loci* candidatos foi feita mediante a detecção de outliers, isto é, de marcadores de AFLP com níveis de diferenciação entre populações anormalmente elevados (selecção direccional) ou reduzidos (selecção balanceada) de acordo com o expectável num cenário de neutralidade. Foram utilizados dois métodos de detecção, um método frequencista e um método Bayesiano, e embora ambos tenham detectado um proporção semelhante de outliers (3-4%), apenas alguns dos outlieres foram detectados por ambos, denotando diferenças na sensibilidade dos dois métodos. Vários dos AFLPs detectados como outliers foram também associados com a variação na temperatura, na insolação ou na precipitação registadas ao longo do gradiente, sugerindo que estas variáveis poderão ser importantes forças selectivas ao nível da adaptação local do sardão.

Devido à susceptibilidade dos métodos de detecção de outliers aos erros de tipo I (falsos-positivos), que embora podendo ser controlados, dificilmente poderão ser totalmente eliminados, os outliers deverão ser tratados como *loci* candidatos, potencialmente influenciados pela selecção, que deverão ser posteriormente validados por outros meios. Como os AFLPs são marcadores gerados a partir da fragmentação do genoma pela acção de enzimas de restrição, tratam-se de marcadores com localização desconhecida no genoma e a sequência de DNA que constitui cada fragmento permanece completamente anónima durante todo o processo de genotipagem, sendo estes distinguidos apenas pelas diferenças de tamanho e genotipados como marcadores dominantes. Assim sendo, é muito importante que depois de indentificar os AFLPs com comportamento outlier, estes sejam investigados por forma a indentificar as sequências que os constituem e determinar a que genes poderão pertencer e quais as suas possíveis funções.

O isolamento de AFLPs com tamanho específico de entre dezenas de outros fragmentos com tamanho semelhante é tecnicamente exigente. Sete outliers foram isolados, clonados e sequenciados com sucesso, mas nenhum deles parece fazer parte de uma região codificante, sendo o polimorfismo de tamanho dos fragmentos explicado pela presença de indels ou elementos repetitivos (microssatélites). Para cada outlier sequenciado foram desenhados *primers* internos de forma a converter estes *loci* em marcadores codominantes e poder amplificá-los a partir do genoma não digerido. Devido ao reduzido tamanho dos fragmentos, apenas para três dos sete outliers sequenciados (*mk75*, *mk209* e *mk245*) foi possível desenvolver *primers* capazes de amplificar tanto os alelos dominantes como os alelos recessivos. Para o *locus mk75*, um outlier associado à variação na precipitação, foram detectados um haplótipo dominante conservado, com uma deleção de nove pares de bases, e oito haplótipos recessivos. A frequência do alelo dominante é superior em *L. l. iberica* enquanto que em *L. l. nevadensis* ele se encontra ausente. O *locus mk209*, também associado com a precipitação, apresentou dois haplótipos dominantes, caracterizados pela inserção de quatro bases (TGGA), e sete haplótipos recessivos. Todas os indivíduos de *L. l. nevadensis* sequenciados para o *locus mk209* apresentaram apenas haplótipos dominantes, enquanto que estes estão ausentes em todas as amostras sequenciadas para as restantes subespécies. Relativamente ao *locus mk245*, detectado em forte associação com a variação nas temperaturas máximas ao longo da Península Ibérica, a sequenciação revelou apenas um haplótipo dominante, contendo

um microssatélite com seis repetições de GTT, e oito haplótipos recessivos com três a cinco repetições de GTT. O haplótipo dominante não foi encontrado nos indivíduos de *L. l. iberica* nem de *L. l. nevadensis*, sendo que ambas as subespécies apresentaram apenas sequências com três repetições de GTT, embora estas se encontrem em extremos opostos do gradiente de temperatura. Os outliers *mk75*, *mk209* e *mk245* foram ainda amplificados e sequenciados com sucesso em espécies próximas (*Lacerta tangitana*, *L. pater*, *L. schreibei*, *L. agilis* e *Iberolacerta monticola*), evidenciando que apesar da variação de tamanho nos elementos repetitivos, as zonas que os flanqueiam permanecem bastante conservadas entre espécies. Tratando-se de regiões não codificantes, os outliers sequenciados para o lagarto ocelado poderão actuar como elementos reguladores da actividade de alguns genes ou poderão estar em desequilíbrio de *linkage* com outros genes que serão o verdadeiro alvo de selecção. Em qualquer dos casos, serão necessários mais recursos genómicos para compreender o papel destes outliers na evolução do lagarto ocelado.

A análise de genes candidatos, com efeitos conhecidos nas características fenotípicas de outras espécies, pode constituir uma alternativa ao *genome scan* para a identificação de genes importantes na adaptação local do lagarto ocelado. A coloração dorsal constitui uma das características morfológicas que varia substancialmente entre as subespécies de *L. lepida*, tendo possíveis consequências adaptativas ao nível da camuflagem ou da eficiência da termorregulação. O receptor da melanocortina 1 (*Mc1r)* é um gene envolvido na síntese de melanina e, como tal, constitui um importante candidato para a variação na proporção de escamas pretas ou castanhas entre os lagartos ocelados. A análise do *Mc1r* em *L. lepida* revelou uma susbtituição derivada e não conservativa na cadeia de aminoácidos (T162I), que se encontra associada com a coloração acastanhada de *L. l. nevadensis*, sugerindo que a mutação poderá conduzir a uma perda parcial da função do gene. Uma segunda substituição (S172C) foi detectada em associação com a prevalência de escamas pretas em *L. l. lepida* e em *L. l. iberica*. No entanto, não foi detectada qualquer mutação no gene *Mc1r* associada com a maior proporção de escamas pretas em *L. l. iberica*, implicando que tal differença se deva a mutações regulatórias que afectem a expressão do gene *Mc1r* ou a mutações noutros genes envolvidos na pigmentação. Os resultados da análise do gene *Mc1r* no lagarto ocelado constituem a primeira contribuição para a determinação da base genética da variação na coloração desta espécie e serão úteis no delineamento da investigação futura. As consequências

funcionais das mutações detectadas neste estudo deverão ser testadas com ensaios *in vitro* de forma a confirmar a sua associação com os fenótipos de cor do sardão.

Os dados obtidos neste estudo para o lagarto ocelado a partir de um grande número de marcadores nucleares confirmam a previsão inicial de que a evolução da espécie é congruente com a perspectiva génica do processo de especiação, encontrando-se cada subspécie em differentes estádios de divergência. A análise da estrutura genética do lagarto ocelado foi realizada com base em 318 marcadores neutrais de AFLPs, 23 marcadores não-neutrais de AFLPs (ambos gerados pelo *genome scan*) e em oito microssatélites. A divergência de *L. l. nevadensis* é bem suportada tanto pelos marcadores neutrais como pelos não-neutrais, confirmando que a subespécie se encontra nos estádios finais do processo de especiação. Por outro lado, a divergência de *L. l. iberica* é sobretudo explicada pelos marcadores não-neutrais, enquanto que a homogeneidade genética ao nível dos marcadores neutrais implica a ocorrência generalizada de fluxo genético, sugerindo que a subespécie se encontra nos estádios iniciais da especiação ecológica, quando o processo de divergência é ainda reversível. Relativamente à subspécie nominal, e embora estudos anteriores tenham detectado vários clades mitocondriais na área geográfica de *L. l. lepida*, estes não são totalmente suportados pelos marcadores nucleares analisados neste estudo. A incongruência entre os marcadores mitocondriais e nucleares pode ser justificada pela recente divergência dos referidos clades e por *incomplete lineage sorting* ao nível dos marcadores nucleares, embora a ocorrência de fluxo genético nas zonas de contacto entre os clades também possa contribuir para a homogeneização da variação genética entre as populações de *L. l. lepida*.

Palavras-chave: gradiente ambiental, *genome scan*, adaptação local, especiação, genes candidatos

ABSTRACT

The determination of the genetic basis of adaptive traits in natural populations is fundamental to better understand how populations adaptively diverge in heterogeneous environments and eventually give rise to new species. This work describes the efforts to identify candidate loci influenced by selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient in the Iberian Peninsula, strongly affected by climatic variables. Two subspecies are recognized at the opposite extremes of the gradient, *L. l. iberica* in the northwest and *L. l. nevadensis* in the southeast, with morphological differences that suggest their local adaptation. Candidate loci were detected through an AFLP genome scan. Detection of candidate loci with a frequentist method and with Bayesian method resulted in a similar proportion of outliers (3-4%), but only a few loci were detected by both methods, denoting differences in methods' sensitivity. Several outliers were associated with variation in temperature, insolation or precipitation along the gradient, suggesting the importance of these variables as selective pressures for local adaptation. Seven outliers were successfully characterized, being noncoding, with internal indels or repetitive elements as causes of length polymorphism, indicating that they might act as regulatory elements or are in linkage with the actual target of selection. The analysis of *Mc1r*, a candidate gene for coloration, revealed a nonconserved and derived substitution (T162I) associated with the brownish colour phenotype of *L. l. nevadensis*, suggesting a putative partial loss of function. Another substitution (S172C) was associated with the presence of black scales in both *L. l. lepida* and *L. l. iberica*, but no mutations were associated with the higher melanization of *L. l. iberica*. Analysis of genetic structure showed that *L. l. nevadensis* divergence is well supported by both neutral and non-neutral loci, confirming that the subspecies is at the final stages of its speciation process. *L. l. iberica* divergence from the nominal subspecies is mostly explained by a few adaptive loci, indicating that *L. l. iberica* might be at the early stages of ecological speciation.

Keywords: candidate genes, environmental gradient, genome scan, local adaptation, speciation

CHAPTER 1

General Introduction

GENERAL INTRODUCTION

1.1 – Speciation

The revolutionary idea that new species could arise from pre-existing ones by natural selection, proposed originally by Darwin and Wallace (1858), and later in Darwin's "*On the Origin of Species*" (1859), led to the emergence of the field of evolutionary biology. The following decades brought important developments in the understanding of genetics and heredity, but also extensive debate and controversy on the species concept and speciation mechanisms driving to the emergence of new species. Evolutionary biologists could not find to date a consensual and universal species concept (Coyne & Orr 2004). The biological species concept (Mayr 1942, 1995) describes a species as a group of interbreeding natural populations which are reproductively isolated from other such groups, and still remains the most widely accepted species concept. As a consequence, and because gene flow normally prevents divergence among populations and therefore the evolution of reproductive isolation, geographical isolation was thought to be an important barrier to trigger speciation, imposing a complete (allopatric speciation) or partial (parapatric speciation) absence of gene flow, whereas speciation in the presence of gene flow (sympatric speciation) was considered an extremely rare mode of speciation (Coyne & Orr 2004).

As the knowledge in molecular biology progressed, the rise of the neutral theory of molecular evolution reduced the importance of natural selection as a trigger for speciation (Coyne & Orr 2004), claiming that most mutations had a neutral or nearly neutral selective effect and were fixed by random drift (Kimura 1983). However, the past decade brought a renewed interest in speciation mechanisms driven by natural selection, assuming that divergence in nonallopatry is not a rare event (Via 2001; Dieckmann *et al.* 2004; Nosil 2008; Wolf *et al.* 2010). Wu (2001) introduced a new perspective for the evolution of divergence at the genomic level by proposing the genic view of the speciation process. According to this perspective, speciation can occur among differently adapted populations in the absence of complete reproductive isolation. In the early stages of speciation, populations will diverge at a few loci that are locally advantageous, whereas divergence in non-adaptive regions of the genome will be prevented by

homogenizing gene flow. At this point, divergence between incipient species can still be reversed if perturbations in the habitat occur that facilitate massive hybridization (Wu 2001). However, if the accumulation of divergence extends to genes that promote non random mating or if hybrids have lower fitness, the reproductive isolation will evolve as a by-product of adaptive divergence and the diverging groups will became new species (Wu 2001, Nosil *et al.* 2009; Via 2009).

Speciation is a process that normally takes a long time, during which the spatial circumstances might change, i.e. initial differentiation can take place in allopatry but reinforcement in latter stages can occur in parapatry (Butlin *et al.* 2008). Therefore, current research in speciation is focused on the driving forces that trigger the evolution of divergence and reproductive isolation among emerging species rather than in the geographical modes of speciation (Via 2001; Dieckmann *et al.* 2004).

Among animals, three main driving forces for speciation are currently recognized by evolutionary biologists: drift (or chance), natural selection and sexual selection (Butlin *et al.* 2008). Genetic drift corresponds to the change in allelic frequencies by chance and not by their adaptive value. Speciation under such circumstances must take place in allopatry to counteract the homogenizing effect of gene flow, and because reproductive isolation evolves under neutral conditions, it takes a long time under geographical isolation to achieve complete reproductive isolation between the new species (Sobel *et al.* 2010). According to the scarce empirical evidence collected to date, speciation events caused by drift alone are likely to be rare (Coyne & Orr 2004; Sobel *et al.* 2010). In some cases, new species might arise by chance from hybridization events between older species, resulting in the duplication of the genome (allopolyploidy) or in the recombination between genomes of parent species with the maintenance of a diploid genome (homoploid hybridization) (Mallet 2007). The new hybrid species becomes reproductively isolated from parent species due to genomic incompatibilities. This process has been frequently reported in plants but is has been rarely detected in animals (Mallet 2007).

A significant role for sexual selection in speciation has been detected in several birds, insects and fishes. Sexual selection results from assortative mating due to male-male competition or to sex-specific preference over morphological (e.g. colour, sexual ornaments) or behavioral traits (e.g. calling songs, nuptial parade) on the opposite sex, thus promoting premating isolation. The cichlid fish radiation has been a classic example for speciation in sympatry by sexual selection, yet, some authors argue that sexual

selection might have acted alongside ecological speciation in cichlids and that empirical evidence suggest that it is unlikely that sexual selection may often act alone in the formation of new species (Ritchie 2007).

Ecological speciation occurs when adaptive traits are favoured by natural selection, contributing to the evolution of barriers to gene flow between populations living in different ecological conditions, as a response to biotic and abiotic factors such as climate, resources, habitat structure, competition, predation or pathogens (Schluter 2001; Rundle & Nosil 2005). The action of divergent natural selection between environments leads to the fixation of alleles that are advantageous in one environment but not in the other (Schluter 2009). Ecological speciation might occur either in sympatry or in allopatry. When ecological divergence evolves in allopatry, a secondary contact zone can be later established if divergent populations become sympatric through dispersal or range shifts (Schluter 2001; Rundle & Nosil 2005). In a parapatric scenario, adaptive divergence can evolve along an environmental gradient, where an ancestral population gives rise to phenotypically distinct descendant species that are locally adapted and spatially segregated along the gradient, resulting in a stepped phenotypic cline (Doebeli & Dieckmann 2004).

Schulter (2001, 2009) suggests another way of speciation by natural selection besides ecological speciation, the mutation-order speciation, although few empirical examples support this speciation mode so far. Under mutation-order speciation, populations become divergent under similar natural selection pressures because different mutations arise by chance in each population and reach fixation through natural selection, even though they might be advantageous in both environments.

The evolution of reproductive isolation between emerging species might often involve a combination of both selection and stochastic forces, regardless of the mechanism that triggered the initial steps of population divergence (drift, natural selection or sexual selection) (van Doorn *et al.* 2009; Sobel *et al.* 2010).

1.2 – Detection of genes under selection

The identification and characterization of genes under selection is essential to understand the genetic mechanisms underlying the evolution of adaptive phenotypic

divergence, reproductive isolation and the formation of new species. Gathering empirical evidence for the genetic basis of adaptive traits from many taxa will make possible to address several unanswered questions, which have been much debated, through comparative genomics. For instance, how often evolution of adaptive traits results from a few genes of large effect or from many genes of small effect (Orr 2001)? What is the relative role of structural (i.e. in coding regions) versus regulatory mutations (i.e. mutations in regulatory regions affecting gene expression) in adaptive evolution (Stern & Orgogozo 2008)? Do adaptive alleles emerge from standing genetic variation or as new mutations (Barrett & Schluter 2007)? Does phenotypic convergence (i.e. the independent evolution in the same environment of the same phenotypic trait in different taxa) arise from convergence at the genetic level (Manceau *et al.* 2010)? How predictable is genetic evolution (Gompel & Prud'homme 2009)?

Several approaches have been used in a variety of organisms to detect genetic targets of selection and to investigate genomic regions or specific genes with implications on adaptation to the environment or involved in sexual selection. The identification of genes under selection is particularly challenging in non-model species (i.e. species with limited genomic resources). In the last decade, much progress has been made in the development of molecular tools, which have been applied to wild populations from a growing number of non-model species, offering the opportunity to investigate speciation mechanisms in natural ecological backgrounds where species are evolving. The following sections will detail some of the molecular tools and genomic approaches which have been successfully applied in non-model species.

1.2.1 – QTL mapping

Quantitative trait locus (QTL) mapping, also termed as linkage mapping, consists in the identification of chromosomal regions that underlie phenotypic variation. QTL mapping has been traditionally used in model species and requires the crossing of individuals with distinct phenotypes to follow the inheritance of segregating traits in a large progeny (Ellegren & Sheldon 2008). Individuals must be scored for phenotypic traits of interest and genotyped for many genetic markers across the genome in order to detect genetic markers that co-segregate with phenotypic traits. Thus, the demanding need

for genetic markers requires a wealthy amount of genomic resources, a condition that is not met by most non-model species. QTL mapping is feasible in non-model species as long as individuals from natural populations can be bred and selected for divergent phenotypes under controlled conditions (Ellegren & Sheldon 2008). QTL mapping studies for the parallel armour plate reduction in threespine stickleback (Colisimo *et al.* 2004), the wing colour pattern in butterfly *Heliconius erato* (Tobler *et al.* 2005) or for albinism in cavefish *Astyanax mexicanus* (Protas *et al.* 2006) provide some recent examples. However, some authors propose that QTL mapping without controlled breeding can be performed in species with long-term studies, that provide accurate pedigree information from free-living wild populations, along with detailed phenotypic and genotypic records (Beraldi *et al.* 2007; Slate *et al.* 2010). Nevertheless, chromosomal regions identified by QTL mapping are often large and potentially contain several genes. Further fine-scale linkage mapping is required to refine the search and narrow down the number of candidate loci affecting the phenotypic trait of interest (e.g. Colisimo *et al.* 2005; Ferguson *et al.* 2010).

1.2.2 – Genome scans

Genome scans are a commonly used tool in population genomics that consist in the comparison of relative levels of differentiation between populations based in a large number of unlinked genomic markers (Storz 2005). This strategy does not require the prior identification of candidate genes or the definition of relevant phenotypic traits and is instead a useful approach to identify candidate genes under selection. The detection of selection footprints through genome scans relies on the assumption that neutral loci across the genome will be similarly affected by demographic and historical effects (bottlenecks, population expansion, migration), while loci under selection will behave differently, exhibiting unusually higher (divergent selection) or lower (balancing selection) divergence between populations (Luikart *et al.* 2003). Loci exhibiting outlier behavior are therefore considered as candidate loci potentially under selection and can be selected for follow-up studies in order to confirm such assumptions and understand how they affect phenotypes and fitness or which selection forces are acting over them (Vasemägi & Primmer 2005).

Genome scans usually rely in molecular markers such as microsatellites (e.g. Mäkinen *et al.* 2008), single nucleotide polymorphisms (SNPs) (e.g. Namroud *et al.* 2008) or amplified fragment length polymorphisms (AFLPs) (e.g. Bonin *et al.* 2006). The development of a high number of microsatellites or SNPs is normally expensive and time consuming. AFLPs are probably the type of markers that can be most rapidly and cost effectively developed for any non-model species (Bench & Akesson 2005). The technique consists in the fragmentation of the genomic DNA in thousands of small fragments with two different restriction enzymes, normally by combining a rare cutter, like *Eco*RI, with a frequent cutter such as *Mse*I (Vos *et al.* 1995; Bonin *et al.* 2005). The following step of the method involves the attachment of adapters (short oligonucleotides) to the sticky ends of digested fragments. The sequence content and length of digested fragments beyond the restriction site of the enzymes is unknown. Moreover, a random number of fragments result from digestion in both ends by the same enzyme while other fragments result from the digestion of both enzymes (one at each end). The latter are the fragments for which the AFLP technique was designed to amplify and genotype (Vos *et al.* 1995). In order to achieve the amplification of these fragments, the polymerase chain reaction (PCR) is first performed with preselective primers, one specific for each enzyme's cutting end, whose sequence is complementary to the adaptors and to the enzyme restriction site, plus a selective base. This step will greatly reduce the amount of fragments to genotype because only fragments containing the selective base will be amplified, thus reducing the complexity of the analysis. A second selective PCR is then performed using the product of the first PCR with primers that are equal to the preselective ones plus one to three additional selective bases. Multiple sets of polymorphic AFLP markers can be generated for the same individual by conducting several selective PCRs with primers bearing different combinations of selective bases (Bench & Akesson 2005; Bonin *et al.* 2005). Unlike microsatellites and SNPs, which are multiallelic and codominant markers, AFLPs are normally scored as dominant biallelic markers due to the anonymity of the sequence content. Thus, the follow-up studies with AFLP outliers require the isolation and sequencing of outlier AFLP fragments, which is technically demanding and particularly challenging in non-model species (Wood *et al.* 2008).

1.2.3 – Transcriptome analysis

Genome-wide variation in gene expression among natural populations with divergent phenotypes can be assessed through hybridization intensity in microarrays, (Ellegren & Sheldon 2008). This technique requires the development of species-specific arrays containing probes for thousands of genes, a process that can be technically demanding. Sometimes it is possible to use cross-species microarrays developed from closely-related species. The measurement of gene expression requires the isolation of messenger RNA (mRNA) from living tissues, which is often an invasive procedure in animals, frequently requiring the sacrifice of the sampled individuals. The amount of mRNA is tissue-specific and may change with time, stress or other environmental factors. Moreover, microarrays are prone to several sources of technical variance that can compromise the precision of gene expression, thus requiring the analysis of several replicates (Whitehead & Crawford 2006). All these limitations reduce the attractiveness of the technique for many non-model species. Even though, some studies have successfully used gene expression microarrays to target candidate genes associated with adaptive traits by differential expression in non-model species, such as the fish *Fundulus heteroclitus* (Oleksiak *et al.* 2002), Darwins' finches (Abzhanov *et al.* 2006) or the whitefish, *Coregonus clupeaformis* (Derome *et al*. 2006; Derome & Bernatchez 2006). Recent studies have implemented an alternative strategy to microarrays for the detection of genes with differential expression using a cDNA-AFLP genome scan, which combines the use of mRNA (reverse transcribed into complementary DNA, cDNA) with an AFLP protocol (Martínez-Fernández *et al.* 2010). This strategy overcomes the need for the development of massive genomic information in non-model species for the construction of species-specific arrays.

1.2.4 – Candidate genes

The investigation of candidate genes to assess the genetic basis of adaptive phenotypes in non-model species is sometimes an effective shortcut to genome scans or QTL mapping. For non-model species not suitable for maintenance in captivity, nor to conduct controlled crosses, and with very scarce genomic resources, a candidate gene

approach may provide an effective and rewarding alternative to determine the genetic variation underlying adaptive phenotypes, using preexisting knowledge about genes affecting similar phenotypic traits in other species. However, the candidate genes approach is necessarily biased towards genes that are already characterized (Hoekstra 2006; Jensen *et al.* 2007) and only genome scans or quantitative trait loci analysis can highlight unsuspected targets of selection in the genome. Nevertheless, the study of vertebrate pigmentation genes in wild populations has provided several successful examples, benefiting from the vast knowledge accumulated from model organisms about the genes involved in pigmentation pathways (Hoekstra 2006; Protas & Patel 2008; Hubbard *et al.* 2010). Hoekstra *et al.* (2006) investigated the melanocortin-1 receptor (*Mc1r*), a gene involved in the melanin synthesis, and identified a derived amino acid mutation in association with adaptive colour pattern in beach mice (*Peromyscus polionotus*). The involvement of *Mc1r* in beach mice coloration was confirmed by functional assays *in vitro*, showing that the amino acid change reduces the activity of the receptor (Hoekstra *et al.* 2006). Complementing QTL mapping with candidate genes' genotypes can also enhance the power of QTL analysis to highlight the genes underlying phenotypic traits, as demonstrated in beach mice (Steiner *et al.* 2007) and cavefish (Gross *et al.* 2009).

1.2.5 – Next generation sequencing

The emergence of new DNA sequencing technologies from 2005 onwards has opened new and exciting prospects for the investigation of adaptation in non-model species (Ansorge 2009; Stapley *et al.* 2010). The next generation sequencing technologies (NGS) use nanotechnology to generate millions of small sequence reads simultaneously, overcoming the traditional automated sequencers, which implement the Sanger method through capillary electrophoresis, and are limited to the parallel sequencing of 96 samples (Hudson 2008; Mardis 2008). Several mutually competitive NGS platforms became commercially available. The best known and most popular platforms are the Roche 454 FLX Titanium system, the Illumina's Genome Analyser (Solexa) and SOLiD from Applied Biosystems (Stapley *et al.* 2010). NGS dramatically reduce the time and cost

needed to obtain huge amounts of data from the genome or transcriptome of any organism, offering the opportunity to study species that are relevant for questions in evolutionary biology but that are distantly related to model species with sequenced genomes (Hudson 2008). This has been the case for several species, such as the marine snail *Littorina saxatilis*, the butterfly *Heliconius melpomene* or the tree-spined stickleback, *Gasterosteus aculeatus,* where NGS has been applied to study adaptation to the environment and speciation (Ferguson *et al.* 2010; Galindo *et al.* 2010; Hohenlohe *et al*. 2010). Genome scans and QTL mapping can greatly benefit from the implementation of NGS, by generating thousands of genetic markers in non-model species, such as SNPs or microsatellites, much faster than by traditional methods (Slate *et al*. 2010; Stapley *et al.* 2010). Moreover, the implementation of NGS for many individuals from different populations simultaneously corresponds to a genotype-by-sequencing approach, where the sequence content of genetic markers emerges from anonymity, thus overcoming limitations raised by the use of anonymous and dominant markers such as AFLPs (Rowe *et al.* 2011).

The NGS are very young in the field, and like all new technologies, they still have some weaknesses and limitations (McPherson 2009). The large amount of data generated by NGS requires large computation and storing capacity and some studies report difficulties in dealing with repetitive sequences or indels in large genome assemblies based on short length sequences (Schatz *et al.* 2010; Gnerre *et al.* 2011). Assemblies of NGS sequences are more reliable when a reference genome is available, but most species have no reference genome available yet. These problems shall be overcome with progress in technology and in bioinformatic tools used for data analysis. The overall cost per base is lower for NGS when compared to Sanger sequencing, but it is still limiting for most low budget projects. However, a third generation of sequencing technologies is under development, with the Single Molecular Real Time (SMRT) DNA sequencer from Pacific Biosciences, which proposes to produce much longer read lengths (up to 10,000 bases) than current NGS platforms or Sanger sequencing (up to 1,000 bases) (McCarthy 2010). Such long reads will make *de novo* assembly of non-model species genome feasible, reliable and probably even cheaper than it is today.

1.2.6 – Validation of candidate loci

Once candidate regions of the genome potentially under selection have been targeted by statistical inference through one or several of the approaches described above combined, further studies are needed to understand the causes of selection signatures. Sequence analysis of candidate genes' coding regions is useful to identify amino acid changes and their association with phenotypic variation. The consequences of changes in amino acids with different physico-chemical properties in the conformation and functionality of the protein can be inferred based on homology to protein structure in well-studied species (e.g. Wheat *et al.* 2006). Furthermore, whenever possible, functional assays might provide solid evidence for the consequences of alternative alleles in the phenotype. Such functional assays include *in vitro* studies, where proteins can be generated in bacteria, in cultured eukaryotic cells or in a cell free system, and tested for levels of enzyme activity on substrates (Jensen *et al.* 2007). Functional consequences of *Mc1r* mutations has been tested this way in beach mice (Hoekstra *et al.* 2006), in lizards (Rosenblum *et al.* 2010) and even in an extinct species, the mammoth (Römpler *et al.* 2006). Because proteins encoded by genes are often involved in complex metabolic networks, the link between the functional effects of mutations in a single gene and the consequences in individual fitness is rarely direct due to the effect of epistasy (i.e. when effects of one gene are modified by other genes) (Storz & Wheat 2010). Measures of single-gene expression are also important to validate candidate genes, since genes with a conserved amino acid chain can still affect the phenotype by changes in gene expression, which are triggered by regulatory mutations. This was the case of the *Agouti* gene in beach mice (Steiner *et al.* 2007) and in deer mice (Linnen *et al.* 2009). Functional assays *in vivo* are perhaps more powerful than *in vitro* assays, because the function is tested in the genetic background where it evolved (Jensen *et al.* 2007). However, *in vivo* assays are restricted to a limited group of species where transgenic technologies are possible. Recent studies in cavefish *Astyanax mexicanus* provide an interesting example for *in vivo* functional assays in non-model species. The functional effect of *Mc1r* mutations was tested by gene knockdown in zebrafish (*Danio rerio*) embryos, while *in vitro* transcribed RNA derived from alternative *Mc1r* alleles in cavefish were injected in those embryos to evaluate changes in pigmentation during early development (Gross *et al.* 2009).

1.3 – Lizards as models for selection and adaptation

Lizards are a very successful group of vertebrates that have diversified to fill a wide variety of ecological niches (Pianka & Vitt 2003). Several lizard species have become model organisms in ecological studies (e.g. species from the genus *Anolis*), where differences in body size or shape and coloration have been frequently considered as local adaptations to the habitat (Losos *et al.* 1994; Ogden & Thorpe 2002; Thorpe & Stenson 2003; Thorpe *et al.* 2008). However, when compared to the other groups of vertebrates, lizards are far behind in the application of population genomics and genome sequencing. The first lizard genome was finally released in 2010, from *Anolis carolinensis* (Alföldi *et al.* 2011), and hopefully more lizard genomes will follow, bridging a long-standing gap in vertebrate's genomics and opening new perspectives for the investigation of the genetic basis of adaptive traits (Schneider 2008). Although the genetics of adaptation in lizards is still largely unknown, this is a group with great potential for speciation studies (Camargo *et al.* 2010). The species chosen for the present study was *Lacerta lepida*, a non-model lizard species with spatial morphological and genetic variation, suitable to study local adaptation. Current knowledge in *L. lepida* characteristics and evolution will be detailed in the following sections.

1.3.1 – *Lacerta lepida*

Lacerta lepida Daudin, 1802 is a lizard species from the Lacertidae family (species of this family are native from Africa, Europe or Asia) and belongs to a group of lizard species commonly designated as ocellated lizards or jewelled lizards, due to the characteristic eyespots on their dorsum and flanks. Mayer & Bischoff (1996) reviewed their taxonomy and proposed the exclusion of the subgenus Timon (where ocellated lizards were included) from genus Lacerta, and its upgrade to a full genus based on morphological and karyological peculiarities. Therefore, *L. lepida* is currently known as *Timon lepidus*, but the former designation is still in use in recently published works (Díaz *et al.* 2006; Paulo *et al.* 2008; Miraldo *et al.* 2011), and will be used throughout this thesis.

Ocellated lizards are large sized lizards with a distribution that is restricted in Europe to almost all of the Iberian Peninsula and some regions in the South of France and North of Italy, while in Africa they can be found in Morocco and in northern Algeria and Tunisia (Fig. 1). Ocellated lizards from Europe and North Africa were until recently considered as the same species, but the morphological and genetic divergence between lizards from each side of the Strait of Gibraltar led to the recognition of North African lizards in Tunisia and Algeria as *Lacerta pater* Lataste, 1880 and in Morocco as *Lacerta tangitana* Boulenger, 1887, while only European ocellated lizards remained as *L. lepida* (Mateo *et al.* 1996; Fig. 2).

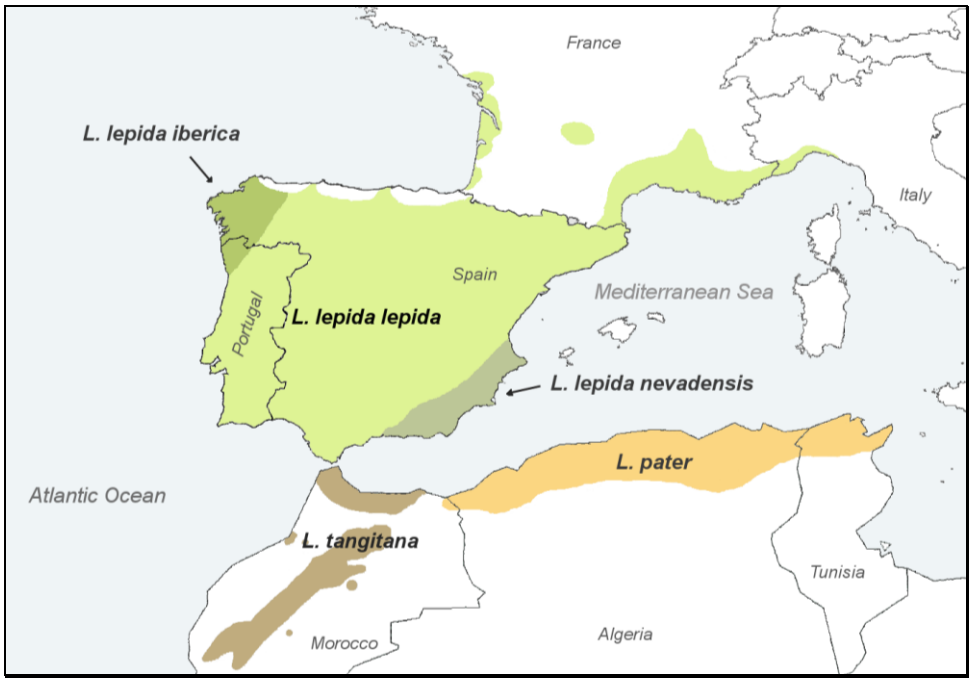


Fig. 1 Map of the western Mediterranean with the distribution of ocellated lizards (adapted from Paulo (2001)). *Lacerta lepida* (green) is the single species of ocellated lizards living in Europe. Distribution of subspecies *L. l. iberica* (in the northwest of the Iberian Peninsula) and *L. l. nevadensis* (in the southeast of the Iberian Peninsula) are recognizable with darker tones of green, while light green corresponds to the nominal subspecies distribution, *L. l. lepida*. In Africa, two species can be found: *L. tangitana* (brown) in Morroco and *L. pater* (light orange) in Algeria and Tunisia.

There are four subspecies currently recognized within *L. lepida*: *Lacerta lepida lepida* Daudin, 1802; *Lacerta lepida nevadensis* Buchholz, 1963; *Lacerta lepida iberica* (López Seoane, 1884) and *Lacerta lepida oteroi* Castroviejo & Mateo, 1998. The nominal subspecies occupies most of the distribution area, occurring in typically Mediterranean habitat. It is replaced by *L. l. nevadensis* in the southeastern region of the Iberian Peninsula, along the Mediterranean Sea coastline, including most of the Spanish provinces of Almeria, Murcia and Alicante. *L. l. lepida* is also replaced by *L. l. iberica* in the northwest of the Iberian Peninsula, along the western coast of Galicia and north of Portugal (Fig. 1). The fourth subspecies, *L. l. oteroi*, is restricted to the small island of Sálvora, located near the Galicia coast. This isolated subspecies exhibits some differences in morphology and in karyotype, but it resembles *L. l. iberica* from the nearby coast (Castroviejo & Mateo 1998). The present work will focus on the three subspecies inhabiting the mainland in the Iberian Peninsula (Fig. 2), and therefore all mentions to *L. lepida* subspecies hereafter will not be extensive to *L. l. oteroi*.

Fig. 2 European and African ocellated lizards. From Europe: (1) *Lacerta lepida lepida*; (2) *Lacerta lepida iberica* and (5) *Lacerta lepida nevadensis*. From Africa: (3) *Lacerta tangitana* and (4) *Lacerta pater*.

Ocellated lizards are diurnal, poikilothermic and ground-dwelling lizards that forage actively on a large variety of mainly arthropod prey (Castilla & Bawens 1989;

Busack & Visnaw 1989). Morphological and genetic divergence has been documented between the three mainland subspecies in the Iberian Peninsula and some authors argue that it might be important for the local adaptation to the contrasting climatic and ecological conditions registered in each subspecies' distribution range (Mateo 1998; Mateo & Castroviejo 1990; Mateo & López-Jurado 1994). Clinal variation has been detected in several morphological traits in *L. lepida*, following a northwest-southeast climatic gradient across the Iberian Peninsula. This gradient is affected by the irregular distribution of precipitation and temperature (Fig. 3), which in turn affects the vegetation cover and even the properties of the soil. Therefore, the distribution of each *L. lepida* subspecies coincides with different bioclimatic regions (Fig. 4).

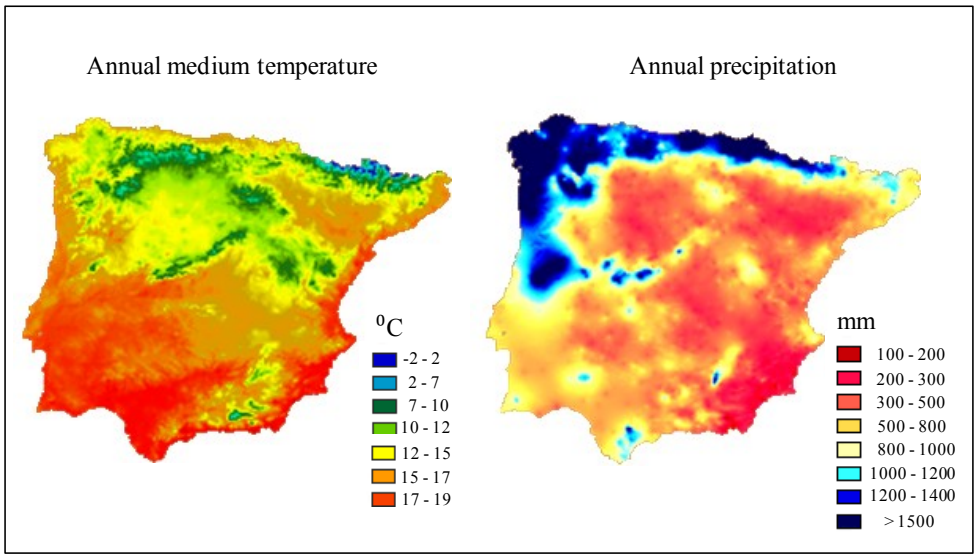


Fig. 3 Distribution of annual medium temperature and annual precipitation in the Iberian Peninsula according to measures recorded from 1950 to 1999. Adapted from the Digital Climatic Atlas of the Iberian Peninsula (Ninyerola *et al.* 2005, http://www.opengis.uab.es/WMS/iberia/en_index.htm).

Most of the distribution area of the nominal subspecies is characterized by a Mediterranean climate, with hot and dry summers. In the southeast, coinciding with most of *L. l. nevadensis* distribution, the Mediterranean climate assumes xeric properties, with reduced (< 300 mm) and irregular annual rainfall (IGN 1992). On the opposite direction, in the northwest of the Iberian Peninsula, the distribution of *L. l. iberica* is associated with a temperate climate with abundant annual rainfall (> 2000 mm) and mild summers. This

region presents the lowest annual hours of sunshine in the Iberian Peninsula (< 2000 h; IGN 1992). The following sections will describe in more detail the differences between the three *L. lepida* subspecies as provided in available literature.

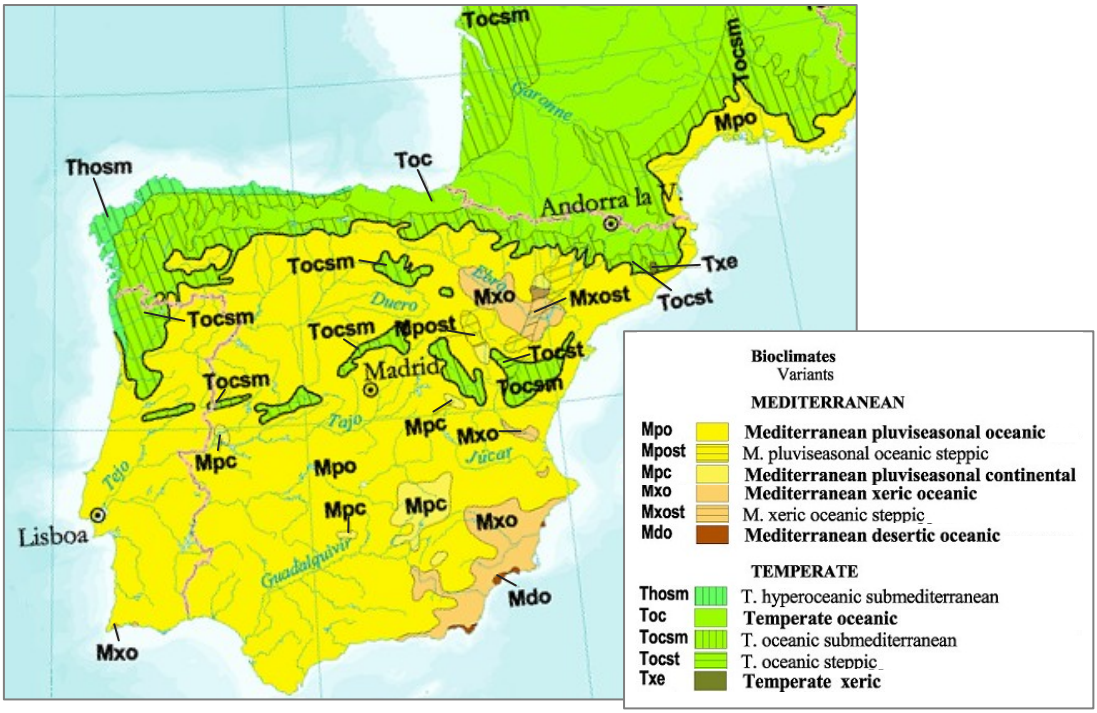


Fig. 4 Bioclimatic map of the Iberian Peninsula and south of France (adapted from Rivas-Martínez *et al.* 2004, http://www.ucm.es/info/cif/form/bi_med.htm). Most of the Iberian Peninsula is characterized by a Mediterranean bioclimate (Mpo variant), except in the northwest, which presents a Temperate bioclimate (Toc variant). In the southeast of the Iberian Peninsula, a Mediterranean bioclimate with xeric properties can be recognized (Mxo variant).

1.3.1.1 – Variation in morphology

The most conspicuous differences in morphology between the three *Lacerta lepida* subspecies are related to biometry and colour pattern (Fig. 2). According to Mateo & Castroviejo (1990), the average snout-to-vent length (SVL) measured in *L. l. iberica* (males: 147.16 ± 18.42 mm; females: 130.78 ± 11.69 mm) was notoriously inferior to the measures obtained in the nominal subspecies (males: 166.36 ± 27.39 mm; females:

151.08 ± 21.84 mm) and in *L. l. nevadensis* (males: 175.40 ± 25.21 mm; females: 153.23 ± 17.19 mm). The head is longer and broader in *L. l. nevadensis* males and they present more femoral pores in the hind limbs than males from the other subspecies (Mateo & Castroviejo 1990). Lizards belonging to *L. l. nevadensis* are also the ones with more pronounced differences in colour pattern. They normally have no dorsal black scales, which are replaced by brown/grey scales in the colour pattern. The dorsal ornamentation composed by a pattern of more or less perfect rings of dark scales enclosing yellow/green scales is often absent in this subspecies, especially in large males. However, when dorsal circles are well defined, they are normally arranged in two longitudinal series in *L. l. nevandensis*, whereas they are arranged in several transversal series in *L. l. iberica*. The transversal banding is common in juveniles but adult lizards from *L. l. iberica* maintain this trait whereas it normally disappears with age in the nominal subspecies (Mateo 1988). The coloration presented by *L. l. iberica* is darker and results from the higher proportion of black scales in the dorsal pattern in this subspecies (Mateo & Castroviejo 1990; Mateo & López-Jurado 1994). The presence of ornamentation in the neck and in the limbs is commonly observed in *L. l. iberica* but is nearly absent in *L. l. nevadensi*s (Mateo 1988). Ocellated lizards have conspicuous blue eyespots in their body flanks, which have probably an important role for social signaling (Font *et al.* 2009). The blue scales of lateral eyespots are bordered by a ring of black scales in *L. l. iberica* and in *L. l. lepida*, but not in *L. l. nevadensis*, because black scales are lacking in this subspecies' colour pattern (Mateo 1988).

Mateo (1988) proposed that the smaller body size and darker coloration presented by *L. l. iberica* could be driven by thermoregulatory limitations in the northwest, where sunshine hours and temperatures are lower, and such morphological traits could contribute to reduce the amount of solar exposition required to heat the body. The same author proposed that the brownish/greyish dorsal coloration in *L. l. nevadensis* might be effective for crypsis with the soil in the southeast, because large portions of the landscape have no vegetation cover.

Dentition is another morphological trait that varies among *L. lepida* subspecies (Fig. 5). The number of teeth increases from the southeast to the northwest: 16-19 teeth in *L. l. nevadensis* (average = 17.5); 17-22 teeth in *L. l. lepida* (average = 18.6) and 19-26 teeth in *L. l. iberica* (average = 21.3). However, the degree of morphological specialization among teeth increases in the opposite direction, with *L. l. iberica* presenting

homogeneous teeth both in size and symmetry, whereas *L. l. nevadensis* has both large and small teeth, with one, two or three cusps (Mateo 1988; Castroviejo & Mateo 1998; Fig. 5). Differences among subspecies in teeth number and morphology might be related with slight variation in diet composition. *L. l. nevadensis* diet has been reported as less diverse, with a preference for Tenebrionidae beetles and a shift towards frugivory in scarcity periods in the dry zones from Almeria and Alicante (Busack & Visnaw 1989; Hodar *et al.* 1996), whereas in populations from *L. l. lepida*, in central Spain, Scarabidae beetles were the most abundant in ocellated lizards' diet (Castilla 1989). Athough Coleoptera are the main prey for ocellated lizards, Gastropoda are specially abundant in humid regions and have a relevant contribution to *L. l. iberica* diet (Mateo 1988).

