UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL


# On the evolutionary history of the iberian Sooty Copper butterflies

Renata Filipa Ribeiro Martins

Mestrado em Biologia Evolutiva e do Desenvolvimento

2011

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL


# On the evolutionary history of the iberian Sooty Copper butterflies

Renata Filipa Ribeiro Martins

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por :

- Professor Doutor Octávio Paulo

- Mestre Eduardo Marabuto

2011

**“Il faut bien que je supporte deux ou trois chenilles si je veux connaître les papillons. Il paraît que c'est tellement beau.”** – Antoine de Saint-Exupéry *in* Le Petit Prince

# Nota prévia

A presente tese encontra-se escrita na língua inglesa para que seja possível a sua divulgação pela comunidade científica internacional. Apenas a secção de agradecimentos estará em português para que a sua leitura seja possível a todas as pessoas mencionadas. A revista científica internacional *Molecular Ecology* foi escolhida como modelo para a estruturação das referências bibliográficas.

# Agradecimentos

Tenho de agradecer, em primeiro lugar, ao Professor Octávio Paulo e também à Doutora Sofia Seabra, como colaboradora deste projecto, por me terem aceitado como sua aluna na realização desta tese, apesar de todas as dificuldades a ela inerentes e pelas suas sugestões e interpretações ao meu trabalho.

Ao Eduardo Marabuto tenho a agradecer todo o projecto, já que este não existiria sem a sua ideia inicial. Obrigada pela confiança de depositares em mim as tuas borboletas, fiz o melhor para lhes fazer justiça.

Agradeço a todos os que ajudaram na amostragem e trabalho de campo: Ana Sofia, Tiago Magalhães, Manuel Dinis Cortes, Pedro Pires e Fernando Romão.

A todo o grupo Cobig2, um obrigado por toda a ajuda, desde problemas básicos de computador aos problemas intermináveis de laboratório, entre outros. À Sara, Catarina, Ana Sofia, Joana, Patrícia, Francisco, Bruno, Vera e Carla, obrigado pela paciência e pela boa-disposição. Queria agradecer em particular à Inês Modesto e Diogo Silva a grande ajuda que me deram nos tempos mais difíceis, por ficarem comigo a tentar o impossível para resolver os problemas da tese.

Aos que passaram o ano também no 3º piso do C2 entre quem partilhámos as dores e os triunfos de uma tese, um obrigado e um boa sorte!

Um grande e sentido agradecimento é devido aos meus amigos que me aturaram (em demasia, até) nos altos e baixos que se apresentam quando se constrói uma dissertação. Obrigada por todo o apoio que tinham disponível sempre que precisei, pelos momentos de descontracção que me souberam pela vida durante este ano. Os almoços e cafés e lanches e jantares foram mesmo importantes para manter a sanidade mental que ainda me resta! Obrigado Adriana, Cátia, Francisco, Gonçalo, João, Marta, Sofia e Susana! Para os que não estiveram lá todos os dias, mas cuja força foi também muito importante vai um grande obrigado: Pio, Marta L. e Lena!

O mais importante agradecimento é aos meus pais, claro. Por me terem permitido e dado todas as condições para frequentar o mestrado e, mais importante, por me educarem, por todo o apoio e amor.

# Resumo

As grandes alterações climáticas do Pleistocénico desempenharam um papel central na distribuição actual dos padrões biológicos de diversidade que encontramos hoje em dia. As glaciações provocaram grandes alterações na distribuição das espécies que teriam de encontrar condições climáticas adequadas para a sua sobrevivência. Para isso, muitas espécies deslocaram a sua área de distribuição para latitudes inferiores, onde encontrariam refúgios climáticos, já que estariam impossibilitadas de usar os habitats das regiões mais setentrionais. Particularmente a Europa estaria coberta por grandes calotes de gelo no norte presentes também nas zonas montanhosas, mas manteria habitats mais quentes no sul. A Península Ibérica foi, indiscutivelmente, um dos refúgios de maior importância devido à sua heterogeneidade geográfica e climática, que providenciaria habitats diversos desde florestas de coníferas até semi-desertos. Inúmeros estudos têm-se debruçado sobre esta importância e as evidências observadas para vários grupos animais, de répteis a mamíferos e invertebrados, têm demonstrado que esta península era não só um refúgio importante mas seria também de onde expandiriam as espécies que iriam colonizar o norte da Europa, durante os períodos interglaciais.

Estes movimentos de contracção e expansão das espécies foram constantes e cíclicos e levaram ao aparecimento de várias zonas de contacto secundário. Estas zonas caracterizam-se pela sobreposição de taxons diferentes, sejam eles filogeneticamente próximos ou não. Nestas zonas de contacto secundário é possível estudar vários processos que têm intrigado os investigadores evolutivos: como se mantém a identidade da espécie face a contacto com outros taxa numa área de simpatria; que barreiras existem que possam evitar ou impedir totalmente o cruzamento com entidades distintas e que forças selectivas actuam sobre os genótipos e fenótipos híbridos que podem resultar nestas circunstâncias, entre outros.

Algumas espécies são particularmente sensíveis a pequenos distúrbios nas condições do habitat e as borboletas são especialmente importantes em estudos deste género já que, na sua grande maioria, as espécies de Lepidoptera são especialistas ecológicos estrictos. Para além desta particularidade, as borboletas apresentam um conjunto de características que as tornam bons modelos de estudo dos padrões filogeográficos. Têm

por um lado uma grande capacidade de dispersão a nível global, mantendo algum sedentarismo quando analisamos em termos individuais e por outro, excepto nalguns casos, mantêm populações grandes e constantes, que não confundirão sinais filogeográficos com sinais de estruturação demográfica.

Neste trabalho foi escolhida a espécie *Lycaena bleusei*, uma borboleta endémica da Península Ibérica, com uma distribuição restricta às zonas montanhosas do Sistema Central Ibérico. Esta espécie foi considerada como uma subespécie da Acobreada escura, *Lycaena tityrus*, até muito recentemente, mas foi-lhe atribuído o nível de espécie devido não só à sua distribuição discrepante, mas também a diferenças morfológicas subtis. O facto de esta história evolutiva ser pouco conhecida e da dificuldade de perceber os limites entre as duas espécies terá mantido a lacuna de conhecimento ecológico, comportamental e genético que caracteriza esta espécie. Alguns autores têm, recentemente, sugerido que a falta de prospecção da área ocupada pela *L. bleusei* pode estar a camuflar uma hipotética zona de simpatria em Portugal, onde as duas putativas espécies-irmãs estarão em contacto. De facto, e apesar de já ser considerada como uma espécie independente, estudos genéticos que incluam esta espécie são ainda raros ou mesmo inexistentes.

Assim, utilizando a *Lycaena bleusei* como modelo, esta tese de mestrado pretende centrar-se na resolução das relações filogenéticas da espécie, enquadrando-a no género *Lycaena* e também na clarificação da sua história evolutiva, tendo por base a comparação com a espécie largamente distribuída na Europa, *L. tityrus*. A possibilidade de existir uma banda geográfica de simpatria entre estes dois taxon foi também avaliada, visando a possibilidade de ocorrer hibridação e introgressão neste grupo. Para tal, foi feita amostragem na área de distribuição da *Lycaena bleusei*, que cobriu não só a área de distribuição alopátrica, mas também a área de suposta simpatria. Populações de *Lycaena tityrus* foram também amostradas nesta zona, para serem utilizadas como base de comparação em termos ecológicos e genéticos, assim como algumas populações gregas. A análise foi feita com recurso a dois marcadores moleculares, um mitocondrial e um nuclear, citocromo oxidase I (COI) e Elongation factor 1 alpha (EF-1α) respectivamente. Estes marcadores têm sido utilizados em diversos estudos filogenéticos e a sua capacidade de resolução das relações nos diferentes níveis de classificação (*i.e.* subespécie, espécie, género, etc.) está bem descrita na literatura actualmente.

Os dois loci utilizados provaram ser suficientes para distinguir as duas espécies em estudo, sendo possível atribuir a cada uma delas uma linhagem genética individual, quer mitocondrial quer nuclear, de forma clara e indubitável. Desta forma, foi possível posicionar pela primeira vez a espécie *L. bleusei* dentro do seu género com base em análises genéticas, aparecendo como esperado num clade monofilético com a sua espécie-irmã, *L. tityrus*. De acordo com as ideias descritas por outros autores, as espécies deste género que se encontram distribuídas pela região palearctica formam também um clade monofilético e as espécies americanas estão também agrupadas desta forma, excepto *L. helloides* e *L. cupreus* que se incluem no clade anterior. A análise dos sinais filogeográficos, dada quer pelas árvores quer pelas redes haplotípicas, indica uma clara separação das duas espécies, permitindo assim afirmar que, de facto, como sugerido por outros autores estamos perante duas espécies distintas, com uma diferenciação bem marcada (distância genética de 3,3% com o COI e de 2,3% com o EF-1α). A divergência entre as duas espécies parece ter ocorrido há cerca de 1 milhão de anos, correspondendo a um período interglacial relativamente extenso, o Waaliano. A par destes resultados, vem também a evidência de que cada espécie apresenta uma história demográfica independente, sendo que a *Lycaena bleusei* parece apresentar um aumento de efectivo populacional, correspondente a uma expansão demográfica que se terá iniciado há cerca de 30 000 anos. A sua espécie-irmã, *L. tityrus*, apresenta por outro lado um efectivo populacional que se tem mantido constante desde a sua divergência.

No decorrer deste trabalho foi possível analisar detectar uma zona de simpatria, que mostrou ser mais alargada do que inicialmente previsto. Foram encontradas populações mistas das duas espécies na encosta oeste da Serra da Estrela, numa banda geográfica que se estende até ao norte de Portugal, onde inicialmente só se conhecia a espécie *L. tityrus*. Assim, à luz das condições actuais e ao facto de termos obtido um sinal genético que indica expansão demográfica da espécie *L. bleusei*, sugere-se que esta espécie está em crescimento demográfico, ocupando nichos ecológicos que anteriormente poderiam conter apenas populações de *L. tityrus*.

É comummente aceite que hibridação é um evento relativamente comum entre os Lepidópteros e muitos são os casos já descritos na bibliografia e assim, na presença de uma zona de simpatria, era esperada a ocorrência de eventos de hibridação e introgressão. De

entre toda a amostragem, dois indivíduos amostrados a norte de Portugal mostraram incongruências entre as duas linhagens genéticas utilizadas, linhagem mitocondrial (e características morfológicas) da espécie *L. tityrus* mas linhagem nuclear de *L. bleusei*. Os resultados obtidos neste trabalho parecem indicar que se trata de um processo raro ou incomum, podendo estar associado a questões genéticas e/ou a características morfológicas, comportamentais e ecológicas.

**Palavras-chave**: Filogeografia, Ciclos climáticos, *Lycaena*, Simpatria, Hibridação.

# Abstract

From an evolutionary point of view the study of genetic diversity and the way it is distributed in the geographic space has always been of interest. Phylogeography tries to understand how the present distribution patterns of species and genetic lineages were shaped. The major climatic changes from the Quaternary, characterized by glacial and interglacial periods, led to species movements towards the South to find suitable habitats and climatic conditions during the glacial periods. One of the most important refugium in Europe was the Iberian Peninsula, where species could find a great variety of habitats. Studying species that are found today within this peninsula provides insights on how these global-scale events affected the distribution of genetic variation.

In this research thesis, two butterfly species were used, *Lycaena tityrus* and *Lycaena bleusei*, whose evolutionary history and taxonomic classification have always been problematic. The Iberian Sooty Copper, *L. bleusei*, was considered, until very recently, as a subspecies of the Sooty Copper, *L. tityrus*, but was brought to species level for its distinct distribution and morphologic characteristics. Two molecular markers were employed, one mitochondrial gene – Cytochrome oxidase subunit I and one nuclear gene – Elongation factor 1α to assess: 1) phylogenetic relationships between these species among the genus *Lycaena*, clarifying their taxonomic relationship and 2) phylogeography and genetic structure of both species. Demographic changes and divergence times were also calculated, taking into account the suggested times for the Pleistocene changes, in order to understand how these climatic events influenced current species distributions and genetic architecture.

We also used the genetic histories provided by each loci to account for incongruence between the mitochondrial and nuclear lineages. A secondary contact zone is described in northern Portugal and mixed populations were found. From all the individuals, two proved to have incongruent molecular histories suggesting the presence of hybridization events between the two copper species.



**Keywords**: Phylogeography, Climatic cycles, *Lycaena*, Simpatry, Hybridization.

# Table of contents

# 1. Introduction

## *1.1 Past Climatic Events: shaping the present*

The distribution and evolution of species is extremely variable in time and space. The Quaternary period has been characterized by several glacial (cold) and interglacial (warm) periods through the Pleistocene and was initiated with the establishment of the Arctic ice cap. Since then ice sheets have advanced and receded in cyclic time periods. Reconstruction of paleoclimates has been facilitated by the increase of studies and recovery of fossils, mainly pollen and plant but also insects. In the latter, beetles have been of extreme importance, due to the clean preservation, since it is often possible to recover intact genitalia, which facilitates species identification and also development of new methods to reconstruct the environmental shifts throughout the climatic changes in a finer-scale (Elias 1996, 2000; Hewitt 1999).

Major shifts were felt differently throughout the globe, depending mostly on the conditions of each region. During glaciations, ice sheets advanced considerably and temperature and vegetation zones shifted towards the Equator. Mountain blocks (as the Alps and the Andes, for example) were also covered with ice and this accumulation led to a drop in sea level. As a consequence, land bridges formed and allowed passage to some species. In warmer areas, some species descended in altitude and rain forests were reduced and confined, allowing the expansion of desert and savannah. With these changes some species, or some populations, became extinct in some of their ranges and others, mainly temperate species, dispersed and found refugia, from where they could expand again in each interglacial period (Hewitt 2004; Schmitt 2007). All these range and demographic shifts would have occurred repeatedly, allowing adaptation to new environments and also to new neighbors. This has clear stochastic and selective effects in the genetic architecture and variation of species. Some populations and lineages went extinct, alleles are lost in bottlenecks and founder events and, with time, new mutations occur and spread through selection and demographic expansion (Hewitt 2004). These climatic changes are one of the great shapers of the distribution patterns of genetic variability observed nowadays.

