UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Evolutionary history of *Philaenus spumarius* (Hemiptera, Aphrophoridae) and the adaptive significance and genetic basis of its dorsal colour polymorphism

Doutoramento em Biologia

Especialidade em Biologia Evolutiva

Ana Sofia Bartolomeu Rodrigues

Tese orientada por:

Professor Doutor Octávio Paulo

Doutor Chris Jiggins

Documento especialmente elaborado para a obtenção do grau de doutor

2016

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Evolutionary history of *Philaenus spumarius* (Hemiptera, Aphrophoridae) and the adaptive significance and genetic basis of its dorsal colour polymorphism

Doutoramento em Biologia

Especialidade em Biologia Evolutiva

Ana Sofia Bartolomeu Rodrigues

Júri:

Presidente:

- Doutora Maria da Luz da Costa Pereira Mathias, Professora Catedrática da Faculdade de Ciências da Universidade de Lisboa, Presidente do júri por subdelegação de competências

Vogais:

- Doutor Thomas Schmitt, University Professor (W3) da Faculty of Natural Sciences I da Martin Luther Universität Halle Wittenberg (Alemanha)
- Doutor Diogo Francisco Caeiro Figueiredo, Professor Catedrático da Escola de Ciências e Tecnologia da Universidade de Évora
- Doutora Maria Alice da Silva Pinto, Professora Adjunta da Escola Superior Agrária do Instituto Politécnico de Bragança
- Doutor José Alberto de Oliveira Quartau, Professor Catedrático Aposentado da Faculdade de Ciências da Universidade de Lisboa
- Doutor Octávio Fernando de Sousa Salgueiro Godinho Paulo, Professor Auxiliar da Faculdade de Ciências da Universidade de Lisboa
- Doutora Sofia Gonçalves Seabra, Professora Auxiliar Convidada da Faculdade de Ciências da Universidade de Lisboa

Documento especialmente elaborado para a obtenção do grau de doutor

2016

TABLE OF CONTENTS

NOTE:

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

AGRADECIMENTOS / ACKNOWLEDGEMENTS

Gostaria de agradecer a todos aqueles que contribuiram para a realização deste trabalho e sem o apoio dos quais esta tese não teria sido possível.

Ao Professor Doutor Octávio S. Paulo e à Doutora Sofia Seabra por me terem sugerido este projeto de doutoramento e pelo apoio no planeamento e realização do mesmo. Por terem partilhado os seus conhecimentos e me atrairem para o mundo da Biologia Evolutiva. Pelo seu entusiasmo e encorajamento face aos muitos desafios colocados ao longo deste trabalho. Agradeço ainda pela confiança depositada nas minhas capacidades e pela oportunidade de aprender um pouco mais sobre este inseto fascinante.

To Dr Chris Jiggins for kindly accepting to be my co-adviser and for giving me access to his laboratory at Zoology Department of Cambridge University. For contributing with his expertise to this project. Thank you so much for your patience to attend all my requests and to comment the drafts of my thesis and manuscripts.

To the Financial support provided from National Funds through FCT – Fundação para a Ciência e a Tecnologia, Portugal under projects PTDC/BIA-BEC/098783/2008 and UID/BIA/00329/2013, and grant SFRH/BD/73879/2010.

A todos aqueles que colaboraram e me ajudaram na recolha de amostras no campo: Sara E. Silva, Eduardo Marabuto, José A. Quartau, Paulo A.V. Borges, Rui Nunes, Luís Ferreira, Carla Pereira, Paulo Fonseca, Vera Nunes, Filipe Ribeiro, Paula Simões, Andreia Penado, Selçuk Yurtserver, Olli Halkka, Liisa Halkka, K. Halkka, Mike Wilson, Vinton Thompson, Andy Beckenbach, Martin Corley, Vera d' Urso, Federico Lessio, Jerome Constant e MajaDerlink.

To everyone that helped me in the laboratory work at Faculdade de Ciências and at Zoology Department of Cambridge University, specially to Sofia Seabra, Sara E. Silva, Carla Ribeiro and Dr. Simon Baxter.

À Sara, uma amiga que acompanhou de perto todos os altos e baixos desta minha caminhada e com quem tenho partilhado esta aventura ciêntifica desde o 1º ano da faculdade. Obrigada pelos bons momentos passados no campo e por aturares os meus “stresses”. Obrigada por me ajudares no laboratório e pela disponibilidade que demonstraste quando te pedi para reveres os manuscritos da tese e artigos. Sobretudo, obrigada pela tua amizade.

À Carla Ribeiro, uma amiga com quem partilhei esta aventura. Obrigada pela tua ajuda no laboratório e não só. Agradeço pela tua boa disposição e alegria que me contagiaram nos momentos menos bons. Não me esqueço das boleias que me deste sempre que os senhores revisores e maquinistas dos transportes públicos resolviam tirar uns dias para "descansar". Obrigada por teres sido uma companheira nesta viagem, pelas palavras de conforto e incentivo.

Ao Francisco Pina Martins, agradeço pela disponibilidade e paciência demonstrada aquando dos meus "stresses" bioinformáticos. Obrigada por me ajudares a entrar neste mundo agora não tão assustador. Hoje posso dizer que sei escrever um script, ler código e resolver problemas básicos de programação. Contudo, não sou uma "expert" e sei que ainda te vou chatear muito, afinal és o único capaz de pôr o PC a funcionar apenas por estares presente.

À Catarina Dourado, por todo o carinho e amizade. Pelos bons momentos que passámos juntas e pelas palavras de conforto. Obrigada pelos momentos de descontração e procrastinação. Agradeço-te a ti e à Claúdia pelo tempo passado na vossa companhia e que me permitiu esquecer um pouco os "stresses" do trabalho. Pelas nossas aventuras e passeios pela Europa que espero continuar a partilhar com vocês.

Ao Eduardo, pela imprescindível ajuda no trabalho de campo. Pela sua boa disposição e palavras de incentivo. Mas principalmente por ser um bom amigo.

À Telma, pelas suas palhaçadas que me fizeram rir nos bons e maus momentos. Pelo incentivo e apoio. Pelas nossas conversas sobre tudo e sobre nada.

Aos meus restantes amigos e colegas do CoBig, agradeço pela amizade e pelos momentos de boa disposição, pela entreajuda e partilha de conhecimentos e sobretudo pelo encorajamento nos momentos mais difíceis da tese.

A toda a minha família, em especial à minha mãe e à minha irmã, pelo apoio e incentivo que sempre me deram ao longo destes anos e por estarem comigo nos momentos mais difícies da realização desta tese. Mãe obrigada pela tua força e coragem e por todos os sacrifícios que fizeste por mim e que me permitiram chegar até aqui.

A todos aqueles que me esqueci de mencionar mas sem os quais esta tese não seria possível.

ABSTRACT

Understanding the genetic basis of adaptive traits and how natural populations adapt to their environment, are fundamental problems in evolutionary biology. Colour polymorphisms are good systems in which these issues can be addressed. This work exploits the adaptive significance of the dorsal colour polymorphism in *Philaenus spumarius* and describes the efforts to identify genomic region(s) linked to the dorsal colour variation in this species. It also involves the investigation of the evolutionary pattern of *P. spumarius*. The phylogeographic results showed that main demographic and evolutionary events for the European populations seem to have occurred during Pleistocene, probably as a consequence of the main climatic oscillations that characterised this period. Evidence of recent gene-flow among Mediterranean peninsulas, as well as a close relationship between Iberia and North Africa, a probable British origin for the populations of the Azores and New Zealand, and indication that both western and north-eastern Europe colonised North America, were also found. Captivity experiments, testing the adaptive function of *P. spumarius'* dorsal colour polymorphism, indicated a higher longevity, a higher number of oviposition events, and a higher number of eggs laid for *trilineatus* than for *marginellus* and *typicus*. A total of 1,837 genomic markers (SNPs) and 928 loci were obtained through RAD sequencing for 33 individuals of three colour phenotypes (*trilineatus*, *marginellus* and *typicus*), and a genome wide association study performed to identify regions related to dorsal colour variation. Single and multi-association analyses identified a total of 60 SNPs associated with dorsal colour phenotypes but none of these SNPs showed homology with colour genes described in other insect species. A stronger differentiation of the *trilineatus* colour phenotype was also found with these markers. New genomic and transcriptomic resources were developed, constituting important tools and a basis for future research in this species. The resources correspond to a genome draft (25% of the total estimated genome (5.3 Gb)) and to an 81.4 Mb transcriptome assembly. The analysis of the candidate gene *yellow* found no association with dorsal colour phenotypes. Although suggesting that *yellow* is not involved in colour variation in *P. spumarius*, a possible effect of this gene can not be totally excluded without further analyses. Phylogenetic analyses found contrasting patterns between *yellow* and the neutral *ITS2*, indicating that, contrary to *ITS2, yellow* is conserved within the genus. A higher haplotype and nucleotide diversity was observed for *P. spumarius*. This could be related to a higher sample size or to the different ecology of the species. The present thesis showed that *P. spumarius* constitutes a potential good model system to study adaptation and the evolutionary mechanisms involved in the maintenance of polymorphisms.

Keywords: adaptation, association study, candidate genes, colour variation, spittlebug.

RESUMO

Compreender a base genética de características adaptativas e o modo como as populações naturais evoluem e se adaptam ao seu ambiente é fundamental e constitui um dos temas mais interessantes para a Biologia Evolutiva. Os polimorfismos de coloração são bastante comuns em animais e plantas e, quando combinados com dados genéticos e ecológicos, podem ser excelentes sistemas para estudar não só o processo de adaptação mas também as alterações moleculares subjacentes à variação fenotípica. O presente trabalho teve como principais objetivos a exploração da importância adaptativa de um polimorfismo de coloração e a identificação de genes ou regiões genómicas potencialmente envolvidos na variação de cor. O organismo escolhido para este estudo foi a espécie *Philaenus spumarius* (L.) (Hemiptera, Aphrophoridae), vulgarmente conhecida como cigarrinha-da-espuma. Este inseto, que se encontra amplamente distribuído pela região Holártica, apresenta vários padrões de coloração da zona dorsal do corpo, com 16 fenótipos melânicos e não melânicos descritos, e cuja base genética e significado adaptativo permanecem por investigar.

Para melhor compreender o processo de adaptação e, neste caso, para melhor estudar a base molecular que está por trás do polimorfismo dorsal de coloração de *P. spumarius,* é importante perceber a história evolutiva da espécie. Ao comparar-se o padrão de genes neutrais com o padrão de genes adaptativos, os efeitos da história evolutiva e/ou demografia são passíveis de ser distinguidos dos efeitos da seleção natural. Embora existam vários estudos genéticos sobre a biogeografia da espécie, estes focam-se, maioritariamente, em populações Europeias e carecem de estimativas de tempo para os principais eventos evolutivos e demográficos. Assim, uma primeira fase deste trabalho envolveu a investigação do padrão biogeográfico da cigarrinha-da-espuma, usando marcadores mitocondriais (Citocromo Oxidase I, Citocromo Oxidase II e Citocromo *b*) e nucleares ("Elongation Factor 1α"). Pela primeira vez obtiveram-se estimativas de tempo para os principais eventos demográficos e evolutivos ocorridos para esta espécie nas penínsulas Mediterrânicas e, analisaram-se os seus padrões de colonização fora da Eurasia, nomeadamente, no norte de África e na América do norte. Os resultados indicam que a divergência entre as populações de *P. spumarius* é recente, tendo ocorrido no Pleistoceno Médio/Superior, e está muito provavelmente ligada às alterações climáticas do período Quaternário. Os marcadores mitocondriais mostram que terá ocorrido uma primeira separação da cigarrinha-da-espuma em duas grandes linhagens: a linhagem "ocidental", na região Mediterrânica, e a linhagem "oriental", na zona da Anatolia/Cáucaso. Posteriormente, ter-se-á dado a diferenciação da linhagem "ocidental" em duas sub-linhagens: a linhagem

"Mediterrâneo-ocidental" na Península Ibérica e a linhagem "Mediterrâneo-oriental" localizada na região dos Balcãs. Este padrão, no entanto, difere do padrão nuclear, o que sugere a existência de cruzamentos e/ou separação incompleta de linhagens. Os eventos de divergência que tiveram lugar nas várias penínsulas Mediterrânicas (refúgios a sul) terão, provavelmente, ocorrido durante os períodos glaciares do Pleistoceno. Após esses períodos, as populações de *P. spumarius* ter-se-ão expandido em direcção ao norte, a partir da Península Ibérica para o centro e oeste da Europa, e da Anatolia/Cáucaso (ou zonas mais a este) para o leste e centro da Europa. Este estudo também detetou a existência de fluxo genético entre as penínsulas Mediterrânicas e uma relação próxima entre as populações Ibéricas e do norte de África. As populações dos Açores e Nova Zelândia tiveram muito provavelmente uma origem a partir das populações Britânicas e a colonisação da América do norte parece ter sido feita por indivíduos pertencentes não só à linhagem ocidental mas também pertencentes à linhagem presente no norte da Europa.

Para tentar perceber porque é que o polimorfismo de cor se mantém nas populações naturais de *P. spumarius* e qual a sua importância adaptativa, foram realizadas experiências em cativeiro nas quais se testaram diferenças na sobrevivência, sucesso reprodutor e duração de maturação dos ovos em três fenótipos de cor (*typicus*, *trilineatus* e *marginellus*). Curiosamente, foi observada uma maior longevidade para a forma *trilineatus*, assim como um maior número de eventos de oviposição e maior número total de ovos. O estudo sugere que, na manutenção deste polimorfismo poderão estar envolvidos vários factores e que as diferenças observadas entre o *trilineatus* e os restantes fenótipos são, possivelmente, uma maneira destes indíviduos compensarem uma maior taxa de ataque por parasitóides e/ou uma maior refletância da radiação solar, ambos reportados em estudos anteriores.

No presente trabalho foi investigada a base genética deste polimorfismo. O objetivo consistiu em tentar identificar uma ou várias regiões genómicas associadas à variação do padrão de coloração dorsal em *P. spumarius*. Duas abordagens foram usadas. A primeira abordagem implicou um estudo de associação no qual se usaram marcadores ("Single Nucleotide Polymorphisms (SNPs)) distribuídos ao longo do genoma, obtidos através da técnica "Restriction Associated DNA (RAD) sequencing". Com este método de sequenciação, um conjunto de 1837 marcadores foi obtido para 33 indivíduos e, associações com três padrões de cor (*typicus*, *trilineatus* e *marginellus*), foram testadas. As análises identificaram um total de 60 SNPs associados com o padrão de coloração dorsal e revelaram uma maior diferenciação dos indivíduos *trilineatus*. Os indivíduos deste morfotipo também se revelaram os mais diferenciados em várias características fisiológicas e de história de vida testadas nas experiências

de cativeiro. Não foi encontrada homologia entre as regiões associadas e genes da cor já descritos para outras espécies. Os dados sugerem que *loci* de grande efeito, correspondendo a várias regiões do genoma, podem estar envolvidos na variação encontrada entre os três fenótipos investigados. Para além disso, indicam que uma arquitetura genética complexa pode estar a controlar quer a variação dos padrões de cor quer características da história de vida e que, a selecção natural pode não estar a atuar diretamente na cor. Para ajudar na caracterização das regiões genómicas associadas à variação dos padrões de coloração, o genoma parcial, correspondendo a 25% dos 5.3 Gb estimados, e o transcriptoma de *P. spumarius* (81.4 Mb) foram sequenciados e "assemblados". Dos SNPs associados com a cor, 35% alinharam com o genoma e 10% com o transcriptoma indicando que, caso o transcriptoma esteja bem representado, a maioria dos SNPs associados está em regiões não codificantes.

Uma abordagem de genes candidatos foi igualmente usada para investigar genes que estão envolvidos na coloração em outros insectos, e que podem potencialmente contribuir para o padrão de coloração em *P. spumarius*. Os padrões de cor em *P. spumarius* variam desde totalmente melânicos a totalmente pálidos. Em *Drosophila*, o gene *yellow* é um gene envolvido na síntese de melanina e, como tal, constitui um potencial candidato para a variação dos padrões de cor em *P. spumarius* e nas restantes espécies do género, também elas polimórficas para o padrão de coloração. Uma possível associação entre o *yellow* e os morfotipos *typicus*, *trilineatus* e *marginellus* foi testada mas nenhuma relação foi encontrada, sugerindo que este gene pode não estar diretamente envolvido na variação de cor nesta espécie. No entanto, como só uma parte do gene foi investigada, o seu envolvimento não pode ser totalmente excluído. A análise filogenética envolvendo o *P. spumarius* e algumas espécies próximas detetou a existência de padrões contrastantes entre este gene e o gene nuclear *Internal Transcribed Spacer 2* (*ITS2*). Isto indica que, contrariamente ao *ITS2,* que separa as espécies em dois grandes grupos, o gene *yellow* é conservado dentro do género. Para o *yellow* foi encontrado o mesmo haplotipo em quase todas as espécies de *Philaenus* excepto para o *P. maghresignus* e *P. arslani*. Uma maior diversidade genética foi observada para o *P. spumarius*, podendo ser resultado da diferente ecologia da espécie. Este estudo demonstrou que, apesar de a aplicação de uma abordagem de genes candidatos, numa espécie como o *P. spumarius,* ser difícil, é importante investigar outros genes que possam estar envolvidos na determinação deste polimorfismo.

A presente tese mostrou que a cigarrinha-da-espuma, *P. spumarius,* é potencialmente um bom modelo para estudar o processo de adaptação e os mecanismos evolutivos envolvidos na manutenção dos polimorfismos nas populações naturais.

Palavras-chave: adaptação, cigarrinha-da-espuma, estudo de associação, polimorfismo de coloração.

CHAPTER 1

General Introduction

1. GENERAL INTRODUCTION

1.1 A view of the adaptation process

The idea that natural selection plays an important role in the process of adaptation was first proposed by Charles Darwin and Alfred Wallace (1858) and later on Darwin's book *On the Origins of Species by Means of Natural Selection* (1859). Understanding how populations adapt to their environment, and the genetic basis underlying this process are challenging and central problems in evolutionary biology. In the past few years, the scientific knowledge on this area has improved, mainly due to the developing of new sequencing technologies and molecular tools. Long-standing questions about the genetic mechanisms behind phenotypic evolution are now beginning to be addressed.

Does adaptive evolution implicate small or major effect genes/loci? Are there many or few genes/loci involved? The earlier view of genetic basis of adaptation defended by Fisher (Fisher, 1930) was based on the idea that phenotypic evolution and adaptation of organisms to their environment occurred through accumulation of small effect mutations arising in a large number of genes with additive effects (reviewed in Orr & Coyne, 1992). Empirical work, however, applying either Quantitative Trait Locus (QTL) analysis or experimentally testing evolution in microbes put this idea into question (Orr, 2005). Several studies suggested the involvement of large effect loci other than small effect loci, for example, on the evolution of body armour or pelvic structure in threespine stickleback *Gasterosteus aculeatus* (Shapiro *et al.*, 2006; Chan *et al.*, 2010), on the loss of larval trichomes in *Drosophila* species (Sucena & Stern, 2000; McGregor *et al.*, 2007), and on the colour in mice (Hoekstra *et al.*, 2005). Therefore, these studies indicate that adaptive change may also implicate mutations of large effect and not only small ones as previous accepted. However, how frequent large effect loci are in relation to polygenic traits is still an open question.

Another interesting question is related with the source of adaptive genetic variation. There is evidence that adaptation of populations to a novel or changing environments can occur either by selection of beneficial alleles from standing genetic variation or by selection of new mutations (Barrett & Schluter, 2008). Depending on the source of adaptive variation, different evolutionary pathways and distinct genetic outcomes are expected. The importance of *de novo* mutations was demonstrated, for example, on insect adaptation to chemical pesticides and herbicides (Walsh *et al.*, 2001; Weill *et al.*, 2004) and on the evolution of microorganisms in response to a change in the environmental conditions (Elena & Lenski, 2003; Burke, 2012). On the other hand, recent studies on ecologically important genes of the oldfield mice *Peromyscus*

polionotus (Steiner *et al.*, 2007) and of threespine stickleback *G. aculeatus* (Jones *et al.*, 2012) indicated that standing genetic variation plays an important role in facilitating rapid adaptation to novel environments. Besides, the fixation of beneficial alleles with small phenotypic effects and the spread of recessive alleles is more likely when adaptation relies on standing variation other than on *de novo* mutations (Barrett & Schluter, 2008). It was also suggested that adaptive alleles transmitted from one species to another by interbreeding or introgression (adaptive introgression), can be an alternative source for adaptation (Hedrick, 2013).

The relative importance of coding and regulatory changes in phenotypic evolution has also been a target of interest in evolutionary biology. Changes in protein-coding sequences were commonly thought to be the primary source for adaptive divergence. Examples include temperate climate adaptation in *Drosophila melanogaster* that was found to be associated with *de novo* mutations in sequences of several metabolic genes (Sezgin *et al.*, 2004), and beach mice's colour variation which is linked to a mutation in the coding region of the melanocortin-1-receptor gene (Mc1r) (Steiner *et al.*, 2007). However, over the last few years, it has been suggested that mutations in regulatory elements are more likely to contribute to adaptive divergence. This is because changes in regulatory regions can produce tissue-specific expression patterns while avoiding deleterious pleiotropic effects (Wittkopp & Kalay, 2011; Olson-Manning *et al.*, 2012).

Adaptive evolution can also result from mutations involving either a loss of function (e.g., reduced armor in threespine stickleback fish (Chan *et al.*, 2010), and lack of pigment in cavefish (Protas *et al*., 2006)) or a gain of a new function (e.g. adaptive variation in beach mice (Steiner *et al.*, 2007), and wing patterns variation in butterflies (Martin *et al.*, 2014)). In *Drosophila*, regulatory changes have been shown to be involved in the loss of trichomes on the larval cuticle (Sucena & Stern, 2000; McGregor *et al.*, 2007), in the gain of melanic wing spots (Wittkopp *et al.*, 2002a; Gompel *et al.*, 2005) and in changes in abdominal pigmentation (Rogers *et al.*, 2013).

Alternatively, populations can respond to environmental change through plasticity, that is, through phenotypic changes that do not depend on genetic change. Phenotypic plasticity may facilitate rapid and effective response to a changing environment such as the introduction of a novel predator (e.g. *Daphnia melanica*, Scoville & Pfrender, 2010) or climate change (e.g. the great tit *Parus major*, Charmantier *et al.*, 2008). Besides it could be a good option for species where adaptive change is slow, as result of their long generation times (Chevin *et al.*, 2010; Vander Wal *et al*., 2012). Plasticity may be important in the short term but, owing to fitness costs, adaptive genetic changes will still be essential, especially when species face

persistent environmental changes that overcome their capability to respond through phenotypic plasticity (Anderson *et al.*, 2012).

To disentangle evolutionary mechanisms underlying adaptation it is also important to understand how alleles at various loci interact (epistatic effects), the degree to which variation in a particular gene affects multiple traits (pleiotropy), and the possible interactions between genotype and environment (Radwan & Babik, 2012).

In most cases, addressing these questions has been challenging, mainly because many adaptive traits are quantitative and do not have a simple genetic basis. Colour patterns, in turn, are a source of enduring fascination for evolutionary biologists. They are very conspicuous (normally a discrete trait), in many cases they have a simple genetic basis, which make them easily discernible to the human eye (e.g flower colour polymorphism in *Ipomoea purpurea*, Zufall & Rausher, 2003).

1.2 Colour polymorphisms and the study of adaptation

Colour polymorphisms are usually defined as the presence of two or more distinct colour phenotypes, genetically determined, in a temporary or permanent balance within a single interbreeding population, the rarest of which being too frequent to be solely the result of recurrent mutation (Huxley, 1955). In balanced polymorphisms, the frequencies of the colour phenotypes tend to be stable over long time periods while in transient polymorphisms, the phenotypes frequencies do not remain in equilibrium over extended periods (Oxford & Gillespie, 2001). By acting as a visual/warning signal, used for example in mate choice or in predator avoidance, or by being involved in the thermoregulation, colour patterns may affect the fitness of individuals (Forsman *et al.*, 2008). Colour patterns, namely the melanic, also play a role in immune response (Wilson *et al.*, 2001), and in wound healing and cuticular hardening (Sugumaran, 2009). Besides, when interacting with other physiological and/or ecological traits, colour polymorphisms can act as camouflage (having a protective role) or influence the habitat choice, the dispersal capability, the distribution ranges or the adaptation to a changing or novel environment, thus influencing the ecological success and evolutionary dynamics of populations and species (Lozier *et al.*, 2016).

Therefore, colour polymorphisms are an extraordinary system for studying the process of adaptation and the micro-evolutionary forces maintaining the genetic variation underlying phenotypic traits (Gray & McKinnon, 2007). Several mechanisms are suggested to be involved in the maintenance of colour variation in natural populations (Gray & McKinnon, 2007).

Colour patterns can be maintained through negative frequency-dependent selection, where there is an advantage of the rare phenotypes over the common ones, resulting from processes such as predation or sexual selection (Ayala & Campbell, 1974; Punzalan *et al.,* 2005; Svensson *et al.,* 2005; Kusche & Meyer 2014). The maintenance of colour polymorphisms in natural populations can also result from heterozygote advantage, where heterozygous individuals have higher fitness than homozygous individuals (Vercken *et al.,* 2010). If different populations are subjected to different environmental conditions that favour different colour morphs, the polymorphism may be maintained due to an exchange of migrants, a mechanism called divergent selection with gene-flow (Jones *et al.,* 1977; Oxford 1985, 2005; Hoesktra *et al.*, 2004; Comeault *et al.*, 2015). There are also cases in which different colour phenotypes result in almost the same fitness values, explaining the maintenance of the polymorphism in some species (Roulin *et al.,* 2003). Colour variation is found in many animals and in plants, and investigation in this area have contributed to our understanding of key evolutionary processes at both population and species levels (see Bond 2007 for references). In vertebrates, there are several studies on the adaptive significance of colour polymorphisms in fishes (e.g. Munday *et al.,* 2003), amphibians (e.g. Hoffman & Blouin, 2000), reptiles (e.g. Perez i de Lanuza *et al.,* 2013), birds (e.g. Roulin, 2004), and mammals (e.g. Hoekstra *et al.*, 2005). In invertebrates, the evolution of colour variation was exploited, for example, in land snails (Jones *et al.*, 1977; Ożgo, 2012), spiders (Oxford & Shaw, 1986; Oxford & Gillespie, 1996a,b), grasshoppers (Tsurui *et al.*, 2010), butterflies (Brown & Benson, 1974; Bishop *et al.*, 1978; Nijhout, 2003) and ladybirds (Michie *et al.*, 2010).

1.3 Genetic basis of colour variation

The investigation of colour patterns, because they are conspicuous and easy to characterise, have greatly contributed to our understanding of the genetic mechanisms behind the maintenance of phenotypic variation and their role in the adaptation process. In many species, colour variation has a simple Mendelian inheritance, usually involving alleles at a single locus. That is the case of the garden pea plant *Pisum sativum*, used in Mendel's experiments, whose flower colour variation was found to be controlled by a single loci (*A*) (Hellens *et al.*, 2010). Another example is the pea aphid *Acyrthosiphon pisum*, which has two described morphs found to be controlled by a single autosomal locus with two alleles (*P* and *p*) (Caillaud & Losey, 2010). The presence or absence of yellow carotenoid pigments in the petals of pink-flowered *Mimulus lewisii* and its red-flowered sister species *Mimulus cardinalis* also

involves a single major locus (YUP) (Bradshaw & Schemske, 2003).

More complex genetic architectures, often involving two or more interacting loci of large effect, have been reported for numerous species as well. Steiner *et al.* (2007) found that differences in colour pattern of two sub-species of *P. polionotus* mice are associated with three loci, two of large and one of small effect. Similarly, loci with large phenotypic effects affecting adaptive colour variation were found in *Timema cristinae* stick insects (Comeault *et al.,* 2014).

A small number of major loci controlling colour pattern have been identified also in *Heliconius* species (Jiggins & McMillan, 1997; Joron et *al*., 2006a). In *Heliconius melpomene*, a cluster of three tightly linked (HmN, HmYb and HmSb) and other unlinked loci are associated with distinct wing colour pattern elements in this species (Joron *et al*., 2006b). On the other hand, in its related species *Heliconius numata,* colour variation is controlled by a single locus *P*, acting as a *supergene* (Joron *et al.* 1999*,* 2011). In *Papilio* species, the wing pattern phenotype was also found to be controlled by a *supergene* (Clarke & Sheppard, 1972; Nijhout, 2003), and Kunte *et al.* (2014) found that a single gene, *doublesex*, with closely linked mutations, controls the *supergene* mimicry in *Papilio polytes*.

In the early 1970s, *supergenes* were defined as "*coadapted combinations of several or many genes locked in inverted sections of chromosomes and therefore inherited as single units*" (Dobzhansky, 1970). A more recent and integrated view defines a *supergene* as a group of "*multiple linked functional genetic elements that allows switching between discrete, complex phenotypes maintained in a stable local polymorphism*" (Thompson & Jiggins, 2014). *Supergenes* are normally maintained due to reduced recombination resulting from either chromosomal rearrangements (e.g. inversions) or selection of co-adapted loci with epistatic or pleiotropic effects (Thompson & Jiggins, 2014). In the Hawaiian happy-face spider *Theridion grallantor*, the genetic control of colour pattern polymorphism involving abdomen and carapace was found to differ considerably between Maui (Oxford & Gillespie, 1996a) and Hawaiian populations (Oxford & Gillespie, 1996b). If a *supergene* may be controlling colour variation on Maui population, disruptive events like recombination and chromosomal rearrangements could have unlinked the two loci controlling colour pattern on Hawaiian populations (Oxford & Gillespie, 2001). Evidence that balanced polymorphisms can result from tight genetic linkage between multiple functional loci has been shown for birds (Tuttle *et al.,* 2016) and plants (see examples in Schwander *et al.*, 2014) as well. Also in the land snails *Cepaea hortensis* and *Cepaea nemoralis*, loci controlling variation in shell colour (*C*) and banding patterns (*B*) are known to be inherited together as a *supergene* (Jones *et al.*, 1977). A recent study on the *C. nemoralis'* polymorphism identified a set of RAD (restricted site-

associated DNA) markers linked to this species' *supergene*, and obtained a genetic map for the region, thus contributing for further research on the identity of the *supergene*'s loci and the evolution and maintenance of *Cepaea*'s polymorphism (Richards *et al.*, 2013).

The genetic basis of adaptive colour variation has also been widely studied in *Drosophila* species (Pool & Aquadro, 2007).

Recent advances has been made in the identification of functional or regulatory region(s) responsible for colour pattern variation. The following section present a brief description of the several colour genes already characterised in some species.

1.4 Colour genes for pigmentation variation

Pigmentation studies in mammals and other vertebrates have identified over 100 genes involved in colour variation (Hoekstra, 2006; Mills & Paterson, 2009) and have shown that the melanin pathway is highly conserved across vertebrates (Hubbard *et al.*, 2010). Most of those genes was identified through candidate gene and QTL mapping approaches which proved very successful in identifying the molecular basis underlying adaptive variation (Hoekstra, 2006; Takeda & Matsuoka, 2008). The Mc1r is one of the colour genes most studied but there are other genes responsible for colouration patterns in vertebrates (Protas & Patel, 2008; Bourgeois *et al.*, 2016). This gene is involved in melanin synthesis, and several different mutations in Mc1r were found to be responsible for adaptive colour variation in the beach mice (Hoesktra *et al.*, 2006), in lizards (Rosenblum *et al.*, 2004; Nunes *et al.*, 2011) and also in birds (Baião *et al.*, 2007). These studies have contributed to our understanding of the genetic basis of adaptive traits, specifically the adaptive significance of colour pattern variation in wild populations.

Unlike vertebrates, the molecular mechanisms underlying insect melanisation patterns are still poorly understood, but several genes and pathways have already been identified to be involved in colouration and pigmentation, mainly in *Drosophila* and some Lepidoptera species. Genetic crosses, the construction of genomic and transcriptomic libraries (BAC, UAS, EST), linkage maps, analysis of expression, association tests and candidate gene approaches, are some of the techniques that have been used to identify and locate genes/loci controlling colour variation in these species (e.g. *Drosophila*, see True *et al.*, 1999, 2005; Wittkopp *et al.*, 2002a and *Heliconius* butterflies, see Baxter *et al.*, 2009). Genome and transcriptome sequencing technologies have also contributed to the identification of colour gene/loci (e.g. *Heliconius* butterflies, Nadeau *et al.*, 2016).

The melanin synthesis process in *Drosophila* (Wittkopp *et al.*, 2003) involves the

conversion of tyrosine to DOPA. This reaction is catalysed by the tyrosine hydroxylase (TH), encoded by the *pale* gene (True *et al.*, 1999). DOPA is subsequently transformed into either DOPA-melanin (black melanin) or dopamine-melanin (brown melanin). The formation of DOPA-melanin is catalysed by phenoloxidase (PO) and by members of the *yellow* gene family (Wittkopp *et al.*, 2002b). On the other hand, conversion from DOPA to dopamine-melanin involves the transformation of DOPA into dopamine by DOPA decarboxilase, which is encoded by the *Ddc* gene, and subsequently into dopamine-melanin by PO. Dopamine is also the precursor of light pigments, yellow sclerotin and colourless or transparent sclerotin. The enzyme NBAD (N-b-alanyl dopamine) synthase, a product of the *ebony* gene (Wittkopp *et al.*, 2002a), can convert dopamine to NBAD, which is then used in the production of yellow sclerotin. However, this reaction can be reverted, and some NBAD can be converted into dopamine again by an NBAD hydrolase that is thought to be encoded by the *tan* gene (True *et al.*, 2005). Besides, a family of arylalkylamine-N-acetyl transferases (aaNATs) can also converts dopamine to N-acetyl dopamine, which serves as a precursor for colourless or transparent sclerotin.