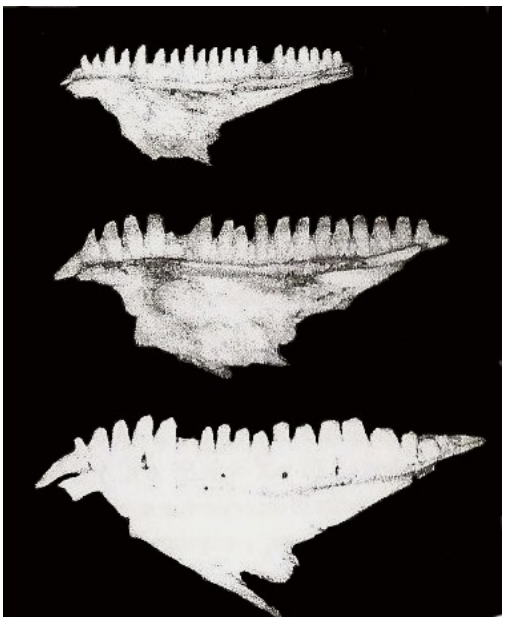


Fig. 5 Maxilla from adult males belonging to different *Lacerta lepida* subspecies. From top to bottom: *L. l. iberica*, *L. l. lepida* and *L. l. nevadensis* (adapted from Mateo 1988).

Analysis of the karyotype in *L. lepida* revealed three distinct patterns which differ in the number and position of the nucleolar organizers (NORs). The same pattern is found in *L. l. lepida* and *L. l. nevadensis*, with only one pair of NORs at the 5th pair of chromosomes, while three patterns were found in lizards from *L. l. iberica*: i) the same pattern as in the other subspecies, with one pair of NORs at the 5th pair; ii) one pair of

NORs at the 11th pair of chromosomes and iii) two pairs of NORs, at the 5th and the 11th pairs of chromosomes (Castroviejo & Mateo 1998; Mateo *et al.* 1999).

1.3.1.2 – Variation in reproductive strategy

Lacerta lepida females reach their sexual maturity normally around 32-33 months of age, after the third winter since birth. Ocellated lizards emerge from winter hibernation in March and matings occur from April to June. Females lay a single clutch (with 16.9 eggs on average) until the end of July (Castilla & Bauwens 1989). Newborns appear in September and the hibernation period resumes in October until the following spring. In the northwest, *L. l. iberica* females mature a year before, when they reach 21 months of age, starting their sexual activity with a relatively smaller body size (Castilla & Bauwens 1989; Mateo & Castanet 1994). However, *L. l. iberica* females seem to have a shorter life span which rarely exceeds two breeding seasons (Mateo & Castanet 1994). *L. l. iberica* females born and raised in captivity, in the same conditions as *L. l. lepida* and *L. l. nevadensis,* maintain a significant lower body size after their third year of life, suggesting that size is genetically determined (Mateo & Castanet 1994). Females from *L. l. nevadensis* produce smaller clutches (12.3 eggs on average), but composed by larger eggs, and may produce more than one clutch per year, resulting in an extended and asynchronous laying period (Castilla & Bauwens 1989; Mateo & Castanet 1994).

The length of the annual cycle of activity and reproduction differs between subspecies, probably as the result of lizards' adjustment to the local climatic conditions. In the southeast, temperature rises earlier in the spring, leading to a shorter hibernation period. Copulations in *L. l. nevadensis* may start as early as February and the laying period extends from March until June (Mateo & Castanet 1994). In the northwest, the end of lizards' hibernation may be delayed in the year until the spring weather warms up. The first copulations in *L. l. iberica* were observed only in May and pregnant females were not detected until the end of June (Mateo & Castanet 1994).

1.3.1.3 – Genetic variation

The first studies concerning the genetic variability within European ocellated lizards were reported in Mateo (1988) and Mateo *et al.* (1996), based in allozymes' diversity, and demonstrated that *L. l. nevadensis* was the most differentiated group within *L. lepida.* More recent studies provided detailed phylogenetic and phylogeopraphic patterns based in additional genetic data (mitochondrial DNA and nuclear genes) and confirmed *L. l. nevadensis* as the most divergent subspecies (Paulo 2001; Paulo *et al.* 2008; Miraldo 2009; Miraldo *et al.* 2011). These studies detected further genetic structure and geographical subdivisions among *L. lepida* populations. According to divergence time estimates from Paulo *et al*. (2008), the split between European and African ocellated lizards occurred around 11.33 million years ago (Ma). The split between *L. l. nevadensis* (clade N, Fig. 6) and the remaining European ocellated lizards (clade L, Fig. 6) is estimated in 9.43 Ma, during the Miocene, and resulted in two reciprocally monophyletic clades (Paulo *et al*. 2008). The divergence within clade L occurred more recently, during the climatic cycles of the Quaternary (between 2.0 and 0.5 Ma), giving rise to six subclades (clades L1-L5 and Guadalquivir; Paulo 2001; Paulo *et al.* 2008; Miraldo *et al.* 2011; Fig. 6). The oldest split within clade L separated clades L2 and L3 from the remaining clades. Interestingly, clade L3 corresponds to the subspecies *L. l. iberica* from the northwest of the Iberian Peninsula, while clade L2 is geographically separated by several hundred kilometers from clade L3 and is composed by populations from the south of Portugal (Fig. 7). These genetic patterns suggest a history of population's contraction and allopatric divergence of *L. lepida* in different refugia during the glacial periods of the Pleistocene. Miraldo *et al.* (2011) detected signs of recent demographic and spatial expansion in each clade and identified several zones of secondary contact between geographically adjacent clades. The contact zone between *L. l. nevadensis* and *L. l. lepida* (clade L4) was investigated in more detail by Miraldo (2009). The genetic patterns obtained with both mitochondrial and nuclear (microsatellites) DNA were consistent with a narrow contact zone and limited gene flow, attesting for the deep divergence level between these subspecies, which are following independent evolutionary trajectories.

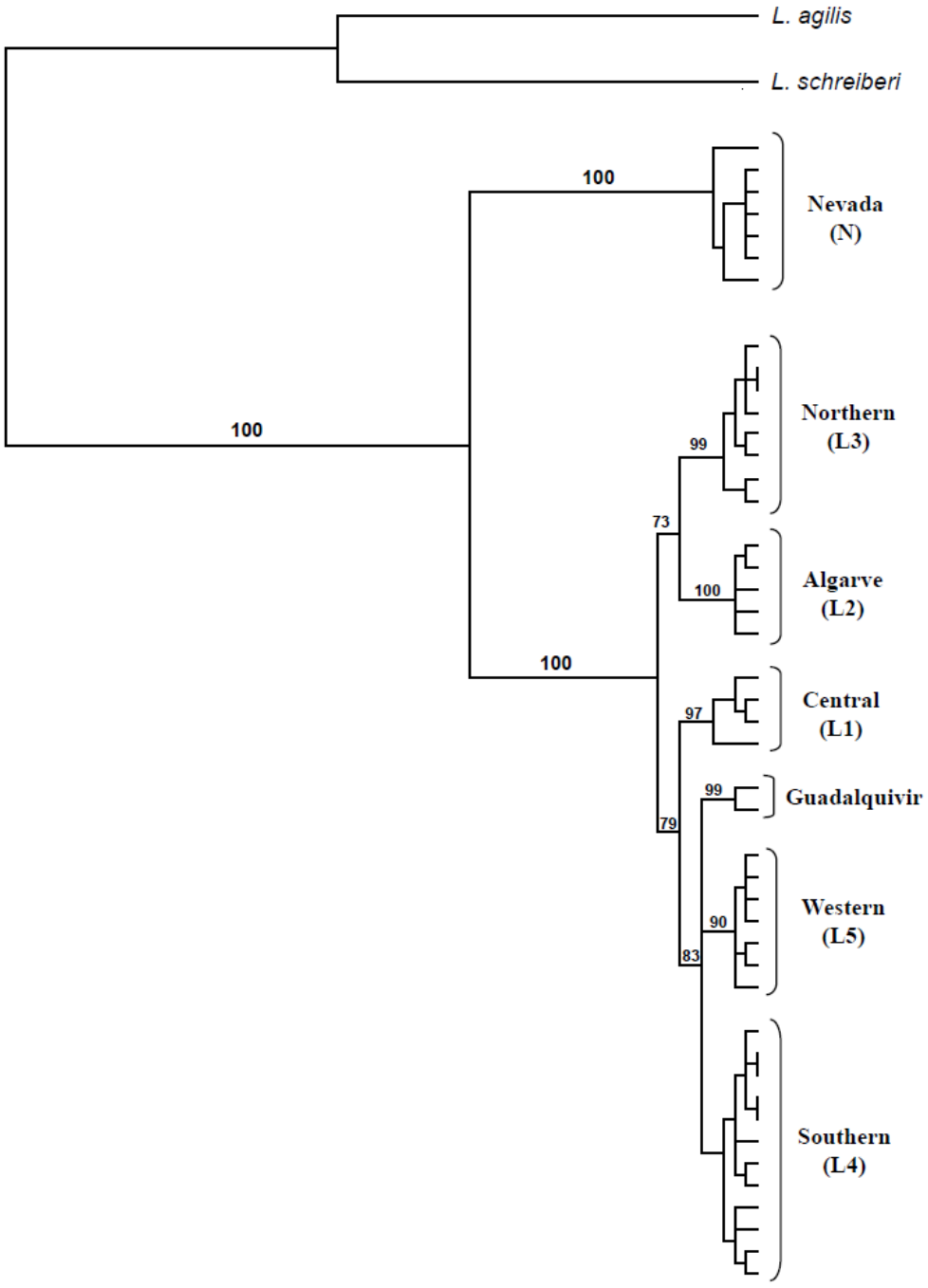


Fig. 6 Maximum likelihood phylogenetic tree (with molecular clock enforced) based on mitochondrial DNA cytochrome *b* fragments (adapted from Paulo 2001). Numbers above branches correspond to bootstrap support values obtained from 100 pseudo-replicates. Seven clades were detected within *Lacerta lepida*.

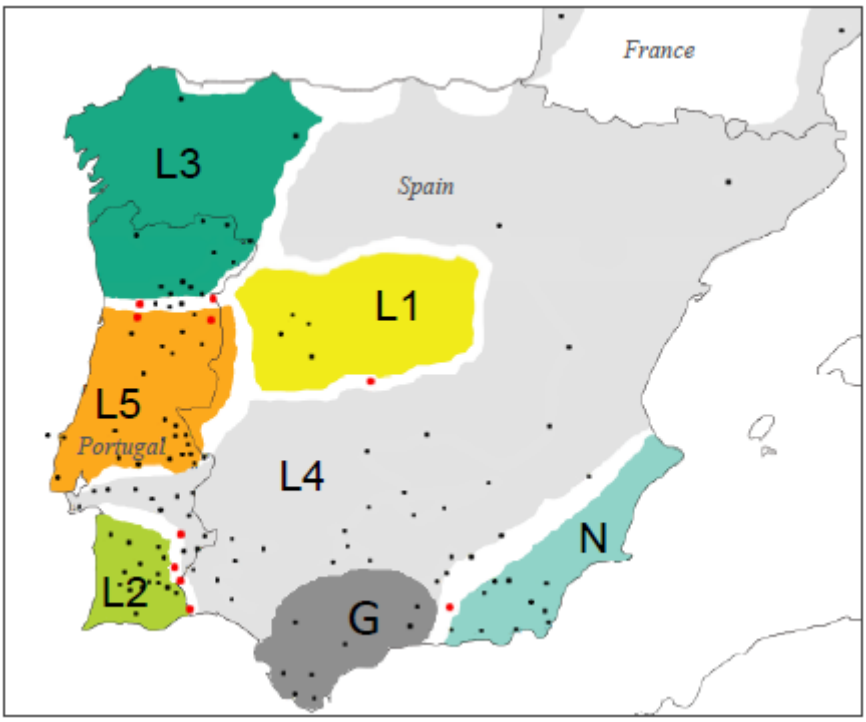


Fig. 7 Geographical distribution of *Lacerta lepida* mitochondrial clades as inferred from cytochrome *b* sequences (adapted from Miraldo *et al.* 2011). Sampled localities are represented by black dots and red dots. Red dots indicate localities where two or more phylogroups were detected in sympatry by Miraldo *et al.* (2011). The phylogroup "G", represented in dark grey, was detected by Paulo (2001) but was not confirmed by Miraldo *et al.* (2011) phylogeographic analysis.

1.4 – Objectives and thesis structure

The main objective of this work is to gather empirical evidence for the speciation process in spatially structured populations, in the absence of abrupt environmental barriers for the occurrence of gene flow, through the identification and investigation of genes or genomic regions involved in the local adaptation. The species chosen as a model is the ocellated lizard (*Lacerta lepida*), which has scarce genomic resources available. The species has a wide distribution range in the Iberian Peninsula, with ecological and climatic heterogeneity and three recognized parapatric subspecies. Previous studies indicate that variation in several morphological traits might be a response to local selective pressures. The phylogeny and phylogeography of the species corroborates the morphological differentiation at the genetic level and demonstrates that populations are spatially structured, with a level of genetic divergence that reflects several divergence timescales and a complex evolutionary history. This work addresses, for the first time, the genetic variation in ocellated lizards from an adaptive perspective, investigating the role of natural selection in the evolutionary history of the species. Two main approaches were chosen for the detection of genes under selection: a population genomic approach using an AFLP genome scan and a candidate gene approach through the investigation of *melanocortin-1-receptor* (*Mc1r*), a candidate gene for melanin-based coloration.

To achieve the main objective, the following specific objectives were established:

(i) To statistically detect candidate loci under selection (outlier loci) through a genome scan with AFLP markers in European ocellated lizards sampled across an environmental gradient;

(ii) To infer possible selective pressures acting over AFLP outliers by testing for associations between AFLP outliers' band frequency and variation in environmental variables along the environmental gradient;

(iii) To isolate and sequence AFLP markers detected as outliers in the ocellated lizards genome scan, bringing them out of anonymity, infer their homology with known genes and evaluate outliers' haplotypic variation among ocellated lizards;

(iv) To investigate the genetic basis of dorsal colour variation in ocellated lizards through the analysis of sequence variation in a candidate gene involved in the melanin synthesis pathway, the *Mc1r*, and look for associations between amino acid changes in *Mc1r* and colour phenotypes of ocellated lizards;

(v) To compare the genetic structure in European ocellated lizards as inferred from neutral markers (microsatellites and AFLPs) and from loci suspected to be under the effect of selection (AFLP outliers), gaining some insights about gene flow and heterogeneity in nuclear genomic divergence along the environmental cline.

To address the above specific objectives, the thesis is organized in four data chapters. The first data chapter (chapter 2) provides the results from the AFLP genome scan for selection in European ocellated lizards and from tests for association between AFLP markers and environmental variables (objectives i and ii). The second data chapter (chapter 3) describes the efforts to isolate, sequence and characterize AFLP outliers (objective iii), identified as candidate loci potentially under selection in chapter 2. The third data chapter (chapter 4) presents the sequence analysis of a candidate gene (*Mc1r*) in European ocellated lizards and its association with dorsal colour phenotypes (objective iv). Finally, the fourth data chapter (chapter 5) presents a comparative analysis of microsatellite data with AFLP data, using neutral markers and outlier markers in the analysis of population structure of European ocellated lizards (objective v). The implications of the findings of this thesis are debated in the general discussion (chapter 6) and the thesis ends with the enunciation of the main achievements and the new questions raised by this work, indicating which topics should be addressed in future research.

1.5 – References

Abzhanov A, Kuo WP, Hartmann C *et al.* (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, **442**, 563-567.

Alföldi J, Di Palma F, Grabherr M *et al.* (2011) The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*, **477**, 587-591.

Ansorge WJ (2009) Next-generation DNA sequencing techniques. *New Biotechnology*, **25**, 195-203.

Barret RDH, Schluter D (2007) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38-44.

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, **14**, 2899-2914.

Beraldi D, McRae AF, Gratten J *et al.* (2007) Mapping quantitative trait loci underlying fitness-related traits in a free-living sheep population. *Evolution*, **61**, 1403-1416.

Bonin A, Pompanon F, Taberlet P (2005) Use of amplified fragment length polymorphism (AFLP) markers in surveys of vertebrate diversity. *Methods in enzymology*, **395**, 145-161.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773-783.

Busack S, Visnaw JA (1989) Observations on the natural history of *Lacerta lepida* in Cádiz province, Spain. *Amphibia-Reptilia*, **10**, 201-213.

Butlin RK, Galindo J, Grahame JW (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **363**, 2997-3007.

Camargo A, Sinervo B, Sites JW (2010) Lizards as model organisms for linking phylogeographic and speciation studies. *Molecular Ecology*, **19**, 3250-3270.

Castilla, AM (1989) *Autoecología del lagarto ocelado (Lacerta lepida)*. PhD thesis, Universidad Autónoma de Madrid, Madrid.

Castilla AM, Bauwens D (1989) Reproductive characteristics of the lacertid lizard *Lacerta lepida*. *Amphibia-Reptilia*, **10**, 445-452.

Castroviejo J, Mateo JA (1998). Una nueva subespecie de *Lacerta lepida* Daudin 1802 (Sauria, Lacertidae) para la isla de Salvora. *Publicaciones de la Asociación de Amigos de Doñana*, **12**, 1-21.

Colisimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928-1933.

Colisimo PF, Peichel CL, Nereng K *et al.* (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biology*, **2**, 635-641.

Coyne JA, Orr HA (2004) *Speciation*. Sunderland, MA: Sinauer Associates.

Darwin C (1859) *The origin of species*, 6th ed. London, UK: John Murray.

Darwin C, Wallace AR (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the Proceedings of the Linnean Society of London. Zoology*, **3**, 46-50.

Derome N, Bernatchez L (2006) The transcriptomics of ecological convergence between 2 limnetic Coregonine fishes (Salmonidae). *Molecular Biology and Evolution*, **23**, 2370-2378.

Derome N, Duchesne P, Bernatchez L (2006) Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. *Molecular Ecology*, **15**, 1239-1249.

Díaz JA, Monasterio C, Salvador A (2006) Abundance, microhabitat selection and conservation of eyed lizards (*Lacerta lepida*): a radiotelemetric study. *Journal of Zoology*, **268**, 295-301.

Dieckmann U, Doebeli M, Metz JAJ, Tautz D (eds) (2004) *Adaptive speciation*. Cambridge, UK: Cambridge University Press.

Doebeli M, Dieckmann U (2004) Adaptive dynamics of speciation: spatial structure. In: Dieckmann U, Doebeli M, Metz JAJ, Tautz D (eds) (2004) *Adaptive speciation*. Cambridge, UK: Cambridge University Press.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169-175.

Ferguson L, Lee SF, Chamberlain N (2010) Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the *HmYb/Sb* locus. *Molecular Ecology*, **19**, 240-254.

Font, E, Pérez I De Lanuza G, Sampedro C (2009) Ultraviolet reflectance and cryptic sexual dichromatism in the ocellated lizard, *Lacerta (Timon) lepida* (Squamata: Lacertidae). *Biological Journal of the Linnean Society,* **97**, 766-780.

Galindo J, Grahame JW, Butlin RK (2010) An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *Journal of Evolutionary Biology*, **23**, 2004-2016.

Gnerre S, MacCallum I, Przybylski D *et al.* (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 1513-1518.

Gompel N, Prud'homme B (2009) The causes of repeated genetic evolution. *Developmental Biology,* **332**, 36-47.

Gross JB, Borowsky R, Tabin CJ (2009) A novel role for *Mc1r* in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genetics*, **5,** e1000326.

Hodar JA, Campos F, Rosales BA (1996) Trophic ecology of the Ocellated Lizard *Lacerta lepida* in an arid zone of southern Spain: Relationships with availability and daily activity of prey. *Journal of Arid Environments*, **33**, 95-107.

Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, **97**, 222-234.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101-104.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hubbard JK, Uy JAC, Hauber ME, Hoekstra HE, Safran RJ (2010) Vertebrate pigmentation: from underlying genes to adaptive function. *Trends in Genetics*, **26**, 231-239.

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources*, **8**, 3-17.

IGN (1992) Atlas Nacional de España, Sección II: El medio terrestre - Grupo 9-Climatología, 1ª edicion. Instituto Geografico Nacional, Spain.

Jensen JD, Wong A, Aquadro CF (2007) Approaches for identifying targets of positive selection. *Trends in Genetics*, **23**, 568-577.

Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge, UK: Cambridge University Press.

Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an adaptive allele in deer mice. *Science*, **325**, 1095-1098.

Losos JB, Irschick DJ, Schoener TW (1994) Adaptation and constraint in the evolution of specialization of Bahamian *Anolis* lizards. *Evolution*, **48**, 1786-1798.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981-994.

Mäkinen HS, Cano JM, Merilä J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Molecular Ecology*, **17**, 3565-3582.

Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279-283.

Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE (2010) Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **365**, 2439-2450.

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, **24**, 133-141.

Martínez-Fernández M, Bernatchez L, Rolán-Alvarez E, Quesada H (2010) Insights into the role of differential gene expression on the ecological adaptation of the snail *Littorina saxatilis*. *BMC Evolutionary Biology*, **10**, 356.

Mateo JA (1988) *Estudio Sistemático y Zoogeográfico de los Lagartos Ocelados, Lacerta lepida Daudin, 1802 y Lacerta pater (Lataste 1880) (Sauria, Lacertidae)*. PhD thesis, University of Sevilla, Sevilla.

Mateo JA, Castanet J (1994) Reproductive strategies in 3 Spanish populations of the ocellated lizard, *Lacerta Lepida* (Sauria, Lacertidae). *Acta Oecologica*, **15**, 215-229.

Mateo JA, Castroviejo J (1990) Variation morphologique et révision taxonomique de l'espèce *Lacerta lepida* Daudin, 1802 (Sauria, Lacertidae). *Bulletin du Muséum National D'Histoire Naturelle, Paris*, **12**, 691-706.

Mateo JA, López-Jurado LF (1994) Variaciones en el color de los lagartos ocelados; aproximación a la distribución de *Lacerta lepida nevadensis* Buchholz 1963. *Revista Española de Herpetología*, **8**, 29-35.

Mateo JA, López-Jurado LF, Cano J, Guillaume CP (1999) Intraspecific variability in nucleolar organiser position in *Lacerta (Timon) lepida*. *Amphibia-Reptilia*, **20**, 73-77.

Mateo JA, López-Jurado LF, Guillaume CP (1996) Proteic and morphological variations in ocellated lizards (Lacertidae): a complex of species across the Strait of Gibraltar. *Comptes Rendus de l' Academie des Sciences Serie III-Sciences de la vie,* **319**, 737-746.

Mayer W, Bischoff W (1996) Beiträge zur taxonomischen Revision der Gattung *Lacerta* (Reptilia: Lacertidae). Teil 1: Zootoca, Omanosaura, Timon und Teira als eigenständige Gattungen. *Salamandra*, **32**, 163-170.

Mayr E (1942) *Systematics and the origin of species*. New York, NY: Columbia University Press.

Mayr E (1995) Species, classification, and evolution. In: Arai R, Kato M, Doi Y (eds). *Biodiversity and Evolution*. Natural Science Museum Foundation, Tokyo.

McCarthy A (2010) Third generation DNA sequencing: Pacific Biosciences' single molecule real time technology. *Chemistry & Biology*, **17**, 675-676.

McPherson JD (2009) Next-generation gap. *Nature Methods*, **6**, S2-S5.

Miraldo A (2009) *Phylogeography and population dynamics of secondary contact zones of Lacerta lepida in the Iberian Peninsula*. PhD thesis, University of East Anglia, Norwich.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, **17**, 3599-3613.

Ninyerola M, Pons X, Roure JM (2005) Atlas Climático Digital de la Península Ibérica. Metodología y aplicaciones en bioclimatología y geobotánica. ISBN 932860-8-7, Universidad Autónoma de Barcelona, Bellaterra.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103-2106.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375-402.

Ogden R, Thorpe RS (2002) Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 13612-13615.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature genetics*, **32**, 261-266.

Orr HA (2001) The genetics of species differences. *Trends in Ecology and Evolution*, **16**, 343-350.

Paulo OS (2001). *The phylogeography of reptiles of the Iberian Peninsula*. PhD thesis, University of London, London.

Paulo OS, Pinheiro J, Miraldo A *et al.* (2008) The role of vicariance vs. dispersal in shaping genetic patterns in ocellated lizard species in the western Mediterranean. *Molecular Ecology*, **17**, 1535-1551.

Pianka ER, Vitt LJ (2003) *Lizards: Windows to the evolution of diversity*. University of California Press, Berkeley, CA.

Protas ME, Hersey C, Kochanek D *et al.* (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics*, **38**, 107-111.

Protas ME, Patel NH (2008) Evolution of coloration patterns. *Annual Review of Cell and Developmental Biology*, **24**, 425-446.

Ritchie MG (2007) Sexual selection and speciation. *Annual Review of Ecology, Evolution and Systematics,* **38**, 79-102.

Rivas-Martínez S, Penas A, Díaz TE (2004), Bioclimatic map of Europe. Servicio Cartográfico de la Universidad de León, España (http://www.ucm.es/info/cif/form/bi_med.htm).

Römpler H, Rohland N, Lalueza-Fox C *et al.* (2006) Nuclear gene indicates coat-color polymorphism in mammoths. *Science*, **313**, 62.

Rosenblum EB, Römpler H, Schöneberg T, Hoekstra HE (2010) Molecular and functional basis of phenotypic convergence in white lizards at White Sands. P*roceedings of the National Academy of Sciences of the United States of America,* **107**, 2113-2117.

Rowe HC, Renaut S, Guggisberg A (2011) RAD in the realm of next-generation sequencing technologies. *Molecular Ecology*, **20**, 3499-3502.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336-352.

Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second-generation sequencing. *Genome Research*, **20**, 1165-1173.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372-380.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737-741.

Schneider CJ (2008) Exploiting genomic resources in studies of speciation and adaptive radiation of lizards in the genus *Anolis*. *Integrative and Comparative Biology*, **48**, 520-526.

Slate J, Santure AW, Feulner PGD *et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, **26**, 275-284.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. *Evolution*, **64**, 295-315.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology and Evolution*, **25**, 705-712.

Steiner CC, Weber JN, Hoekstra HE (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biology*, **5**, 1880-1889.

Stern D, Orgogozo V (2008) The loci of evolution: how predictable is genetic evolution? *Evolution*, **62**, 2155-2177.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671-688.

Storz JF, Wheat CW (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution*, **64**, 2489-2509.

Thorpe RS, Stenson G (2003) Phylogeny, paraphyly and ecological adaptation of the colour and pattern in the *Anolis roquet* complex on Martinique. *Molecular Ecology*, **12**, 117-132.

Thorpe RS, Surget-Groba Y, Johansson H (2008) The relative importance of ecology and geographic isolation for speciation in anoles. *Philosophical Transaction of the Royal Society of London, Series B: Biological Sciences*, **363**, 3071-3081.

Tobler A, Kapan D, Flanagan NS *et al.* (2005) First-generation linkage map of the warningly colored butterfly *Heliconius erato*. *Heredity*, **94**, 408-417.

van Doorn GS, Edelaar P, Weissin FJ (2009) On the origin of species by natural and sexual selection. *Science*, **326**, 1704-1707.

Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623-3642.

Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381-390.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the Unites States of America*, **106**, 9939-9946.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP - A new technique for DNA-fingerprinting. *Nucleic Acids Research*, **23**, 4407-4414.

Wheat CW, Watt WB, Pollock DD, Schulte PM (2006) From DNA to fitness differences: sequences and structures of adaptive variants of Colias phosphoglucose isomerase (PGI). *Molecular Biology and Evolution*, **23**, 499-512.

Whitehead A, Crawford DL (2006) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, **15**, 1197-1211.

Wolf JBW, Lindell J, Backström N (2010) Speciation genetics: current status and evolving approaches. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **365**, 1717-1733.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123-3135.

Wu C-I 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851-865.

CHAPTER 2

Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient

Vera L. Nunes[1], Mark A. Beaumont[2,3], Roger K. Butlin[4], Octávio S. Paulo[1]

[1] Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, DBA/FCUL, 1749-016 Lisboa, Portugal

[2] School of Animal and Microbial Sciences, University of Reading, P.O. Box 228, Reading, RG6 6AJ, United Kingdom

[3] Schools of Mathematics and Biological Sciences, University of Bristol, UK

[4] Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, England

Molecular Ecology (2011) 20, 193-205

Molecular Ecology (2011) **20**, 193–205 doi: 10.1111/j.1365-294X.2010.04936.x

Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient

VERA L. NUNES,* MARK A. BEAUMONT,†‡ ROGER K. BUTLIN§ and OCTÁVIO S. PAULO*
**Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, DBA/FCUL, 1749-016 Lisboa, Portugal, †School of Animal and Microbial Sciences, University of Reading, PO Box 228, Reading RG6 6AJ, UK, ‡Schools of Mathematics and Biological Sciences, University of Bristol, Bristol, UK, §Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK*

Abstract

Identification of loci with adaptive importance is a key step to understand the speciation process in natural populations, because those loci are responsible for phenotypic variation that affects fitness in different environments. We conducted an AFLP genome scan in populations of ocellated lizards (*Lacerta lepida*) to search for candidate loci influenced by selection along an environmental gradient in the Iberian Peninsula. This gradient is strongly influenced by climatic variables, and two subspecies can be recognized at the opposite extremes: *L. lepida iberica* in the northwest and *L. lepida nevadensis* in the southeast. Both subspecies show substantial morphological differences that may be involved in their local adaptation to the climatic extremes. To investigate how the use of a particular outlier detection method can influence the results, a frequentist method, DFDIST, and a Bayesian method, BayeScan, were used to search for outliers influenced by selection. Additionally, the spatial analysis method was used to test for associations of AFLP marker band frequencies with 54 climatic variables by logistic regression. Results obtained with each method highlight differences in their sensitivity. DFDIST and BayeScan detected a similar proportion of outliers (3–4%), but only a few loci were simultaneously detected by both methods. Several loci detected as outliers were also associated with temperature, insolation or precipitation according to spatial analysis method. These results are in accordance with reported data in the literature about morphological and life-history variation of *L. lepida* subspecies along the environmental gradient.

Keywords: adaptive divergence, AFLP, environmental gradient, landscape genetics, natural selection, outlier loci

Received 16 September 2009; revision received 7 October 2010; accepted 14 October 2010

Introduction

Uncovering the genetic basis of important adaptive traits in natural populations is a major goal for evolutionary biology in order to better understand how populations adaptively diverge in heterogeneous environments and eventually give rise to new species. It is now well accepted that differentiation can occur in the presence of gene flow if adaptively driven (Schluter 2009). This mode of speciation is likely to produce genomically heterogeneous divergence, unlike the classic view of allopatric speciation, where the whole genome should behave as a cohesive unit in the development of reproductive isolation (Wu 2001; Nosil *et al.* 2009; Via 2009). When populations face different environments, ecologically based selection can arise, leading to local adaptations. These may select for assortative mating or other pre-zygotic forms of isolation, reducing and perhaps eventually eliminating gene flow between the emerging species (Gavrilets 2004).

Correspondence: Vera L. Nunes, Fax: +351 217500028;
E-mail: vlnunes@fc.ul.pt

© 2010 Blackwell Publishing Ltd

Identifying which regions of the genome are under selection and understanding what selective pressures are acting upon natural populations is challenging, in particular for non-model species. Although searching for selection signatures in candidate genes for adaptive traits has been fruitful in some studies (Rosenblum *et al.* 2004; Hoekstra *et al.* 2006), this kind of approach depends on the availability of well-characterized genes and is limited to well-studied metabolic pathways that underlie measurable phenotypic traits. Association of phenotypes with genotypes by quantitative trait loci analysis also relies on well-studied species, which are easy to manipulate and cross (Stinchcombe & Hoekstra 2008), although future prospects in this field are promising (see Ellegren 2008). In the last few years, population genomic approaches have become accessible and popular for searching for genes influenced by selection, even in non-model organisms (Black *et al.* 2001; Luikart *et al.* 2003). In particular, for species with little genetic information, scans using AFLP markers have become a fast and economic tool to survey several hundreds of random loci across the whole genome (Bensch & Akesson 2005), although a full coverage of the species' genome can hardly be attained with this kind of approach. Demography and the neutral evolutionary history of populations affect neutral loci across the genome in the same way, while loci under selection and closely linked loci will exhibit an outlier pattern of variation (Luikart *et al.* 2003). This strategy has been successfully applied in natural populations to detect candidate loci underlying adaptation to altitude (Bonin *et al.* 2006) or temperature (Jump *et al.* 2006), or to investigate ecotype-based differentiation (Wilding *et al.* 2001; Campbell & Bernatchez 2004; Egan *et al.* 2008; Herrera & Bazaga 2008; Nosil *et al.* 2008) and species boundaries (Murray & Hare 2006; Savolainen *et al.* 2006; Minder & Widmer 2008). However, as well as identifying outliers as candidate loci under selection, it is crucial to investigate which selective forces are acting upon them (Joost *et al.* 2007) and disentangle the possible functional roles of outlier loci (Vasemagi & Primmer 2005; Jensen *et al.* 2007). As genome scans by AFLPs rely on anonymous markers, follow-up analyses are needed to identify and characterize those loci. This task may still be challenging in non-model organisms, but some studies have already accomplished it (Minder & Widmer 2008; Wood *et al.* 2008).

The species studied here, *Lacerta lepida*, is the only species of ocellated lizard occurring in Europe. It is also designated *Timon lepidus* since Mayer & Bischoff (1996) upgraded the subgenus *Timon* (we will continue to use the previous designation hereafter, as it is still the more widely used). The species' distribution is limited to the Iberian Peninsula and some regions in the South of France and North of Italy (Fig. 1). The species occurs in a wide set of environmental conditions. Mateo (1988) emphasized the coincidence of a climatic gradient running from southeast (SE) to northwest (NW) across the Iberian Peninsula with the major morphological differences observed among populations, coinciding also with the geographical distribution of three currently

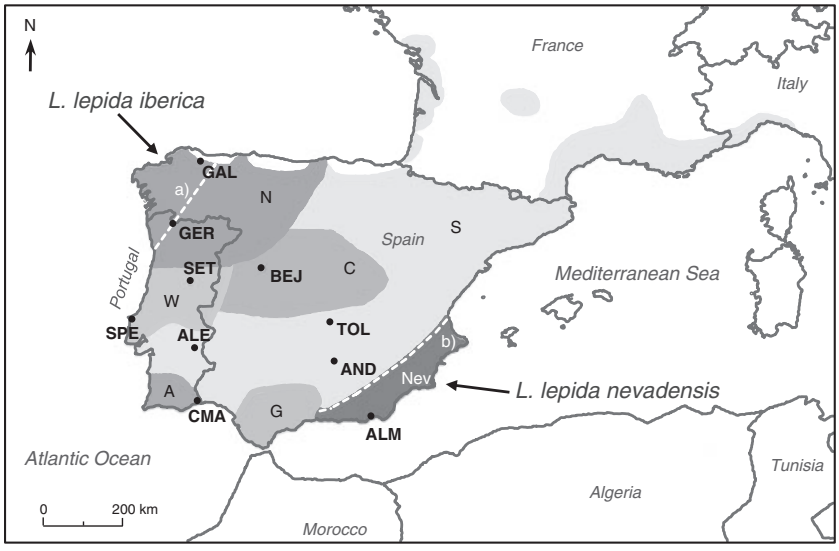


Fig. 1 Map of the western Mediterranean region showing the distribution area of *Lacerta lepida* in Europe. Distribution limits of *L. lepida* subspecies based on morphology are indicated by dashed lines: (a) *L. l. iberica*; (b) *L. l. nevadensis*. *L. l. lepida* occupies the remaining shaded area in Iberian Peninsula, France and Italy. Predicted geographical distribution of each mitochondrial clade (cyt *b*) (Paulo 2001; Paulo *et al.* 2008) is represented by different grey shades. Sampled populations are represented by black dots (codes as in Table 1).

© 2010 Blackwell Publishing Ltd

accepted subspecies. The nominal subspecies, *L. lepida lepida*, occurs in the typical Mediterranean climate that characterizes most of the distribution area. *L. lepida nevadensis* occupies the most arid areas in the SE along the Mediterranean Sea coast and *L. lepida iberica* is limited to the occidental coast of Galicia, in the NW, characterized by a rainy Atlantic climate (Mateo & Castroviejo 1990). Several morphological traits show clinal variation, including average body size (decrease in size towards NW), colour pattern (increase of black scales in the dorsum in NW and their replacement by brown scales in SE) and dentition (morphological differentiation among teeth increases but their number decreases, both from NW to SE) (Mateo 1988; Mateo & Castroviejo 1990; Mateo & López-Jurado 1994). Some karyological (Mateo *et al.* 1999), allozyme (Mateo *et al.* 1996) and even reproductive strategy (Castilla & Bauwens 1989; Mateo & Castanet 1994) differences between subspecies have also been recorded. Neither the genetic basis nor the effect on fitness have yet been investigated for the morphological traits that show clinal variation, but some of these morphological differences may lead us to suspect selective forces shaping each subspecies' local adaptation to major climatic and ecological differences (the latter being in turn influenced by climatic conditions, such as vegetation cover, food availability and thermoregulatory conditions). This effect may be more dramatic at the opposite extremes of the cline, where *L. l. iberica* and *L. l. nevadensis* replace the nominal and widespread subspecies, *L. l. lepida*.

Studies conducted on *L. lepida* with the cytochrome *b* mitochondrial gene (Paulo 2001) reveal this species as highly structured (seven clades), with strong geographical subdivisions and limited gene flow. The separation of *L. l. nevadensis* (clade Nev), with about 13% mtDNA sequence divergence, is the oldest split in the Iberian Peninsula, estimated to have occurred about 9.43 million years ago (Ma), while subspecies *L. l. iberica* (clade N), with <3% mtDNA sequence divergence from the nominal subspecies, is estimated to be derived from a Plio-Pleistocenic splitting event, corresponding to a much more recent differentiation, around 2 Ma (Paulo *et al.* 2008).

The genetic and geographical structure of ocellated lizards in the Iberian Peninsula probably reflects the existence of multiple refugia during the last glaciations, which would have suffered several demographic contractions and expansions through the successive climatic oscillations (Paulo *et al.* 2008). Given the unique geographical characteristics of the Iberian Peninsula, the severity of the climatic changes affecting each population would depend on the latitude, the topography and the influence of Atlantic Ocean or Mediterranean Sea proximity. Therefore, ocellated lizards from NW and SE extremes of the Iberian Peninsula evolved under different environmental conditions.

Here, we present the results of a genome scan using AFLPs, with populations collected along the SE–NW environmental gradient in the Iberian Peninsula, to detect candidate loci influenced by selection. Several methods for outlier detection are now available, but the consequences for the results of choosing a particular detection method have not been fully investigated. If, as in this case, the purpose of the study is to detect a set of candidate loci influenced by selection for future confirmatory investigations, it is important that by choosing a specific detection method, the results will not be dramatically different from other methods. Here, a frequentist method, DFDIST (Beaumont & Nichols 1996), and a Bayesian method, BayeScan (Foll & Gaggiotti 2008), were used for outlier detection. Results from both methods were compared, assessing the advantages and limitations of the use of each method. To investigate which selective pressures may be acting upon ocellated lizards along the climatic gradient, associations of AFLP band frequencies with climatic variables, such as temperature or precipitation, were tested by logistic regression as implemented in the spatial analysis method (SAM) (Joost *et al.* 2008). By comparing and combining results from the three methods, we aimed to obtain a list of candidate loci for further investigation of ocellated lizard's adaptation to different environments along the Iberian Peninsula.

Materials and methods

Sampling and DNA isolation

A total of 10 populations of *Lacerta lepida* were sampled in the Iberian Peninsula along a SE–NW transect and in a north–south transect along the Atlantic coast, covering the distribution of the three subspecies and all cyt *b* mitochondrial clades, except the Gualdaquivir clade (Paulo 2001) (Fig. 1). Population locations, sample sizes and corresponding mtDNA clades are listed in Table 1. In each location, tissue samples from tail were collected from free-living adult lizards from both sexes, in approximately equal proportions. The animals were immediately released back into the wild. Whole genome DNA was extracted from tail tissue using the Jetquick Tissue DNA kit (Genomed).

Environmental data

At least one measure of GPS coordinates was recorded in the field for each sampled location. Lizards were collected no more than 3 km away from the GPS coordinates. Climatic data was obtained from public

© 2010 Blackwell Publishing Ltd

Table 1 Populations used in the genome scan. Sample sizes (N) and mitochondrial clade (cyt *b*) (Paulo 2001; Paulo *et al.* 2008) are indicated for each location. Values of annual mean temperature (Temp), annual precipitation (Prec), annual insolation (Ins) and annual relative humidity (Hum) are also presented

Population	Code	Latitude	Longitude	Taxa	N	mtDNA clade (Cyt *b*)	Temp (°C)	Prec (mm)	Ins (h)	Hum (%)
Galicia	GAL	43° 21′ 60″ N	7° 22′ 03″ W	*L. l. iberica*	19	N (Northern)	10.6	1233	1800	75
Gerês	GER	41° 43′ 23″ N	8° 06′ 50″ W	*L. l. iberica*	23	N (Northern)	10.8	1228	2000	75
Béjar	BEJ	40° 40′ 15″ N	5° 36′ 32″ W	*L. l. lepida*	22	C (Central)	10.5	565	2500	70
Serra da Estrela	SET	40° 19′ 24″ N	7° 36′ 44″ W	*L. l. lepida*	22	W (Western)	10.3	1563	2500	70
Peniche	SPE	39° 19′ 41″ N	9° 20′ 45″ W	*L. l. lepida*	16	W (Western)	15.1	616	2400	85
Toledo	TOL	39° 15′ 32″ N	3° 44′ 01″ W	*L. l. lepida*	22	S (Southern)	13.9	477	2700	60
Alentejo	ALE	38° 35′ 55″ N	7° 33′ 40″ W	*L. l. lepida*	13	S (Southern)	16.5	622	3000	70
Andalucia	AND	38° 16′ 51″ N	3° 37′ 02″ W	*L. l. lepida*	16	S (Southern)	15.9	646	2700	60
Castro Marim	CMA	37° 14′ 08″ N	7° 26′ 43″ W	*L. l. lepida*	25	A (Algarve)	17.3	488	2900	80
Almería	ALM	36° 49′ 54″ N	2° 31′ 32″ W	*L. l. nevadensis*	18	Nev (Nevadensis)	17.8	226	2900	65

databases for GPS coordinates of the 10 sampling sites in the Iberian Peninsula. Precipitation (annual and monthly values) and temperatures (annual and monthly values of maximum, mean and minimum temperatures) were obtained from the Digital Climatic Atlas of the Iberian Peninsula (Ninyerola *et al.* 2005, available at http://www.opengis.uab.es/WMS/iberia/en_index.htm). It presents a continuous distribution of values with a resolution of 200 m, obtained by spatial interpolation from data collected by meteorological stations for the period 1950–1999. Two other climatic variables, annual insolation (hours of sunshine) and annual relative humidity, were obtained from the Atlas of Portugal (APA 1975, scale 1:1 000 000) and the Atlas of Spain (IGN 1992, scale 1:4 500 000), where data is represented by isoline maps. In the Atlas of Portugal, for insolation (data from 1931–1960) and humidity (data from 1938–1970), the isoline of lower value flanking each sampling point was chosen. The same strategy was used to collect humidity data in the Atlas of Spain (data from 1956–1985), but for insolation, as isolines delimit intervals of 200 h instead of 100 h as in the Atlas of Portugal, the median value between flanking isolines was chosen for sampled locations in Spain. Mean annual values at each population main coordinates for temperature, precipitation, insolation and relative humidity are listed in Table 1.

AFLP genotyping

AFLP markers were obtained with a modified version of the original protocol in Vos *et al.* (1995). Digestion of 50–200 ng of genomic DNA was performed with 10 units each of restriction enzymes, *Eco*RI and *Mse*I, and 2× NEBuffer 2 (NEB) in 20 μL reaction for 3 h at 37 °C, followed by enzyme inactivation at 70 °C for 15 min. Ligation of adaptors was conducted for 16 μL of digested DNA in a 40-μL reaction with 1× T4 DNA Ligase Reaction Buffer, 75 pmol of each adaptor and 40 units of T4 DNA Ligase (NEB). The ligation reaction took place overnight at 16 °C and products were diluted 10-fold. For selective amplification, reactions of 10 μL were performed with 3.3 μL of template, 1× PCR buffer (Promega), 0.5 units of *Taq* polymerase (Promega), 1.5 mM $MgCl_2$, 0.3 mM of DNTPs, 1 pmol of *Mse*I selective primer and 0.2 pmol of *Eco*RI selective primer labelled with fluorescent dye (6-FAM or VIC). AFLP markers were generated for the following eight primer combinations (*Eco*RI-*Mse*I): ACA-CAG, AAC-CAC, ACT-CTG, AAC-CTC, ACT-CTT, AAG-CAC, ACA-ACA, and AAG-CTA. Fragments were separated by electrophoresis on an ABI PRISM 310 Genetic Analyser (Applied Biosystems) with Genescan-500 LIZ as internal size standard. Only polymorphic markers were considered and presence versus absence of peaks was called automatically by GeneMapper 3.7 (Applied Biosystems), setting a fluorescent signal detection threshold of 100 units to avoid background noise. All loci for all individuals were visually inspected and corrected for peak miscalls.

The error rate was estimated as the ratio of mismatches in 22 replicates (10.8% of total samples) to total number of replicated markers (Pompanon *et al.* 2005). After removal of markers with high mismatch rates between replicates, the error rate for 392 markers (35–64 per primer combination) was estimated at 5.6%, lower than the maximum value of 10% recommended by Bonin *et al.* (2007). The size range of AFLP markers was 62–303 bp, and 92% of markers had a fragment size above 100 bp.