Since 1987, when Avise first employed the term *phylogeography* (Avise *et al.* 1987), interpretative studies of the genetic consequences of these events have multiplied

quickly. These studies try to understand how the genetic patterns we find today can explain evolutionary history (either at species level and above or below species level) comparing, for example, intraspecific phylogeographic patterns from several taxa within the same region (Taberlet *et al.* 1998). Phylogeography uses molecular and geographic data to infer the role of historic factors in the distribution of current patterns of biodiversity (Avise 2000).

Geographic and phylogenetic distribution of genetic variability can be identified through the use of molecular markers. Most DNA sequences show a relatively low level of divergence and did not have enough time to mutate since the beginning of the Quaternary. With that, phylogeographic structure described for many natural populations seems to arise from *lineage sorting* within populations with fluctuating effective sizes ($N_e$) and DNA marker distribution differences among populations are probably due to sorting of mutations generated in the past (Randi 2007). Estimating the coalescence times *i.e.* tracing all alleles of a gene shared by all members of a population to the most recent common ancestor (MRCA) seems to be the most suitable framework for phylogeography (Rosenberg & Nordberg 2002). Because the mtDNA, as a genomic compartment, is just ¼ of the nuclear compartment, the average TMRCA is just N generation instead of 4N generation, consequently the two types of markers can infer different patterns and events.

## *1.2 Iberian Peninsula: refugia from the ice*

During cold periods, the Mediterranean Sea and the Black Sea acted as barriers to the dispersal of terrestrial biota towards the south and in warmer periods, mountain blocks acted as barriers towards the north. Several studies on animal and plant organisms points out that a large number of those which are widespread throughout Europe nowadays, had at least one refugium in southern Europe during glaciations peaks. When the weather warmed and the ice retreated, species expanded from their southern refugia towards the north. Each species responded individually to this temperature change as each needed to find its particular climatic niche. Southerner populations went extinct as their tolerance range also moved north or persiste in interglacial refugia on the top of mountains or buffered valleys and this movement is normally associated with a loss of genetic variation.

Southern Europe brings many opportunities for species to find a suitable habitat during climatic changes, due to the varied and rugged topography, warmer climate and thus, abundance of different habitats. This has clear consequences on the dispersal, genetic variation and retention of variability over time (Hewitt 1999). From several studies, it seems clear that there were three major refugium areas: the Iberian, Italic and Balkan peninsulas and a potential less known fourth one in the Caucasus/Black sea region. Populations from the three peninsulas would have limited to no contact during glaciations maximums, leading to a disjunction into three separate genetic groups, as found in many studied organisms, albeit admitting some exceptions (Schmitt 2007). Understanding the genetic variation of species found nowadays within these peninsulas provides insights on the recent history of these global-scale events. Shallow divergences are expected in such cases, indicating that deeper genetic signals were eroded during range shifts (Paulo *et al.* 2001).

Particularly, the Iberian Peninsula seems to have constituted a key refugium within Europe with greatest importance during the Pleistocene, being not only a differentiation center but also a repository of species after the ice ages. This fact has been demonstrated by innumerous phylogeographic studies and also, indirectly, by the high levels of endemism both in animal and plant species (about 31% of the 900 European endemic species). This may have been potentiated by its geographic features, with its mountain systems offering microclimates that allow survival of populations with an altitudinal range shift and also because it is subject to different regional climates, from evergreen subtropical forests to semi-deserts. It is easy to understand that this refugium should not be considered homogeneous, but instead should rather offer different suitable habitats with a wide range of characteristics (Gómez & Lunt 2007).

The post-glacial colonization routes from the three refugia show great concordance and this conservation and repetition in many animal and plant species allowed the establishment of (four) main patterns of colonization routes (Schmitt 2007). As a consequence of these different expansion routes, with geographic barriers to species distribution, several suture or contact zones are now considered in Europe and are not distributed randomly throughout the continent. The term contact zone was first employed by Remington (1968) to describe a geographic band where there is a geographic overlap of

biota, either two species or subspecies, that can hybridize in that area. Nowadays, the concept incorporates also different genetic lineages within the same formal taxon.

Where two sister taxa overlap, two extreme outcomes are possible. Either the interacting lineages exhibit complete reproductive isolation or they interbreed freely. However, between these two extremes there are some intermediate consequences that arise from secondary contact namely hybridization, reinforcement and introgression. Studying the mechanisms underlying a secondary contact zone has always been of interest to evolutionary biologists. These areas have been considered "a natural laboratory" for studies of evolutionary processes such as species formation or the extent of reproductive isolation but also the ability to maintain identity in the face of hybridization, which ultimately will lead to the ongoing discussion of boundaries between inter and intra-specific differentiation (Kuchta 2007).

Some species have the ability to give us more adequate data to test biogeographical patterns than others, but butterflies seem to gather a set of characteristics that render this group a pool of great model organisms. Many butterfly species meet the criteria used when selecting a study species for past climate events, such as a high ability to disperse to newly emerging suitable habitats while still maintaining some sedentarism at an individual level to minimize the distortion of the phylogeographic signal through migration, for example. Also, when established, populations shall be relatively large and stable to reflect true phylogeographic signals rather than population structure (Schmitt 2007). The number of studies concerning species dispersal throughout time that employ butterfly species has greatly increased. The rapid response of the butterflies to fine-scale habitat disturbance is well-known and used to understand ecological consequences of both anthropogenic and biotic changes, since most butterflies depend on very specific habitat conditions (see, for example, Warren *et al.* 2001; Hoyle & James 2005; Wallisdevries & van Swaay 2006). In fact, this arthropod group is also being used in many generalized studies of global warming and local climate changes (see, for example, Parmesan *et al.* 1999; Wilson *et al.* 2007 and Habel *et al.* 2011).

## *1.3 The Sooty Coppers* Lycaena tityrus *and* Lycaena bleusei *as models*

Genus *Lycaena* belongs to the Lycaeninae subfamily of the Lycaenidae family, the second largest butterfly family with ca. 5000 species. This subfamily has two recognized tribes: Lycaenini Leach, 1815 and Heliophorini Geyer, 1832. The true coppers of the genus *Lycaena* are placed in the former, forming a group of around 80 species with allied genera Anthamanthia, Hyrcanana and Phoenicurusia of Palearctic distribution (Bozano & Weindenhoffer 2001).

The sole use of morphological characters has led, so far, to a bewildering subdivision of the subfamily, and particularly the tribe Lycaenini. Even the subdivision of Lycaeninae into two tribes is only weakly supported by morphological characters (Jong & van Dorp 2006). Following the same boundary definition problems, the genus *Lycaena* has also been controversial and attempted divisions into subgenera have promptly failed, due to the great homogeneity within this group. Many genera and subgenera have been proposed to define associations between species according to morphological differences in male genitalia, however, as most species have intermediate characters it is better to considered the genus as an unit. It seems that the consensus is now in the arrangement of the *Lycaena* species into eleven morphological species groups proposed by Bozano & Weidenhofer (2001) for Palearctic species and Brower (2008) for the Nearctic, South African and New Zealand species.

*Lycaena tityrus* Poda, 1761 is a widespread temperate zone butterfly, ranging from Western Europe to central Asia (Karl *et al.* 2009; Bozano & Weindenhoffer 2001). The species is bivoltine with two discrete generations per year in most parts of its range, although populations with one or three annual generations may occur (Maravalhas 2003). Central European low-altitude populations are typically bivoltine and are on the wing from April to September. *Lycaena tityrus* is commonly known as the Sooty Copper due to its dark brown upperside coloration, contrasting with most of the other more brightly coloured Copper butterflies (genus *Lycaena*) in the Palearctic. Wingspan is about 28-30 mm and the underside is lightly coloured with small black markings and a row of orange spots near the wing margin. The species exhibits a distinct sexual dimorphism, where males are generally dark brown with more pronounced discal and postdiscal blackspots on the forewings and

faint marginal orange lunules in both wings. Female upperside wings show a more pronounced orange coloration and, typical of Copper butterflies, bear more rounded wings (Figure 1).


**Figure 1** *Lycaena tityrus* on its natural habitat: upperside wing (A) and underside wing (B).

The Sooty Copper colonizes different types of wetland as well as natural unimproved grassland such as swampy clearings or mountainous canyons and ridges, between 100 – 1000 meters, although alpine populations can be found up to 2020 m. (Karl 2008). Adult butterflies predominantly feed on composite plants (Compositae). Larvae of the last brood enter diapause, overwintering half-grown in the third instar. Pupation occurs after the completion of four larval instars and the onset of Spring. Main larval host-plants are docks, especially common surrel (*Rumex acetosa*), but also *R. acetosella* and *R. scutatus* (Karl 2008; Maravalhas 2003). Although this species is still relatively widespread in most parts of Europe, the intensification of agriculture caused a clear decline in population numbers associated with local and regional extinctions in some parts of its range. Thus, in many parts of Europe, this species is considered to be vulnerable (van Swaay & Warren 1999).

There is one subspecies assigned to *Lycaena tityrus*, *L. t. subalpinus* Speyer, 1851. Its taxonomic status has been largely discussed since there is, in one hand, a clear geographic isolation from the nominal species in the Alps, where it seems to have reached some differentiation but, on the other hand, these morphological and ecological differences are not so clear in the populations of the Pyrenees (Bozano & Weidenhoffer 2001).

The Iberian Sooty Copper, *Lycaena bleusei* Oberthür, 1884 was until very recently regarded as a segregate taxon within *L. tityrus* as *L. t. bleusei*, but was raised to species

level due to its well defined diagnostic morphological characters and geographic isolation within the Iberian Central Mountain System (Cassulo *et al.* 1989; Bozano & Weindenhoffer 2001). Despite its restricted distribution the Iberian Sooty Copper is listed as Least Concern in IUCN Red List, but the concerns for the destruction of suitable habitats in mountainous region should also apply in this case.

Morphologically similar to *L. tityrus* in the first brood, the second generation, with a peak in August, presents a small but distinct tail in the inferior wing and more pronounced black spots on a bright orange background of the upperside wings (Figure 2).


**Figure 2** *Lycaena bleusei* on its natural habitat: upperside wing (A) and underside wing (B).

Due to the confounding taxonomic notion of the Iberian Sooty Copper, its range as well as of *L. tityrus* is still poorly defined in the Iberian Peninsula. It is certainly the only species from this genus found exclusively in the Iberian Central System, from Sierra de Guadarrama, Spain to the west in Serra da Estrela and Serra da Gardunha, Portugal at 700-1000 m of altitude. García-Barros *et al.* (2004) suggested that the distribution of *L. bleusei* and *L. tityrus* could converge in a sympatric area north of Serra da Estrela, however without data to support this fact. The formal finding of *L. bleusei* itself in Portugal is fairly recent, after having been located during recent field work some *L. bleusei* populations in Serra da Estrela in areas close to known *L. tityrus* populations (Marabuto *et al.* 2004) (Figure 3).

Although *Lycaena bleusei* is considered a good species there is no phylogenetic analyses of the genus *Lycaena* that include individuals from this species, thus its real positioning is still unattended. It is expected to behave as a sister species of *Lycaena tityrus*, clustering with the Sooty Copper in a monophyletic group.

Both species belong in the *L. virgaureae* species group, with *L ottomanus* Lefébvre, 1830 and *L. alciphron* Rottemburg, 1775. All the species in this group have a fairly similar distribution to *L. tityrus*, with the exception of *L. ottomanus* that is more restricted to the south east of Europe, where it is now considered a vulnerable species (van Swaay & Warren 1999).

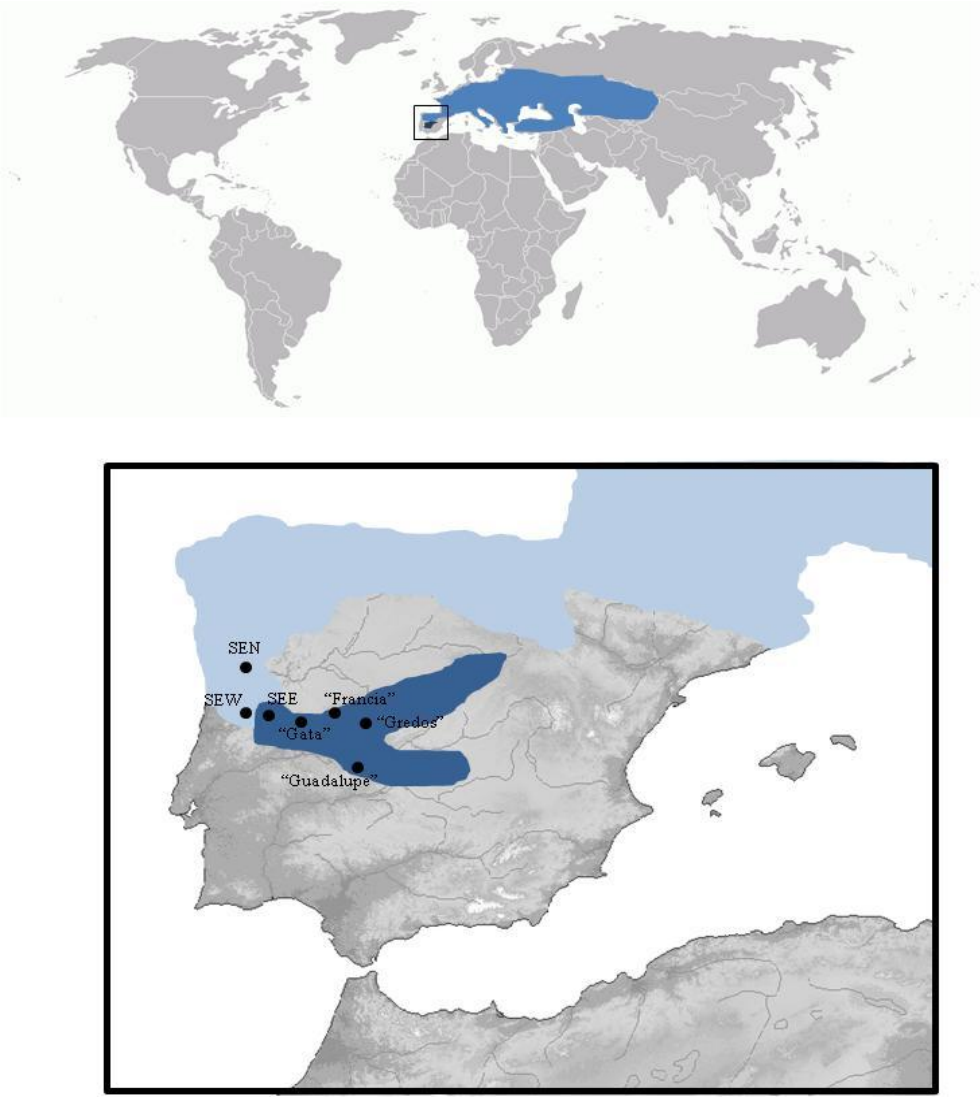


**Figure 3** Map of the distribution of *Lycaena tityrus* and *Lycaena bleusei* with detail on the Iberian Peninsula. Names on map correspond to sampling groups: SEN – Serra da Estrela North; SEW – Serra da Estrela West; SEE – Serra da Estrela East.