Colour genes already characterised in *Drosophila* can be divided into two groups. There are the regulatory genes (e.g *doublesex*, *Abdominal-B*, *wingless, decapentaplegic*, *engrailed, optomotor-blind* and *bric-a-brac*) that control the distribution of pigments in space and time, and the structural genes that are directly involved in the synthesis of melanin based pigments (see Wittkopp *et al.*, 2003; Wittkopp & Beldade, 2009 for a review). Three of the most interesting structural genes are *yellow*, *ebony* and *tan.* The *yellow* gene promotes the formation of black melanin (Wittkopp *et al.*, 2002a), *ebony* is responsible for yellow sclerotin production (Wittkopp *et al.*, 2002a) while *tan* induces the brown pigmentation by catalysing *ebony* reverse reaction (True *et al.*, 2005; Wittkopp *et al.*, 2009). Evolutionary changes in regulatory elements have been demonstrated to alter the expression of *yellow* and *ebony* pigmentation genes and consequentially to cause differences in pigmentation pattern in *Drosophila melanogaster* (Wittkopp *et al.*, 2002a). *Cis*-regulatory change at the *yellow* gene is involved in the evolution of male-specific wing pigmentation spot in *Drosophila biarmipes* (Gompel *et al.*, 2005). In another study, regulation of *yellow* was found to contribute to divergent pigmentation patterns in three distantly related drosophilids (*D. melanogaster*, *Drosophila subobscura*, and *Drosophila virilis*) (Wittkopp *et al.*, 2002b). Moreover, Wittkopp *et al.* (2009) found that pigmentation differences within and between *Drosophila* species are caused in part by alleles of the *tan* gene, that share non-coding changes, and by changes in alleles of the *ebony* gene. These changes appear to rely on standing genetic variation in a common ancestor.

yellow and *ebony* are also involved in larval colour pattern in the silk worm *Bombyx mori* (Futahashi *et al.,* 2008). In *Papilio* spp. wing colour variation is controlled by the regulatory genes *doublesex* and *invected*/*engrailed* (Clark *et al.,* 2008; Kunte *et al.,* 2014). The *Distal-less* gene is linked to variation in eyespots size in *Bicyclus anynana* (Beldade *et al.*, 2002). *Antennepedia*, *invected*/*engrailed*, *spalt* and *notch* are also involved in eyespot pattern in *B. anynana* and in other nymphalid butterflies (Brunetti *et al.*, 2001; Oliver *et al.*, 2012).

Novel and unexpected genes were found to control wing colour variation in *Heliconius* species. *Cortex* gene, a member of a conserved cell cycle regulator family, appears to have adopted a novel function controlling colour pattern in *Heliconius* and probably across the Lepidoptera (Nadeau *et al*., 2016). In the peppered moth *Biston betularia*, an insertion of a tranposable element into the gene *cortex* is responsible for the melanisation pattern in this species (Hof *et al.*, 2016). Two distinct *cis*-regulatory loci (Wallbank *et al.,* 2016) regulate the expression of the transcription factor *optix*, which in turn is associated with red wing pattern variation across *Heliconius* (Reed *et al.,* 2011).

Candidate gene approach has been a very common method to study the genetic basis of adaptive colour variation but other methodologies, including genomic approaches, have been applied to a growing number of species.

1.5 From genetics to genomics

Not long ago, due to technological and cost constraints, investigating the evolutionary history of populations and species, establishing genotype-phenotype associations, mapping adaptive loci or identifying gene function, was limited not only to a few organisms but also to narrow regions of the genome. However, the emergence and advances in sequencing technologies (next-generation sequencing, NGS) over the last decade, and the consequent development of population genomics has opened a new and exciting era in the area of Biology (Ellegren, 2014). Scientists have now the capability to sample the entire genome and to generate huge amounts of genomic data at relatively low cost and less time consuming. Studies are now possible on a wide range of species for which, in some cases, there is extensive knowledge of ecological and evolutionary history, but few genomic resources (Amores *et al.,* 2011; Baxter *et al.,* 2011; Rowe *et al.,* 2011; Wagner *et al.,* 2013). This unlimited access to genome sequences from a multitude of species has brought the field of evolutionary research to a level where relevant evolutionary processes, including the genetic basis of adaptation, can now be investigated from a whole-genome perspective, and the effects of forces acting on the

entire genome distinguished from those influencing only particular loci (Black *et al.*, 2001; Stinchcombe & Hoekstra, 2008). In the past few years, the amount of genome sequences that became available in public databases has increased. In the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) (April 2016), for example, information about genome sequence assemblies is listed for 3,002 eukaryotes and most of them represent non-model species. The idea of getting whole-genome sequences from any organism is thus very promising. However, for some species, particularly those with large genome sizes, whole-genome sequencing is still difficult and extremely expensive. In these cases, reduced-complexity methodologies are, sometimes, the best approach. By allowing to sequence only a targeted fraction of the genome, they have been used to generate markers genome-wide (Narum *et al.*, 2013). Restricted site-associated DNA sequencing (RADseq) method (Baird *et al.* 2008) is a reduced complexity method of sampling that allows to increase the number and type of molecular markers, reduce the costs and increase the speed analyses. This method proved to be suitable for genetic analysis, including genotype-phenotype association mapping, phylogeography and population genomics. It is also useful in organisms lacking a reference genome or with complex histories, and in studies where a complete sequence genome is not always possible or suitable (Davey & Blaxter 2010; Hohenlohe *et al.*, 2010; Amores *et al.,* 2011; Baxter *et al.,* 2011; Davey *et al.,* 2011; Rowe *et al.,* 2011; Hohenlohe *et al.*, 2011, 2013).

Transcriptome sequencing is an efficient way to obtain functional data for non-model organisms or for those with genome characteristics incompatible with whole genome sequencing. This technique allows to focus the analysis on the transcribed portion of the genome (Parchman *et al.*, 2010). There are several studies that successfully used transcriptomic based data to gene discover and annotation (Fergunson *et al.*, 2010; Wei *et al.*, 2011), comparative genomics (Yim *et al.*, 2014), and for investigating genetic variation associated with adaptive traits in population genomics (Berden *et al.*, 2015; Gugger *et al.*, 2016).

The analysis of candidate genes, that are directly sequenced, is also effective in finding the genetic basis of phenotypic traits in non-model species, specially those that can not be manipulated in laboratory or for which few genetic resources are available. This technique uses an *a priori* knowledge about the effects that a particular gene has on a trait in other species. However, there are some disadvantages including bias in the choice of the candidate and long lists of numerous candidate genes that could be, ultimately, unsuccessful (Zhu, 2007). Population structure can also interfere with candidate studies that, normally, require sample size to be very large. Because there are traits that can be influenced by many different genes, identifying the independent contribution of a single gene in the biological pathway may be

difficult as well. Even so, this method has proved to be very useful, for example, on the discovery of the molecular basis behind colour variation, e.g. in European ocellated lizards (Nunes *et al.,* 2011), in the Réunion grey white-eye (Bourgeois *et al.,* 2016), and also in the Mocker swallowtail (Clark *et al.*, 2008).

A more broad and unbiased approach is to use markers encompassing the entire genome. Genome-wide association studies (GWAS) have become very common for investigating the genetic basis of natural variation and ecological important traits in wild populations (Atwell *et al.*, 2010, Parchman *et al.*, 2012, Hetch *et al.*, 2013, Shirasawa *et al.*, 2013, Comeault *et al.*, 2014). By screening the entire genome looking for common genetic variation, GWAS have contributed to the discovery of many candidate loci involved in adaptive traits (Hecht *et al.*, 2013; Takahashi *et al.*, 2013). Contrary to candidate gene approach that are inevitably biased towards a gene already characterised, GWAS can highlight unknown targets of selection in the genome. These studies, normally require the developing of a large number of polymorphic markers.

Combining the different techniques and sequencing technologies has, thus, great potential to the study of the evolutionary mechanisms and the genetic basis underlying adaptation.

1.6 Meadow spittlebugs as models to study colour polymorphisms

1.6.1 The meadow spittlebug *Philaenus spumarius*

During the last decades several insect species have been used as models in genetic and ecological studies and to investigate the mechanisms underling evolution, adaptation and speciation. The meadow spittlebug *P. spumarius* (Linnaeus, 1758) (Hemiptera, Aphrophoridae) is a small insect whose body length ranges from 5.3 and 6.9 mm (Fig. 1). This species is hemimetabolous emerging from eggs into nymphs and from those into adults. It is widely distributed across the Holartic region, being found throughout Europe, in several parts of Asia and in North Africa as well (Halkka & Halkka, 1990). In North America, the species occupies two broad areas along the eastern and western sea boards of the United States of America and Canada, where it became a crop pest (Weaver & King, 1954). The meadow spittlebug also colonised the Azorean islands of São Miguel and Terceira (Quartau *et al.*, 1992; Borges *et al.*, 2005) and was accidentally introduced in Hawaii (Davis & Mitchell, 1946) and New Zealand (Thompson, 1984). Its distribution limits are determined by humidity and temperature, factors which are particularly relevant in the earlier stages of this species' life cycle (Weaver & King,

1954). A northward range shift was detected in some North American populations by Karban & Strauss (2004), probably as response to the rising temperature caused by climate change. The spittlebug is a highly polyphagous insect xylem feeder, occurring in a variety of terrestrial plant communities (Quartau & Borges, 1997; Yurtsever, 2000). Several invertebrates and vertebrates are known to attack eggs, nymphs and the adults (Halkka & Halkka, 1990; Yurtsever, 2000).

Fig. 1: The meadow spittlebug *Philaenus spumarius*. Nymph and adult (phenotype TYP).
Photos by Eduardo Marabuto.

1.6.2 Dorsal colour variation

Philaenus spumarius shows an adult dorsal colour pattern balanced polymorphism widely studied (Halkka & Halkka, 1990). Thirteen main phenotypes are generally referred in the literature (Fig. 2). These can be characterised as non-melanic forms, essentially pale brown with a dark patterning – *populi* (POP), *typicus* (TYP), *trilineatus* (TRI), *vittatus* (VIT) and *praeustus* (PRA) – and melanic forms, dark brown with pale marks in various combinations – *marginellus* (MAR), *flavicollis* (FLA), *gibbus* (GIB), *leucocephalus* (LCE), *lateralis* (LAT), *quadrimaculatus* (QUA), *albomaculatus* (ALB) and *leucopthalmus* (LOP) (Halkka & Halkka 1990; Stewart & Lees 1996; Quartau & Borges 1997; Yurtsever, 2000). Three other rare phenotypes were also described: *ustulata* (UST), *hexamaculata* (HEX) and *marginellus/flavicollis* (MAR/FLA) (Yurtsever, 2000).

Crossing experiments have revealed the Mendelian inheritance of this trait that is mainly under the control of an autosomal locus (*p*) with seven alleles. These alleles have complex dominance and co-dominance relationships and are probably regulated by other loci (Halkka *et al.,* 1973; Stewart & Lees, 1988). The allele p^T controls the phenotypes TRI + VIT (a variation of TRI) while TYP and POP are controlled by p^t. The melanic phenotype MAR is controlled by the allele p^M and the allele p^L controls the phenotype LAT. The allele p^C controls the phenotypes FLA + GIB + LCE, forming the "C" group. The allele p^O controls the

phenotypes QUA + ALB + LOP corresponding to the "O" group. The expression of FLA is also controlled by another allele, p^F. The genetic basis of PRA is unknown. The dominance hierarchy differs in both sexes in Finnish populations (Halkka *et al.,* 1973). The allele p^T was found to be top dominant in both sexes but in females the allele p^t was the bottom recessive allele while in males this allele was dominant over the melanic alleles. On the other hand, in British populations, Stewart & Lees (1988) found no difference in the dominance pattern between sexes. In both sexes, the allele p^T is top dominant and the allele p^t is the bottom recessive allele. However, they have observed a dominance of the allele p^M over the allele p^F in females and the opposite in males. Figure 3 summarises the dominance hierarchy among alleles according to Stewart & Lees (1988). Halkka & Lallukka (1969) suggested that the colour genes may be linked to genes involved in response to the physical environment through epistatic interactions, constituting a *supergene*, and selection may not be directly related to colour.

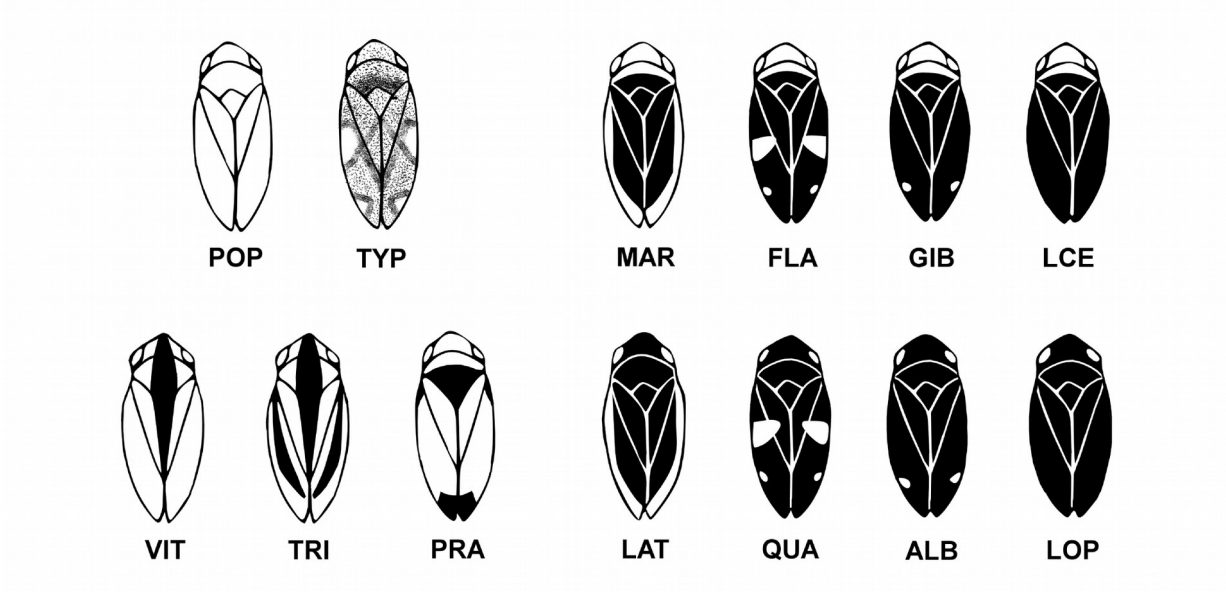


Fig. 2: The thirteen commonly referred dorsal colour phenotypes of *Philaenus spumarius*. Full names of the phenotypes are detailed in the text. Adapted from Silva *et al.* (2015).

The occurrence and frequency of the dorsal colour phenotypes differ among natural populations. In the majority of populations TYP is the most frequent morph. Melanic phenotypes are usually expressed in females only, and in lower frequencies. Besides, in several populations, some melanics tend to be rare or absent (Halkka, 1964; Whittaker, 1972; Thompson & Halkka, 1973; Honěk, 1984; Boucelham *et al.*, 1988; Quartau & Borges, 1997; Zeybekoglu *et al.*, 2004; Yurtserver *et al.*, 2001, 2010). This pattern of frequency distribution of non-melanics and melanics morphs normally exhibits close resemblance in many

populations (Halkka & Halkka, 1990), with only a few local populations sharply deviating from this pattern (e.g. several urban areas of southern Great Britain; see Lees & Dent, 1983; Stewart & Lees, 1988; Stewart & Lees, 1996). Halkka *et al.* (1975a) demonstrated that phenotype frequencies remain stable for long periods of time and that there is a rapid re-establishment of previous frequencies after transfer experiments as result of balancing selection. The selective pressures behind the maintenance of this balanced polymorphism are somehow puzzling but habitat composition, climatic conditions (including altitudinal and latitudinal gradients), industrial melanism and predation are pointed as some of the selective forces (revised in Quartau & Borges 1997; Yurtsever, 2000). Apostatic/visual selection, a form of negative frequency-dependent selection, has been suggested to be involved in the selection regime of the colour polymorphic *P. spumarius* populations. However, in some European populations the colour polymorphism persist despite predation (Halkka & Kohila, 1976; Harper & Whittaker, 1976), indicating that other types of selective pressures than mere visual selection are contributing to this polymorphism. Geographic/clinal variation in the frequencies of the colour phenotypes has been reported in several studies. The southern populations tend to have a higher number of colour morphs than the northern ones (Halkka, 1987**)**. A northward increase in the frequencies of some melanic forms were observed in Finnish populations, in the eastern central Europe (Halkka, 1964; Halkka *et al.*, 1975b), in Asia (Whittaker, 1972) and also in North America (Thompson, 1988), suggesting that thermal melanism may play a role in the maintenance of this polymorphism. Curiously, for MAR (melanic) the frequency vary in the opposite direction from north to south (Halkka, 1987; Thompson, 1988). The latitude was also associated with colour frequencies in *P. spumarius* populations of eastern United States mostly due to climatic factors (Boucelham *et al.*, 1988). Positive correlation between altitude and colour morphs frequencies of melanics in eastern European populations was also found (Halkka *et al.*, 1980). A negative association (probably a consequence of thermal selection) between the frequency of TRI and the increasing altitude and latitude was observed for this phenotype in Scotland (Berry & Willmer, 1986), in Scandinavia (Halkka *et al.*, 1974; Boucelham *et al.*, 1987) and in North America (Boucelham *et al.*, 1988), and is probably related to the highest radiation reflectance and lower temperature excess of TRI. In Scandinavian populations, a west-east cline was also reported for TYP probably as a result of increasing humidity from west to east (Halkka *et al.*, 1974). In some areas, phenotypic frequencies may be associated with the type of habitat more specifically with the type of vegetation due to multi-niche selection (Halkka, 1987, Halkka *et al.*, 2001). Quartau & Borges, 1997 observed differences in the number of melanics in Portuguese populations related to

habitat composition. However, in Turkish populations no association was found between phenotypes and habitat (Yurtsever, 2001). Industrial melanism most likely resulting from high pollution levels is another factor reported to be responsible for the variation in morph frequencies in some British populations of England and Wales (Lees & Dent, 1983; Lees & Stewart, 1987; Stewart & Lees, 1996). High levels of melanic phenotypes were observed for both sexes in spittlebug populations closer to the industrial areas and in urban regions where the effect of pollution is higher. On the other hand, no correlation between pollution and melanism was found in the industrial areas of Chicago (USA) (Thompson & Halkka, 1973) and former Czechoslovakia (Honěk, 1984**)**. Apart from selection, the founder effect and genetic drift were also associated with variation of the phenotypic frequencies in this species. In populations of New Zealand (Thompson, 1984) and Azores (São Miguel) (Quartau *et al.*, 1992), for example, only three morphs (POP, TYP and LOP) were found and this can be explained by the loss of genetic variability due to founder effect and drift resulting from the recent colonisation of the islands. This effect was also detected in some islands in the Golf of Finland (Halkka & Halkka, 1990). In this case, the effects of stochastic processes such as drift and gene flow seem to be overcome by the effect of natural selection (Halkka *et al.*, 2001).

All this points to the idea that the maintenance of this colour polymorphism in *Philaenus spumarius* is probably the result of several selective pressures acting together. However, its adaptive significance is still poorly understood.

1.6.3 Evolutionary history of the species

The evolutionary history of *P. spumarius* has been investigated and recent studies on its genetic diversity, using mitochondrial and nuclear genes, indicate a phylogeographic structure across Europe with two main mitochondrial lineages, a north-eastern lineage ranging from eastern Asia to central and northern Europe and a south-western lineage distributed in western and southern Europe and also in the Middle East (Rodrigues, 2010; Seabra *et al.*, 2010; Maryańska-Nadachowska *et al.*, 2012a). The nuclear gene revealed the existence of three main lineages: the northeastern/Eurasiatic lineage, the southeastern/east Mediterranean – Caucasus lineage and southwestern/Iberian lineage (Maryańska-Nadachowska *et al.*, 2012a). According to these lineages two main routes of post-glacial colonisation of higher latitudes in Europe were suggested for the species, a western colonisation from the Iberian and the Italian peninsulas to the United Kingdom and an eastern colonisation from the Middle East to Finland. A recent work on Asian populations of Turkey and Iran found high levels of genetic diversity

within these populations, the majority of the populations belonging to the south-western clade with only a single population from north-eastern Turkey found to be a highly divergent group within the north-eastern clade (Maryańska-Nadachowska *et al.*, 2015). The same study suggested south-western Asia may be the region of origin of the *P. spumarius*, or at least one of the main refugia for the species during the Pleistocene glaciations, from where it spread across almost the whole Palaearctic region. Some insights were also given about North American populations which probably originated from western European populations (Rodrigues, 2010; Seabra *et al*., 2010; Maryańska-Nadachowska *et al.*, 2012a). A secondary contact zone along the Carpathians mountains was also investigated and the results showed the existence of hybrids indicating that individuals belonging to the main mitochondrial and nuclear lineages meet in the region and interbreed (Lis *et al.*, 2014).

These genetic studies have contributed to our understanding of *P. spumarius* current biogeographic pattern and colonisation routes in Europe. However, the current studies lack estimates of time for the main evolutionary events and are mainly focused in European populations. As a widely distributed species, to extend these studies to populations out of Eurasia is also important. Furthermore, biogeographic studies are also essential to comprehend the process of adaptation in *P. spumarius*, particularly the maintenance of its dorsal colour polymorphism. This is because, by comparing the pattern of neutral genes with the pattern of adaptive genes, the non-selective effects of evolutionary history and/or demography can be distinguished from the effects of natural selection in natural populations.

1.6.4 *Philaenus* species

As observed in *P. spumarius*, the other *Philaenus* species exhibit identical variation in dorsal colour/pattern (Drosopoulos, 2003). Halkka & Halkka (1990) suggested that this variation may be an old polymorphism maintained through the speciation process, probably due to balancing selection, known for keeping genetic variability within species (Reininga *et al.*, 2009). Until few years ago, only three species were recognised by taxonomists: *Philaenus spumarius* (Linnaeus, 1758), *Philaenus tesselatus* (Melichar, 1889) and *Philaenus signatus* (Melichar, 1896). The taxonomic status of *P. tesselatus* has been suffering considerable revision since its description, having been considered a subspecies of *P. spumarius,* geographically circumscribed to southern Iberia and Northern Africa (Halkka & Lallukka, 1969). More recent studies, based on differences in male genitalia suggested *P. tesselatus* as a distinct species of *P. spumarius* (Drosopolous & Quartau, 2002). However, despite the reported

morphological differences, recent genetic studies showed no differentiation between both taxa (Maryańska-Nadachowska *et al.,* 2010, 2012a), and its taxonomic position remains unsolved.

Meanwhile, more five species were described in the Mediterranean region: *Philaenus loukasi* present in Greece (Drosopoulos & Asche, 1991), *Philaenus arslani* in Lebanon (Abdul-Nour & Lahoud, 1996), *Philaenus maghresignus* distributed through southern Iberia and North Africa (Drosopoulos & Remane, 2000), *Philaenus italosignus* distributed through southern Italy and Sicily (Drosopoulos & Remane, 2000) and *Philaenus tarifa* present in southern Iberia (Remane & Drosopoulos, 2001). In 2013, two additional species were described in Iran: *Philaenus elbursianus* (Tishechkin, 2013) and *Philaenus iranicus* (Tishechkin, 2013). These species are currently separated into three groups based on male genitalia morphology:

- The *spumarius* group: *P. spumarius*, *P. tesselatus*, *P. loukasi* and *P. arslani*;
- The *signatus* group: *P. signatus*, *P. maghresignus*, *P. italosignus*, *P. tarifa* and *P. elbursianus*;
- The group comprising the subgenus *Gyrurus*: *P. iranicus*.

Philaenus spumarius, the most common and widely distributed species within the genus occurs in sympatry with the other species. On the other hand, the remaining species, are allopatric or parapatric with each other. Ecological data showed that *P. spumarius* can be found in a variety of host plants (Halkka & Halkka, 1990; Quartau & Borges, 1997; Yurtsever, 2000) while the other Mediterranean species are mostly specialists regarding food and oviposition plants (Drosopoulos, 2003). Cytogenetic analyses in Mediterranean species found differences in the number of chromosomes and in the type of sex determination, with four groups of species differing in karyotype (Maryańska- Nadachowska *et al*., 2008, 2012b, 2013). These findings seem to be in agreement with two of three main groups proposed based on male genitalia morphology (Drosopoulos & Remane, 2000), and also with host plant preferences division (Drosopoulos, 2003). However, the two species of Iran were not part of the analyses. Nothing is known about their karyotype, and food/oviposition habits. Genetic studies, that did not include the Iranian species, support the closer relationship between *P. spumarius* and *P. tesselatus* and also between *P. arslani* and *P. loukasi* but the relationship among the other Mediterranean species is not clear (Maryańska-Nadachowska *et al*., 2010; Rodrigues, 2010; Seabra *et al.*, 2010).

The spittlebugs group, due to its dorsal colour polymorphism, constitutes a potential good model system to test hypothesis about adaptation, speciation and maintenance of polymorphisms in natural populations. For this study in particular, the meadow spittlebug *P. spumarius* was selected because of its ubiquitous distribution, its genetic variation and mainly

because of its widely studied dorsal colour variation (Halkka & Halkka, 1990).

1.7 Objectives and thesis structure

The main purpose of this thesis is to address the genetic basis underlying the adaptive variation in natural populations. Colour polymorphisms are good systems in which adaptation can be investigated.

The study system is the meadow spittlebug *P. spumarius,* an insect species for which several questions related to its dorsal colour/pattern balanced polymorphism and to its evolutionary history remain to be exploited and investigated. Why is this polymorphism maintained in natural populations? What is its adaptive significance? What is the genetic basis of this colour pattern variation? Which gene(s) or genomic regions are involved? Are colour genes described in other insect species implicated?

This thesis will involve (i) analysing the evolutionary pattern of *P. spumarius*; (ii) exploiting the adaptive significance of its dorsal colour polymorphism; and, (iii) testing genotype-phenotype associations aiming to identify a genomic region or several genomic regions linked to this species' dorsal colour variation.

The evolutionary pattern will be analysed using phylogeographic methods applied to the sequence variation in two mitochondrial and one nuclear markers. The adaptive significance will be exploited trough a laboratory experimental procedure to understand differential survival and reproductive success of three of the colour phenotypes. Finally, the identification of a possible candidate genomic region or regions linked to colour variation will be investigated by using two main approaches: a genome-wide association approach using Single Nucleotide Polymorphisms (SNPs) obtained through Restriction Associated DNA (RAD) sequencing on individuals of the same three colour phenotypes used in captivity experiments; and a candidate gene approach to investigate genes involved in melanin-based colouration in other insect species, for which sequence variation across dorsal colour phenotypes in *Philaenus spumarius*, and in other congeneric species, will be analysed.

The specific objectives of this work are:

1 – To study the evolutionary history of *Philaenus spumarius* by providing, for the first time, time estimates of the main demographic and evolutionary events occurred in the Mediterranean peninsulas, and by analysing colonisation patterns out of Eurasia;

2 – To explore the adaptive significance of the colour variation of *P. spumarius* by conducting an experiment in captivity under semi-natural conditions and obtaining data on the fitness of three dorsal colour phenotypes: *typicus* (the most common, non-melanic recessive phenotype), *trilineatus* (the non-melanic dominant phenotype with the highest radiation reflectance and with thermal properties different from the melanic and TYP phenotypes (Berry & Willmer, 1986)), and *marginellus* (the most common melanic phenotype found in the population used for the study);

3 – To test for associations between SNPs obtained with RAD sequencing and the same three dorsal colour phenotypes (*typicus*, *trilineatus* and *marginellus*) used in the captivity study;

4 – To sequence and assemble a first draft genome of *P. spumarius*, an important genomic resource to help in the characterisation of the genomic regions found to be associated with colour variation;

5 – To analyse the sequence variation of *P. spumarius* in a candidate gene involved in melanin synthesis pathway in other insects, the *yellow* gene, by looking for possible associations between nucleotide/amino acid changes in *yellow* sequences and the dorsal colour morphs, *typicus*, *trilineatus* and *marginellus.* And also to investigate *yellow* phylogenetic pattern in other species of the genus.

The present thesis is organised in six main chapters. In Chapter 1, a general overview of the thematic of the thesis is given (Introduction). The phylogeographic analysis results (objective 1) are provided in the Chapter 2 (Rodrigues *et al.*, Plos One). Chapter 3 (Silva & Rodrigues *et al.*, Ecological Entomology) presents the experimental approach for investigating fitness differences between colour types (objective 2). Chapter 4 (Rodrigues *et al.*, submitted) provides the association study using RAD sequencing loci (objective 3), as well as the development of genomic and transcriptomic resources for *Philaenus spumarius* (objective 4). The study of the candidate gene *yellow* is presented in Chapter 5, including the association with colour types and the phylogenetic analysis of this gene (objective 5). Finally, the findings of the thesis and their future implications are discussed in Chapter 6 as well as final remarks about the work.

1.8 References

Abdul-Nour H, Lahoud L (1996) Révision du genre *Philaenus* Stål, 1964 au Liban avec la description d'une nouvelle espèce: *P. arslani* n. sp. (Homoptera, Cercopidae). *Nouv. Rev. Entomol.*, **12**, 297–303.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. *Genetics*, **188**, 799–808.

Anderson JT, Inouye DW, McKinney AM, Colautti RI, Mitchell-Olds T (2012) Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3843–3852.

Atwell S, Huang YS, Vilhjálmsson BJ *et al.* (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, **465**, 627–631.

Ayala FJ, Campbell CA (1974) Frequency-Dependent Selection. *Annual Review of Ecology and Systematics*, **5**, 115–138.

Baiao PC, Schreiber E, Parker PG (2007) The Genetic Basis of the Plumage Polymorphism in Red-Footed Boobies (*Sula sula*): a Melanocortin-1 Receptor (MC1R) Analysis. *Journal of Heredity*, **98**, 287–292.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, **3**, e3376.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

Baxter SW, Davey JW, Johnston JS *et al.* (2011) Linkage Mapping and Comparative Genomics Using Next-Generation RAD Sequencing of a Non-Model Organism (PK Ingvarsson, Ed,). *PLoS ONE*, **6**, e19315.

Baxter S, McMillan O, Chamberlain N, Ffrench-Constant R, Jiggins C (2009) Prospects for Locating Adaptive Genes in Lepidopteran Genomes. In: *Molecular Biology and Genetics of the Lepidoptera* (eds Goldsmith MR, Marec F), p. 105. CRC Press.

Beldade P, Brakefield PM, Long AD (2002) Contribution of Distal-less to quantitative variation in butterfly eyespots. *Nature*, **415**, 315–318.

Berdan EL, Mazzoni CJ, Waurick I, Roehr JT, Mayer F (2015) A population genomic scan in *Chorthippus grasshoppers* unveils previously unknown phenotypic divergence. *Molecular Ecology*, **24**, 3918–3930.

Berry AJ, Willmer PG (1986) Temperature and the colour polymorphism of *Philaenus spumarius* (Homoptera: Aphrophoridae). *Ecological Entomology*, **11**, 251–259.

Bishop JA, Cook LM, Muggleton J (1978) Polymorphisms for Melanism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **281**, 489–515.

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

Bond AB (2007) The Evolution of Color Polymorphism: Crypticity, Searching Images, and Apostatic Selection. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 489–514.

Borges PAV, Cunha R, Gabriel R *et al.* (2005) A list of terrestrial fauna (Mollusca and Arthropoda) and flora (Bryophyta, Pteridophyta and Spermatophyta) from the Azores. *Direcção Regional do Ambiente and Universidade dos Açores, Horta, Angra do Heroísmo and Ponta Delgada*, 317pp.

Boucelham M, Hokkainen H., Raatikainen M (1988) Polymorphism of *Philaenus spumarius* (L.) (Homoptera, Cercopidae) in different latitudes, altitudes and habitats in the U.S.A. *Annals Entomologici Fennici*, **54**, 49–54.

Boucelham M, Raatikainen M (1987) The effect of altitude, habitat and geographical location on the polymorphism of *Philaenus spumarius* (L.) (Hom. Cercopidae). *Annales Entomologici Fennici*, **53**, 80–86.

Bourgeois YXC, Bertrand JAM, Delahaie B *et al.* (2016) Candidate Gene Analysis Suggests Untapped Genetic Complexity in Melanin-Based Pigmentation in Birds. *Journal of Heredity*, **107**, 327–335.

Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176–178.

Brown KS, Benson WW (1974) Adaptive Polymorphism Associated with Multiple Mullerian Mimicry in *Heliconius numata* (Lepid. Nymph.). *Biotropica*, **6**, 205.

Brunetti CR, Selegue JE, Monteiro A *et al.* (2001) The generation and diversification of butterfly eyespot color patterns. *Current Biology*, **11**, 1578–1585.

Burke MK (2012) How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proceedings of the Royal Society B: Biological Sciences*.

Caillaud MC, Losey JE (2010) Genetics of Color Polymorphism in the Pea Aphid, *Acyrthosiphon pisum*. *Journal of Insect Science*, **10**, 1–13.

Chan YF, Marks ME, Jones FC *et al.* (2010) Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science*, **327**, 302–305.