Outlier detection

A modification of the FDIST2 software (Beaumont & Nichols 1996) for dominant markers, DFDIST (http://

© 2010 Blackwell Publishing Ltd

www.rubic.rdg.ac.uk/~mab/stuff), was used to detect outlier loci, potentially influenced by selection, as loci with unusually high or low F_{ST} values when compared to neutral expectations. The first step of the analysis consists of the calculation of the neutral target F_{ST} for the empirical distribution, after excluding loci with critical frequency for the most common allele equal to or above 0.98. Second, a null distribution of F_{ST} close to the empirical distribution is obtained by coalescent simulation (50 000 realizations). Simulations were performed with a mean F_{ST} similar to the trimmed mean F_{ST}. The latter is intended to be an estimate uninfluenced by outlier loci (Bonin *et al.* 2006). Therefore, it was computed by removing the 30% highest and the 30% lowest F_{ST} values observed in the empirical distribution. Different values for the $4N_e\mu$ parameter were initially tested for the simulations (0.04, 0.06 and 0.4, corresponding to an N_e of 10 000, 15 000 and 100 000, respectively), but results remained robust as previously reported (Beaumont & Nichols 1996) and a value of 0.06 was chosen to perform all the simulations. Finally, empirical and simulated distributions were compared so that loci lying outside the upper and lower confidence levels were highlighted as outliers, and thus as candidate loci influenced by selection. A global analysis was performed with all 10 populations sampled in the Iberian Peninsula (Fig. 1). Both 95% and 99% confidence intervals (CI) were considered. However, the concerns about high false detection rates (type-I error) by DFDIST pointed out by several authors (Bonin *et al.* 2007; Caballero *et al.* 2008; Pérez-Figueroa *et al.* 2010) suggest the use of more restrictive significance levels to control for false positives. Therefore, a false discovery rate (FDR) of 5% was adopted for DFDIST analysis by applying the Benjamini & Hochberg (1995) method, as implemented by Chiurugwi *et al.* (2010).

Detection of outliers was also performed with the BayeScan software (http://www-leca.ujf-grenoble.fr/logiciels.htm), which implements the method of Foll & Gaggiotti (2008). This method directly estimates the probability that each locus is subject to selection using a Bayesian method. The method uses population-specific and locus-specific components of F_{ST} coefficients and assumes that allele frequencies follow a Dirichlet distribution. BayeScan takes all loci into account for the analysis and seems to be robust when dealing with complex demographic scenarios for neutral genetic differentiation (Foll & Gaggiotti 2008). As in DFDIST analysis, the 10 populations sampled in the Iberian Peninsula were used in a global analysis, performed with default parameters. Outliers detected with posterior probabilities above 0.99, but also above 0.95 or 0.90, were taken into consideration.

Association with environmental variables

The program SAM (Joost *et al.* 2008), available at http://www.econogene.eu/software/sam, was used to test for associations between the frequency of AFLP bands' presence/absence and data from environmental variables at sample locations. SAM computes multiple univariate logistic regression models. To ensure the robustness of the method, likelihood ratio (G) and Wald statistical tests are implemented to assess the significance of coefficients calculated by the logistic regression function. A model is considered significant only if the null hypothesis is rejected by both tests, after Bonferroni correction. For both tests, the null hypothesis is that the model with the examined variable does not explain the observed distribution better than a model with a constant only (Joost *et al.* 2007).

The 392 AFLP markers were tested against 54 climatic variables consisting of monthly precipitation and temperature (maximum, mean and minimum), and annual precipitation, temperature, insolation and relative humidity. All 10 populations from the Iberian Peninsula were used in the SAM analysis. The significance threshold was set to 4.72E-7, corresponding to a 99% CI after Bonferroni correction.

Neutral differentiation and genetic structure

Loci under divergent selection are expected to show larger differentiation among populations than neutral loci. To compare levels of differentiation between neutral AFLP markers and outliers potentially under selection, pairwise F_{ST} was calculated in AFLP-SURV 1.0 (Vekemans 2002). AFLP markers were grouped in two data sets: neutral loci and outlier loci. To minimize the presence of loci under weak selection or false negatives in the neutral loci data set, only AFLP markers that were never detected as outliers in any of the DFDIST (95% CI), BayeScan (with posterior probability above 0.90) or SAM (99% CI with Bonferroni correction) analyses were considered. The outlier loci data set corresponds to a conservative list intended for future validation studies, with outliers detected at more restrictive significance levels. The outliers considered were detected with a posterior probability higher than 0.99 by BayeScan or with a 5% FDR by DFDIST.

In order to evaluate whether the AFLP markers displayed the same genetic structure as that obtained with mtDNA (Table 1) or if population clustering would follow the geographical location of the climatic extremes, both neutral and outlier loci data sets were tested with analyses of molecular variance (AMOVAs) in Arlequin 3.5 (Excoffier *et al.* 2005). Analyses were performed with populations grouped in (i) a single group; (ii) in

© 2010 Blackwell Publishing Ltd

three groups corresponding to each subspecies, which also correspond to the major climatic regions; and (iii) in six groups corresponding to each mtDNA clade.

Results

Outlier detection

A total of 392 polymorphic AFLP markers were successfully scored for 10 populations of ocellated lizards sampled along the climatic gradient in the Iberian Peninsula. The same data set was used in a global analysis for outlier detection, both with DFDIST and BayeScan.

DFDIST analysis produced 26 outliers at 99% CI and 62 at 95% CI. Among these outliers, some presented lower F_{ST} than expected under neutrality (one at 99% CI and 14 at 95% CI), corresponding to candidate loci influenced by balancing selection, which promotes locus polymorphism and the maintenance of similar allele frequencies across populations. The remaining outliers presented higher F_{ST} than expected under neutrality, being therefore potentially influenced by directional selection. When applying a 5% FDR, only 16 detections remained significant, all corresponding to outliers potentially influenced by directional selection, representing 4.1% of the investigated loci.

BayeScan analysis detected 30 outliers with posterior probability above 0.90, 20 above 0.95 and only 12 loci remained as outliers when considering posterior probabilities higher than 0.99, corresponding to 3.1% of the investigated loci. Outliers detected by BayeScan were all candidate loci potentially under divergent selection. The correspondence between loci detected as outliers by DFDIST (5% FDR) and by BayeScan (with posterior probability above 0.99) was limited. Among the 23 loci detected as outliers, only five (21.7%) were detected by both programs (Table 2). Even when less stringent significance criteria were considered (10% FDR for DFDIST, posterior probability > 0.90 for BayeScan), the proportion of loci simultaneously detected as outliers by both detection methods remained low (26%).

Table 2 List of 23 outliers selected for future validation studies. The list includes outliers detected by DFDIST with a FDR of 5% or with posterior probability above 0.99 by BayeScan. For each outlier, values of posterior probability from BayeScan (values above 0.99 in bold) and *P*-values from DFDIST (significant values with a 5% FDR in bold) are indicated. When outliers were detected by spatial analysis method (SAM), the climatic variables most strongly associated with locus' band frequency are indicated: maximum temperature (T_{max}), mean temperature (T_{mean}), minimum temperature (T_{min}), insolation (Ins), precipitation (Prec) and relative humidity (Hum)

Outlier	DFDIST *P*-value	BayeScan posterior probability	SAM
30	**0.000**	0.935	
52	0.031	**1.000**	T_{max}
75	0.024	**0.999**	Prec
140	**0.000**	0.604	
201	**0.001**	0.402	
209	**0.000**	0.684	Prec
220	0.106	**0.998**	
223	**0.000**	0.979	
228	**0.000**	0.908	
235	0.042	**0.998**	
245	**0.001**	**1.000**	T_{max}
291	**0.002**	0.509	
297	**0.000**	0.661	Prec
301	**0.000**	0.946	Prec, Ins, T_{med}
311	**0.001**	**0.999**	T_{max}, Ins
315	**0.000**	**1.000**	Ins
323	0.225	**0.992**	T_{max}
340	**0.000**	0.907	
347	**0.000**	0.710	
351	0.059	**0.991**	
353	0.037	**0.999**	Hum
386	**0.000**	**1.000**	
390	**0.002**	**0.999**	T_{max}, Ins

Association with environmental variables

AFLP marker frequency variation along the Iberian Peninsula was tested in SAM for associations with annual and monthly values of precipitation, temperature (maximum, mean and minimum), annual insolation and annual relative humidity. Significant associations were detected for 20 (5%) of 392 loci at 99% CI with Bonferroni correction (Table 3). SAM highlighted 15 loci associated with precipitation. Fifteen loci were also associated with maximum temperatures (T_{max}), but only five loci were associated with mean temperatures (T_{mean}) and only three loci with minimum temperatures (T_{min}). Eleven loci showed an association with insolation and only two loci were associated with relative humidity. Loci 317, 315 and 245 show the highest number of associations (with 37, 26 and 22 out of 54 variables, respectively), followed by loci 311 and 75 (14 and 10 associations, respectively). At the other end of the spectrum, loci 209, 213 and 353 were associated with only one variable.

The strongest associations were observed for locus 245 with T_{max} from June and locus 315 with annual insolation. Each locus found in association with T_{max} was at least associated with T_{max} from July and August, the hottest months of the year, except for loci 317 and 388. Locus 317 was associated with T_{max} for all months, June to September excepted. Associations with T_{mean} and T_{min} were also preferentially related to summer months, with the exception of locus 317.

© 2010 Blackwell Publishing Ltd

Table 3 Association between AFLP loci frequency and climatic variables along the Iberian Peninsula as detected by spatial analysis method (SAM). Significant associations above 99% CI with Bonferroni correction are denoted by [+] or [−] (* 99.9% CI, ** 99.99% CI, *** 99. 999% CI and **** 99.9999% CI). When the probability of the AFLP band presence is higher for higher values of the variable, it is denoted by [+]. When the probability of the AFLP band presence is higher for lower values of the variable, it is denoted by [−]

	Locus	245	315	311	353	52	231	34	390	304	314	75	317	301	297	370	388	323	349	209	213
Precipitation	Annual		[+]									[+]			[−]						
	January																				
	February		[+]*									[+]*									
	March																				
	April		[+]									[+]		[−]							
	May		[+]										[+]								
	June		[+]*										[+]*	[−]							
	July	[−]	[+]***	[−]**			[+]*	[−]	[+]	[+]*	[+]*	[+]	[+]			[−]					
	August		[+]**	[−]**			[+]			[+]	[+]	[+]	[+]								
	September	[−]*	[+]***	[−]		[−]							[+]								
	October		[+]**									[+]*		[−]							
	November													[−]	[+]					[+]	
	December											[+]		[−]	[+]						
Temperature (maximum)	Annual	[+]***	[−]*	[+]		[+]							[−]								
	January												[−]								
	February												[−]								
	March	[+]*											[−]								
	April	[+]**	[−]			[+]							[−]								
	May	[+]***	[−]**	[+]*		[+]*							[−]								
	June	[+]****	[−]**	[+]**		[+]**	[−]	[+]	[−]	[−]	[+]										
	July	[+]***	[−]**	[+]***		[+]*	[−]*	[+]*	[−]*	[−]*	[+]*	[−]				[+]		[+]	[+]		
	August	[+]**	[−]**	[+]***		[+]*	[−]*	[+]*	[−]*	[−]*	[+]*	[−]				[+]	[−]	[+]	[+]		
	September	[+]***	[−]**	[+]**		[+]**	[−]	[+]	[−]								[−]				
	October	[+]*	[−]										[−]								
	November												[−]								
	December												[−]								
Insolation	Annual	[+]	[−]****	[+]***		[+]	[−]*	[+]*	[−]*			[−]	[−]	[+]			[−]				
Humidity	Annual				[+]***																[+]

Whenever a locus exhibited associations with temperature or insolation, but also with precipitation, the probability for the band presence was higher for higher temperatures and insolation hours, but lower for higher levels of precipitation, or vice-versa (Table 3). For example, for locus 315, the probability of the band's presence increased for higher values of precipitation and decreased for higher temperatures and insolation, while for locus 245 the opposite trends were observed. This observation is consistent with the inverse correlation between precipitation and temperature/insolation along the Iberian Peninsula, resulting in a gradient from north and northwest to south that decreases for precipitation but increases for temperature and insolation.

Comparing DFDIST and BayeScan with SAM analyses

Among the 16 outliers detected by DFDIST (with 5% of FDR), seven loci were also highlighted by SAM as significantly associated with some of the climatic variables tested (Table 2). For BayeScan, eight of the 12 outliers detected with posterior probabilities higher than 0.99 were highlighted by SAM. Only four loci highlighted by SAM (99% CI with Bonferroni correction) in association with climatic variables were simultaneously detected as outliers by DFDIST (5% FDR) and BayeScan (posterior probability > 0.99). These include loci 315, 245 and 311, which are among the loci associated with the highest number of climatic variables by SAM, as well as locus 390 (Table 2). All of the four outliers were associated most strongly with either T_{max} or annual insolation, or with both equally.

For follow-up research, a conservative list of 23 outliers was selected, where outliers detected by DFDIST (5% FDR) or BayeScan (posterior probability > 0.99) were retained (Table 2).

Neutral differentiation and genetic structure

Pairwise F_{ST} values between all 10 populations were calculated independently with neutral loci (318 loci)

© 2010 Blackwell Publishing Ltd

and outlier loci (23 loci). The global F_{ST} for neutral loci was 0.05 and reached 0.24 for outlier loci. When considering the neutral loci, the highest values of F_{ST} were registered between ALM and the other nine populations (Table 4), with values ranging from 0.10 (ALM-SET) to 0.16 (ALM-GAL). For the outlier loci data set, an increase in F_{ST} was observed for almost all population pairs. The increase in F_{ST} for outlier loci was much more dramatic in population pairs involving ALM, GAL or GER, with values of F_{ST} that were, on average, five times higher than F_{ST} values obtained with the neutral loci. The highest F_{ST} value based on outlier loci was found between populations ALM and GAL (F_{ST} = 0.712), located at opposite extremes of the climatic gradient (Table 4).

Variance components from analyses of molecular variance (AMOVAs) performed with neutral and outlier loci data sets are summarized in Table 5. The percentage of variance explained among groups was considerably increased when outlier loci were used instead of neutral loci, reaching 46.85% when populations were grouped according to the subspecies, whose geographical distribution also coincides with the major climatic regions.

Discussion

The AFLP genome scan in *Lacerta lepida* revealed several candidate loci under divergent selection as expected from the high morphological and genetic differentiation observed for *L. l. iberica* and *L. l. nevadensis* subspecies, located at the opposite extremes of a climatic gradient. Moreover, a significant proportion of those loci also revealed statistically significant associations with climatic variables along the gradient.

Outlier detection

The proportion of outliers obtained with DFDIST (4.1%) is close to the proportion of outliers detected with BayeScan with posterior probability above 0.99 (3.1%) and also similar to the 5–10% reported in the generality of AFLP genome scans employing DFDIST (Nosil *et al.* 2009). However, most of these studies are not directly comparable, because chosen confidence levels may vary (α = 0.05, Savolainen *et al.* 2006; α = 0.01, Jump *et al.* 2006; α = 0.0005, Herrera & Bazaga 2008) as well as the study design (global analysis, Herrera & Bazaga 2008; pairwise comparisons, Jump *et al.* 2006; or

Table 4 Pairwise population F_{ST} (lower diagonal for 318 neutral loci and upper diagonal for 23 outliers)

	GAL	GER	BEJ	SET	SPE	TOL	ALE	AND	CMA	ALM
GAL	—	0.114	0.334	0.384	0.440	0.508	0.306	0.349	0.409	0.712
GER	0.040	—	0.141	0.139	0.203	0.283	0.085	0.097	0.190	0.608
BEJ	0.057	0.021	—	0.056	0.044	0.173	0.076	0.074	0.089	0.578
SET	0.049	0.010	0.016	—	0.035	0.107	0.015	0.003	0.068	0.579
SPE	0.051	0.021	0.048	0.022	—	0.181	0.067	0.065	0.092	0.595
TOL	0.064	0.034	0.029	0.014	0.033	—	0.080	0.069	0.083	0.626
ALE	0.077	0.032	0.038	0.017	0.027	0.036	—	0.000	0.047	0.550
AND	0.078	0.030	0.042	0.025	0.034	0.032	0.007	—	0.046	0.539
CMA	0.051	0.041	0.032	0.026	0.037	0.025	0.051	0.048	—	0.592
ALM	0.163	0.106	0.124	0.104	0.112	0.127	0.107	0.112	0.132	—

Table 5 Analysis of molecular variance (AMOVA) of the neutral loci and outlier loci data sets according to three grouping criteria: single group, by subspecies and by mtDNA clade (see Table 1 for each population's subspecies and mtDNA clade)

		318 neutral loci			23 outlier loci		
		Variance components (%)			Variance components (%)		
Grouping criteria	Number of groups	Within populations	Among populations within groups	Among groups	Within populations	Among populations within groups	Among groups
Single group	1	91.51	8.49		59.41	40.59	
By subspecies	3	88.01	4.05	7.95	45.99	7.16	46.85
By mtDNA clade	6	91.09	4.04	4.87	57.34	5.87	36.8

© 2010 Blackwell Publishing Ltd

both, Bonin *et al.* 2006; see Nosil *et al.* 2009 for a review).

The same data set was used for both programs, but *P*-values from DFDIST and posterior probabilities obtained by BayeScan for the same loci are not directly comparable. Nevertheless, the correspondence between the most extreme outliers obtained with both programs is limited, probably reflecting the differences in their methodology. This behaviour is similar to the one evidenced by Beaumont & Balding (2004), who noted from the results of simulations that, while on average the power of the frequentist and Bayesian methods were similar, there was not necessarily a strong overlap in detected markers within simulated data sets (in which the markers under selection were known). The main difference between DFDIST and Bayescan is that the latter allows for variable within-population F_{ST}, whereas in the former it is assumed to be the same in all populations. This difference did not have a major impact in the simulations of Beaumont & Balding (2004), who simulated many-fold differences in F_{ST} among their populations, yet the overall power of both methods were quite similar. Both methods have a tendency to show false positives when there are correlations in allele frequencies among populations, because of shared recent ancestry or isolation by distance effects (Robertson 1975; Excoffier *et al.* 2009). Although Beaumont & Nichols (1996) detected no strong effect of isolation by distance or heterogeneous levels of gene flow between populations on outlier detection, the recent study by Excoffier *et al.*(2009) demonstrated some cases where neglecting population structure could lead to high rates of false positives. They implemented a new methodology for outlier detection where population structure can be taken into account for building the null distribution of F_{ST}, but the method is not yet available for dominant marker data.

In the present results, the most serious concerns about the effect of population structure in outlier detection are raised by the inclusion of the ALM population in global analyses. This population belongs to *L. l. nevadensis*, which started its divergence from the other ocellated lizards in the Iberian Peninsula long ago and has accumulated the highest levels of neutral divergence, as reported by cytochrome *b* mitochondrial gene analysis (Paulo 2001; Paulo *et al.* 2008). The exclusion of the ALM population from the global analysis did not have a major effect on results from BayeScan, but results from DFDIST were more severely affected (Table S1, Supporting information). On the other hand, ALM is also located at the SE extreme of the climatic gradient, and due to morphological differences observed in this subspecies, we expect that along with neutral differentiation, populations from the SE have also accumulated adaptive divergence by natural selection. Therefore, it is difficult to assess the extent to which the outliers detected as a result of the inclusion of the ALM population on global analysis are false positives.

Some studies have employed multiple pairwise comparisons among populations (see Nosil *et al.* 2009). Such comparisons are less susceptible to problems caused by unknown complexity in the true population structure and can strengthen evidence for candidate loci where independent comparisons can be made across the environmental transition (as in Wilding *et al.* 2001; for example). However, they are problematic when all pairs of samples are compared because comparisons are no longer independent and there is no way to correct for the large total number of comparisons. For the present data set, only five independent comparisons could be made. The proportion of outliers and the lists of AFLP markers detected when pairwise comparisons were made were very variable among methods and between global or pairwise analyses within the same detection method (Table S2, Supporting information).

Overall, DFDIST seems to be more sensitive than BayeScan to changes in the input and in stringency criteria. Caballero *et al.* (2008) also raised several concerns about the sensitivity of DFDIST and a recent simulation study by Pérez-Figueroa *et al.* (2010), comparing the efficiency of DFDIST, DETSELD and BayeScan to detect loci under directional selection with dominant markers, showed that BayeScan appears to perform more efficiently under a wide range of scenarios than the other methods. The results presented here show that using the same data set with different outlier detection methods can produce quite different results. If the purpose of a genome scan is to target candidate loci influenced by selection for further research, the most appropriate approach is probably the use of more than one method and to combine the outlier lists to avoid the risk of losing interesting candidates.

When applying SAM analysis to the AFLP data set, several loci detected as outliers by SAM in association with climatic variables were also highlighted as outliers by DFDIST or BayeScan. It is important to note that SAM is individual-centred, being independent of any notion of population and making no presumption as to the genotypic structure of the populations to which the sampled individuals belong (Joost *et al.* 2008). However, detection of significant associations by SAM depends on the variables tested. Therefore, the existence of associations with other environmental variables not tested here, such as soil colouration, or food availability, cannot be ruled out. To overcome this limitation, it seemed appropriate to set a list of

© 2010 Blackwell Publishing Ltd

outliers for future validation that includes loci detected by DFDIST or BayeScan but not by SAM. Therefore, 23 outlier loci (6% of the investigated loci) stand out as strong candidate loci, potentially under divergent selection.

SAM is a useful method to search for associations with many environmental variables simultaneously, providing some insights into which selective forces may be in play. Since many environmental variables are closely correlated with others, information on the ecological requirements of the species is particularly important to assess which variables have the potential to be the major selective forces. Moreover, the variables tested must be ecologically relevant for the species. Otherwise, any variable that ranges along the same axis as the genetic data will turn out to be statistically associated with it, without being necessarily a selective force. For instance, if dispersal routes coincide with the environmental gradient, associations of allele frequencies with environmental variables could result from isolation by distance or founding effect rather than selection. For ocellated lizards, the colonization routes from glacial refugia were predictably located along river basins and valleys, which are mostly oriented southwest–northeast in the Iberian Peninsula (nearly perpendicular to the climatic gradient axis), rather than across the mountain systems that shape the landscape. Other methods currently available for dominant data to find associations with environmental variables, like generalized estimating equations, can take spatial autocorrelation into account by assuming that genetic similarity is normally higher for neighbouring individuals than distant ones (Poncet *et al.* 2010).

Neutral differentiation and genetic structure

Neutral loci differentiation reflects the existence of population genetic structure as predicted by mitochondrial divergence data in Paulo (2001) and Paulo *et al.* (2008), but this structure is much less strongly supported by AFLP markers. Mitochondrial clades expected within the nominal subspecies' distribution, *L. l. lepida*, are barely recognizable with neutral AFLP markers. The same situation was reported for the nuclear gene β-fibrinogen by Paulo *et al.* (2008), who argued that differences from the mtDNA pattern could result from the higher dispersal rate of males and/or the retention of ancestral polymorphism.

Genetic differentiation with outlier loci was much higher than with neutral loci, in particular for populations located at the opposite ends of the climatic gradient, ALM (*L. l. nevadensis*) and GAL (*L. l. iberica*), following the expected pattern of loci under selection.

Association with environmental variables

Results from the SAM analysis give us some clues about the possible selective forces acting upon outlier loci along the environmental gradient in the Iberian Peninsula. However, associations detected here denote correlations between environmental variables and AFLP band's frequency at each locus, not necessarily a causal relationship between them.

The strongest associations are registered for maximum monthly temperatures (T_{max}) and insolation. It is noteworthy that most T_{max} monthly variables associated with outlier locus frequencies correspond to the months when *L. lepida*'s activity is higher, from April to October (Busack & Visnaw 1989). As lizards are diurnal and ectothermic, temperature and insolation are known to play a major role in *L. lepida*'s seasonal and daily activities. Adult *L. lepida* are known to be active at the surface when ambient temperatures range from 15.6 to 42 °C (Busack & Visnaw 1989). During colder months, between November and February, they remain inactive while in the remaining months, exposure to sunshine is required to heat their body for daily activity and food search. Moreover, insects constitute *L. lepida*'s main food resource (Busack & Visnaw 1989), and their availability also depends on temperature, as they are ectothermic as well.

In the Iberian Peninsula, while the south registers hot summers, temperatures in the NW, where the *L. l. iberica* subspecies can be found, register much milder maximal temperatures, even in summer, and with much less annual insolation than in the south. Clearly, populations of *L. l. iberica* subspecies face more adverse climatic conditions, and its smaller and darker body may be a selective response to make thermoregulation faster and more efficient. Busack & Visnaw (1989) reported that sub-adults of *L. lepida* displaying the darker juvenile coloration seem to be more tolerant to cold and less tolerant to heat than adults, displaying surface activity at ambient temperatures from 11.1 to 30 °C. Clusella-Trullas *et al.* (2007) cites several studies in lizards showing that in most cases melanization increases solar absorption under cool conditions and allows melanistic animals to reach their thermal optimum more rapidly than lighter ones. Moreover, comparative studies on *L. lepida* subspecies (Mateo & Castanet 1994) show that when developing under the same controlled conditions, smaller body size can still be observed for *L. l. iberica*, indicating that besides possible environmental constraints, it is also genetically determined.

Several associations with precipitation were also found, although this variable probably does not affect lizards' activity as directly as temperature or insolation. Nevertheless, the asymmetry in precipitation distribu-

© 2010 Blackwell Publishing Ltd

tion at the two extremes of the Iberian Peninsula can influence the reproductive cycle, the availability of food and strongly influences the vegetation cover (Hodar *et al.* 1996). Precipitation decreases from north to south, with a meridional asymmetry, leading to higher precipitation close to the Atlantic and lower close to the Mediterranean Sea (IGN 1992). In the NW, annual precipitation can reach values as high as 2000 mm. At the opposite extreme of the climatic gradient, in the SE, *L. l. nevadensis* subspecies survives with less than 300 mm of annual precipitation. The dry climate produces an arid landscape with scarce vegetation cover, where the proportion of bare ground can reach more than 40% (Hodar *et al.* 1996) and the distinct dorsal colour of *L. l. nevadensis* provides a good camouflage in a landscape where lizards are probably frequently exposed to predatory birds and mammals. Additionally, Mateo & Castanet (1994) proposed that irregular rainfall in the SE, and consequent irregular and limiting trophic resources, could explain the observed differences in *L. l. nevadensis* reproductive strategies, with annual asynchrony of clutches, an extended laying period and even the ability to lay more than one annual clutch, as opposed to the single annual clutch registered in the rest of *L. lepida*'s distribution.

Two loci were associated with relative humidity. Although the percentage of humidity in the air is normally higher in the NW than in the SE, locations in the south of the Iberian Peninsula, along the Atlantic coastline, can reach relative humidity values as high as in the NW. This is the case for SPE and CMA populations (Table 1), but morphological or life-history trait variation for these populations has not been extensively documented to date. The actual role of humidity in *L. lepida*'s local adaptation requires further investigation.

Future directions

This study highlighted a list of candidate loci suspected to be under the influence of selection in ocellated lizards, which are currently being isolated and sequenced. The characterization of outliers and the determination of their functional role, if any, is necessary to uncover their involvement in local adaptation of *L. lepida*, with particular interest for divergence of subspecies located at cline extremes. If outlier's sequence has no homology with any known gene, it may belong to an unknown regulatory region or simply a non-coding fragment that is in linkage with the actual target of selection. While in the first case, conducting further studies with the candidate gene under selection may be straightforward, in the two other cases, it implies the extension of the sequenced fragment to find the cause of marker polymorphism and map the outlier fragment in the genome. The latter situation requires the availability of more detailed genomic information on the species to proceed with the investigation, but the use of next generation sequencing offers the possibility to overcome this limitation in the near future (see review by Slate *et al.* 2010).

This study also highlights the usefulness of complementing genome scans for selection in natural populations with ecological data in order to identify selective pressures potentially acting upon candidate loci at a local scale. Public databases constitute a good source of environmental information at sample coordinates, but more detailed information, ecologically relevant for the species studied, needs to be collected in the field, such as vegetation cover, soil colouration, diversity and abundance of food sources or predation pressure. Combining ecological, phenotypic and genomic data will be very fruitful to test hypotheses regarding adaptation and speciation.

Acknowledgements

We thank A. Bonin and five anonymous referees for their helpful comments on the manuscript. This work was supported by a doctoral scholarship from FCT (SFRH/BD/21306/2005) to Nunes VL and the FCT project POCI/BIA/59288/2004- ''Process of speciation: accessing the genic view.'' coordinated by Paulo OS. All samples were collected under appropriate licenses in each country.

References

APA (1975) Atlas do Ambiente Digital. Agência Portuguesa do Ambiente (Portugal).

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **263**, 1619–1626.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, **57**, 289–300.

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, **14**, 2899–2914.

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441–469.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for

© 2010 Blackwell Publishing Ltd

molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Busack S, Visnaw JA (1989) Observations on the natural history of *Lacerta lepida* in Cádiz province, Spain. *Amphibia-Reptilia*, **10**, 201–213.

Caballero A, Quesada H, Rolan-Alvarez E (2008) Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, **179**, 539–554.

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Castilla AM, Bauwens D (1989) Reproductive characteristics of the lacertid lizard *Lacerta lepida*. *Amphibia-Reptilia*, **10**, 445–452.

Chiurugwi T, Beaumont MA, Wilkinson MJ, Battey NH (2010) Adaptive Divergence and speciation among sexual and pseudoviviparous populations of *Festuca*. *Heredity*, In Press.

Clusella-Trullas S, van Wyk JH, Spotila JR (2007) Thermal melanism in ectotherms. *Journal of Thermal Biology*, **32**, 235–245.

Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution*, **62**, 1162–1181.

Ellegren H (2008) Comparative genomics and the study of evolution by natural selection. *Molecular Ecology*, **17**, 4586–4596.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Herediy*, **103**, 285–298.

Foll M, 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.

Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princenton University Press, Princeton, NJ

Herrera CM, Bazaga P (2008) Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. *Molecular Ecology*, **17**, 5378–5390.

Hodar JA, Campos F, Rosales BA (1996) Trophic ecology of the Ocellated Lizard *Lacerta lepida* in an arid zone of southern Spain: relationships with availability and daily activity of prey. *Journal of Arid Environments*, **33**, 95–107.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101–104.

IGN (1992) Atlas Nacional de España, Sección II: El medio terrestre—Grupo 9—Climatología, 1ª edicion. Instituto Geografico Nacional (Spain).

Jensen JD, Wong A, Aquadro CF (2007) Approaches for identifying targets of positive selection. *Trends in Genetics*, **23**, 568–577.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.

Joost S, Kalbermatten M, Bonin A (2008) Spatial analysis method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources*, **8**, 957–960.

Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology*, **15**, 3469–3480.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Mateo JA (1988) *Estudio Sistemático y Zoogeográfico de los Lagartos Ocelados, Lacerta lepida (Daudin, 1802) y Lacerta pater (Lataste 1880) (Sauria, Lacertidae)*. University of Sevilla, Sevilla.

Mateo JA, Castanet J (1994) Reproductive strategies in 3 Spanish populations of the ocellated lizard, *Lacerta Lepida* (Sauria, Lacertidae). *Acta Oecologica*, **15**, 215–229.

Mateo JA, Castroviejo J (1990) Variation morphologique et révision taxonomique de l'espèce *Lacerta lepida* Daudin, 1802 (Sauria, Lacertidae). *Bulletin du Muséum National D'Histoire Naturelle, Paris*, **12**, 691–706.

Mateo JA, López-Jurado LF (1994) Variaciones en el color de los lagartos ocelados; aproximación a la distribución de *Lacerta lepida nevadensis* Buchholz 1963. *Revista Española de Herpetología*, **8**, 29–35.

Mateo JA, López-Jurado LF, Guillaume CP (1996) Proteic and morphological variations in ocellated lizards (Lacertidae): a complex of species across the Strait of Gibraltar. *Comptes Rendus de l' Academie des Sciences Serie III-Sciences de la vie-Life Sciences*, **319**, 737–746.

Mateo JA, López-Jurado LF, Cano J, Guillaume CP (1999) Intraspecific variability in nucleolar organiser position in *Lacerta (Timon) lepida*. *Amphibia-Reptilia*, **20**, 73–77.

Mayer W, Bischoff W (1996) Beiträge zur taxonomischen Revision der Gattung *Lacerta* (Reptilia: Lacertidae) Teil 1: Zootoca, Omanosaura, Timon und Teira als eigenständige Gattungen. *Salamandra*, **32**, 163–170.

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Molecular Ecology*, **17**, 1552–1563.

Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Molecular Ecology*, **15**, 4229–4242.

Ninyerola M, Pons X, Roure JM (2005) *Atlas Climático Digital de la Península Ibérica. Metodología y aplicaciones en bioclimatología y geobotánica*. ISBN 932860-8-7, Universidad Autónoma de Barcelona, Bellaterra.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: ''Isolation by adaptation'' and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

© 2010 Blackwell Publishing Ltd

Paulo OS (2001). *The phylogeography of reptiles of the Iberian Peninsula*. University of London, London.

Paulo OS, Pinheiro J, Miraldo A *et al*. (2008) The role of vicariance vs. dispersal in shaping genetic patterns in ocellated lizard species in the western Mediterranean. *Molecular Ecology*, **17**, 1535–1551.

Pérez-Figueroa A, García-Pereira M, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.

Poncet B, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology*, **19**, 2896–2907.

Robertson A (1975) Gene frequency distribution as a test of selective neutrality. *Genetics*, **81**, 775–785.

Rosenblum EB, Hoekstra HE, Nachman MW (2004) Adaptive reptile color variation and the evolution of the *Mc1r* gene. *Evolution*, **58**, 1794–1808.

Savolainen V, Anstett MC, Lexer C *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature*, **441**, 210–213.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Slate J, Santure AW, Feulner PGD *et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, **26**, 275–284.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.

Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.

Vekemans X (2002) *AFLP-SURV version 1.0. Distributed by the author*. Laboratoire de Génetique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9939–9946.

Vos P, Hogers R, Bleeker M *et al*. (1995) AFLP—a new technique for DNA-fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123–3135.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

VLN, RKB and OSP are interested in using population genomics approaches to gain insights into the genetic basis of adaptive speciation. MAB is interested in understanding how genetic data can be used both to study adaptation and selection and also to elucidate the demographic history of populations.

Supporting information

Additional supporting information may be found in the online version of this article:

Table S1 Comparison of results from outlier detection by DFDIST and BayeScan with and without ALM population (*L. lepida nevadensis)*

Table S2 Results from outlier detection with DFDIST and BayeScan in five independent pairwise analyses along a climatic gradient in the Iberian Peninsula

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2010 Blackwell Publishing Ltd

CHAPTER 2

Supporting information

Table S1 Comparison of results from outlier detection by DFDIST and BayeScan with and without ALM population (*L. lepida nevadensis)*.For DFDIST analyses, *P*-values significant with 5% FDR in DFDIST are highlighted in bold. (*) denotes loci not considered in DFDIST analysis because the frequency of the most common allele (band presence or band absence) was above 0.98, the critical frequency threshold set for loci exclusion in DFDIST. For BayeScan analyses, loci with posterior probability above 0.99 are highlighted in bold.

	DFDIST global analysis *P*-value			BayeScan global analysis Posterior probability	
Locus	With ALM	Without ALM	Locus	With ALM	Without ALM
30	**0.000**	0.063	170	0.955	**0.990**
140	**0.000**	0.353	34	0.966	**0.992**
192	0.025	**0.000**	320	0.975	**0.995**
201	**0.001**	(*)	351	**0.991**	0.974
209	**0.000**	(*)	323	**0.992**	**0.998**
223	**0.000**	0.041	220	**0.998**	0.983
228	**0.000**	0.680	235	**0.998**	**0.991**
245	**0.001**	**0.000**	311	**0.999**	**1.000**
291	**0.002**	0.206	75	**0.999**	**0.998**
297	**0.000**	(*)	353	**0.999**	**0.996**
301	**0.000**	0.936	390	**0.999**	**1.000**
304	0.014	**0.000**	52	**1.000**	**1.000**
311	**0.001**	**0.000**	245	**1.000**	**1.000**
315	**0.000**	**0.000**	315	**1.000**	**1.000**
340	**0.000**	0.538	386	**1.000**	**0.996**
347	**0.000**	(*)			
386	**0.000**	0.004			
390	**0.002**	0.002			

Table S2 Results from outlier detection with DFDIST and BayeScan in five independent pairwise analyses along a climatic gradient in the Iberian Peninsula (population codes as in Table 1). Outliers that remain significant with a 5% FDR in DFDIST are denoted with an asterisk. Among loci detected in pairwise analyses in DFDIST with *P*-value <0.01, only loci 245, 315, 386 and 390 were also detected by DFDIST global analysis with 10 populations. Among loci detected by BayeScan pairwise analyses, loci 52, 220, 235, 315 and 323 were also detected by BayeScan global analysis (with posterior probability above 0.99).

	DFDIST		BayeScan	
Pairwise comparison	Locus	*P*-value <0.01	Locus	Posterior probability >0.90
GAL-BEJ	34	0.006	52	0.988
	52	0.006	220	0.969
	170	0.005	231	0.986
	231	0.000 *	235	0.907
	235	0.001	250	0.913
	250	0.000	315	0.990
	304	0.002	320	0.952
	314	0.004	323	0.969
	315	0.001	349	0.930
	320	0.002		
	326	0.005		
	349	0.004		
GER-SPE	75	0.003	__	__
	240	0.003		
	245	0.002		
	351	0.003		
	386	0.003		
	390	0.004		
TOL - SET	75	0.000 *	__	__
	245	0.003		
	351	0.004		
	353	0.004		
	375	0.000 *		
ALE - CMA	214	0.007	__	__
	315	0.007		
ALM - AND	44	0.008	37	0.944
			209	0.944

CHAPTER 3

Challenges and pitfalls in the isolation and characterization of anonymous AFLP markers in non-model species: lessons from an ocellated lizard genome scan

Vera L. Nunes[1], Mark A. Beaumont[2,3], Roger K. Butlin[4], Octávio S. Paulo[1]

[1] Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, DBA/FCUL, 1749-016 Lisboa, Portugal

[2] School of Animal and Microbial Sciences, University of Reading, P.O. Box 228, Reading, RG6 6AJ, United Kingdom

[3] Schools of Mathematics and Biological Sciences, University of Bristol, UK

[4] Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, England

In prep.

Abstract

In the last few years dozens of studies have documented the detection of loci influenced by selection from genome scans in a wide range of non-model species. Many of those studies used AFLP markers, which became popular for being easily applicable to any organism. However, because they are anonymous markers, AFLP impose many challenges for their isolation and identification. Most recent AFLP genome scans used capillary electrophoresis, which adds even more obstacles to the isolation of bands with a specific size for sequencing. These caveats may justify the extremely low number of studies that moved from the detection of outlier AFLP markers to their actual isolation and characterization. We document our efforts to characterize a set of outlier AFLP markers from a previous genome scan with capillary electrophoresis in ocellated lizards (*Lacerta lepida*). Seven AFLP outliers were successfully isolated, cloned and sequenced. Their sequences are noncoding and show internal indels or polymorphic repetitive elements (microsatellites). Three outliers were converted into co-dominant markers by using specific internal primers to sequence and screen population variability from undigested DNA. Amplification in closely related lizard species was also achieved, revealing remarkable interspecific conservation in outlier loci sequences. We stress the importance of following up AFLP genome scans to validate selection signatures of outlier loci, but also report the main challenges and pitfalls that may be faced during the process.

Keywords: AFLP sequence, capillary electrophoresis, dominant markers, outlier's validation, selection

Introduction

The use of population genomic approaches to search for genomic regions potentially under selection has gained much popularity in the last decade. Amplified fragment length polymorphism (AFLP) markers have been used frequently for genome scans in several non-model species (e.g. Bonin *et al.* 2006; Minder & Widmer 2008; Apple *et al.* 2010; Croucher *et al.* 2011). First described by Vos *et al.* (1995), the AFLP technique consists of the digestion of genomic DNA with restriction enzymes, the ligation of adaptors to the digested fragments, and the amplification by PCR of these fragments using selective primers that anchor in the adaptors. Hundreds or even thousands of polymorphic AFLP markers, distributed across the whole genome, can be easily and affordably genotyped for many populations of any species, but the sequence content of each AFLP marker remains unknown throughout the whole process. Despite the recent revolution in sequence technology that makes the cost of complete genome sequences affordable, AFLPs will probably remain as the molecular marker of choice for population genomics in non-model species where reference genomes are still not available (Gaggiotti 2010; Stapley *et al.* 2010). RAD sequencing is one of the newly emergent sequencing technologies and was recently applied in a high-density SNP-based genome scan, suggesting that this genotype-by-sequencing approach might replace AFLP scans in the future, overcoming marker anonymity (Hohenlohe *et al.* 2010; Rowe *et al.* 2011).

Recent efforts have been made to improve the reliability of existing methods for statistical detection of outlier loci in AFLP genome scans, by controlling factors that inflate the false positive rate, such as homoplasy, population structure and history or multiple comparisons (Caballero *et al.* 2008; Excoffier *et al.* 2009; Pérez-Figueroa *et al.* 2010). However, the profusion of methodologies employed among genome scan studies to detect outliers and the use of variable criteria for outlier classification (use of one or several methods simultaneously; variable significance thresholds; population pairwise comparisons versus global analyses) makes it difficult to compare results from different taxa (Butlin 2010). We could learn much more from such studies if AFLP markers detected as outliers were brought out from anonymity and further research were conducted towards the identification of genes linked with such outliers and their implications for local adaptation. Unfortunately, and despite the more than 25 AFLP genome scans for selection in non-model species available in the literature, only two

studies report their attempts to isolate and sequence AFLP markers identified as candidate loci under selection (Minder & Widmer 2008; Wood *et al.* 2008). While the generation of hundreds of AFLP markers with a single primer pair is technically straightforward, the isolation of a particular AFLP fragment can be technically demanding and time consuming, often involving the need for fragment cloning. The use of capillary electrophoresis (CE) has become popular for separating fluorescently labelled AFLP fragments, with gains in both resolution (fragments migrating with a difference in size of one single base can be accurately distinguished and scored) and sensitivity (even fragments amplified with lower efficiency can be easily visualized) as compared to traditional 6% polyacrylamide gels with silver staining (Polanco *et al.* 2005; Apple *et al.* 2010). Nevertheless, CE adds extra difficulties to the isolation of AFLP fragments because they can no longer be directly excised from the denaturing matrix. These drawbacks may help to explain why most AFLP markers identified as candidate loci potentially under selection in genome scans still remain completely anonymous.

Polanco *et al.* (2005) proposed a method to isolate AFLP fragments by CE. The method is not particularly practical and implies that amplified fragments previously separated in the automated sequencer must be re-run and monitored to interrupt the migration at the moment where the desired fragment is detected through its fluorescence emission. The capillary is then removed from the machine and broken with precision at the detection window where the fragment is supposed to be. From there, the fragment can be re-amplified and cloned by standard procedures. The alternative is to use gel matrices to separate the fragments previously genotyped by CE and excise the desired band from the gel. This was the approach chosen by Minder & Widmer (2008), using high resolution gels (Spreadex, Elchrom) to isolate outlier markers previously scored by CE.

Despite the troublesome nature of the AFLP marker's isolation and identification, their importance to gain insights into the genomic regions responsible for adaptive evolution in non-model organisms should not be neglected. The AFLP technique seems to be prone to generate polymorphic fragments within intergenic regions, thus limiting its power to detect structural mutations in functional genes. Even when an AFLP marker matches with a coding region in a non-model species, there is a good chance that the gene is not characterized and annotated yet in other taxa. Thus, genome scans can provide new candidate genes or regulatory elements with importance in adaptation that were unsuspected before. That was the case reported in Wood *et al.* (2008), who found no

differentiation in flanking regions of sequenced outliers, indicating that indel polymorphisms detected within outlier sequences (with characteristics of transposable elements) could be the actual targets of selection, perhaps affecting the expression of downstream loci.

Here we report the results and main difficulties faced when trying to isolate and characterize a set of outliers resulting from a previous AFLP genome scan conducted by Nunes *et al.* (2011) in the European ocellated lizard (*Lacerta lepida*). The species is widespread throughout the Iberian Peninsula in a variety of ecological conditions that are strongly influenced by the distribution of precipitation and temperature ranges. Morphological and genetic differentiation in populations from the northwest and southeast of the Iberian Peninsula are strong enough to consider those populations as distinct subspecies: *Lepida lepida iberica* and *Lepida lepida nevadensis*, respectively (Mateo & Castroviejo 1990; Paulo *et al.* 2008). The first lives in a rainy and less warm weather regime while the second inhabits a region that experiences hot summers and the lowest annual rainfall across the species' distribution range. Detection of selection with DFDIST (Beaumont & Nichols 1996) and BayeScan (Foll & Gaggiotti 2008) produced a combined list of 23 outliers (5.9% of investigated loci) targeted for further validation. Nunes *et al.* (2011) also tested for associations between AFLP band frequency and variation in climatic variables across the Iberian Peninsula with the spatial analysis method (SAM) (Joost *et al.* 2007). Several loci detected as outliers were also associated with temperature, insolation or precipitation. The present study reports our efforts to characterize and validate a subset of 12 outliers out of the 23 outliers highlighted by our AFLP genome scan, which includes the five AFLP markers with the most extreme outlier behaviour, detected by both DFDIST and BayeScan (Nunes *et al.* 2011).