Morphological characters used to define the morphologic species groups have been, however, considered to bear insufficient resolution for a correct phylogenetic inference (Jong & van Dorp 2006) and only recently molecular markers have been employed to infer the genetic relationships within the genus *Lycaena*. The phylogenetic reconstruction of van

Dorp *et al.* (2004, 2006), using two mtDNA genes, COI and COII, focused mainly in the relative positions of the south African and New Zealander species within the tribe and the genus and it was suggested that the two tribes of the Lycaeninae are not monophyletic and the inclusion of the south African species forces the genus into a paraphyly (Jong & van Dorp 2006). Relationships among *Lycaena* suggest three major groups within the genus, the bulk of diversity in the Palearctic, a smaller Nearctic and a third, New Zealand species (van Dorp 2004). However, as the authors admit, the whole study might be compromised due to outgroup choices or the non-inclusion of enough species spanning the whole diversity within the genus. Further, in the light of these results, the authors themselves advise to take caution before taking any dramatic steps in the definition of new divisions by, for example, confirming their results with a nuclear gene.

### *1.4 Selection of molecular markers*

Prior to the generalized utilization of molecular markers, systematics studies of most animal and plant groups were traditionally achieved with morphological data. With genetic markers easily available, contrasting classifications with the previous established ones are becoming frequent. Most species occupy very discrete non-overlapping clusters in the morphological space, and this fact has been used to evoke reproductive isolation between taxa that maintain this morphological gap in areas of sympatry (Oliver *et al.* 2007). However, many species may present variation in the morphological characters and the presence of intermediates or too subtle morphological differences can confound the species classification. In this case, two hypotheses can be drawn; either we are in the presence of polytypic species/subspecies or these individuals are hybrids. Since reproductive isolation should produce a pattern of genetic discontinuity between the two species, assaying the genetic diversity and differentiation should be a helpful tool to resolve the hybridization *vs.* intraspecific morphological variation concept.

When dealing with closely related species it is important to select a molecular marker with a relatively fast mutation rate that can resolve phylogenetic relationships within a short temporal scale. Mitochondrial DNA has been the classic choice for many phylogenetic and phylogeographic studies, since its applicability became easily available and affordable. Its popularity may derive from the relatively ease of isolation and

amplification, even in poorly preserved samples (Caterino *et al.* 2000). Cytochrome oxidase subunit I (COI) has proven to be a fit gene to resolve relationships between species and slightly above species level and is thus widely applied in phylogenetic studies (Hebert *et al.* 2004; Hajibabaei *et al.* 2006; Zink & Barrowclough 2008). Its mutation rate is high enough to distinguish closely related taxa still not achieving a saturation point where homoplasy predominate, thus making it useful in phylogeographic studies as well.

Phylogenetic studies using nuclear protein-coding genes are far fewer in number than studies applying mtDNA. However, the advantages of using combined data from both markers, mtDNA and nuclear DNA, are well-defined and most recent studies have abandoned the notion of defining genetic relationships based solely on one single molecular marker. Of the nuclear protein-coding genes, Elongation Factor 1α (EF-1α) has been the most used. Its sequences have proven very useful for studies among species groups and genera within (sub)families and support relationships that are usually congruent with the information given by mitochondrial COI-II. Because this is a protein-coding gene it is subject to purifying selection and so amino acids in EF-1α are highly conserved and third codon positions provide essentially all the phylogenetic information (Caterino *et al.* 2000) Several studies in Lepidoptera successfully applied this gene in phylogenetic and phylogeographic studies (Cho *et al.* 1995; Reed & Sperling 1999; Kim *et al.* 2010).

Another interesting feature of analyzing the outcome of two molecular markers is the confrontation of both evolutionary histories, either agreeing or contradicting each other. Since this study trawled into the relationships between two sister taxa either considered species or subspecies and their supposed biological interaction, the possibility of hybridization and introgression between *Lycaena tityrus* and *Lycaena bleusei* is also being assessed. If there is an incongruence between a nuclear and a mitochondrial gene it is an indication of a possible introgression between individuals of both taxa (Mallet *et al.* 2010). Put simply, if a mitochondrial lineage attributed to one species is found in organisms with a nuclear lineage of the other species then it becomes clear that there has been some exchange of genetic material between the two pools. In order to understand this question, genetic distances between the taxa involved given by each marker have to be high enough so as not to allow for ambiguities and to clearly separate both taxa in order to attribute to each one a specific lineage.

# 2. Main Goals

This work intends to investigate phylogenetic relationships among the species within genus *Lycaena* and the evolutionary and demographic history of a small-range species, *Lycaena bleusei* in contrast with a widespread and ecologically similar species, *Lycaena tityrus*. The possible existence of a sympatric area where the two closely related species, initially thought to have complete geographical isolation, are in contact, also brings the possibility to understand mechanisms than can either maintain two species as separate identities or change the genetic pool of any or both species.

Therefore, this research project aims to achieve the following goals:

- To understand the relative position of the species *Lycaena bleusei* within the genus *Lycaena*;
- To infer the phylogenetic relationship between *Lycaena tityrus* and *Lycaena bleusei*, taking in consideration genetic and morphologic congruencies;
- To investigate the evolutionary history of *Lycaena bleusei* and *Lycaena tityrus* throughout the past climatic events;
- To assess the evidence of hybridization and introgression between the two species along the sympatric distribution area in the Iberian Peninsula.

# 3. Materials and Methods

## *3.1 Sampling and sample preparation*

Specimens of *Lycaena tityrus* and *Lycaena bleusei* were collected in eight sampling sites (comprehending several populations) and identified based on morphologic characters. The sampling sites were chosen in order to cover the hypothetic sympatric area from Serra da Estrela and farther north but also the surrounding allopatric distribution of *L. bleusei*. Samples of *Lycaena alciphron*, *L. virgurae*, *L. phlaeas* and *Cigaritis alardi* were also collected and included for phylogenetic analyses. Specimens were collected alive using a entomological net and kept at -20ºC until DNA extraction. Several representative species of the genus available at NCBI Genbank were included in the phylogenetic analyses within the genus *Lycaena* (Table 5, Supplementary Data).

## *3.2 Molecular analyses*

For DNA extraction two legs per individual were used and another one was kept as stock. The rest of the body was preserved for future morphological analyses. DNA extraction was performed with E.Z.N.A® Tissue DNA Kit (Omega, Biotek) following the manufacturer's protocol. Two loci were amplified for each sample, the mitochondrial gene cytochrome oxidase subunit I (COI) and the nuclear coding gene Elongation Factor 1α (EF-1α). The amplification of the mtDNA fragment was performed with the primers LEPF (5'- AAT CAA CCA ATC ATA AAG ATA TTG G -3') and LEPR (5'- TAA ACT TCT GGA TGT CCA AAA AAT -3') (Hajibabaei *et al.* 2006). Polymerase Chain Reaction (PCR) were performed in a final volume of 25 µL, with 3 µL of DNA (10-50 ng/µL), 1x Colorless GoTaq® Flexi Buffer, 0.04 U of GoTaq® DNA Polymerase, 1.8 mM $MgCl_2$, 0.1 mM of dNTPs and 0.4 µM of each primer, according to the following protocol: an initial denaturation step of 5 min at 94ºC, followed by 35 cycles at 94ºC for 30 sec, 53ºC for 45 sec, 72ºC for 1 min with a final extension step of 72ºC for 7 min. Nuclear gene EF-1α was amplified with the primers ef44 (5'- GCY GAR CGY GAR CGT GGT ATY AC -3') and ELF1R (5'- GTT TCA ACT CTG CCT ACK GGC AC -3') (Kim *et al.* 2010) in a PCR mix

with a final volume of 25 µL, with 3 µL of DNA (10-50 ng/µL), 1x Colorless GoTaq® Flexi Buffer, 0.02 U of GoTaq® DNA Polymerase, 2 mM $MgCl_2$, 0.1 mM of dNTPs and 0.32 µM of each primer, according to the following protocol: an initial denaturation step of 94ºC for 7 min, followed by 35 cycles of 94ºC for 20 sec, 56ºC for 30 sec, 72ºC for 40 sec with a final extension step of 72ºC for 7 min. PCR products were checked to confirm correctly sized products on 0.5% agarose gels, stained with 20 000x Red Safe™ Nucleic Acid Staining Solution (iNtRON Biotechnology, Inc). The fragments were purified using SureClean (Bioline) following manufacturer's protocol and sequencing reactions were performed with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biossystems, ABI) with an initial denaturation step at 96ºC for 1 min, followed by 25 cycles of 96ºC for 10s, annealing step of 50ºC for 5s and a final extension step of 4 min at 60ºC. Both fragments were sequenced in both directions using the respective amplification primers on an ABI PRISM 310 Genetic Analyser (Applied Biosystems) and base composition of the DNA sequences were obtained with Sequencing Analysis Software 5.2 (Applied Biosystems). Chromatograms were manually checked for errors in Sequencher v.4.0.5 (Gene Codes Co.) and nucleotide ambiguities, considered as heterozigous sites, were classified accordingly to IUPAC ambiguity code. The obtained sequences were aligned with Clustal X v.2.1 (Larkin *et al.* 2007).

### *3.3 Phylogenetic analysis*

To infer the phylogenetic relationships of *L. bleusei* within the genus *Lycaena* and to compare the genetic information given by each molecular marker we performed maximum parsimony (MP) and maximum likelihood (ML) analyses using PAUP* v4.0b10 (Swofford 2001). Also a Bayesian inference (BI) method was applied using MrBayes (Huelsenbeck & Ranquist 2001) for each locus. Additionaly, a combined dataset for both genes was carried out with Concatenator v.1.1.0 (Pina-Martins & Paulo 2008) and was also analyzed with MP, ML and BI. The most fit evolution model for each gene was selected with Modeltest v.3.7 (Posada & Crandall 1998) for PAUP* analyses and MrModeltest v.2.3 (Nylander 2004) for MrBayes analyses, under Akaike Information Criterion (AIC). For the MP analyses a heuristic search was performed using 1000 random addition sequence

replicates and clade support was obtained by performing 1000 replicates of nonparametric bootstrap. For the ML analyses another 1000 random addition sequence replicates heuristic search was performed also assessing clade support by performing 1000 replicates of nonparametric bootstrap. The Bayesian inference analyses were performed using two runs of four Monte Carlo Markov Chain (MCMC) were iterated for 2 million generations each, with a sampling frequency of 150 generations, with a burnin of 100. Bayesian posterior probabilities (bpp) were calculated using the 'sumt' command of MrBayes.

### *3.4 Phylogeography and population genetic structure*

For the individual analyses of each species with the nuclear locus, haplotypes estimations of multiple heterozigotes were performed with PHASE v.2.1.1 (Stephens & Scheet 2005).

Maximum parsimony median-joining haplotypic networks were then constructed using Network v.4.6 (Bandelt *et al.* 1999) for both species and population genetic differentiation was assessed with ARLEQUIN v.3.5 (Excoffier *et al.* 2005) for each species and each locus separately. AMOVA analyses, nucleotide diversity ($\pi$), haplotypic diversity (h) and $F_{st}$ pairwise distances were also assessed with ARLEQUIN. The groups were created based on the sampling sites. We also calculated both Fu's Fs and Tajima's D neutrality tests in order to evaluate demographic expansion or presence of selection and the results were then confirmed in DNAsp v.5 (Librado & Rozas 2009).

### *3.5 Divergence dating and demographic analyses*

For divergence dating we performed a Bayesian MCMC analyses in *BEAST (Heled and Drummond 2010), an extension of the BEAST v.1.6.2 software package (Drummond & Rambaut 2007), using a concatenated matrix of both molecular markers. The required species assignment was performed based on morphological characters and later confirmed with molecular data. Migration rate between the two species was assessed with software MDIV (Nielson & Wakeley 2001). The fittest evolution models used in these analyses were those suggested by Modeltest, as mentioned above. Based on preliminary

runs using uncorrelated lognormal molecular clocks, the "CoefficientOfVariation" parameter for the EF-1α partition revealed a probability distribution abutting 0, suggesting that it does not significantly deviate from a strict clock assumption. Therefore, an uncorrelated lognormal molecular clock was used for the COI partition and a strict molecular clock was applied to the EF-1α partition in our final model build. To convert genetic distance into geological time units, we used a 2.3% per million year substitution rate, calibrated by Brower (1994). This substitution rate was fixed for the COI partition and estimated for the EF-1α partition using an uniform prior [0, 1]. The final results were obtained by combining three independent runs with $1x10^6$ generations, sampled every $1000^{th}$ iteration, using LOGCOMBINER v1.6.2 (Drummond & Rambaut 2007). Convergence and mixing for all parameters were assessed in TRACER v1.5 (Rambaut & Drummond 2007) by visually inspecting the trace log and estimating the Effective Sample Size (ESS); ESS for most parameters were > 200 indicating an adequate mixing.