Charmantier A, McCleery RH, Cole LR *et al.* (2008) Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. *Science*, **320**, 800–803.

Chevin LM, Lande R, Mace GM (2010) Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. *PLoS Biology*, **8**.

Clark R, Brown SM, Collins SC *et al.* (2008) Colour pattern specification in the Mocker swallowtail *Papilio dardanus*: the transcription factor invected is a candidate for the mimicry locus H. *Proceedings. Biological sciences / The Royal Society*, **275**, 1181–1188.

Clarke CA, Sheppard PM (1972) The Genetics of the Mimetic Butterfly *Papilio polytes* L. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **263**, 431–458.

Comeault AA, Flaxman SM, Riesch R *et al.* (2015) Selection on a Genetic Polymorphism Counteracts Ecological Speciation in a Stick Insect. *Current Biology*, **25**, 1975–1981.

Comeault AA, Soria-Carrasco V, Gompert Z *et al.* (2014) Genome-Wide Association Mapping of Phenotypic Traits Subject to a Range of Intensities of Natural Selection in *Timema cristinae* *. *The American Naturalist*, **183**, 711–727.

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature reviews. Genetics*, **12**, 499–510.

Davis AL, Mitchell CJ (1946) Host Records of *Philaenus spumarius* (Linn.) at Kilauea, Hawaii National Park (Homoptera: Cercopidae). *Proc. Haw. Ent Soc.*, **12**, 515–516.

Dobzhansky T (1970) *Genetics of the Evolutionary Process*. Columbia University Press: New York, New York, USA.

Drosopoulos S (2003) New data on the nature and origin of colour polymorphism in the

spittlebug genus Philaenus (Hemiptera : Aphorophoridae). *Annales De La Societe Entomologique De France*, **39**, 31–42.

Drosopoulos S, Asche M (1991) Biosystematic studies on the spittlebug genus *Philaenus* with the description of a new species. *Zoological Journal of the Linnean Society*, **101**, 169–177.

Drosopoulos S, Quartau JA (2002) The spittle bug *Philaenus tesselatus* Melichar, 1899 (Hemiptera, Auchenorrhyncha, Cercopidae) is a distinct species. *Zootaxa*, **68**, 1–8.

Drosopoulos S, Remane R (2000) Biogeographic studies on the spittlebug *Philaenus signatus* melichar, 1986 species group (hemiptera: aphrophoridae) with the description of two new allopatric species. *Ann. Soc. Entomol. Fr.*, **36**, 269–277.

Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature reviews. Genetics*, **4**, 457–69.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.

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

Fisher RA (1930) *The genetical theory of natural selection*. Oxford University Press.

Forsman A, Ahnesjö J, Caesar S, Karlsson M (2008) A model of ecological and evolutionary consequences of color polymorphism. *Ecology*, **89**, 34–40.

Futahashi R, Fujiwara H (2005) Melanin-synthesis enzymes coregulate stage-specific larval cuticular markings in the swallowtail butterfly, *Papilio xuthus*. *Development Genes and Evolution*, **215**, 519–529.

Futahashi R, Sato J, Meng Y *et al.* (2008) yellow and ebony are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*. *Genetics*, **180**, 1995–2005.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature*, **433**, 481–487.

Gray SM, McKinnon JS (2007) Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution*, **22**, 71–79.

Gugger PF, Cokus SJ, Sork VL (2016) Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genetics & Genomes*, **12**, 15.

Halkka O (1964) Geographical, spatial and temporal variability in the balanced polymorphism of *Philænus spumarius*. *Heredity*, **19**, 383–401.

Halkka O (1987) Colour morphs of *Philaenus spumarius* are ecomorphs and are dissimilary distributed in different of the habitat. *Proc. 6-th Auchen. Meeting. Turin, Itali*, 347–358.

Halkka O, Halkka L (1990) Population-Genetics Of the Polymorphic Meadow Spittlebug, *Philaenus spumarius* (L). *Evolutionary Biology*, **24**, 149–191.

Halkka O, Halkka L, Raatikainen M (1975a) Transfer of individuals as a means of investigating natural selection in operation. *Hereditas*, **80**, 27–34.

Halkka O, Halkka L, Raatikainen M, Hovinen R (1973) The genetic basis of balanced polymorphism in *Philaenus* (Homoptera). *Hereditas*, **74**, 69–80.

Halkka O, Halkka L, Roukka K (2001) Selection often overrides the effects of random processes in island populations of *Philaenus spumarius* (Homoptera). *Biological Journal of the Linnean Society*, **74**, 571–580.

Halkka O, Kohila T (1976) Persistence of visual polymorphism, despite a low rate of predation, in *Philaenus spumarius* (L.) (Homoptera, Aphrophoridae). *Annales Zoologici Fennici*, **13**, 185–188.

Halkka O, Raatikainen M, Halkka L (1974) Radial and peripheral clines in northern polymorphic populations of *Philaenus spumarius*. *Hereditas*, **78**, 85–96.

Halkka O, Raatikainen M, Vilbaste J (1975b) Clines in the colour polymorphism of *Philaenus spumarius* in Eastern Central Europe. *Heredity*, **35**, 303–309.

Halkka O, Vilbaste J, Raatikainen M (1980) Colour gene allele frequencies correlated with altitude of habitat in *Philaenus* populations. *Hereditas*, **92**, 243–246.

Hallka O, Lallukka R (1969) The origin of balanced polymorphism in the spittlebugs (Philaenus, Homoptera). *Annales Zoologici Fennici*, **6**, 431–434.

Harper G, Whittaker JB (1976) The Role of Natural Enemies in the Colour Polymorphism of *Philaenus spumarius* (L.). *The Journal of Animal Ecology*, **45**, 91.

Hecht BC, Campbell NR, Holecek DE, Narum SR (2013) Genome-wide association reveals

genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, **22**, 3061–3076.

Hedrick PW (2013) Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606–4618.

Hellens RP, Moreau C, Lin-Wang K *et al.* (2010) Identification of Mendel's White Flower Character (GP Copenhaver, Ed,). *PLoS ONE*, **5**, e13230.

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

Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution; international journal of organic evolution*, **58**, 1329–1341.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science*, **313**, 101–104.

Hoekstra HE, Krenz JG, Nachman MW (2005) Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations. *Heredity*, **94**, 217–228.

Hof AE van't, Campagne P, Rigden DJ *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, **534**, 102–105.

Hoffman EA, Blouin MS (2000) A review of colour and pattern polymorphisms in anurans. *Biological Journal of the Linnean Society*, **70**, 633–665.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

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

Hohenlohe PA, Day MD, Amish SJ *et al.* (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002–3013.

Honek A (1984) Melanism in populations of *Philaenus spumarius* (Homoptera, Aphrophoridae) in Czechoslovakia. *Věstník Československé Společnosti Zoologické*, 241–247.

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

Huxley J (1955) Morphism and evolution. *Heredity*, **9**, 1–52.

Jiggins CD, McMillan WO (1997) The genetic basis of an adaptive radiation: warning colour in two *Heliconius* species. *Proceedings of the Royal Society B: Biological Sciences*, **264**, 1167–1175.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Jones JS, Leith BH, Rawlings P (1977) Polymorphism in *Cepaea*: A Problem with Too Many Solutions? *Annual Review of Ecology and Systematics*, **8**, 109–143.

Joron M, Frezal L, Jones RT *et al.* (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, **477**, 203–6.

Joron M, Jiggins CD, Papanicolaou A, McMillan WO (2006a) *Heliconius* wing patterns: an evo-devo model for understanding phenotypic diversity. *Heredity*, **97**, 157–167.

Joron M, Papa R, Beltrán M *et al.* (2006b) A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies (MAF Noor, Ed,). *PLoS Biology*, **4**, e303.

Joron M, Wynne IR, Lamas G, Mallet J (1999) Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. *Evolutionary Ecology*, **13**, 721–754.

Karban R, Strauss SY (2004) Physiological tolerance, climate change, and a northward range shift in the spittlebug, *Philaenus spumarius*. *Ecological Entomology*, **29**, 251–254.

Kunte K, Zhang W, Tenger-Trolander A *et al.* (2014) doublesex is a mimicry supergene. *Nature*, **507**, 229–232.

Kusche H, Meyer A (2014) One cost of being gold: selective predation and implications for the maintenance of the Midas cichlid colour polymorphism (Perciformes: Cichlidae). *Biological Journal of the Linnean Society*, **111**, 350–358.

Lees DR, Dent CS (1983) Industrial Melanism in the Spittlebug *Philaenus spumarius* (L) (Homoptera, Aphrophoridae). *Biological Journal Of the Linnean Society*, **19**, 115–129.

Lees DR, Stewart AJA (1987) Localized industrial melanism in the spittlebug *Philaneus spumarius* (L) (Homoptera: Aphrophoridae) in Cardiff docks, south Wales. *Biol. J. Linn. Soc.*, **31**, 333–345.

Lis A, Maryańska-Nadachowska A, Lachowska-Cierlik D, Kajtoch U (2014) The Secondary Contact Zone of Phylogenetic Lineages of the *Philaenus spumarius* (Hemiptera: Aphrophoridae): An Example of Incomplete Allopatric Speciation. *Journal of Insect Science*, **14**, 227–227.

Lozier JD, Jackson JM, Dillon ME, Strange JP (2016) Population genomics of divergence among extreme and intermediate color forms in a polymorphic insect. *Ecology and Evolution*, **6**, 1075–1091.

Martin A, McCulloch KJ, Patel NH *et al.* (2014) Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. *EvoDevo*, **5**, 7.

Maryańska-Nadachowska A, Drosopoulos S, Lachowska D, Kajtoch Ł, Kuznetsova VG (2010) Molecular phylogeny of the Mediterranean species of *Philaenus* (Hemiptera: Auchenorrhyncha: Aphrophoridae) using mitochondrial and nuclear DNA sequences. *Systematic Entomology*, **35**, 318–328.

Maryańska-Nadachowska A, Kajtoch Ł, Lachowska D (2012) Genetic diversity of *Philaenus spumarius* and *P. tesselatus* (Hemiptera, Aphrophoridae): implications for evolution and taxonomy. *Systematic Entomology*, **37**, 55–64.

Maryańska-Nadachowska A, Kuznetsova VG, Drosopoulos S, Lachowska-Cierlik D (2008) A chromosomal analysis of eight Mediterranean species of *Philaenus*. *Bulletin of Insectology*, **61**, 133–134.

Maryańska-Nadachowska A, Kuznetsova VG, Karamysheva T V. (2013) Chromosomal location of rDNA clusters and TTAGG telomeric repeats in eight species of the spittlebug genus *Philaenus* (Hemiptera: Auchenorrhyncha: Aphrophoridae). *European Journal of Entomology*, **110**, 411–418.

Maryańska-Nadachowska A, Kuznetsova VG, Lachowska D, Drosopoulos S (2012) Mediterranean species of the spittlebug genus *Philaenus*: modes of chromosome evolution. *Journal of Insect Science*, **12**, 54.

Maryańska-Nadachowskaa A, Sanaieb E, Kajtocha Ł (2015) High genetic diversity in southwest Asian populations of *Philaenus spumarius* (Hemiptera: Auchenorrhyncha). *Zoology in the Middle East*, **61**, 264–272.

McGregor AP, Orgogozo V, Delon I *et al.* (2007) Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature*, **448**, 587–590.

Michie LJ, Mallard F, Majerus MEN, Jiggins FM (2010) Melanic through nature or nurture: Genetic polymorphism and phenotypic plasticity in *Harmonia axyridis*. *Journal of Evolutionary Biology*, **23**, 1699–1707.

Mills MG, Patterson LB (2009) Not just black and white: Pigment pattern development and evolution in vertebrates. *Seminars in Cell & Developmental Biology*, **20**, 72–81.

Munday PL, Eyre PJ, Jones GP (2003) Ecological mechanisms for coexistence of colour polymorphism in a coral-reef fish: An experimental evaluation. *Oecologia*, **137**, 519–526.

Nadeau NJ, Pardo-Diaz C, Whibley A *et al.* (2016) The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, **534**, 106–110.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Nijhout HF (2003) Polymorphic mimicry in *Papilio dardanus*: mosaic dominance, big effects, and origins. *Evolution and Development*, **5**, 579–592.

Nunes VL, Miraldo A, Beaumont MA, Butlin RK, Paulo OS (2011) Association of Mc1r variants with ecologically relevant phenotypes in the European ocellated lizard, *Lacerta lepida*. *Journal of evolutionary biology*, **24**, 2289–98.

Oliver JC, Tong X-L, Gall LF, Piel WH, Monteiro A (2012) A Single Origin for Nymphalid Butterfly Eyespots Followed by Widespread Loss of Associated Gene Expression. *PLoS Genetics*, **8**, e1002893.

Olson-Manning CF, Wagner MR, Mitchell-Olds T (2012) Adaptive evolution: evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, **13**, 867–877.

Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–127.

Orr HA, Coyne JA (1992) The Genetics of Adaptation: A Reassessment. *The American Naturalist*, **140**, 725–742.

Oxford GS (1985) A countrywide survey of colour morph frequencies in the spider *Enoplognatha ovata* (Clerck) (Araneae: Theridiidae): evidence for natural selection. *Biological Journal of the Linnean Society*, **24**, 103–142.

Oxford GS (2005) Genetic drift within a protected polymorphism: enigmatic variation in color-morph frequencies in the candy-stripe spider, *Enoplognatha ovata*. *Evolution; international journal of organic evolution*, **59**, 2170–2184.

Oxford GS, Gillespie RG (1996a) Genetics of a colour polymorphism in *Theridion grallator* (Araneae: Theridiidae), the Hawaiian happy-face spider, from Greater Maui. *Heredity*, **76**, 238–248.

Oxford GS, Gillespie RG (1996b) The effects of genetic background on the island-specific control of a colour polymorphism in *Theridion grallator* (Araneae: Theridiidae), the Hawaiian happy-face spider. *Heredity*, **76**, 257–266.

Oxford GS, Gillespie RG (2001) Portraits of Evolution: Studies of Coloration in Hawaiian Spiders. *BioScience*, **51**, 521.

Oxford GS, Shaw MW (1986) Long-term variation in colour-morph frequencies in the spider *Enoplognatha ovata* (Clerck) (Araneae: Theridiidae): natural selection, migration and intermittent drift. *Biological Journal of the Linnean Society*, **27**, 225–249.

Ożgo M (2012) Shell polymorphism in the land-snail *Cepaea nemoralis* (L.) along a West-East transect in continental Europe. *Folia Malacologica*, **20**, 181–253.

Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC genomics*, **11**, 180.

Parchman TL, Gompert Z, Mudge J *et al.* (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

Perez i de Lanuza G, Font E, Carazo P (2013) Color-assortative mating in a color-polymorphic lacertid lizard. *Behavioral Ecology*, **24**, 273–279.

Pool JE, Aquadro CF (2007) The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. *Molecular Ecology*, **16**, 2844–2851.

Protas ME, Hersey C, Kochanek D *et al.* (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature genetics*, **38**, 107–11.

Protas ME, Patel NH (2008) Evolution of Coloration Patterns. *Annual Review of Cell and Developmental Biology*, **24**, 425–446.

Punzalan D, Rodd FH, Hughes KA (2005) Perceptual Processes and the Maintenance of Polymorphism Through Frequency-dependent Predation. *Evolutionary Ecology*, **19**, 303–320.

Quartau JA, Borges PA V (1997) On the colour polymorphism of *Philaenus spumarius* (L.) in Portugal. *Miscel-lánia Zoológica*, **2**, 19–30.

Quartau A, Borges P, André G (1992) *Phileanus spumarius* (Linaeus, 1975) new to the Azores (Homoptera, Auchenorrhyncha, Cercopidae). *Actas do Congresso Iberico de Entomologia, 1*, 129–136.

Radwan J, Babik W (2012) The genomics of adaptation. *Proceedings of the Royal Society B: Biological Sciences*.

Reed RD, Papa R, Martin A *et al.* (2011) optix Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. *Science*, **333**, 1137–1141.

Reininga JM, Nielsen D, Purugganan MD (2009) Functional and geographical differentiation of candidate balanced polymorphisms in *Arabidopsis thaliana*. *Molecular Ecology*, **18**, 2844–2855.

Remane R, Drosopoulos S (2001) *Philaenus tarifa* sp. n.: An additional spittlebug species from southern Spain (Homoptera, Auchenorrhyncha Cercopidae). *Mitteilungen aus dem Museum fur Naturkunde in Berlin - Deutsche Entomologische Zeitschrift*, **48**, 277–279.

Richards PM, Liu MM, Lowe N *et al.* (2013) RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology*, **22**, 3077–3089.

Rodrigues ASB (2010) Estudo dos padrões filogenéticos e filogeográficos em *Philaenus spumarius* (Hemiptera, Aphrophoridae) e espécies próximas.

Rogers WA, Salomone JR, Tacy DJ *et al.* (2013) Recurrent modification of a conserved cis-regulatory element underlies fruit fly pigmentation diversity. *PLoS genetics*, **9**, e1003740.

Rosenblum EB, Hoekstra HE, Nachman MW (2004) Adaptive reptile color variation and the evolution of the Mc1r gene. *Evolution; international journal of organic evolution*, **58**, 1794–808.

Roulin A (2004) The evolution, maintenance and adaptive function of genetic colour

polymorphism in birds. *Biological reviews of the Cambridge Philosophical Society*, **79**, 815–848.

Roulin A, Ducret B, Ravussin PA, Altwegg R (2003) Female colour polymorphism covaries with reproductive strategies in the tawny owl *Strix aluco*. *Journal of Avian Biology*, **34**, 393–401.

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

Schwander T, Libbrecht R, Keller L (2014) Supergenes and complex phenotypes. *Current Biology*, **24**, R288–R294.

Scoville AG, Pfrender ME (2010) Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. *Proceedings of the National Academy of Sciences*, **107**, 4260–4263.

Seabra SG, Pina-Martins F, Marabuto E *et al.* (2010) Molecular phylogeny and DNA barcoding in the meadow-spittlebug *Philaenus spumarius* (Hemiptera, Cercopidae) and its related species. *Molecular Phylogenetics and Evolution*, **56**, 462–467.

Sezgin E, Duvernell DD, Matzkin LM *et al.* (2004) Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. *Genetics*, **168**, 923–931.

Shapiro MD, Marks ME, Peichel CL *et al.* (2006) Corrigendum: Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **439**, 1014–1014.

Shirasawa K, Fukuoka H, Matsunaga H *et al.* (2013) Genome-Wide Association Studies Using Single Nucleotide Polymorphism Markers Developed by Re-Sequencing of the Genomes of Cultivated Tomato. *DNA Research*, **20**, 593–603.

Steiner CC, Weber JN, Hoekstra HE (2007) Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes (MA. Noor, Ed,). *PLoS Biology*, **5**, e219.

Stewart AJA, Lees DR (1988) Genetic control of colour/pattern polymorphism in British populations of the spittlebug *Philaenus spumarius* (L.) (Homoptera: Aphrophoridae). *Biological Journal of the Linnean Society*, **34**, 57–79.

Stewart a. J a., Lees DR (1996) The Colour/Pattern Polymorphism of *Philaenus spumarius* (L.) (Homoptera: Cercopidae) in England and Wales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **351**, 69–89.

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

Sucena E, Stern DL (2000) Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proceedings of the National Academy of Sciences*, **97**, 4530–4534.

Sugumaran M (2009) Complexities of cuticular pigmentation in insects. *Pigment Cell & Melanoma Research*, **22**, 523–525.

Svensson EI, Abbott J, Härdling R (2005) Female Polymorphism, Frequency Dependence, and Rapid Evolutionary Dynamics in Natural Populations. *The American Naturalist*, **165**, 567–576.

Takahashi T, Sota T, Hori M (2013) Genetic basis of male colour dimorphism in a Lake Tanganyika cichlid fish. *Molecular Ecology*, **22**, 3049–3060.

Takeda S, Matsuoka M (2008) Genetic approaches to crop improvement: responding to environmental and population changes. *Nature Reviews Genetics*, **9**, 444–457.

Thompson V (1984) Color polymorphism in the introduced spittlebug *Philaenus spumarius* (Homoptera, Aphrophoridae) in New-Zealand. *New Zealand Entomologist*, **8**, 86–88.

Thompson V (1988) Parallel Colour form Distributions in European and North American Populations of the Spittlebug *Philaenus spumarius* (L.). *Journal of Biogeography*, **15**, 507.

Thompson V, Halkka O (1973) Color polymorphism in some North American *Philaenus spumarius* (Homoptera: Aphrophoridae) populations. *American Midland Naturalist*, **89**, 348–359.

Thompson MJ, Jiggins CD (2014) Supergenes and their role in evolution. *Heredity*, **113**, 1–8.

Tishechkin DY (2013) Two new species of the genus *Philaenus* (Homoptera, Aphrophoridae) from Iran. *Entomological Review*, **93**, 73–76.

True JR, Edwards KA, Yamamoto D, Carroll SB (1999) *Drosophila* wing melanin patterns form by vein-dependent elaboration of enzymatic prepatterns. *Current biology : CB*, **9**, 1382–1391.

True JR, Yeh SD, Hovemann BT *et al.* (2005) *Drosophila* tan encodes a novel hydrolase

required in pigmentation and vision. *PLoS genetics*, **1**.

Tsurui K, Honma A, Nishida T (2010) Camouflage effects of various colour-marking morphs against different microhabitat backgrounds in a polymorphic pygmy grasshopper *Tetrix japonica*. *PLoS ONE*, **5**.

Tuttle EM, Bergland AO, Korody ML *et al.* (2016) Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Current Biology*, 1–7.

Vercken E, Clobert J, Sinervo B (2010) Frequency-dependent reproductive success in female common lizards: a real-life hawk–dove–bully game? *Oecologia*, **162**, 49–58.

Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Vander Wal E, Garant D, Festa-Bianchet M, Pelletier F (2012) Evolutionary rescue in vertebrates: evidence, applications and uncertainty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**, 20120090–20120090.

Wallbank RWR, Baxter SW, Pardo-Diaz C *et al.* (2016) Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. *PLoS biology*, **14**, e1002353.

Walsh SB, Dolden TA, Moores GD *et al.* (2001) Identification and characterization of mutations in housefly (Musca domestica) acetylcholinesterase involved in insecticide resistance. *Biochem J*, **359**, 175–181.

Weaver C, King D (1954) Meadow spittlebug, *Philaenus leucophthalmus* (L.). *Ohio Agricultural Experiment Station. Research bulletin*, **741**, 1–99.

Wei W, Qi X, Wang L *et al.* (2011) Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics*, **12**, 451.

Weill M, Malcolm C, Chandre F *et al.* (2004) The unique mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. *Insect Molecular Biology*, **13**, 1–7.

Whittaker JB (1972) Polymorphism in *Philaenus spumarius* (Homoptera) in the U.S.S.R. *Oikos*, **23**, 366–369.

Wilson K, Cotter SC, Reeson AF, Pell JK (2001) Melanism and disease resistance in insects.

Ecology Letters, **4**, 637–649.

Wittkopp PJ, Beldade P (2009) Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy. *Seminars in Cell and Developmental Biology*, **20**, 65–71.

Wittkopp PJ, Carroll SB, Kopp A (2003) Evolution in black and white: Genetic control of pigment patterns in *Drosophila*. *Trends in Genetics*, **19**, 495–504.

Wittkopp PJ, Kalay G (2011) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, **13**, 59–69.

Wittkopp PJ, Stewart EE, Arnold LL *et al.* (2009) Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science*, **326**, 540–544.

Wittkopp PJ, True JR, Carroll SB (2002a) Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development*, **129**, 1849–58.

Wittkopp PJ, Vaccaro K, Carroll SB (2002b) Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Current Biology*, **12**, 1547–1556.

Yim KM, Brewer MS, Miller CT, Gillespie RG (2014) Comparative transcriptomics of maturity-associated color change in hawaiian spiders. *Journal of Heredity*, **105**, 771–781.

Yurtsever S (2000) On the polymorphic meadow spittlebug, *Philaenus spumarius* (L.) (Homoptera: Cercopidae). *Turkish Journal of Zoology*, **24**, 447–459.

Yurtsever S (2001) Colour/pattern polymorphism of the meadow spittelbug *Philaenus spumarius* (Homoptera, Cercopidae) in Northwest Turkey. *Biologia-Bratislava*, **56**, 497–501.

Yurtsever S, Akdeniz N, Karahalil B (2010) The colour/pattern polymorphism of *Philaenus spumarius* (Hemiptera: Cercopidae) in the north-western Black Sea region of Turkey. *Biologia*, **65**, 309–315.

Zeybekoglu Ü, Yurtsever S, Turgut F (2004) Polymorphism of *Philaenus spumarius* L. (Hemiptera: Cercopidae) in the Samsun (Mid-Black Sea Region) populations of Turkey. *Annales de la Société Entomologique de France (Nouvelle série)*, **40**, 277–283.

Zhu M (2007) Candidate Gene Identification Approach: Progress and Challenges. *International Journal of Biological Sciences*, **3**, 420–427.

Zufall RA, Rausher MD (2003) The genetic basis of a flower color polymorphism in the common morning glory (Ipomoea purpurea). *The Journal of heredity*, **94**, 442–8.

CHAPTER 2

New mitochondrial and nuclear evidences support recent demographic expansion and an atypical phylogeographic pattern in the spittlebug *Philaenus spumarius* (Hemiptera, Aphrophoridae)

Ana S.B. Rodrigues[1*], Sara E. Silva[1], Eduardo Marabuto[1], Diogo N. Silva[1], Mike R. Wilson[2], Vinton Thompson[3], Selçuk Yurtsever[4], Antti Halkka[5], Paulo A.V. Borges[6], José A. Quartau[1], Octávio S. Paulo[1], Sofia G. Seabra[1]

[1] Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal and Platform for Enhancing Ecological Research & Sustainability.

[2] Department of Natural Sciences, National Museum of Wales, Cardiff, United Kingdom.

[3] Metropolitan College of New York, New York, United States of America.

[4] Biology Department, Science Faculty, Trakya University, Edirne, Turkey.

[5] Department of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland.

[6] Azorean Biodiversity Group, Centro de Investigação e Tecnologias Agrárias dos Açores and Platform for Enhancing Ecological Research & Sustainability, Universidade dos Açores, Departamento de Ciências Agrárias, Angra do Heroísmo, Terceira, Portugal.

PlosOne (2014) 9, e98375

OPEN ACCESS Freely available online

PLOS ONE

New Mitochondrial and Nuclear Evidences Support Recent Demographic Expansion and an Atypical Phylogeographic Pattern in the Spittlebug *Philaenus spumarius* (Hemiptera, Aphrophoridae)

Ana S. B. Rodrigues[1]*, Sara E. Silva[1], Eduardo Marabuto[1], Diogo N. Silva[1], Mike R. Wilson[2], Vinton Thompson[3], Selçuk Yurtsever[4], Antti Halkka[5], Paulo A. V. Borges[6], José A. Quartau[1], Octávio S. Paulo[1], Sofia G. Seabra[1]

1 Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, Departamento de Biologia Animal/Platform for Enhancing Ecological Research & Sustainability, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal, **2** Department of Natural Sciences, National Museum of Wales, Cardiff, United Kingdom, **3** Metropolitan College of New York, New York, New York, United States of America, **4** Biology Department, Science Faculty, Trakya University, Edirne, Turkey, **5** Department of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland, **6** Azorean Biodiversity Group, Centro de Investigação e Tecnologias Agrárias dos Açores and Platform for Enhancing Ecological Research & Sustainability, Universidade dos Açores, Departamento de Ciências Agrárias, Angra do Heroísmo, Terceira, Portugal

Abstract

Philaenus spumarius is a widespread insect species in the Holarctic region. Here, by focusing on the mtDNA gene COI but also using the COII and Cyt *b* genes and the nuclear gene EF-1α, we tried to explain how and when its current biogeographic pattern evolved by providing time estimates of the main demographic and evolutionary events and investigating its colonization patterns in and out of Eurasia. Evidence of recent divergence and expansion events at less than 0.5 Ma ago indicate that climate fluctuations in the Mid-Late Pleistocene were important in shaping the current phylogeographic pattern of the species. Data support a first split and differentiation of *P. spumarius* into two main mitochondrial lineages: the "western", in the Mediterranean region and the "eastern", in Anatolia/Caucasus. It also supports a following differentiation of the "western" lineage into two sub-lineages: the "western-Mediterranean", in Iberia and the "eastern-Mediterranean" in the Balkans. The recent pattern seems to result from postglacial range expansion from Iberia and Caucasus/Anatolia, thus not following one of the four common paradigms. Unexpected patterns of recent gene-flow events between Mediterranean peninsulas, a close relationship between Iberia and North Africa, as well as high levels of genetic diversity being maintained in northern Europe were found. The mitochondrial pattern does not exactly match to the nuclear pattern suggesting that the current biogeographic pattern of *P. spumarius* may be the result of both secondary admixture and incomplete lineage sorting. The hypothesis of recent colonization of North America from both western and northern Europe is corroborated by our data and probably resulted from accidental human translocations. A probable British origin for the populations of the Azores and New Zealand was revealed, however, for the Azores the distribution of populations in high altitude native forests is somewhat puzzling and may imply a natural colonization of the archipelago.

Citation: Rodrigues ASB, Silva SE, Marabuto E, Silva DN, Wilson MR, et al. (2014) New Mitochondrial and Nuclear Evidences Support Recent Demographic Expansion and an Atypical Phylogeographic Pattern in the Spittlebug *Philaenus spumarius* (Hemiptera, Aphrophoridae). PLoS ONE 9(6): e98375. doi:10.1371/journal.pone.0098375

Editor: Wolfgang Arthofer, University of Innsbruck, AUSTRIA

Received February 6, 2014; **Accepted** May 1, 2014; **Published** June 3, 2014

Funding: Financial support: FCT, Portugal (Grant SFRH/BPD/26365/2006; Grant SFRH/BD/73879/2010; Project PTDC/BIA-BEC/098783/2008). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: asorodrigues@fc.ul.pt

Introduction

Distribution patterns of animals and plants have faced dramatic changes throughout time and are influenced by ecological requirements and historical factors. In the northern hemisphere, Quaternary long-term glacial (cold) and interglacial (warm) climatic cycles that started about 2.6 million years (Ma) ago [1] have strongly influenced the species distributions and range sizes and, as a consequence, have affected the genetic structure of their populations [2,3]. Evidence from numerous studies suggests that southern European regions of Iberia, Italy and the Balkans and areas near the Caucasus and western Asia, acted as glacial refugia for temperate species during cold periods [4,5,6]. Recent work indicates that temperate refugia might not have been restricted to the three southern peninsulas and that cryptic northern refugia might have existed in central, western, eastern and even northern Europe in the Late Pleistocene [7,8,9]. The relative impact of the post-glacial colonization history and more recent processes such as gene flow and population fluctuations, strongly depend on the dispersal mode and ability of the species [10,11].

Genetic analyses have proven to be useful for a more detailed understanding of post-glacial expansions of several animals and

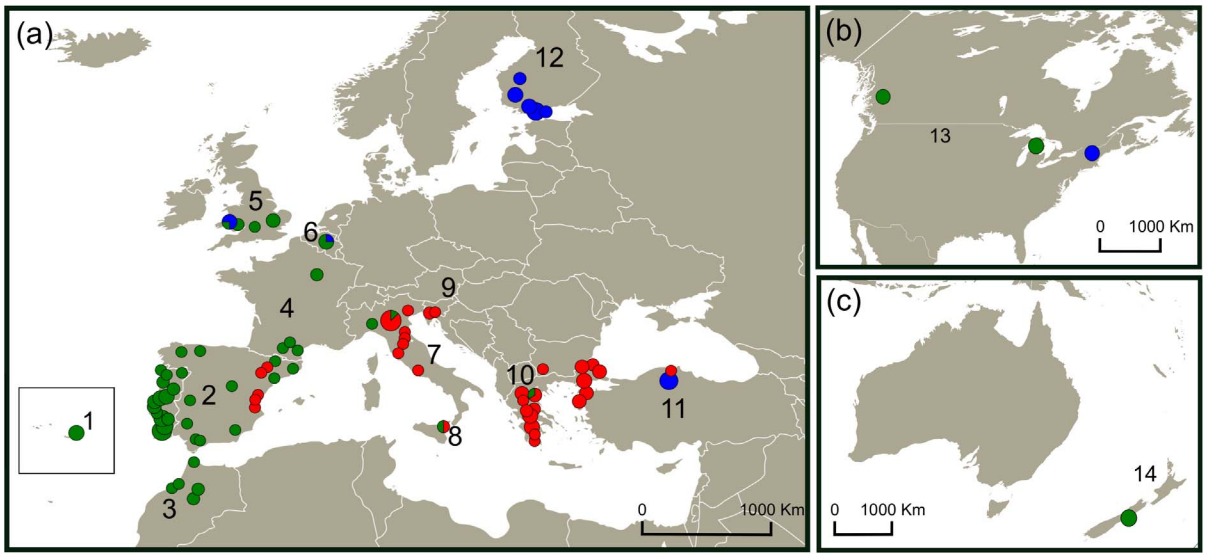


Figure 1. Sampling locations of *Philaenus spumarius* in (a) Europe and Anatolia (b) North America and (c) New Zealand in each geographic region. 1 – Azores; 2 – Iberian Peninsula; 3 – Morocco; 4 – France; 5 – United Kingdom; 6 – Belgium; 7 – Italian Peninsula; 8 – Sicily; 9 – Slovenia; 10 – Balkans (Bulgaria; Greece; European Turkey); 11 – Anatolian Peninsula; 12 – Finland; 13 – North America (Canada and United States of America); 14 – New Zealand. Circle sizes are proportional to the number of individuals. Circles: green – "western-Mediterranean" mtDNA group; red – "eastern-Mediterranean" mtDNA group; blue – "eastern" mtDNA group. Circle sizes are proportional to the number of samples.
doi:10.1371/journal.pone.0098375.g001

plants [3,12]. Mitochondrial DNA (mtDNA), due to its particular characteristics, has been widely used in determining population dynamics and phylogeographic divergence in recent times, such as the Quaternary period [13]. Nevertheless, the signal of deeper history can be obscured by homoplasy or saturation resulting from high mutation rate. On the other hand, reconstructing evolutionary histories using individual genes (gene trees) could lead to misrepresentation of population or species histories because in this case mtDNA, which reflects matrilineal history, might not represent the overall lineage history of the species. Also, if multiple population divergences or speciation events were closely spaced in time, a single gene tree might be 'incorrect' by chance due to the random nature of lineage sorting during the coalescence process [14]. Therefore, the use of multiple types of molecular markers is recommended.