Materials and Methods

Isolation and cloning of outlier AFLP markers

Twelve AFLP markers considered as candidate loci potentially under selection were chosen for isolation (see outlier list in Table 1 and the corresponding combinations of selective primers), including the five outliers detected by both DFDIST and BayeScan (see Nunes *et al.* 2011). Samples for which the band corresponding to the target outlier

was previously scored as present were re-amplified by PCR from digested DNA using the same conditions as in Nunes *et al.* (2011), but using *Eco*RI selective primer without fluorescent label (to avoid interference in downstream steps) and a Green GoTaq® Flexi PCR buffer (Promega) for direct loading of PCR products into agarose gels. For each sample, three PCR replicates (10 µL x 3) were loaded together in the same lane of a 1.5% agarose gel stained with ethidium bromide. Because PCR generates many fragments with similar size and agarose gel resolution is insufficient to isolate a single band, three contiguous slices of gel were excised within a size range of 50-100 bp that included the desired outlier size. Each gel slice was purified separately with GENECLEAN®II kit (MP Biomedicals) to recover the DNA fragments. To confirm the recovery of the desired AFLP marker, gel purified fragments were used as template for a PCR with fluorescent labelled primers as in Nunes *et al.* (2010), and PCR products were then separated by CE on an ABI Prism 310 (Applied Biosystems). After confirming the amplification of the target outlier from the gel purified fragments, they were cloned with TOPO TA Cloning ® Kit (Invitrogen), following the manufacturer's instructions. Single colonies were randomly selected to construct clone libraries.

Library screening

Because each cloning reaction was expected to include multiple fragments of similar size to the outlier fragment, a quick but efficient library screening procedure was needed to identify clones bearing inserts with the size of the desired outlier, dramatically reducing the number of clones to be sequenced. Therefore, each colony from a library was amplified by PCR with universal primers M13 in a total reaction volume of 15 µL. The amplified clones were readily used as template for another PCR with fluorescent labeled selective primers using the same conditions as Nunes *et al.* (2011), but scaled to a final volume of 5 µL. The amplified inserts from individual colonies were pooled together in sets of 12 and separated by CE. If fragments with the expected size were present within a pool of inserts, the respective clones were run separately to identify which of the 12 clones was bearing the insert of the expected size. This way, only clones with inserts of the desired size (confirmed by CE) were sequenced with M13 primers, using standard protocols (BigDye Terminator v.3.1, Applied Biosystems) on an ABI

PRISM 310 (Applied Biosystems). Sequences were edited in Sequencher v.4.0.5 (Gene Codes Co.) and deposited in GenBank (see Table 1 for accession numbers).

Outlier sequence characterization

Cloned sequences were aligned with sequences of *Eco*RI and *Mse*I selective primers to check for mismatches in selective bases. GenBank was searched for sequences homologous to each clone insert sequence using BLASTN. Sequences were also inspected for the presence of open reading frames (ORF) that could indicate that the sequence might correspond totally or partially to a coding region. Since most outlier sequences were rich in repetitive elements, alignments with each other were tried to rule out the possibility that they would belong to the same locus, although varying in length.

An internal primer pair for each sequenced outlier was designed as close to the sequence ends as possible using Primer 3 (Rozen & Skaletsky 2000). The Reddy *et al.* (2008) method for genome-walking was employed to extend outlier fragment sequences into their flanking regions but all attempts failed. To investigate sources of polymorphism between the dominant allele (fragment scored as present) and the recessive allele (scored as absent), we combined unlabeled *Eco*RI or *Mse*I selective primer with the complementary outlier-specific primer in two independent amplifications in an attempt to obtain the full sequence from recessive alleles. Digested DNA from samples where the outlier was scored as absent (homozygous for the recessive allele) was used for PCR with 1x PCR buffer (Promega), 0.75 U GoTaq® DNA polymerase (Promega), 2.0 mM $MgCl_2$, 0.12 mM dNTPs and 0.4 µM of each primer in a final volume of 15 µL. The cycling conditions used were 3 min at 94 ºC, 35x (30 s at 94 ºC, 30 s at outlier specific primers annealing temperature (Table 1), and 30s at 72 º C) followed by 10 min at 72 ºC. Purified products (Sureclean, Bioline) were sequenced in both directions using standard protocols (BigDye Terminator v.3.1, Applied Biosystems) on an ABI PRISM 310 (Applied Biosystems).

Internal primer pairs designed for each outlier were tested in undigested DNA from samples of each *Lacerta lepida* subspecies (*L. l. nevadensis*, *L. l. lepida* and *L. l. iberica*), previously genotyped in Nunes *et al.* (2011), to evaluate primer efficiency, to characterize outliers as co-dominant markers and to corroborate AFLP genotypes. Sequences from

dominant alleles are expected to be conserved within the same species, because the presence of multiple mutations or indels in the dominant allele would affect the migration rate of the fragment in the electrophoresis profile and, consequently, the fragment would no longer be scored as the same AFLP marker. The opposite is true for recessive alleles because any fragment different enough to migrate faster or slower than the dominant allele will be scored as absent for the AFLP marker in question. This means that several recessive allele haplotypes (differing both in length and in nucleotide content) could be found in the same species or population. Additionally, and because AFLP fragments are dominant markers, all individuals for which an AFLP marker was scored as present, must at least carry one copy of the dominant allele, but the second allele is unknown and it may correspond to a second copy of the dominant allele or to any possible haplotype for the recessive allele. Therefore, we sequenced several samples from each *L. lepida* subspecies previously genotyped for the outlier AFLP markers to investigate the intraspecific variation in length and nucleotide composition of their sequences.

For outliers whose internal primers worked properly on undigested genomic DNA, cross-species amplification was tested in DNA samples from African ocellated lizard species, *L. pater* and *L. tangitana*, collected in Tunisia (Tabarka) and Morocco (Azrou), respectively (see collection details in Paulo *et al.* 2008). Cross-species amplification was also tested in two other Iberian lizard species (one *L. schreiberi* sample from Paulo *et al.* (2008) and one *Iberolacerta monticola* sample from Moreira *et al.* (2007)) and another European lizard (one *L. agilis* sample from Paulo *et al.* (2008)). PCR reactions and sequencing were performed as above.

Sequences were edited in Sequencher v.4.0.5 (Gene Codes Co.). Sequences of each allele from samples that were heterozygous in length were reconstructed according to guidelines from Flot *et al*. (2006). Base ambiguities were resolved with PHASE 2.1.1 (Stephens *et al.* 2001; Stephens & Scheet 2005). We ran the algorithm five times (1000 iterations with the default values) with different random number seeds, and the same haplotypes were consistently recovered in each run. Phased alleles from each individual were aligned with CLUSTAL W (Thompson *et al.* 1994) as implemented in Bioedit (Hall 1999) and gap length for repetitive element alignment was adjusted manually. Sequences from haplotypes detected in each lizard species and each outlier AFLP marker were deposited in GenBank (accession numbers JQ310676-JQ310742).

Nucleotide diversity (π) and haplotype diversity (H) for each outlier were determined for each ocellated lizard species and subspecies in ARLEQUIN 3.5 (Excoffier *et al.* 2005). Neutrality was tested with Tajima's *D* test (Tajima 1989) for each ocellated lizard species or subspecies. To infer the relationships among haplotypes, a minimum spanning network was constructed for each outlier marker with the Median Joining method (Bandelt *et al.* 1999) in NETWORK 4.51 (www.fluxus-engineering.com). The input file was converted from fasta to nexus format with CONCATENATOR 1.1.0 (Pina-Martins & Paulo 2008).

Results

AFLP marker isolation and sequence

All 12 outlier AFLP markers were successfully isolated from agarose gel slices, re-amplified and cloned. Clone libraries were screened by CE (Fig.1) and clones with the expected size were retrieved for only seven (58%) of the AFLP outliers (Table 1). Inserts in clone sequences included the full AFLP fragment flanked by *Eco*RI and *Mse*I adaptors sequence. In no case was a mismatch detected in the *Eco*RI or *Mse*I primer selective bases. Sequences of outliers did not align with each other, which indicate that all of them belong to independent loci. After cutting the adaptors out of the outlier sequences, the inserts were blasted against the GenBank database. Among the seven outliers sequenced, only three returned significant hits (Table 1). Their sequences were homologous with the green anole (*Anolis carolinensis*) or with the Indian python (*Python molurus*) whole genome shotgun sequences, but these species have no known genes annotated around the homologue of the outlier sequence. Because *A. carolinensis* and *P. molurus* are distantly related with ocellated lizards, possible inferences on the significance of the homologies detected here are very limited.

No open reading frames could be detected in outlier fragment sequences. Their sequences are likely to be non-coding regions and some are quite rich in repetitive elements. A specific primer pair was designed for each outlier based on clone sequences. These primers were used to amplify the fragments directly from undigested genomic DNA. The amplification was successful for five loci, *mk75*, *mk209*, *mk245*, *mk386* and

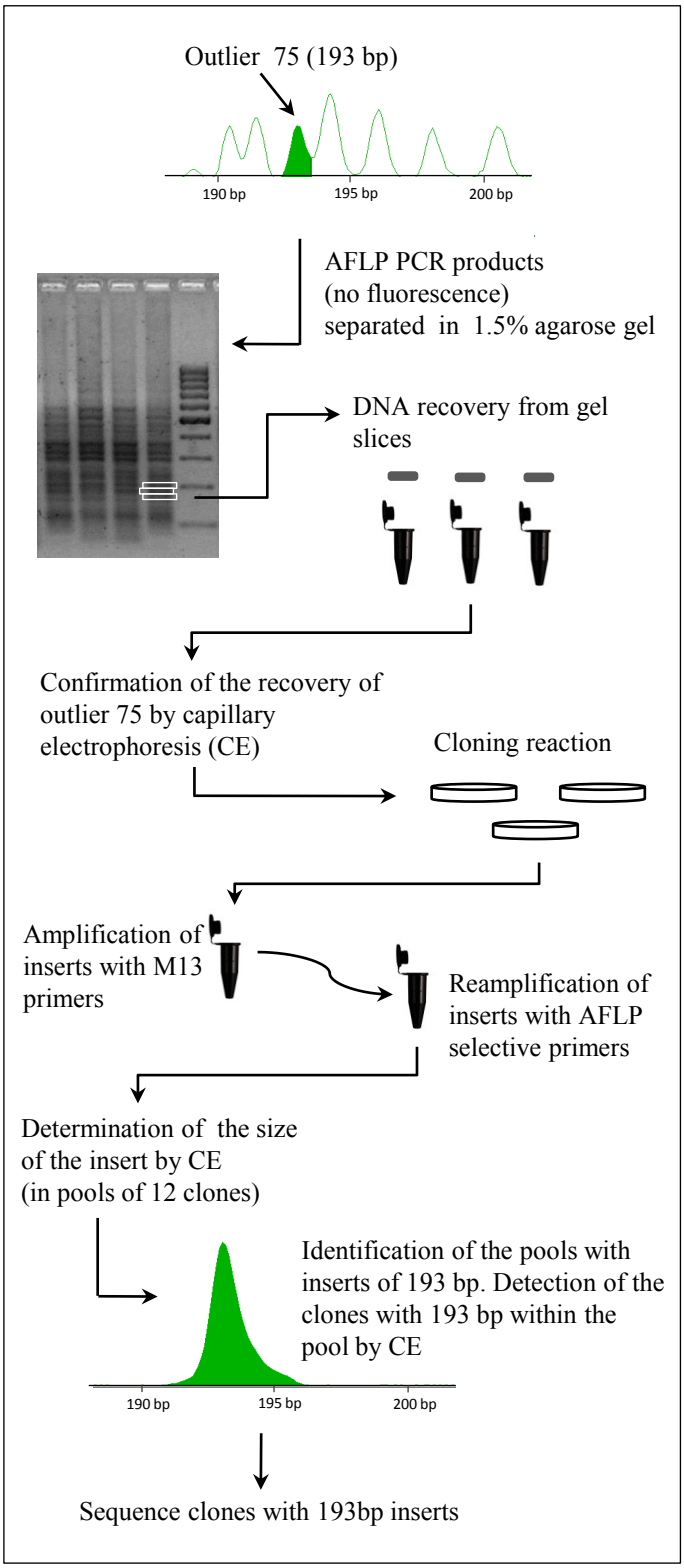


Fig. 1- Schematic representation of the steps used to isolate and sequence outlier AFLP marker 75, with a band size of 193 base pairs.

Table 1 Outlier loci isolated for sequence validation. For successfully cloned and sequenced outliers, sequences were blasted against GeneBank and internal primers pairs were designed. Temperature of annealing used for PCR amplification (Ta) is indicated for each primer pair.

Outlier	Size (bp)	Primer combination	Accession n.	Primer sequence 5'--> 3'	Ta (ºC)	Best BLASTN hits
75	193	AAC-CAC	JQ268560	mk75L1- AACAAGTAATACAAGCTCCAATGTG	58	No relevant hits
				mk75R1- GACACCATGGATAGTCTTTGC		
140	244	ACT-CTG	JQ268561			Score 3e-9, coverage 49% , *Anolis carolinensis*, WGS, AAWZ02004274
209	303	AAC-CTC	JQ268562	mk209L1- GCAACTGCTGTTCTTGCAC	60	Score 3e-15, coverage 24%, *Python molurus*, WGS, AEQU010237015
				mk209R1- GTCAACTGCGGCACGTAAG		
235	193	ACT-CTT				
245	250	ACT-CTT	JQ268563	mk245L1- TTTGAAGGAGTGGGAGAAGG	58	No relevant hits
				mk245R1- GCTGCAGGTCCTCTGTAAGAC		
297	278	AAG-CAC				
301	293	AAG-CAC				
311	95	ACA-ACA	JQ268564			Score 2e-9, coverage 70%, *Anolis carolinensis*, WGS, AAWZ02012507
315	111	ACA-ACA				
323	178	ACA-ACA				
386	207	AAG-CTA	JQ268565	mk386L1- TTGTAACAGATGGAGAACTGAGG	56	No relevant hits
				mk386R1- GATGACCCCGAGAAATATGC		
390	231	AAG-CTA	JQ268566	mk390L1*- ACATGCAGTTTACATTCTTTGC	53	No relevant hits
				mk390R1- ACATAATGTTATTTGGGTTACTTGC		

(*) only anneals with the dominant allele, while the mismatch of some bases with the recessive allele prevents its successful amplification.

mk390, but failed for *mk140* and *mk311*. In the case of *mk311*, the length of the clone insert was too small and it was not possible to find regions suitable for a primer pair design. An alternative approach was attempted by designing complementary primers for each strand, anchored in the only suitable region for primer design. The idea was to use a combination of *Eco*RI or *Mse*I selective primer with the specific primers for antisense amplification, but it was not possible to find appropriate conditions for successful amplification.

Because sources of length polymorphism in outlier markers can result from internal indels or mutations at one of the enzymes' restriction site or selective bases, it is important to obtain the full length of the recessive allele sequence to compare with the dominant allele sequence. Full length sequences from recessive alleles were successfully obtained from digested DNA (combining AFLP selective primers with outlier specific primers) for outlier *mk209*, *mk245* and *mk386* (Fig. S1, Supporting information). In *mk209*, an indel of five bases preceded by three consecutive single nucleotide polymorphisms (SNPs) explained the length polymorphism, while for *mk245*, the dominant allele carried a microsatellite composed by six GTT repeats but the recessive allele had only three GTT repeats. For both *mk209* and *mk245*, sources of polymorphism were located within the segment amplified with the internal specific primer pairs.

The amplification of *mk386* with internal specific primers resulted in a single SNP differing between dominant and recessive alleles. A single base replacement is normally not enough to cause detectable changes in CE migration rate and therefore other sources of length polymorphism should be present. The amplification of the full length sequence from the recessive allele revealed a deletion of three bases just before the binding site of the first specific primer which probably accounts for the length polymorphism.

All attempts to isolate the full sequence of the *mk75* recessive allele failed, which indicates that mutations might be present in AFLP primer selective bases or in the *Eco*RI or *Mse*I restriction sites. Nevertheless, length polymorphism between dominant and recessive alleles of *mk75* can be explained by an insertion of nine base pairs in the recessive allele, located within the fragment amplified with *mk75* internal primers (Fig. S1, Supporting information). As for *mk390*, amplification of the sequence end next to the *Mse*I adaptor failed, suggesting mutations in the recessive allele that prevent the correct annealing of the *Mse*I selective primer. Nevertheless, eight SNPs and a single-base deletion were detected between the dominant and recessive *mk390* allele sequences.

Because the single-base deletion and three of these SNPs are located right in the mk390L1 primer binding site, only dominant alleles can be amplified with the *mk390* internal specific primer pair (Fig. S1, Supporting information). For both *mk386* and *mk390*, there were no alternative binding sites to design internal primers suitable to amplify both dominant and recessive alleles from genomic DNA. However, the *mk390* locus was amplified and sequenced in African ocellated lizards with the *mk390* internal primer pair, which indicates that at least the dominant allele is present and conserved in *L. pater* and *L. tangitana* (accession numbers JQ310741-JQ310742).

Intra and interspecific variation in outlier sequences

Several samples from each *L. lepida* subspecies (*L. l. nevadensis*, *L. l. lepida* and *L. l. iberica*) previously genotyped with AFLP markers were sequenced for outliers *mk75*, *mk209* and *mk245*. No discordances between band score genotypes and sequences obtained for *mk75* were detected. This means that sequences from samples where the band *mk75* was scored as absent were carrying two recessive alleles as expected, while samples where the band was scored as present had either two dominant alleles or one dominant allele together with a recessive allele. Although only a small sub-sample of individuals with *mk75* scored as present were sequenced, homozygote individuals were only detected in *L. l. iberica* populations, while all samples from other populations were heterozygous, carrying the expected dominant allele sequence, but also a recessive allele. Marker *mk75* band frequency recorded in the *L. lepida* AFLP genome scan increased from southern to north-western populations of the Iberian Peninsula, especially in *L. l. iberica* populations (Fig. 2). The results from *mk75* sequences seem to indicate that the probability for a sample to be homozygous for the dominant allele is also higher in north-western populations, as expected.

Amplification of *mk75* in African ocellated lizards (*L. pater* and *L. tangitana*) revealed the absence of the dominant allele, except for one sample from Morocco that was heterozygous with one dominant allele and the most frequent recessive allele in *L. lepida* (Fig. 3 A). The same specific primers were able to amplify the *mk75* fragment in *L. schreiberi*, *L. agilis* and in *Iberolacerta monticola*, retrieving remarkably conserved sequences. A total of 10 recessive allele haplotypes were detected in ocellated lizards,

differing in single mutations from each other, while only a single *mk75* dominant allele haplotype could be found, that seems to be derived from a single deletion event of 9 bp.

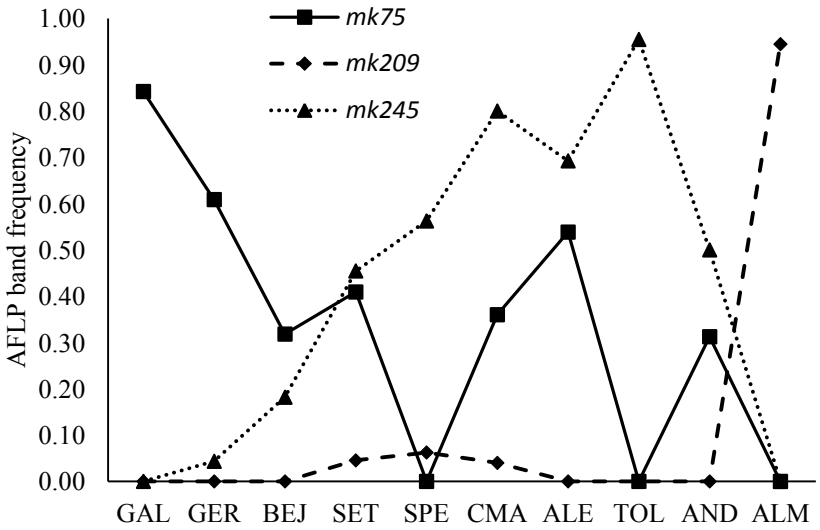


Fig. 2 Frequency of band presence observed in each *Lacerta lepida* population for outlier AFLP loci *mk75*, *mk209* and *mk245*. GAL and GER populations belong to *L. l. iberica*, ALM belongs to *L. l. nevadensis* and the remaining populations belong to the nominal subspecies, *L. l. lepida*.

Sequencing of *L. lepida* samples for *mk209* resulted in two dominant allele haplotypes, differing in one mutation (Fig. 3 B). Dominant alleles were exclusive to *L. l. nevadensis* samples. Outlier *mk209* was scored as present in genome scan genotyping at a high frequency in the ALM population (*L. l. nevadensis*) but was nearly or completely absent in all other populations (Fig. 2). We sequenced the three samples from outside the ALM population where the outlier mk209 was scored as present but none of them had any copy of a dominant allele sequence. This observation indicates that homoplasic fragments must have been responsible for the erroneous positive score in these samples. Five recessive allele haplotypes were found in *L. lepida* and they result mostly from length variation in a repetitive element of Gs that follows the indel responsible for *mk209* polymorphism (Table 2; Fig. S2, Supporting information). The homologous *mk209* fragments obtained in African ocellated lizards were more variable in length of repetitive elements, resulting in 11 haplotypes that were detected only in Africa (Table 2). All samples from *L. pater* and some from *L. tangitana* share the insert TGGA with *L. l.*

Table 2 Sequence diversity measures for ocellated lizard samples sequenced for locus *mk75*, *mk209* and *mk245*. Detected haplotypes were distinguished in dominant (corresponding to the scored AFLP band) and recessive alleles. The frequency observed for the dominant alleles is indicated, as well as the number of segregating sites and indel sites. Haplotype diversity (H), nucleotide diversity (π) and Tajimas's D test values are also presented.

Outlier	Species	Alleles	Dominant haplotypes	Recessive haplotypes	Dominant allele frequency	H	Seg. sites	Indel sites	π	Tajima's D	Repetitive elements
mk75	*Lacerta lepida*	54	1	8	0.28	0.826	5	9	0.033		
	L. l. iberica	18	1	2	0.56	0.621	4	9	0.039	0.962	
	L. l. lepida	24	1	6	0.17	0.837	4	9	0.029	1.095	
	L. l. nevadensis	12	0	4	0.00	0.803	3	0	0.010	1.823	
	Lacerta tangitana	14	1	2	0.07	0.275	3	9	0.012	-1.278	
	Lacerta pater	6	0	2	0.00	0.333	1	0	0.002	-0.933	
mk209	*Lacerta lepida*	46	2	5	0.26	0.738	5	3	0.015		
	L. l. iberica	12	0	4	0.00	0.652	0	3	0.004	---	$(G)_{6\text{-}9}A(GA)_3$
	L. l. lepida	22	0	4	0.00	0.541	1	2	0.003	-1.162	$(G)_{6\text{-}8}A(GA)_3$
	L. l. nevadensis	12	2	0	1.00	0.545	1	0	0.002	1.486	$(G)_8A(GA)_3$
	Lacerta tangitana	14	0	9	0.00	0.934	4	14	0.025	-0.819	$(G)_{8\text{-}10}A(GA)_{3\text{-}7}$
	Lacerta pater	6	0	2	0.00	0.533	1	7	0.017	0.851	$(G)_{7\text{-}8}A(GA)_{5\text{-}6}$
mk245	*Lacerta lepida*	40	1	8	0.15	0.654	2	11	0.019		
	L. l. iberica	12	0	1	0.00	0.000	0	0	0.000	---	$(GTT)_3$
	L. l. lepida	16	1	5	0.38	0.842	1	9	0.026	0.650	$(GTT)_{3\text{-}6}$
	L. l. nevadensis	12	0	4	0.00	0.561	1	2	0.004	-1.141	$(GTT)_3$
	Lacerta tangitana	8	3	2	0.63	0.893	1	19	0.035	0.334	$(GTT)_{3\text{-}9}$
	Lacerta pater	6	1	3	0.17	0.867	1	12	0.033	0.851	$(GTT)_{6\text{-}10}$

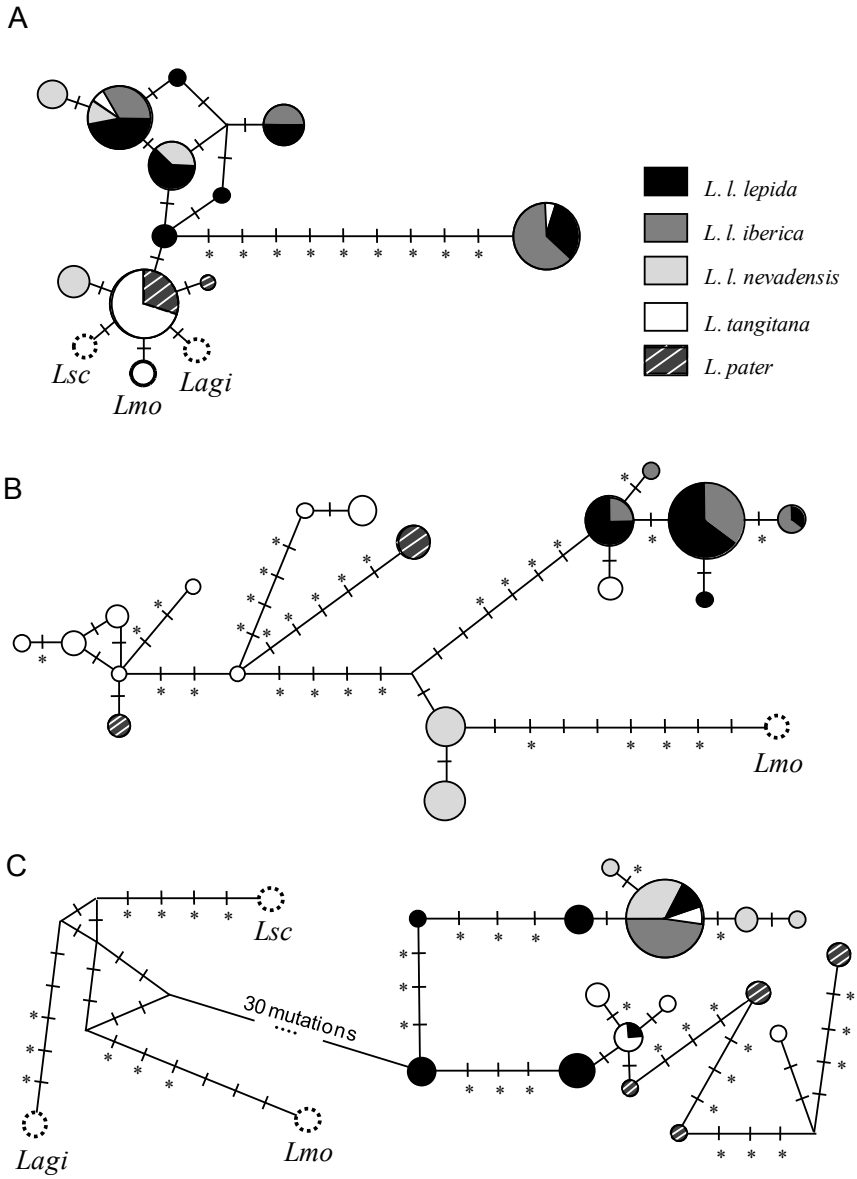


Fig. 3- Minimum spanning haplotype network for the *mk75* (A), *mk209* (B) and *mk245* (C). The size of the circles is proportional to sample size. Each mutation between haplotypes is represented by a dash. Asterisks denote mutations at indel sites. Non-ocellated lizard haplotypes are represented in white dashed circles: *L. agilis* (*Lag*), *Iberolacerta monticola* (*Imo*) and *L. shreiberi* (*Lsc*).

nevadensis samples (Fig. S2, Supporting information). The same individual from Morocco carrying European alleles from *mk75* also carries an *mk209* haplotype that differs in a single mutation from the second most frequent haplotype in European ocellated lizards (Fig. 3 B). An homologous and conserved fragment of *mk209* was amplified for *Iberolacerta monticola*, but it was not possible to obtain the homologous fragment from *L. schreiberi* nor *L. agilis*.

Three samples scored with the *mk245* band as present were sequenced and the presence of one or two copies of the dominant allele (with six repeats of GTT) in each sample was confirmed. Marker *mk245* was absent in populations from the genome scan from both ends of the environmental gradient (ALM and GAL) and reached higher frequencies in populations from the south of the Iberian Peninsula (CMA=0.80, TOL=0.95) (Fig. 2). Our sequencing efforts were then directed to understand how variable in length this microsatellite was in recessive alleles. We detected alleles with three to ten GTT repeats (Table 2). All samples sequenced from *L. l. iberica* and *L. l. nevadensis* were fixed for haplotypes with three repeats of GTT while 75% of sequenced samples from *L. l. lepida* were heterozygous for GTT repeat number. As for the African ocellated lizards, we detected four dominant allele haplotypes (with six GTT repeats) and five recessive allele haplotypes. Once more, the sample from Morocco that had European-like alleles both for *mk75* and *mk209* also had an European allele for *mk245*, with three GTT repeats, and was the only African sample with less than six GTT repeats (Table 2; Fig. 3 C). Regions flanking the *mk245* microsatellite were quite conserved. Although the number of GTT repeats both in *L. l. iberica* and *L. l. nevadensis* was always three, haplotype diversity was higher for *L. l. nevadensis*, while all *L. l. iberica* samples corresponded to the same haplotype (Table 2; Fig. 3 C). Marker *mk245* was amplified in *L. schreiberi*, *L. agilis* and *Iberolacerta monticola* with success. These three species had a long insertion of 29 bp right before the microsatellite sequence. The number of GTT repeats detected was four in *L. agilis*, five in *I. monticola* and six in *L. schreiberi*.

Tajimas' D values calculated to test for neutrality in each marker and each species or subspecies ranged from -1.48 to 1.82, but none differed significantly from zero (Table 2).

AFLP electrophoresis profiles from sequenced samples for markers *mk75*, *mk209*, *mk245*, *mk386* and *mk390* were inspected to identify bands that could correspond to the

expected size of the recessive alleles. For *mk75*, no markers with 202 bp (193bp from the dominant allele plus a 9 bp insertion) were detected in profiles from sequenced samples. We cannot exclude the existence of more sources of length polymorphism in *mk75* because sequences from the recessive allele were not complete and mutations at recognition sites of one or both restriction enzymes may be present. The same situation is true for *mk390*. As for marker *mk209*, recessive alleles detected in *L. lepida* sequences vary in expected band size from 297 to 300 bp. An AFLP marker with 298 bp is present in *mk209* band profiles, but the band is present in all individuals (Fig. S3, Supporting information). This indicates that recessive alleles from mk209 are probably co-migrating with another fragment with the same size. The same happens in *mk245* and *mk386* profiles, where other AFLP fragments match the predicted size of recessive alleles but their presence is not always in agreement with the scoring expectations based on sequenced haplotypes. It is likely that they co-migrate with other non-homologous fragments.

Discussion

AFLP outlier isolation and characterization

Our attempts to isolate outlier AFLP markers in ocellated lizards were successful for seven out of twelve outliers, reflecting the technical demands of the isolation process. However, several observations may help to avoid some pitfalls when isolating outlier markers from AFLP genome scans performed by CE. The first observation is that the chances of successful isolation are higher for fragments with larger sizes. Isolation of small fragments (<150 bp) is less prone to success because the fragment size distribution in AFLP profiles is asymmetrical, with higher density of small-sized fragments. The fragment that is meant to be isolated is probably closely flanked by fragments of similar size that can hardly be resolved in gel electrophoresis. Additionally, it has been suggested that the chances of homoplasy (co-migration of non-homologous bands with the same size) are higher for smaller fragments (Vekemans *et al.* 2002; Caballero *et al.* 2008).

Polyacrylamide (Wood *et al.* 2008) or Spreadex gels (Minder & Widmer 2008) can be used as somewhat expensive but higher resolution alternatives to regular agarose gels.

The decision on which gel matrix to use for band excision will depend on the availability of the required equipment and associated costs rather than on gel resolution power. Also some adjustments should be made in loading volumes of PCR products on the gel to increase the concentration of AFLP bands as compared to fluorescence in CE and compensate for the loss of DNA during purification of PCR products from gels.

Brugmans *et al.* (2003) proposed an alternative method for AFLP markers isolation, which corresponds to an AFLP-mediated mini-sequencing that reduces the density of bands in AFLP profiles prior to the excision of the desired band, without co-isolation of adjacent fragments. This approach consists of the use of 12 *Mse*I degenerate primers to determine, by subtractive amplification, the fourth, the fifth and the sixth selective base following the *Mse*I restriction site. The band is then amplified with the appropriate *Mse*I primer with six selective bases, excised from the gel, re-amplified and used for direct sequencing. This method may greatly improve the efficiency of fragment recovery from gels, even for smaller fragments. Nevertheless, small fragments like *mk311* still impose some other limitations. Validation of small-sized outliers might be compromised if no specific primers can be designed from the fragment sequence, imposing the use of genome walking strategies to expand the known sequence to its flanking regions in order to find suitable regions for primer design.

All AFLP outliers that were successfully isolated from agarose gels were cloned. A simple strategy was used to distinguish positive clones from clones bearing close migrating or other non-specific fragments. Clones were amplified with universal primers and then used in a second PCR to amplify the insert with fluorescently labelled AFLP selective primers. Insert size was determined by CE, thus assuring that only clones bearing the insert of the expected size were sequenced. None of the seven outliers sequenced in ocellated lizards belong to a coding region, but this is not a surprising observation. Mutational constrains in coding regions and the size of AFLP markers, that rarely exceeds 500 bp, implies that most polymorphic AFLP markers probably fall in non-coding regions, that comprise most of the genome. Consequently, AFLP markers with outlier behaviour will rather be in close linkage with the gene under selection than inside the gene sequence itself (Stinchcombe & Hoekstra 2008; Butlin 2010). In accordance with these predictions, follow-up studies of AFLP genome scans available to date (Minder & Widmer 2008; Wood *et al.* 2008) detected mostly non-coding outlier fragments, containing repetitive or transposable elements.

Once an AFLP marker is sequenced and specific internal primers can be designed to amplify the fragment directly from undigested genomic DNA, we can access the sources of length polymorphism. In ocellated lizards, the sources of length polymorphism in investigated outliers were mostly internal indels (*mk75*, *mk209* and *mk386*) or polymorphic microsatellites (*mk245*). The failure of the amplification of the full fragment sequence of recessive alleles in *mk75* and *mk390* might indicate additional polymorphism at selective bases or enzyme restriction sites, but it would be necessary to extend the sequence towards their flanking regions by genome walking to access such sources of polymorphism.

With the knowledge of AFLP allele sequences, outliers can be converted in co-dominant markers with the development of specific primers, thus overcoming the disadvantages that dominant markers pose for population genomic analyses. A recent study by Foll *et al.* (2010) presented new developments on AFLP scoring methodology that account for band intensity in CE to genotype bands as co-dominant markers, where band presence can be classified as homozygous if the band is more intense, or as heterozygous if the band is less intense. This approach opens new perspectives on AFLP usefulness in population genomics, but the method depends on the use of high quality electrophoresis profiles, because band intensity might be highly variable among samples, reflecting technical variance in the generation of band profiles.

Another observation from our ocellated lizard genome scan follow-up is that outlier fragments may often include one or more internal repetitive elements of variable length (*mk209* and *mk245*), thus resulting in several length-polymorphic alleles for the same locus, in the same population, and generated by the same selective primer combination. If the repetitive element is formed by mononucleotide units like in *mk209*, alleles differing in one unit will probably be indistinguishable from each other in AFLP electrophoresis profiles, where resolution below two base pairs is poor. If, however, the AFLP marker contains a microsatellite composed of trinucleotide units as for *mk245*, several band sizes are expected within the same population and they might bias AFLP band scores due to homoplasy, if allele size coincides with other non-homologous markers, or might lead to the non-independence of some polymorphic markers, if alleles of different size but from the same locus are scored as independent markers.

The risk for statistical bias caused by homoplasy and non-independence of anonymous markers like AFLPs has been noted before (Bonin *et al.* 2007, Caballero *et al.* 2008). Species with larger genomes are richer in repetitive elements and transposable elements. These regions have fewer mutational constraints and because of that are prone to contain AFLP restriction sites and generate several fragments of variable sizes. It has been suggested that smaller AFLP fragments (< 150 bp) should be avoided to minimize homoplasy (Bonin *et al.* 2007). However, fragments with repetitive elements can easily achieve larger sizes, as in *mk209* (303 bp) and in *mk245* (250 bp). Only a minority of AFLP genome scan studies have tried to control for non-indepedent markers (Murray & Hare 2006; Poncet *et al.* 2010; Paris *et al.* 2010), and they have done so by testing for linkage disequilibrium among AFLP markers. When an AFLP locus contains a repetitive element that can generate more than two length variable alleles, like *mk245*, it becomes difficult to target bands that actually belong to the same locus within AFLP profiles. Therefore, interpretation of results from AFLP outlier detection should be made with caution and special efforts should be made to investigate the sequence content of outliers in order to validate their selection signature (Butlin 2010).

Outlier sequence variation in ocellated lizards

Data from *mk75* sequences reveal that the dominant band detected under selection in the genome scan corresponds to a quite conserved allele with a deletion of 9 bp that reaches high frequencies in *L. l. iberica* in accordance with AFLP outlier 75 scoring frequencies. The allele could also be found in heterozygous samples from *L. l. lepi*da but was not detected in *L. l. nevadensis*. Locus 75 was detected as an outlier only by BayeScan and its band presence was found in association with higher levels of precipitation, as registered in the northwest of the Iberian Peninsula, where *L. l. iberica* is found (Nunes *et al.* 2011). However, genomic resources available from reptile species in GenBank were not sufficient to find homology with the noncoding sequence of *mk75* alleles, and therefore, the reason for the selection signature remains unknown and needs further investigation.

For *mk209*, the TGGA indel justifies the length polymorphism of outlier 209. Sequence data indicate that the presence of TGGA is fixed in *L. l. nevadensis* but absent in all other European ocellated lizards. Locus 209 was detected as an outlier only by

DFDIST and its band presence was found in association with low levels of precipitation, as found in the southeast of the Iberian Peninsula, where *L. l. nevadensis* lives (Nunes *et al.* 2011). Because the neutral divergence of *L. l. nevandensis* is deep (Paulo *et al.* 2008; Miraldo *et al.* 2011), the role of drift might have been relevant for the fixation of *mk209* in *L. l. nevandensis*. Nevertheless, more genomic resources are needed to understand which genes might be linked with locus *mk209* and its importance in *L. l. nevadensis* evolution.

Sequence data from *mk245* are perhaps the most surprising. A microsatellite composed by GTT repeats is responsible for locus 245 length polymorphism. Locus 245 was detected as an outlier by both DFDIST and BayeScan and was the one with the strongest association with maximum temperature variation along the Iberian Peninsula (Nunes *et al.* 2011). The dominant allele corresponds to six repeats of GTT and is absent from *L. l. iberica* and *L. l. nevadensis* samples, as expected from band frequency scored for locus 245 in the genome scan. The most striking observation is that *mk245* alleles found in *L. l. iberica* and *L. l. nevadensis* have all three GTT repeats, but the ecological settings faced by these subspecies are actually the most contrasted ones across the species range. Sequences of *mk245* in samples from *L. l. lepida* populations are polymorphic and show not only the dominant allele but also alleles with three to five GTT repeats. If locus *mk245* is linked with genes that respond to higher temperatures, then we would not expect that *L. l. nevadensis* would have the same fixed allele as *L. l. iberica*, which lives in a much colder region.

Microsatellites detected in loci *mk209* and *mk245* are even more variable in ocellated lizards from North Africa, but their flanking regions remain quite conserved across species. Haplotypes found in *L. pater* and *L. tangitana* for the three investigated loci (*mk75*, *mk209* and *mk245*) are normally distinct from the ones found in European ocellated lizards, with one remarkable exception: a single individual from Morocco (*L. tangitana*) carries European haplotypes for the three loci. Given the long time (about 11 million years) of divergence and isolation of this species from European lizards (Paulo *et al.* 2008), it is unexpected to find such a level of ancestral polymorphism retained in a single individual at the nuclear level. Perhaps this might be indicative of past episodes of secondary contact between African and European lizards, but it remains extremely speculative with the current knowledge.

Future directions

The sequence data obtained so far for the investigated outliers is restricted and neither confirms nor denies the existing evidence that they might be candidates influenced by selection. However, none of them correspond to coding DNA, implying that they may be involved either in the regulation of genes under selection or simply be in linkage disequilibrium with the actual target of selection. To further investigate the possible role of these outliers in ocellated lizards local adaptation, we need to develop more genomic resources in this species. Lizards have been important models in ecological and evolutionary studies (Camargo *et al.* 2010), but genomic resources available for reptile species are particularly scarce when compared with other vertebrate groups. Genome walking from the sequences of outlier loci is an important step to analyze differentiation in outlier flanking regions. A useful tool, potentially applicable in any non-model species, might be the construction of a bacterial artificial chromosome (BAC) library to probe with outlier sequences, as performed in Wood *et al.* (2008) when performing the follow-up of the AFLP genome scan in *Littorina saxatilis*. Alternatively, genomic resources could be obtained with next generation sequencing (NGS) technologies for *L. lepida* or for one of its closely related species, like *Iberolacerta monticola*, helping to bridge the gap between outlier sequences and genes that surround them in the genome. Markers *mk75*, *mk209* and *mk245* were successfully amplified not only in ocellated lizard species from north Africa, revealing that these markers are conserved above the species level, but also above the genus level, since the three markers could be amplified in *Iberolacerta monticola*. Markers *mk75* and *mk209* could even been amplified in *L. schreiberi* and *L. agilis*, showing that the flanking regions of these markers are quite conserved among Lacertidae species. The transferability of AFLP markers between closely related species opens new options for further follow-up studies. Today NGS provides a fast and cost-effective way to discover thousands of SNPs (co-dominant markers) that can be used to construct a linkage map (Slate *et al.* 2010). If a closely related species to the genome scan target species offers better chances for maintenance and breeding in captivity (simpler ecological requirements or shorter generation time), then it might be a good choice to invest in the generation of NGS data from that species.

Acknowledgements

This research was supported by a doctoral scholarship from FCT (SFRH/BD/21306/2005) to Nunes VL and the FCT project POCI/BIA/59288/2004- "Process of speciation: accessing the genic view." coordinated by Paulo OS.

References

Apple JL, Grace T, Joern A, Amand PS, Wisely SM (2010) Comparative genome scan detects host-related divergent selection in the grasshopper *Hesperotettix viridis*. *Molecular Ecology*, **19**, 4012-4028.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16,** 37-48.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **263**, 1619-1626.

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**, 3737-58.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773-83.

Brugmans B, van der Hulst RGM, Visser RGF, Lindhout P, van Eck HJ (2003) A new and versatile method for the successful conversion of AFLP markers into simple single locus markers. *Nucleic Acids Research*, **31**, e55.

Butlin RK (2010) Population genomics and speciation. *Genetica*, **138**, 409-418.

Caballero A, Quesada H, Rolan-Alvarez E (2008) Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, **179**, 539-554.

Camargo A, Sinervo B, Sites JW (2010) Lizards as model organisms for linking phylogeographic and speciation studies. *Molecular Ecology*, **19**, 3250-3270.

Croucher PJP, Oxford GS, Lam A, Gillespie RG (2011) Stabilizing selection maintains exuberant colour polymorphism in the spider *Theridion californicum* (Araneae, Theridiidae). *Molecular Ecology*, **20**, 206-218.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285-298.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1,** 47-50.

Flot JF, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes*, **6**, 627-630.

Foll M, Fischer MC, Heckel G, Excoffier L (2010) Estimating population structure from AFLP amplification intensity. *Molecular Ecology*, **19**, 4638-4647.

Foll M, 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.

Gaggiotti OE (2010) Bayesian statistical treatment of the fluorescence of AFLP bands leads to accurate genetic structure inference. *Molecular Ecology*, **19**, 4586-4588.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41,** 95-98.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955-69.

Mateo JA, Castroviejo J (1990) Variation morphologique et révision taxonomique de l'espèce *Lacerta lepida* Daudin, 1802 (Sauria, Lacertidae). *Bulletin du Muséum National D'Histoire Naturelle, Paris*, **12**, 691-706.

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Molecular Ecology*, **17**, 1552-1563.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Moreira PL, Nunes VL, Martín J, Paulo OS (2007) Copulatory plugs do not assure high first male fertilisation success: sperm displacement in a lizard. *Behavioral Ecology and Sociobiology*, **62**, 281-288.

Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Molecular Ecology*, **15**, 4229-4242.

Nunes VL, Beaumont MA, Butlin RK, Paulo OS (2011) Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Molecular Ecology*, **20**, 193-205.

Paris M, Boyer S, Bonin A *et al.* (2010) Genome scan in the mosquito *Aedes rusticus*: population structure and detection of positive selection after insecticide treatment. *Molecular Ecology*, **19**, 325-337.

Paulo OS, Pinheiro J, Miraldo A *et al.* (2008) The role of vicariance vs. dispersal in shaping genetic patterns in ocellated lizard species in the western Mediterranean. *Molecular Ecology*, **17**, 1535-1551.

Pérez-Figueroa A, García-Pereira M, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267-2276.

Pina-Martins F, Paulo OS (2008) Concatenator: sequence data matrices handling made easy. *Molecular Ecology Resources*, **8**, 1254-1255.

Polanco C, González AI, de la Puente R, Somalo S, Ruiz ML (2005) Cloning of AFLP markers detected by fluorescence capillary electrophoresis. *Plant Molecular Biology Reporter*, **23**, 271-277.

Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology*, **19**, 2896-2907.

Reddy PS, Mahanty S, Kaul T *et al.* (2008) A high-throughput genome-walking method and its use for cloning unknown flanking sequences. *Analytical Biochemistry*, **381**, 248-253.

Rowe HC, Renaut S, Guggisberg A (2011) RAD in the realm of next-generation sequencing technologies. *Molecular Ecology*, **20**, 3499-3502.

Rozen S, Skaletsky HJ (2000) Primer 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* (eds Krawetz S, Misener S), pp 365-386. Humana Press, Totowa, New Jersey.

Slate J, Santure AW, Feulner PGD *et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, **26**, 275-284.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology and Evolution*, **25**, 705-712.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, **76**, 449-462.

Stephens M, Smith NJ, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978-989.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158-170.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673-4680.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139-151.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP - A new technique for DNA-fingerprinting. *Nucleic Acids Research*, **23**, 4407-4414.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123-3135.