The demographic histories of *L. bleusei* and *L. tityrus* were reconstructed separately using a Bayesian Skyline Plot (BSP) analyses, implemented in BEAST v1.6.2 (Drummond & Rambaut 2007). Due to the very low informative content of the COI partition at the species level, it was not possible to include this partition in the analyses. Thus, both BSP analyses were performed only with the EF-1α partition using an uncorrelated lognormal molecular clock. Given the absence of information regarding the substitution rate of Ef-1α for lycenids, we have fixed the clock rate with "1" in order to produce estimates in substitutions/site units. Final demographic reconstructions for each species were performed with three independent runs of $1x10^6$ generations, sampled at every $1000^{th}$ generation, and combined using LOGCOMBINER v1.6.2. The performance of the MCMC runs were assessed using Tracer v1.5 (Rambaut & Drummond 2007), which was also used to create the BSP with the mean and corresponding credibility intervals of the estimated demographic parameters.

# 4. Results

## *4.1 Genetic Diversity*

For all 68 ingroup individuals (44 *L. bleusei* and 24 *L. tityrus*) both mitochondrial gene and nuclear gene fragments were correctly amplified. From the 598 bp of the COI fragment, 575 sites were constant while 23 were polymorphic, of which 5 were singletons and 18 were parsimoniously informative. A total of 12 haplotypes were encountered (including both species). For the nuclear EF-1α gene a fragment of 585 bp was amplified of which 558 sites were monomorphic and 27 were variable: 9 singletons and 18 parsimoniously informative sites. For the nuclear gene 32 haplotypes were encountered, however when the allelic phase is considered, only 30 haplotypes arise.

The reading frame for both fragments was determined by alignment with *Drosophila yakuba*, for which both COI and EF-1a sequences are available and no stop codons were found. Synonymous and non-synonymous mutations were also assessed and there was no evidence for selection.

Table 1 resumes haplotypic richness (h) and nucleotide diversity (π), as well as Tajima's D and Fu's Fs values for each sampling site for both species and both loci. Generally, Portuguese populations of the Iberian Sooty Copper show high haplotypic richness and very low nucleotide diversity, both with COI and EF-1α analyses. Considering mtDNA, the highest haplotypic diversity for *L. tityrus* is found in populations from Greece and for *L. bleusei* is found in "Gata" (h = 1,000). However, in this sampling area only two individuals are taken into consideration, each with a different haplotype, so this may not reflect the true genetic variation present in that area. The lowest value is found in "Gredos", where only two haplotypes are present in a relatively large sample. For *L. tityrus* in Serra da Estrela West and *L. bleusei* in "Guadalupe" there is only one haplotype, so both h and π values are zero. None of the neutrality tests revealed to be statistically significant and there is no signal from the mtDNA for demographic expansion or signature of selection on either species. For the two sampling sites which are represented by a single haplotype these tests are not applicable.

**Table 1**. Haplotypic diversity (h), nucleotide diversity ($\pi$) and neutrality tests for each sampling site of *Lycaena tityrus* and *Lycaena bleusei*.

| | COI | | | | | | EF-1α | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Locality** | **n** | **no. of haplotypes** | **h** | **π** | **Tajima's D** | **Fu's Fs** | **n (alleles)** | **no. of haplotypes** | **h** | **π** | **Tajima's D** | **Fu's Fs** |
| *Lycaena tityrus* | | | | | | | | | | | | |
| Serra da Estrela North | 10 | 2 | 0.533 + 0.095 | 0.00098 + 0.00016 | 1.30268 | 1.029 | 20 | 3 | 0.426 + 0.122 | 0.00078 + 0.00025 | -0.44022 | -0.377 |
| Serra da Estrela West | 6 | 1 | 0 | 0 | - | - | 12 | 6 | 0.813 + 0.074 | 0.00205 + 0.00033 | -0.15715 | -2.408 |
| Greece | 5 | 3 | 0.700 + 0.218 | 0.00134 + 0.00050 | -0.9725 | -0.829 | 10 | 4 | 0.778 + 0.091 | 0.00190 + 0.00038 | 0.17186 | -0.657 |
| *Lycaena bleusei* | | | | | | | | | | | | |
| Serra da Estrela North | 6 | 4 | 0.867 + 0.129 | 0.00368 + 0.00138 | -0.93169 | -0.325 | 12 | 8 | 0.924 + 0.057 | 0.00430 + 0.00097 | 0.20194 | -3.164** |
| Serra da Estrela West | 12 | 4 | 0.742 + 0.084 | 0.00160 + 0.00025 | 1.28955 | -0.719 | 24 | 6 | 0.841 + 0.037 | 0.00242 + 0.00026 | 0.16038 | -0.951** |
| Serra da Estrela East | 5 | 2 | 0.600 + 0.175 | 0.00100 + 0.00029 | 1.22474 | 0.626 | 10 | 5 | 0.844 + 0.080 | 0.00239 + 0.00039 | -0.03789 | -1.430** |
| "Gata" | 2 | 2 | 1.000 + 0.500 | 0.00167 + 0.00084 | - | -0.000 | 4 | 3 | 0.833 + 0.222 | 0.00570 + 0.00152 | 0.17969 | 0.888 |
| "Gredos" | 9 | 2 | 0.389 + 0.164 | 0.00065 + 0.00027 | 0.15647 | 0.477 | 18 | 5 | 0.693 + 0.086 | 0.00228 + 0.00064 | -0.77589 | -0.530 |
| "Francia" | 7 | 2 | 0.476 + 0.171 | 0.00080 + 0.00029 | 0.55902 | 0.589 | 14 | 8 | 0.824 + 0.098 | 0.00242 + 0.00050 | -0.89259 | -4.732** |
| "Guadalupe" | 4 | 1 | 0 | 0 | - | - | 8 | 4 | 0.750 + 0.139 | 0.00220 + 0.00068 | -0.72673 | -0.729 |

** $P < 0.05$

Regarding the nuclear gene, the number of alleles (2n) was considered. In this case, the highest haplotypic diversity is found in Serra da Estrela North for *Lycaena bleusei*, even though all values are similarly high, and in Serra da Estrela West for *Lycaena tityrus*. For the nuclear gene analyses of both species, $\pi$ is once again very low, demonstrating that, although there is a higher number of haplotypes these are not very divergent. Neutrality tests, in this case, show a different result since *L. bleusei* samples from Serra da Estrela (North, West and East) and also from "Francia" show significant Fu's Fs values, suggesting demographic changes of *L. bleusei* but not *L. tityrus*.

## *4.2 Phylogeny of the genus Lycaena*

Phylogenetic analyses of the genus *Lycaena* based on Maximum Parsimony, Maximum Likelihood and Bayesian Inference for both molecular markers generally presented the same basic phylogenetic pattern. Here, we chose to present the Bayesian Inference analyses, for each loci and for the concatenated data (Figure 4, Figure 5 and Figure 6). Within *Lycaena* two major groups were unveiled: a Palearctic group and a Nearctic group. The first includes all Lycaena species with a Palearctic distribution, including all European and Asian species. *Lycaena li* (China) stands out, as it did not group monophyletically with the remaining Asian (*L. solski* and *L. alpheraki*) species, appearing in a basal position. The Palearctic group includes also two American species, *L. helloides* and *L. cupreus* and this result is constant throughout the analyses and well supported. The Nearctic group encompasses all American species, except those mentioned above, with two internal clades. Although, crown groups are usually well supported by bootstrap statistics and Bayesian posterior probabilities (bpp), most basal relationships are poorly resolved and have low support. The best resolved tree was MP analysis of EF-1α which presents a topology with low polytomies (Figure 11, Appendix).

*L. bleusei* and *L. tityrus* grouped in a monophyletic clade which is closely related with the species in their morphologic group. A clear separation between *L. bleusei* and *L. tityrus* is shown, suggesting sufficient differentiation to classify them as separate sister taxa and not subspecies (see also bellow). This monophyletic group is constant throughout all

phylogenetic analyses and generally supported by high values of bootstrap and Bayesian posterior probabilites.

Overall, species in the Palearctic group were arranged accordingly to their morphologic group (Table 6, Appendices), suggesting concordance between morphology and genetic differentiation. However, not all the species within each morphologic groups were included in the phylogenetic analyses and, thus, this result must not be generalized.

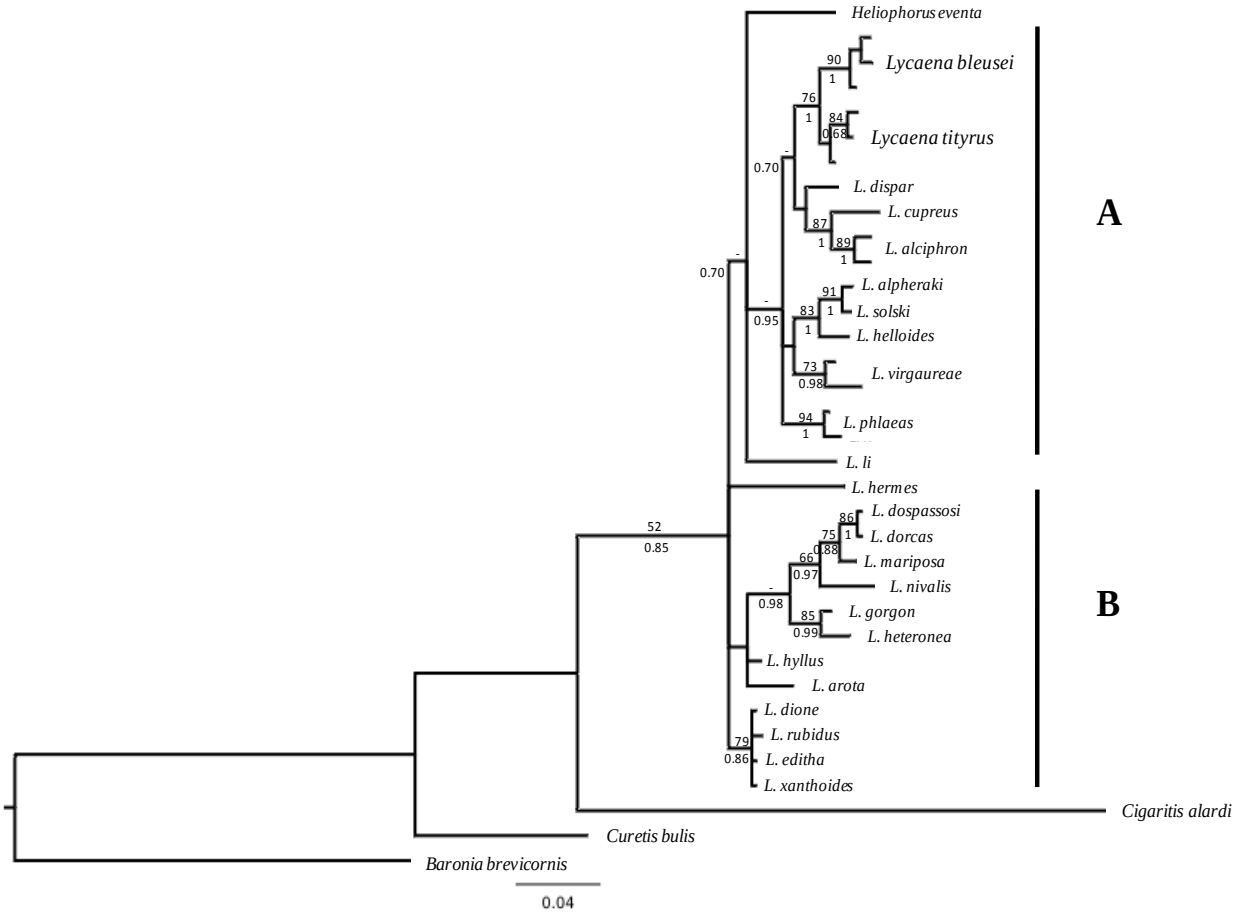


**Figure 4** Bayesian Inference consensus tree based on the EF-1α dataset. Values above branches correspond to ML bootstrap values (only values > 50% are shown) and values below branches correspond to bayesian posterior probability . **A** – Palearctic group and **B** – Nearctic group.

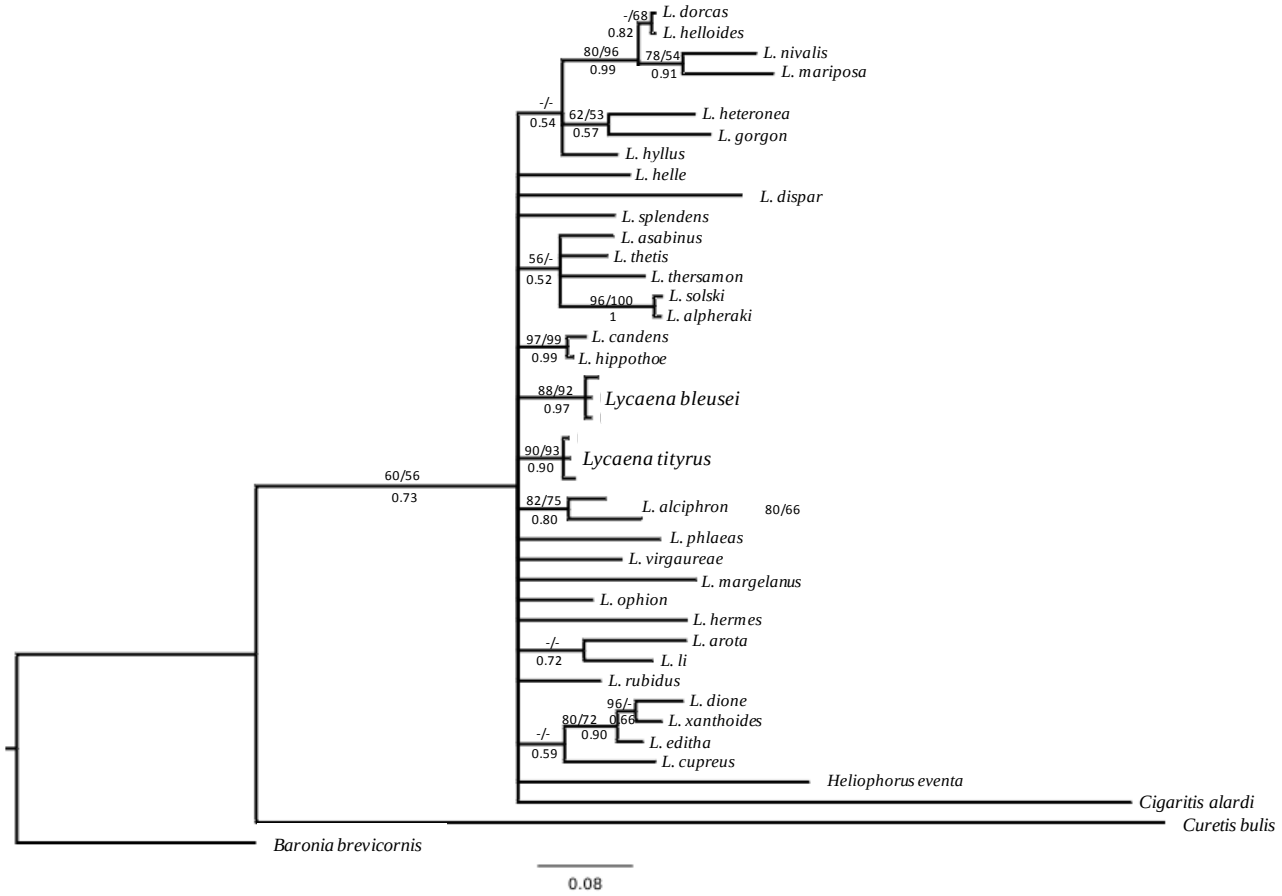


**Figure 5** Bayesian Inference consensus tree based on mtDNA data. Values above branches correspond to ML/MP bootstrap values (only values > 50% are shown) and values below branches correspond to bayesian posterior probability.