Insects have been widely used as models for animal biogeographical studies (e.g., [15,16,17]). The meadow spittlebug *Philaenus spumarius* (Linnaeus, 1758) (Hemiptera, Aphrophoridae) is a widely investigated species, very suitable for genetic and ecological studies. It is a highly polyphagous insect which can be found in a variety of terrestrial plant communities and habitats, being the most common species within the genus *Philaenus* [18,19]. It is widespread across the Palaearctic region from where it is native [20] having also colonized the Azores [21,22] and has been introduced in the Nearctic region [20] and New Zealand [23]. The meadow spittlebug is very sensitive to humidity and temperature, especially in the earlier stages of its life cycle, which limits its range [24]. A remarkable example was reported for some North American populations where a northward range shift, probably as a result of climatic changes, was detected by [25]. This species shows a well studied dorsal colour polymorphism with eleven main described phenotypes which can be divided in melanic and non-melanic forms [20]. The phenotype frequencies differ among populations, probably due to the effects of natural selection under different habitats, climatic conditions and predation pressure (reviewed in [18,19]). Recent studies on the genetic diversity of *P. spumarius* have given insights on its evolutionary history suggesting two routes of post-glacial colonization of higher latitudes in Europe and indicating a probable western European origin for North American populations [26,27].

In the present study we tried to explain how and when the current biogeographic pattern of *P. spumarius* evolved by (i) providing time estimates of the main demographic and evolutionary events with focus on the populations occurring in the main Mediterranean peninsulas; and, (ii) investigating the colonization patterns out of Eurasia, namely of north-western Africa, North America, and the islands of the Azores and New Zealand.

Material and Methods

Ethics Statement

The field sampling was carried out on private lands with owners' permissions. The studied species, *Philaenus spumarius*, is considered a widespread species across the Palaearctic and the Nearctic regions, being a crop pest in some locations of USA and Canada. It is not an endangered or protected species.

Sampling

A total of 196 specimens of *P. spumarius* were collected or sent by collaborators between 2007 and 2011 from 75 sampling locations across Europe, two from Anatolia, five from North Africa, three from North America and one from New Zealand (Fig. 1 and Table S1). Adult insects were captured using a sweep net suitable for low-growing vegetation and an entomological aspirator. In some cases, larval stages were collected by hand. Specimens were preserved in absolute ethanol or dried in silica gel and stored at room temperature.

DNA extraction, amplification and sequencing

Entire larval stage specimens were used for DNA extraction while in the adults the wings and abdomen were removed and only the thorax and head were used. Genomic DNA was extracted

using the E.Z.N.A. Tissue DNA Isolation kit (Omega Bio-Tek) and a 800 bp fragment of the 3'-end of the mitochondrial gene cytochrome *c* oxidase subunit I (COI) was amplified by polymerase chain reaction (PCR) using the primers C1-J-2195 (5′–TTGATTTTTTGGTCATCCAGAAGT–3') and TL2-N-3014 (5′–TCCAATGCACTAATCTGCCATATTA–3') [28]. Due to DNA degradation in the New Zealand samples, a new set of primers, COI-PspF (5'–GTATAGATGTTGATACACGTGC–3') and COI-PspR (5'–TCCAGTAAATAAAGGGTATC–3') was designed to amplify an informative smaller fragment with 300 bp of COI that included the variable sites that differentiate the different haplogroups. Fragments with 500 bp of the mitochondrial genes cytochrome *c* oxidase subunit II (COII) and cytochrome *b* (cyt *b*) were also amplified using the primers TL2-J3033 (5'–GATATGGCAGAAATAGTGCA–3') and C2-N3665 (5'–CCACAAATTTCTGAACACTG–3') and CB-N3665 (5'–GTCCTACCATGAGGTCAAATATC–3') and CB-N11526 (5'–TTCAACTGGTCGTGCTCC–3'), respectively [29]. The three mitochondrial genes were sequenced initially in a subset of samples and similar genetic patterns and level of polymorphism were observed for the three mitochondrial genes on a preliminary analysis (results not shown). Thus, cytochrome *c* oxidase subunit I gene was sequenced for all individuals and cyt *b* and COII were only sequenced in a representative subset. The nuclear gene elongation factor-1α (EF-1α) is widely used in insect genetic studies. Therefore, we chose to sequence it in a subset of individuals that covered all the geographical areas of the study. A 700 bp fragment of the nuclear gene EF-1α was amplified using the primers M3 (5′–CACATYAACATTGTCGTSATYGG–3') and rcM44.9 (5′–CTTGATGAAATCYCTGTGTCC–3') [30]. For COI gene, PCR was performed in a 12.5 μL reaction volume containing: 1 μM of each primer, 0.1 mM dNTPs, 1 mM $MgCl_2$, 2.5 μL 5x Colorless GoTaq Flexi Buffer, 0.02U GoTaq DNA Polimerase (Promega) and approximately 30 ng of DNA. The PCR conditions were: an initial denaturation step at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 30 sec, annealing at 50 °C for 45 sec and extension at 72 °C for 1 min, with a final extension period at 72 °C for 7 min. The same PCR conditions were used for COII and cyt *b* genes except for annealing temperature where a touch up between 52.5 °C and 56 °C for COII and between 47 °C and 54 °C for cyt *b* was performed. Nuclear EF-1α gene PCR was performed in a 20 μL reaction volume containing: 0.6 μM of each primer, 0.2 mM dNTPs, 1.125 mM $MgCl_2$, 0.8 μL BSA (10 μg/mL), 4.0 μL 5x Colorless GoTaq Flexi Buffer, 0.05U GoTaq DNA Polimerase (Promega) and approximately 30 ng of DNA. PCR conditions used were: an initial denaturation step at 95 °C for 5 min, followed by 40 cycles of denaturation at 95 °C for 45 sec, annealing at 59 °C for 35 sec and extension at 72 °C for 1 min, with a final extension period at 72 °C for 10 min. All PCR products were purified with SureClean (Bioline) following the manufacturer's protocol, sequenced using the forward and the reverse primers with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analysed on a genetic analyser ABI PRISM 310 (Applied Biosystems).

Molecular and population structure analyses

Sequences were verified and edited using the software Sequencher v. 4.0.5 (Gene Codes Corporation) and BioEdit v. 7.0.9 [31]. They were then aligned using Mafft v. 7.029b (http://mafft.cbrc.jp/alignment/software/) and converted in the appropriate format with Concatenator v. 1.1.0 [32]. For nuclear EF-1α sequences, haplotype phase from heterozygous individuals for base positions and length-variable regions was determined using Champuru v. 1.0 [33]. Phylogenetic analysis using the Maximum Parsimony (MP), Maximum Likelihood (ML) and the Bayesian inference (BI) methods were performed for concatenated mtDNA genes (COI, COII and cyt *b*) and for the nuclear gene EF-1α in Paup v. 4.0.d99 [34] and in MrBayes v. 3.1.2 [35]. For MP and ML analysis a heuristic search was performed using 100 replicates and branch support was obtained by performing 1000 replicates of non-parametric bootstrap. Gaps were treated as a fifth base in MP. The BI analysis was performed using the Monte Carlo Markov Chain (MCMC) method iterated for 2 000 000 generations, with a sampling frequency of 1500 generations and a burn-in of 1000. For each dataset the best fit model of sequence evolution was estimated using Modeltest v. 3.7 [36] under the Akaike information criterion (AIC). Elongation factor-1α sequences of *P. spumarius* and *P. italosignus* from [26] and available at NCBI Genbank were added to our nuclear matrix and included in the phylogenetic analysis (GenBank accession numbers: JF309079 and JF309081-JF309095). *Philaenus italosignus* was used as outgroup in all phylogenetic analysis. Polymorphic sites and mtDNA haplotypes for COI, COII and cyt *b* genes were calculated using Mega v. 5.0 [37] and a median-joining haplotype network was constructed using Network v. 4.5.0.1 (Fluxus Technology Ltd. 2004). For COI mtDNA gene, haplotype (*h*) and nucleotide diversities (π) were calculated for each geographical region (defined as numbers in Fig. 1) and an analysis of molecular variance (AMOVA) was performed using Arlequin v. 3.5 [38] to assess population genetic structure of *P. spumarius*. The groupings were based in the several sub-regions of Europe, America, Africa and Asia. This analysis produces estimates of variance components and *F*-statistic analogues, designated as Φ-statistics, reflecting the correlation of haplotypes at different levels of hierarchical subdivision. Groupings with the highest significant Φ_{CT} value in AMOVA should reflect the most probable geographical subdivisions [39].

Divergence time estimates

We used the software package Beast v. 1.7.0 [40] and the mtDNA gene COI to estimate divergence times of nodes of interest, as well as their demographic history via Bayesian Skyline plots (BSPs). For each dataset the best fit model of sequence evolution was estimated using Modeltest v. 3.7 under the Akaike information criterion (AIC) and a piece-wise constant Bayesian skyline tree prior was selected with 10 groups. Two additional analyses of our data with 5 and 15 groups to assess the impact of the number of groups on the final result were conducted. These analyses did not reveal a significant impact on the overall result, whether on the shape of the Bayesian Skyline plot or on the estimation of divergence times. Preliminary runs using the uncorrelated lognormal relaxed clock revealed a posterior distribution of the σ_r (''CoefficientOfVariation'') parameter consistently abutting 0, suggesting that the COI partitions do not significantly deviate from a strict clock assumption. Therefore, we employed a strict molecular clock for each dataset with a normal prior distribution on the substitution rate with a mean of 0.0354, following the conserved rate of 3.54% per million years as suggested by [41], and a standard deviation of 0.005 to account for rate uncertainty. Two Markov chains Monte Carlo (MCMC) of 50 000 000 generations, sampled at every 5000th iteration, were conducted and combined with LogCombiner v. 1.6.1 [40]. Tracer v. 1.4 [42] was used to assess the convergence and mixing for all model parameters and to create the Bayesian Skyline plots.

Demographic analyses and neutrality tests

Neutrality tests of Tajima's *D* [43] and Fu's *F* statistics [44] were performed using ARLEQUIN. These statistics are widely used with molecular data to detect changes in population size and/or estimating deviations from neutrality, assuming a constant population size at mutation-drift equilibrium. Thus, significant negative values of Tajima's *D* and Fu's *Fs* are considered to be evidence of expanding populations. Signatures of population expansion can also be detected through the frequency distribution of the number of pairwise differences between haplotypes and thus statistics based on the mismatch distribution and taking into account the Sudden Expansion Model [45] were also performed to detect and estimate the time of population growth. Estimated expansion values were obtained using ARLEQUIN and graphics of frequency distribution using DNASP v. 5 [46]. To test the observed mismatch distribution goodness-of-fit to the Sudden Expansion Model and to obtain confidence intervals around the estimated mode of mismatch distribution, 1000 permutation replicates were used [47]. Statistically significant differences between observed and expected distributions were evaluated with the sum of the square deviations (*SSD*) and Harpending's raggedness index (*hg*) [48,49].

Timing of the demographic expansion as well as the 95% confidence interval for each mitochondrial haplogroup was estimated by converting the expansion time parameter τ, generated by ARLEQUIN, to time (t) in years using the formula $\tau = 2ut$, where u is the mutation rate per nucleotide per year multiplied by sequence length (i.e. number of nucleotides), and t is the time since population expansion in years [45,49]. We assumed a generation time of one year [24] and the conserved evolutionary rate of 3.54% per million years suggested by [41] for insect mitochondrial gene COI.

Results

A total fragment of 539 bp was obtained for the mitochondrial gene COI in 190 samples. The remaining samples from Finland, Turkey, Canada (GenBank accession numbers: KJ699232–KJ699234) and New Zealand were not included in the main analysis due to their reduced size. In the 190 individuals there were a total of 71 haplotypes (GenBank accession numbers: KC111886 – KC111956) of which 44 occurred only once (Table S2). Of a total of 539 sites sequenced, 53 were polymorphic but only 26 were parsimony informative. For the mtDNA COII (495bp) and cyt *b* (434bp) genes, 47 individuals were sequenced and a total of 14 (GenBank accession numbers: KF280589 – KF280602) and 18 haplotypes (GenBank accession numbers: KF280603 – KF280620) were found, respectively. As commonly observed for insects [28], nucleotide sequences were A+T rich (approximately 71%). No gaps or early stop codons were detected in the 3 mtDNA genes sequences suggesting that all of them are functional mitochondrial DNA copies.

From the 24 individuals sequenced, we were able to successfully sequence a fragment of the nuclear gene EF-1α for only 13 individuals of *P. spumarius*. Almost all sequences exhibited double peaks due to the frameshift resulting from indels (length-variable regions) located at several sites of the intron in the EF-1α gene. Ten individuals were heterozygous in respect to indels and/or to base positions and the phased haplotypes (*alleles*) were differentiated by adding the letter *a* or *b* at the end of name (Table S1) (GenBank accession numbers: KF280621 – KF280642).

Phylogenetic and population structure analyses

Phylogenetic trees obtained for concatenated mtDNA genes and for a subset of *P. spumarius* individuals by the three methods, MP, ML and BI, presented a congruent topology. Maximum likelihood (Fig. 2), MP (not shown) and BI (not shown) trees, revealed the existence of two main haplotype groups: the "western" and the "eastern". The "western" is divided in the "western-Mediterranean" and the "eastern-Mediterranean" sub-groups. The same phylogeographic pattern was found in the COI median-joining haplotype network (Fig. 3) and also in the COII and Cyt *b* median-joining haplotype networks (Fig. S1 and S2). The "eastern" haplogroup includes haplotypes from a wide geographical area, including northern Anatolia (Cerkes), Finland, Belgium, the UK (Aberdare – Wales) and eastern North America (New Hampshire – USA). In the "western-Mediterranean" group, the most common haplotype (H29) and several derived haplotypes, differing by one or two mutational steps, are shared between populations from the Iberian Peninsula, Morocco, France, Belgium, Italian Peninsula, Sicily and one individual from Balkans (H18). A group of haplotypes derived from H29 (H23, H24, H25, H70 and H71) includes samples from the Azores, western North America (British Columbia – Canada), eastern North America (Michigan – USA) and the UK, differing by two or three mutations. In the "eastern-Mediterranean" group, a similar star-like pattern is present with rare haplotypes connected to the most common (H57), usually by one mutational step. This group encompasses populations from the Balkans (Greece, Bulgaria and European Turkey), Slovenia, Italian Peninsula and Sicily. This lineage is also present in five samples from the eastern part of the Iberian Peninsula (H56) (Figs. 1 and 2).

The haplotype median-joining network based on smaller sequences of 289 bp of the COI gene, used in this analysis to include the samples from New Zealand, showed a total of 31 haplotypes and a pattern congruent to that observed for the 539 bp of COI region, with the same three distinct groups ("western-Mediterranean", "eastern-Mediterranean" and "eastern"). The haplotype H14 belonging to the "western" haplogroup was found in the three New Zealand individuals and this same haplotype was shared with seven individuals from the UK, five from the Azores, three from Canada and four from the USA (Fig. S3).

In the MP (Fig. 4) and BI (not shown) phylogenetic trees obtained for nuclear gene EF-1α three main groups could also be distinguished: *clade* A, *clade* B and *clade* C. However, these groups were not totally congruent to the groups found for the concatenated mtDNA genes. The *clade* A includes samples from our "eastern-Mediterranean" haplogroup and also from Georgia, Bulgaria, Hungary, Greece and Italy (E2 *clade*, in [26]). The *clade* B includes individuals from our "western-Mediterranean" group and from Portugal, Spain and Italy (E3 *clade*, in [26]). The *clade* C, however, is constituted by individuals from our three main mitochondrial groups and from Russia, Norway, Alps, Crimea, Poland and Ukraine (E1 *clade*, in [26]). Although the three groups have good bootstrap support, the branching order is unsolved since there is very low support for the branch clustering the *clades* A and B (Fig. 4). We also observed that both alleles of the most heterozygous samples are clustered within the same *clade* (*Clade* C – Arrabida_5, Spain_2, UK_7, Belgium_1, Slovenia_1.1 and USA_2 and *Clade* B – Morocco_6.1), with the exception for the Azores_1, Italy_2 and Keçan_N1 samples which have one allele in the *clade* C and the other allele in the *clades* A or B.

Genetic variability and diversity

Mitochondrial haplotype diversity of COI was generally high ($h > 0.6000$), except for the Azores ($h = 0.4000$) and Slovenia ($h =$

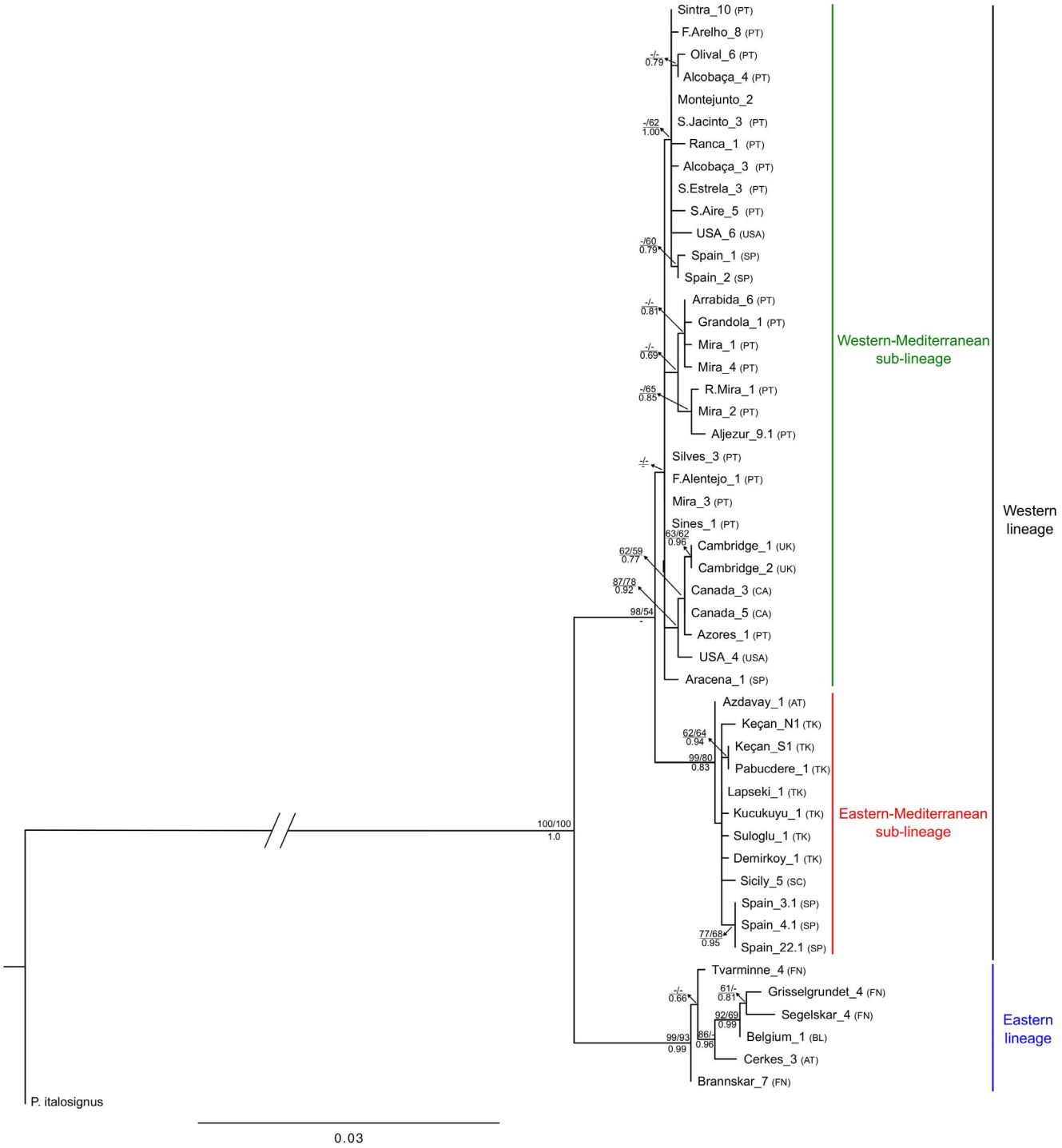


Figure 2. Maximum Likelihood tree based on the 3 concatenated mtDNA genes (COI, COII and cyt *b*) (1527bp). Values above branches correspond to MP and ML bootstrap values (only values > 50% are shown) and values below branches correspond to Bayesian posterior probability. PT – Portugal; SP – Spain; UK – United Kingdom; BL – Belgium; FN – Finland; SC – Sicily; TK – European Turkey; AT – Anatolia; USA – United States of America; CA – Canada.
doi:10.1371/journal.pone.0098375.g002

0.0000) (Table 1), that may be related to the low sample size. On the other hand, Anatolia and the Balkans have the highest values of haplotype diversity (h = 0.9048 and h = 0.9128, respectively). Interestingly, Finland and North America also have high haplotype diversity (h = 0.8971 and h = 0.8939). Analysing nucleotide diversity, Anatolia, United Kingdom and North America have the highest values (π = 0.009895, π = 0.009318 and π = 0.008236, respectively). The high values of both haplotype and nucleotide diversities detected in Anatolia and North America are likely a result of the presence of different mitochondrial lineages in these regions. The AMOVA performed for the *P. spumarius* groupings defined revealed that most of the

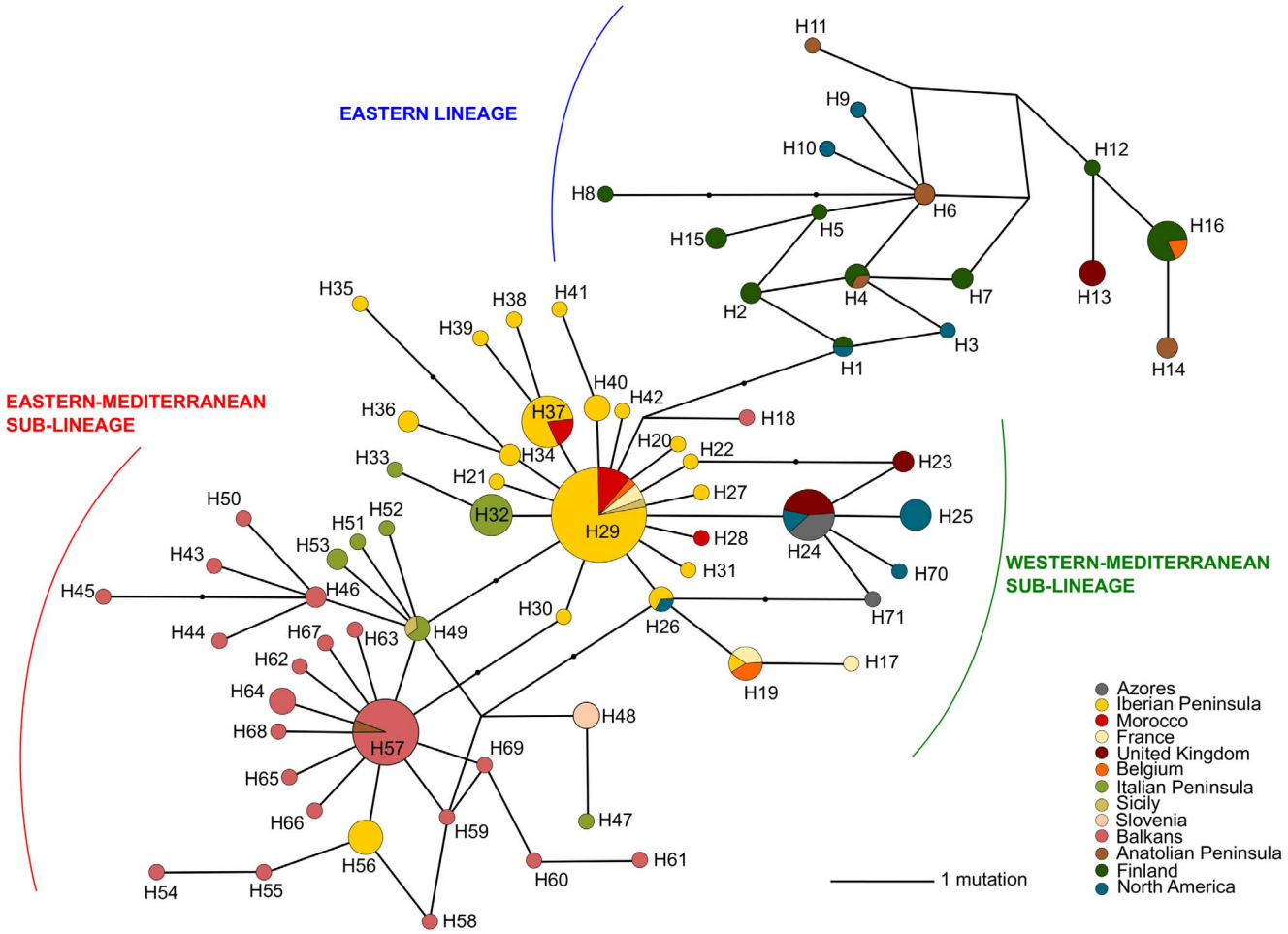


Figure 3. Median-joining haplotype network of *Philaenus spumarius* sampled geographic regions for mitochondrial gene COI (539bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their sizes are in proportion to the number of mutations.
doi:10.1371/journal.pone.0098375.g003

genetic variation (47% and 43.87%) is explained by genetic differences within populations and not by geographic subdivisions (Table S3).

Divergence times

The mean ages and 95% highest posterior density (HPD) determined for the TMRCA of mtDNA COI haplogroups are presented in Table S4. Estimated divergence times of all mitochondrial groups are less than 0.5 Ma. The "eastern" mtDNA group seems to be the oldest having diverged from the "western" mtDNA group at approximately 0.270 Ma ago and then begun its diversification around 0.190 Ma ago (0.374–0.056 Ma ago). The split of the "western" mtDNA group into the "eastern-Mediterranean" and the "western-Mediterranean" subgroups was estimated to have occurred at approximately 0.146 Ma ago (0.243–0.067 Ma ago), while the TMRCA of both groups was quite similar and was estimated to be around 0.080 Ma ago. The confidence interval associated with our time estimates, however, is broad and the mutation rate of 3.54% per million years, in which these calculations are based, was estimated for COI in tenebrionid beetles [41]. Therefore, our results should be treated with caution and regarded as the best approximations given the current methods and calibrations [50,51].

Demographic analyses and neutrality tests

The demographic history of *P. spumarius* populations was analysed separately for the three COI groups. The distribution of pairwise nucleotide differences (mismatch distribution) showed that the "western-Mediterranean" and the "eastern-Mediterranean" groups exhibited a smooth and unimodal shape while the "eastern" group revealed a slightly bimodal curve (Fig. S4). All distributions, except for the "eastern" group distribution, were consistent with sudden and spatial population expansions. The observed raggedness index was low for all groups and both P_{SSD} and P_{RAG} showed that the observed distributions did not differ significantly from those expected under a sudden and a spatial population expansion model (Table 2). Negative significant deviations from neutrality were detected with Tajima's D and Fu's F statistics for the "western-Mediterranean" and the "eastern-Mediterranean" mtDNA groups, which corroborate the hypothesis of past population expansion events. The "eastern" group presented a non significant p-value with Tajima's D test although the Fu's F was significant (Table 3) indicating that it may have undergone negligible population growth. Demographic reconstructions (BSPs) for all mtDNA groups suggest a trend of population growth (Fig. S5), with a more evident demographic expansion in the "western-Mediterranean" and the "eastern-

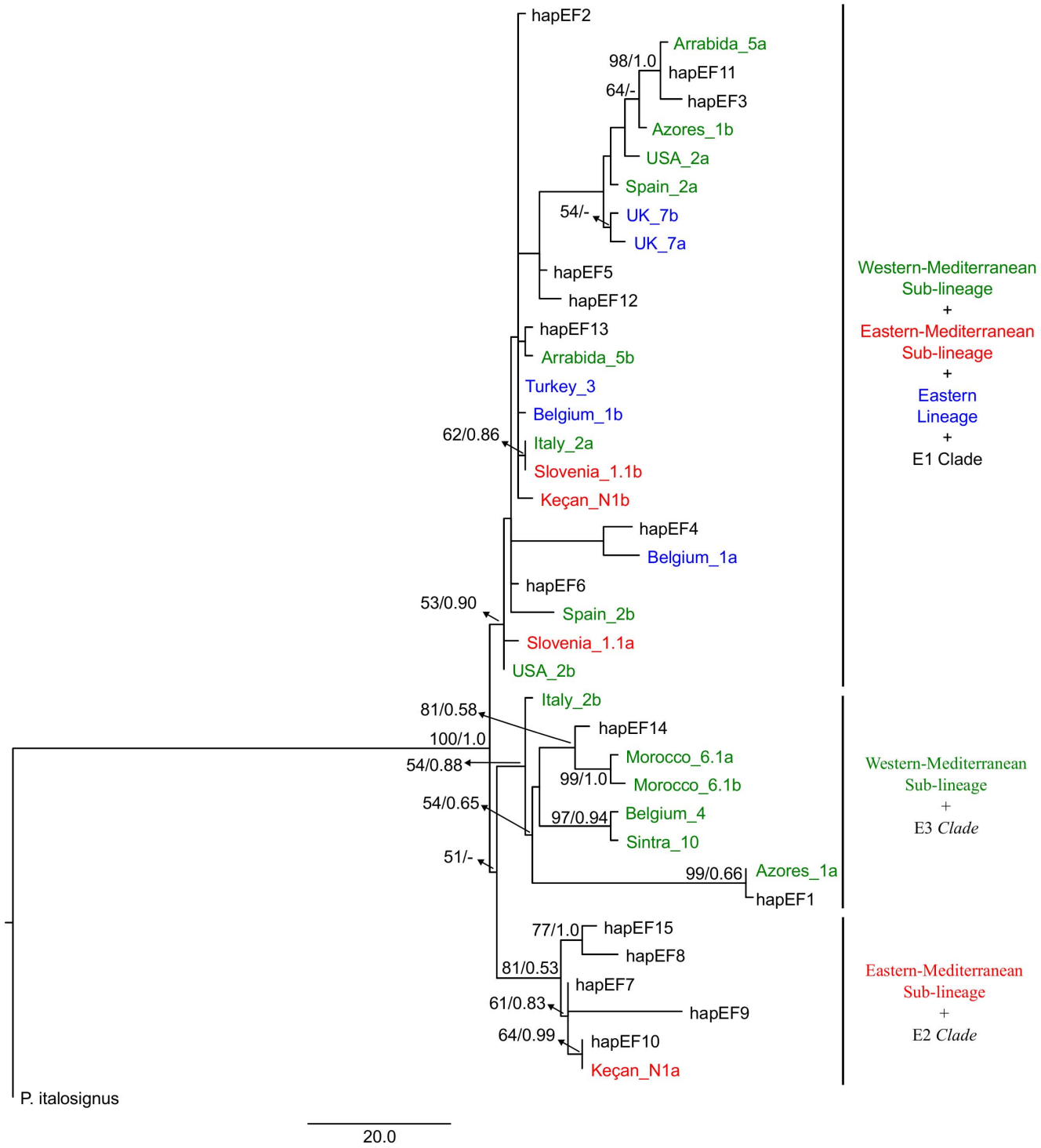


Figure 4. Maximum Parsimony tree based on nuclear gene elongation factor-1α. Values above branches correspond to MP bootstrap (only values > 50% are shown) and Bayesian posterior probability values. Black: GenBank sequences (see [26]); blue individuals correspond to the eastern mtDNA group; red individuals correspond to the eastern-Mediterranean mtDNA group and green individuals to the western-Mediterranean mtDNA group.
doi:10.1371/journal.pone.0098375.g004

Mediterranean" mtDNA groups, and very slight or absent population growth for the "eastern" lineage.

Assuming a slight population growth for the "eastern" group, the timing of demographic expansion was estimated to have occurred at approximately 0.121 Ma ago (0.195–0.041 Ma ago), while for the "western-Mediterranean" and the "eastern-Mediterranean" haplogroups it was more recent, at 0.055 Ma ago (0.085–0.021 Ma ago) and 0.058 Ma ago (0.079–0.027 Ma ago), respectively. The spatial expansion for the three mtDNA groups was estimated to be slightly more recent than demographic

Table 1. Number of individuals, number of haplotypes and genetic diversity indices calculated for geographic regions of *Philaenus spumarius* and for mitochondrial gene Cytochrome *c* oxidase I (COI).