CHAPTER 3

Supporting information

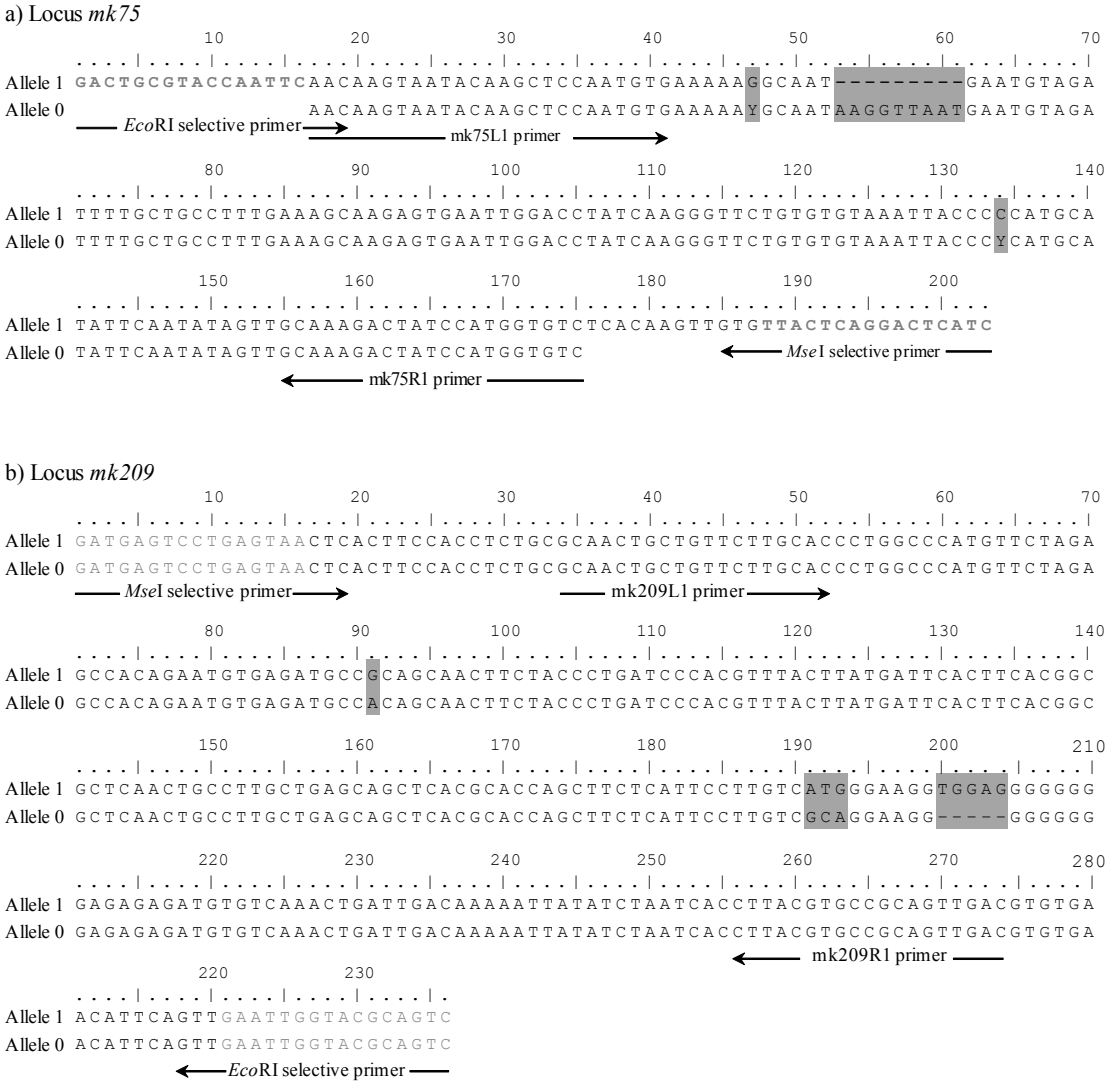
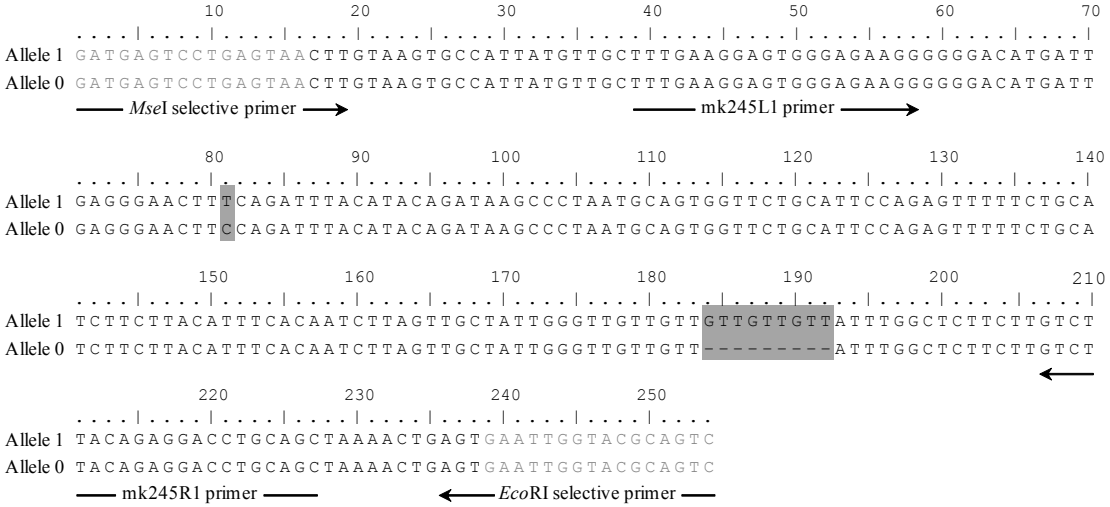
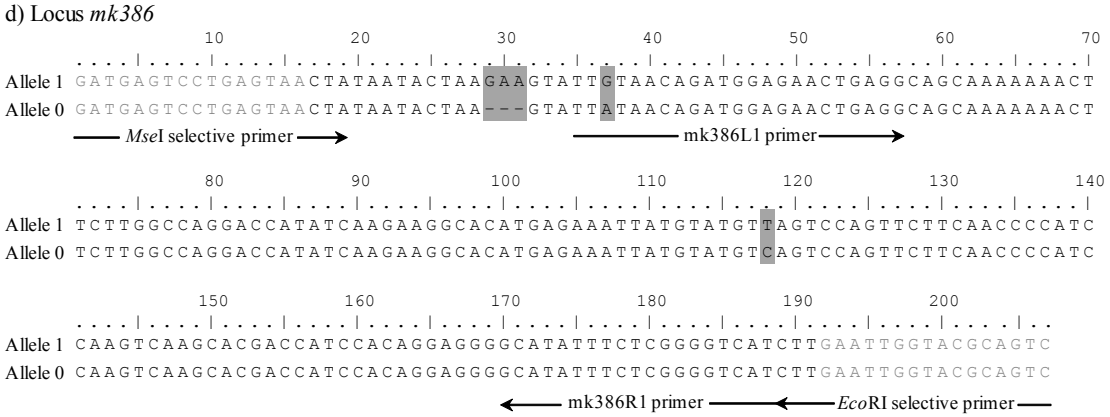


Fig. S1 Alignment of sequences from a dominant allele (allele 1), retrieved by clone sequence, with a recessive allele (allele 0), obtained with a combination of AFLP selective primers with specific internal primers. Indels and SNPs are highlighted in grey shades. The binding sites of MseI and *Eco*RI selective primers and of locus specific primers are indicated (arrows).

c) Locus *mk245*



d) Locus *mk386*



e) Locus *mk390*

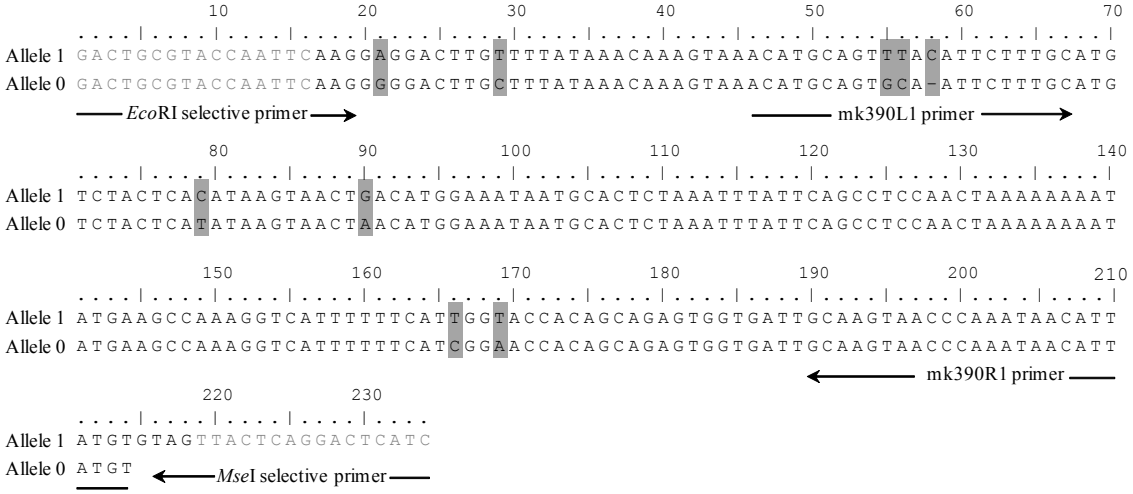


Fig S1 (continued)

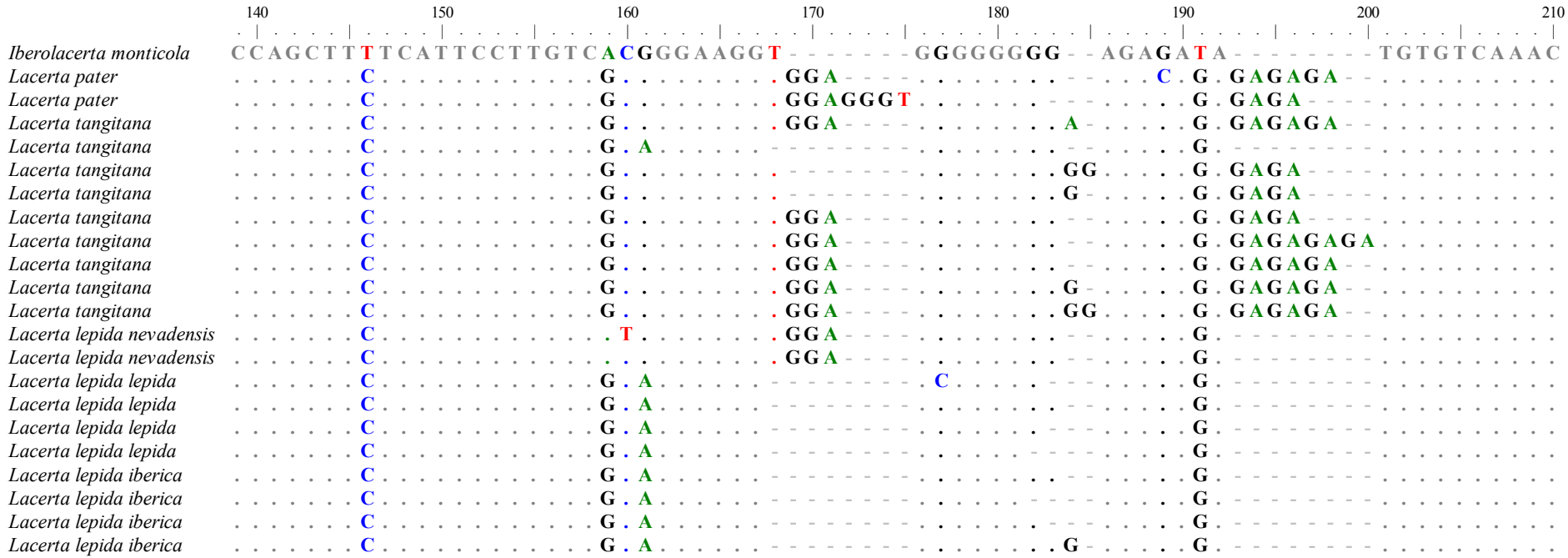


Fig S2- Partial alignment from haplotypes detected in locus *mk209* sequences. Dots denote conserved bases and dashes denote gaps. A repetitive element composed by Gs at positions 176-185 is variable in length. Another repetitive element composed by GA units at positions 187-200 is variable in African ocellated lizards (*L. pater* and *L. tangitana*).

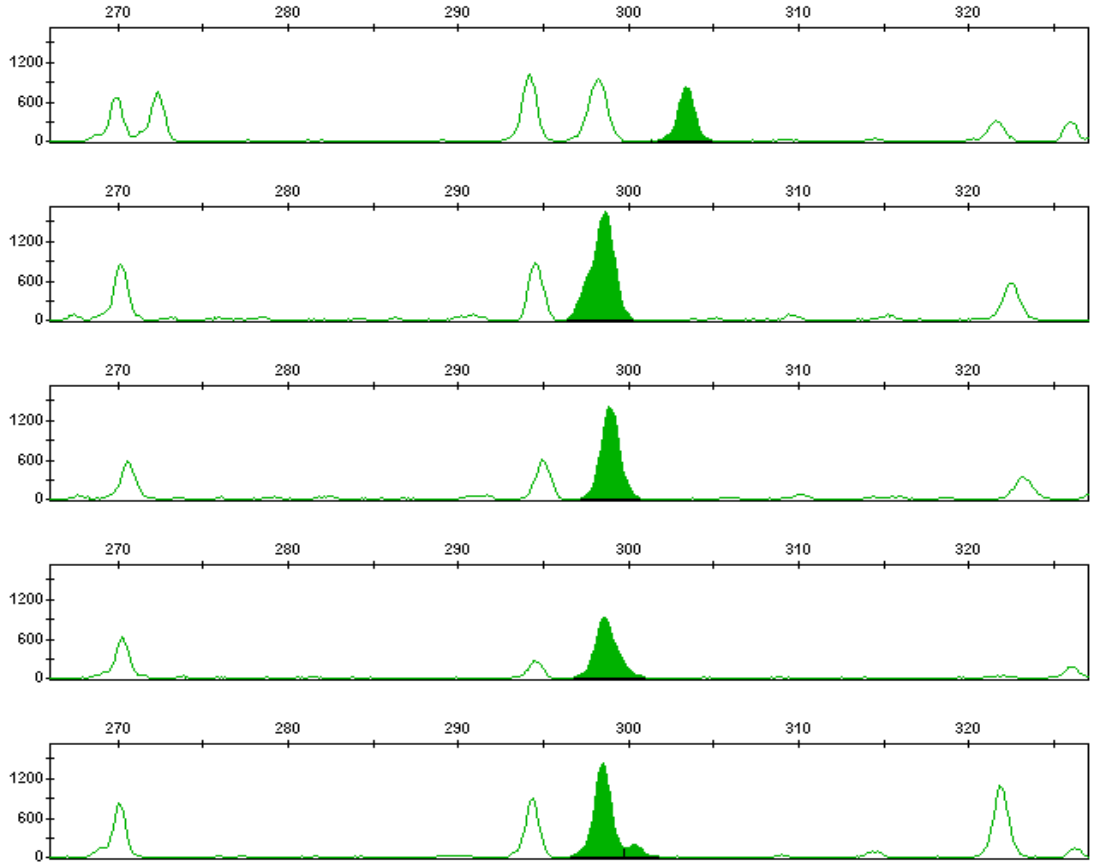


Fig. S3- AFLP profile caption from GeneMapper of five samples sequenced for locus *mk209*. The peak scored as *mk209* (dominant allele) had 303 base pairs (bp) and is shaded in green on the first row sample, but absent from all other four samples. Shaded peaks on rows two to five correspond to the expected size of *mk209* recessive alleles found by sequencing: two alleles on the second row, with 297 and 298 bp; two alleles with 298bp on the third row; two alleles on the fourth row with 298 and 299 bp; two alleles on the fifth row with 298 and 300 bp. AFLP band with 298 bp was present in all samples in the genome scan and therefore was not considered as a polymorphic fragment. Most recessive alleles from *mk209* seemed to have comigrated with another homplasic fragment with 298 bp.

CHAPTER 4

Association of *Mc1r* variants with ecologically relevant phenotypes in the European ocellated lizard, *Lacerta lepida*

Vera L. Nunes[1], Andreia Miraldo[2], Mark A. Beaumont[3,4], Roger K. Butlin[5], Octávio S. Paulo[1]

[1] Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, DBA/FCUL, 1749-016 Lisboa, Portugal

[2] Metapopulation Research Group, Department of Biological and Environmental Sciences, P.O. Box 65, FI-00014, University of Helsinki, Finland

[3] School of Animal and Microbial Sciences, University of Reading, P.O. Box 228, Reading, RG6 6AJ, United Kingdom

[4] Schools of Mathematics and Biological Sciences, University of Bristol, UK

[5] Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, England

Journal of Evolutionary Biology (2011) 24, 2289-2298

doi: 10.1111/j.1420-9101.2011.02359.x

Association of *Mc1r* variants with ecologically relevant phenotypes in the European ocellated lizard, *Lacerta lepida*

V. L. NUNES*, A. MIRALDO†, M. A. BEAUMONT‡§, R. K. BUTLIN¶ & O. S. PAULO*

*Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, DBA/FCUL, Lisboa, Portugal
†Metapopulation Research Group, Department of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland
‡School of Animal and Microbial Sciences, University of Reading, Reading, UK
§Schools of Mathematics and Biological Sciences, University of Bristol, Bristol, UK
¶Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

Keywords:

adaptation;
candidate genes;
colour variation;
contact zone;
molecular evolution;
natural selection;
population genetics.

Abstract

A comprehensive knowledge on the genetic basis of coloration is crucial to understand how new colour phenotypes arise and how they contribute to the emergence of new species. Variation in *melanocortin-1 receptor* (*Mc1r*), a gene that has been reported as a target for repeated evolution in a wide range of vertebrate taxa, was assessed in European ocellated lizards (*Lacerta lepida*) to search for associations with melanin-based colour phenotypes. *Lacerta lepida* subspecies' distribution is associated with the three major bio-climatic regions in the Iberian Peninsula. A nonconserved and derived substitution (T162I) was associated with the *L. l. nevadensis* phenotype (prevalence of brown scales). Another substitution (S172C) was associated with the presence of black scales in both *L. l. lepida* and *L. l. iberica,* but no mutations were found to be associated with the higher proportion of black in *L. l. iberica*. Extensive genotyping of *Mc1r* along the contact zone between *L. l. nevadensis* and *L. l. lepida* revealed low gene flow (only two hybrids detected). The implications of these findings are discussed in the context of previous knowledge about the evolutionary history of ocellated lizards.

Introduction

Species and subspecies were once described based on morphological traits alone. Nowadays, the use of molecular tools has become widespread in species other than model organisms, allowing the investigation into the genetic basis underlying morphological variation and its evolution. Coloration has been one of the most extensively studied traits at the molecular level (Hoekstra, 2006). Variation in colour within and between species often has important implications for adaptation to the environment such as in concealment, thermoregulation, mimicry and warning signals (Roulin, 2004; Protas & Patel, 2008). Mammals' and birds' coat colour depends on the amount of eumelanin (black/brown) and phaeomelanin (red/yellow) pigments, produced in melanocytes (Hoekstra, 2006). In poikilothermic vertebrates, besides melanophores (pigment cells able to produce eumelanin, but not phaeomelanin as in mammals), two additional pigment cell types contribute to the diversity of colours and patterns visible to the human eye: xanthophores/erythrophores (yellow to red carotenoid or pteridine pigments) and light-reflecting iridophores (iridescent or silvery structural colours) (Grether *et al.*, 2004). The spatial arrangement and architectural combination of each type of pigment cells determine the skin colour of reptiles, fishes and amphibians (Grether *et al.*, 2004).

Genes and pathways underlying nonmelanin-based colour patterns in poikilothermic vertebrates remain poorly understood (Hubbard *et al.*, 2010). The melanin synthesis pathway is an exception, as it is quite conserved among vertebrates, and extensive study in model species has led to more than 100 well-characterized genes with known effects on coloration (Hoekstra, 2006). *Melanocortin-1 receptor* (*Mc1r*), in particular, has been extensively studied to date, and associations between changes in its

Correspondence: Vera L. Nunes, Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, Edifício C2-2.3.22, Campo Grande, 1749-016 Lisboa, Portugal.
Tel.: +351 217500000; fax: +351 217500028;
e-mail: vlnunes@fc.ul.pt

© 2011 THE AUTHORS. *J. EVOL. BIOL.* **24** (2011) 2289–2298
JOURNAL OF EVOLUTIONARY BIOLOGY © 2011 EUROPEAN SOCIETY FOR EVOLUTIONARY BIOLOGY

coding sequence and differences in pigmentation have been documented in more than 20 vertebrate species, mostly mammals and birds (Gompel & Prud'homme, 2009). Because *Mc1r* is a small intronless gene of about 950 base pairs (bp) and functionally conserved among vertebrates, there has been a bias towards its study over other candidate genes underlying coloration (Mundy, 2005; Gompel & Prud'homme, 2009). Nevertheless, growing evidence suggests that *Mc1r* may in fact be a recurrent target for evolutionary changes with effect on pigmentation (Gompel & Prud'homme, 2009), being responsible for convergent phenotypes in closely, but also in distantly, related species like mice and mammoths (Manceau *et al.*, 2010). *Mc1r* function is specialized for melanin synthesis, and the high number of variants associated with body colour indicates that many amino acid substitutions are likely to affect protein function without deleterious consequences (Mundy, 2005; Gompel & Prud'homme, 2009). Recent studies provided evidence for the causal effects of *Mc1r* mutations on melanin production by functional assays in cavefish *Astyanax mexicanus* (Gross *et al.*, 2009) and in lizards *Aspidoscelis inornata* and *Sceloporus undulatus* (Rosenblum *et al.*, 2010), reinforcing the importance of *Mc1r* as a candidate gene for coloration in vertebrate groups other than mammals or birds.

Many lizard species are colour polymorphic, enabling an efficient camouflage within the background against visual predators. The species considered in this study, *Lacerta lepida* (synonym of *Timon lepidus* according to Mayer & Bischoff (1996)), is such an example. *Lacerta lepida* is presently the only ocellated lizard species inhabiting the European continent. It is widespread throughout the Iberian Peninsula, extending its range into some discontinuous areas in southern France and northern Italy. Two closely related ocellated lizard species are found in the north of Africa: *Lacerta pater* and *Lacerta tangitana*. In the Iberian Peninsula, three *L. lepida* subspecies were initially recognized based on morphological divergence alone: *L. lepida lepida*, *L. lepida nevadensis* and *L. lepida iberica* (Mateo & Castroviejo, 1990). Their taxonomy was later supported by molecular, cytogenetic and genetic evidence (Mateo *et al.*, 1996, 1999; Paulo *et al.*, 2008; Miraldo *et al.*, 2011). Each *L. lepida* subspecies presents a distinct colour phenotype and occurs in association with a single bioclimatic region of the species' distribution (Mateo & Castroviejo, 1990). The nominal subspecies, *L. l. lepida*, occurs almost all over the species distribution area, in typical Mediterranean habitats. Its dorsal coloration is composed of an irregular pattern of black scales over a green to yellowish background, where green/yellow scales are enclosed by black scales in more or less well-defined circles (Fig. S1, Supporting information). The nominal subspecies is replaced by *L. l. iberica* in the north-west of the Iberian Peninsula, in a region affected by the Atlantic climate, with mild temperatures and abundant annual rainfall. *Lacerta l. iberica* has a smaller body size and darker dorsal coloration, with predominance of black scales (Mateo & Castroviejo, 1990). The third subspecies, *L. l. nevadensis*, replaces *L. l. lepida* on the south-eastern coast of the Peninsula. This region is characterized by a reduced and irregular amount of annual precipitation (< 300 mm), which leads to an increased aridity with sparse shrub-like vegetation and large portions of exposed soil (Mateo & Castroviejo, 1990; Hodar *et al.*, 1996). Coloration of *L. l. nevadensis* is remarkably distinct, with a reduced or faded pattern of green/yellow scales and a dominance of brown rather than black scales, producing a phenotype that blends in with the dry landscape (Mateo & Castroviejo, 1990; Mateo & López-Jurado, 1994).

Investigations into genes underlying coloration polymorphism in reptiles have so far been limited to *Mc1r* (Rosenblum *et al.*, 2004; Raia *et al.*, 2010). The characterization of the *Mc1r* locus in three White Sands lizard species (*Aspidoscelis inornata*, *Sceloporus undulatus* and *Holbrookia maculata*) highlighted a single derived amino acid replacement in each species associated with blanched phenotypes (Rosenblum *et al.*, 2004), leading to partial loss of function of *Mc1r* in *A. inornata* and *S. undulatus* (Rosenblum *et al.*, 2010). Here, we investigate *Mc1r* variation in ocellated lizards to shed light on the genetic basis of melanin-based colour phenotypes found in the Iberian Peninsula and contribute to disentangling the importance of *Mc1r* in the evolution of ocellated lizard coloration. Special attention has been given to the variation of *Mc1r* in the vicinity of the contact zone between the nominal subspecies and *L. l. nevadensis*, as this pair exhibits the most contrasting colour phenotypes.

Materials and methods

Colour phenotype characterization

The colour phenotype of European ocellated lizards was characterized through the visual inspection of film photographs from 1997 of the dorsal pattern of 82 adult lizards. The animals were captured at six locations in the Iberian Peninsula: Toledo, Béjar, Castro Marim and Peniche (within *L. l. lepida* distribution range), Galicia (*L. l. iberica*) and Almeria (*L. l. nevadensis*; Fig. 1). Each lizard was placed in a box with a glass cover to fully expose the dorsal side of the animal. A piece of millimetric paper was attached to the glass as size reference. Photographs were taken with a Canon A1 camera (Canon Inc., Tokyo, Japan) using 35-mm Kodak Gold film. Tail tissue samples were collected from the individuals mentioned above and from 19 additional lizards from Serra da Estrela (*L. l. lepida*), Gerês (*L. l. iberica*), Azrou (*L. tangitana*) and Tabarka (*L. pater*; Fig. 1), but no photographic record is available from these. All animals were immediately released back into the wild.

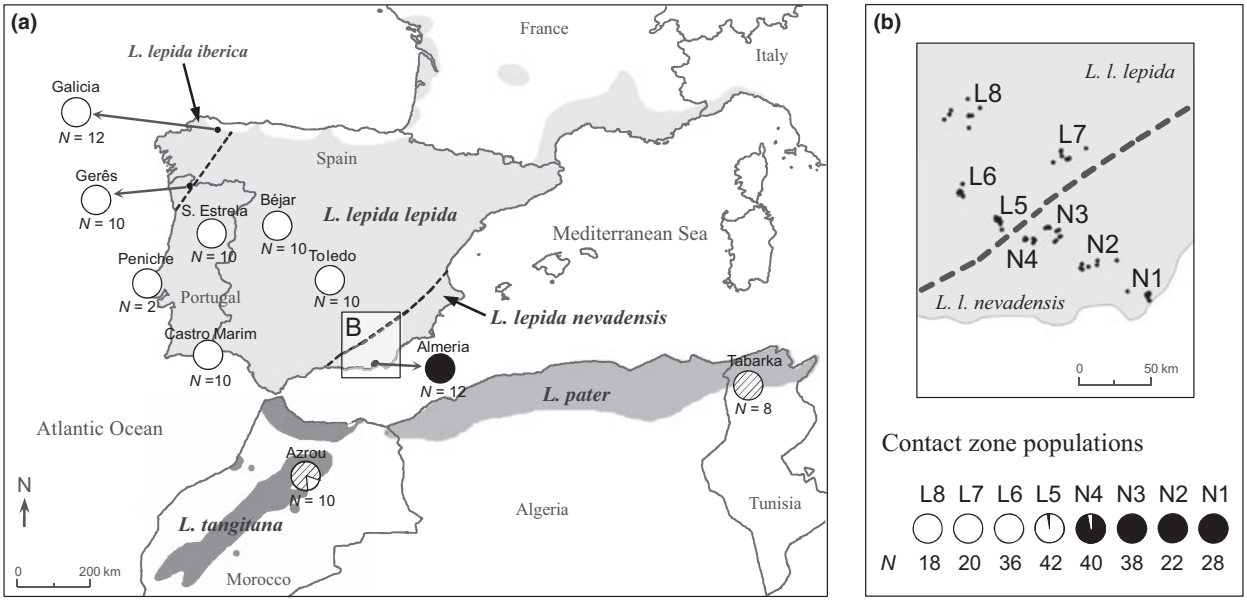


Fig. 1 Map of the western Mediterranean region (a) showing the distribution of ocellated lizards: *Lacerta lepida* (Europe), *L. tangitana* (Morocco) and *L. pater* (Algeria and Tunisia). In the Iberian Peninsula, known distribution limits of *L. lepida* subspecies (*L. l. iberica*, *L. l. lepida* and *L. l. nevadensis*) based on morphology are indicated by dashed lines. A transect of sampled locations across the contact zone between *L. l. nevadensis* (N1–N4) and *L. l. lepida* (L5–L8) is presented in more detail in the zoom box (b). In (a) and (b), pie charts denote the proportion of sampled alleles (*N*) in each location that correspond to each *Melanocortin-1 receptor* haplotype, based on three single-nucleotide polymorphisms, at positions 477 (synonymous), 485 and 514 (nonsynonymous): G-C-T (white), A-T-A (black) and A-C-A (stripped pattern).

To characterize the prevalence of melanin-based colours (black/brown) in lizards' dorsal patterns, all scales present in 1 cm^2 of the mid-dorsal region of each individual were counted and classified according to their colour in one of three possible categories: black, brown or green/yellow scales. The partial melanization index (Mateo, 1988) was calculated as the ratio of the number of black scales over the number of green/yellow scales in 1 cm^2 of the mid-dorsal region. Phenotypic data from both sexes were treated together because no significant sexual dimorphism in dorsal pattern has been described to date (Mateo, 1988; Font *et al.*, 2009). Sample sizes by sex are very small for some populations, but our data reveals no significant differences in counts of scales between males and females (Fisher's exact test, $P > 0.05$), although body size, and as a consequence scale size, is generally higher in males.

Mc1r genotyping and analyses

Whole genome DNA was extracted from tail tissue using Jetquick Tissue DNA kit (Genomed, Löhne, Germany). An additional set of 122 DNA samples was made available from the work of Miraldo (2009), belonging to eight locations (located 20–50 km apart from each other) along a transect perpendicular to the putative contact zone between *L. l. lepida* and *L. l. nevadensis* subspecies in the south-east of the Iberia Peninsula (locations N1–N4 and L5–L8 from Fig. 1). Populations N1 to N4 are fixed for *L. l. nevadensis* haplotypes and populations L5 to L8 are fixed for *L. l. lepida* haplotypes in cytochrome *b* (Miraldo, 2009). DNA samples from other Lacertidae species from the Iberian Peninsula, available from previous studies, were used as outgroups for *Mc1r* analysis: two samples from *L. schreiberi* (Paulo *et al.*, 2001), four from *Iberolacerta monticola* (Moreira *et al.*, 2007) and one from *L. agilis* (Paulo *et al.*, 2008).

A central portion of 450 bp from the *Mc1r* gene was successfully amplified with primers F2 (5′-TACTACTTCATCTGCTGCCTGGC-3′) and R1 (5′-CCCAGSAGGATGGTGAGGGTG-3′) (Rosenblum *et al.*, 2004). PCRs were performed in 25 μL of total volume with 1× PCR buffer (Promega), 1 U *Taq* polymerase (Promega, Madison, WI, USA), 2.0 mM $MgCl_2$, 0.12 mM dNTPs and 0.4 μM of each primer. The cycling conditions used were 94 °C for 3 min, 35 × (94 °C for 30 s, 60 °C for 45 s and 72 °C for 60 s) and 72 °C for 10 min. Purified products (Sureclean; Bioline, London, UK) were sequenced in both directions for a subset of 37 samples, representative of all sampled locations, using standard protocols (BigDye Terminator v.3.1; Applied Biosystems Foster City, CA, USA) on an ABI PRISM 310 (Applied Biosystems). Sequences were edited in Sequencher v.4.0.5 (Gene Codes Co., Ann Arbor, MI, USA) and deposited in GenBank (accession numbers JF732930–JF732966). PHASE 2.1.1 (Stephens *et al.*, 2001; Stephens & Scheet, 2005) was used to infer haplotype phase of six individuals that were heterozygous at multiple sites. We ran the

algorithm five times (1000 iterations with the default values) with different random number seeds, and the same haplotypes were consistently recovered in each run.

Alignments with other *Mc1r* sequences from other vertebrates available in GenBank were performed with CLUSTAL W (Thompson *et al.*, 1994). These included 11 lizard species (*Podarcis bocagei* GU180965, *Aspidoscelis inornata* AY586066, *Mabuya wrightii* GU180949, *Phelsuma astriata* GU180957, *Tarentola mauritanica* HM014691, *Urocotyledon inexpectata* GU180930, *Phrynosoma platyrhinos* AY586113, *Holbrookia maculata* AY586106, *Sceloporus undulatus* AY586127, *Uta stansburiana* AY586159, *Anniella pulchra* AY586034), three snake species (*Thamnophis sirtalis* AY586157, *Morelia boeleni* FJ865133, *Crotalus tigris* EU526278), one bird (*Gallus gallus* AY220305), one fish (*Takifugu rubripes* AY227791) and one mammal (*Bos taurus* AF445641).

Nucleotide diversity (π) and haplotype diversity (H) were determined for each ocellated lizard species and subspecies in DNAsp v5.10 (Librado & Rozas, 2009). Linkage disequilibrium and tests of neutrality were also performed in DNAsp. Neutrality was tested with Tajima's D test (Tajima, 1989) for each ocellated lizard species or subspecies, and the McDonald–Kreitman test (McDonald & Kreitman, 1991) was used to search for selection signatures in three pairs of pooled samples: (i) *L. l. nevadensis* × *L. pater*, (ii) *L. l. nevadensis* × *L. tangitana* and (iii) *L. l. nevadensis* × *L. l. lepida* and *L. l. iberica*. To infer the relationships among haplotypes, a minimum spanning network was constructed with the median-joining method (Bandelt *et al.*, 1999) in NETWORK 4.51 (http://www.fluxus-engineering.com). The input file was converted from fasta to nexus format with CONCATENATOR (Pina-Martins & Paulo, 2008).

Further investigations concerning the geographical distribution of *Mc1r* variants across *L. lepida*'s range, particularly at the contact zone between *L. l. lepi*da and *L. l. nevadensis*, were conducted by genotyping the remaining set of DNA samples (139 lizards) for three single-nucleotide polymorphisms (SNPs) at sites 477, 485 and 514, suspected to be diagnostic for *L. l. nevadensis*. Genotypes were obtained from single-strand sequences of *Mc1r*, using the same protocol as above. Samples suspected to be heterozygous for any of the SNPs were sequenced for the complementary strand to confirm their genotype. Associations between SNP genotypes and colour phenotypes were tested with contingency tables using Fisher's exact test (Fisher, 1935) in R (http://www.R-project.org).

Results

Colour phenotype

Despite individual variation in dorsal colour pattern, it is possible to visually recognize differences in melanin-based colours (black/brown) associated with each of the *Lacerta lepida* subspecies (Fig. S1, Supporting information). The colour phenotype observed in *L. l. nevadensis* individuals from Almeria shows the most conspicuous differences from all others. Unlike populations from *L. l. lepida* and *L. l. iberica*, all *L. l. nevadensis* individuals analysed lack green scales on the head, the hind legs and the tail, having only brown/grey scales on these body parts. Another characteristic trait of the *L. l. nevadensis* phenotype is the tendency for a decrease in the green dorsal pattern. Among the eight lizards analysed from Almeria, green scales in the dorsal pattern clearly reduce in number or disappear completely near the hind leg insertions and the neck in six individuals, leaving the geometrical figures formed by the combination of dark and green scales faded or absent. The same faded pattern was not observed in lizards from the *L. l. lepida* or *L. l. iberica* subspecies. The partial melanization index resulted in contrasting values for each subspecies: 0.56 in *L. l. nevadensis*, 1.36–1.94 in *L. l. lepida* and 3.33 in *L. l. iberica* (Fig. 2). These values reflect the differences in black scale proportions between subspecies. The *L. l. nevadensis* phenotype corresponds to the lowest proportion of black scales (19.6%), as these are replaced by brown scales (43.7%), whereas in the *L. l. iberica* phenotype, the proportion of black scales reaches the highest value (74.6%; Fig. 3).

Mc1r sequence diversity

Sequences of the *Mc1r* gene of 450 bp were obtained from 30 ocellated lizards from Europe and Africa and

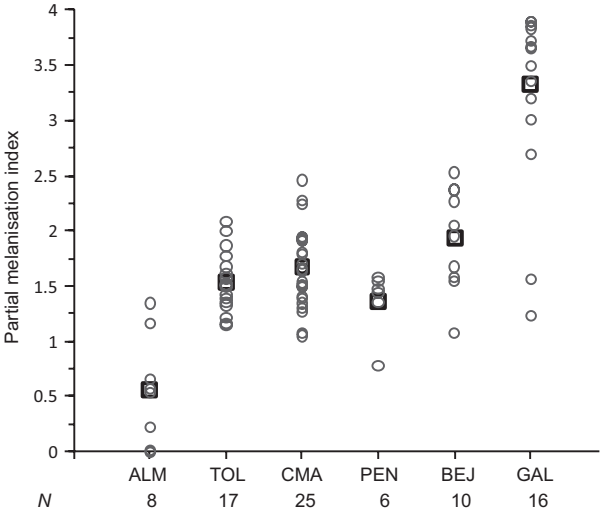


Fig. 2 Partial melanization index (PMI) from six populations of *Lacerta lepida* (N = sample size): Almeria (ALM) from *L. l. nevadensis*, Toledo (TOL), Castro Marim (CMA), Peniche (PEN) and Béjar (BEJ) from *L. l. lepida* and Galicia (GAL) from *L. l. iberica*. Average values of PMI are represented by open squares. The PMI was calculated as the ratio of black scales over light scales (green or yellow) counted in 1 cm^2 of mid-dorsal skin.

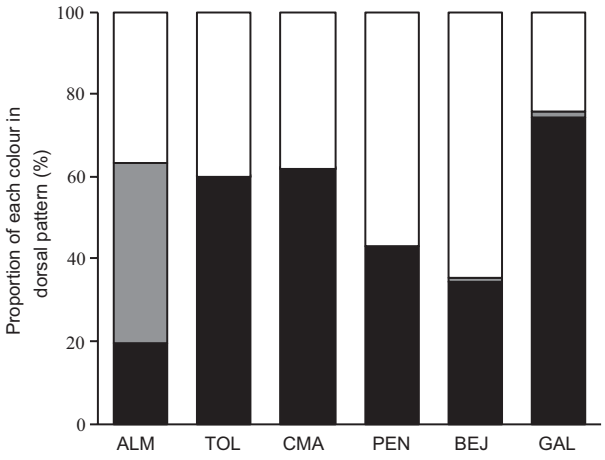


Fig. 3 Average proportion per population of black (in black), brown (in grey) and green/yellow (in white) scales counted in 1 cm^2 of mid-dorsal skin in European ocellated lizards. Each bar represents one population (sample size as in Fig. 2). Almeria (ALM) belongs to *L. l. nevadensis*, Galicia (GAL) is from *L. l. iberica* and the remaining populations are from *L. l. lepida* subspecies.

from another seven lacertid lizards from the Iberian Peninsula. Sequences aligned without gaps from nucleotide 268 to 717 of the 945-bp *Mc1r* gene from little striped whiptail lizard, *Aspidoscelis inornata*, from North America, one of the most closely related species available in GenBank (AY586066) with the full gene sequence. Nucleotide diversity values were similar for the European species *L. lepida* ($\pi = 0.00656$) and the African species *L. pater* ($\pi = 0.00637$) and *L. tangitana* ($\pi = 0.00514$), but when considering each *L. lepida* subspecies alone, *L. l. nevadensis* showed the highest value ($\pi = 0.00333$; Table 1). The highest haplotype diversity was observed in *L. pater* ($H = 0.933$), whereas *L. lepida* showed a lower value ($H = 0.748$), very close to the value observed in *L. l. nevadensis* alone ($H = 0.742$; Table 1).

We observed 13 segregating sites in *Mc1r* from European ocellated lizards, two of them corresponding to nonsynonymous mutations, at positions 485 and 514. The mutation at site 485 was a C-T transition in the second position of codon 162 (T162I) that leads to the replacement of a threonine for an isoleucine residue in the second intracellular protein domain, involving changes in its polarity. A threonine residue at this position was found in all ocellated lizards and other lacertid species investigated, except in *L. l. nevadensis*, where only isoleucine was found (Fig. 4). The second nonsynonymous mutation, at site 514, was located only 29 base pairs away from the mutation T162I and corresponds to an A-T transversion in the first position of codon 172 (S172C), replacing a serine by a cysteine residue. The latter is a conservative amino acid replacement, located in the fourth transmembrane domain of the protein, which follows the second intracellular loop, where mutation T162I is positioned. The mutation S172C occurs in a remarkably conserved position in reptiles, as most of the species with *Mc1r* sequences available to date share a serine residue whereas *L. l. iberica* and *L. l. lepida* have a derived cysteine residue (Fig. 4).

Besides nonsynonymous mutations, three other synonymous changes at positions 351, 411 and 477 were also fixed in European subspecies, separating *L. l. nevadensis* from the other two subspecies (Fig. 5). All five mutations segregate in complete linkage disequilibrium ($N = 46$, Fisher's exact test, $P < 0.001$ with Bonferroni correction). However, *L. l. lepida* and *L. l. iberica* cannot be distinguished from each other by any fixed mutation. Although six haplotypes were found among these two subspecies, five have low frequencies (3–15%) and most of them differ by one synonymous mutation alone from the most frequent haplotype. By contrast, *Lacerta pater* (from Tunisia, Africa) showed a high number of haplotypes, especially considering the small sample size (four lizards), and all haplotypes were unique to the species (Fig. 5). Finally, the main haplotype in *Lacerta tangitana* (from Morocco) differs from *L. l. lepida* by five mutations, including the S172C mutation. However, one heterozygous *L. tangitana* lizard was detected with the most common haplotype of *L. l. lepida* and another allele differing from the previous one by a single mutation (Fig. 5). With the exception of this sample, the cysteine residue in mutation S172C (site 514) was exclusively found in Europe, in both *L. l. lepida* and *L. l. iberica* subspecies, where it was fixed in our sample.

No signature of selection was detected with the McDonald–Kreitman test. For Tajima's *D* test, values were slightly positive for all ocellated lizard species or

Table 1 Number of alleles, haplotypes, segregating (Seg.) sites, synonymous (Syn.) and nonsynonymous (Nonsyn.) substitutions in each ocellated lizard species and subspecies. Haplotype diversity (H), nucleotide diversity (π) and Tajimas's D test values are also presented.

Species	No. of alleles	No. of haplotypes	H	Seg. sites	Syn. sub	Nonsyn. sub	π	Tajima's D
Lacerta lepida	46	10	0.748	13	11	2	0.00656	
L. l. iberica	8	2	0.429	1	1	0	0.00095	0.33350
L. l. lepida	26	5	0.594	4	4	0	0.00165	−0.78167
L. l. nevadensis	12	4	0.742	4	4	0	0.00333	0.46585
Lacerta pater	6	5	0.933	6	5	1	0.00637	0.52043
Lacerta tangitana	8	3	0.464	6	4	2	0.00532	0.15875

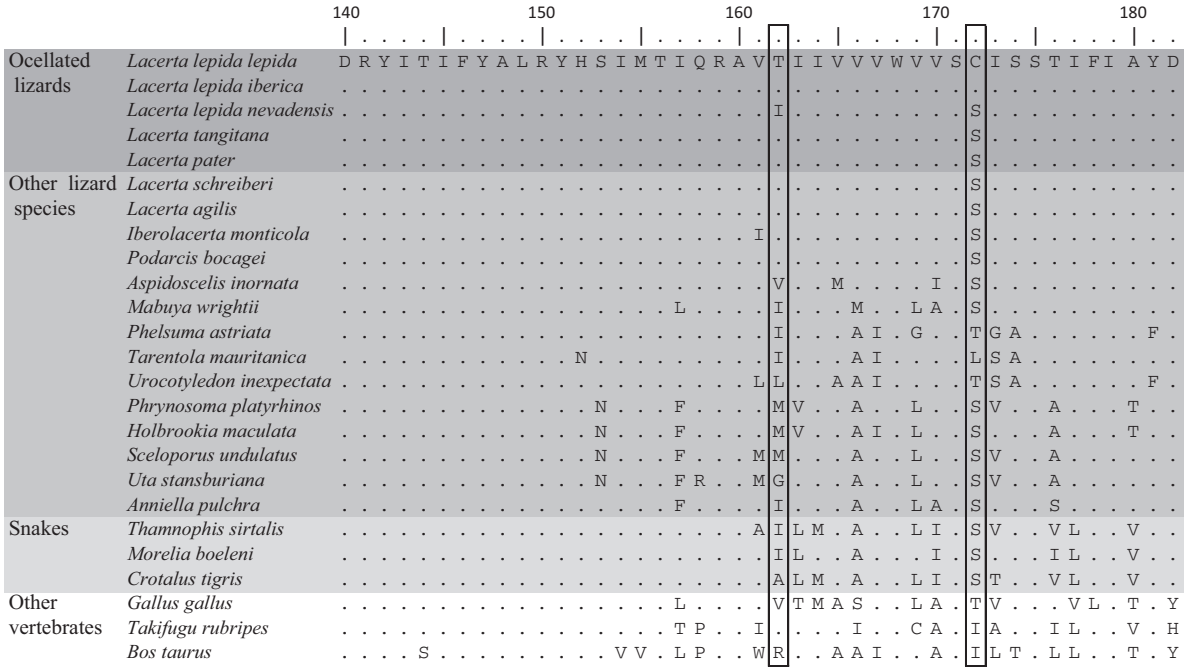


Fig. 4 Partial alignment of *Melanocortin-1 receptor* amino acid chain from ocellated lizards (*Lacerta lepida*, *L. tangitana* and *L. pater*) with other vertebrates, mainly reptile species. The positions where amino acid changes were detected in *L. lepida* are highlighted with boxes (positions 162 and 172).