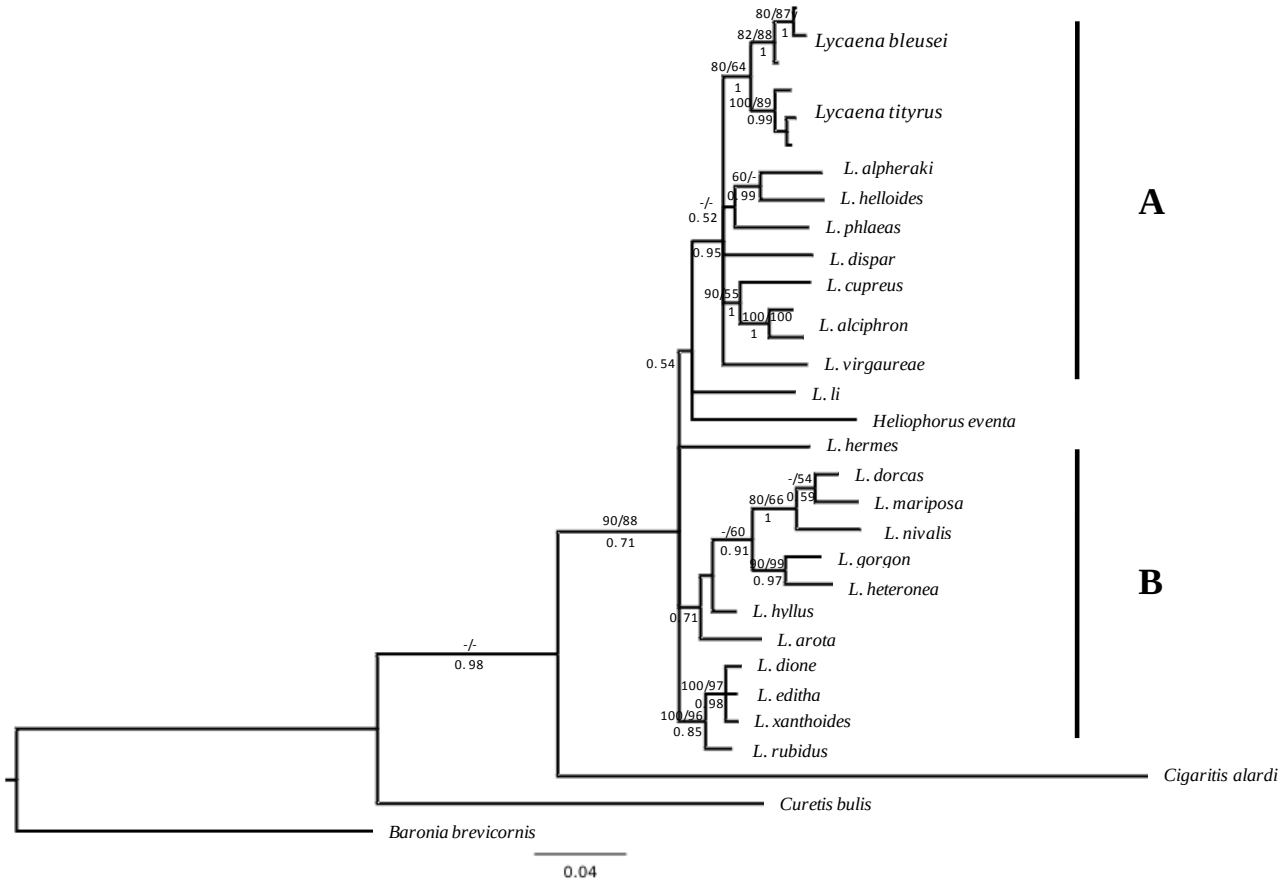


**Figure 6** Bayesian Inference consensus tree based on combined dataset of mtDNA and nuclear loci. Values above branches correspond to ML/MP bootstrap values (only values > 50% are shown) and values below branches correspond to bayesian posterior probability. **A** – Palearctic group and **B** – Nearctic group.

## *4.3 Phylogeography and population genetic divergence*

The nuclear gene EF-1α yielded a complex haplotypic network, reflecting the higher number of haplotypes (30), while mtDNA produced a simpler network corresponding to the 12 haplotypes previously mentioned (Figure 7A and 7B respectively).

Overall, both haplotypic networks show the same basic pattern. The haplotypes grouped in two different clades designated here as *L. bleusei* and *L. tityrus*, being the distance between these groups large enough to clearly separate the two species, with no intermediate haplotypes or suggested median vectors. MtDNA grouped all individuals as expected, accordingly to the morphological differences between the two taxa. However, in the haplotypic network given by EF-1α one haplotype, which grouped with *L. bleusei*, corresponds in fact to a pair of individuals classified initially as *L. tityrus* (red arrows in

Figure 7) sampled in northern Portugal. Genetic distances between the two species corroborate this result, showing maximum values of 3.3% and 2.3% and minimum of 2.6% and 1% for the mitochondrial and nuclear loci respectively.

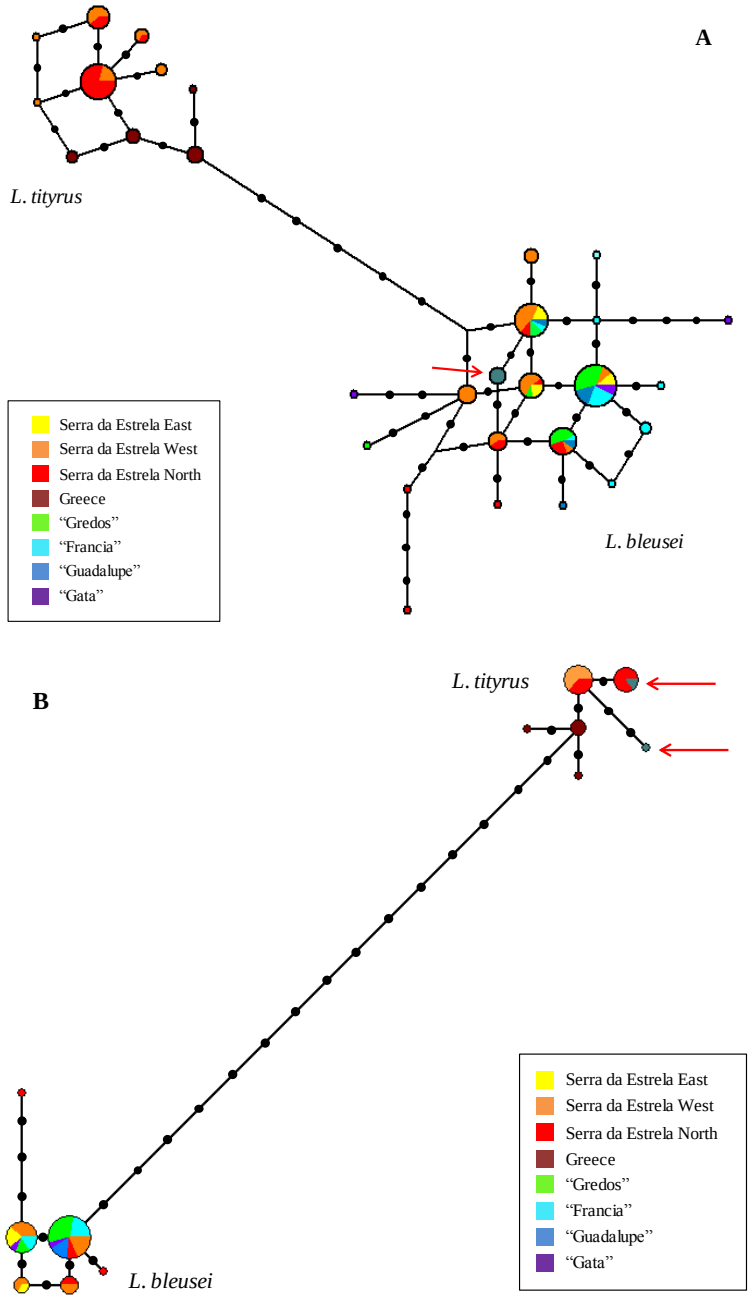


**Figure 7** Median Joining network based on A) EF-1α haplotypic data and B) COI data. Circle size is proportional to the frequency of each haplotype. Colors correspond to sampling sites described in the legend.

There is no apparent geographical pattern on the distribution of the haplotypes. Haplotypes from each group are distributed across all sampling sites, except four (EF-1α) and three (COI) haplotypes that are exclusively found in Greek populations. Some expected haplotypes determined with Phase v.2.1.1 also appear exclusively in some populations, however this result has to be taken with caution since there is always some degree of uncertainty on the construction of the most probable haplotypes.

The Analysis of Molecular Variance (AMOVA) suggests greater variation within populations than among populations, except for *Lycaena tityrus* analyses with the mitochondrial gene (Table 2), suggesting that haplotypes are shared among most populations, as seen in the haplotypic networks. Each sampling area is heterogeneous in its genetic composition, usually with occurrence of more than one haplotype within each population, however, in geographic space, they behave as a homogeneous group.

**Table 2.** Analysis of molecular variance (AMOVA) for both species and both loci.

| | | COI | | | EF-1α | | |
|---|---|---|---|---|---|---|---|
| | **d.f** | **Sum of squares** | **Variance of components** | **Percentage of variance** | **Sum of squares** | **Variance of components** | **Percentage of variance** |
| *Lycaena tityrus* | | | | | | | |
| Among populations | 2 | 3.382 | 0.21161 | 51.41 | 2,025 | 0.09022 | 19.32 |
| Within populations | 19 | 3.800 | 0.20000 | 48.59 | 7,157 | 0.37669 | 80.68 |
| **Total** | 21 | 7.182 | 0.41161 | | 9,182 | 0.46691 | |
| *Lycaena bleusei* | | | | | | | |
| Among populations | 6 | 3.510 | 0.04807 | 14.31 | 2,756 | -0.00195 | -0.042 |
| Within populations | 38 | 10.934 | 0.28774 | 85.69 | 17,911 | 0.47135 | 100.42 |
| **Total** | 44 | 14.444 | 0.33581 | | 20,667 | 0.46939 | |

** $P < 0.05$

An exception to this pattern is found in mtDNA of *L. tityrus* populations in which there are similar levels of variation within and among populations. In this case there is a lower number of haplotypes in each population but these are usually not shared between the three main sampling areas.

This result is further supported by $F_{st}$ values between each population (Table 3). *L. tityrus* shows significant divergence between Portuguese and Greek populations (COI) and

between Serra da Estrela West and Greece (EF-1α). These values remained statistically significant after Bonferroni correction for multiple testing.

**Table 3.** Pairwise $F_{st}$ values between populations of *Lycaena tityrus*. Numbers below diagonal correspond to COI and above correspond to EF-1α.

| *L. tityrus* | Serra da Estrela North | Serra da Estrela West | Greece |
|---|---|---|---|
| Serra da Estrela North | - | 0.09444 | 0.33333** |
| Serra da Estrela West | 0.50089** | - | 0.16306** |
| Greece | 0.40157** | 0.70588** | - |

** P < 0.05

For *L. bleusei*, $F_{st}$ values were overall lower and some negative values were obtained (Table 4). Negative values should be interpreted has zero. Significant values suggested differentiation between populations from Serra da Estrela North *vs.* East; Serra da Estrela East and "Gredos" (COI) and also Serra da Estrela West and "Gredos" (EF-1α). These values, however, revealed to be not statistically significant after Bonferroni correction for multiple testing.

**Table 4.** Pairwise $F_{st}$ values between populations of *Lycaena bleusei*. Numbers below diagonal correspond to COI and above correspond to EF-1α.

| *L. bleusei* | Serra da Estrela North | Serra da Estrela East | Serra da Estrela West | "Gredos" | "Francia" | "Gata" | "Guadalupe" |
|---|---|---|---|---|---|---|---|
| Serra da Estrela North | - | -0.0578 | -0.02412 | -0.00052 | 0.00000 | 0.00000 | -0.04348 |
| Serra da Estrela East | 0.25917 ** | - | -0.04361 | 0.01154 | 0.01869 | -0.03604 | -0.05263 |
| Serra da Estrela West | 0.04204 | 0.04938 | - | 0.08118** | 0.02070 | 0.00244 | 0.01538 |
| "Gredos" | 0.18182 | 0.45433** | 0.11525 | - | 0.00895 | -0.01348 | -0.14592 |
| "Francia" | 0.12753 | 0.35895 | 0.03735 | -0.13318 | - | -0.07692 | -0.07692 |
| "Gata" | 0.26241 | 0.66574 | 0.28205 | -0.00599 | 0.05724 | - | -0.14286 |
| "Guadalupe" | -0.09081 | -0.04294 | -0.32663 | -0.15826 | -0.31250 | 0.3842 | - |

** P < 0.05

## *4.4 Divergence time estimates and demographic analyses*

The MDIV analysis of the concatenated dataset revealed a very low migration rate between *L. tityrus* and *L. bleusei* (Figure 8). The posterior distribution for this parameter was close to 0, as determined by visual inspection, suggesting that the null hypothesis of no migration could not be rejected. Therefore, the assumption of complete isolation of the *BEAST analyses was met (Heled & Drummond 2010).