Geographic regions	Number of individuals	Number of haplotypes	Haplotype diversity (*h*)	Nucleotide diversity (π)
Morocco	7	3	0.6667 +/− 0.1598	0.001414 +/− 0.001338
Iberian Peninsula	63	19	0.7798 +/− 0.0493	0.003133 +/− 0.002062
Azores	5	2	0.4000 +/− 0.2373	0.000742 +/− 0.000944
Western Europe	9	4	0.7500 +/− 0.1121	0.006597 +/− 0.004194
Slovenia	3	1	0.0000 +/− 0.0000	0.000000 +/− 0.000000
Italy	17	8	0.8162 +/− 0.0815	0.004666 +/− 0.002957
Balkans	40	21	0.9128 +/− 0.0303	0.003825 +/− 0.002431
Anatolian Peninsula	7	5	0.9048 +/− 0.1033	0.009895 +/− 0.006223
United Kingdom	10	3	0.6889 +/− 0.1038	0.009318 +/− 0.005592
Finland	17	9	0.8971 +/− 0.0534	0.006603 +/− 0.003950
North America	12	8	0.8939 +/− 0.0777	0.008236 +/− 0.004917

Western Europe: Belgium and France; Italy: Italian peninsula and Sicily.
doi:10.1371/journal.pone.0098375.t001

expansion, at 0.080 Ma ago (0.152–0.035 Ma ago) for the "eastern" group, at 0.056 Ma ago (0.075–0.028 Ma ago) for the "eastern-Mediterranean" and at 0.054 Ma ago (0.074–0.022 Ma ago) for the "western-Mediterranean" group.

Discussion

Biogeographical patterns, divergence time and demographic events in *Philaenus spumarius*

Our time estimates indicate that the evolutionary history of *P. spumarius* is most likely related to climate changes of the Pleistocene epoch (~2.588–0.0117 Ma ago [1]). Divergence within species is estimated to be recent (no more than 0.5 Ma) occurring most probably in the Middle/Late Pleistocene. The biogeographical pattern of *P. spumarius* obtained from mtDNA genes shows the differentiation of two main mtDNA lineages, the "western" in the Mediterranean region and the "eastern" in Anatolia/Caucasus. Within the "western" lineage we observed two sub-lineages: the "western-Mediterranean" centred in the Iberian Peninsula and the "eastern-Mediterranean" centred in the Balkans (Fig 1). This pattern was first found by [27], who gave more emphasis to the westernmost lineages, and was later corroborated by [26] that brought some insights regarding possible refugia located in eastern Europe and western Asia. According to our data, during the Mindel or Riss glacial period, a first split between the western and eastern populations seems to have occurred with the diversification of the "western" lineage in the Mediterranean region, and of the "eastern" lineage maybe in Anatolia and surrounding area of the Caucasus, or even in territories of western Asia, as also suggested by [26].During the following interglacial, the "eastern" lineage seems to have suffered a negligible population growth compared with a more significant demographic expansion of the "western" lineage, which appears to have later retracted to two Mediterranean refugia, the Iberian Peninsula and the Balkans, where it diverged into two sub-lineages (the "western-Mediterranean" and the "eastern-Mediterranean"), maybe during the Würm glacial. After that period, the "eastern-Mediterranean" lineage centred in the Balkans seems to have expanded to the Italian Peninsula. The land bridge which existed in the northern and central part of the present Adriatic Sea between the Italian and the Balkan peninsulas, during the Quaternary cold periods [6], would have made the contact between these two peninsulas easier. The expansion was followed by a slight differentiation in Italy from the Balkans. The COI haplotype network shows that the dispersal of haplotypes from the Iberian Peninsula to the north of Italy has also occurred, maybe either crossing mountains or along the Mediterranean coast of Spain and France as suggested for *Cicada orni* [52]. Although a weak flier, it is also possible that *P. spumarius* dispersed over the sea facilitated by wind (anemohydrochoric dispersal) since this mode of dispersal has already been observed in this species [20].

The expansion dates estimated here have wide confidence intervals. However, the lowest boundaries of these dates are about 0.015 Ma ago, suggesting that the demographic and spatial expansion of this species may have occurred earlier than Holocene. Our data and reference [26] suggest the current geographic pattern of the species seems to result from postglacial range expansion from the Iberian Peninsula to the central and north-western Europe and, from the Anatolia/Caucasus (and eventually from western Asia) to east, north and central Europe, thus seemingly not following one of the common four paradigms [12,53]. Although a northern expansion from Balkans cannot be completely ruled out, the current data indicates that the Carpathians may have represented a geographic barrier to the northern expansion of Balkans populations. Further detailed sampling and genetic analysis of the Carpathian region would be important to test this hypothesis.

Contact zones in Europe have been recorded for several European temperate species (reviewed in [54]). According to our results, the mtDNA lineages are geographically separated in most part of the range of *P. spumarius* but came into contact in some geographic regions (Fig. 1). This suggests the existence of recent admixture (secondary contact of diverged lineages) between mtDNA lineages in populations from these regions, also corroborated by [26]. The presence of the "eastern-Mediterranean" sub-lineage in some populations from the eastern part of the Iberian Peninsula (haplotype H56) and of the "western" sub-lineage in Balkans (haplotype H18) indicates that recent migrations between Mediterranean refugia may have occurred during the Quaternary period as reported in the olive fly *Bactrocera oleae* [55]. However, incomplete lineage sorting of an ancestral polymorphism cannot

Table 2. Parameters from the mismatch distribution for *Philaenus spumarius* COI groups.

	Mismatch Analysis						
	Demographic Expansion						
	Parameters						
	θ_0 (CI = 95%)	θ_1 (CI = 95%)	τ (CI = 95%)	SSD	P_{SSD}	Raggedness	P_{rag}
Eastern Group	0.00176 (0.000–1.366)	14.41895 (8.628–99999)	4.64844 (1.561–7.461)	0.00620	0.55700	0.01760	0.84400
Western-Mediterranean Group	0.00703 (0.000–0.729)	25.15625 (4.970–99999)	2.11523 (0.801–3.238)	0.00133	0.58200	0.03281	0.65100
Eastern-Mediterranean Group	0.04395 (0.000–0.698)	115.625 (9.687–99999)	2.21094 (1.041–3.016)	0.00144	0.55300	0.03706	0.59400
	Spatial Expansion						
	Parameters						
	θ (CI = 95%)	M (CI = 95%)	τ (CI = 95%)	SSD	P_{SSD}	Raggedness	P_{rag}
Eastern Group	1.26863 (0.001–3.921)	23.44725 (10.230–99999)	3.07576 (1.335–5.833)	0.00933	0.30800	0.01760	0.85800
Western-Mediterranean Group	0.02484 (0.001–1.066)	42.03757 (8.630–99999)	2.05086 (0.863–2.820)	0.00129	0.57000	0.03281	0.66500
Eastern-Mediterranean Group	0.07874 (0.001–0.958)	351.4398 (18.276–99999)	2.15680 (1.062–2.877)	0.00146	0.45200	0.03706	0.60600

Numbers in parenthesis are the upper and lower bound of 95% CI (1000 bootstrap replicates).
θ_0 and θ_1: pre-expansion and post-expansion populations size; τ: time in number of generations elapsed since the sudden/demographic expansion and spatial expansion episodes; SSD: sum of squared deviations; Raggedness: raggedness index following [49]; P_{SSD} and P_{RAG}.: probability that expected mismatch distributions (1000 bootstrap replicates) be larger than observed mismatch distributions.
doi:10.1371/journal.pone.0098375.t002

Table 3. Tajima's *D* and [44] Fu's *Fs* test values and their statistical significance for *Philaenus spumarius* Cytochrome *c* oxidase I mtDNA groups.

	Neutrality Tests	
	Tajima's *D* test	**Fu's *Fs* test**
Eastern Group	0.15553	−6.28375**
Western-Mediterranean Group	−1.77941*	−23.61561***
Eastern-Mediterranean Group	−1.75709*	−26.22826***

*: indicates significant values at P<0.05; **: indicates significant values at P<0.01 and ***: indicates significant values at P<0.001.
doi:10.1371/journal.pone.0098375.t003

be ruled out as another possible explanation for the current mtDNA pattern of *P. spumarius*. Our data also indicate the existence of incomplete lineage sorting and/or admixture in the nuclear gene EF-1α. Although nuclear *clades* A and B correspond well to the "eastern-Mediterranean" and to the "western-Mediterranean" mtDNA sub-lineages, respectively, *clade* C is a mix of individuals from the three mitochondrial lineages. Heterozygous individuals whose alleles grouped within different *clades* were found, a fact not detected by [26], since they analysed homozygous individuals only. In the nuclear gene there was a lack of support for the monophyly of the "western" lineage ("western-Mediterranean and "eastern-Mediterranean" mitochondrial sub-lineages). Taken together, our results suggest that the current biogeographic pattern of *P. spumarius* may be the result of both secondary admixture and incomplete lineage sorting.

Also quite interesting is the uncommon [12] high genetic diversity detected in *P. spumarius* populations from northern Europe (Scandinavia) indicating that the north of Europe was colonized by populations that may have survived in several extra-Mediterranean glacial refugia in addition to the "classical" Mediterranean ones [7].

Gene-flow between the Iberian Peninsula and Morocco

The presence of the "western-Mediterranean" sub-lineage in Morocco suggests a close relationship between these *P. spumarius* populations and the Iberian Peninsula. This close relationship is also corroborated by the nuclear data. There is evidence that the Strait of Gibraltar has not been an effective barrier to the dispersal, having been, in fact, the route of dispersal for many species from Africa to Europe and vice-versa [56,57,58]. Lowered sea levels during glacial periods possibly facilitated exchange across the Strait of Gibraltar [59]. It is quite possible that, during such lower sea level periods, individuals from the Iberian Peninsula reached North Africa via anemohydrochoric dispersal. Contrarily to the thermophilous species *Cicada barbara*, a common cicada in southern Portugal and Spain [60], the Rif Mountains did not appear to have acted as a geographical barrier to the dispersal of *P. spumarius* through Morocco, since this latter can be found in a variety of terrestrial habitats and even at altitudes above 1700m (e.g. Mt Parnassus, in Greece: observations by Ana Rodrigues, Sara Silva and Eduardo Marabuto). Although the haplotypes found in samples from Morocco were the same as some of the ones found in the Iberian Peninsula, indicating that they belong to *P. spumarius*, the analysis of four male genitalia from these populations revealed similarities with *P. tesselatus* [61] and showed the necessity of further investigation on the taxonomic status of these species, as previously suggested by [26].

The UK and the origin of the North American and insular populations

Our analyses suggest the presence of at least two mitochondrial lineages in the UK, the "western-Mediterranean" and the "eastern", and support a British origin for the populations of the Azores and New Zealand, and a multiple origin for the North American populations (Figs. 1 and 2). Populations from the Azores and New Zealand seem to have originated from only one of the British lineages here represented (the "western-Mediterranean" lineage). North American populations seem to have a mixed origin: a British origin, already suggested in the preliminary study by [27], from the "western-Mediterranean" lineage present in the UK, and an Iberian origin, due to the close relationship between one haplotype (H26) from the eastern United States and the Iberian Peninsula haplotypes. In fact, multiple translocations from different localities from western Europe have already been suggested for North American populations [26]. The close relationship of some North American haplotypes (New Hampshire) to the ones found in Anatolia and Finland indicates another origin from the "eastern" mitochondrial lineage, and that was never reported before. Nevertheless, a morphological variation in North American populations was already reported by [62]. The author shows that *P. spumarius* populations from New Hampshire and adjacent areas of North American exhibit morphological variation in male genitalia features and attributes this variation to hybridization between *P. spumarius* subspecies from different parts of Europe. Verification of whether there is any correlation between Hamilton's morphological subspecies and haplotype variation would require a parallel investigation of morphology beyond the scope of this work. The colonization of New Zealand and North America was probably recent and resulted from non-intentional anthropogenic introductions. This recent colonization could explain the spittlebug pest status in some locations of the USA and Canada, where *P. spumarius* reaches high densities, perhaps as result of the lack of competitors and predators [19,24]. Concerning the Azorean populations, the fact that they only occur in high altitude native habitats (e.g. in the geologically older areas of S. Miguel), very far from human altered habitats, is somewhat puzzling and we cannot exclude the possibility of a natural colonization by long-distance dispersal.

Conclusion

P. spumarius is one of the most widespread insects in Europe. We successfully provided time estimates of the main demographic and evolutionary events for the populations occurring in the main Mediterranean peninsulas, and in addition interpreted the colonization patterns out of Eurasia, namely of north-western Africa, North America, and the islands of Azores and New Zealand. This combination of a well analysed phylogeographic

and demographic pattern with the multiple transcontinental colonization events, some putatively natural, others recent non-intentional anthropogenic introductions, with ecosystem level consequences, make this species well placed for understanding the long term effects of invasive species and their post-invasion evolution.

Supporting Information

Figure S1 Median-joining haplotype network of a set of *Philaenus spumarius* sampled geographic regions for mitochondrial gene COII (495bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their size is in proportion to the number of mutations.
(TIF)

Figure S2 Median-joining haplotype network of a set of *P. spumarius* sampled geographic regions for mitochondrial gene cyt *b* (434bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their size is in proportion to the number of mutations.
(TIF)

Figure S3 Median-joining haplotype network of *P. spumarius* sampled geographic regions for mitochondrial gene COI (289bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their size is in proportion to the number of mutations.
(TIF)

Figure S4 Mismatch distribution of mtDNA COI haplotypes for each of the three *P. spumarius* haplogroups. (a) Eastern lineage; (b) Western-Mediterranean sub-lineage and (c) Eastern-Mediterranean sub-lineage. The expected frequency is based on a population growth-decline model, determined using DnaSP and is represented by a continuous line. The observed frequency is represented by a dotted line. Parameter values for the mismatch distribution are given in Table 2.
(TIF)

Figure S5 Bayesian skyline plots showing the historical demographic trends for each main *Philaenus spumarius* mtDNA group detected using COI gene. Along the y-axis is the expressed population size estimated in units of Neμ (Ne: effective population size, μ: mutation rate per haplotype per generation). The y-axis is in a log-scale. Solid lines represent median estimates and blue lines represent the 95% high probability density (HPD) intervals.
(TIF)

Table S1 Analysed samples of *Philaenus spumarius* with description of the sampling locations and indication of the corresponding mtDNA Cytochrome *c* oxidase I (COI), Cytochrome *c* oxidase II (COII), Cytochrome *b* haplotype/code and Elongation Factor-1α code (EF-1α).
(PDF)

Table S2 Haplotype distribution within *P. spumarius* geographic regions for mitochondrial gene COI. The total number of haplotypes per geographic region and the total number of individuals per haplotype are also shown. Western Europe: Belgium and France; Italy: Italian peninsula and Sicily.
(PDF)

Table S3 Analyses of molecular variance (AMOVA) among regions of *P. spumarius* based on COI data.
(PDF)

Table S4 Divergence time estimates in million years (Ma) from the most recent common ancestor of each main *Philaenus spumarius* mtDNA COI haplogroup estimated using a mean mutation rate of 3.54% per million years as suggested by [41]**.**
(PDF)

Acknowledgments

We thank for samples and help in the work: Francisco Pina-Martins, Genage André, Andreia Penado, Vera Nunes, Carla Pereira, Filipe Ribeiro, Daniel Pires, Renata Martins, Rui Nunes, Luís Ferreira, Paulo Fonseca, Paula Simões, Carla Rego, Olli Halkka, Liisa Halkka, Kaisa Roukka, Andrew Beckenbach, Martin Corley, Vera d' Urso, Federico Lessio, Jérôme Constant and Maja Derlink. We thank Thomas Schmitt and anonymous referees for helpful comments on a previous versions of the manuscript. We would like to pay homage to Professor Olli Halkka who dedicated a substantial part of his life to the study of this interesting species and with whom we had the privilege to collaborate. His work continues to be a source of inspiration for our studies.

Author Contributions

Conceived and designed the experiments: ASBR SGS OSP. Performed the experiments: ASBR SGS. Analyzed the data: ASBR DNS. Contributed reagents/materials/analysis tools: ASBR SES EM MRW VT SY AH PAVB JAQ SGS OSP. Wrote the paper: ASBR. Revised the text: SES EM DNS MRW VT SY AH PAVB JAQ OSP SGS.

References

1. Gibbard PL, Head MJ, Walkers MJC, The Subcommission on Quaternary Stratigraphy (2010) Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. Journal of Quaternary Science 25: 96–102.
2. Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58: 247–276.
3. Hewitt G (2011) Mediterranean peninsulas: the evolution of hotspots. In: Zachos FE, Habel JC, editors. Biodiversity hotspots: distribution and protection of conservation priority areas. Springer Berlin Heidelberg. pp. 123–147.
4. Bhagwat SA, Willis KJ (2008) Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? Journal of Biogeography 35: 464–482.
5. Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. Trends in Ecology & Evolution 23: 564–571.
6. Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7: 453–464.
7. Schmitt T, Varga Z (2012) Extra-Mediterranean refugia: the rule and not the exception? Frontiers in Zoology 9: 22.
8. Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. Trends in Ecology & Evolution 16: 608–613.
9. Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. Proceddings of the Royal Society Biological Sciences 277: 661–671.
10. Besold J, Schmitt T, Tammaru T, Cassel-Lundhagen A (2008) Strong genetic impoverishment from the centre of distribution in southern Europe to peripheral Baltic and isolated Scandinavian populations of the pearly heath butterfly. Journal of Biogeography 35: 2090–2101.
11. Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity 77: 282–291.
12. Hewitt GM (1999) Post-glacial re-colonization of european biota. Biological Journal of the Linnean Society 68: 87–112.
13. Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Molecular Ecology 10: 537–549.
14. Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology 17: 2107–2121.
15. Arensburger P, Buckley TR, Simon C, Moulds M, Holsinger KE (2004) Biogeography and phylogeny of the New Zealand cicada genera (Hemiptera:

Cicadidae) based on nuclear and mitochondrial DNA data. Journal of Biogeography 31: 557–569.
16. Ribera I, Vogler AP (2004) Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). Molecular Ecology 13: 179–193.
17. Schmitt T, Seitz A (2001) Allozyme variation in *Polyommatus coridon* (Lepidoptera: Lycaenidae): identification of ice-age refugia and reconstruction of post-glacial expansion. Journal of Biogeography 28: 1129–1136.
18. Quartau J, Borges P (1997) On the colour polymorphism of *Philaenus spumarius* (L.) (Homoptera, Cercopidae) in Portugal. Miscellanea Zoologica20: 19–30.
19. Yurtsever S (2000) On the polymorphic meadow spittlebug, *Philaenus spumarius* (L.) (Homoptera: Cercopidae). Turkish Journal of Zoology 24: 447–459.
20. Halkka O, Halkka L (1990) Population genetics of the polymorphic meadow spittlebug, *Philaenus spumarius* (L.). Evolutionary Biology (USA) 24: 149–191.
21. Quartau J, Borges P, André G (1992) *Phileanus spumarius* (Linaeus, 1975) new to the Azores (Homoptera, Auchenorrhyncha, Cercopidae). Actas do Congresso Iberico de Entomologia 1: 129–136.
22. Borges PAV, Costa A, Cunha R, Gabriel R, Gonçalves V, et al. (2010) A list of the terrestrial and marine biota from the Azores. Princípia, Cascais432 pp
23. Thompson V (1984) Color polymorphism in the introduced spittlebug *Philaenus spumarius* (Homoptera, Aphrophoridae) in New Zealand. New Zealand Entomologist 8: 86–88.
24. Weaver C, King D (1954) Meadow spittlebug *Philaenus leucophthalmus* (L.). Ohio Agricultural Experimental Station Research Bulletin 741: 1–99.
25. Karban R, Strauss SY (2004) Physiological tolerance, climate change, and a northward range shift in the spittlebug, *Philaenus spumarius*. Ecological Entomology 29: 251–254.
26. Maryanska-Nadachowska A, Kajtoch L, Lachowska D (2012) Genetic diversity of *Philaenus spumarius* and *P. tesselatus* (Hemiptera, Aphrophoridae): implications for evolution and taxonomy. Systematic Entomology 37: 55–64.
27. Seabra SG, Pina-Martins F, Marabuto E, Yurtsever S, Halkka O, et al. (2010) Molecular phylogeny and DNA barcoding in the meadow-spittlebug *Philaenus spumarius* (Hemiptera, Cercopidae) and its related species. Molecular Phylogenetics and Evolution 56: 462–467.
28. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, et al. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87: 651–701.
29. Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics 37: 545–579.
30. Downie DA, Gullan PJ (2004) Phylogenetic analysis of mealybugs (Hemiptera: Coccoidea: Pseudococcidae) based on DNA sequences from three nuclear genes, and a review of the higher classification. Systematic Entomology 29: 238–260.
31. Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.
32. Pina-Martins F, Paulo OS (2008) CONCATENATOR: sequence data matrices handling made easy. Molecular Ecology Resources 8: 1254–1255.
33. Flot J (2007) Champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. Molecular Ecolology Notes 7: 974–977.
34. Swofford D (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
35. Huelsenbeck J, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics 17: 754–755.
36. Posada D, Crandall K (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817–818.
37. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731–2739.
38. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47–50.
39. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479–491.
40. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29: 1969–1973.
41. Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-aegean trench calibration. Molecular Biology and Evolution 27: 1659–1672.
42. Rambaut A, Drummond AJ (2007) Tracer v1.4. Available: http://beast.bio.ed.ac.uk/Tracer.
43. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.
44. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915–925.
45. Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9: 552–569.
46. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.
47. Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 152: 1079–1089.
48. Harpending H, Sherry S, Rogers A, Stoneking M (1993) The genetic structure of ancient human populations. Current Anthropology 34: 483–496.
49. Harpending HC (1994) Signature of ancient population-growth in a low-resolution mitochondrial-DNA mismatch distribution. Human Biology 66: 591–600.
50. Graur D, Martin W (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends in Genetics 20: 80–86.
51. Heads M (2005) Dating nodes on molecular phylogenies: a critique of molecular biogeography. Cladistics 21: 62–78.
52. Seabra SG, Quartau JA, Bruford MW (2009) Spatio-temporal genetic variation in sympatric and allopatric Mediterranean Cicada species (Hemiptera, Cicadidae). Biological Journal of the Linnean Society 96: 249–265.
53. Habel JC, Schmitt T, Muller P (2005) The fourth paradigm pattern of post-glacial range expansion of European terrestrial species: the phylogeography of the Marbled White butterfly (Satyrinae, Lepidoptera). Journal of Biogeography 32: 1489–1497.
54. Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. Genetica 139: 617–638.
55. Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F (2005) Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). Molecular Ecology 14: 2729–2738.
56. Franck P, Garnery L, Loiseau A, Oldroyd BP, Hepburn HR, et al. (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. Heredity 86: 420–430.
57. Horn A, Roux-Morabito G, Lieutier F, Kerdelhue C (2006) Phylogeographic structure and past history of the circum-Mediterranean species *Tomicus destruens* Woll. (Coleoptera: Scolytinae). Molecular Ecology 15: 1603–1615.
58. Schmitt T, Rober S, Seitz A (2005) Is the last glaciation the only relevant event for the present genetic population structure of the meadow brown butterfly *Maniola jurtina* (Lepidoptera: Nymphalidae)? Biological Journal of the Linnean Society 85: 419–431.
59. Rodriguez-Sanchez F, Perez-Barrales R, Ojeda F, Vargas P, Arroyo J (2008) The Strait of Gibraltar as a melting pot for plant biodiversity. Quaternary Science Reviews 27: 2100–2117.
60. Pinto-Juma GA, Quartau JA, Bruford MW (2008) Population structure of *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco based on mitochondrial DNA analysis. Bulletin of Entomological Research 98: 15–25.
61. Drosopoulos S, Quartau JA (2002) The spittlebug *Philaenus tesselatus* Melichar, 1899 (Hemiptera, Auchenorrhyncha, Cercopidae) is a distinct species. Zootaxa 68: 1–8.
62. Hamilton KGA (1979) Synopsis of North American Philaenini (Rhynchota: Homoptera: Cercopidae) with a new genus and four new species. The Canadian Entomologist 111: 127–141.

CHAPTER 2

Supporting Information

Appendix S1

Table S1 - Analysed samples of *Philaenus spumarius* with description of the sampling locations and indication of the corresponding mtDNA Cytochrome *c* oxidase I (COI), Cytochrome *c* oxidase II (COII), Cytochrome *b* haplotype/code and Elongation Factor-1α code (EF-1α).

Location (Collector)	COI Haplotype/Code	COII Haplotype/Code	Cyt b Haplotype/Code	EF-1α Code	GPS Coordinates
Portugal					
Quinta da Ranca, Vinhais, Trás-os-Montes (Seabra)	H27/Ranca_1	H7/Ranca_1	H9/Ranca_1		41°48'48.80"N; 6°59'44.51"W
Viana do Castelo, Minho (Seabra)	H30				41°42'40.10"N; 8°51'20.95"W
Parque Nacional Peneda Gerês, Minho (Rodrigues, Silva, Marabuto)	H29				41°43'44.34"N; 8° 9'46.46"W
São Jacinto, Aveiro (Seabra)	H29				40°40'36.19"N; 8°43'13.53"W
São Jacinto, Aveiro (Marabuto)	H29/S.Jacinto_3	H3/S.Jacinto_3	H9/S.Jacinto_3		40°40'36.19"N; 8°43'13.53"W
Bom Sucesso, Foz do Arelho (Quartau, Seabra & Penado)	H37				39°25'2.95"N; 9°13'39.18"W
Bom Sucesso, Foz do Arelho (Quartau, Seabra & Penado)	H36				39°25'2.95"N; 9°13'39.18"W
Bom Sucesso, Foz do Arelho (Rodrigues, Marabuto, Pereira & Seabra)	H29/ F. Arelho_8	H10/ F. Arelho_8	H9/ F. Arelho_8		39°25'2.95"N; 9°13'39.18"W
Covão da Ametade, Serra da Estrela (Seabra)	H29				40°19'33.85"N; 7°34'19.08"W
Santa Combadão, Serra da Estrela (Marabuto)	H29/S.Estrela_3	H3/S.Estrela_3	H9/S.Estrela_3		40°23'51.71"N; 8° 7'49.97"W
Olival, Ourém (Quartau)	H29				39°42'38.51"N; 8°36'4.39"W
Olival, Ourém (Quartau)	H37/ Olival_6	H9/ Olival_6	H9/ Olival_6		39°42'38.51"N; 8°36'4.39"W
Évora de Alcobaça, Alcobaça (Nunes)	H37				39°30'55.84"N; 8°58'20.40"W
Évora de Alcobaça, Alcobaça (Nunes)	H29/ Alcobaça_3	Alcobaça_3	Alcobaça_3		39°30'55.84"N; 8°58'20.40"W
Évora de Alcobaça, Alcobaça (Nunes)	H37/ Alcobaça_4	H3/ Alcobaça_4	H9/ Alcobaça_4		39°30'55.84"N; 8°58'20.40"W
Évora de Alcobaça, Alcobaça (Nunes)	H29				39°30'55.84"N; 8°58'20.40"W
Serra d' Aire (Seabra)	H21/S.Aire_5	H4/S.Aire_5	H9/S.Aire_5		39°31'56.98"N; 8°31'52.56"W
Serra d' Aire (Seabra)	H37				39°31'56.98"N; 8°31'52.56"W
Serra Montejunto (Marabuto)	H29/Montejunto_2	H3/Montejunto_2	H9/Montejunto_2		39°10'24.53"N; 9° 2'55.66"W
Gouveia, Sintra (Quartau, Seabra & Penado)	H36				38°50'15.75"N; 9°25'20.77"W
Fontanelas, Sintra (Rodrigues, Seabra & Pereira)	H29/ Sintra_10	H3/ Sintra_10	H9/ Sintra_10	Sintra_10	38°50'15.75"N; 9°25'20.77"W
Arrábida, Setúbal (Fonseca)	H29			Arrabida_5	38°28'20.09"N; 8°59'41.77"W
Arrábida, Setúbal (Fonseca)	H37/Arrabida_6	H8/Arrabida_6	H6/Arrabida_6		38°28'20.09"N; 8°59'41.77"W
Vale do Gaio, Torrão, Alentejo (Seabra)	H37				38°14'57.66"N; 8°17'49.32"W
Grândola, Alentejo (Seabra)	H29				38°10'5.83"N; 8°36'34.40"W
Grândola, Alentejo (Seabra)	H22				38°10'5.83"N; 8°36'34.40"W
Grândola, Alentejo (Quartau & Seabra)	H38/Grandola_1	H8/Grandola_1	H6/Grandola_1		38°10'5.83"N; 8°36'34.40"W
Grândola, Alentejo (Quartau & Seabra)	H26				38°10'5.83"N; 8°36'34.40"W
Beringel, Ferreira do Alentejo (Quartau & Simões)	H29/F.Alentejo_1	H3/F.Alentejo_1	H4/F.Alentejo_1		38° 3'23.09"N; 7°59'5.45"W
Beja, Alentejo (Marabuto)	H37				38° 5'22.11"N; 7°58'26.53"W
Sines, Alentejo (Rodrigues & Marabuto)	H29/Sines_1	H3/Sines_1	H4/Sines_1		37°57'58.98"N; 8°52'25.14"W
Cercal, Alentejo (Seabra)	H40				37°46'39.12"N; 8°39'46.80"W
Ribeira do Torgal, Odemira, Alentejo (Ribeiro & Pires)	H20				37°39'35.14"N; 8°37'40.86"W
Ribeira do Torgal, Odemira, Alentejo (Ribeiro & Pires)	H29				37°39'35.14"N; 8°37'40.86"W
Ribeira do Torgal, Odemira, Alentejo (Ribeiro & Pires)	H29				37°39'35.14"N; 8°37'40.86"W

Table S1 - Analysed samples of *Philaenus spumarius* with description of the sampling locations and indication of the corresponding mtDNA Cytochrome *c* oxidase I (COI), Cytochrome *c* oxidase II (COII), Cytochrome *b* haplotype/code and Elongation Factor-1α code (EF-1α) (*cont.*).

Location (Collector)	COI Haplotype/Code	COII Haplotype/Code	Cyt b Haplotype/Code	EF-1α Code	GPS Coordinates
Portugal					
Santa Clara a Velha, Rio Mira, Alentejo (Ribeiro)	H40/R.Mira_1	H8/R.Mira_1	H8/R.Mira_1		37°30'38.22"N; 8°28'27.23"W
Santa Clara a Velha, Rio Mira, Alentejo (Ribeiro)	H39/Mira_1	H8/Mira_1	H6/Mira_1		37°30'38.22"N; 8°28'27.23"W
Santa Clara a Velha, Rio Mira, Alentejo (Ribeiro)	H40/Mira_2	H8/Mira_2	H7/Mira_2		37°30'38.22"N; 8°28'27.23"W
Santa Clara a Velha, Rio Mira, Alentejo (Rodrigues & Marabuto)	H29/Mira_3	H3/Mira_3	H4/Mira_3		37°30'38.22"N; 8°28'27.23"W
Santa Clara a Velha, Rio Mira, Alentejo (Rodrigues & Marabuto)	H37	H8	H5		37°30'38.22"N; 8°28'27.23"W
Aljezur, Alentejo (Rodrigues & Marabuto)	H41/Aljejur_9.1	H8/Aljejur_9.1	H7/Aljejur_9.1		37°18'11.22"N; 8°47'59.88"W
Silves, Algarve (Seabra)	H29				37°11'40.31"N; 8°27'55.28"W
Silves, Algarve (Seabra)	H29				37°11'40.31"N; 8°27'55.28"W
Silves, Algarve (Seabra)	H29/Silves_3	H3/Silves_3	Silves_3		37°11'40.31"N; 8°27'55.28"W
Silves, Algarve (Seabra)	H29				37°11'40.31"N; 8°27'55.28"W
Silves, Algarve (Seabra)	H29				37°11'40.31"N; 8°27'55.28"W
Silves, Algarve (Seabra)	H29				37°11'40.31"N; 8°27'55.28"W
Barranco do Velho, Algarve (Quartau)	H29				37° 8'15.13"N; 8° 1'24.27"W
São Miguel, Açores (Borges)	H24/ Azores_1	H3/ Azores_1	H13/ Azores_1	Azores_1	37°47'46.86"N; 25°11'4.50"W
São Miguel, Açores (Borges)	H24				37°47'46.86"N; 25°11'4.50"W
São Miguel, Açores (Borges)	H71				37°47'46.86"N; 25°11'4.50"W
São Miguel, Açores (Borges)	H24				37°47'46.86"N; 25°11'4.50"W
São Miguel, Açores (Borges)	H24				37°47'46.86"N; 25°11'4.50"W
Finland					
Tvarminne (Halkka)	H2/ Tvarminne_4	H2/ Tvarminne_4	H1/ Tvarminne_4		59°50'37.60"N; 23°14'21.68"E
Tvarminne (Halkka)	H16				59°50'37.60"N; 23°14'21.68"E
Tvarminne (Halkka)	H16				59°50'37.60"N; 23°14'21.68"E
Tvarminne (Halkka)	H16				59°50'37.60"N; 23°14'21.68"E
Norra Grisselgrundet (Halkka)	Grisselgrundet_4	Grisselgrundet_4	Grisselgrundet_4		59°50'17.48"N; 23°14'46.87"E
Norra Grisselgrundet (Halkka)	H16				59°50'17.48"N; 23°14'46.87"E
Windskar (Halkka)	H2				59°49'40.56"N; 23°12'42.74"E
Brannskar (Halkka)	H5/ Brannskar_7	H2/ Brannskar_7	H3/ Brannskar_7		59°50'39.51"N; 23°16'22.50"E
Brannskar (Halkka)	H15				59°50'39.51"N; 23°16'22.50"E
Brannskar (Halkka)	H15				59°50'39.51"N; 23°16'22.50"E
Brannskar (Halkka)	H16				59°50'39.51"N; 23°16'22.50"E
Segelskar (Halkka)	H8/ Segelskar_4	H2/ Segelskar_4	H3/ Segelskar_4		59°45'52.22"N; 23°22'24.53"E
Segelskar (Halkka)	H16				59°45'52.22"N; 23°22'24.53"E
Punkalaidun (Halkka)	H1				61° 7'0.75"N; 23° 5'35.54"E
Punkalaidun (Halkka)	H4				61° 7'0.75"N; 23° 5'35.54"E
Punkalaidun (Halkka)	H7				61° 7'0.75"N; 23° 5'35.54"E
Punkalaidun (Halkka)	H12				61° 7'0.75"N; 23° 5'35.54"E
Haaparnaki-Keuruu (Halkka)	H4				62°15'32.75"N; 24°42'28.08"E
Haaparnaki-Keuruu (Halkka)	H7				62°15'32.75"N; 24°42'28.08"E

Table S1 - Analysed samples of *Philaenus spumarius* with description of the sampling locations and indication of the corresponding mtDNA Cytochrome *c* oxidase I (COI), Cytochrome *c* oxidase II (COII), Cytochrome *b* haplotype/code and Elongation Factor-1α code (EF-1α) (*cont.*).