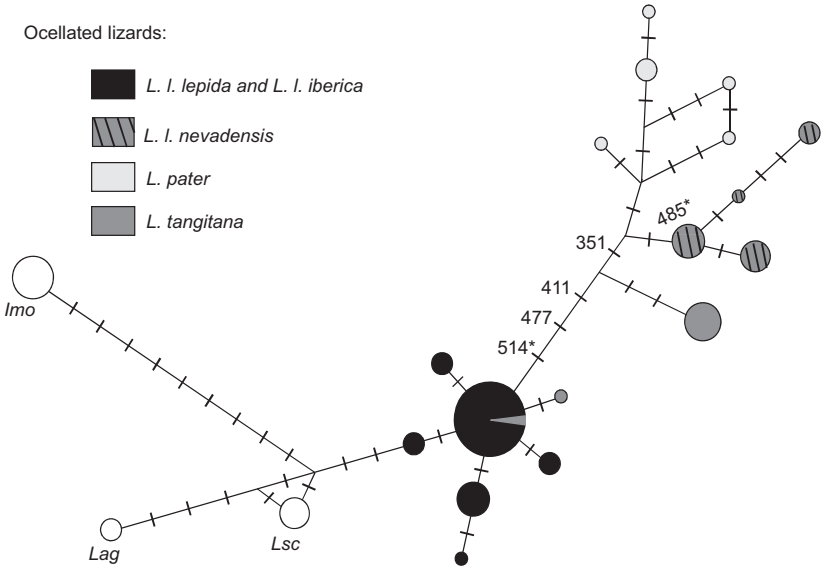


Fig. 5 Minimum spanning haplotype network for the *Melanocortin-1 receptor* gene from 37 lizards. The size of the circles is proportional to sample size. Each mutation between haplotypes is represented by a dash. Site number is indicated for mutations that are fixed for *Lacerta lepida nevadensis* and for *L. l. lepida* and *L. l. iberica* phenotypes. Asterisks denote nonsynonymous changes. Nonocellated lizard haplotypes are represented in white: *L. agilis* (*Lag*), *Iberolacerta monticola* (*Imo*) and *L. shreiberi* (*Lsc*).

subspecies, except for *L. l. lepida* (Table 1), but no value was significantly different from zero.

Association of *Mc1r* with phenotype

Association tests were conducted between *Mc1r* and European ocellated lizard phenotypes. As *L. l. iberica* only differs from *L. l. lepida* phenotypically in the relative proportion of black scales and no fixed mutations in *Mc1r* were found between them, phenotypes were grouped as '*Nev*' (*L. l. nevadensis*) or '*non-Nev*' (*L. l. iberica* and *L. l. lepida*) for association tests. Only three *Mc1r* SNPs were considered, corresponding to the nonsynonymous mutations at sites 485 (T162I) and 514

(S172C), as well as one closely located synonymous SNP at site 477, in linkage disequilibrium with the previous ones. Samples of *L. l. nevadensis* (*Nev*) shared the homozygous genotype for the three SNPs, AA (477)-TT (485)-AA (514), whereas the remaining *L. lepida* samples (non-*Nev*) exhibited the alternative homozygous genotype, GG-CC-TT (Fig. 1), leading to a significant association of phenotypes with *Mc1r* genotype ($N = 160$, Fisher's exact test, $P = 2.2e^{-16}$). Despite the extensive sampling near the contact zone between *L. l. lepida* and *L. l. nevadensis* (Fig. 1b), only two heterozygous individuals were detected, one from population L5 and another from N4. The forward and reverse sequences of these individuals confirmed the heterozygote state at the five polymorphic positions found to be fixed in each subspecies (accession numbers JF732967 and JF732968). The allelic phases for these two hybrids most likely correspond to a haplotype from *L. l. lepida* and another from *L. l. nevadensis*. Inferences about the dominance effect of each allele on *L. l. nevadensis* or *L. l. lepida* phenotypes were not possible due to lack of detailed information on colour traits for individuals sampled across the contact zone.

Discussion

The three subspecies of ocellated lizards with parapatric distributions in the Iberian Peninsula exhibit clear differences in melanin-based colour content and distribution over the lizards' dorsum. The degree of subspecific differences in colour is consistent with the level of genetic differentiation measured by mitochondrial DNA, with an estimated divergence time of 9.43 million years (My) for *L. l. nevadensis* but only around 2 My for *L. l. iberica* (Paulo *et al.*, 2008; Miraldo *et al.*, 2011). The most divergent colour phenotype corresponds to *L. l. nevadensis*, with the prevalence of brown over black scales and a clear reduction in green colour content over the dorsal pattern. In the same way, variation at the candidate gene *Mc1r* is in accordance with the higher level of morphological divergence of *L. l. nevadensis*, as compared to the other Iberian subspecies, which are indistinguishable from each other in the *Mc1r* amino acid chain. A derived and nonconservative amino acid change, T162I, was perfectly associated with the *L. l. nevadensis* phenotype in our sample, with an isoleucine residue fixed at this position. A second amino acid change, S172C, segregates in linkage disequilibrium with the previous one, but the serine residue that occupies this position in *L. l. nevadensis* is shared with all other reptiles investigated to date, except for *L. l. lepida* and *L. l. iberica* that have a cysteine residue instead. Therefore, whereas for mutation T162I the isoleucine is associated with the prevalence of brown scales (putatively due to a partial loss of function of the *Mc1r* receptor), in mutation S172C the cysteine residue seems to be associated with a higher melanization (a putative gain of function), as registered in *L. l. lepida* and *L. l. iberica* phenotypes. In northern Africa, only one lizard from *L. tangitana* has a cysteine residue at mutation S172C, but information on this species' colour variation is scarce and unclear, because previous studies focused on pattern rather than colour variability (Mateo, 1988, 1990; Mateo *et al.*, 1996).

No evidence for positive selection was detected in the present *Mc1r* data set. However, the opportunity for detection of selection with the present tests was slim in such short sequences, with only two amino acid changes (see Hughes, 2007). Nevertheless, it has been previously shown that a single amino acid substitution can have a dramatic effect on phenotype and thus have significant adaptive consequences (Hoekstra *et al.*, 2006). The derived haplotype of *L. l. nevadensis* with a isoleucine residue in mutation T162I is a nonconservative change and is exactly the same amino acid change found in association with the blanched phenotype of little striped whiptail lizard, *Aspidoscelis inornata* (mutation T170I), resulting in a partial loss of function (Rosenblum *et al.*, 2010). Although *A. inornata* is not closely related to ocellated lizards, if future functional studies can confirm T162I as a partial loss-of-function mutation in *L. l. nevadensis*, this might represent another convergence example in *Mc1r* evolution. Both mutations are located in the fourth transmembrane domain of the *Mc1r* protein, and implications of such an amino acid replacement for the physical properties of the protein in *L. l. nevadensis* may be similar to the functional consequences seen in *A. inornata*. Some human variants of *Mc1r* associated with red hair and fair skin that result in diminished function of the protein also have mutations that are homologous or in the immediate vicinity of T162I mutation in ocellated lizards (R160W, R162P and R163Q, homologous to amino acid positions 159, 161 and 162 in the *L. lepida Mc1r* sequence, respectively) (Garcia-Borron *et al.*, 2005). Another example is the cavefish *Aystyanax mexicanus*, where mutation R164C (homologous to residue 159 in *L. lepida*) was shown to be functionally responsible for a brown derived phenotype in cave populations (Gross *et al.*, 2009).

Variation at *Mc1r* sequences indicates low gene flow between *L. l. lepida* and *L. l. nevadensis*, as suggested by the presence of five diagnostic mutations, fixed for all haplotypes detected in each subspecies (including the nonsynonymous mutations referred to above) and by the detection of an extremely low number of hybrids across the contact zone. The phenotype of *L. l. nevadensis* seems to have camouflage advantages in the arid landscapes of the south-eastern Iberian Peninsula. However, neither the causal relationship between *Mc1r* variants and phenotype nor the relative contribution of genetic drift and selection in the evolution of the observed colour pattern can be assessed without further investigations into association in segregating populations or into the fitness and functional effects of genes underlying such colour traits.

A long history of divergence between *L. l. nevandensis* and *L. l. lepida* (Paulo *et al.*, 2008; Miraldo *et al.*, 2011), by the combined effect of selection and genetic drift, helps to explain why such a low number of hybrids were detected in the current contact zone. The two hybrids detected in the present study were sampled at the nearest locations to the subspecies' distribution limits (N4 and L5), located just about 25 km apart. Both hybrids inherited mitochondrial haplotypes from the subspecies of their population of origin, on each side of the contact zone, meaning that hybridization probably resulted from copulation of local females by immigrant males. As ocellated lizard males are territorial, even when young (Castilla, 1989), lower-ranking males may be forced to disperse and seek suboptimal territories, as the best ones are already occupied by older and larger males (Paulo, 1988). For populations closer to the contact zone, male-biased dispersal may increase the opportunity for hybridization, but if their different phenotypes have fitness implications, it may prevent further blending of genomes from the two lineages. Furthermore, slight differences in intraspecific signals between subspecies may promote assortative mating, therefore reducing the opportunity for gene flow between *L. l. lepida* and *L. l. nevadensis*. An alternative hypothesis for the predominance of the brown/grey colour in *L. l. nevadensis* would be the effect on colour variation of genes underlying other traits, in which case the fixed differences in colour between the subspecies would be a by-product not a result of direct selection.

The second amino acid replacement, S172C, found in association with the *L. l. lepida* and *L. l. iberica* phenotype of increased melanin content, corresponds to a derived but conservative change and therefore is less likely to have functional implications. Nevertheless, it is located in a remarkably conserved position among reptiles and belongs to the second intracellular loop of the *Mc1r* protein, a region that seems to have an important role in proper coupling, normal processing and intracellular traffic according to results from functional studies on *Mc1r* human variants (Garcia-Borron *et al.*, 2005). Only functional assays can confirm or reject the functional consequences of the S172C replacement. The involvement of other genes in the melanin-based phenotype in ocellated lizards must also be investigated. The lack of association of the *L. l. iberica* phenotype with *Mc1r* variants suggests that increased melanization in this subspecies as compared to the nominal subspecies may result from regulatory mutations or missed structural mutations in *Mc1r* extremities that were not sequenced in this study or in other genes with important functions in pigmentation. Studies in beach mice have shown that the differences in coat colour result mainly from the interaction of two pigmentation genes, *Mc1r* and *Agouti* (Steiner *et al.*, 2007), whereas melanism in deer mice is associated with mutations in *Agouti* but not in *Mc1r* (Kingsley *et al.*, 2009). Likewise, investigations into convergent coat colour polymorphism in gophers and in some populations from rock pocket mice failed to detect associations with *Mc1r* variation (Hoekstra & Nachman, 2003; Wlasiuk & Nachman, 2007), whereas in Soay sheep, light and dark coat colour are associated with mutations in *Tyrp1*, another gene involved in the melanin synthesis pathway (Gratten *et al.*, 2007).

Finally, melanin-based coloration can only explain variation in the complex colour pattern of ocellated lizards to a limited extent. The genetic mechanisms underlying the formation and distribution of other pigment cells responsible for nonmelanin pigments and structural colours in vertebrates, like xanthophores and iridophores, and how they interfere with melanin-based colours in the overall colour pattern remain largely unknown (Hubbard *et al.*, 2010). Further research on the molecular basis of each colour pigment type is much needed to fully address questions related to diversification and evolution of adaptive colour phenotypes in nonmammalian vertebrates.

Acknowledgments

This research was supported by doctoral scholarships from FCT to Nunes VL (SFRH/BD/21306/2005) and to Miraldo A (SFRH/BD/1696/2004) and by the FCT project POCI/BIA/59288/2004 'Process of speciation: accessing the genic view' coordinated by Paulo OS. All samples were collected under appropriate licences in each country. We thank Juan Pablo de la Vega, Luis Garcia, Gabriel Marin, Eugenia Zarza-Franco and Brent Emerson for helping with fieldwork and two anonymous reviewers for their comments on a previous version of this manuscript.

References

Bandelt, H.-J., Forster, P. & Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**: 37–48.

Castilla, A.M. 1989. *Autoecología del lagarto ocelado (Lacerta lepida)*. PhD thesis, Universidad Autónoma de Madrid, Madrid.

Fisher, R.A. 1935. The logic of inductive inference. *J. R. Stat. Soc. Ser. A* **98**: 39–54.

Font, E., Pérez I De Lanuza, G. & Sampedro, C. 2009. Ultraviolet reflectance and cryptic sexual dichromatism in the ocellated lizard, *Lacerta (Timon) lepida* (Squamata: Lacertidae). *Biol. J. Linn. Soc.* **97**: 766–780.

Garcia-Borron, J.C., Sanchez-Laorden, B.L. & Jimenez-Cervantes, C. 2005. Melanocortin-1 receptor structure and functional regulation. *Pigment Cell Res.* **18**: 393–410.

Gompel, N. & Prud'homme, B. 2009. The causes of repeated genetic evolution. *Dev. Biol.* **332**: 36–47.

Gratten, J., Beraldi, D., Lowder, B.V., McRae, A.F., Visscher, P.M., Pemberton, J.M. *et al.* 2007. Compelling evidence that a single nucleotide substitution in *TYRP1* is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proc. R. Soc. B Biol. Sci.* **274**: 619–626.

Grether, G.F., Kolluru, G.R. & Nersissian, K. 2004. Individual colour patches as multicomponent signals. *Biol. Rev.* **79**: 583–610.

Gross, J.B., Borowsky, R. & Tabin, C.J. 2009. A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Gen.* **5**: e1000326.

Hodar, J.A., Campos, F. & Rosales, B.A. 1996. Trophic ecology of the ocellated lizard *Lacerta lepida* in an arid zone of southern Spain: relationships with availability and daily activity of prey. *J. Arid Environ.* **33**: 95–107.

Hoekstra, H.E. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**: 222–234.

Hoekstra, H.E. & Nachman, M. 2003. Different genes underlie adaptive melanism in different populations of rock pocket mice. *Mol. Ecol.* **12**: 1185–1194.

Hoekstra, H.E., Hirschmann, R.J., Bundey, R.A., Insel, P.A. & Crossland, J.P. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101–104.

Hubbard, J.K., Uy, J.A.C., Hauber, M.E., Hoekstra, H.E. & Safran, R.J. 2010. Vertebrate pigmentation: from underlying genes to adaptive function. *Trends Genet.* **26**: 231–239.

Hughes, A.L. 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* **99**: 364–373.

Kingsley, E.P., Manceau, M., Wiley, C.D. & Hoekstra, H.E. 2009. Melanism in *Peromyscus* is caused by independent mutations in Agouti. *PLoS ONE* **4**: e6435.

Librado, P. & Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Manceau, M., Domingues, V.S., Linnen, C.R., Rosenblum, E.B. & Hoekstra, H.E. 2010. Convergence in pigmentation at multiple levels: mutations, genes and function. *Philos. Trans. R. Soc. B* **365**: 2439–2450.

Mateo, J.A. 1988. *Estudio sistemático y zoogeográfico de los lagartos ocelados, Lacerta lepida (Daudin, 1802) y Lacerta pater (Lataste 1880) (Sauria, Lacertidae)*. PhD thesis, University of Sevilla, Sevilla.

Mateo, J.A. 1990. Taxonomy and evolution of the North African ocellated lizard, *Lacerta pater* (Lataste, 1880) (Sauria: Lacertidae). *Bonn. Zool. Beitr.* **41**: 203–212.

Mateo, J.A. & Castroviejo, J. 1990. Variation morphologique et révision taxonomique de l'espèce *Lacerta lepida* Daudin, 1802 (Sauria, Lacertidae). *Bull. Mus. Natl. Hist. Nat. Paris* **12**: 691–706.

Mateo, J.A. & López-Jurado, L.F. 1994. Variaciones en el color de los lagartos ocelados; aproximación a la dstribución de *Lacerta lepida nevadensis* Buchholz 1963. *Rev. Esp. Herpetol.* **8**: 29–35.

Mateo, J.A., Lopez-Jurado, L.F. & Guillaume, C.P. 1996. Proteic and morphological variations in ocellated lizards (Lacertidae): a complex of species across the Strait of Gibraltar. *C. R. Acad. Sci. III* **319**: 737–746.

Mateo, J.A., Lopez-Jurado, L.F., Cano, J. & Guillaume, C.P. 1999. Intraspecific variability in nucleolar organiser position in *Lacerta (Timon) lepida*. *Amphib. Reptil.* **20**: 73–77.

Mayer, W. & Bischoff, W. 1996. Beiträge zur taxonomischen revision der gattung *Lacerta* (Reptilia: Lacertidae) teil 1: Zootoca, Omanosaura, Timon und Teira als eigenständige gattungen. *Salamandra* **32**: 163–170.

McDonald, J.H. & Kreitman, M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**: 652–654.

Miraldo, A. 2009. *Phylogeography and population dynamics of secondary contact zones of Lacerta lepida in the Iberian Peninsula*. PhD thesis, University of East Anglia, Norwich.

Miraldo, A., Hewitt, G.M., Paulo, O.S. & Emerson, B.C. 2011. Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evol. Biol.* **11**: 170.

Moreira, P.L., Nunes, V.L., Martín, J. & Paulo, O.S. 2007. Copulatory plugs do not assure high first male fertilisation success: sperm displacement in a lizard. *Behav. Ecol. Sociobiol.* **62**: 281–288.

Mundy, N.I. 2005. A window on the genetics of evolution: *Mc1r* and plumage colouration in birds. *Proc. R. Soc. B Biol. Sci.* **272**: 1633–1640.

Paulo, O.S. 1988. *Estudo eco-etológico da população de Lacerta lepida (Daudin 1802) (Sauria, Lacertidae) da ilha da Berlenga*. Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Paulo, O.S., Dias, C., Bruford, M.W., Jordan, W.C. & Nichols, R.A. 2001. The persistence of pliocene populations through the pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. *Proc. R. Soc. Lond. B Biol. Sci.* **268**: 1625–1630.

Paulo, O.S., Pinheiro, J., Miraldo, A., Bruford, M.W., Jordan, W.C. & Nichols, R.A. 2008. The role of vicariance vs. dispersal in shaping genetic patterns in ocellated lizard species in the western Mediterranean. *Mol. Ecol.* **17**: 1535–1551.

Pina-Martins, F. & Paulo, O.S. 2008. Concatenator: sequence data matrices handling made easy. *Mol. Ecol. Resour.* **8**: 1254–1255.

Protas, M.E. & Patel, N.H. 2008. Evolution of coloration patterns. *Ann. Rev. Cell Dev. Biol.* **24**: 425–446.

Raia, P., Guarino, F.M., Turano, M., Polese, G., Rippa, D., Carotenuto, F. *et al.* 2010. The blue lizard spandrel and the island syndrome. *BMC Evol. Biol.* **10**: 289.

Rosenblum, E.B., Hoekstra, H.E. & Nachman, M.W. 2004. Adaptive reptile color variation and the evolution of the Mc1r gene. *Evolution* **58**: 1794–1808.

Rosenblum, E.B., Römpler, H., Schöneberg, T. & Hoekstra, H.E. 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proc. Natl Acad. Sci. USA* **107**: 2113–2117.

Roulin, A. 2004. The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. *Biol. Rev.* **79**: 815–848.

Steiner, C.C., Weber, J.N. & Hoekstra, H.E. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* **5**: 1880–1889.

Stephens, M. & Scheet, P. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* **76**: 449–462.

Stephens, M., Smith, N.J. & Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**: 978–989.

Tajima, F. 1989. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.

Wlasiuk, G. & Nachman, M.W. 2007. The genetics of adaptive coat color in gophers: coding variation at *Mc1r* is not responsible for dorsal color differences. *J. Hered.* **98**: 567–574.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Dorsal colour pattern of European ocellated lizard subspecies.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 6 May 2011; revised 22 June 2011; accepted 23 June 2011

CHAPTER 4

Supporting information

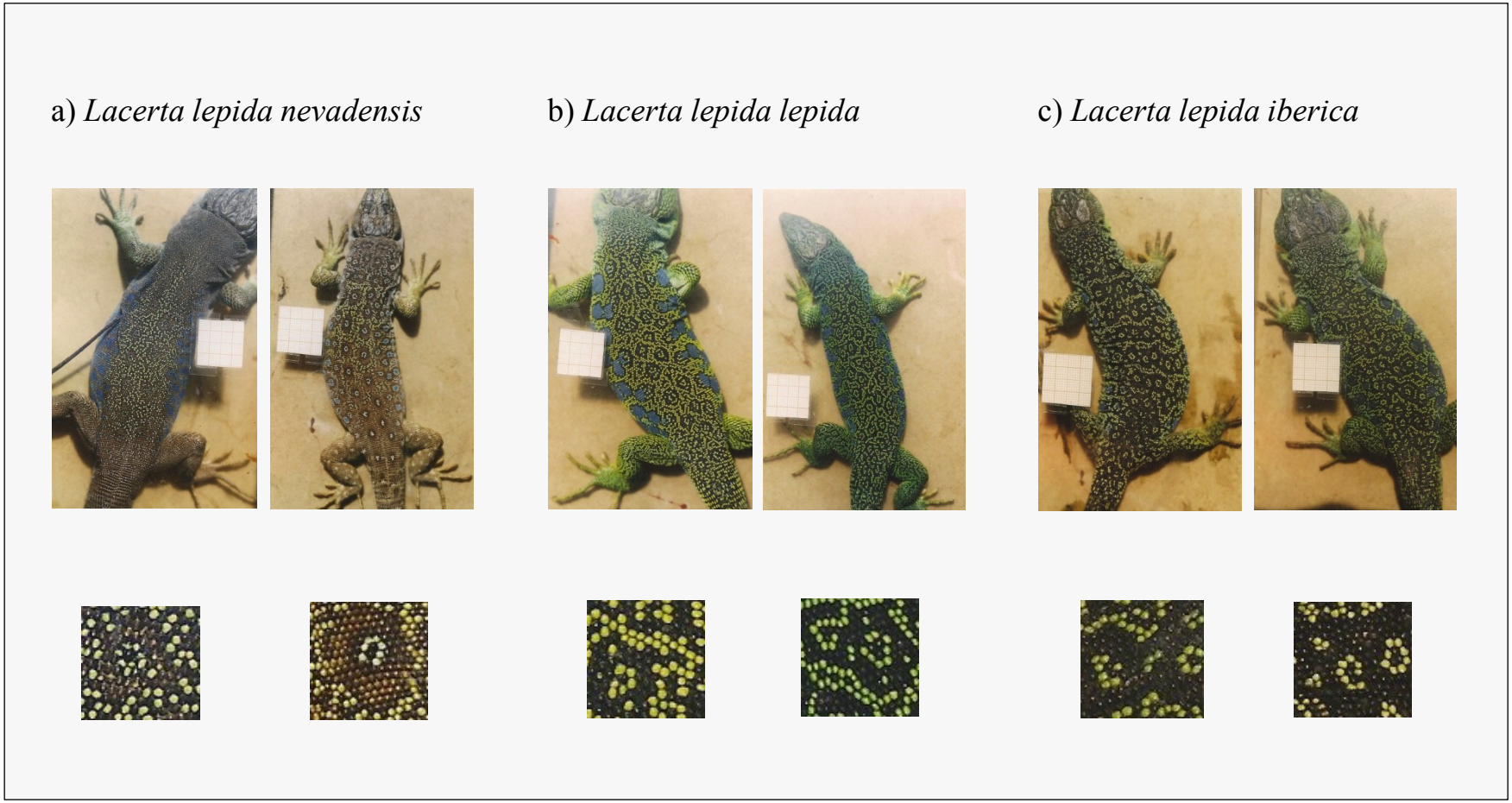


Fig. S1 Dorsal colour pattern of european ocellated lizard subspecies. (a) *Lacerta lepida nevadensis,* from Almeria, in the southeast of the Iberian Peninsula; (b) *Lacerta lepida lepida,* from Toledo, in the center of the Iberian Peninsula and (c) *Lacerta lepida iberica,* from Galicia, in the northwest of the Iberian Peninsula. A male and a female of each subspecies is presented. For each lizard, one cm^2 of the mid-dorsal region is magnified to show in detail the colour of dorsal scales.

CHAPTER 5

Analysis of neutral versus non-neutral nuclear loci provides evidence for incipient ecological speciation within European ocellated lizards, *Lacerta lepida*

Vera L. Nunes[1], Mark A. Beaumont[2,3], Roger K. Butlin[4], Octávio S. Paulo[1]

[1] Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, DBA/FCUL, 1749-016 Lisboa, Portugal

[2] School of Animal and Microbial Sciences, University of Reading, P.O. Box 228, Reading, RG6 6AJ, United Kingdom

[3] Schools of Mathematics and Biological Sciences, University of Bristol, UK

[4] Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, England

In prep.

Abstract

European ocellated lizards survived the dramatic climatic oscillations of the Quaternary, probably within multiple refugia, but are presently distributed all over the Iberian Peninsula. Three subspecies with parapatric distributions are presently recognized and exhibit morphological variation that is associated with environmental heterogeneity. Here we compare population genetic structure in *Lacerta lepida* as inferred from 8 STRs, 318 neutral AFLPs and 23 outlier AFLPs (suspected to be under the influence of directional selection). *L. l. nevadensis* divergence is well supported by both neutral and non-neutral loci, confirming that this subspecies, which has been evolving in the southeast of the Iberian Peninsula, is in the final stages of its speciation process. Within the nominal subspecies, *L. l. lepida*, which occupies most of the Iberian Peninsula, it is possible to recognize a weak substructure with STRs but not with AFLPs. Nevertheless, the clades inferred with mitochondrial data within *L. l. lepida* in previous studies are not fully supported with STRs. The third subspecies, *L. l. iberica*, cannot be distinguished from the nominal subspecies with neutral AFLP markers but both STRs and outlier AFLPs support its subspecific status. While *L. l. iberica* divergence in STRs could be explained by genetic drift (with overall reduction in allelic richness and heterozygosity), divergence at putatively adaptive loci indicates that *L. l. iberica* might be at the early stages of ecological speciation, where loci with adaptive advantages in the northwest of the Iberian Peninsula are being selected in spite of gene flow among neutral loci.

Keywords: AFLPs, hierarchical structure, local adaptation, neutrality, selection, STRs

Introduction

The climatic oscillations of the Quaternary had an important role in the evolutionary history of many species that persist today in Europe. Phylogeographic studies in a wide range of plant and animal species suggest that some southern areas of the European continent served as important refugia during glacial periods (Hewitt 1996, 1999; Taberlet *et al.* 1998). Surviving populations were able to expand and disperse under favourable conditions during interglacial periods, re-establishing gene flow between divergent gene pools in secondary contact zones. The present patterns of population genetic structure and diversity were necessarily affected by the repeated episodes of species range expansion and contraction promoted by the glacial cycles. While some populations or lineages became isolated and lost genetic variability due to drift or went extinct during range contraction, adaptive mutations could have increased in frequency as a result of selection in refugial populations and spread during range expansions (Hewitt 2004).

The Iberian Peninsula was one of the most important glacial refuge areas in southern Europe for several species. Its unique topography and climatic features resulted in a patchy landscape and offered favourable conditions for the persistence of species in several separate regions of Iberia, promoting population divergence and speciation in allopatry (Gómez & Lunt 2007; Schmitt 2007). High endemism and complex geographic structure have been reported in many Iberian species (Gómez & Lunt 2007; Feliner 2011). Traditionally, phylogeographic studies have been based on mitochondrial genes and a few nuclear genes. However, information provided by mitochondrial genes often does not agree with patterns obtained from nuclear data because mitochondrial DNA is maternally inherited as opposed to nuclear genes, which are biparentaly inherited, implying the need for much more time to reach complete lineage sorting in the latter. The use of a few nuclear genes is probably not representative of the whole genome diversity and a multilocus sampling strategy should be preferred. The combined use of different types of markers with different mutation rates can also enrich the detail and robustness of phylogeographic studies in species with a complex evolutionary history (Godinho *et al.* 2008; Sequeira *et al.* 2008).

Microsatellites (or STRs) and amplified fragment length polymorphisms (AFLPs) have been widely used as nuclear markers for multilocus analyses in phylogeographic studies. Microsatellites are highly polymorphic markers due to their multiallelic and

codominant nature, and for that reason, even a small number of loci can be highly informative to infer genetic diversity and differentiation among populations. As for AFLPs, they are less informative because they behave as dominant markers and are scored as biallelic, according to the presence or absence of fragments with a specific size. While dominance can be a drawback for AFLPs in genetic analyses, the affordability of the technique allows the generation and scoring of several hundreds of markers randomly distributed across the genome with the same effort and a similar cost as a few dozens of STRs (Bensch & Akesson 2005). The development of AFLP markers for species with little genetic information available is technically straightforward when compared to the development of new STRs. The use of STRs developed for closely-related species is also often preferred to the development of new species-specific STRs.

Simulation studies conducted by Mariette *et al.* (2002) showed that ability of microsatellites to express the genetic variability of populations can vary across different evolutionary scenarios, while AFLPs are more robust and do not exhibit a large range of variation. However, these simulations were limited to fully neutral evolutionary scenarios. Neutral markers have been preferred for genetic variation analysis but the information they provide might be biased or incomplete. When adaptations to different environmental conditions are selected at a local scale, in the absence of physical barriers to gene flow, genetic divergence will be mostly restricted to adaptive loci. In the early stages of ecological divergence, gene flow can still occur freely among neutral regions of the genome, while at more advanced stages of speciation, gene flow becomes restricted until the point where new species with complete reproductive isolation are formed (Wu 2001; Nosil *et al.* 2009; Via 2009). There is a growing interest to complement phylogeographic studies with non-neutral markers and evaluate adaptive genetic variation (Colbeck *et al.* 2011; Kirk & Freeland 2011; Richter-Boix *et al.* 2011). Advances in population genomics made it possible to easily genotype hundreds of loci across the genome and statistical tests to detect loci under selection have become more sophisticated (Beaumont & Nichols 1996; Foll & Gaggiotti 2008; Excoffier *et al.* 2009). It is now possible and advisable to assess and compare the genetic structure of populations with both neutral nuclear loci and loci suspected to be under the effect of selection.

The European ocellated lizard (*Lacerta lepida*; synonym: *Timon lepidus*) is one of the temperate reptile species that experienced the harsh climatic conditions during the glaciations and persisted in the Iberian Peninsula. The species is currently distributed all

over Iberia, and in some parts of southern France and northern Italy. Recent studies demonstrate that the species is highly structured, showing evidence of multiple refugia during the Pleistocene glaciations, followed by recent demographic and spatial expansions (Paulo *et al.* 2008; Miraldo *et al.* 2011). Six distinct mitochondrial lineages (based on cytochrome *b* sequences) were detected across the species' distribution (Paulo 2001; Miraldo *et al.* 2011). Moreover, considerable morphological variation was detected within the species (Mateo & Castroviejo 1990). Variation in colour pattern, body size and dentition was found to be clinal, following an environmental gradient along the Iberian Peninsula, with a northwest-southeast orientation (Mateo 1988; Mateo & Castroviejo 1990; Mateo & López-Jurado 1994; Nunes *et al.* 2011a). Populations at each end of the cline have been recognized as subspecies: *L. lepida iberica* in the northwest and *L. l. nevadensis* in the southeast. A genome scan for selection with AFLP markers was conducted recently on European ocellated lizards, showing that nearly 6% of the investigated loci might be under the effect of directional selection (Nunes *et al.* 2011b). Some of those loci were also associated with climatic variables such as temperature, precipitation or insolation, supporting the importance of the environment in shaping the adaptive variance within the species (Nunes *et al.* 2011b).

Here we investigate the genetic structure of European ocellated lizards in the Iberian Peninsula with two types of nuclear markers: STRs and AFLPs. We compare the use of hundreds of weakly informative AFLPs scattered randomly across the genome with the use of eight highly informative STRs. Additionally, we compare the genetic structure patterns inferred with neutral and non-neutral STR and AFLP markers to disentangle the importance of adaptive loci for the genetic structure of ocellated lizards along the environmental cline.

Materials and Methods

Samples and collection sites

A total of 10 populations of *Lacerta lepida* were sampled in the Iberian Peninsula along a southeast-northwest transect and a north-south transect along the Atlantic coast, covering the distribution of three subspecies (*L. l. iberica*, *L. l. nevadensis* and the nominal subspecies *L. l. lepida*), and all six mitochondrial lineages identified by Miraldo

et al. (2011). Population locations, sample sizes and corresponding mitochondrial DNA clades are listed in Table 1. Tissue samples were collected from the tails of free-living animals that were immediately released back into the wild. Whole genome DNA was extracted with the Jetquick Tissue DNA kit (Genomed).

STRs and AFLPs genotyping

Lacerta lepida populations were genotyped for eight STR loci, all but one of which were isolated and characterized in other Lacertidae species: Pb73 and Pb66 from *Podarcis bocagei* (Pinho *et al.* 2004), B4, C9 and D1 from *Podarcis muralis* (Nembrini & Oppliger 2003), Lv-4-72 from *Lacerta vivipara* (Boudjemadi *et al.* 1999), Lvir17 from *Lacerta viridis* (Böhme *et al.* 2005) and LIZ24 (Paulo, unpublished data), which was isolated from both *Lacerta schreiberi* and *L. lepida* DNA. Amplification of microsatellites was performed in reactions of 10 μL with approximately 50 ng of DNA, 1x PCR buffer (Promega), 0.3 units of *Taq* polymerase (Promega), 2.0 mM $MgCl_2$, 0.15 mM of dNTPs and 0.5 μM of each primer (forward primers labelled with either 6-FAM, NED or HEX fluorescent dyes). The cycling conditions used were 3 min at 94 ºC, 30 x (30 s at 94 ºC, followed by 30 s at locus specific annealing temperature (Table S1, Supporting information), and 30 s at 72 ºC), with a final extension of 45 min at 72 ºC. Two pairs of loci were amplified in multiplex (Pb73 and D1; C9 and Pb66). PCR products were loaded in multiplex (combining 6-FAM, NED and HEX labelled PCR products) in an ABI PRISM 310 (Applied Biosystems) with Genescan-500 ROX as the internal size standard. Allele size was determined with GeneMapper 3.7 (Applied Biosystems).

AFLP markers were developed for the same populations of European ocellated lizards with a modified version of the original protocol from Vos *et al.* (1995) as documented in Nunes *et al.* 2011b, using eigth *Eco*RI-*Mse*I primer combinations (ACA-CAG, AAC-CAC, ACT-CTG, AAC-CTC, ACT-CTT, AAG-CAC, ACA-ACA and AAG-CTA). In total, 392 polymorphic markers were scored in each lizard either as present or absent.

Table 1 Populations from *Lacerta lepida* genotyped for microsatellite loci and AFLP markers. Population code, sample size (N) and geographical coordinates are given for each population, as well as the corresponding subspecies and the mitochondrial DNA clade (mtDNA), as inferred from cytochrome *b* sequences (Cyt *b*) by Miraldo *et al.* (2011).

Population	Code	N	Latitude	Longitude	Subspecies	mtDNA (Cyt *b*) clade
Galicia	GAL	19	43º 21' 60" N	7º 22' 03" W	*L. l. iberica*	L3 (Northern)
Gerês	GER	23	41º 43' 23" N	8º 06' 50" W	*L. l. iberica*	L3 (Northern)
Béjar	BEJ	22	40º 40' 15" N	5º 36' 32" W	*L. l. lepida*	L1 (Central)
Serra da Estrela	SET	22	40º 19' 24" N	7º 36' 44" W	*L. l. lepida*	L5 (Western)
Peniche	SPE	16	39º 19' 41" N	9º 20' 45" W	*L. l. lepida*	L5 (Western)
Castro Marim	CMA	25	37º 14' 08" N	7º 26' 43" W	*L. l. lepida*	L2 (Algarve)
Alentejo	ALE	13	38º 35' 55" N	7º 33' 40" W	*L. l. lepida*	L4 (Southern)
Toledo	TOL	22	39º 15' 32" N	3º 44' 01" W	*L. l. lepida*	L4 (Southern)
Andalucia	AND	16	38º 16' 51" N	3º 37' 02" W	*L. l. lepida*	L4 (Southern)
Almería	ALM	18	36º 49' 54" N	2º 31' 32" W	*L. l. nevadensis*	N (Nevadensis)

Data analysis

For microsatellite loci, MICRO-CHECKER 2.2.3 (van Oosterhout *et al.* 2004) was used to test for the existence of stuttering errors, allele dropout and the presence of null alleles. The number of alleles per locus and population, allele frequencies, and allelic richness (A_R, number of alleles per population standardized for the minimum sample size, which was 13 for the present dataset, El Mousadik & Petit 1996) were determined with FSTAT 2.9.3.2 (Goudet 2001). Expected (He) and observed (Ho) heterozygosity were calculated in ARLEQUIN 3.5 (Excoffier *et al.* 2005). Deviations from Hardy-Weinberg equilibrium (HWE) by heterozygote deficit, using F_{IS} (Weir & Cockerham 1984), were tested in GENEPOP 4.0 (Rousset 2008), as well as tests for linkage disequilibrium (LD) between each pair of loci. All tests were performed with 10 000 dememorization steps, 1000 batches and 10 000 iterations per batch. We applied the Bonferroni correction to all multitest analyses. BayeScan 1.0 (Foll & Gaggiotti 2008) was used to test if any of the microsatellite loci were under the effect of selection in European ocellated lizards. Loci with an associated posterior probability over 0.95 where considered significant. BayeScan assumes that loci affected by directional selection will show larger genetic differentiation than neutral loci while loci under balancing selection will be less differentiated. The

method is effective under different demographic histories and different levels of genetic drift between the populations (Foll & Gaggiotti 2008). For AFLP markers, tests for selection were conducted with BayeScan (Foll & Gaggiotti 2008) and DFDIST (Beaumont & Nichols 1996) and were reported in Nunes *et al.* (2011b). The combined results from both detection methods resulted in a list of 23 markers considered as strong outliers, and therefore as candidate loci potentially under selection (directional selection). Only AFLP loci never detected as outliers in either BayeScan or DFDIST were retained as neutral markers, resulting in a set of 318 loci. The two sets of AFLP markers were analysed here independently to investigate genetic structure in European ocellated lizards.

Genetic differentiation between pairs of populations with STRs was estimated with F_{ST} in ARLEQUIN. Although R_{ST} (Slatkin 1995) is especially suited for STR loci to estimate genetic differentiation, F_{ST} was preferred to R_{ST} in the present study because F_{ST} is more conservative when the number of scored loci is small (less than 20; Gaggiotti *et al.* 1999). For neutral and outlier AFLPs, pairwise F_{ST} was computed with AFLP-SURV 1.0 (Vekemans 2002). Mantel tests were performed to assess the significance of correlations between genetic and geographical distances with IBD 1.52 (Bohonak 2002) using 10000 randomizations. Linearized F_{ST} estimates, $F_{ST}/(1- F_{ST})$, were used as genetic distances (Rousset 1997). Pairwise geographical distances were estimated with the haversine formula in the calculator available at http://www.movable-type.co.uk/scripts/latlong.html, using latitude/ longitude coordinates from sampled locations. The haversine formula estimates the shortest distance over the earth's surface ("as-the-crow-flies") between two points.

Principal coordinates analyses (PCoA) were performed in GENALEX 6.3 (Peakall & Smouse 2005) for each dataset and hierarchical analysis of molecular variance (AMOVA) was performed in ARLEQUIN. Population structure was assessed with STRUCTURE 2.3.3 (Pritchard *et al.* 2000). This program uses Bayesian simulations to estimate the posterior probabilities of assignment of individuals to each of a given number of groups (K). AFLP data input was prepared as recommended in Falush *et al.* (2007) for dominant data, setting the recessive allele option. To determine the best number of clusters, 10 independent runs (burnin period of 10^5 and 10^6 iterations) were performed for $K = 1$ to $K = 11$, assuming the admixture ancestry model and correlated allele frequencies among samples (information about the individuals' sampling sites was not included as a parameter for the simulations). The value of K with the highest ΔK

according to Evanno *et al.* (2005) was assumed to be the most probable number of clusters and was compared with the number of clusters inferred from Ln Pr ($X|K$) values (Pritchard *et al.* 2000). We used DISTRUCT 1.1 (Rosenberg *et al.* 2004) to obtain graphical displays for the best run (with higher Ln Pr ($X|K$)) from the 10 independent runs performed for the most probable number of clusters.

Results

Microsatellite diversity

Ocellated lizards from 10 populations from the Iberian Peninsula were successfully genotyped for eight microsatellite loci. The average number of alleles per locus detected among the 196 individuals analysed was 17.4, varying from six alleles in locus LIZ24 to 23 alleles in locus D1 (Table S1, Supporting information). Observed heterozygosity (H_O) averaged over all loci was only 0.35 in the GAL population but ranged from 0.60 to 0.77 in the remaining populations. The GAL population, located near the northwest limit of the ocellated lizard distribution in the Iberian Peninsula, was also the one with the lowest values of allelic richness (1.00-5.65) while the highest values (9.74-14.00) were found mostly in populations from the south of the Iberian Peninsula (populations AND, ALM and ALE; Table 2). Most alleles detected in *L. l. iberica* are also present in *L. l. lepida* populations. Only two alleles at low frequency detected in the GER population were unique to *L. l. iberica*. *L. l. lepida* showed a broader range of allele size in most loci when compared to the other susbspecies (Fig. S1, Supporting information), with 22 alleles that were exclusive to *L. l. lepida* populations, among which AND, SET, ALE and SPE show the highest number of unique alleles (4-6 alleles). Although ALM was the only sampled population of *L. l. nevadensis*, it exhibited a remarkable number of unique alleles (22 alleles). In locus C9 in particular, six of the nine alleles detected in *L. l. nevadensis* were unique to this subspecies (Fig. S1, Supporting information). These are also the alleles with the largest sizes observed for locus C9 in all genotyped lizards. Additionally, both *L. l. iberica* populations (GAL and GER) were monomorphic for locus C9, exhibiting solely the smallest sized allele observed in this locus (Fig. S1, Supporting information).

Tests of HWE by heterozygote deficit (F_{IS}) were significant for the ALE population (loci Lvir17 and Pb73) and for the AND population (locus Lv-4-72) (Table 2). Null

Table 2 Allelic richness (A_R), expected (H_E) and observed (H_O) heterozygosity and inbreeding coefficient (F_{IS}) for each STR locus and population. Statistically significant values of F_{IS} are marked with an asterisk and shaded in grey (P<0.00625 after Bonferroni correction).

Locus	B4				C9				D1			
Pop.	A_R	H_E	H_O	F_{IS}	A_R	H_E	H_O	F_{IS}	A_R	H_E	H_O	F_{IS}
GAL	3.81	0.587	0.474	0.198	1.00	__	__	__	4.97	0.744	0.737	0.010
GER	7.07	0.793	0.826	-0.042	1.00	__	__	__	9.15	0.884	0.783	0.117
BEJ	8.41	0.822	0.727	0.118	3.53	0.388	0.364	0.064	11.32	0.923	0.955	-0.035
SET	9.18	0.815	0.864	-0.061	4.12	0.450	0.364	0.196	10.15	0.884	0.955	-0.082
SPE	5.75	0.704	0.750	-0.068	4.00	0.704	0.625	0.115	11.15	0.901	0.938	-0.042
CMA	9.12	0.842	0.960	-0.144	4.21	0.356	0.360	-0.012	8.89	0.869	0.880	-0.013
ALE	10.00	0.874	0.692	0.215	4.00	0.348	0.308	0.119	11.00	0.926	0.923	0.004
TOL	8.98	0.860	0.864	-0.004	3.59	0.549	0.636	-0.164	10.73	0.888	1.000	-0.130
AND	11.78	0.903	0.875	0.032	3.94	0.427	0.250	0.423	13.59	0.938	0.750	0.205
ALM	5.79	0.767	0.667	0.134	8.57	0.889	0.889	0.000	11.22	0.903	0.889	0.016

Locus	LIZ24				Lvir17				Lv-4-72			
Pop.	A_R	H_E	H_O	F_{IS}	A_R	H_E	H_O	F_{IS}	A_R	H_E	H_O	F_{IS}
GAL	1.00	__	__	__	5.65	0.781	0.722	0.077	2.99	0.569	0.684	-0.209
GER	1.97	0.162	0.087	0.470	7.68	0.831	0.913	-0.101	3.82	0.656	0.826	-0.267
BEJ	2.97	0.354	0.318	0.104	6.27	0.789	0.818	-0.039	8.16	0.848	0.773	0.090
SET	2.58	0.246	0.273	-0.110	8.78	0.870	0.727	0.167	9.32	0.904	0.818	0.097
SPE	3.97	0.599	0.625	-0.045	8.37	0.780	0.875	-0.126	7.69	0.813	0.688	0.158
CMA	1.52	0.040	0.040	__	6.00	0.801	0.640	0.204	7.66	0.845	0.800	0.054
ALE	4.00	0.443	0.385	0.137	14.00	0.942	0.692	0.273*	10.00	0.880	0.846	0.040
TOL	3.96	0.292	0.318	-0.093	5.72	0.776	0.818	-0.056	7.89	0.853	0.818	0.042
AND	4.78	0.514	0.375	0.277	9.47	0.871	0.750	0.143	8.43	0.869	0.563	0.360*
ALM	1.72	0.056	0.056	__	4.44	0.657	0.889	-0.367	11.43	0.911	0.778	0.150

Locus	Pb66				Pb73			
Pop.	A_R	H_E	H_O	F_{IS}	A_R	H_E	H_O	F_{IS}
GAL	2.37	0.104	0.105	-0.014	2.37	0.104	0.105	-0.014
GER	5.51	0.729	0.739	-0.015	4.92	0.770	0.652	0.156
BEJ	7.91	0.844	0.818	0.031	7.22	0.797	0.727	0.089
SET	7.79	0.835	0.818	0.021	7.94	0.840	0.864	-0.028
SPE	5.75	0.776	0.813	-0.048	8.87	0.875	0.875	0.000
CMA	8.24	0.856	0.880	-0.028	5.53	0.764	0.880	-0.155
ALE	10.00	0.917	0.923	-0.007	9.00	0.883	0.615	0.312*
TOL	9.56	0.883	0.773	0.127	7.33	0.807	0.636	0.215
AND	12.59	0.905	0.750	0.176	8.71	0.875	0.688	0.220
ALM	9.49	0.784	0.833	-0.065	9.74	0.865	0.722	0.169

alleles were detected in populations ALE (loci Lvir17 and Pb73) and AND (loci C9, D1 and Lv-4-72), with low frequencies, ranging between 0.08-0.15 according to the *Brookfield 1* equation (Brookfield 1996). Individuals sampled at ALE and AND were collected in a larger geographical area than the other populations and it is possible that some of those individuals do not belong to the same interbreeding population. Therefore, deviations at HWE and excess of homozygotes might result from a broader sampling area for these populations. No significant LD was detected among microsatellite loci.