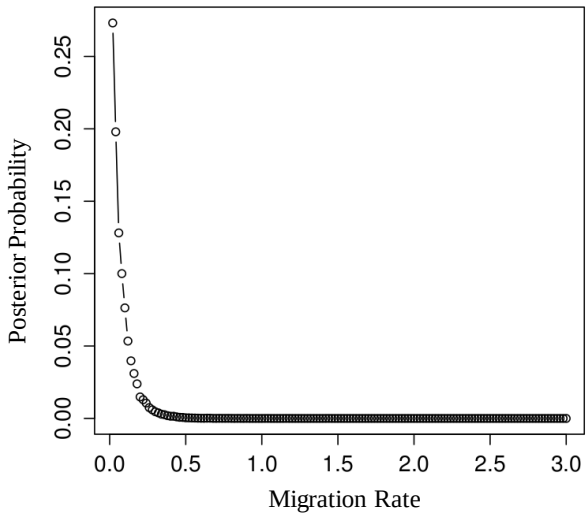


**Figure 8** The posterior probability distribution of the migration rate between *Lycaena tityrus* and *Lycaena bleusei*.

To understand the timeline of speciation events, a time-calibrated species tree was generated from the *BEAST analysis based on a concatenated dataset (Figure 9). The topology of the species tree was well supported and coincident with the previous MP, ML and BI phylogenetic analyses. This analysis allowed the co-estimation of the TMRCA (Time to most recent common ancestor), including 95% HDP (High Posterior Density) credibility intervals of both *L. tityrus* and *L. bleusei* populations separately, as well as their divergence time. The TMRCA of *L. bleusei* was estimated to be 120 400 years (95% HDP, 31 900 to 229 900) and the TMRCA of *L. tityrus* was estimated to be between 20 874 and 179 800 (95% HDP) with a mean of 92 818 years. Both species seem to have diverged from each other around 999 300 years (95% HDP, 487 800 to 1 584 300).

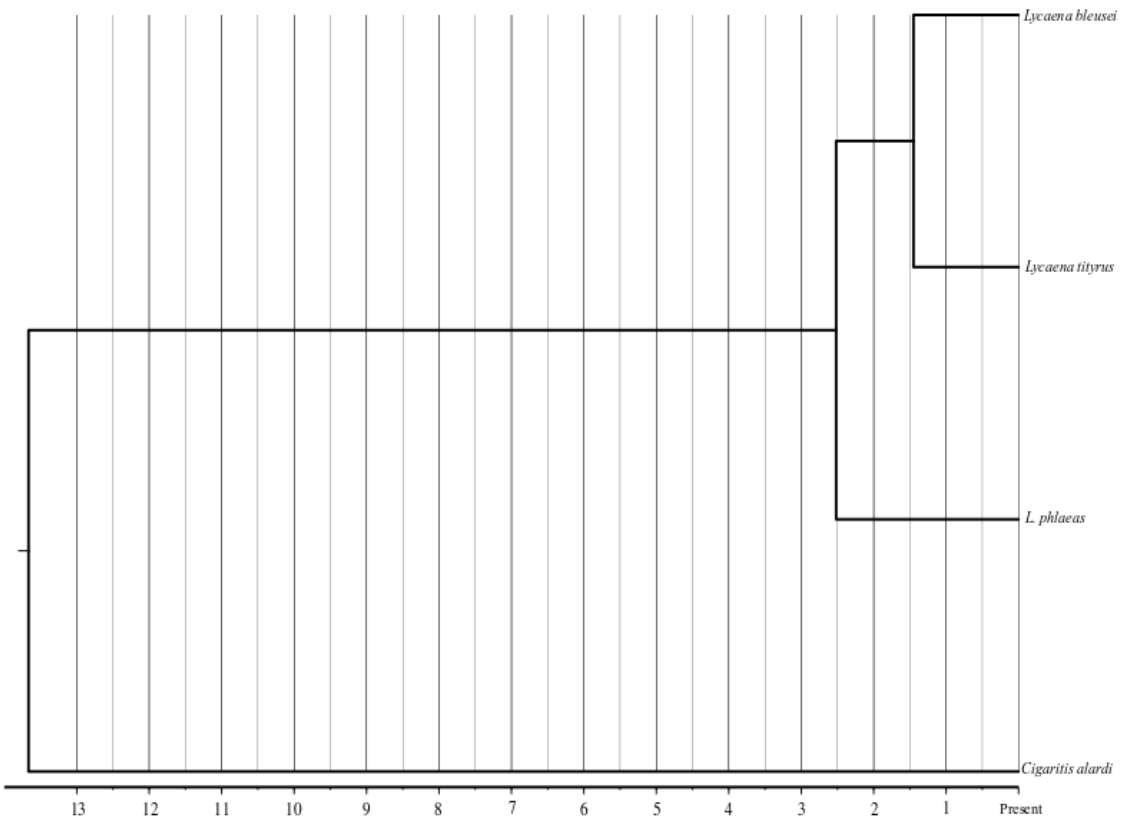


**Figure 9** Time calibrated species tree based on the *BEAST analysis of the concatenated dataset. Scale numbers correspond to Million years before present

Demographic reconstructions of both *L. tityrus* and *L. bleusei* populations revealed distinct demographic histories for each species (Figure 10). While there is no significant evidence for a demographic expansion in *L. tityrus* according to the BSP (Figure 10A) and the Fu's *Fs* test described above, there is a clear increase in the effective population size of *L. bleusei* in a fairly recent time. The BSP (Figure 10B) suggests that this species began to expand at roughly one fourth of the time since its divergence from *L. tityrus*, which corresponds to ~30 000 years before present by extrapolating from the TMRCA of *L. bleusei*. However, given that different molecular data were used between the BEAST Skyline and *BEAST analyses, hence different models were employed, this extrapolation should be cautiously interpreted as a rough approximation.

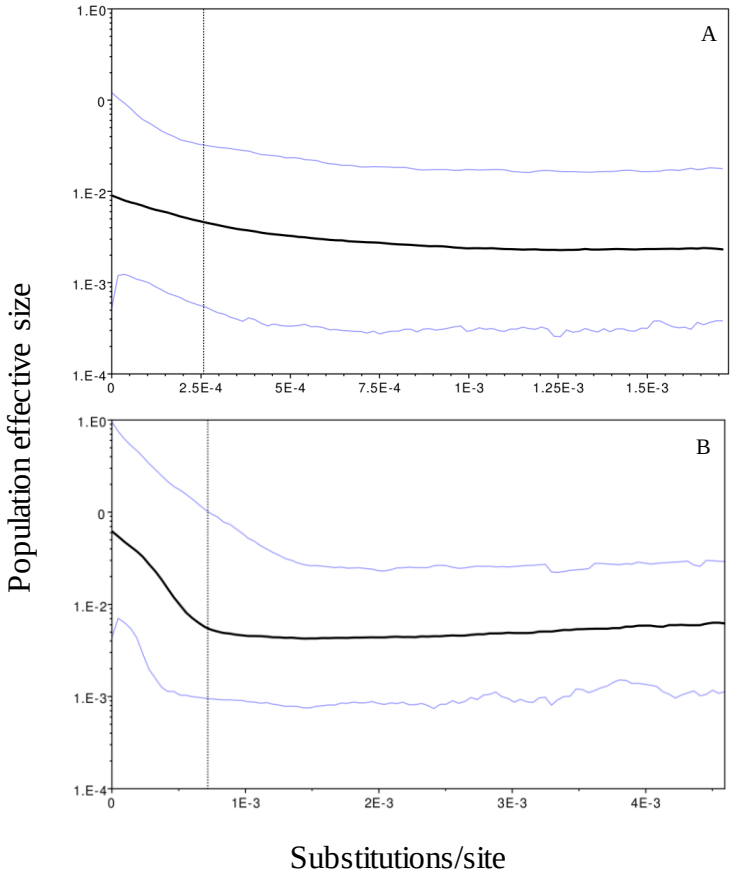


**Figure 10** Bayesian Skyline Plot of the fluctuations in the population size of A) *Lycaena tityrus* and B) *Lycaena bleusei*. Black lines represent the median estimate and the gray lines correspond to the upper and lower 95% High Posterior Density (HPD).

### *4.5 Sympatric distribution area and hybridization*

Because both taxa were expected to occur very close together in central Portugal, an area of about ~270 km$^2$ was prospected in order to evaluate suitable habitats for both iberian Sooty Coppers, covering Serra da Estrela and northern Portugal. Several areas typically attributed to *L. tityrus* were sampled and many revealed the presence of sympatric populations of *Lycaena tityrus* and *Lycaena bleusei*, unveiling the finding of a completely unknown (though expected) geographic area where both taxa are in strict syntopy and sympatry.

Ecological data collected in the field, such as habitat type and size, weather conditions, food source and butterfly community do not suggest any significant differences of habitat usage and in sympatric populations specimens were found within the same

slightly overgrown, flower-rich meadows and river borders, overlapping completely their territories.

Overall, genetic analyses did not show any incongruence between genetic lineages of the individuals in these areas. The majority of analyzed specimens were genetically classified in agreement with the initial morphologic identification done during field collections and only two individuals did not follow this pattern. Morphologically scored as *Lycaena tityrus*, the two males from Boticas, northern Portugal show a dark forewing upperside with small black spots and absence of hindwing tail. MtDNA of these specimens matched *L. tityrus* samples from elsewhere, in agreement with morphology, one of the specimens bearing a typically northern haplotype and the other presenting three substitutions from this haplotype, but still grouping clearly within *L. tityrus* group (Figure 7B). Unexpectedly, the nuclear sequences of both individuals differed only in a singleton from the more common *L. bleusei* sequences, and these specimens were placed in the later group (Figure 7A).

# 5. Discussion

This work has two clear implications: a) the two taxa previously defined based on morphologic characters are indeed two isolated gene pools with divergent molecular and ecological histories and b) isolation between the two species occurred earlier than first expected.

Phylogenetic analyses of the genus *Lycaena* yielded generally the same basic pattern described by other authors (van Dorp 2004 and Jong & van Dorp 2006), with two groups, one Palearctic and one Nearctic. Tree resolution and branch support was less than desired and the presence of politomies was constant throughout the analyses, even with the concatenation of both loci in the same matrix (Figure 4, Figure 5 and Figure 6). When dealing with closely related species and recently divergent as Lycaenidae, these problems are bound to happen, as sequences probably had insufficient time to diverge, thus providing insufficient resolution for deeper phylogenetic relationships. This comes in agreement, however, with the rising evidence of a rapid speciation ability within the Lepidoptera and especially in the Lycaenidae (Nice 1999, 2005).

As mentioned above, the two major phylogenetic groups proposed by van Dorp (2004) for the genus *Lycaena* are also found in these analyses, although excluding African and New Zealander species. The inclusion of North American species within the Palearctic clade is only recognized by that author for *L. cupreus* that always groups monophyletically with *L. alciphron*. The positioning of *L. helloides* in the Palearctic group contradicts Dorp work, where it grouped with the North American species. This arrangement however, was consistent throughout the analysis and was supported by high bootstrap and Bayesian posterior probability values. According to this topology, and taking into consideration that these American species have a derived position, we can infer that the two species have colonized North America in two independent events and are descendant of the Palearctic species. Bonzano & Weindenhofer (2001) proposed several morphological groups according to phenotype and male genitalia for the Palearctic copper species, which were mostly recovered in this work. Thus, morphological analysis broadly reflects true phylogenetic relationships recovered here, although we could only include a few representative species of each group.The two sister taxa *Lycaena bleusei* and *Lycaena*

*tityrus*, although fairly similar ecologically and morphologically, are characterized by distinct genetic clusters, as show by the haplotypic networks (Figure 7) and from evidence that there is no haplotype sharing between the two species, with the exception of a singular hybridization event (red arrows in Figure 7). This evidence is further supported by high genetic distances between species (3.3% and 2.6% for COI and EF-1α respectively) and by overall low distance values within species that range 0-0.2% with either loci. Between the two taxa, partitioning of genetic variance is due to differences among individuals within populations rather than differences between geographically disperse populations, meaning that individuals from distant populations are at least as closely related as individuals from the same populations. This usually occurs in a scenario where there are high levels of gene flow between populations of the same species. In this case, although we are working with flying insects, lycenid males are usually territorial and tend to remain in their established territories *i.e.* they are perchers (see for example Scott 1974; Douwes 1975; Cordero & Soberón 1990; Rutowski 1991 and Takeuchi & Imafuku 2005) and, thus, gene flow shall occur through female migration. The only exception in this case is for *L. tityrus* with COI, because the genetic divergence between Portuguese and Greek populations is, as expected, high.

Isolation between the two taxa, as shown by Bayesian Inference (Figure 9), seems to indicate that *Lycaena bleusei* and *Lycaena tityrus* diverged from each other ~1 Myr before present, corresponding to the Waalian interglacial period. This estimation has however to be taken carefully as there is a large time interval to account for. The Waalian interglacial is one of the longest warm periods of the mid Pleistocene and usually marks the beginning of the more accentuated climatic fluctuations that were to take place in the late Pleistocene (Anderson & Borns Jr. 1994).

Individually, *Lycaena bleusei* presented high values of haplotype richness with low nucleotide diversity, which is consistent with recent demographic changes, supported by significant Fu's Fs values. When there is an increase in effective population size, new mutations are kin to arise promoting new haplotypes, which did not have enough time to diversify. In the case of this species, which seems to be broadly panmitic throughout its range but with several isolates, populations are generally small. These populations are now

expanding, probably with high levels of migration between isolates which did not allow to reach a mutation-drift equilibrium. Bayesian Skyline Plot confirmed these evidences and suggests a fairly recent expansion in this species, beginning at around 30 000 years before present.

As to *Lycaena tityrus*, there seems to be no evidence of such demographic expansion or contraction and the demographic analysis suggests that this species was demographically constant since its divergence. However, for this species, there is limited data available and the majority of its distribution was not sampled. The inclusion of more individuals from populations in northernmost Iberia and central Europe should help to clarify the demographic history of the Sooty Copper and how it may have responded to the cyclic climatic changes that occurred in the Pleistocene.