Location (Collector)	COI Haplotype/Code	COII Haplotype/Code	Cyt b Haplotype/Code	EF-1α Code	GPS Coordinates
Turkey					
Cerkes (Yurtserver)	Cerkes_3	H1/ Cerkes_3	H2/ Cerkes_3		40°48'59.24"N; 32°54'9.50"E
Cerkes (Yurtserver)	H4				40°48'59.24"N; 32°54'9.50"E
Cerkes (Yurtserver)	H6				40°48'59.24"N; 32°54'9.50"E
Cerkes (Yurtserver)	H6				40°48'59.24"N; 32°54'9.50"E
Cerkes (Yurtserver)	H11				40°48'59.24"N; 32°54'9.50"E
Cerkes (Yurtserver)	H14			Turkey_3	40°48'59.24"N; 32°54'9.50"E
Cerkes (Yurtserver)	H14				40°48'59.24"N; 32°54'9.50"E
Azdavay (Yurtserver)	H57/Azdavay_1	H11/Azdavay_1	H16/Azdavay_1		41°38'29.26"N; 33°17'52.54"E
Keçan (North) (Yurtserver)	H57/ Keçan_N1	H13/Keçan_N1	H17/Keçan_N1	Keçan_N1	40°53'29.60"N; 26°38'42.81"E
Keçan (North) (Yurtserver)	H64				40°53'29.60"N; 26°38'42.81"E
Keçan (South) (Yurtserver)	H57/Keçan_S1	H12/Keçan_S1	H16/Keçan_S1		40°44'41.83"N; 26°36'6.98"E
Keçan (South) (Yurtserver)	H64				40°44'41.83"N; 26°36'6.98"E
Lapseki (Yurtserver)	H46/ Lapseki_1	H12/ Lapseki_1	H16/ Lapseki_1		40°19'3.46"N; 26°43'46.86"E
Lapseki (Yurtserver)	H57				40°19'3.46"N; 26°43'46.86"E
Lapseki (Yurtserver)	H59				40°19'3.46"N; 26°43'46.86"E
Kucukuyu (Yurtserver)	H55/ Kucukuyu_1	H12/ Kucukuyu_1	H16/ Kucukuyu_1		39°36'35.48"N; 26°33'2.11"E
Kucukuyu (Yurtserver)	H57				39°36'35.48"N; 26°33'2.11"E
Kucukuyu (Yurtserver)	H67				39°36'35.48"N; 26°33'2.11"E
Suloglu (Yurtserver)	H57/ Suloglu_1	H14/ Suloglu_1	H16/ Suloglu_1		41°46'52.25"N; 26°53'1.74"E
Suloglu (Yurtserver)	H57				41°46'52.25"N; 26°53'1.74"E
Suloglu (Yurtserver)	H62				41°46'52.25"N; 26°53'1.74"E
Pabucdere (Yurtserver)	H50/ Pabucdere_1	H14/ Pabucdere_1	H16/ Pabucdere_1		41°38'18.49"N; 27°54'54.61"E
Pabucdere (Yurtserver)	H57				41°38'18.49"N; 27°54'54.61"E
Pabucdere (Yurtserver)	H64				41°38'18.49"N; 27°54'54.61"E
Demirkoy (Yurtserver)	H54/ Demirkoy_1	H12/ Demirkoy_1	H18/ Demirkoy_1		41°52'34.74"N; 27°46'8.04"E
Demirkoy (Yurtserver)	H57				41°52'34.74"N; 27°46'8.04"E
Italy					
Tardaria, Etna, Sicily (d'Urso)			H4		37°43'53.05"N; 14°59'8.12"E
Tardaria, Etna, Sicily (d'Urso)	H29				37°43'53.05"N; 14°59'8.12"E
Bosco di Acisantantonio, Etna, Sicily (d'Urso)	H49/Sicily_5	H12/Sicily_5	Sicily_5		37°41'16.21"N; 15° 9'30.01"E
Torcello, Veneze (Quartau)	H47				45°29'47.21"N; 12°25'11.44"E
Via Appia, Rome (Quartau)	H53				41°44'33.26"N; 12°42'44.66"E
Emilia, Parma (Quartau)	H32				44°48'12.33"N; 10°20'45.45"E
Emilia, Parma (Quartau)	H32				44°48'12.33"N; 10°20'45.45"E
Emilia, Parma (Quartau)	H32				44°48'12.33"N; 10°20'45.45"E
Emilia, Parma (Quartau)	H32				44°48'12.33"N; 10°20'45.45"E
Emilia, Parma (Quartau)	H32				44°48'12.33"N; 10°20'45.45"E
Emilia, Parma (Quartau)	H32				44°48'12.33"N; 10°20'45.45"E
Emilia, Parma (Quartau)	H33			Italia_2	44°48'12.33"N; 10°20'45.45"E
Emilia, Parma (Quartau)	H51				44°48'12.33"N; 10°20'45.45"E

Table S1 - Analysed samples of *Philaenus spumarius* with description of the sampling locations and indication of the corresponding mtDNA Cytochrome *c* oxidase I (COI), Cytochrome *c* oxidase II (COII), Cytochrome *b* haplotype/code and Elongation Factor-1α code (EF-1α) (*cont.*).

Location (Collector)	COI Haplotype/Code	COII Haplotype/Code	Cyt b Haplotype/Code	EF-1α Code	GPS Coordinates
Italy					
Florence, Toscana (Marabuto)	H49				43°49'21.87"N; 11°21'9.66"E
Mt. Calvi, Toscana (Marabuto)	H53				43° 5'9.00"N; 10°36'28.00"E
Greve in Chienti, Toscana (Marabuto)	H49				43°35'2.00"N; 11°18'50.00"E
Bologna (Marabuto)	H52				44° 2'45.00"N; 11°17'44.00"E
Villaromagnana, Alps (Lessio)	H32				44°50'59.56"N; 8°53'25.52"E
United Kingdom					
Cardiff, Wales (Wilson)	H24				51°28'57.16"N; 3°11'0.41"W
Cardiff, Wales (Wilson)	H24				51°28'57.16"N; 3°11'0.41"W
Aberdare, Wales (Wilson)	H13			UK_7	51°42'49.13"N; 3°26'43.27"W
Aberdare, Wales (Wilson)	H13				51°42'49.13"N; 3°26'43.27"W
Aberdare, Wales (Wilson)	H13				51°42'49.13"N; 3°26'43.27"W
Aberdare, Wales (Wilson)	H24				51°42'49.13"N; 3°26'43.27"W
Cambridge, England (Borges)	H23/Cambridge_1	H3/Cambridge_1	H12/Cambridge_1		52°12'22.80"N; 0° 7'29.41"E
Cambridge, England (Borges)	H23/Cambridge_2	H3/Cambridge_2	H12/Cambridge_2		52°12'22.80"N; 0° 7'29.41"E
Cambridge, England (Borges)	H24				52°12'22.80"N; 0° 7'29.41"E
Oxfordshire, England (Corlev)	H24				51°39'30.43"N; 1°35'5.11"W
Canada					
Burnaby Mt. British Columbia (Beckenbach)	H24				53°43'36.00"N; 127°38'51.43"W
Burnaby Mt. British Columbia (Beckenbach)	H24/Canada_3	H3/Canada_3	H12/Canada_3		53°43'36.00"N; 127°38'51.43"W
Burnaby Mt. British Columbia (Beckenbach)	H70				53°43'36.00"N; 127°38'51.43"W
Burnaby Mt. British Columbia (Beckenbach)	Canada_5	Canada_5	Canada_5		53°43'36.00"N; 127°38'51.43"W
United States of America					
Michigan (Fonseca)	H25			USA_2	44°18'49.13"N; 85°35'6.26"W
Michigan (Fonseca)	H25				44°18'49.13"N; 85°35'6.26"W
Michigan (Fonseca)	H25/USA_4	H5/USA_4	H4/USA_4		44°18'49.13"N; 85°35'6.26"W
Michigan (Fonseca)	H25				44°18'49.13"N; 85°35'6.26"W
Michigan (Fonseca)	H26/USA_6	H3/USA_6	H11/USA_6		44°18'49.13"N; 85°35'6.26"W
Wonalancet – New Hampshire (Thompson)	H1				43°53'58.93"N; 71°21'33.47"W
Wonalancet – New Hampshire (Thompson)	H3				43°53'58.93"N; 71°21'33.47"W
Wonalancet – New Hampshire (Thompson)	H9				43°53'58.93"N; 71°21'33.47"W
Wonalancet – New Hampshire (Thompson)	H10				43°53'58.93"N; 71°21'33.47"W
Spain					
C. R. Fauna et Ed. Amb. "Los Hornos" (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H30				39°26'14.10"N; 6°17'17.46"W
Valdenoches, Castille La Mancha (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H56/Spain_3.1	H12/Spain_3.1	H15/Spain_3.1		40°41'11.03"N; 3° 5'18.87"W
Caspe, Aragon (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H56/Spain_4.1	H12/Spain_4.1	H15/Spain_4.1		41°21'41.30"N; 0° 6'17.83"W
Tolva, Aragon (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H26				42° 6'42.89"N; 42° 6'42.89"N
El Pont de Suert, Catalunha (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H35				42°24'46.98"N; 0°44'22.89"E
Forgais de Montclús, Catalunha (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H19				41°43'42.20"N; 2°26'13.21"E
Chert, Valencia (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H56				40°30'55.36"N; 0° 8'27.38"E

Table S1 - Analysed samples of *Philaenus spumarius* with description of the sampling locations and indication of the corresponding mtDNA Cytochrome *c* oxidase I (COI), Cytochrome *c* oxidase II (COII), Cytochrome *b* haplotype/code and Elongation Factor-1α code (EF-1α) (*cont.*).

Location (Collector)	COI Haplotype/Code	COII Haplotype/Code	Cyt b Haplotype/Code	EF-1α Code	GPS Coordinates
Spain					
San Agustin, Aragon (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H56				40° 8'24.18"N; 0°43'0.50"W
Cheste, Valencia (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H56/Spain_22.1	H12/Spain_22.1	H15/Spain_22.1		39°31'9.33"N; 0°40'8.49"W
Vélez-Rubio, Andalusia (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H29				37°38'49.42"N; 2° 5'13.65"W
Ronda, Andalusia (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H29				36°43'52.11"N; 5°10'55.79"W
El Gastor, Andalusia (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H29				36°50'16.20"N; 5°20'51.02"W
Valdés, Asturias (Rodrigues, Silva & Nunes)	H34/Spain_1	H3/Spain_1	H10/Spain_1		43°32'49.08"N; 6°31'28.92"W
Meira, Galiza (Rodrigues, Silva & Nunes)	H34/Spain_2	H3/Spain_2	H9/Spain_2	Spain_2	43°13'44.40"N; 7°17'26.64"W
France					
Saint-Antonin-Noble-Val (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H19				44° 9'29.52"N; 1°43'16.45"E
Lautrec (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H29				43°42'47.14"N; 2° 7'6.04"E
Fitou (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	H19				42°52'49.11"N; 2°59'34.12"E
Aube, Lusigny-sur-Barse (Constant)	H17				48°15'7.19"N; 4°16'7.76"E
Aube, Lusigny-sur-Barse (Constant)	H29				48°15'7.19"N; 4°16'7.76"E
Belgium					
Namur (Constant)	H16/ Belgium_1	H2/ Belgium_1	H2/ Belgium_1	Belgium_1	50°27'55.51"N; 4°52'3.33"E
Namur (Constant)	H19				50°27'55.51"N; 4°52'3.33"E
Namur (Constant)	H19				50°27'55.51"N; 4°52'3.33"E
Namur (Constant)	H29			Belgium_4	50°27'55.51"N; 4°52'3.33"E
Morocco					
Near Rabat (Rodrigues, Silva, Marabuto & Ferreira)	H28				33°46'37.56"N; 7°13'58.92"W
Rabat (Rodrigues, Silva, Marabuto & Ferreira)	H29				34° 0'41.70"N; 6°42'32.94"W
Rif (Rodrigues, Silva, Marabuto & Ferreira)	H29			Morocco_6.1	35°51'58.14"N; 5°24'30.30"W
Ifrane (Rodrigues, Silva, Marabuto & Ferreira)	H37				33°30'41.82"N; 5° 5'34.08"W
Ifrane Centre (Rodrigues, Silva, Marabuto & Ferreira)	H37				33°31'58.86"N; 5° 6'7.86"W
Azrou (Rodrigues, Silva, Marabuto & Ferreira)	H29				33°26'57.78"N; 5°13'55.14"W
Azrou (Rodrigues, Silva, Marabuto & Ferreira)	H29				33°29'39.48"N; 5°15'47.46"W
Greece					
Bralou (Rodrigues, Silva, Marabuto)	H57				38°44'35.70"N; 22°26'55.50"E
Iti National Park (Rodrigues, Silva, Marabuto)	H63				38°44'6.96"N; 22°22'10.20"E
Mt Olympus (Rodrigues, Silva, Marabuto)	H43				40° 6'22.32"N; 22°27'33.66"E
Mt Olympus (Rodrigues, Silva, Marabuto)	H68				40° 5'35.94"N; 22°25'10.62"E
Mt Olympus (Rodrigues, Silva, Marabuto)	H18				40° 5'3.12"N; 22°24'25.26"E
Meteora (Rodrigues, Silva, Marabuto)	H57				39°43'11.76"N; 21°38'14.28"E

Table S1 - Analysed samples of *Philaenus spumarius* with description of the sampling locations and indication of the corresponding mtDNA Cytochrome *c* oxidase I (COI), Cytochrome *c* oxidase II (COII), Cytochrome *b* haplotype/code and Elongation Factor-1α code (EF-1α) (*cont.*).

Location (Collector)	COI Haplotype/Code	COII Haplotype/Code	Cyt b Haplotype/Code	EF-1α Code	GPS Coordinates
Greece					
Mt Vourinos (Rodrigues, Silva, Marabuto)	H66				40° 6'28.50"N; 21°40'21.78"E
Mt Vourinos (Rodrigues, Silva, Marabuto)	H57				40° 9'26.94"N; 21°43'36.36"E
Mt Vourinos (Rodrigues, Silva, Marabuto)	H65				40°12'1.68"N; 21°39'30.00"E
Mt Giona (Rodrigues, Silva, Marabuto)	H57				38°40'29.40"N; 22°18'31.62"E
Mt Giona (Rodrigues, Silva, Marabuto)	H44				38°40'6.24"N; 22°18'31.62"E
Mt Parnassus (Rodrigues, Silva, Marabuto)	H57				38°36'50.16"N; 22°34'33.00"E
Mt Parnassus (Rodrigues, Silva, Marabuto)	H46				38°34'42.24"N; 22°34'30.30"E
Mt Parnassus (Rodrigues, Silva, Marabuto)	H57				38°37'23.94"N; 22°33'12.96"E
Mt Parnassus (Rodrigues, Silva, Marabuto)	H69				38°33'19.98"N; 22°34'44.04"E
Mt Menalo (Rodrigues, Silva, Marabuto)	H45				37°37'50.64"N; 22°19'34.56"E
Mt Menalo (Rodrigues, Silva, Marabuto)	H57				37°38'26.10"N; 22°16'1.62"E
Mt Menalo (Rodrigues, Silva, Marabuto)	H60				37°40'49.38"N; 22°13'46.26"E
Mt Menalo (Rodrigues, Silva, Marabuto)	H58				37°37'37.32"N; 22°17'39.18"E
Mt Taygetus (Rodrigues, Silva, Marabuto)	H61				36°53'20.10"N; 22°21'4.68"E
Mt Parnonas (Rodrigues, Silva, Marabuto)	H57				37°11'13.44"N; 22°33'30.96"E
Slovenia					
Dragonja (Derlink)	H48			Slovenia_1	45°27'18.00"N; 13°42'4.68"E
Dragonja (Derlink)	H48				45°27'18.00"N; 13°42'4.68"E
Gorice pri Famljah (Derlink)	H48				45°40'21.05"N; 14° 0'52.99"E
Bulgaria					
Petrich (Paulo)	H57				41°25'60.00"N; 23° 1'0.00"E
New Zealand					
Lincoln, South Island (Yurtserver)	H14 FigS3				43°38'40.41"S; 172°28'9.98"E
Lincoln, South Island (Yurtserver)	H14 FigS3				43°38'40.41"S; 172°28'9.98"E
Lincoln, South Island (Yurtserver)	H14 FigS3				43°38'40.41"S; 172°28'9.98"E

Table S2 – Haplotype distribution within *P. spumarius* geographic regions for mitochondrial gene COI. The total number of haplotypes per geographic region and the total number of individuals per haplotype are also shown. Western Europe: Belgium and France; Italy: Italian peninsula and Sicily.

Geographic Regions	Haplotypes																																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Azores																								4												
Iberian Peninsula																			1	1	1	1				2	1		28	1	1			2	1	2
Morocco																												1	4							
Western Europe																1	1		4										3							
United Kingdom													3										2	5												
Italy																													1			7	1			
Slovenia																																				
Balkans																		1																		
Anatolian Peninsula				1		2					1			2																						
Finland	1	2		2	1		2	1				1			2	5																				
North America	1		1						1	1														2	4	1										
Total	2	2	1	3	1	2	2	1	1	1	1	1	3	2	2	6	1	1	5	1	1	1	2	11	4	3	1	1	36	1	1	7	1	2	1	2

Geographic Regions	Haplotypes																																		
	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	Total
Azores																																			1	2
Iberian Peninsula	9	1	1	3	1	1														5																19
Morocco	2																																			3
Western Europe																																				4
United Kingdom																																				3
Italy											1		3		1	1	2																			8
Slovenia												3																								1
Balkans							1	1	1	2				1				1	1		17	1	1	1	1	1	1	3	1	1	1	1	1			20
Anatolian Peninsula																					1															6
Finland																																				9
North America																																		1		8
Total	11	1	1	3	1	1	1	1	1	2	1	3	3	1	1	1	2	1	1	5	18	1	1	1	1	1	1	3	1	1	1	1	1	1	1	

Table S3 – Analyses of molecular variance (AMOVA) among regions of *P. spumarius* based on COI data.

Source of variation	d.f.	Sum of squares	Variance of components	% of variation	Fixation Indices (Φ)
**[North America][Azores][North Africa][Northern Europe] [Western Europe] [South-western Europe][Central Europe][South-eastern Europe][South-western Asia]*					
Among groups	8	196.948	0.73116 Va	27.73	Φ_{CT} 0.2773
Among regions within groups	4	29.265	0.66613 Vb	25.26	Φ_{SC} 0.2526
Within regions	177	219.194	1.23930 Vc	**47.00**	**Φ_{ST} 0.4700**
Total	189	445.568	2.63658		
***[North America][Azores][North Africa][Northern Europe] [United Kingdom] [Southern Europe][Central Europe][South-western Asia]*					
Among groups	7	155.904	0.77953 Va	27.59	Φ_{CT} 0.2759
Among regions within groups	5	70.309	0.80628 Vb	28.54	Φ_{SC} 0.2854
Within regions	177	219.356	1.23930 Vc	**43.87**	**Φ_{ST} 0.4387**
Total	189	445.568	2.82511		

* ***North America:*** United States of America and Canada; ***North Africa:*** Morocco; ***Northern Europe:*** Finland; ***Western Europe:*** United Kingdom, France and Belgium; ***South-western Europe:*** Iberian Peninsula; ***Central Europe:*** Slovenia; ***South-eastern Europe:*** Italian Peninsula, Sicily and Balkans; ***South-western Asia:*** Anatolia Peninsula.

** ***North America:*** United States of America and Canada; ***North Africa:*** Morocco; ***Northern Europe:*** Finland; ***Southern Europe:*** Iberian Peninsula, Italian Peninsula, Sicily and Balkans; ***Central Europe:*** Belgium, France and Slovenia; ***South-western Asia:*** Anatolia Peninsula.

Table S4 - Divergence time estimates in million years (Ma) from the most recent common ancestor of each main *Philaenus spumarius* mtDNA COI haplogroup estimated using a mean mutation rate of 3.54% per million years as suggested by Papadopoulou *et al.* (2010).

	Lower 95% HPD	Mean	Upper 95% HPD
Eastern-Mediterranean	0.031	0.084	0.153
Western-Mediterranean	0.026	0.079	0.148
Eastern	0.056	0.190	0.374
Eastern vs western-Mediterranean	0.118	0.269	0.450
Eastern-Mediterranean vs western-Mediterranean	0.067	0.146	0.243
Eastern-Mediterranean vs eastern	0.120	0.270	0.447
Combined	0.122	0.270	0.448

Appendix S2

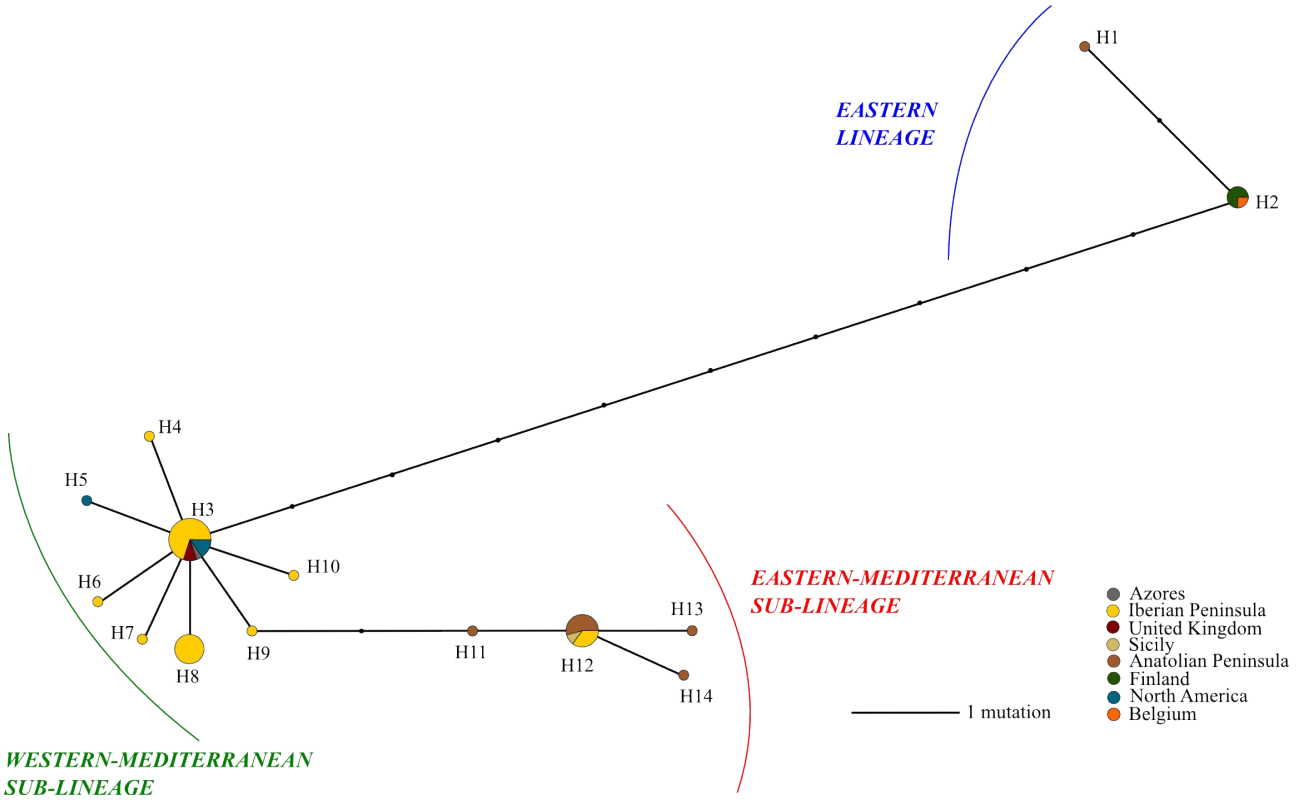


Figure S1 – Median-joining haplotype network of a set of *Philaenus spumarius* sampled geographic regions for mitochondrial gene COII (495bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their size is in proportion with the number of mutations.

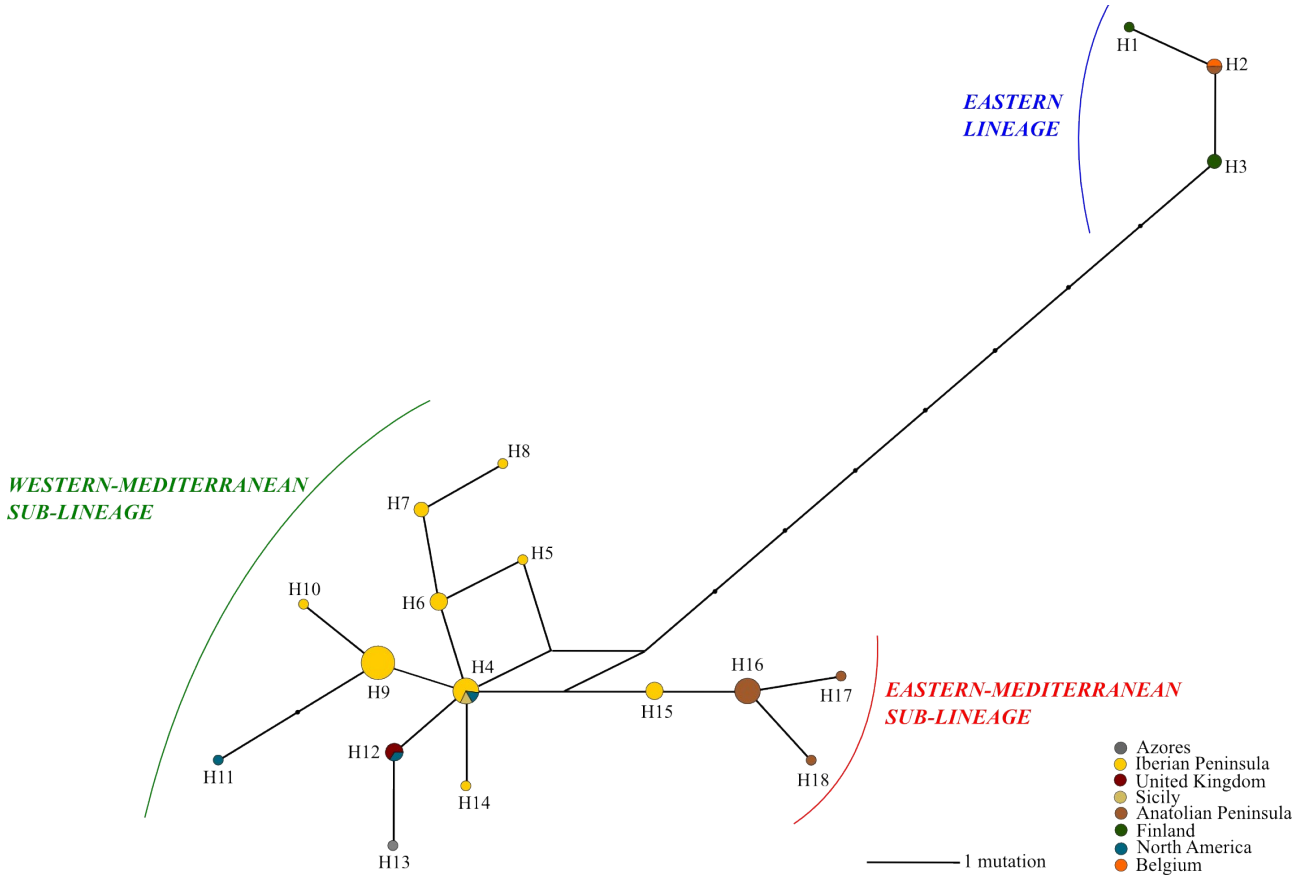


Figure S2 – Median-joining haplotype network of a set of *P. spumarius* sampled geographic regions for mitochondrial gene cyt *b* (434bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their size is in proportion with the number of mutations.

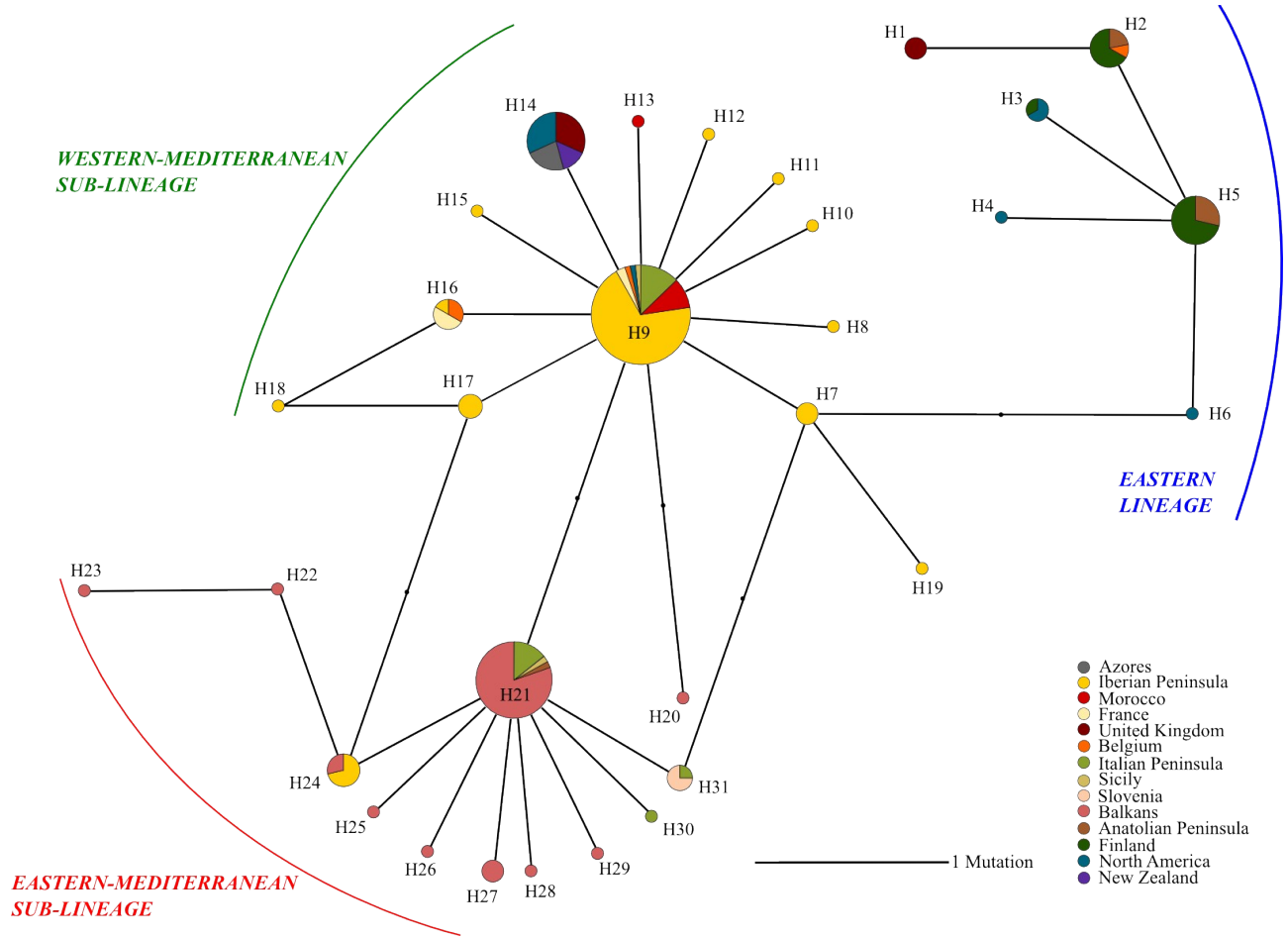


Figure S3 – Median-joining haplotype network of *P. spumarius* sampled geographic regions for mitochondrial gene COI (289bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their size is in proportion with the number of mutations.