Tests for selection with BayeScan indicated that locus C9 seems to be under strong directional selection while locus D1 appears under the effect of balancing selection. To ensure that selection tests were not affected by hierarchical genetic structure within the data (see Excoffier *et al.* 2009), BayeScan analysis was repeated after the exclusion of *L. l. nevadensis* (ALM) and also after the exclusion of both *L. l. nevadensis* (ALM) and *L. l. iberica* populations (GAL and GER) (Table 3). In the analysis without *L. l. nevadensis*, loci C9 and D1 maintain the same selection signature. When performing the analysis without *L. l. nevadensis* and *L. l. iberica* populations, none of the microsatellites showed strong selection signatures. However, when pairwise analyses were performed between all seven *L. l. lepida* populations, locus C9 appeared to be under strong directional selection (with posterior probability > 0.95) in 19 of the 21 population pairs, while only three pairs indicated that locus D1 is affected by balancing selection and another three pairs showed that locus LIZ24 is under directional selection.

Genetic structure: STRs vs. AFLPs, neutral vs. non-neutral loci

All types of analysed nuclear markers were in agreement about the population pair with the highest divergence. This pair consists of ALM, from *L. l. nevadensis*, and GAL, from *L. l. iberica* (F_{ST} = 0.33 with 8 STRs and F_{ST} = 0.32 if loci D1 and C9 are removed, F_{ST} = 0.16 with 318 neutral AFLPs, F_{ST} = 0.71 with 23 outlier AFLPs). As these populations are geographically the farthest apart in this study (835 Km), and also belong to different subspecies, we expect a reduction in gene flow caused not only by distance but also by reproductive isolation between subspecies. Therefore, besides testing for IBD among the full dataset of 10 populations with a Mantel test, we performed the test for the nominal subspecies populations alone, to discard biases caused by the high level of divergence of *L. l. nevadensis* and *L. l. iberica* subspecies. The results of the Mantel test

were significant (P<0.05) for all sets of markers tested (r = 0.55 with 8 STRs, r = 0.59 with 318 neutral AFLPs and r = 0.65 with 23 outlier AFLPs, Table 4). When testing for *L. l. lepida* populations alone, a trend of IBD was present but was significant only for STR loci (r = 0.67, Table 4).

Table 3 Results from tests for selection for microsatellite loci with BayeScan. Three analyses were conducted. First, the analysis was performed for all populations and then repeated without ALM population (*L. l. nevadensis*) and without populations from *L. l. iberica* (GAL and GER) and from *L. l. nevadensis* (ALM). Loci D1 and C9 revealed significant selection signatures, with a posterior probability over 0.95 (values in bold).

Dataset	Locus	Probability	log10(BF)	Alpha	F_{ST}	Selection
10 populations	Pb73	0.227	-0.532	-0.094	0.115	
(all populations are included)	D1	**0.953**	1.307	-0.663	0.073	balancing
	C9	**1.000**	3.699	1.279	0.288	directional
	Pb66	0.207	-0.584	-0.060	0.117	
	LV-4-72	0.249	-0.480	-0.161	0.110	
	LIZ24	0.842	0.727	0.837	0.221	
	B4	0.358	-0.254	0.303	0.143	
	Lvir17	0.427	-0.128	0.347	0.150	
9 populations	Pb73	0.237	-0.509	0.069	0.127	
(without ALM)	D1	**0.956**	1.333	-0.710	0.071	balancing
	C9	**0.995**	2.335	1.243	0.276	directional
	Pb66	0.225	-0.537	-0.089	0.116	
	LV-4-72	0.234	-0.514	0.126	0.131	
	LIZ24	0.824	0.670	0.833	0.216	
	B4	0.229	-0.528	0.166	0.133	
	Lvir17	0.238	-0.506	-0.125	0.113	
7 populations (without ALM, GAL and GER)	Pb73	0.250	-0.476	0.004	0.064	
	D1	0.743	0.462	-0.586	0.039	
	C9	0.828	0.683	0.885	0.142	
	Pb66	0.295	-0.379	-0.225	0.055	
	LV-4-72	0.279	-0.412	-0.185	0.056	
	LIZ24	0.683	0.333	0.756	0.127	
	B4	0.371	-0.228	0.326	0.079	
	Lvir17	0.351	-0.267	0.302	0.080	

Table 4 Mantel tests to detect isolation by distance with data from 8 STR loci, 318 neutral AFLP markers and 23 outlier AFLPs. Significant *P*-values are shown in bold. The analysis was performed with all populations and repeated with populations from *L. l. lepida* subspecies only.

	Markers	R^2	r	*P*-value	Intercept	Slope
L. lepida	8 STRs	0.41	0.64	**0.0007**	-0.117	$6.532\ e^{-4}$
(10 pop.)	318 AFLPs	0.34	0.59	**0.0027**	-0.044	$2.675\ e^{-4}$
	23 AFLPs	0.42	0.65	**0.0001**	-0.876	$3.559\ e^{-3}$
L. l. lepida	8 STRs	0.18	0.43	**0.0017**	-0.012	$2.164\ e^{-4}$
(7 pop.)	318 AFLPs	$1.99\ e^{-3}$	0.04	0.3957	$-3.265\ e^{-3}$	$1.178\ e^{-4}$
	23 AFLPs	0.07	0.27	0.0930	-0.076	$5.229\ e^{-4}$

The principal coordinates analysis (PCoA) performed with each set of nuclear markers showed individuals from ALM (*L. l. nevadensis*) to form a well-defined group. The distinction was best defined by the 23 outlier AFLPs (Fig. 1), with the first two axes explaining 62.1% of the variance. This value dropped to 49% for STRs and for the 318 neutral AFLPs. Individuals from GAL (*L. l. iberica*) were also well separated from *L. l. lepida* samples according to STRs and the 23 outlier AFLPs, but many individuals from GER, which belongs to *L. l. iberica* as well, were located among *L. l. lepida* individuals in the PCoA. The set of 318 neutral AFLPs had much less resolution for the PCoA analysis as compared to the other sets of nuclear markers, since not even GAL individuals could be distinguished easily from all other *L. l. lepida* samples, which were tightly clustered (Fig. 1)

Results from analyses of molecular variance (AMOVAs) are presented in Table 5. For all three sets of nuclear markers, the percentage of variance among groups was higher when populations were divided into two groups, [*L. l. nevadensis*] and [*L. l. lepida* + *L. l. iberica*], followed by the analyses with three groups corresponding to each subspecies. These results suggest the existence of a hierarchical structure, where *L. l. nevadensis* is the most divergent group, followed by *L. l. iberica*. AMOVA results did not support the existence of structured groups within *L. l. lepida*, since the percentage of variance among groups of populations from each mitochondrial clade (L1, L2, L4 or L5; see Table 1) was lower than the percentage of variance among populations within the same clade.

Analyses of population structure with STRUCTURE for the eight STR loci showed that $K = 3$ was the most probable number of clusters, where each cluster corresponded to a subspecies (Fig. 2). Simulations performed with *L. l. lepida* populations alone suggested K = 4, which is the number of clusters expected according to the number of mitochondrial clades recognized within the subspecies (L1, L2, L4 and L5), but the clustering pattern was weakly supported due to extensive admixture between groups (Fig. S2, Supporting information). However, to ensure that the clustering was not strongly biased by locus C9, suspected to be under directional selection, and by locus D1, apparently under balancing selection, the analysis was repeated with both loci removed from the dataset. The clustering pattern changed considerable when only six STRs were used. The most probable number of clusters changed to $K = 4$, where *L. l. iberica* and *L. l. nevadensis* remained as independent clusters, but *L. l. lepida* populations became subdivided into two clusters. These two clusters exhibited extensive admixture and grouped BEJ, SET and SPE in one cluster and CMA, TOL, AND and ALE in another, although most ALE individuals had nearly 50% chance of being assigned to each of the two clusters (Fig. 2). *L. l. lepida* clustering broadly separated the southernmost populations from the central and western ones, where ALE showed an intermediate position in this geographical assemblage. The number of clusters inferred from the 318 neutral AFLPs was $K = 2$. One of the clusters included only ALM individuals (*L. l. nevadensis*) while the remaining nine sampled populations clustered together in the second group (Fig. 2). As for the 23 outlier AFLPs potentially under selection, the most likely number of clusters was $K = 2$ according to the Evanno *et al.* (2005) method (Fig. S3, Supporting information), separating ALM individuals from all other populations as with the 318 neutral AFLP markers. Nevertheless, unlike the 318 neutral AFLPs, which failed to cluster *L. l. iberica* individuals (GAL and GER) in the same group at $K = 3$, the structure pattern obtained with the 23 outlier loci with K = 3 (Fig. 2) recovered a grouping pattern very similar to the one obtained with STRs, clustering populations according to their respective subspecies, although with evidence of considerable admixture in several individuals from *L. l. lepida* with *L. l. iberica*.

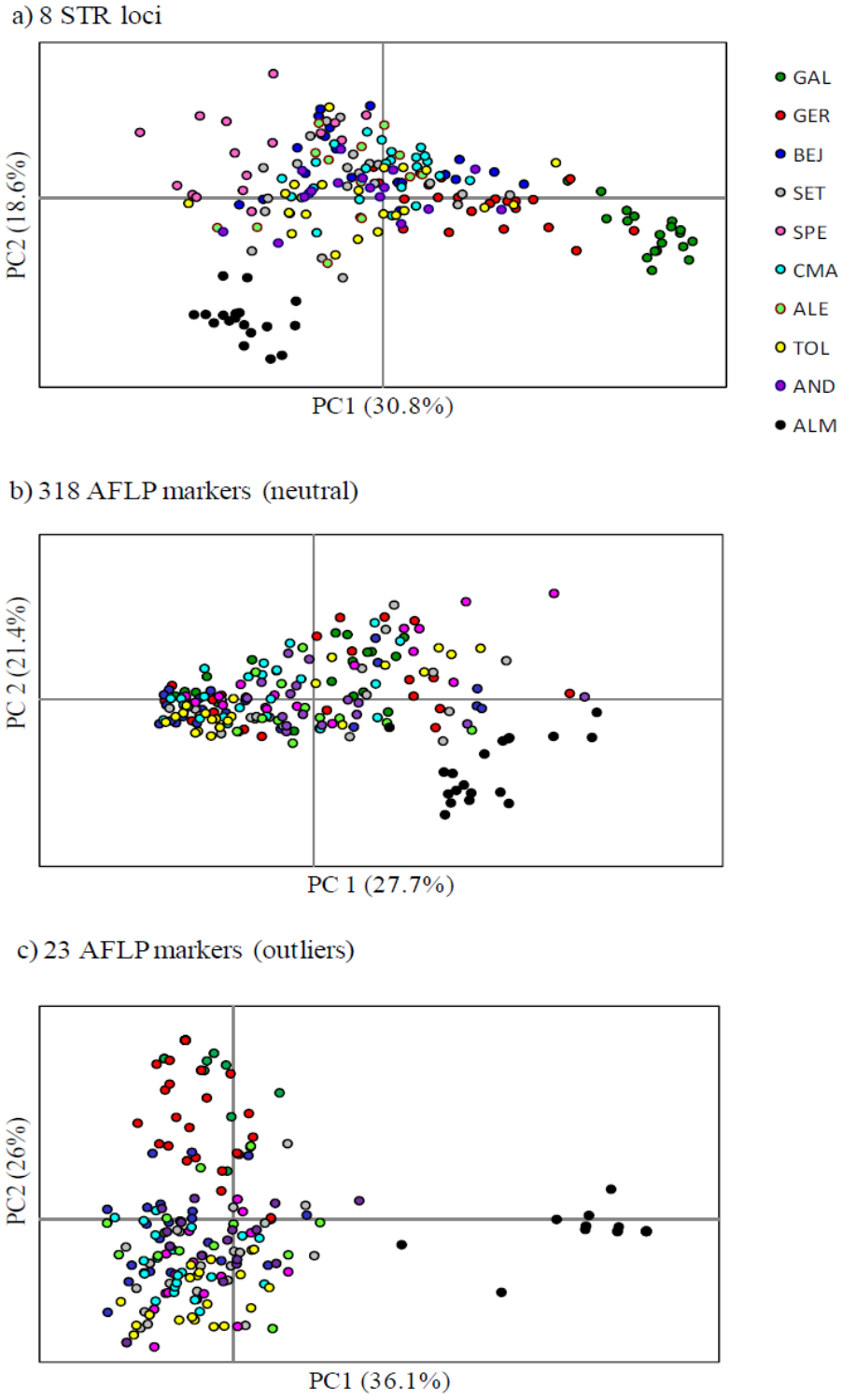


Fig. 1 – Principal coordinate analysis (PCoA) performed with 196 lizards collected from 10 populations of *Lacerta lepida*. Three types of nuclear markers were used: 8 STRs (a), 318 neutral AFLP markers (b) and 23 outlier AFLP markers (c). Populations GAL and GER belong to *L. l. iberica* and ALM is from *L. l. nevadensis*, whereas the remaining populations belong to the nominal subspecies, *L. l. lepida*.

Table 5 Analyses of molecular variance (AMOVAs) performed with STRs, neutral AFLPs and outlier AFLPs. Populations were grouped according to five scenarios of population structure: (i) a single group; (ii) two groups where *L. l. nevadensis* (ALM population) is in one group and the remaining populations are in the other; (iii) three groups according to the subspecies; (iv) six groups corresponding to each mitochondrial clade (N and L1 to L5), as defined by Miraldo *et al.* (2011); (v) four groups including only *L. l. lepida* (seven populations), divided according to their mitochondrial clade (L1, L2, L4 and L5; Miraldo *et al.* 2011). Asterisks denote significance at $P < 0.05$ (*) or $P < 0.001$ (**).

			Variance components (%)		
Dataset	# groups	Grouping structure	Within populations	Among populations within groups	Among groups
8 STRs	1	single group	89.25	10.75**	_
	2	[*L. l. nevadensis*], [*L. l. lepida* + *L. l. iberica*]	82.47**	8.19**	9.34
	3	[*L. l. nevadensis*], [*L. l. lepida*], [*L. l. iberica*]	85.54**	5.84**	8.62*
	6	*L. lepida* mtDNA clades (10 Populations)	88.85**	6.37**	4.78*
	4	*L. l. lepida* mtDNA clades (7 populations)	94.89**	4.38*	0.73
318 AFLPs (neutral loci)	1	single group	91.51	8.49**	_
	2	[*L. l. nevadensis*], [*L. l. lepida* + *L. l. iberica*]	79.82**	4.48**	15.69
	3	[*L. l. nevadensis*], [*L. l. lepida*], [*L. l. iberica*]	88.01**	4.05**	7.95*
	6	*L. lepida* mtDNA clades (10 Populations)	91.09**	4.04**	4.87*
	4	*L. l. lepida* mtDNA clades (7 populations)	95.67**	4.21**	0.12
23 AFLPs (outliers)	1	single group	59.41	40.59**	_
	2	[*L. l. nevadensis*], [*L. l. lepida* + *L. l. iberica*]	32.26**	11.59**	56.14
	3	[*L. l. nevadensis*], [*L. l. lepida*], [*L. l. iberica*]	45.99**	7.16**	46.85**
	6	*L. lepida* mtDNA clades (10 Populations)	57.34**	5.87**	36.8*
	4	*L. l. lepida* mtDNA clades (7 populations)	87.14**	7.13**	5.72*

Discussion

The eight STRs used in this study, with the exception of LIZ24, were developed for other Lacertidae species, but most of them showed high levels of polymorphism in *L. lepida* populations. Populations from the south of the Iberian Peninsula showed higher allelic richness, which is in accordance with Miraldo *et al.* (2011) and supports the importance

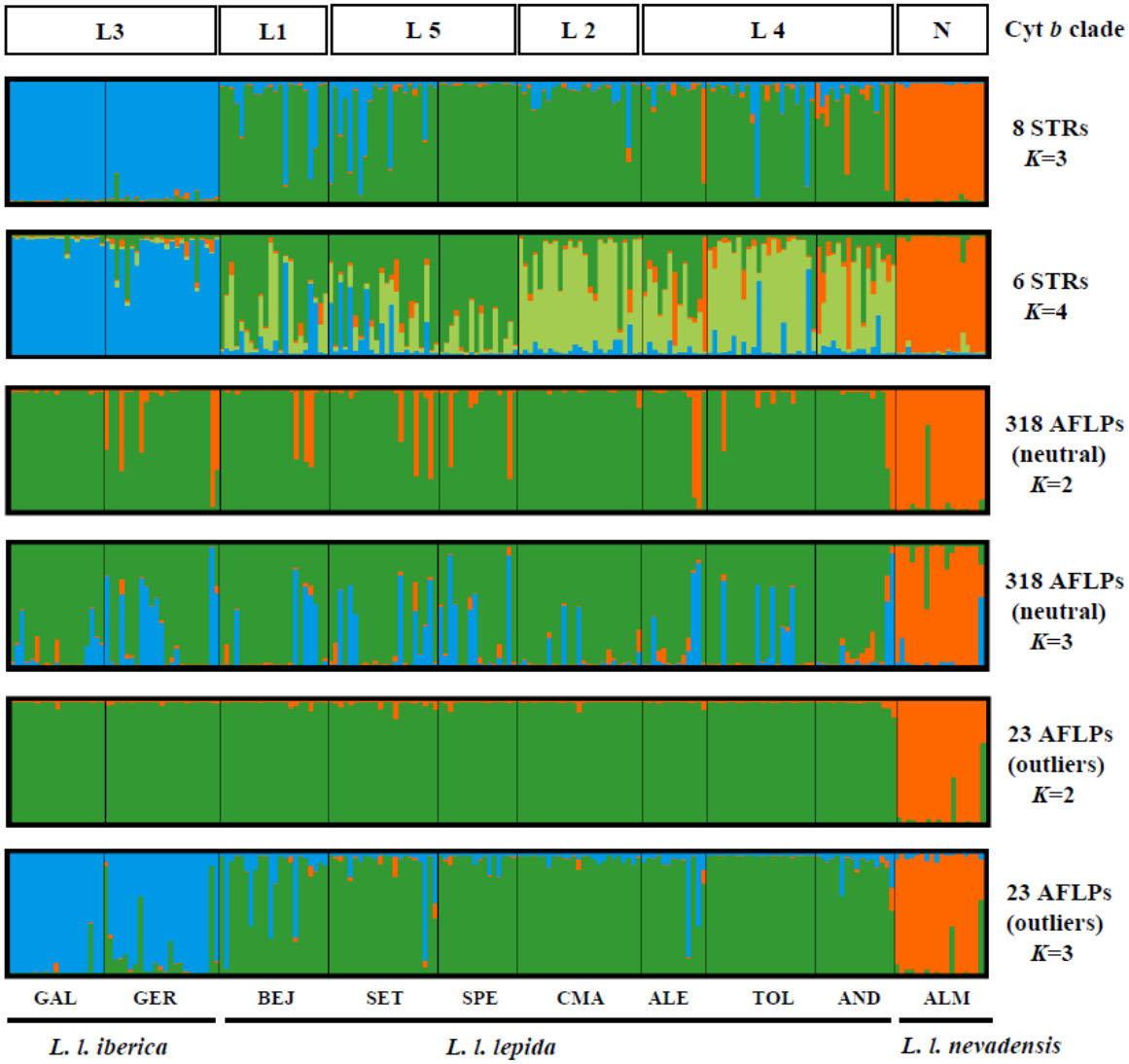


Fig. 2 Inference of *K* clusters by STRUCTURE. Each individual is corresponds to a vertical bar representing the probability of assignment to each cluster (indicated by different colours). Codes at the bottom correspond to population labels and the corresponding subspecies is also indicated. The mitochondrial clade corresponding to each population is indicated on the top (as inferred by Miraldo *et al.* (2011)). Top to bottom: eight STRs (including locus C9 and locus D1, detected under selection; $K = 3$); six neutral STRs ($K = 4$); 318 neutral AFLP markers ($K = 2$ and $K = 3$) and 23 outlier AFLP markers ($K = 2$ and $K = 3$).

of the southern region as a glacial refuge for *L. lepida* during the climatic oscillations of the Quaternary. The allelic richness and heterozygosity measures suffer a marked decrease for most STRs in the GAL population, from *L. l. iberica*, with a single allele reaching fixation at the two least polymorphic loci, C9 and LIZ24. Alleles found in the

GAL population were a subsample of the alleles detected in the nominal subspecies. These observations suggest that the consequences of genetic drift might have been more severe in the northwest, that was probably more exposed to the effects of the glacial periods due to its higher latitude. The GAL population is located close to the northern limits of the species' range and probably results from a recent range expansion of lizards that survived at lower latitudes (Miraldo *et al.* 2011), leading to a reduction in microsatellite diversity. As for *L. l. nevadensis*, the high proportion of private alleles observed in the ALM population is congruent with a long period of divergence with restricted gene flow. Despite the peripheral and restricted distribution range of the subspecies when compared to the wide distribution range of the nominal subspecies, *L. l. nevadensis* had high levels of genetic diversity, as reflected by the high allelic richness and heterozygosity values at most STRs.

Analyses of genetic structure, irrespective of the type of nuclear markers used, with or without the effect of selection, all distinguished *L. l. nevadensis* lizards as a well-defined group, while the distinction of *L. l. iberica* populations was supported by STR loci and by the 23 outlier AFLPs, but not by the 318 neutral AFLPs. Within the *L. l. lepida* subspecies, the subdivisions corresponding to the four mitochondrial clades previously identified by Paulo (2001), and later confirmed by Miraldo *et al.* (2011), were not fully supported by the present nuclear loci datasets. The lack of concordance between mitochondrial genes and nuclear markers within *L. l. lepida* populations is not surprising, since Miraldo *et al.* (2011) detected similar patterns of lack of structure with a nuclear gene, β-fibrinogen. Because the divergence between *L. l. lepida* mitochondrial clades was recent, incongruence with nuclear loci may be expected from the retention of ancestral polymorphism and an incomplete lineage sorting of STRs and AFLP genotypes, as well as admixture since secondary contact.

Tests for selection in STRs detected balancing selection in locus D1 and directional selection in locus C9. However, these results must be taken with caution, because the number of loci tested here is very small and the performance tests made by Foll & Gaggiotti (2008) for BayeScan were conducted with hundreds of STRs. The authors also noted that the inclusion of loci with different mutation rates in the analysis can affect the performance of BayeScan by increasing the false-positive rate and that separate analyses should be performed on di-, tri- and tetranucleotide STRs. The STRs used in this study comprise three trinucleotide loci (C9, D1 and LIZ24) and four dinucleotide loci. The split

of such a small number of STRs into two separate datasets for the BayeScan analysis would compromise even further the reliability of the results. In the case of locus C9, the signature of directional selection detected here probably arose from the nearly non-overlapping allele sizes and the remarkable differences in allele frequencies between subspecies (Fig. S1, Supporting information). However, the divergence accumulated between these subspecies under neutrality could explain the evolution of this pattern in locus C9 without invoking the action of directional selection. We know from mitochondrial DNA sequence data that *L. l. nevadensis* has a deep divergence from the remaining Iberian subspecies, over 9 million years ago (Ma), and that *L. l. iberica* also has considerable genetic differentiation from *L. l. lepida* (divergence time estimated in nearly 2 Ma) (Paulo *et al.* 2008; Miraldo *et al.* 2011). Nevertheless, inferences of genetic structure with STRUCTURE were repeated without STR loci suspected to be under selection. On one hand, we removed the masking effect over weak genetic structure that locus D1 could produce under balancing selection, by promoting population similarity. On the other hand, we removed the contribution of locus C9 to the hierarchical structure, by strengthening the level of divergence of *L. l. nevadensis* (due to the considerable number of unique alleles found in ALM population) and *L. l. iberica* (because populations of this subspecies are monomorphic at locus C9). Nevertheless, when C9 and D1, the remaining six STRs were insufficient to consider populations from each *L. l. lepida* mitochondrial clade as reliable and independent clusters. It was only possible, at best, to detect a weak genetic structure that distinguished southern populations (mitochondrial clades L1 and L5) from populations in the centre and west of the Iberian Peninsula (mitochondrial clades L2 and L4).

AFLPs are considered less informative than STRs due to its biallelic and dominant scoring. The use of a large number of AFLP markers could compensate for the lower resolution of individual loci. Mariette et al. (2002) suggested that four to ten times more dominant markers should be used to obtain the same efficiency as with codominant markers, such as STRs. The number of neutral AFLP markers used here greatly outnumbered the small set of STRs. Yet, the 318 neutral AFLP markers were insufficient to distinguish *L. l. iberica* samples from the nominal subspecies, whereas the small set of 23 AFLP outliers, considered as candidate loci under the effect of directional selection, was enough to recognise *L. l. iberica* individuals from *L. l. lepida*. Several examples in the literature show that results from STRs and AFLPs are frequently not in full

agreement, but while we could expect that resolution obtained from AFLPs would be normally lower than from STRs, some studies detected the opposite trend (Woodhead *et al.* 2005; Garoia *et al.* 2007; Kato *et al.* 2011). Some authors argue that when the heterogeneity within the genome is high but the sampling effort with nuclear markers is low, the correlation between results from different types of markers is expected to be low (Mariette *et al.* 2002). The use of a small set of highly polymorphic STRs might not be representative of the whole genome diversity (Väli *et al.* 2008). Therefore, the large number of AFLPs used in this study should be more representative of the genomic diversity than STRs in ocellated lizards, since AFLP markers are expected to be randomly scattered across the genome.

Assuming that the AFLPs investigated here are representative of European ocellated lizards' genome, they suggest that divergence of *L. l. iberica* from the nominal subspecies results mainly from a few adaptive loci. We should note that *L. l. iberica*'s divergence in neutral STRs is accompanied by an overall decrease in variability. This scenario is compatible with an incipient speciation process occurring in *L. l. iberica* populations, following the genic view of the speciation process proposed by Wu (2001). Some morphological and life history traits observed in *L. l. iberica* support the hypothesis of an adaptive response to the ecological conditions in the northwest, with abundant rainfall, lower temperatures and lower insolation. Such climatic conditions are less favourable for ectothermic animals such as ocellated lizards that depend on solar exposure to thermoregulate and gather energy for daily activities. The body size of *L. l. iberica* is on average smaller (Mateo & Castroviejo 1990) and its coloration is darker (Mateo & López-Jurado 1994; Nunes *et al.* 2011a), with possible benefits for thermoregulatory efficiency (Clusella-Trullas *et al.* 2007). *L. l. iberica* females start reproducing around one year earlier and at a smaller body size than in the other subspecies, producing smaller clutches (Mateo & Castanet 1994). Further investigations are required to test for the adaptive value and the genetic inheritance of such traits but the insights provided by outlier AFLP markers indicate that the quest is worthwhile.

All types of nuclear markers investigated, including the outlier AFLP markers, attest the clear divergence of *L. l. nevadensis* from other subspecies. The long divergence time (before the Quaternary glaciations) was enough to accumulate genetic differentiation not only in mitochondrial DNA but also in neutral nuclear loci (Paulo 2001; Paulo *et al.* 2008; Miraldo *et al.* 2011) and in adaptive loci. The subspecies shows morphological and

life history variation that is suspected to be an adaptive response to the arid environment faced in the southeast of the Iberian Peninsula (with high temperatures and the lowest and most irregular annual rainfall in Iberia). For instance, the dorsal pattern of *L. l. nevadensis* is brownish, with clear advantages for camouflage in an arid landscape (Mateo & López-Jurado 1994; Nunes *et al.* 2011a). The first evidence for the genetic basis of colour variation in ocellated lizards was given by sequence analysis of the *melanocortin-1 receptor* gene (involved in melanin synthesis), which revealed one derived amino acid replacement associated with *L. l. nevadensis* colour phenotype (Nunes *et al.* 2011a). What might have started as divergence in a few adaptive loci has long ago spread to adjacent regions of the genome. Miraldo (2009) found evidence for a narrow secondary contact zone between *L. l. nevadensis* and the nominal subspecies with restricted gene flow and suggested selection against hybrids. *L. l. nevadensis* is probably in the final stages of speciation, although with reproductive isolation still incomplete. Some authors have already suggested the upgrade of the subspecies taxonomic status to the species level, as *Lacerta nevadensis* (Paulo *et al.* 2008).

Acknowledgements

This research was supported by a doctoral scholarship from FCT (SFRH/BD/21306/2005) to Nunes VL and the FCT projects POCTI/BSE/47999/2002 – "Coalescent methods applied to populations' analyses with microsatellites" and POCI/BIA/59288/2004- "Process of speciation: accessing the genic view", both coordinated by Paulo OS.

References

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **263**, 1619-1626.

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, **14**, 2899-2914.

Böhme MU, Berendonk TU, Schlegel M (2005) Isolation of new microsatellite loci drom the Green Lizard (*Lacerta viridis viridis*). *Molecular Ecology Notes*, **5**, 45-47.

Bohonak AJ (2002) IBD (Isolation By Distance): a program for analyses of isolation by distance. *Journal of Heredity*, **93**, 153-154.

Boudjemadi K, Martin O, Simon J-C, Estoup A (1999) Development and cross-species comparison of microsatellite markers in two lizards species, *Lacerta vivipara* and *Podarcis muralis*. *Molecular Ecology*, **8**, 518-520.

Brookfield J FY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453-455.

Clusella-Trullas S, van Wyk JH, Spotila JR (2007) Thermal melanism in ectotherms. *Journal of Thermal Biology*, **32**, 235-245.

Colbeck GJ, Turgeon J, Sirois P, Dodson J (2011) Historical introgression and the role of selective vs. neutral processes in structuring nuclear genetic variation (AFLP) in a circumpolar marine fish, the capelin (*Mallotus villosus*). *Molecular Ecology*, **20**, 1976-1987.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical Applied Genetics*, **92**, 832-839.

Evanno G, Regnaut S, 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, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Herediy*, **103**, 285-298.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online,* **1,** 47-50.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574-578.

Feliner GN (2011) Southern European glacial refugia: A tale of tales. *Taxon*, **60**, 365-372.

Foll M, 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.

Gaggiotti OE, Lange O, Rassmann K, Gliddons C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, **8**, 1513-1520.

Garoia F, Guarniero I, Grifoni D, Marzola S, Tinti F (2007) Comparative analysis of AFLPs and SSRs efficiency in resolving population genetic structure of Mediterranean *Solea vulgaris*. *Molecular Ecology*, **16**, 1377-1387.

Godinho R, Crespo EG, Ferrand N (2008) The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Molecular Ecology*, **17**, 4670-4683.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 155-188. Springer, Netherlands.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). http://www.unil.ch/izea/softwares/fstat.htlm. Updated from Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87-112.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. P*hilosophical Transactions of the Royal Society of London, Series B:Biological Sciences*, **359**, 183-195.

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.

Kirk H, Freeland JR (2011) Applications and implications of neutral versus non-neutral markers in molecular ecology. *International Journal of Molecular Sciences*, **12**, 3966-3988.

Mariette S, Le Corre V, Austerlitz F, Kremer A (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Molecular Ecology*, **11**, 1145-1156.

Mateo JA (1988) *Estudio Sistemático y Zoogeográfico de los Lagartos Ocelados, Lacerta lepida (Daudin, 1802) y Lacerta pater (Lataste 1880) (Sauria, Lacertidae)*. University of Sevilla, Sevilla.

Mateo JA, Castanet J (1994) Reproductive strategies in 3 Spanish populations of the ocellated lizard, *Lacerta Lepida* (Sauria, Lacertidae). *Acta Oecologica*, **15**, 215-229.

Mateo JA, Castroviejo J (1990) Variation morphologique et révision taxonomique de l'espèce *Lacerta lepida* Daudin, 1802 (Sauria, Lacertidae). *Bulletin du Muséum National D'Histoire Naturelle, Paris*, **12** , 691-706.

Mateo JA, López-Jurado LF, Cano J, Guillaume CP (1999) Intraspecific variability in nucleolar organiser position in *Lacerta (Timon) lepida*. *Amphibia-Reptilia*, **20**, 73-77.

Miraldo A (2009). *Phylogeography and population dynamics of secondary contact zones of Lacerta lepida in the Iberian Peninsula*. University of East Anglia, Norwich.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Nembrini M, Oppliger A (2003) Characterization of microsatellite loci in the wall lizard *Podarcis muralis* (Sauria: Lacertidae). *Molecular Ecology Notes*, **3**, 123-124.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375-402.

Nunes VL, Beaumont MA, Butlin RK, Paulo OS (2011a) Association of *Mc1r* variants with ecologically relevant phenotypes in the European ocellated lizard, *Lacerta lepida*. *Journal of Evolutionary Biology*, **24**, 2289-2298.

Nunes VL, Beaumont MA, Butlin RK, Paulo OS (2011b) Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Molecular Ecology*, **20**, 193-205.

Paulo OS (2001). *The phylogeography of reptiles of the Iberian Peninsula*. University of London, London.

Paulo OS, Pinheiro J, Miraldo A *et al.* (2008) The role of vicariance vs. dispersal in shaping genetic patterns in ocellated lizard species in the western Mediterranean. *Molecular Ecology*, **17**, 1535-1551.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288-295.

Pinho C, Sequeira F, Godinho R, Harris DJ, Ferrand N (2004) Isolation and characterization of nine microsatellite loci in *Podarcis bocagei*. *Molecular Ecology Notes*, **4**, 286-288

Pritchard JK, Stephens M, Donelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Richter-Boix A, Quintela M, Segelbacher G, Laurila A (2011) Genetic analysis of differentiation among breeding ponds reveals a candidate gene for local adaptation in *Rana arvalis*. *Molecular Ecology*, **20**, 1582-1600.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137-138.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219-1228.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.

Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4**, 11.

Sequeira F, Alexandrino J, Weiss S, Ferrand N (2008) Documenting the advantages and limitations of different classes of molecular markers in a well-established phylogeographic contex: lessons from the Iberian endemic Golden-striped salamander, *Chioglossa lusitanica* (Caudata: Salamandridae). *Biological Journal of the Linnean Society*, **95**, 371-387.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457-462.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.

Väli U, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, **17**, 3808-3817.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

Vekemans X (2002) AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génetique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9939-9946.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP - A new technique for DNA-fingerprinting. *Nucleic Acids Research*, **23**, 4407-4414.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Woodhead M, Russel J, Squirrell J et al. (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. *Molecular Ecology*, **14**, 1681-1695.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851-865.

CHAPTER 5

Supporting information

Table S1 Microsatellite loci genotyped for *Lacerta lepida* populations. Primer sequences (with indication of fluorescence dye used: 6-FAM, HEX or NED), PCR annealing temperature (T_a), size range and number of alleles (N_A) detected in *L. lepida* are indicated for each locus.

Locus	GenBank ac. n.	Primer sequence (5' --> 3')	T_a (ºC)	Size (bp)	N_A
B4	AY147826	F: HEX-AATCTGCAATTCTGGGATGC R: AGAAGCAGGGGATGCTACAG	61	124-168	22
C9	AY147822	F: 6-FAM-CATTGCTGGTTCTGGAGAAAG R: CCTGATGAAGGGAAGTGGTG	58	133-169	13
D1	AY147830	F: NED-GAGTGCCCAAGACAGTTGTAT R: GAGGTCTTGAATCTCCAGGTG	58	134-209	23
LIZ24	n.a.	F: 6-FAM-TCAGTCCAAATATCTCTACAGG R: AGATGAGCAGCATATAGTGATG	50	115-139	6
Lvir17	AJ783631	F: 6-FAM-AGCTCTGGATCGAGACAACCTGG R: TCTCTGAAGGAGACCGGCTCC	61	213-269	22
Lv-4-72	AF100290	F: HEX-CCCTACTTGAGTTGCCGTC R: CTTTGCAGGTAACAGAGTAG	63	106-140	17
Pb66	AY545227	F: NED- GGACAGCTAGTCCCATGGCTTAC R: GGATTGCTGTCACCAGTCTCCCC	58	150-190	19
Pb73	AY545228	F: 6-FAM-GCCCATGTCACTTCAGGTAGAAGC R: GAAAACTAGGAGTTAGGGAGAAGG	58	118-152	17

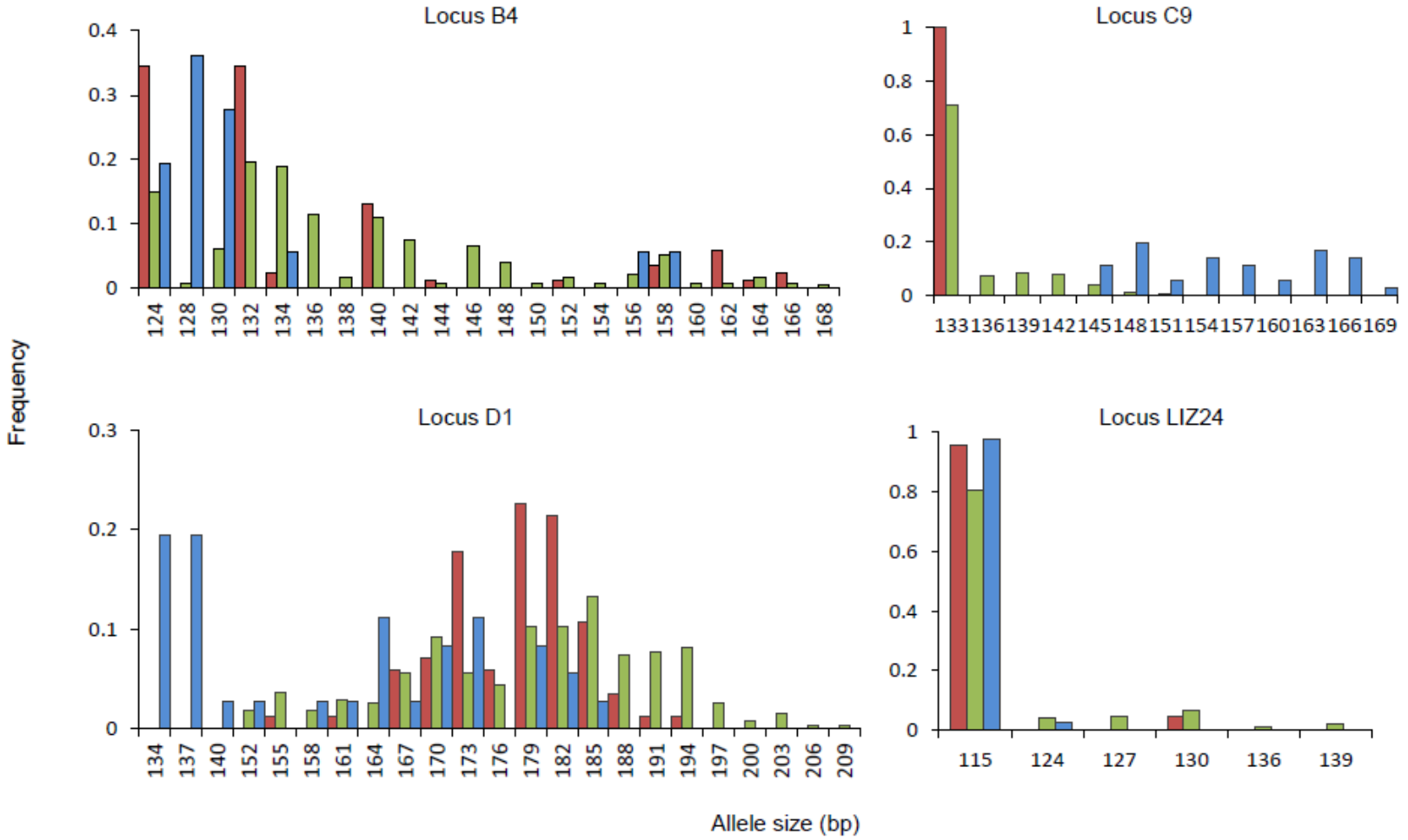


Fig. S1 Allele frequencies from each microsatellite locus for each *Lacerta lepida* subspecies: *L. l. iberica* (red), *L. l. lepida* (green) and *L. l. nevadensis* (blue). Each bar corresponds to a single allele.

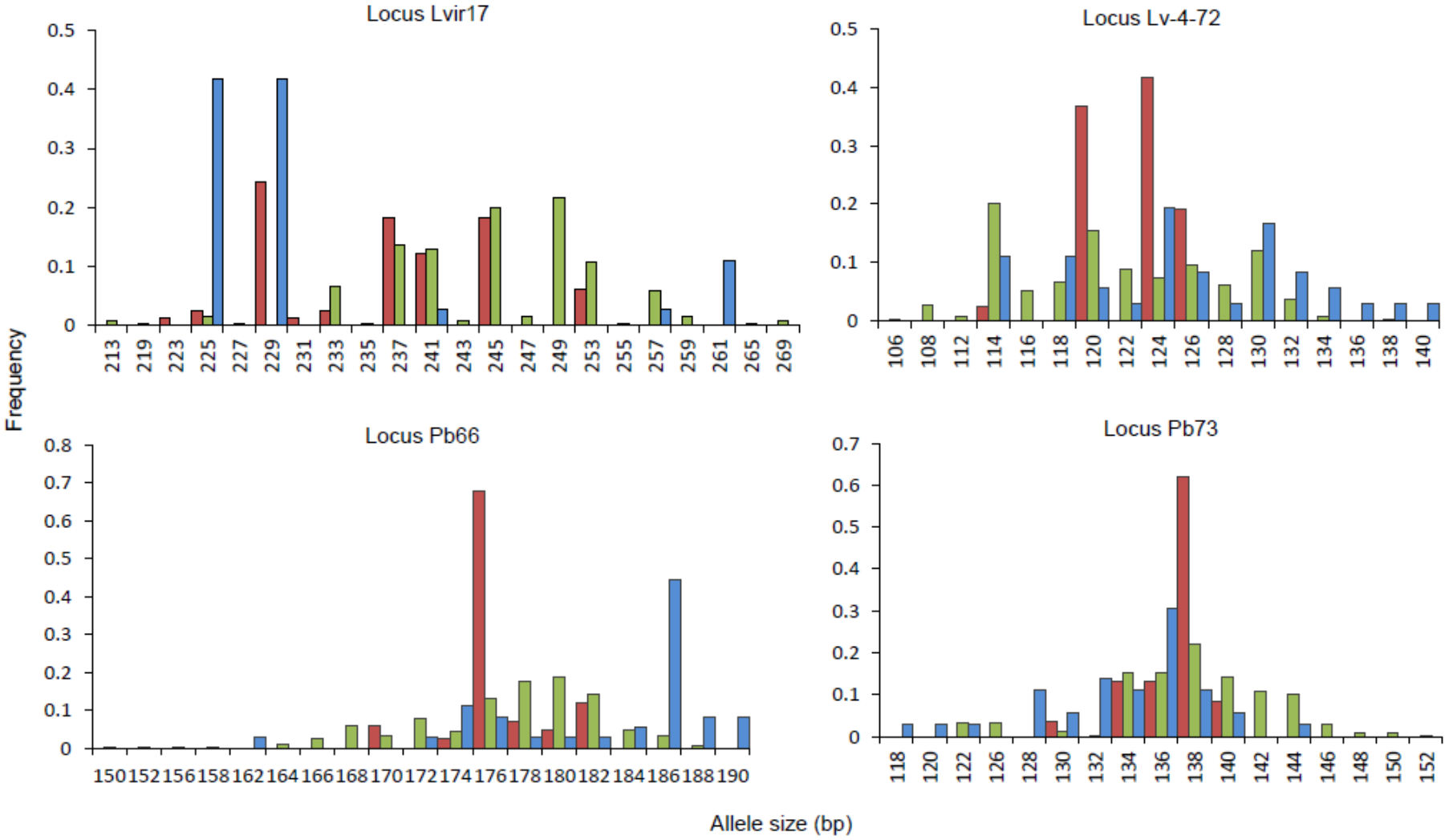


Fig. S1 (continued).