Overall, COI shows a lower variability than initially expected, since numerous studies use this locus to determine phylogeographic patterns and demographic histories (Kim *et al.* 2010, Dincă *et al.* 2010). However, genetic distances among and between species are within the values expected for Lycaenidae (0-3%) and mtDNA was able to clearly differentiate the two genetic clusters. Since, in Lepidoptera, females are the heterogametic sex (ZW) there might be selective pressures on the genes present in the W chromosome (Kunte *et al.* 2011). If females are selected based on this chromosome and, taking into consideration that mitochondria are maternally inherited, we can considerer indirect selection of mitochondrial DNA lineages.

The nuclear loci, EF-1α, albeit showing no evidence for selection, showed high levels of diversity and was proven better for the analyses. This is even more surprising when we consider that the fragment amplified in this work corresponds to an exon of a housekeeping gene. Codifying genes are usually more conserved as mutations can be deleterious and tend to be eroded from the population through purifying selection. Although there are no evidences for selective pressures in these species, several studies have been able to identify, through reconstruction of ancestral sequences, such selective pressures on basal branches of the phylogenies (Chang & Donoghue 2000, Yang & Bielawski 2000, Anisimova & Kosiol 2009). Thus, further investigation is needed to

evaluate if such phenomenon may have occurred in this case, investigating the ratio dN/dS (non-synonymous *vs*. synonymous mutations) throughout the whole *Lycaena* phylogeny.

## *5.1 Barriers to gene flow*

Genetic (and morphologic) variance partitioning supports the hypothesis of at least two isolated gene pools in this group and *Lycaena bleusei* and *Lycaena tityrus* have predominantly allopatric distributions and different phenotypes. Given the genetic divergence, geographical ranges and calibrated times for divergence we can provide preliminary inferences on probable barriers to gene flow. In the Waalian interglacial period (0.9 – 3 Myrs bp), the Iberian Peninsula should have the same basic physiographic as nowadays, thus it seems that there were not different physical barriers keeping the species apart. Looking at the present conditions and the ecology of the Sooty Copper species complex favoring flower-rich meadows, close to rivers and water beds we should predict that the differentiation between the two taxa occurred about 1Mya within the central Iberian Mountain System, avoiding the dry and hot patches of central Iberia. The same barriers that kept both populations apart should be the same that are broadly promoting the general allopatry nowadays. During the long Waalian interglacial, with maximum temperatures similar to present ones, isolation of a Sooty Copper population within the Iberian Central Mountain System subject to slightly different selective pressures and drift might have led to its differentiation from the individuals that were seeking suitable habitats further north.

## *5.2 Secondary contact and hybridization*

During field work, a sympatric zone in Portugal from the central mountain system (western Serra da Estrela) to the northern margins of the Douro River (northern Portugal) was found. Throughout this relatively small geographic area there are mixed populations of *L. bleusei* and *L. tityrus*. This represents a notably extension of the known distribution area of *Lycaena bleusei* which is now known outside the central mountainous area, even in more Atlantic influenced ranges, such as Serra de Montemuro. Associated with the genetic evidence of an expansion, it seems clear that Iberian Sooty Copper adults are colonizing

habitats that were initially inhabited by *Lycaena tityrus* alone. This expansion follows an east-west direction, as only *L. bleusei* is found in eastern parts of the distribution. From empirical observations there seems to be no differences in ecological or behavioral preferences of both species. In this context we should expect, as described for many butterfly species, signals for hybridization and introgression (Chianchi *et al.* 2003, Whinnett *et al.* 2005, Kronforst *et al.* 2006, Dasmahapatra *et al.* 2010). From the pool of individuals collected in this sympatric area, only two revealed a genetic evidence of hybridization, showing morphological and mtDNA sequence of *L. tityrus* and nuclear sequence of *L. bleusei*. This proves that hybridization can occur between these two species, even if it is a rare event. Migration estimates suggested that between *L. bleusei* and *L. tityrus* there was no gene flow, which comes to reinforce that these are two independently evolving and formally separate biological entities. Studies on other butterfly species admit rare hybridization events and Haldane's rule seems to have a strong effect in Lepidoptera, on the inviability and sterility of heterogametic hybrids, which should also be investigated in the future (Sperling 1993, Davies *et al.* 1997, Schilthuizen *et al.* 2011 but see Wahlberg *et al.* 2009).

# 6. Final Remarks

Phylogenetic and phylogeographic analyses performed throughout this work were able to meet the main goals proposed in the beginning of this thesis. Overall, the methods employed for the analyses of the genetic information given by two independent loci, with two different evolutionary histories, provided adequate data from which explicatory and interesting results were drawn for the evolutionary history of the Iberian Sooty Copper.

Phylogenetic trees, although with different topology, presented the same basic pattern on the position of the species within the genus *Lycaena*. The Palearctic group includes both *Lycaena tityrus* and *Lycaena bleusei*, as should be expected from previous considerations on these species. This analyses provided the first insight on the differentiation between the sister species and pointed always to a clear phylogenetic distinction. Further, the interpretation of genetic diversity and how the haplotypes were distributed through the populations sampled, showed that there is no haplotype sharing between the two Sooty Coppers and that each species has its own evolutionary history, genetic variation and differentiation between populations. Molecular analyses showed also introgression events between *Lycaena bleusei* and *Lycaena tityrus*, in the sympatric area described for the first time in this work, which were until now completely unknown.

These results allow drawing some important conclusions for the scientific knowledge of this Iberia endemic species, *Lycaena bleusei* while providing important information for the widely distributed Sooty Copper, *Lycaena tityrus*, as well.

Two different Sooty Copper taxa, which have for a long time been disputably separated or lumped into the same species, *Lycaena tityrus*, have in the last years been finally considered to be distinct species on their own, given their nearly constant defining morphological characters and mostly allopatric distribution. However, before this study there was no molecular evidence of such differentiation, which could lead at any time to reversion to past taxonomic decisions. *Lycaena tityrus*, and *Lycaena bleusei*, upon their molecular study appear now as two independently evolving species with a clear genetic differentiation and with different demographic histories. While *L. bleusei* probably evolved in a relatively warm period within the Central Iberian Mountain System after separation

from the main Sooty Copper populations, *L. tityrus* might have either persisted in Iberia or elsewhere in temperate Europe during the several glacial periods of the Pleistocene. This geographical isolation allied to different selective pressures allowed for independent histories and differentiation in the absence of gene flow, even during the several interglacials that followed periods of maximum cold. There is now a solid ground in the attribution of a species status to *L. bleusei*, in spite of a recorded hybridization event.
As the ranges of both taxa approached each other very closely in western Iberia, namely central Portugal, a narrow secondary contact zone between both taxa was expected, but the discovery of a wide sympatry and syntopic area spanning most central and northern Portugal coupled with the existence of hybrids opens new possibilities in the study of this species complex.

When we work with fairly unkown and unstudied species some problems are bound to arise. Sampling butterfly populations for which there is little information of population size, behavior, ecological preferences and vulnerability does not render generally large sample sizes. A wider and more extensive prospection of the suitable habitats for *Lycaena bleusei* is in order, not only to describe more accurately the range of this species and ecological needs (collecting variables *in loco*) but also to clarify the extension of the secondary contact zone described in this project. This data should also be useful in the predictions of future habitat availability and occupancy, allowing modulation of habitat fragmentation and loss with predictive models that would ultimately lead to definition of the conservation status of this Iberia endemic species. Comparing and contrasting the same data for the Sooty Copper would also allow predicting future interactions (e.g. competition) and range changes between these two sister-taxa.

Another interesting aspect that could be included in the continuation of this work would be the addition of the subspecies *L. tityrus subalpinus*, in order to, again, clarify the phylogenetic relationships within the genus *Lycaena* and among close related taxa in the species group of *Lycaena virgaureae*. A different result should be expected in this case, however, since we are looking at a lower taxonomic level *i.e.* the genetic divergence from the nominal *L. tityrus* should present lower values than those resulting from these projects analyses.

During laboratory work, another pair of primers for the Phosphoglucose isomerase (PGI) gene was used. The optimization of the PCR mix and temperature cycles, however, was not achieved timely. The inclusion of a nuclear non-neutral gene, shown to have adaptive value in other butterfly species, could carry additional information on selective pressures that are absent in the analyses of neutral molecular markers. This gene has proven to be involved in flight metabolic rate, dispersal rate, fecundity and local population growth rate but can also provide insights into gene – environment interactions, local adaptation and morphologically cryptic dispersal of phenotypes (Orsini *et al.* 2009; Wheat *et al.* 2009; Wheat 2010), which could render a different and very important point of view in the study of evolutionary dynamics of *Lycaena bleusei* and *Lycaena tityrus*.

# 7. Bibliographic references


Andersen B, Borns Jr. H (1997) The Ice Age World. An introduction to Quaternary history and research with emphasis on North America and Northern Europe During the last 2.5 Million years. Scandinavian University Press, Oslo, Norway.

Anisimova M, Kosiol C (2009) Investigating protein-coding sequence evolution with probabilistic codon substitution models. *Molecular Biology and Evolution*, **26**, 255-271.

Avise J (2000) Phylogeography: the history and formation of species. Avise J (ed). Harvard University Press, Cambridge, United Kingdom. Chapter 1, 3-36.

Avise J, Arnold J, Ball R, Bermingham E, Lamb T, Neigel J, Reeb C, Saunders N (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematic. *Annual Review of Ecology and Systematics*, **18**, 489-522.

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

Bozano G, Weidenhoffer Z (2001) Guide to the butterflies of the Palearctic region. Lycaenidae part I subfamily Lycaeninae. Bozano G (ed). Omen Artes s.a.s., Milan, Italy.

Brower A (1994) Phylogeny of Heliconius butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). *Molecular Phylogenetics and Evolution*, **3**, 159-174.

Brower A (2008) Lycaena Fabricius 1807. The Coppers. Version 29 April 2008 (under construction). http://tolweb.org/Lycaena/111507/2008.04.29in The Tree of Life Web Project, http://tolweb.org/

Cassulo L, Mensi P, Balletto E (1989) Taxonomy and evolution in *Lycaena* (subgenus *Heodes*) (Lycaenidae). *Nota Lepidopterologica*, **Supplement 1**, 23-25.

Caterino M, Cho S, Sperling F (2000) The current state of insect molecular systematics: A thriving tower of Babel. *Annual Review of Entomology*, **45**, 1-54.

Chang B, Donoghue M (2000) Recreating ancestral proteins. *TREE*, **15**, 109-114.

Chianchi R, Ungaro A, Marini M, Bullini L (2003) Differential patterns of hybridization and introgression between the swallowtails *Papilio machaon* and *P. hospiton* from Sardinia and Corsica islands (Lepidoptera, Papilionidae). *Molecular Ecology*, **12**, 1461-1471.

Cho S, Mitchell A, Regier J, Mitter C, Poole R, Friedlander T, Zhao S (1995) A highly Conserved Nuclear Gene for low-level phylogenetics: *Elongation Factor-1α* Recovers morphology-based tree for Heliothine Moths. *Molecular Biology and Evolution*, **12**, 650-656.

Cordero C, Soberón J (1990) Non-resource based territoriality in males of the butterfly *Xamia Xami* (Lepidoptera: Lycaenidae). *Journal of Insect Behavior*, **3**, 719-732.

Dasmahapatra K, Lamas G, Simpson F, Mallet J (2010) The anatomy of a 'suture zone' in Amazonian butterflies: a coalescent-based test for vicariant geographic divergence and speciation. *Molecular Ecology*, **19**, 4283-4301.

Davies N, Aiello A, Mallet J, Pomiankowski A, Silberglied R (1997) Speciation in two neotropical butterflies: extending Haldane's rule. *Proceedings of the Royal Society of London*, **264**, 845-851.

Dincă V, Zakharov E, Hebert P, Vila R (2010) Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. *Proceedings of the Royal Society*, **278**, 347-355.

Douwes P (1975) Territorial behavior in *Heodes virgaureae* L. (Lep., Lycaenidae) with particular reference to visual stimuli. *Norwegian Journal of Entomology*, **22**, 143-153.

Drummond A, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analyses by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Elias S (1996) Late Pleistocene and Holocene seasonal temperatures reconstructed from fossil beetle assemblages in the Rocky Mountains. *Quaternary Research*, **46**, 311-318.

Elias S (2000) Late Pleistocene climates of Beringia, based on analyses of fossil beetles. *Quaternary Research*, **53**, 229-235.

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

García-Barros E, Munguira M, Cano J, Benito H, Garcia-Pereira P, Maravalhas E (2004) Atlas de las mariposas diurnas de la Península Ibérica e islas Baleares (Lepidoptera: Papilionoidea & Hesperioidae). Sociedad Entomológica Aragonesa (SEA) (ed). Universidade Autónoma de Madrid, Madrid, Spain.

Gómez A, Lunt D (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *In*: Phylogeography of Southern European Refugia. Evolutionary perspectives on the origins and conservation of European biodiversity, Weiss F and Ferrand N (ed). First edition, Springer, The Netherlands, Chapter 5, 155-188.

Habel J, Rödder D, Schmitt T, Nèves G (2011) Global warming will affect the genetic diversity and uniqueness of *Lycaena helle* populations. *Global Change Biology*, **17**, 194-205.

Hajibabaei M, Janzen D, Burns J, Hallwachs W, Hebert P (2006) DNA barcodes distinguish species of tropical Lepidoptera. *PNAS*, **103**, 968-971.

Hebert P, Penton E, Burns J, Janzen D, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *PNAS*, **101**, 14812-14817.

Heled J, Drummond A (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**.

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

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Hewitt G (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society*, **359**, 183-195.

Hoyle M, James M (2005) Global warming, human population pressure, and viability of the world's smallest butterfly. *Conservation Biology*, **19**, 1113-1124.

Huelsenbeck J, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754-755.

Jong R, van Dorp K (2006) The African copper connection. *Entomologische Berichten*, **66**, 124-129.

Karl I (2008) Thermal adaptation in butterflies: patterns, significance and mechanisms. *PhD thesis*, Fakultät Biologie, Chemie und Geowissenschaften der Universität Bayreuth, Bayreuth, Germany.