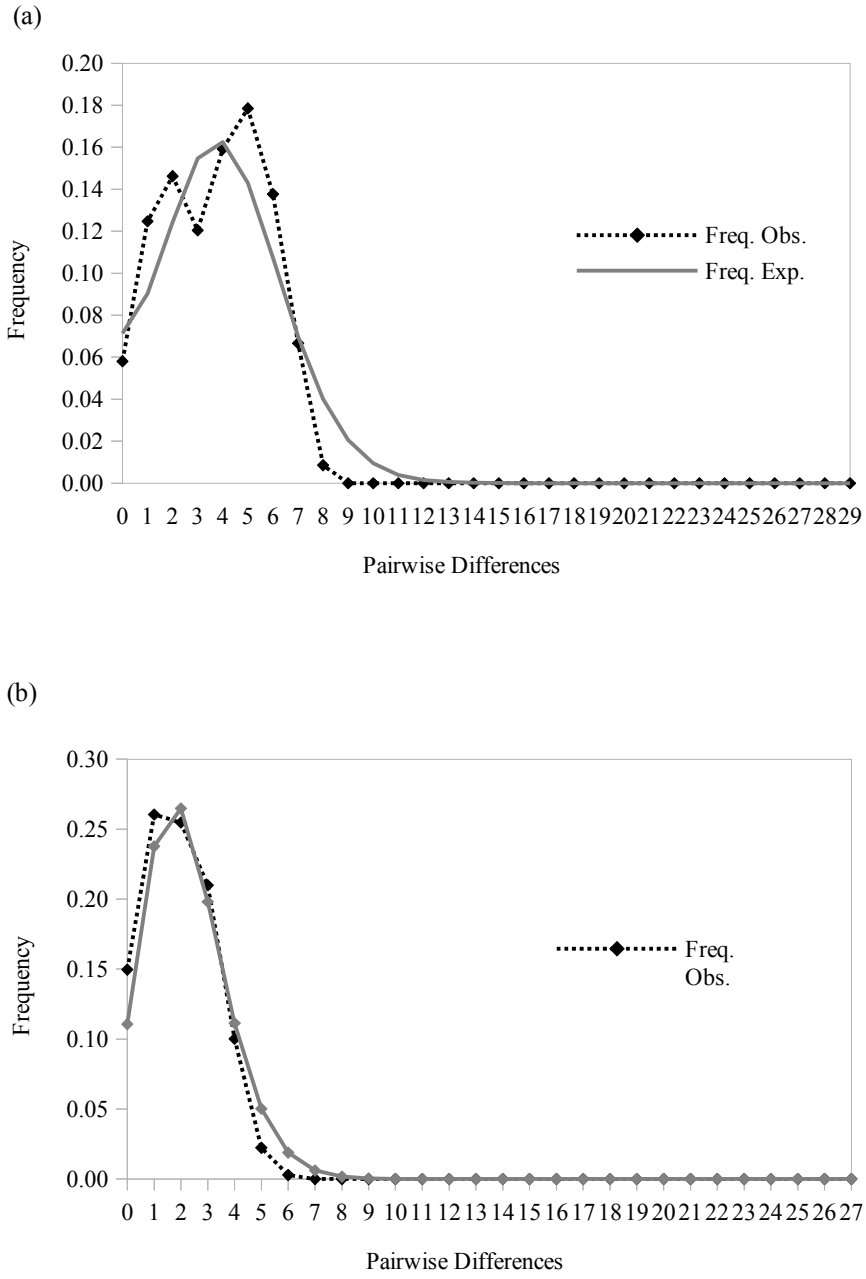


Figure S4 - Mismatch distribution of mtDNA COI haplotypes for each of the three *P. spumarius* haplogroups: (a) Eastern lineage; (b) Western-Mediterranean sub-lineage and (c) Eastern-Mediterranean sub-lineage. The expected frequency is based on a population growth-decline model, determined using DNASP and is represented by a continuous line. The observed frequency is represented by a dotted line. Parameter values for the mismatch distribution are given in Table 3.

(c)

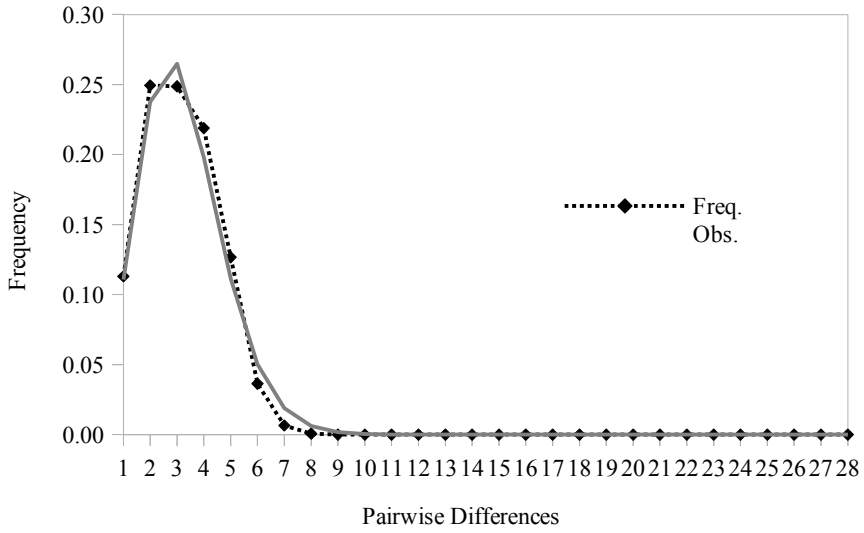


Figure S4 - Mismatch distribution of mtDNA COI haplotypes for each of the three *P. spumarius* haplogroups: (a) Eastern lineage; (b) Western-Mediterranean sub-lineage and (c) Eastern-Mediterranean sub-lineage. The expected frequency is based on a population growth-decline model, determined using DNASP and is represented by a continuous line. The observed frequency is represented by a dotted line. Parameter values for the mismatch distribution are given in Table 3 (*cont.*).

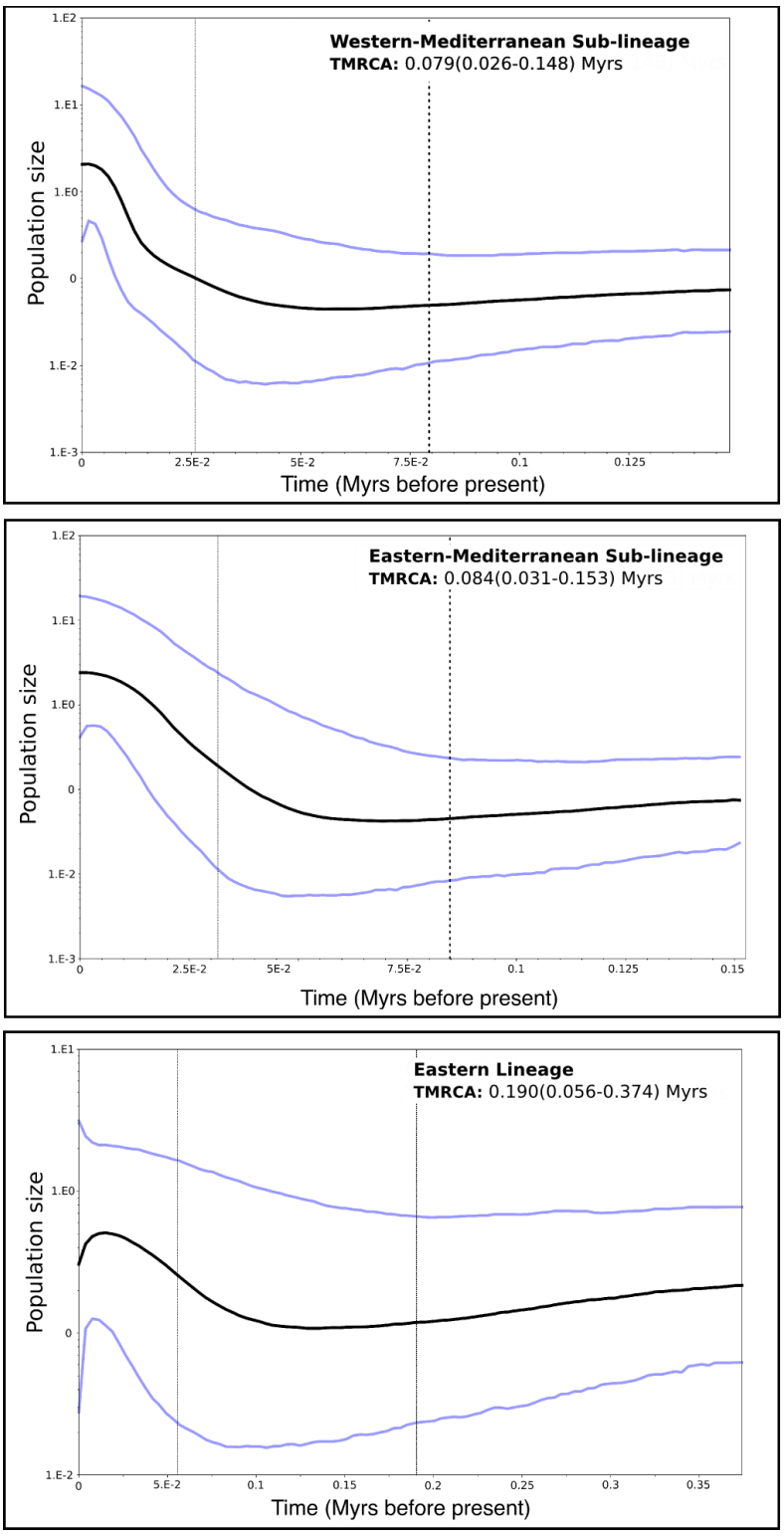


Figure S5 - Bayesian skyline plots showing the historical demographic trends for each main *Philaenus spumarius* mtDNA group detected using COI gene. Along the y-axis is the expressed population size estimated in units of Neμ (Ne: effective population size, μ: mutation rate per haplotype per generation). The y-axis is in a log-scale. Solid lines represent median estimates and blue lines represent the 95% high probability density (HPD) intervals.

CHAPTER 3

Differential survival and reproduction in colour forms of *Philaenus spumarius* (Hemiptera, Aphrophoridae) give new insights to the study of its balanced polymorphism

Sara E. Silva[1*], Ana S.B. Rodrigues[1*], Eduardo Marabuto[1,2], Selçuk Yurtsever[3], Paulo A.V. Borges[4], José A. Quartau[1], Octávio S. Paulo[1], Sofia G. Seabra[1]

[1] CE3C - Centre for Ecology, Evolution and Environmental Changes, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal.

[2] CESAM - Centro de Estudos do Ambiente e do Mar, Departamento de Biologia, Universidade de Aveiro, 3810-193 Aveiro, Portugal.

[3] Biology Department, Science Faculty, Trakya University, 22030 Edirne, Turkey.

[4] CE3C – Centre for Ecology, Evolution and Environmental Changes / Azorean Biodiversity Group and Universidade dos Açores - Departamento de Ciências Agrárias, 9700-042 Angra do Heroísmo, Açores, Portugal.

∗These authors contributed equally.

Ecological Entomology (2015) 40, 759-766

Ecological Entomology

Ecological Entomology (2015), DOI: 10.1111/een.12252

Differential survival and reproduction in colour forms of *Philaenus spumarius* give new insights to the study of its balanced polymorphism

SARA E. SILVA,[1,*] ANA S. B. RODRIGUES,[1,*] EDUARDO MARABUTO,[1,2] SELÇUK YURTSEVER,[3] PAULO A. V. BORGES,[4] JOSÉ A. QUARTAU,[1] OCTÁVIO S. PAULO[1] and SOFIA G. SEABRA[1] [1]cE3c – Centre for Ecology, Evolution and Environmental Changes, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal, [2]CESAM – Centro de Estudos do Ambiente e do Mar, Departamento de Biologia, Universidade de Aveiro, Aveiro, Portugal, [3]Biology Department, Science Faculty, Trakya University, Edirne, Turkey and [4]cE3c – Centre for Ecology, Evolution and Environmental Changes / Azorean Biodiversity Group and Universidade dos Açores – Departamento de Ciências Agrárias, Angra do Heroísmo, Açores, Portugal

Abstract. 1. Colour polymorphisms are common across animals and are often the result of complex selection regimes. *Philaenus spumarius* (Linnaeus) (Hemiptera, Aphrophoridae) shows a widely studied dorsal colour polymorphism with several described phenotypes whose variation in their occurrence and frequency, as well as their maintenance across time, have been reported. Several selective influences have been suggested to play a role, but the mechanisms underlying the maintenance of this polymorphism are still poorly understood.

2. To explore the adaptive significance of the colour polymorphism of *P. spumarius*, an experiment was conducted in captivity under semi-natural conditions to measure survival, reproductive success, and duration of egg maturation.

3. It was found that there was higher longevity, a higher number of oviposition events, and a higher number of eggs laid by *trilineatus* phenotype females than by *typicus* and *marginellus*, supporting previous reports of an increase in *trilineatus* frequency during the season. The duration of egg maturation did not differ among phenotypes.

4. The higher longevity and fertility of the *trilineatus* phenotype may compensate, for example, the higher rate of attack by parasitoids and/or higher solar radiation reflectance in this phenotype, which have already been reported in previous studies, constituting a possible mechanism for the maintenance of this polymorphism.

Key words. Colour polymorphism, meadow spittlebug, reproduction, survival.

Introduction

Animal colour polymorphisms have been a major source for understanding processes affecting adaptation and eco-evolutionary dynamics and the understanding of how these occur and are maintained is one of the central problems in evolutionary biology (e.g. Jones *et al.*, 1977; Oxford, 1985; Grant, 1999). Variation within patterning and colour is known to influence fitness through processes which include thermoregulation (Phifer-Rixey *et al.*, 2008) and inter-/intra-specific communication (e.g. camouflage, Bush *et al.*, 2008; sexual selection, Maan & Cummings, 2008). Several maintenance mechanisms have been described, namely: negative frequency-dependent selection, where rare phenotypes have a fitness advantage over common phenotypes, for example as a strategy to avoid predators or to lower sexual conflict intensity (Ayala & Campbell, 1974; Punzalan *et al.*, 2005; Svensson *et al.*, 2005; Kusche & Meyer, 2014); heterozygote advantage, where individuals with heterozygous genotypes have a higher fitness than individuals

Correspondence: Sara E. Silva, cE3c – Centre for Ecology, Evolution and Environmental Changes, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal. E-mail: saraema.silva@gmail.com

*These authors contributed equally.

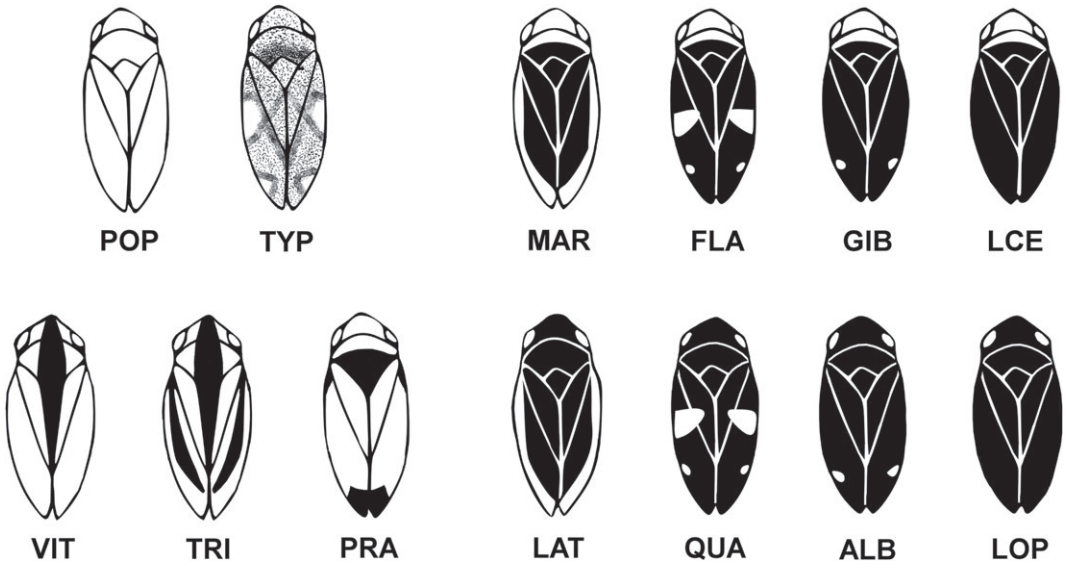


Fig. 1. Thirteen commonly referred dorsal phenotypes of *Philaenus spumarius*, including the phenotypes used in this study [*typicus* (TYP), *trilineatus* (TRI) and *marginellus* (MAR)].

with homozygous genotypes (Vercken *et al.*, 2010); divergence with gene-flow, in which different local conditions may favour different colour phenotypes, but with an exchange of migrants that maintain the polymorphism (Jones *et al.*, 1977; Oxford, 1985); alternative strategies, that result in almost the same fitness values (Roulin *et al.*, 2003).

The meadow spittlebug *Philaenus spumarius* (Linnaeus) (Hemiptera, Aphrophoridae) is a potential good model to test hypotheses about adaptation and maintenance of polymorphisms. It is an abundant, widespread and genetically diverse insect species (Rodrigues *et al.*, 2014) with a widely studied dorsal colour polymorphism (Halkka & Halkka, 1990; Stewart & Lees, 1996; Drosopoulos, 2003). More than 16 different adult dorsal colour/pattern phenotypes are known to occur throughout its distribution range (Yurtsever, 2000), 13 of which are commonly referred to, being generally divided into non-melanic and melanic forms (Halkka & Halkka, 1990; Stewart & Lees, 1996; Quartau & Borges, 1997; Yurtsever, 2000). Five are non-melanic with a pale and limited patterning: *populi* (POP), *typicus* (TYP), *vittatus* (VIT), *trilineatus* (TRI), and *praeustus* (PRA). The remaining are melanic, predominantly black or dark brown, with different combinations of pale markings: *marginellus* (MAR), *flavicollis* (FLA), *gibbus* (GIB), *leucocephalus* (LCE), *lateralis* (LAT), *quadrimaculatus* (QUA), *albomaculatus* (ALB), and *leucopthalmus* (LOP) (Fig. 1). Breeding experiments have revealed the Mendelian inheritance of this trait, which is mainly controlled by a single autosomal locus with seven alleles with complex dominance and co-dominance relationships (Halkka & Halkka, 1990; Stewart & Lees, 1996).

The non-melanic *typicus* generally predominate over the remaining phenotypes. It represents the expression of the bottom recessive allele of this pigmentation locus and its pattern can range from totally pale (*populi* phenotype) to an almost completely melanic type. In contrast, the non-melanic *trilineatus* is produced by the top dominant allele. Melanic phenotypes have other alelles and are usually expressed in females only and in lower frequencies. Some of the melanic phenotypes are rare or absent in several populations (Halkka, 1964; Whittaker, 1972; Thompson & Halkka, 1973; Honěk, 1984; Boucelham *et al.*, 1988; Quartau & Borges, 1997). This general pattern of frequency distribution concerning non-melanics and melanics usually exhibits close resemblance in many populations (Halkka & Halkka, 1990). Only a few local populations sharply deviate from this general pattern (e.g. several urban areas of southern Great Britain; see Stewart & Lees, 1996).

Remarkable geographical, clinal and local variations in the occurrence and frequency of the dorsal phenotypes of *P. spumarius* have been demonstrated (e.g. Thompson, 1984b, 1988; Berry & Willmer, 1986; Halkka & Halkka, 1990; Quartau & Borges, 1997). The association found between high melanic frequencies and cooler climates (higher latitudes, higher altitudes, and cooler habitats) suggests a possible effect of thermal selection (Thompson, 1984b, 1988). Berry and Willmer (1986) experimentally found that thermal selection is possible and verified the frequencies of the melanic patterns to be positively correlated with altitude. However, the same positive correlation was found for *typicus*, and a statistically significant negative correlation with altitude was only found for *trilineatus*. They experimentally demonstrated higher reflectance and lower temperature excesses of *trilineatus* individuals in relation to the other phenotypes, therefore, suggesting *trilineatus* as the only true non-melanic form. Several studies have reported *trilineatus* being more abundant in more humid locations than in dry areas (reviewed in Halkka *et al.*, 2001).

Variation in morph frequencies may even occur within one habitat, owing to the influence of the niche composition, with different phenotypes preferring different vegetation types (Boucelham & Raatikainen, 1984; Halkka & Halkka, 1990; Quartau & Borges, 1997). The selective preference/avoidance of predators (especially birds) or parasites/parasitoids for some

colour patterns can also influence the frequency distribution (particularly by apostatic and/or aposematic selection) (Thompson, 1973, 1984a; Harper & Whittaker, 1976; Halkka & Halkka, 1990). Atmospheric pollution in some industrial areas also seems to influence high melanic frequencies (Lees & Dent, 1983; Lees *et al.*, 1983; Lees & Stewart, 1987).

Halkka *et al.* (1975, 1976) demonstrated that phenotype frequencies remain stable for long periods of time and that there is a rapid re-establishment of previous frequencies after transfer experiments. All these factors and the continuous maintenance of the rarer phenotypes in populations indicate that there is probably a strong selective mechanism involved. Halkka and Lallukka (1969) suggested the colour genes may be linked to genes reacting to the physical environment, constituting a supergene, and selection may not be directly related to colour.

Further processes have to be taken into account to explain the distribution of the phenotypes (Halkka *et al.*, 2001). Stochastic events such as genetic drift may be significant in small isolated, marginal or recently colonised areas, for example in insular populations, leading to loss of genetic variation that is reflected in a reduced number of phenotypes (Halkka *et al.,* 1974; Brakefield, 1990; observations in Azores archipelago by P.A.V.B.). Counteracting with this is migration, as it may in part explain the maintenance of the polymorphism if there are local selective forces acting in different areas (Halkka *et al.*, 2001).

Possibly there is a combination of evolutionary mechanisms acting on the colour polymorphism and more experimental work is needed to disentangle all the factors involved. The present study aims to investigate the adaptive significance of the colour polymorphism, by obtaining data on the fitness of different colour patterns. An experimental approach in captivity under semi-natural conditions was created to perform selected crosses of the phenotypes and measure survival, reproductive success, and the duration of egg maturation. Our main objectives were to test: (i) differential survival and reproductive success among phenotypes, in order to explore if the maintenance of the polymorphism can be explained by higher survival of rare phenotypes that could be counter-balanced by another factor(s) that would favour the common TYP; (ii) the relation between melanism and egg maturation efficiency, in particular to test if egg maturation is faster in melanic females, as only females are melanic in most populations. Three different colour phenotypes were used in these experiments: *typicus* (TYP), the most common, non-melanic recessive phenotype; *trilineatus* (TRI), the non-melanic dominant phenotype that shows the highest radiation reflectance and different thermal properties from the melanic and TYP phenotypes (Berry & Willmer, 1986); and *marginellus* (MAR), the most common melanic phenotype found in the population used for sampling.

Materials and methods

P. spumarius collections and maintenance in captivity

In order to obtain the desired number of individuals per phenotype for survival and reproductive measurements, a total of 3866 larval stage (nymphs) spittlebugs were randomly hand-collected from a Portuguese population at Quinta do Bom Sucesso, Lagoa de Óbidos (Lat 39°25′2.95″N, Long 9°13′39.18″W), on the 31 March (3355 nymphs), 29 April (400 nymphs), and 26 May 2011 (111 nymphs). The sample site is characterised by an extensive semi-open pine tree woodland (*Pinus pinaster*) over a sandy substrate with a predominance of psammophile and peri-litoral scrubland of grey dunes (*Corema album*, *Daphne gnidium*, *Cistus salvifolius*, *Halimium commutatum*, *Ononis ramosissima*, *Stauracanthus genistoides*, *Erica scoparia,* and *Calluna vulgaris*). More exposed areas such as clearings have been severely invaded by *Carpobrotus edulis,* an invasive Aizoaceae from South Africa, which is also used as a food plant by the nymphs of *P. spumarius*. Collected nymphs were kept on potted *C. edulis* (easy to maintain and readily accepted as food source by the spittlebug) in three cages covered with mesh which avoided excess of humidity, fungus contamination, and allows for proper oxygenation. Cages were stored outdoors in the Lisbon's Faculty of Science campus and checked daily for the presence of newly emerged adults, then separated by sex and age (by week of emergence) in distinct cages (one cage for each sex/week). Adults were scored for dorsal colour/pattern according to Halkka *et al.* (1973) and Stewart and Lees (1996). A shade on top of the cages avoided the most intense sunlight directly hitting the cages and prevented rain from soaking the pots.

Survival and reproductive success measurements

From the emerged adults, three main different types of mating pairs were isolated: 36 ♀ TYP × ♂ TYP, 35 ♀ TRI × ♂ TYP and 31 ♀ MAR × ♂ TYP. Crosses were performed by isolating the virgin pairs in small cages with a little portion of the food plant. The mating cage design included mesh-covered plastic cups and a portion of the plant leaf kept inside (and frequently changed). Cages were kept outdoors as well. This setup was checked hourly for the existence of mating pairs (except for a night period when the activity is lower). Once mating pairs were detected, the duration of the copula was recorded, by checking approximately every 10 min. After mating, the male was removed to prevent re-mating and the female was kept individually in a separate pot with the food plant and hay for oviposition. Weekly counts of all eggs laid were made. We defined the batch of one or more eggs laid by one female in one week as an oviposition event (time intervals between successive oviposition events can vary from one to several weeks).

The time between union (isolation of virgin pairs in the small cages) and mating (copula), duration of mating, number of oviposition events, number of eggs, duration of egg maturation (time between copula and oviposition event), and female longevity (time between adult emergence and death) were recorded for each pair.

The differences in measured parameters among phenotype groups were tested with non-parametric Kruskal–Wallis and Mann–Whitney *U*-tests. When appropriate, we controlled for false discovery rate (Benjamini & Hochberg, 1995), to compensate for multiple tests of significance, and consequently adjusted *P*-values were reported. Statistical tests were carried out in R version 3.0.2 (R Development Core Team, 2008).

Table 1. Number of adult *Philaenus spumarius* resulted from the 3866 nymphs captured, by week of final moult, sex and phenotype.

Phenotype	Week 1*	Week 2	Week 3	Week 4	Week 5†	Week 6	Week 7	Week 8‡	Week 9	Total	Freq (%)
Females ($N = 1037$; 53.23%)											
TYP/POP	196	249	195	87	97	39	8	–	33	904	87.17
TRI	13	9	19	5	6	5	3	–	3	63	6.08
MAR	8	9	8	5	4	3	0	–	1	38	3.66
QUA	0	19	5	1	2	0	0	–	0	27	2.60
LOP	0	0	3	0	0	0	0	–	0	3	0.29
ALB	0	0	2	0	0	0	0	–	0	2	0.19
Total	217	286	232	98	109	47	11	–	37	1037	
Males ($N = 911$; 46.77%)											
TYP/POP	214	280	166	78	58	37	8	–	21	862	94.62
TRI	13	7	13	8	2	0	1	–	4	48	5.27
VIT	0	0	0	1	0	0	0	–	0	1	0.11
Total	227	287	179	87	60	37	9	–	25	911	

∗New sampling event at 31 March 2011.
†New sampling event at 29 April 2011.
‡New sampling event at 26 May 2011.
Note that, as *populi* (POP) is considered an extreme form of *typicus* (TYP), they are often grouped together (TYP/POP) (Halkka *et al.*, 1973).
– indicate no data.

Results

P. spumarius phenotype frequency

Of the total of 3866 collected nymphs, 2420 reached an adult stage (63% survival). Of these, 1948 were separated by sex, while still freshly moulted, to ensure they had not mated. The remaining were already too mature and were discarded from the experiment. The number of adult *P. spumarius* by a week of final moult, sex, and phenotype is presented in Table 1. The total number of emerged females was similar to the number of males ($N = 1037$, 53% and $N = 911$, 47%, respectively), as well as the number of emerged females and males in each week. Percentages of emergence for each phenotype, by week, are also similar between both sexes. A percentage of 21–22% of TYP/POP, TRI and MAR females and 25–27% of TYP/POP and TRI males emerged on the first week; 14–28% of TYP/POP, TRI and MAR females and 15–33% of TYP/POP and TRI males emerged on the second week; 21–30% of TYP/POP, TRI and MAR females and 19–27% of TYP/POP and TRI males emerged on the third week; the remaining individuals emerged on weeks 4, 5, 6, 7, and 9 (Table 1).

TYP and POP predominate over the remaining phenotypes both in females and males (F = 87%; M = 95%), followed by TRI (F = 6%; M = 5%). Melanic phenotypes (MAR, QUA, LOP, and ALB) were only found in females and in very low frequencies (<4%). Similar patterns of frequency distribution have been observed in other Portuguese populations (see Quartau & Borges, 1997).

Survival and reproductive success

As the last moult tends to occur at different times for each individual, we tried to use similar numbers of randomly selected specimens of each phenotype from each week. A total of 116 mating pairs were set up and 107 (92%) were successful (the absence of mating in the remaining 9 pairs was as a result of female death or escape). Most copulations occurred in July ($N = 35$) and August ($N = 51$). We also observed copulations in September ($N = 18$), October ($N = 1$), and December ($N = 2$), probably as a result of the absence of effects such as predation and severe wind and rain conditions in captivity (note that the setup was kept under natural temperature and light conditions).

The mean values of measured variables are shown in Table 2. There was a significant difference in longevity of the total number of females (Kruskal–Wallis test, $\chi^2 = 8.5$, d.f. = 2, P-value = 0.0145) among phenotypes, particularly evident in females that oviposited ($\chi^2 = 6.2$, d.f. = 2, P-value = 0.0443). This difference is not evident for females that did not oviposit or did not mate, probably owing to the few number of specimens that were not successful in the mating process. Also, a significant difference can be observed in the number of oviposition events ($\chi^2 = 7.8$, d.f. = 2, P-value = 0.0205) and the number of eggs ($\chi^2 = 6.1$, d.f. = 2, P-value = 0.0461) among phenotypes. Despite these differences in the total number of eggs and oviposition events among phenotypes, the mean number of eggs that can be laid in each oviposition event is very similar among phenotypes ($\chi^2 = 2.3$, d.f. = 2, P-value = 0.3176). However, this number varied from just one egg to a maximum of 89 eggs laid in one oviposition event.

A greater longevity, higher number of oviposition events and a higher number of eggs was observed for TRI females (Table 2, Fig. 2). Paired analysis revealed significant differences between TRI and MAR females for longevity (Mann–Whitney U-test, $W = 579$, P-value = 0.0071, adjusted P-value = 0.0213) and the number of oviposition events ($W = 268$, P-value = 0.0145, adjusted P-value = 0.0432), and marginally significant differences for the number of eggs ($W = 259$, P-value = 0.0300, adjusted P-value = 0.0814). When comparing TRI and TYP, longevity was not significantly different, although marginally for the non-adjusted P-value ($W = 379$, P-value = 0.0789, adjusted P-value = 0.1184), the number of

Table 2. Mean values ± standard deviation (minimum–maximum) of each survival and reproductive variable for each female phenotype TYP, TRI, and MAR.

	TYP		TRI		MAR	
Variables	*N*	Average	*N*	Average	N	Average
Time between union and mating (weeks)	35	1.34 ± 1.81 (0–8)	34	1.97 ± 3.87 (0–20)	26	1.54 ± 1.45 (0–5)
Duration of mating (hours)	22	05.13 ± 02.59 (01.30–11.00)	24	04.40 ± 02.32 (01.45–09.30)	15	04.39 ± 02.06 (01.10–09.20)
Number of oviposition events	22	4.14 ± 2.68 (1–10)	26	6.27 ± 3.37 (1–13)	14	3.64 ± 2.06 (1–8)
Number of eggs	22	68.77 ± 44.56 (7–153)	26	110.23 ± 77.49 (14–322)	14	66.93 ± 47.41 (3–186)
Number of eggs in each oviposition event	22	16.41 ± 11.76 (1–49)	26	17.37 ± 14.92 (2–88)	14	18.74 ± 18.06 (1–89)
Duration of egg maturation first (time between copulation and first oviposition event) (weeks)	22	16.96 ± 4.13 (8–23)	26	18.81 ± 6.18 (9–39)	14	17.21 ± 3.79 (11–23)
Duration of egg maturation second (time between copulation and second oviposition event) (weeks)	17	20.53 ± 3.78 (13–27)	25	21.20 ± 5.70 (12–40)	11	21.73 ± 3.35 (15–27)
Longevity (females that oviposited) (weeks)	19	44.63 ± 9.45 (32–68)	26	54.42 ± 14.81 (36–78)	12	43.58 ± 6.52 (33–59)
Longevity (females that did not oviposit) (weeks)	10	31.20 ± 8.60 (19–46)	8	26.00 ± 6.55 (18–37)	9	27.67 ± 5.27 (22–38)
Longevity (females that did not mate) (weeks)	1	12.00	0	–	3	16.33 ± 2.52 (14–19)
Longevity (total females) (weeks)	30	39.07 ± 12.05 (12–68)	34	47.74 ± 18.03 (18–78)	24	34.21 ± 11.61 (14–59)

N, sample size.

oviposition events was significantly different ($W = 180.5$, P-value = 0.0288, adjusted P-value = 0.0432) and the number of eggs was marginally significantly different ($W = 192$, P-value = 0.0543, adjusted P-value = 0.0814). No significant differences were observed between TYP and MAR females ($W = 447$, P-value = 0.1317, adjusted P-value = 0.1317; $W = 166$, P-value = 0.7055, adjusted P-value = 0.7055; $W = 169$, P-value = 0.6379, adjusted P-value = 0.6379) for the same variables, respectively (Table 2, Fig. 2). No differences in egg maturation time were found among phenotypes (Kruskal–Wallis test, $\chi^2 = 0.9$, d.f. = 2, P-value = 0.6250), as well as in the time between the union of both male and female and mating ($\chi^2 = 1.3$, d.f. = 2, P-value = 0.5170) and in mating duration ($\chi^2 = 0.5$, d.f. = 2, P-value = 0.7891) (Table 2, Fig. 2).