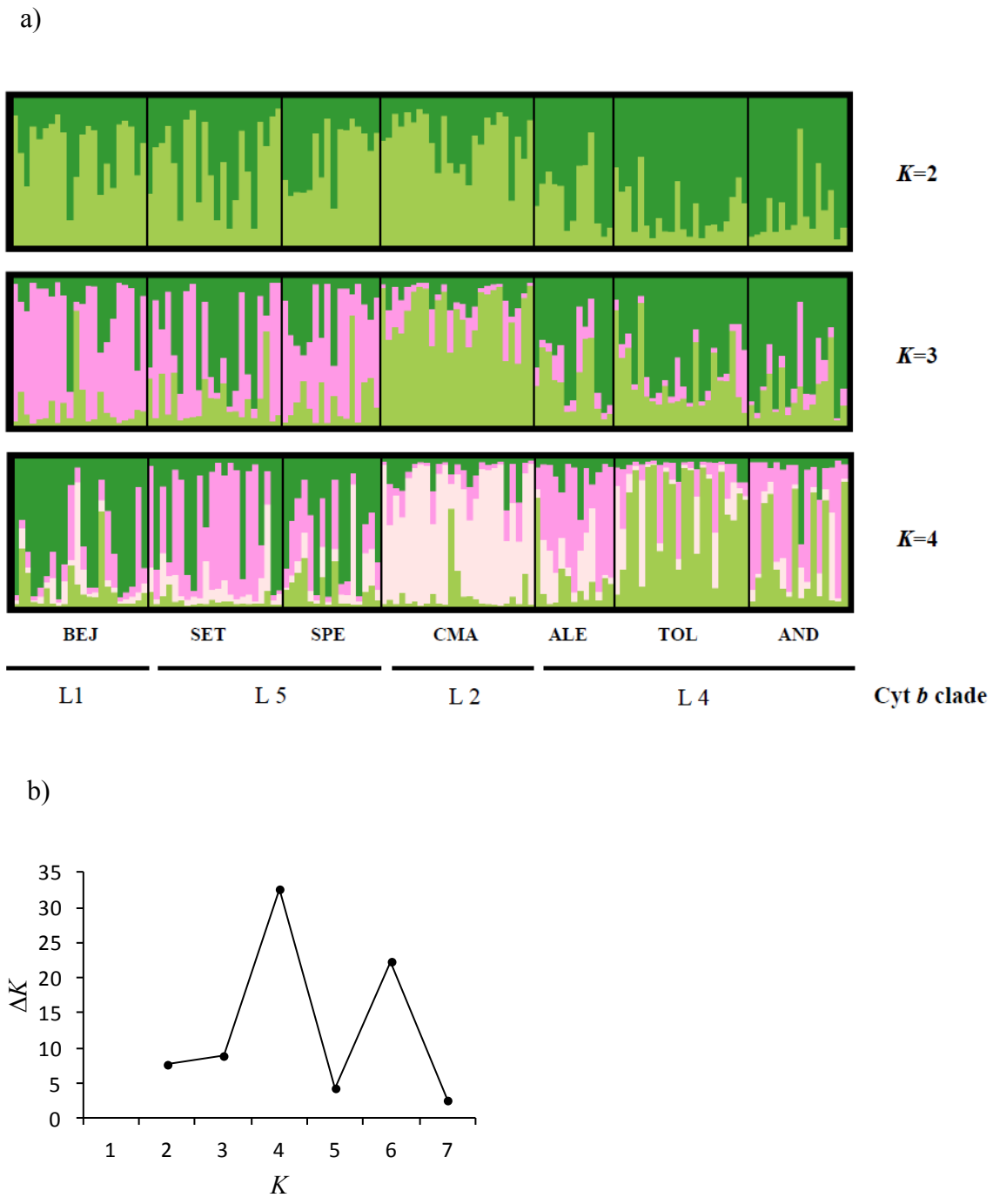


Fig. S2 Population structure (a) inferred in STRUCTURE for *L. l. lepida* populations with eight STRs. Each individual corresponds to a vertical bar representing the probability of assignment to each cluster (represented with different colours). Populations' labels are below, as well as the indication of the cytochrome *b* (cyt *b*) clade that corresponds to each population, according to Miraldo *et al.* (2011). The most probable number of clusters (b) according to Evanno *et al.* (2005) calculations is $K = 4$.

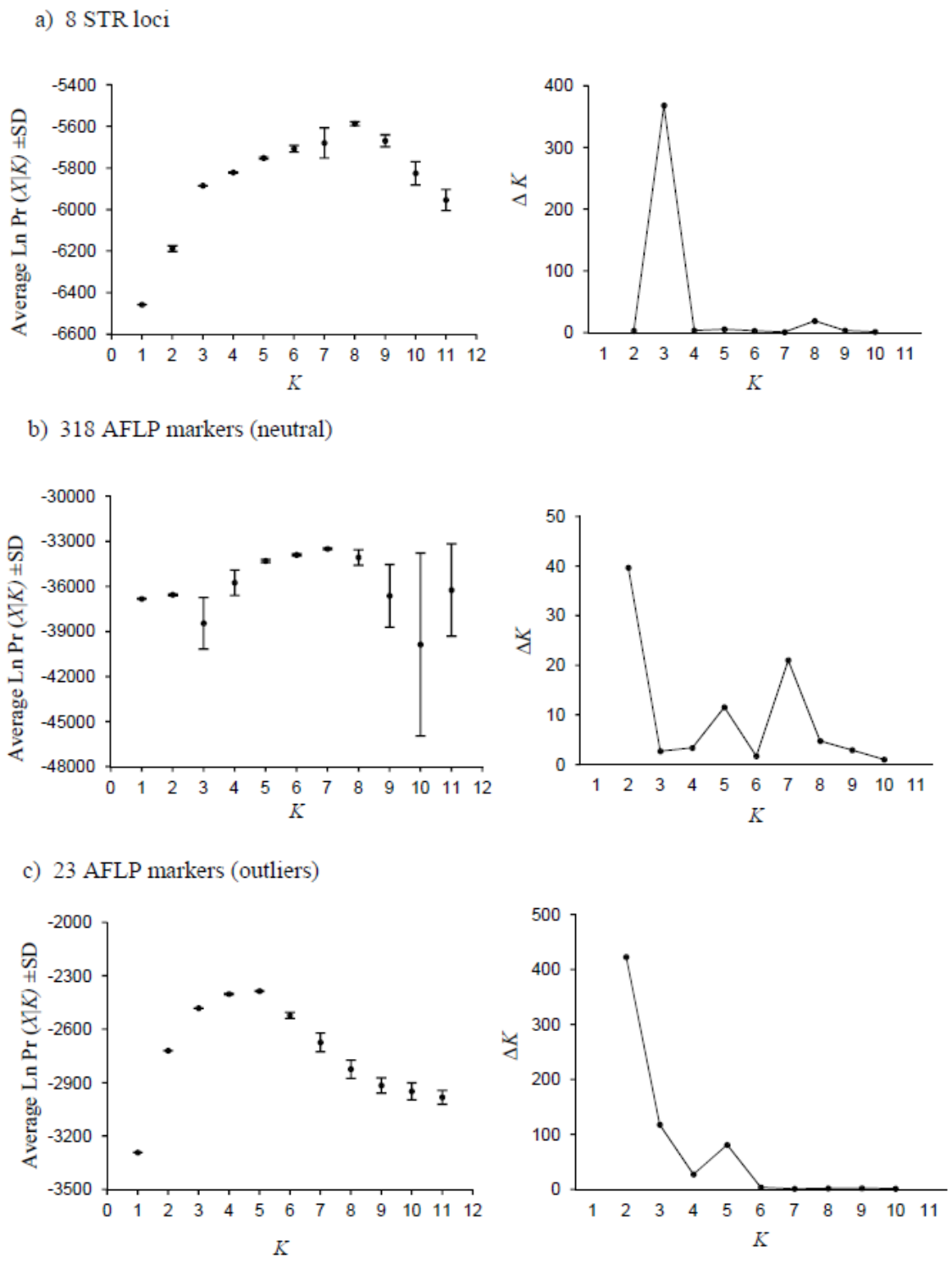


Fig. S3 Plots for average Ln Pr (*X*|*K*) over 10 independent runs in STRUCTURE for *K* = 1 to *K* = 11 and for Δ*K* calculated according to Evanno *et al.* (2005). The results are presented for eight STR loci (a), 318 neutral AFLP markers (b) and 23 outlier AFLP markers (c).

CHAPTER 6

General Discussion

6.1 – General discussion

The present study intended to gain some insights on the speciation process in a spatially structured species along an environmental gradient by detecting and investigating genomic regions responsible for local adaptation. This chapter presents an integrated discussion of the results presented in chapters 2 to 5.

Two approaches were chosen to detect genes under selection in ocellated lizards: a genome scan with AFLP markers and the analysis of a candidate gene for variation in coloration, the *Mc1r*. For species with limited genetic knowledge, a genome scan with AFLPs provides a good starting point in the detection of candidate genes under selection (Bench & Akesson 2005). However, the portion of the genome that can be scanned is limited by the number of AFLP selective primer combinations used (which in turn depend on the time and financial resources available) and by the total number of polymorphic loci considered. In European ocellated lizards, the AFLP genome scan was conducted with 392 polymorphic markers, which were produced by eight selective primer combinations (chapter 2). A total of 196 lizards were genotyped, sampled from 10 populations in the Iberian Peninsula, along a northwest-southeast and North-South transects, covering the environmental heterogeneity of the species range, the distribution of the three parapatric subspecies (*L. l. iberica*, *L. l. lepida* and *L. l. nevadensis*) and all but one mitochondrial clades identified by Paulo (2001). The detection of AFLP markers with outlier behavior (with exceptional higher or lower frequency as compared to neutral expectations) was performed with two of the most commonly used methods: a frequentist method, DFDIST (Beaumont & Nichols 1996) and a Bayesian method, BayeScan (Foll & Gaggiotti 2008). When the detection of AFLP outliers is intended for follow-up studies to investigate the underlying genes, it is important to minimize the false-discovery rate (Bonin *et al.* 2007; Caballero *et al.* 2008; Pérez-Figueroa *et al.* 2010), and the list of outliers should not change dramatically with the statistical method chosen for the analysis. DFDIST simulates the theoretical null distribution of the genetic differentiation between populations and compares it with empirical data to detect outliers with extremely high (directional selection) or low F_{ST} values (balancing selection). Loci with a critical frequency for the most common allele equal to or above 0.98 were discarded from the DFDIST analysis. BayeScan directly estimates the probability that

each locus is subject to selection with a Bayesian method and takes all loci into account for the analysis. BayeScan seems to be robust when dealing with complex demographic scenarios for neutral genetic differentiation (Foll & Gaggiotti 2008; Pérez-Figueroa *et al.* 2010). However, both DFDIST and BayeScan have a tendency to detect false positives when allele frequencies are correlated among populations, due to shared recent ancestry or to the effect of isolation by distance (Robertson 1975; Excoffier *et al.* 2009). Some precautions to minimize the false-positive detection rate proposed in the literature (Caballero *et al*. 2008; Pérez-Figueroa *et al.* 2010) were applied in ocellated lizards' AFLP analyses. Restrictive significance levels were implemented, assuming a false discovery rate of 5% in DFDIST and retaining only outliers detected by BayeScan with a posterior probability above 0.99.

The analyses for outlier detection were performed for all populations simultaneously and not in multiple pairwise comparisons, as preferred by several authors (e.g. Wilding *et al.* 2001; Jump *et al*. 2006; Egan *et al.* 2008; Nosil *et al.* 2008). Pairwise comparisons are less susceptible to problems caused by population structure and can strengthen evidence for candidate loci if they are detected in multiple independent comparisons across the environmental transition. In the ocellated lizards dataset, because only two populations of *L. l. iberica* and one from *L. l. nevadensis* (located at opposite extremes of the environmental gradient) were sampled, only five independent pairwise comparisons could be made. However, results from such comparisons were very different between DFDIST and BayeScan or as compared to global analyses. Moreover, a good true-positive rate is obtained with BayeScan for AFLP markers when at least six populations are compared simultaneously (Foll & Gaggiotti 2008) Therefore, only global analyses from DFDIST and BayeScan were considered and compared.

The inclusion of the only sampled population from *L. l. nevadensis* in the global analyses raises the most serious concerns about bias in the detection of outliers caused by genetic structure, because *L. l. nevadensis* has accumulated the highest levels of neutral divergence in the Iberian Peninsula (Paulo 2001; Paulo *et al.* 2008; Miraldo *et al.* 2011). The exclusion of the *L. l. nevadensis* population from the global analyses did not have a major effect on the results from BayeScan, but the results from DFDIST were more severely affected. However, the inclusion of *L. l. nevadensis* in outlier detection analyses is important, since it shows important morphological differences and is located at the

southeast extreme of the climatic gradient, thus providing an opportunity to detect loci that might have been affected by natural selection.

The proportion of AFLP outliers detected by DFDIST (4.1%) in the global analysis of 10 populations was similar to the one detected by BayeScan (3.1%), but the correspondence between loci detected by both methods was limited, with only 5 loci (21.7% from all outliers) detected by both DFDIST and BayeScan. Since the main objective of ocellated lizards' genome scan was to target candidate loci for further investigation, all 23 outliers detected were considered for follow-up studies, thus avoiding the risk of discarding true-positives. Outliers detected by both methods were all associated with directional selection.

Associations between AFLP markers' frequency and environmental variables along the Iberian Peninsula were tested by logistic regression, as implemented in the spatial analysis method (SAM) (Joost *et al.* 2008), to infer possible selective pressures acting at a local scale in ocellated lizards' populations. SAM is an individual-centred method, making no presumption as to the structure of populations to which sampled individuals belong (Joost *et al.* 2008). Since morphological variation in ocellated lizards seems to be associated with major bioclimatic regions of the Iberian Peninsula, we hypothesize that climatic variables could play an important role as selective pressures leading to the increase in frequency of adaptive loci as a response to local climatic conditions. Thus, a total of 54 environmental variables were tested with SAM for associations with AFLPs frequency in ocellated lizards: annual and monthly precipitation, annual and monthly temperature (maximum, mean and minimum values); annual insolation and annual relative humidity. SAM provides a practical tool to test many environmental variables simultaneously, but results depend on the variables tested, i.e., association with other untested environmental variables cannot be ruled out. Eleven of the 23 outliers detected by DFDIST or BayeScan showed strong associations with some environmental variables, mostly with maximum temperature (particularly of summer months), insolation and precipitation. The strongest association was detected for outlier 245 with maximum temperature in June. Since ocellated lizards are diurnal and ectothermic, variation in temperature and insolation play an important role in their seasonal and daily activities, especially from April to October, when *L. lepida*'s activity is higher (Busack & Visnaw 1989). Thus the asymmetries in climatic variables such as temperature, insolation and precipitation are expect to act as selective pressures in ocellated lizards, although the

association of outlier loci with these variables detected by SAM does not necessarily imply a causal relationship.

The statistical detection of AFLP outliers under selection should not be the end point of a genome scan for selection. Given the vulnerability of methods for outlier detection to type 1 error (false-positives), which can be minimized but hardly eliminated, AFLP outliers should be treated as candidate loci potentially influenced by selection and we should seek ways to identify and characterize these anonymous markers and to validate their selection signature (Butlin 2010). In order to bring outlier loci detected in ocellated lizards genome scan out from anonymity, outlier AFLP fragments previously scored through capillary electrophoresis (CE) were isolated from agarose gels, cloned and sequenced (chapter 3). The CE is commonly used to separate AFLP fragments by size and generate AFLP profiles, but because fragments migrate through a capillary instead of a regular gel matrix, outlier fragments need to be re-run in agarose or polyacrylamid gels to be excised and isolated from the other fragments. Because the relative size of fragments separated through electrophoresis may vary according to the matrix used, excised fragments from gel matrix should be re-amplified and separated by CE (as in chapter 3) to make sure that the isolated fragment is the expected outlier marker and not another close migrating fragment with a similar size.

The isolation process was challenging, due to the difficulty of isolating a specific fragment among dozens of other fragments with a similar size. Only seven out of twelve outliers were successfully isolated and sequenced. Larger AFLP fragments (>150 bp) were easier to isolate because the density of fragments in AFLP profiles is reduced for larger fragments' size. The co-migration of non-homologous fragments (size homoplasy) is more frequent at smaller band's size (Vekemans *et al.* 2002; Caballero *et al.* 2008). Moreover, the validation of a small fragment (95 bp) was compromised by the absence of suitable regions to design a pair of primers for the amplification of the outlier from undigested genomic DNA.

The outliers that were successfully sequenced revealed no homology with any known gene and seem to be non-coding regions, suggesting that they might be in linkage with the actual target of selection or that they belong to a regulatory region. Although only a few follow-up studies from AFLP genome scans were reported to date (Minder & Widmer 2008; Wood *et al.* 2008), they found the same trend, detecting mostly non-coding outlier fragments, rich in repetitive or transposable elements. These findings agree

with theoretical expectations that most polymorphic AFLP markers fall in non-coding regions, due to their smaller mutational constrains as compared to coding regions, thus making non-coding regions prone to contain restriction sites and generate AFLP markers variable in length (Stinchcombe & Hoekstra 2008; Butlin 2010).

Length polymorphism in AFLP markers can result from several types of mutations. When single mutations are present in AFLP primers' binding sites, either in the *Eco*RI or *Mse*I restriction sites or in the selective bases, only the dominant allele can be amplified, thus leaving recessive alleles out from AFLP profiles. If mutations are within the fragment, corresponding to insertions or deletions of several bases or to a variable number of repetitive elements, both dominant and recessive alleles will be amplified but since they differ in size, they are scored as different markers in the AFLP profile (Bensch & Akesson 2005). To investigate the sources of length polymorphism in ocellated lizards' AFLP outliers, an internal primer pair was designed for each sequenced outlier and used in combination with selective AFLP primers to amplify outliers' recessive alleles from digested DNA of homozygous recessive samples (i.e. samples where the outlier was scored as absent). The length polymorphism was mostly caused by internal indels or repetitive elements. Recessive alleles from outliers *mk75* and *mk390* were not fully amplified, suggesting additional sources of polymorphism at *Eco*RI or *Mse*I restriction sites, or at the neighbouring selective bases.

The detection of repetitive elements in sequenced outliers as sources of length polymorphism (*mk209* and *mk245*) justifies the concern reported in previous studies about the effect of homoplasy and non-independence of AFLP markers in the detection of outliers (Bonin *et al*. 2007; Caballero *et al*. 2008). The presence of a trinucleotide microsatellite in outlier *mk245* provides an example of a single AFLP locus that might generate several alleles with different band size within the same population, thus inflating homoplasy if they match the size of other loci, or leading to non-independent markers, if they are erroneously scored as different loci.

Amplification of both dominant and recessive alleles with internal primers from undigested DNA was only possible for *mk75*, *mk209* and *mk245*. For the other four sequenced outliers, the internal primers were not effective to amplify the fragments or the source of polymorphism was in or before the primer binding sites. Several samples from each *L. lepida* subspecies (*L. l. nevadensis*, *L. l. lepida* and *L. l. iberica*) previously

genotyped with AFLP markers were sequenced for outliers *mk75*, *mk209* and *mk245*. No discordances between band score genotypes and sequences obtained for *mk75* or *mk245* were detected, i.e. sequences from samples with the outlier band scored as absent were carrying two recessive alleles as expected, while samples where the band was scored as present had either two dominant alleles or one dominant allele combined with a recessive allele. For *mk209*, three of the sequenced samples where the outlier band was scored as present had no copies of the dominant allele, thus implying that the migration of homoplasic fragments lead to an erroneous scoring of the outlier marker in a few samples from *L. l. lepida*. Therefore, the band corresponding to *mk209* was actually exclusive of *L. l. nevadensis,* although it would not affect the outlier behaviour of *mk209* (it remains as a strong outlier in DFDIST analysis; data not shown).

Sequences from *mk75*, an outlier that was most strongly associated with precipitation, showed that the dominant allele is conserved, with a 9 bp deletion, whereas eight recessive haplotypes were detected in European ocellated lizards. The frequency of the dominant allele haplotype is higher in *L. l. iberica* while it is absent in *L. l. nevadensis*. Outlier *mk209*, also associated with precipitation, showed two dominant allele haplotypes, both with an insertion of four bases (TGGA), and seven recessive haplotypes. *L. l. nevadensis*' sequences from *mk209* were composed only by dominant haplotypes whereas only recessive haplotypes were detected in sequences from *L. l. lepida* or *L. l. iberica*. Sequences from *mk245*, which was strongly associated with maximum temperatures, revealed only one dominant allele haplotype, with a microsatellite composed by six GTT repeats, that was absent in *L. l. iberica* and *L. l. nevadensis*. A total of eight recessive haplotypes were detected, with a variable number of GTT repeats (from three to five repeats). Both *L. l. iberica* and *L. l. nevadensis,* located at opposite ends of the climatic gradient, present only haplotypes with three GTT repeats. If locus *mk245* is linked with genes that respond to higher temperatures, as suggested by SAM analysis, it is interesting that the same microsatellite allele is fixed at populations living at the most contrasted climatic conditions.

Outliers *mk75*, *mk209* and *mk245* were successfully amplified in closely-related species with the same internal primers developed for European ocellated lizards: *Lacerta tangitana*, *L. pater*, *L. schreiberi, L. agilis* and *Iberolacerta monticola*. Although repetitive elements may vary in length in these species, their flanking regions remain quite conserved across species. The data obtained so far from ocellated lizards' genome

scan follow-up confirms that outliers present a level of sequence divergence among populations that justifies their outlier behaviour, but it neither confirms nor denies that they are affected by selection. Further investigation on these loci is needed, requiring the development of additional genomic resources for this species or their close relatives, in order to understand what genomic regions surround or segregate in linkage with these outlier loci.

The benefits of using a candidate gene approach in the investigation of the genetic basis of adaptive traits were well illustrated with the analysis of the *melanocortin-1 receptor* (*Mc1r*) gene in ocellated lizards (chapter 4). The analysis of the dorsal colour in ocellated lizards demonstrated that each parapatric subspecies has clear differences in melanin-based colours (black/brown). Dorsal scales were counted in one cm^2 of the mid-dorsal region from each lizard and classified as black, brown or green/yellow. *L. l. nevadensis* presented the most conspicuous differences in colour, with the lowest proportion of black scales, which were replaced by brown scales, and the frequent exhibition of a faded dorsal pattern that results from a reduction in green scales over the body. *L. l. iberica* was in the opposite extreme of the colour variation cline in European ocellated lizards, with the highest proportion of black scales observed.

The melanin synthesis pathway is well conserved in vertebrates and genes affecting this pathway are well-characterized in vertebrate model species (Hoekstra 2006), thus providing a useful list of candidate genes for the investigation of colour polymorphism in lizards. Most pigmentation genes are composed by several exons separated from each other by introns that can expand for several kilobases, raising technical challenges to access the complete coding sequence in non-model species. *Anolis carolinensis* is the only complete lizard genome available to date and could be used to trace conserved regions in pigmentation genes to design primers for their amplification in *L. lepida*, but *A. carolinensis* is not closely-related to ocellated lizards. Using RNA as the starting material for gene isolation in a non-model species such as *L. lepida* is probably the most effective approach to access the full gene's coding sequence, since it avoids the variable introns. Raia *et al.* (2010) implemented this approach to accesss the full coding sequence of *Mc1r* in the Italian wall lizard, *Podarcis sicula*. Yet, the use of RNA to isolate

candidate genes is technically challenging due to the sensitivity and invasiveness of RNA isolation techniques in animals.

Mc1r was the most promising candidate gene for melanin-based colour variation in ocellated lizards, because *Mc1r* has been associated with colour variation in many species, is functionally conserved in vertebrates and is composed by a single and relatively small exon (Mundy 2005; Gompel & Prud'homme 2009). To isolate the *Mc1r* from genomic DNA in ocellated lizards for the first time, several primers were tested and the best results were achieved with a primer pair designed by Rosenblum *et al.* (2004), which amplified a central portion of the gene in European and African ocellated lizards, and also in other closely-related lizards (*L. schreiberi*, *I. monticola* and *L. agilis*). The isolation of the 5' and 3' ends of *Mc1r* in ocellated lizards was attempted with a genome walking protocol described by Reddy *et al.* (2008), but the few fragments that could be sequenced did not belong to *Mc1r* (results not shown).

The distinction of *L. l. iberica* from *L. l. lepida* based on *Mc1r* haplotypes was not possible, but they clearly distinguished both subspecies from *L. l. nevadensis* trough five diagnostic mutations that segregate in linkage (two of them were nonsynonymous). *L. l. nevadensis* presented a fixed, derived and nonconservative amino acid change, corresponding to the replacement of a threonine for an isoleucine residue in position 162 (T162I). This mutation was perfectly associated with the phenotype of sampled *L. l. nevadensis* lizards, with a prevalence of brown scales, thus suggesting a putative partial loss of function in *Mc1r*. The extension of the portion of *Mc1r* analysed in ocellated lizards to the full coding sequence and functional assays on the alternative alleles are needed to confirm the consequences of mutation T162I for the melanin synthesis in *L. l. nevadensis*. However, investigations on *Mc1r* in the little striped whiptail lizard, *Aspisdoscelis inornata*, detected the same amino acid change at the same protein domain (T170I) in association with a blanched phenotype (Rosenblum *et al.* 2004). The partial loss of function caused by mutation T170I was confirmed with in vitro functional assays (Rosenblum *et al.* 2010), suggesting that mutation T162I might have a similar effect in *L. l. nevadensis* phenotype, in which case it will represent another convergence example in *Mc1r* evolution (Manceau *et al.* 2010).

The second amino acid change detected in *Mc1r* in ocellated lizards corresponds to the replacement of a serine for a cysteine residue at position 172 (S172C). The serine

residue is shared between *L. l. nevadensis* and nearly all lizard species investigated to date for *Mc1r*, whereas the cysteine residue is present in all *L. l. iberica* and *L. l. lepida* samples, associated with black scales (putative gain of function). Because mutation S172C is a conservative amino acid change, it is less likely that it affects the melanin synthesis than mutation T162I, but mutation S172C is located in an functionally relevant domain of the protein (Garcia-Borron *et al.* 2005), in a remarkably conserved position in reptiles.

The lack of association of *Mc1r* haplotypes with *L. l. iberica* colour phenotype suggests that regulatory mutations affecting *Mc1r* expression or missed amino acid replacements in unsequenced extremities from the gene might explain the higher melanization in *L. l. iberica*. Alternatively, *L. l. iberica* colour phenotype can be affected by other pigmentation genes, such as *Agouti* or *Tyrp1*, which are implicated in light and dark coat colour in deer mice (Kingsley *et al*. 2009) and Soay sheep (Gratten *et al.* 2007), respectively.

Selection signatures in *Mc1r* from ocellated lizards were tested with Tajima's D (Tajima 1989) and McDonald-Kreitman test (McDonald & Kreitman 1991) but no evidence for positive selection was detected. Nevertheless, several examples show that a single amino acid substitution can affect the phenotype (e.g. Hoekstra *et al.* 2006; Römpler *et al.* 2006; Rosenblum *et al.* 2010), and most tests lack the power to detect selection with such a small number of amino acid changes (Hughes 2007), even when they might have a large effect on fitness.

A more detailed analysis of the frequency and distribution of *Mc1r* haplotypes was conducted in eight locations, along a transect perpendicular to the putative secondary contact zone between *L. l. nevadensis* and *L. l. lepida*. These locations were previously analysed by Miraldo (2009) for mitochondrial DNA, showing that locations from each subspecies were fixed for haplotypes from their respective mitochondrial lineage. The pattern obtained from *Mc1r* haplotypes is similar with only two exceptions: two heterozygous individuals, one from each of the closest populations to the contact zone, which were carrying a haplotype from *L. l. nevadensis* and another from *L. l. lepida*. These two *Mc1r* hybrids were sequenced for outliers *mk209* and *mk245* (data not shown), but for these loci they carry alleles associated with the subspecies from which they inherited their mitochondrial lineage. Thus, *Mc1r* hybrids are unlikely to represent F1

hybrids, and might result instead from backcross of F1 hybrids with one of the parent subspecies.

Results from *Mc1r* analysis in ocellated lizards are an important contribution for the investigation of the genetic basis of colour variation, but they represent only a first step, useful to delineate the direction of future research. We must assay the functional consequences of *Mc1r* amino acid replacements, test for differences in *Mc1r* expression, and look at the coding sequence and expression levels of other important coloration genes, before we get a clear picture of the genetics underlying ocellated lizards colour polymorphism.

Patterns of genetic structure in ocellated lizards were previously inferred from mitochondrial genes and a few nuclear genes (Paulo 2001; Paulo *et al*. 2008; Miraldo *et al*. 2011). However, information provided by mitochondrial genes often does not agree with patterns from nuclear data, because mitochondrial DNA is maternally inherited whereas nuclear genes are biparentaly inherited, thus requiring much more time to achieve a complete lineage sorting. The use of a few nuclear genes is probably not representative of the whole genome diversity and a multilocus sampling strategy should be preferred. Recent studies reflect a general awareness that phylogeographic studies based only on neutral markers might provide biased and incomplete information, which should be complemented with non-neutral markers to evaluate the adaptive genetic variation (Colbeck *et al.* 2011; Kirk & Freeland 2011; Richter-Boix *et al.* 2011). With the application of a genome scan for selection in ocellated lizards, it is now possible to assess the genetic structure based on hundreds of nuclear neutral loci scattered across the genome (318 neutral AFLP markers) but also with markers exhibiting a selection signature (23 outlier AFLP markers). We compared the genetic structure of *L. lepida* populations based on neutral and non-neutral AFLP markers (the same from chapter 2) with microsatellites (chapter 5). All genetic structure analyses, based in any of the sets of nuclear markers, with or without the effect of selection, highlight *L. l. nevadensis* lizards as a well defined group, while the distinction of *L. l. ibe*rica populations is well supported by microsatellites or outlier AFLPs but not by neutral AFLPs. The subdivisions detected by Paulo (2001) and Miraldo *et al.* (2011) among *L. l. lepida* populations, based on mitochondrial DNA sequences, are not well supported by nuclear loci. Both neutral and

outlier AFLPs cluster all *L. l. lepida* populations together, and only with microsatellites it is possible to detect some weak substructure among *L. l. lepida* populations, clustering southern populations together and distinguishing them from a second cluster with central and western populations, although showing signs of extensive admixture between them.

AFLP markers are dominant and biallelic and thus are considered less informative than microsatellites, which are codominant and multiallelic. Therefore, four to ten times more AFLPs must be used as compared to microsatellites to achieve comparable results (Mariette *et al*. 2002). The number of AFLPs (318 markers) used largely exceeded the number of microsatellites (8 loci). Thus, AFLPs should be more representative of the whole genome diversity of *L. lepida* than a few hypervariable microsatellites (Väli *et al.* 2008). However, neutral AFLPs failed to distinguish *L. l. iberica* from the nominal subspecies. On the other hand, even though microsatellites have enough resolution to separate *L. l. iberica* populations from *L. l. lepida*, the differentiation results mostly from a loss in genetic diversity in *L. l. iberica* populations, with a marked reduction in allelic richness and heterozigosity, and the fixation of a single allele at two of the least polymorphic microsatellites. In the opposite extreme of the environmental cline, *L. l. nevadensis* presents high levels of genetic diversity, with high allelic richness and heterozigosity at most microsatellite loci, and a remarkably high number of private alleles (22 alleles). When using the small set of outlier AFLPs (23 markers) for the genetic structure analysis, both *L. l. iberica* and *L. l. nevadensis* can be clearly distinguished from the nominal species when the existence of three clusters is assumed.

Information collected from nuclear loci in European ocellated lizards in this work (chapters 2-5) confirms early predictions that the evolution of the species conforms to the genic view of the speciation process (Wu 2001), offering a snapshot of different stages of speciation. *Lacerta lepida* is a well structured species, where *L. l. nevadensis* constitutes a monophyletic group that may have been diverging since the Upper Miocene, more than 9 million years ago (Paulo *et al.* 2008; Miraldo *et al.* 2011). The convergence between the African and the Eurasian tectonic plates during the Miocene led to the progressive uplift of the Betic mountains, forming a set of islands that later evolved to a land bridge between the Iberian Peninsula and the north-western coast of Africa (Braga *et al*. 2003). The Betic mountains correspond to a substantial portion of the current distribution of *L. l. nevadensis*, suggesting that the evolution of this subspecies might have begun with a process of rapid adaptive divergence of the first lizards colonizing the Betic mountains.

A similar process has been documented in three different lizard species in the Tularosa Basin of New Mexico (Rosenblum *et al.* 2007; Rosenblum & Harmon 2011). The evolution of cryptic colouration might have been particularly important for ocellated lizards adaptation in the Betic mountains, and *Mc1r* sequence variation suggests the evolution and fixation of an haplotype with a putative partial loss of function, which could explain the prevalence of brown/grey in the dorsal colour pattern of *L. l. nevadensis*. What might have started as divergence in a few adaptive loci has long ago spread to adjacent regions of the genome, being reinforced by the evolution of reproductive isolation. *L. l. nevadensis* is probably in the final stages of their speciation and its upgrade to the species level was already suggested by Paulo *et al.* (2008).

Another stage of speciation is illustrated by *L. l. iberica*. The divergence of *L. l. iberica* from the nominal subspecies is much more recent and probably started in the early Pleistocene, at approximately 1.5 million years ago (Paulo *et al*. 2008; Miraldo *et al*. 2011). The influence of cold weather during glacial cycles of the Pleistocene was more pronounced at northern latitudes in the Iberian Peninsula, which probably led to a retreat of ocellated lizards to southern refuges during colder periods and their further expansion to the north as the climate warmed up and became suitable to thermophilic species such as lizards (Schmitt 2007). The recent expansion of lizards that inhabit the northwest explains the reduction in allelic diversity observed in microsatellite loci. On the other hand, *L. l. iberica* genomic divergence from the nominal subspecies based in AFLPs results mainly from a few adaptive loci, which could have increased in frequency as a result of selection in refugial populations and spread during range expansions (Hewitt 2004). The climatic conditions faced by *L. l. iberica* in the northwest were even more adverse in the past than they are today. Some morphological (Mateo & Castroviejo 1990; Mateo & López-Jurado 1994) and life history variation (Mateo & Castanet 1994) exhibited by *L. l. lepida* might have been crucial for their success in the northwestern periphery of ocellated lizards' distribution. The apparent absence of strong restrictions to gene flow between *L. l. iberica* and *L. l. lepida* at neutral regions of the genome suggests that *L. l. iberica* is still at the early stages of speciation, when the process of divergence might still be reversible (Wu 2001), thus preventing *L. l. iberica*'s speciation process to progress to completion as it has happened with *L. l. nevadensis*.

Finally, divergence detected among *L. l. lepida* populations in mitochondrial DNA (Paulo 2001; Paulo *et al.* 2008; Miraldo *et al.* 2011) is not fully corroborated by the

nuclear markers analysed in this work. Because divergence in mitochondrial DNA within the nominal subspecies is of recent origin (Paulo *et al.* 2008; Miraldo *et al.* 2011), incomplete lineage sorting at nuclear loci might explain the incongruence with mitochondrial DNA, but current gene flow at zones of secondary contact can also contribute to homogenize genetic variation among *L. l. lepida* populations (Miraldo *et al.* 2011).

Overall, the results from this work highlight the importance of environmental heterogeneity for the evolution of ecological divergence between populations in continuous distribution areas. Dramatic environmental changes, such as the climatic oscillations of the Quaternary, might lead to the spatial separation of populations due to habitat loss. In these circumstances, less time is required for the emergence and fixation of locally adapted phenotypic traits, since populations become isolated and their effective size is smaller. However, when the habitat between allopatric isolates is restored due to the amelioration of the climatic conditions, diverging populations will expand and establish contact zones (Hewitt 1999). At this point, if divergence accumulated is small and environmental differences are also small, imposing weak directional selection to parapatric populations, hybridization might gradually dilute incipient divergence acquired in allopatry (Doebeli & Dieckmann 2004; Via 2009). However, at the gradient extremes, early stages of ecological speciation might easily progress towards the evolution of new species if strong directional selection is maintained, preventing the disruption of adaptive genotypes by recombination through the evolution of assortative mating or the maladaptation of hybrid phenotypes (Via 2009).

6.2 – Concluding remarks

The present thesis provided an extensive analysis of genomic variation in the European ocellated lizards and constitutes an important contribution to understand their evolutionary history in the light of natural selection. This study represents the first attempt to identify and characterize genomic regions underlying putative adaptive traits in ocellated lizards, and highlights the many challenges faced when pursuing such task in

a species with scarce genomic resources. The main achievements of the present work were:

1. Hundreds of nuclear markers were developed for ocellated lizards with the AFLP technique and used in a genome scan for selection with populations located along an environmental gradient in the Iberian Peninsula. It was possible to detect 23 anonymous AFLP markers with outlier behaviour for further research, representing candidate loci putatively under the effect of directional selection, which exhibit higher differentiation among populations than expected under a scenario of neutral evolution.

2. Several associations were detected between outlier's frequency and environmental variables such as insolation, temperature or precipitation, which have a major role in the climatic asymmetries along the environmental gradient in the Iberian Peninsula, affecting the landscape and *Lacerta lepida*'s activity cycles. These associations highlight climatic variables as putative selective forces acting over ocellated lizards, but further research is needed to confirm a causal relationship.

3. The characterization of 12 anonymous AFLP outlier loci was attempted through their isolation from agarose gels, followed by cloning and sequencing reactions. Seven of the outlier loci were successfully sequenced but none showed homology with known coding regions, suggesting that most AFLP outliers investigated might act as regulatory regions or be in linkage with the actual target of directional selection. Repetitive elements and indels were identified as the sources of length polymorphism in outlier AFLP fragments. Further genomic resources are needed to understand the relative position of AFLP outliers in the genome of ocellated lizards and their importance in the evolution of the subspecies located at the opposite extremes of the climatic gradient (*L. l. iberica* and *L. l. nevadensis*).

4. The investigation of *melanocortin-1 receptor* (*Mc1r*) as a candidate gene for melanin-based colour variation in ocellated lizards resulted in the identification of two nonsynonymous mutations. The mutation T162I is a derived and nonconservative substitution associated with *L. l. nevadensis* brownish phenotype, in which the isoleucine residue might lead to a putative loss of

function. The mutation S172C is a conservative amino acid change but it corresponds to a remarkably conserved position in reptiles. The cysteine residue in position 172 is associated with the prevalence of black scales in *L. l. lepida* and *L. l. iberica*. Despite the higher proportion of black scales present in the dorsal colouration of *L. l. iberica*, no mutations in *Mc1r* were found that could distinguish *L. l. iberica* from *L. l. lepida*, suggesting that its phenotype might be affected by regulatory mutations or by other pigmentation genes.

5. The three parapatric subspecies of European ocellated lizards represent a good model to study different stages of speciation. *L. l. nevadensis* is at the final stages of speciation, presenting deep morphological and genetic divergence, at both neutral and non-neutral loci, and the reduced gene flow with *L. l. lepida* in the absence of physical barriers suggests that mechanisms of reproductive isolation have evolved. *L. l. iberica* presents less pronounced morphological differentiation, but assuming that the hundreds of AFLP markers investigated here are representative of the ocellated lizards' genomic variation, *L. l. iberica* divergence is mostly explained by a few non-neutral loci. Thus, *L. l. iberica* might represent an early stage of speciation, when divergence is restricted to adaptive loci but the homogeneity at most neutral regions of the genome is maintained due to the small time of divergence and ongoing gene flow.

6.3 – Future directions

The investigation of genes underlying adaptive traits in non-model species is a daunting task. The present work lead to important finds for this major quest in ocellated lizards, but many question remain unanswered and will require the collection of much more detailed phenotypic, ecological and genotypic data.

To further characterize AFLP outlier loci in ocellated lizards and disentangle the causes for their selection signature, additional genomic resources are required to extend the known outlier sequence into their flanking regions, either by using genome walking

strategies, by constructing and screening BAC libraries or through the use of next generation sequencing technologies. Ideally, controlled crosses in captivity should be attempted to conduct QTL analysis or construct linkage maps, although it would require several years and plenty resources to obtain an F2 generation. These analyses would help to find associations between outlier loci and phenotypic traits that might be affected by them.

Regarding *Mc1r* variants, it will be essential to conduct *in vitro* functional assays to demonstrate the effect of amino acid changes in melanin production and prove the putative loss of function of *Mc1r* alleles in *L. l. nevadensis* colour phenotype. A detailed analysis of phenotypic and genetic variation in the secondary contact zone between *L. l. nevadensis* and *L. l. lepida* will prove useful to understand how often hybridization occurs, how phenotypic traits express in hybrids and their effect in fitness. Current understanding about the mechanisms that promote reproductive isolation of *L. l. nevadensis* is truly incipient, but some clues might lead future investigation. First, the role of differences in lateral blue eyespots between subspecies in mate choice should be tested, since blue eyespots are expected to be important in social signaling. Second, if brownish coloration is selected for crypsis with the background, what happens in the contact zone? Is the transition in the landscape at the contact zone as abrupt as the transition observed in both phenotypes and genotypes? Third, testing the efficiency of each colour morph to avoid predation in each side of the contact zone would clarify the adaptive value of colour variation in the southeast of the Iberian Peninsula.

Other candidate genes for coloration should be investigated to uncover the genetic basis for the increased melanization observed in *L. l. iberica*. The analysis of populations along and across the contact zone with the nominal species could bring new insights about the adaptations being selected at northern latitudes in ocellated lizards' and explore their tolerance to colder weather regimes.

6.4 – References

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **263**, 1619-1626.

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, **14**, 2899-2914.

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.

Braga JC, Martin JM, Quesada C (2003) Patterns and average rates of late Neogene - recent uplift of the Betic Cordillera, SE Spain. *Geomorphology*, **50**, 3-26.

Busack S, Visnaw JA (1989) Observations on the natural history of *Lacerta lepida* in Cádiz province, Spain. *Amphibia-Reptilia*, **10**, 201-213.

Butlin RK (2010) Population genomics and speciation. *Genetica*, **138**, 409-418.

Caballero A, Quesada H, Rolan-Alvarez E (2008) Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, **179**, 539-554.

Colbeck GJ, Turgeon J, Sirois P, Dodson J (2011) Historical introgression and the role of selective vs. neutral processes in structuring nuclear genetic variation (AFLP) in a circumpolar marine fish, the capelin (*Mallotus villosus*). *Molecular Ecology*, **20**, 1976-1987.

Doebeli M, Dieckmann U (2004) Adaptive dynamics of speciation: spatial structure. In: Dieckmann U, Doebeli M, Metz JAJ, Tautz D (eds) (2004) *Adaptive speciation*. Cambridge, UK: Cambridge University Press.

Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a

comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution*, **62**, 1162-1181.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285-298.

Foll M, 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.

Garcia-Borron JC, Sanchez-Laorden BL, Jimenez-Cervantes C (2005) Melanocortin-1 receptor structure and functional regulation. *Pigment Cell Research*, **18,** 393-410.

Gompel N, Prud'homme B (2009) The causes of repeated genetic evolution. *Developmental Biology*, **332**, 36-47.

Gratten J, Beraldi D, Lowder BV *et al.* (2007) Compelling evidence that a single nucleotide substitution in *TYRP1* is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **274**, 619-626.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87-112.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **359**, 183-195.

Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, **97**, 222-234.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101-104.

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity*, **99**, 364-373.

Joost S, Kalbermatten M, Bonin A (2008) Spatial analysis method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources*, **8**, 957-960.

Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology*, **15**, 3469-3480.

Kingsley EP, Manceau M, Wiley CD, Hoekstra HE (2009) Melanism in peromyscus is caused by independent mutations in *Agouti*. *PLoS One*, **4**, e6435.

Kirk H, Freeland JR (2011) Applications and implications of neutral versus non-neutral markers in molecular ecology. *International Journal of Molecular Sciences*, **12**, 3966-3988.

Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE (2010) Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, **365**, 2439-2450.

Mariette S, Le Corre V, Austerlitz F, Kremer A (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Molecular Ecology*, **11**, 1145-1156.

Mateo JA, Castanet J (1994) Reproductive strategies in 3 Spanish populations of the ocellated lizard, *Lacerta Lepida* (Sauria, Lacertidae). *Acta Oecologica*, **15**, 215-229.

Mateo JA, Castroviejo J (1990) Variation morphologique et révision taxonomique de l'espèce *Lacerta lepida* Daudin, 1802 (Sauria, Lacertidae). *Bulletin du Muséum National D'Histoire Naturelle, Paris*, **12**, 691-706.

Mateo JA, López-Jurado LF (1994) Variaciones en el color de los lagartos ocelados; aproximación a la distribución de *Lacerta lepida nevadensis* Buchholz 1963. *Revista Española de Herpetología*, **8**, 29-35.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in drosophila. *Nature*, **351**, 652-654.

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Molecular Ecology*, **17**, 1552-1563.

Miraldo A (2009) *Phylogeography and population dynamics of secondary contact zones of Lacerta lepida in the Iberian Peninsula*. PhD thesis, University of East Anglia, Norwich.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Mundy NI (2005) A window on the genetics of evolution: *Mc1r* and plumage colouration in birds. *Proceedings of the Royal Society, Series B: Biological Sciences*, **272**, 1633-1640.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316-336.

Paulo OS (2001). *The phylogeography of reptiles of the Iberian Peninsula*. PhD thesis, University of London, London.

Paulo OS, Pinheiro J, Miraldo A *et al.* (2008) The role of vicariance vs. dispersal in shaping genetic patterns in ocellated lizard species in the western Mediterranean. *Molecular Ecology*, **17**, 1535-1551.

Pérez-Figueroa A, García-Pereira M, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267-2276.

Raia P, Guarino FM, Turano M *et al.* (2010) The blue lizards spandrel and the island syndrome. *BMC Evolutionary Biology*, **10**, 289.

Reddy MK, Nair S, Sopory SK (2002) A new approach for efficient directional genome walking using polymerase chain reaction. *Analytical Biochemistry*, **306**, 154-158.

Richter-Boix A, Quintela M, Segelbacher G, Laurila A (2011) Genetic analysis of differentiation among breeding ponds reveals a candidate gene for local adaptation in *Rana arvalis*. *Molecular Ecology*, **20**, 1582-1600.

Robertson A (1975) Gene frequency distribution as a test of selective neutrality. *Genetics*, **81**, 775-785.

Römpler H, Rohland N, Lalueza-Fox C *et al.* (2006) Nuclear gene indicates coat-color polymorphism in mammoths. *Science*, **313**, 62.

Rosenblum EB, Harmon LJ (2011) "Same same but different": replicated ecological speciation white sands. *Evolution*, **65**, 946-960.

Rosenblum EB, Hickerson MJ, Moritz C (2007) A multilocus perspective on colonization accompanied by selection and gene flow. *Evolution*, **61**, 2971-2985.

Rosenblum EB, Hoekstra HE, Nachman MW (2004) Adaptive reptile color variation and the evolution of the *Mc1r* gene. *Evolution*, **58**, 1794-1808.

Rosenblum EB, Römpler H, Schöneberg T, Hoekstra HE (2010) Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences of the United States of America,* **107**, 2113-2117.

Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4**, 11.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158-170.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Väli U, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, **17**, 3808-3817.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139-151.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the Unites States of America*, **106**, 9939-9946.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611-619.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123-3135.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851-865.