Karl I, Sørensen J, Loeschcke V, Fischer K (2009) HSP70 expression in the Copper butterfly Lycaena across altitudes and temperatures. *Journal of Evolutionary Biology*, **22**, 172-178.

Kim M, Wan X, Kim M, Jeong H, Ahn N, Kim K-G, Han Y, Kim I (2010) Phylogenetic relationships of true butterflies (Lepidoptera: Papiolionoidea) inferred from COI, 16S rRNA and EF-1α sequences. *Molecular Cells*, **30**, 409-425.

Kronforst M, Young L, Blume L, Gilbert L (2006) Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution*, **60**, 1254-1268.

Kuchta S (2007) Contact zones and species limits: hybridization between lineages of the California newt, *Taricha torosa*, in the Southern Sierra Nevada. *Herpetologica*, **63**, 332-350.

Kunte K, Shea C, Aardema M, Scriber J, Juenger T, Gilbert L, Kronforst M (2011) Sex chromosome mosaicism and hybrid speciation among Tiger Swalloetail butterflies. *PLOS Genetics*, **7**.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analyses of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.

Mallet J, Wynne I, Thomas C (2010) Hybridization and climate change: brown argus butterflies in Britain (*Polyommatus* subgenus *Aricia*). *Insect Conservation and Diversity*, **4**, 192-199.

Marabuto E, Pires P, Cardoso J (2004) Contribuição para o conhecimento dos Macrolepidópteros do Parque Natural da Serra da Estrela, Portugal (Lepidoptera). *Boletin. Sociedad Entomológica Aragonesa*, **34**, 171-175.

Maravalhas E (2003) As borboletas de Portugal. Maravalhas E (ed). Apollo Books, Stenstrup, Denmark.

Nice C, Anthony N, Gelembiuk G, Raterman D, FFrench-Constant R (2005) The history and geography of diversification within the butterfly Lycaeides in North America. *Molecular Ecology*, **14**, 1741-1754.

Nice C, Shapiro A (1999) Molecular and morphological divergence in the butterfly genus *Lycaeides* (Lepidoptera: Lycaenidae) in North America: evidence of recent speciation. *Journal of Evolutionary Biology*, **12**, 936-950.

Nielson R, Wakeley JW (2001) Distinguishing migration from isolation: an MCMC approach. *Genetics*, **158**, 885-896.

Nylander J (2004) MrModelTest v2. Evolutionary Biology Center, Uppsala University.

Parmesan C, Ryrholm N, Stefanescu C, Hill J, Thomas C, Descimon H, Huntley B, Kaila L, Kullberg J, Tammaru T, Tennent W, Thomas J, Warren M (1999) poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*, **399**, 579-583.

Paulo OS, Dias C, Bruford MW, Jordan WC, Nichols RA (2001) The persistence of Pliocene populations through the Pleitocene climatic cycles: evidence from the phylogeography of an Iberian lizard. *Proceedings of the Royal Society London*, **268**, 1625-1630.

Pina-Martins F, Paulo OS (2008) Concatenator: Sequence Data Matrices Handling Made Easy. *Molecular Ecology Resources*, **8**, 1254-1255.

Posada D, Crandall K (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

Orsini L, Wheat C, Haag C, Kvist J, Frilander M, Hanski I (2009) Fitness differences associated with *Pgi* SNP genotypes in the Glanville fritillary butterfly (Melitaea cinxia). *Journal of Evolutionary Biology*, **22**, 367-375.

Rambaut A, Drummond A (2007) Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer

Randi E (2007) Phylogeography of South European mammals. *In:* Phylogeography of Southern European Refugia. Evolutionary perspectives on the origins and conservation of European biodiversity, Weiss F and Ferrand N (ed). First edition, Springer, The Netherlands, Chapter 3, 101-126.

Reed R, Sperling F (1999) Interaction of process particions in phylogenetic analyses: an example from the Swallowtail Butterfly genus *Papilio.Molecular Biology and Evolution*, **16,** 286-297.

Remington C (1968) Suture-zones of hybrid interaction between recently joined biotas. *Evolutionary Biology*, **2**, 321-428.

Rosenberg N, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews*, **Genetics 3**, 380-390.

Rutowski R (1991) The evolution of male-locating behavior in butterflies. *American Naturalist*, **138**, 1121-1139.

Schilthuizen M, Giesbers M, Beukeboom LW (2011) Haldane's rule in the 21st century. *Heredity*, **107**, 95-102.

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

Scott A (1974) Mate-locating behavior of butterflies. *American Midland Naturalist*, **91**, 103-117.

Sperling H (1993) Mitochondrial DNA variation and Haldane's rule in the *Papilio glaucus* and *P. troilus* species groups. *Heredity*, **71**, 227-233.

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.

Swofford D (2001) PAUP*. Phylogenetic analysis using parsimony (* and other methods), Version 4.0b10. Sinauser Associates, Sunderland, Massachusetts.

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.

Takeuchi T, Imafuku M (2005) Territorial behavior of a Green Hairstreak *Chrysozephyrus smaragdinus* (Lepidoptera: Lycaenidae): site tenacity and wars of attention. *Zoological Science*, **22**, 989-732.

van Dorp K (2004) Molecular systematic of *Lycaena* F., 1807 (Lepidoptera: Lycaenidae) – Some preliminary results. *Proceedings of Netherland Entomological Society*, **15**, 65-70.

van Swaay C, Warren M (1999) Red Data Book of European Butterflies (Rhopalocera). Nature and Environment (ed). Council of Europe Publishing, Strasbourg.

Wahlberg N, Weingartner E, Warren A, Nylin S (2009) Timing major conflict between mitochondrial and nuclear genes in species relationships of *Polygonia* butterflies (Nymphalidae: Nymphalini). *BMC Evolutionary Biology*, **9**.

Wallisdevries M, van Swaay C (2006) Global warming and excess nitrogen may induce butterfly decline by microclimatic cooling. *Global Change Biology*, **12**, 1620-1626.

Warren M, Hill J, Thomas , Asher J, Fox R, Huntley B, Roy D, Telfer M, Jeffcoate S, Harding P, Jeffcoate G, Willis S, Greatorex-Davies J, Moss D, Thomas D (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature*, **434**, 65-68.

Wheat C (2010) Phosphoglucose isomerase (*Pgi*) performance and fitness effects among Arthropods and its potential role as an adaptive marker in conservation genetics. *Conservation Genetics*, **11**, 387-397.

Wheat C, Haag C, Marden J, Hanski I, Frilander M (2009) Nucleotide Polymorphism at a Gene (*Pgi*) under Balancing Selection in a Butterfly Metapopulation. *Molecular Biology and Evolution*, **27**, 267-281.

Whinnett A. Zimmermann M, Willmott K, Herrera N, Mallarino R, Simpson F, Joron M, Lamas G, Mallet J (2005) Strikingly variable divergence times inferred across an Amazonian butterfly 'suture zone'. *Proceedings of the Royal Society*, **272**, 2525-2533.

Wilson R, Gutiérrez D, Gutiéterrez J, Monserrat V (2007) An elevational shift in butterfly species richness and composition accompanying recent climate change. *Global Change Biology*, **13**, 1873-1887.

Yang Z, Bielawski J (2000) Statistical methods for detecting molecular adaptation. *TREE* , **15**, 469-503.

Zink R, Barrowclough G (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107-2121.

# Appendix

**Table 5.** Palearctic morphologic species groups according to Bozano & Weidenhoffer 2001.

| *L. phlaeas* | *L. dispar* | *L. helle* | *L. virgaureae* | *L. hippothoe* | *L. thersamon* |
|---|---|---|---|---|---|
| *L. kiyokoae* | *L. pavana* | *L. svenhedini* | *L. tityrus* | *L. candens* | *L. alaica* |
| *L. sichuanica* | *L. violacea* | *L. irmae* | *L. bleusei* | | *L. phoebus* |
| | *L. splendens* | *L. li* | *L. ottomanus* | | *L. lampon* |
| | *L. dobrerai* | *L. ouang* | *L. alciphron* | | *L. lamponoides* |
| | *L. kasyapa* | *L. pang* | | | *L. eberti* |
| | *L. aeolus* | *L. tseng* | | | *L. aditya* |
| | *L. standfussi* | | | | *L. solski* |
| | | | | | *L. alpheraki* |
| | | | | | *L. asabinus* |
| | | | | | *L. ochimus* |
| | | | | | *L. thetis* |

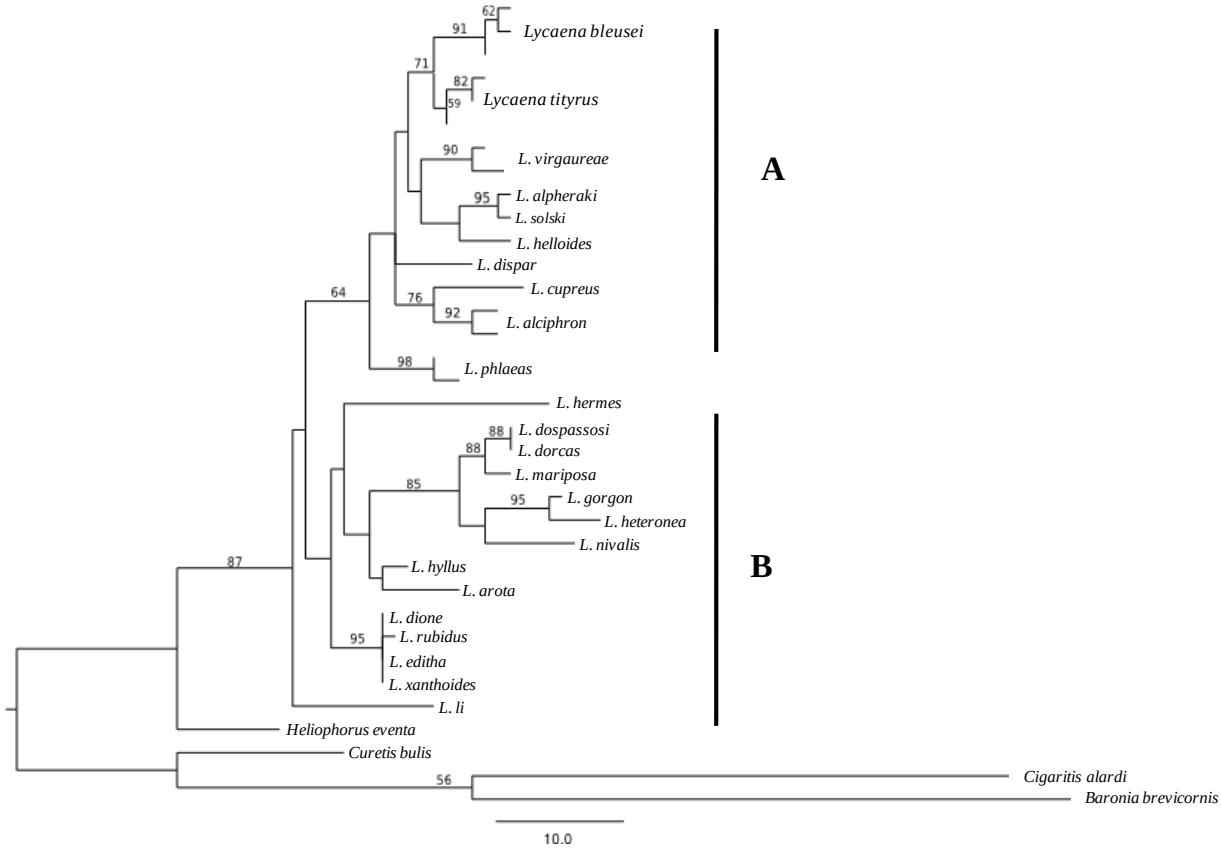


**Figure 11** Maximum parsimony tree based on Ef-1α haplotypic data (one of twenty six trees with equal tree lenght). Values on branches correspond to MP bootstrap values (only values > 50% are shown). **A** correspondes to Palearctic group and **B** corresponds to Nearctic group.

**Table 6.** Genbank Accession numbers of the samples used on the phylogenetic analyses with COI and EF-1α.

| | Genbank Accession Number | |
|---|---|---|
| | **COI** | **EF-1α** |
| ***Baronia brevicornis*** | AF170866 | AF173406 |
| ***Curetis buli*** | DQ018942 | EU024670 |
| ***Heliophorus eventa*** | FJ490484 | FJ490512 |
| ***L. xanthoides*** | FJ490483 | FJ490511 |
| ***L. editha*** | FJ490476 | FJ490504 |
| ***L. rubidus*** | FJ490467 | FJ490495 |
| ***L. dione*** | FJ490480 | FJ490508 |
| ***L. arota*** | FJ490490 | FJ490498 |
| ***L. hyllus*** | FJ490479 | FJ490507 |
| ***L. heteronea*** | FJ490469 | FJ490497 |
| ***L. gorgon*** | FJ490465 | FJ490494 |
| ***L. nivalis*** | FJ490468 | FJ490496 |
| ***L. mariposa*** | FJ490487 | FJ490516 |
| ***L. dorcas*** | GU0969761 | FJ490514 |
| ***L. dospassosi*** | FJ490486 | FJ490515 |
| ***L. hermes*** | FJ490474 | FJ490502 |
| ***L. li*** | FJ490473 | FJ490501 |
| ***L. phlaeas*** | HM391834 | GU372656 |
| ***L. virguareae*** | HM393191 | FJ490505 |
| ***L. helloides*** | FJ490472 | DQ018915 |
| ***L. solski*** | FJ490490 | FJ490520 |
| ***L. alpheraki*** | FJ490464 | FJ490521 |
| ***L. alciphron*** | JN20495 | JN204975 |
| ***L. cupreus*** | FJ490471 | FJ490499 |
| ***L. dispar*** | HQ004651 | GU372655 |
| ***L. helle*** | GU688466 | - |
| ***L. splendens*** | FJ664056 | - |
| ***L. asabinus*** | AY556876 | - |
| ***L. thetis*** | AY557116 | - |
| ***L. thersamon*** | AY556988 | - |
| ***L. candens*** | AY556890 | - |
| ***L. hippothoe*** | GU688462 | - |
| ***L. margelanica*** | FJ490492 | - |
| ***L. ophion*** | FJ663673 | - |