Discussion

In this study, we present evidence for both increased longevity and a higher number of oviposition events and egg counts in the *trilineatus* females compared with the other phenotypes studied. Although P-values and adjusted P-values are in some cases only marginally significant or close to significance at the 0.05 level, this pattern is consistent and gives experimental support to previous field observations where *trilineatus* survives longer, i.e. its frequency increases as the season progresses towards the end of summer (Owen & Wiegert, 1962; Halkka, 1964). In 1976, Harper and Whittaker reported a higher attack rate by the dipteran pipunculid parasitoid *Verrallia aucta* (Fallen) towards *trilineatus* than for other phenotypes (even against the most abundant *typicus*). This would lead to instability of population phenotype frequencies unless, as they suggested, *trilineatus* has an increased survival rate or higher fertility than the other phenotypes. Our study brings evidence for this higher survival and fertility when compared with a very common phenotype (*typicus*) and another scarce phenotype (*marginellus*). In contrast, in several species of insects, melanic individuals are reported to be more resistant to pathogens, probably owing to the fact that melanin is one product of the phenoloxidase cascade that is involved in the immunity function of invertebrates (Wilson *et al.*, 2001; Dubovskiy *et al.*, 2013). The heavy defence investments made by melanic insects seems to carry life-history costs, namely decreased longevity and lower fecundity in comparison with their non-melanic phenotypes in some of these species [e.g. *Spodoptera littoralis* (Boisduval), Cotter *et al.*, 2008; *Galleria mellonella* (Linnaeus), Dubovskiy *et al.*, 2013]. This aspect has not been tested in *P. spumarius* but may be one of the factors maintaining the polymorphism.

Another balancing mechanism that could be acting to explain *trilineatus* phenotype frequencies is its different thermal properties, as demonstrated by Berry and Willmer (1986). The highest radiation reflectance and lower temperature excess of *trilineatus*, is pointed as an explanation for the observed negative correlation between the frequency of this phenotype and altitude and latitude (probably a consequence of thermal selection) (Thompson, 1984b, 1988; Berry & Willmer, 1986; Boucelham & Raatikainen, 1987; Halkka & Halkka, 1990). The evidences for the higher intrinsic survival capacity of *trilineatus* females observed in this study may counter-balance the increased thermal deficits in some low radiation environments. In contrast, individuals of *marginellus* demonstrated to be particularly sensitive to high-temperature periods during this study, as we

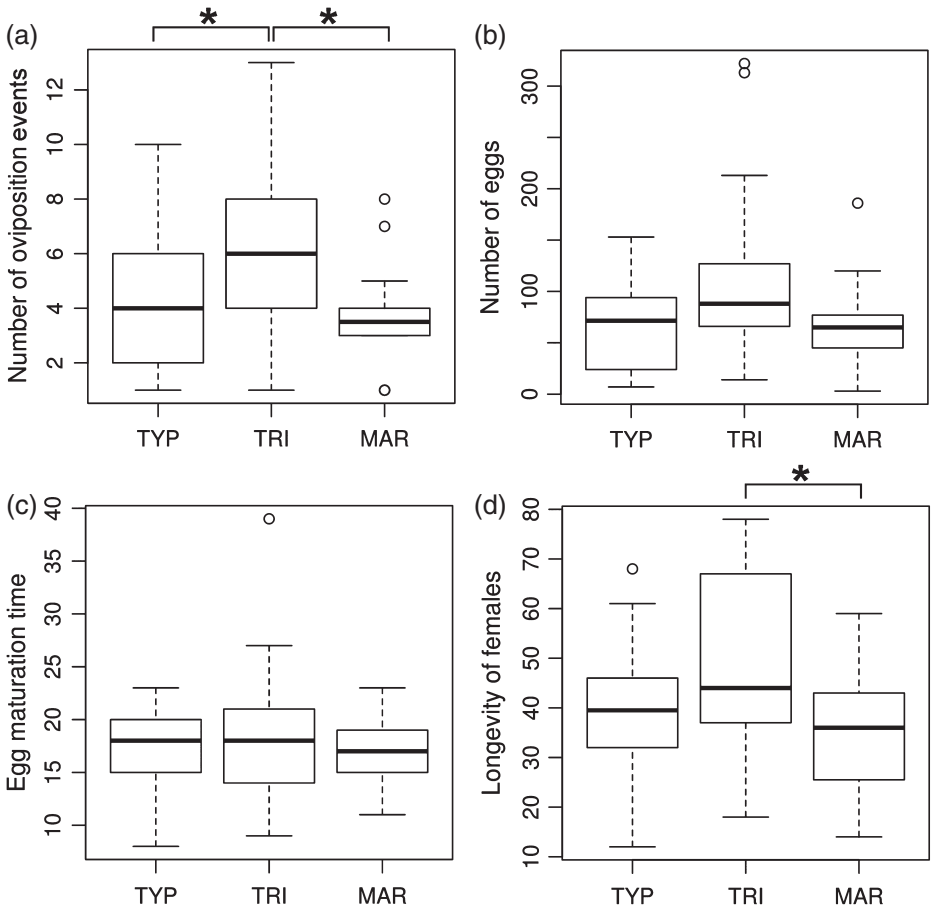


Fig. 2. Boxplots of the (a) number of oviposition events, (b) number of eggs, (c) egg maturation time, and (d) longevity of females. The rectangular box is delimited by the quartiles 25% and 75%, with the median value shown as a tick horizontal line; the whiskers indicate the non-outlier maximum and minimum, and the circles are the outliers. Statistically significant ($P < 0.05$) differences (Mann–Whitney *U*-tests) are indicated with the *.

observed a tendency for increased mortality during these periods, albeit not fully tested. The differences in thermal properties of pale versus dark phenotypes are reported in several insect species and constitute an important adaptive factor (Brakefield & Willmer, 1985; Rhamhalinghan, 1990; True, 2003).

The sex-specific restriction of many melanic patterns (Halkka *et al.*, 1973), and the highest radiation absorbency observed for these phenotypes (Berry & Willmer, 1986), could suggest an influence of the different colour types on the egg maturation time. Brakefield and Willmer (1985) and Rhamhalinghan (1990) reported that the higher fecundity of darker females in some ladybird species could be ascribed to their higher thermal absorption, which would lead to faster egg maturation. However, in this study we observed no differences in the duration of egg maturation among colour phenotypes, suggesting that thermal properties are not directly associated with egg maturation efficiency. Nevertheless, Horsefield (1978) and Berry and Willmer (1986) suggested that colouration seems to be essential to the duration of activity periods through the regulation of upper and lower body temperature thresholds between which the animal can undertake normal activity. The extent of the active period is possibly critical to *P. spumarius* owing to its feeding on nutrient-poor xylem, requiring long feeding periods to gain sufficient nutrition from its food source. This is particularly important to females owing to the greater nutritional requirements for egg production. The role of temperature in influencing egg survival and hatching time of polymorphic phenotypes has also been investigated in three female phenotypes of *Ischnura elegans* (Vander Linden) (Zygoptera: Coenagrionidae) (Bouton *et al.*, 2011). For *I. elegans*, the different colour forms did not differ in their response to temperature treatments.

This study allowed for monitoring the emergence patterns of adults across several weeks. No differences were observed within the emergence times of males and females or of different phenotypes in the weeks between March and May. An early study by Halkka *et al.* (1967) suggested that an early emergence of males, as well as *typicus*, in comparison with other phenotypes, would be favoured by natural selection to prevent predation (apostatic selection) and increase protection of females until egg-laying. However, we have no data regarding the early emergences in the field, as first nymphs were observed in early February (pers. obs. by S.E.S.) and adults were already present in early March.

We consider that *P. spumarius* constitute a still under-studied model system for the study of adaptation, in particular of evolutionary mechanisms maintaining polymorphisms. The long life cycle of *P. spumarius* and its sensibility to factors such as humidity and temperature are the main challenges in laboratory fitness studies on this species. Despite these difficulties, this study managed to give insights on the effect of the colour phenotypes in the survival and reproduction of *P. spumarius*, contributing to the exploration of the adaptive significance of the colour polymorphism in this species. These experiments were carried out by placing all the phenotypes in the same semi-natural conditions, but it would also be important to test differential fitness among phenotypes under different environmental conditions. These would include different plant species and (micro)climates. Moreover, the distribution range of *P. spumarius* in California has been reported to have moved northwards, presumably associated with climate change (Karban & Strauss, 2004). Studying the potential effects of climate change on the distribution and frequency of phenotypes in this species may also constitute an opportunity to test some predictions about the selective effect of increased temperature, UV-radiation, humidity, and pathogens in melanin-based colouration polymorphisms (Roulin, 2014). Also, the range of colour variation of the *typicus* phenotype, from pale to almost melanic, should also be explored in future studies as this may be influencing variation in the studied traits within *typicus*.

Most certainly, a combination of factors is acting on the maintenance of this polymorphism worldwide. A full understanding of the process will require investigating the interactions of selection, migration and drift, and the role of metapopulation dynamics in this system (Brakefield, 1990; Halkka *et al.*, 2001). Interestingly, a computational model developed by Craze (2009), showed that a biased phenotype ratio favouring the homozygous recessive phenotype in relation to the dominant phenotype (as we also see in *P. spumarius*), could be explained by a scenario involving a highly fragmented metapopulation with low migration and with a low level of frequency-dependent selection.

Acknowledgements

Financial support was provided by Fundação para a Ciência e a Tecnologia, Portugal (Project PTDC/BIA-BEC/098783/2008; cE3c Unit FCT funding UID/BIA/00329/2013; Grant SFRH/BPD/26365/2006; Grant SFRH/BD/73879/2010). We would like to thank the following people for help in the construction and maintenance of the material to the experimental setup, and for their comments during this work: Renata Martins, Vera Nunes, Carla Ribeiro, Inês Modesto, Patrícia Brás, and Francisco Pina-Martins. We would also like to thank Olli and Liisa Halkka for earlier comments on the planning of the work, Vinton Thompson and Mike Wilson for comments on the results, and Inês Fragata for help in the analysis. We thank the anonymous referees for helpful comments on earlier version of the manuscript.

References

Ayala, F.J. & Campbell, C.A. (1974) Frequency-dependent selection. *Annual Review of Ecology and Systematics*, **5**, 115–138.

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

Berry, A.J. & Willmer, P.G. (1986) Temperature and the colour polymorphism of *Philaenus spumarius* (Homoptera: Aphrophoridae). *Ecological Entomology*, **11**, 251–259.

Boucelham, M. & Raatikainen, M. (1984) The colour polymorphism of *Philaenus spumarius* (L.) (Homoptera, Cercopidae) in different habitats. *Annales Entomologici Fennici*, **50**, 43–46.

Boucelham, M. & Raatikainen, M. (1987) The effect of altitude, habitat and geographical location on the polymorphism of *Philaenus spumarius* (L.) (Hom. Cercopidae). *Annales Entomologici Fennici*, **53**, 80–86.

Boucelham, M., Hokkanen, H. & Raatikainen, M. (1988) Polymorphism of *Philaenus spumarius* (L.) (Homoptera, Cercopidae) in different latitudes, altitudes and habitats in the U.S.A. *Annales Entomologici Fennici*, **54**, 49–54.

Bouton, N., Iserbyt, A. & van Gossum, H. (2011) Thermal plasticity in life-history traits in the polymorphic blue-tailed damselfly, *Ischnura elegans*: no differences between female morphs. *Journal of Insect Science*, **11**, 1–11.

Brakefield, P.M. (1990) Genetic drift and patterns of diversity among colour-polymorphic populations of the Homopteran *Philaenus spumarius* in an island archipelago. *Biological Journal of the Linnean Society*, **39**, 219–237.

Brakefield, P.M. & Willmer, P.G. (1985) The basis of thermal melanism in the ladybird *Adalia bipunctata.* Differences in reflectance and thermal properties between morphs. *Heredity*, **54**, 9–14.

Bush, A.A., Yu, D.W. & Herberstein, M.E. (2008) Function of bright coloration in the wasp spider *Argiope bruennichi* (Araneae: Araneidae). *Proceedings of the Royal Society B*, **275**, 1337–1342.

Cotter, S.C., Myatt, J.P., Benskin, C.M.H. & Wilson, K. (2008) Selection for cuticular melanism reveals immune function and life-history trade-offs in *Spodoptera littoralis*. *Journal of Evolutionary Biology*, **21**, 1744–1754.

Craze, P.G. (2009) The fate of balanced, phenotypic polymorphisms in fragmented metapopulations. *Journal of Evolutionary Biology*, **22**, 1556–1561.

Drosopoulos, S. (2003) New data on the nature and origin of colour polymorphism in the spittlebug genus *Philaenus* (Hemiptera: Aphorophoridae). *Annales de la Société Entomologique de France*, **39**, 31–42.

Dubovskiy, I.M., Whitten, M.M.A., Kryukov, V.Y., Yaroslavtseva, O.N., Grizanova, E.V., Greig, C. *et al.* (2013) More than colour change: insect melanism, disease resistance and fecundity. *Proceedings of the Royal Society B*, **280**, 20130584.

Grant, P.R. (1999) *Ecology and Evolution of Darwin's Finches*. Princeton University Press, Princeton, New Jersey.

Halkka, O. (1964) Geographical, spatial and temporal variability in the balanced polymorphism of *Philaenus spumarius*. *Heredity*, **19**, 383–401.

Halkka, O. & Halkka, L. (1990) Population genetics of the polymorphic meadow spittlebug, *Philaenus spumarius* (L.). *Evolutionary Biology*, **24**, 149–191.

Halkka, O. & Lallukka, R. (1969) The origin of balanced polymorphism in the spittlebugs (*Philaenus*, Homoptera). *Annales Zoologici Fennici*, **6**, 431–434.

Halkka, O., Raatikainen, M., Vasarainen, A. & Heinonen, L. (1967) Ecology and ecological genetics of *Philaenus spumarius* (L.) (Homoptera). *Annales Zoologici Fennici*, **4**, 1–18.

Halkka, O., Halkka, L., Raatikainen, M. & Hovinen, R. (1973) The genetic basis of balanced polymorphism in *Philaenus* (Homoptera). *Hereditas*, **74**, 69–80.

Halkka, O., Raatikainen, M. & Halkka, L. (1974) The founder principle, founder selection, and evolutionary divergence and convergence in natural populations of *Philaenus*. *Hereditas*, **78**, 73–84.

Halkka, O., Halkka, L. & Raatikainen, M. (1975) Transfer of individuals as a means of investigating natural selection in operation. *Hereditas*, **80**, 27–34.

Halkka, O., Raatikainen, M. & Halkka, L. (1976) Conditions requisite for stability of polymorphic balance in *Philaenus spumarius* (L.) (Homoptera). *Genetica*, **46**, 67–76.

Halkka, O., Halkka, L. & Roukka, K. (2001) Selection often overrides the effects of random processes in island populations of *Philaenus spumarius* (Homoptera). *Biological Journal of the Linnean Society*, **74**, 571–580.

Harper, G. & Whittaker, J.B. (1976) The role of natural enemies in the colour polymorphism of *Philaenus spumarius* (L.). *Journal of Animal Ecology*, **45**, 91–104.

Honěk, A. (1984) Melanism in populations of *Philaenus spumarius* (Homoptera, Aphrophoridae) in Czechoslovakia. *Věstník Československé Společnosti Zoologické*, **48**, 241–247.

Horsefield, D. (1978) Evidence for xylem feeding by *Philaenus spumarius* (L.). *Entomologica Experimentalis et Applicata*, **24**, 95–99.

Jones, J.S., Leith, B.H. & Rawlings, P. (1977) Polymorphism in Cepaea – a problem with too many solutions? *Annual Review of Ecology and Systematics*, **8**, 109–143.

Karban, R. & Strauss, S.Y. (2004) Physiological tolerance, climate change, and a northward range shift in the spittlebug, *Philaenus spumarius*. *Ecological Entomology*, **29**, 251–254.

Kusche, H. & Meyer, A. (2014) One cost of being gold: selective predation and implications for the maintenance of the Midas cichlid colour polymorphism (Perciformes: Cichlidae). *Biological Journal of the Linnean Society*, **111**, 350–358.

Lees, D.R. & Dent, C.S. (1983) Industrial melanism in the spittlebug *Philaenus spumarius* (L.) (Homoptera: Aphrophoridae). *Biological Journal of the Linnean Society*, **19**, 115–129.

Lees, D.R. & Stewart, A.J.A. (1987) Localized industrial melanism in the spittlebug *Philaenus spumarius* (L.) (Homoptera: Aphrophoridae) in Cardiff Docks, south Wales. *Biological Journal of the Linnean Society*, **31**, 333–345.

Lees, D.R., Dent, C.S. & Gait, P.L. (1983) Geographic variation in the colour/pattern polymorphism of British *Philaenus spumarius* (L.) (Homoptera: Aphrophoridae) populations. *Biological Journal of the Linnean Society*, **19**, 99–114.

Maan, M.E. & Cummings, M.E. (2008) Female preferences for aposematic signal components in a polymorphic poison frog. *Evolution*, **62**, 2334–2345.

Owen, D.F. & Wiegert, R.G. (1962) Balanced polymorphism in the meadow spittlebug, *Philaenus spumarius*. *The American Naturalist*, **96**, 353–359.

Oxford, G.S. (1985) A countrywide survey of colour morph frequencies in the spider *Enoplognatha ovata* (Clerck) (Araneae: Theridiidae): evidence for natural selection. *Biological Journal of the Linnean Society*, **24**, 103–142.

Phifer-Rixey, M., Heckman, M., Trussell, G.C. & Schmidt, P.S. (2008) Maintenance of clinal variation for shell colour phenotype in the flat periwinkle *Littorina obtusata*. *Journal of Evolutionary Biology*, **21**, 966–978.

Punzalan, D., Rodd, F.H. & Hughes, K.A. (2005) Perceptual processes and the maintenance of polymorphism through frequency-dependent predation. *Evolutionary Ecology*, **19**, 303–320.

Quartau, A. & Borges, P. (1997) On the colour polymorphism of *Philaenus spumarius* (L.) (Homoptera, Cercopidae) in Portugal. *Miscellanea Zoologica*, **20**, 19–30.

R Development Core Team (2008) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rhamhalinghan, M. (1990) Variations in the internal temperatures of melanics and typicals of *Coccinella septempunctata* L. (Coleoptera: Coccinellidae). *Journal of Advanced Zoology*, **10**, 31–36.

Rodrigues, A.S.B., Silva, S.E., Marabuto, E., Silva, D.N., Wilson, M.R., Thompson, V. *et al.* (2014) New mitochondrial and nuclear evidences support recent demographic expansion and an atypical phylogeographic pattern in the spittlebug *Philaenus spumarius* (Hemiptera, Aphrophoridae). *PLoS ONE*, **9**, e98375.

Roulin, A. (2014) Melanin-based colour polymorphism responding to climate change. *Global Change Biology*, **20**, 3344–3350.

Roulin, A., Ducret, B., Ravussin, P.A. & Altwegg, R. (2003) Female colour polymorphism covaries with reproductive strategies in the tawny owl *Strix aluco*. *Journal of Avian Biology*, **34**, 393–401.

Stewart, A.J.A. & Lees, D.R. (1996) The colour/pattern polymorphism of *Philaenus spumarius* (L.) (Homoptera: Cercopidae) in England and Wales. *Philosophical Transactions of the Royal Society B*, **351**, 69–89.

Svensson, E.I., Abbott, J. & Hardling, R. (2005) Female polymorphism, frequency dependence, and rapid evolutionary dynamics in natural populations. *American Naturalist*, **165**, 567–576.

Thompson, V. (1973) Spittlebug polymorphic for warning coloration. *Nature*, **242**, 126–128.

Thompson, V. (1984a) Polymorphism under apostatic and aposematic selection. *Heredity*, **53**, 677–686.

Thompson, V. (1984b) Distributional evidence for thermal melanic color forms in *Philaenus spumarius*, the polymorphic spittlebug. *American Midland Naturalist*, **111**, 288–295.

Thompson, V. (1988) Parallel colour form distributions in European and North American populations of the spittlebug *Philaenus spumarius* (L.). *Journal of Biogeography*, **15**, 507–512.

Thompson, V. & Halkka, O. (1973) Color polymorphism in some North American *Philaenus spumarius* (Homoptera: Aphrophoridae) populations. *American Midland Naturalist*, **89**, 348–359.

True, J.R. (2003) Insect melanism: the molecules matter. *Trends in Ecology & Evolution*, **18**, 640–647.

Vercken, E., Clobert, J. & Sinervo, B. (2010) Frequency-dependent reproductive success in female common lizards: a real-life hawk-dove-bully game? *Oecologia*, **162**, 49–58.

Whittaker, J.B. (1972) Polymorphism in *Philaenus spumarius* (Homoptera) in the USSR. *Oikos*, **23**, 366–369.

Wilson, K., Cotter, S.C., Reeson, A.F. & Pell, J.K. (2001) Melanism and disease resistance in insects. *Ecology Letters*, **4**, 637–649.

Yurtsever, S. (2000) On the polymorphic meadow spittlebug, *Philaenus spumarius* (L.) (Homoptera: Cercopidae). *Turkish Journal of Zoology*, **24**, 447–459.

Accepted 12 July 2015

Associate Editor: Robby Stoks

CHAPTER 4

Assessing genotype-phenotype associations in three dorsal colour morphs in the meadow spittlebug *Philaenus spumarius* (L.) (Hemiptera: Aphrophoridae) using genomic and transcriptomic resources

Ana S.B. Rodrigues[1*], Sara E. Silva[1], Francisco Pina-Martins[1,2], João Loureiro[3], Mariana Castro[3], Karim Gharbi[4], Kevin P. Johnson[5], Christopher H. Dietrich[5], Paulo A.V. Borges[6], José A. Quartau[1], Chris D. Jiggins[7], Octávio S. Paulo[1L], Sofia G. Seabra[1L]

[1] Computational Biology and Population Genomics Group, cE3c – Centre for Ecology, Evolution and Environmental Changes, DBA/FCUL, Campo Grande, P-1749-016 Lisboa, Portugal.

[2] Centro de Estudos do Ambiente e do Mar (CESAM), DBA/FCUL, Lisboa, Portugal.

[3] Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal.

[4] Edinburgh Genomics, Ashworth Laboratories, King's Buildings, The University of Edinburgh, Edinburgh EH9 3JT, UK.

[5] Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, Illinois, USA.

[6] cE3c – Centre for Ecology, Evolution and Environmental Changes / Azorean Biodiversity Group and Universidade dos Açores – Departamento de Ciências e Engenharia do Ambiente, Angra do Heroísmo, Açores, Portugal.

[7] Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom.

[L] Both authors contributed equally (senior authors)

submitted to BMC Genetics

* Corresponding author:
Ana Sofia Rodrigues

E-mail: ana87bartolomeu@gmail.com

Computational Biology and Population Genomics Group
cE3c – Centre for Ecology, Evolution and Environmental Changes
Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa
Campo Grande
P-1749-016 Lisboa
Portugal

Tel: +351217500000 ext. 22363/22325
Fax: +351217500028

Running title: Genotype-phenotype associations in *P. spumarius*

Keywords: association study, colour polymorphism, *de novo* genome assembly, *de novo* transcriptome assembly, meadow spittlebug

Abstract

Background: Colour polymorphisms are common among animal species. When combined with genetic and ecological data, these polymorphisms can be excellent systems in which to understand adaptation and the molecular changes underlying phenotypic evolution. The meadow spittlebug, *Philaenus spumarius* (L.) (Hemiptera, Aphrophoridae), a widespread insect species in the Holarctic region, exhibits a striking dorsal colour/pattern balanced polymorphism. Although experimental crosses have revealed the Mendelian inheritance of this trait, its genetic basis remains unknown. In this study we aimed to identify candidate genomic regions associated with the colour balanced polymorphism in this species.

Results: By using restriction site-associated DNA (RAD) sequencing we were able to obtain a set of 1,837 markers across 33 individuals to test for associations with three dorsal colour phenotypes (*typicus*, *marginellus*, and *trilineatus*). Single and multi-association analyses identified a total of 60 SNPs associated with dorsal colour morphs. The genome size of *P. spumarius* was estimated by flow cytometry, revealing a 5.3 Gb genome, amongst the largest found in insects. A partial genome assembly, representing 24% of the total size, and an 81.4 Mb transcriptome, were also obtained. From the SNPs found to be associated with colour, 35% aligned to the genome and 10% to the transcriptome. Our data suggested that major loci, consisting of multi-genomic regions, may be involved in dorsal colour variation among the three dorsal colour morphs analysed. However, no homology was found between the associated loci and candidate genes known to be responsible for coloration pattern in other insect species. The associated markers showed stronger differentiation of the *trilineatus* colour phenotype, which has been shown previously to be more differentiated in several life-history and physiological characteristics as well. It is possible that colour variation and these traits are linked in a complex genetic architecture.

Conclusions: The loci detected to have an association with colour and the genomic and transcriptomic resources developed here constitute a basis for further research on the genetic basis of colour pattern in the meadow spittlebug *P. spumarius*.

Background

Understanding the genetic basis underlying phenotypic variation responsible for evolutionary change and adaptation in natural populations remains a major goal and one of the most interesting challenges in evolutionary biology. Not long ago, despite the development of new molecular tools, establishing genotype-phenotype associations, mapping adaptive loci, and identifying gene function, was limited to a few *taxa* due to technological and cost constraints. With the latest advances in sequencing technologies, the relationships between genetic variation and adaptive traits can now be investigated in a broader range of species for which, in some cases, there is extensive knowledge of ecological and evolutionary history, but few genomic resources [1–7]. Moreover, with the development of population genomics it has become possible not only to assess the genetic basis of adaptation directly at a genomic level, but also to distinguish the evolutionary effects of forces acting on the whole genome from those influencing only particular loci [8,9].

Intraspecific colour variation is commonly found in many different *taxa*, including mammals [10], fishes [11], amphibians [12], reptiles [13,14], birds [15,16], and many invertebrates (e.g. land snails, spiders, grasshoppers and butterflies; see [17] for references). Colour patterns may serve a wide variety of adaptive functions, ranging from a visual signal used in mate choice, to crypsis or aposematism to avoid predators, to aiding in the regulation of body temperature [18]. Through their interactions with other physiological and/or ecological traits, colour polymorphisms may also influence the habitat choice, dispersal capability and adaptation to a changing or novel environment, thus influencing the ecological success and evolutionary dynamics of populations and species [19]. When combined with genomic and ecological data, these colour polymorphisms can be an excellent system for understanding adaptation and speciation and for the study of the micro-evolutionary forces that maintain genetic variation [20]. Negative frequency-dependent selection, resulting from processes such as predation or sexual selection [21–23], heterozygote advantage [24], and disruptive selection/divergence with gene-flow [25,26] are some of the mechanisms suggested to be involved in the maintenance of colour polymorphisms. Alternative strategies that result in almost the same fitness values for colour morphs have also been reported [27].

The meadow spittlebug, *Philaenus spumarius* (Linnaeus, 1758) (Hemiptera, Aphrophoridae), a widespread and highly polyphagous sap-sucking insect species in the Holarctic region, shows a well studied balanced polymorphism of dorsal colour/pattern variation [28]. It is the most investigated species of its genus and has high genetic and morphological variation [29]. Sixteen adult colour phenotypes are known to occur in natural

populations [30] but only thirteen are referred in the literature. These are divided into non-melanic (*populi*, *typicus*, *vittatus*, *trilineatus* and *praeustus*) and melanic forms (*marginellus*, *flavicollis*, *gibbus*, *leucocephalus*, *lateralis*, *quadrimaculatus*, *albomaculatus* and *leucopthalmus*) [28,30–32]. The occurrence and frequency of the colour phenotypes differ among populations and may result from different selective pressures such as habitat composition, climatic conditions (including altitudinal and latitudinal gradients), industrial melanism and predation (reviewed in [30,32]). Silva and colleagues [33] have shown higher longevity and fertility of the *trilineatus* phenotype in laboratory conditions, which was also found to have the highest reflectance [34] and to be more prone to parasitoid attacks [35], supporting the idea that complex mechanisms are involved in the maintenance of this polymorphism. Crossing experiments have revealed the Mendelian inheritance of this trait, which is mainly controlled by an autosomal locus *p* with seven alleles, with complex dominance and co-dominance relationships, being likely regulated by other loci [31,36]. The *typicus* phenotype is the most common (over 90% frequency in most populations) and it is the bottom double recessive form. It is believed to be the ancestral form because its main colour pattern characteristics are shared with several other cercopid species [36]. The completely melanic form *leucopthalmus* is dominant over *typicus*, and several other forms, with pale heads and/or spots, are dominant over the completely dark form. The *trilineatus* phenotype, pale with three dark stripes, is controlled by the top dominant allele p^T [36,37]. Halkka and Lallukka [38] suggested the colour genes may be linked to genes involved in response to the physical environment through epistatic interactions, constituting a supergene, and selection may not be directly related to colour. Evidence that balanced polymorphisms can result from tight genetic linkage between multiple functional loci, known as supergenes [39], has been reported in mimetic butterflies [40,41], land snails [42] and birds [43]. In *P. spumarius* the genetic architecture of its balanced dorsal colour polymorphism and the possible existence of a supergene remain to be investigated.

A genome-wide association study has the potential to identify the genetic and/or genomic region(s) associated with these dorsal colour patterns. In this study we used restriction site-associated DNA (RAD) sequencing [1] to obtain a set of Single Nucleotide Polymorphisms (SNPs) that were tested for associations with three dorsal colour phenotypes in *P. spumarius*. The phenotypes used were: *typicus* (TYP), the most common and non-melanic recessive phenotype; *trilineatus* (TRI), the non-melanic dominant phenotype; and *marginellus* (MAR), the most common melanic phenotype found in the population from which samples were collected. The first partial draft genome and transcriptome of *P. spumarius* are presented here

and were used to help the characterisation of the genomic regions found to be associated with colour variation. The size of the genome of this insect species was also estimated by flow cytometry.

Methods

Sampling and DNA extraction

A total of 36 female specimens of *P. spumarius* from three different colour phenotypes – 12 *typicus* (TYP), 12 *trilineatus* (TRI), and 12 *marginellus* (MAR) – were collected from a Portuguese population near Foz do Arelho locality (39°25'2.95"N; 9°13'39.18"W) in 2011. Adult insects were captured using a sweep net suitable for low-growing vegetation and an entomological aspirator (pooter). Specimens were preserved in absolute ethanol and stored at 4 ºC. The wings and abdomen were removed to avoid DNA contamination by endosymbionts, parasitoids and parasites and only the thorax and head were used. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen).

Illumina sequencing of genomic libraries

Three RAD libraries with twelve individuals each were prepared following a modified RAD sequencing protocol [1], using PstI-HF (New England BioLabs) restriction enzyme to digest 300 ng of genomic DNA per sample. Digested DNA was ligated to P1 barcoded adapters using twelve different barcodes for each library. Adapter-ligated fragments were pooled and sheared targeting a 500 bp average fragment size using a sonicator. To remove adapter dimers, libraries were purified with Agencourt AMPure XP (Beckman Coulter) magnetic beads after P2 adapter ligation with a volume DNA/beads ratio of 1:0.8. After end-repair using a commercial kit (New England BioLab), libraries were amplified by Polymerase Chain Reaction (PCR) performing an initial denaturation step at 98 ºC for 30 s, followed by 18 cycles of one denaturation step at 98 ºC for 10 s, annealing at 65 ºC for 30 s, extension at 72 ºC for 30 s and a final 5 min extension step. PCR-enriched libraries were purified with AMPure XP beads and the DNA concentration of each library was quantified in a Qubit 2.0 (Invitrogen). Libraries, in a proportional representation, were paired end sequenced in three lanes of an Illumina HiSeq 2000 at Genepool (Ashworth Laboratories).

SNP calling and genotyping

Raw reads were trimmed, demultiplexed and aligned using the pyRAD software pipeline v3.0.5 [44], which follows the method of [45]. Reads were first clustered by individual and highly similar reads assembled into "clusters" using the programs MUSCLE v3.8.31 [46] and VSEARCH v1.9.3 [47] that allowed reads within "clusters" to vary not only for nucleotide polymorphisms but also for indels. All bases with a Phred quality score below 20 were converted to N (undetermined base). For each individual, consensus sequences based on estimates of the sequencing error-rate and heterozygosity were obtained for each locus. Similarity threshold required to cluster reads together and individuals into a locus was 0.88. Minimum "cluster" depth for each individual was six reads. Only loci with a minimum coverage of nine individuals (25%) were retained in the final dataset. To limit the risk of including paralogs in analysis, loci sharing more than 50% heterozygous sites were not considered and the maximum number of heterozygous sites in a consensus sequence (locus) allowed was five. After clustering sequences, a data matrix for each locus was generated. Further filtering and summary statistics were, posteriorly, performed using VCF Tools v 0.1.13 [48]. Loci were excluded from the final matrix based on (i) a missing data higher than 90% per individual, (ii) a minor allele frequency lower than 5% and (iii) a missing data per loci higher than 25%. Linkage disequilibrium (LD) was also measured using the squared correlation coefficient (r^2) in VCFtools. In association analysis, the detection of statistical associations may be affected when a marker is replaced with a highly correlated one [49]. Taking this into account, highly correlated SNPs in the same locus (r^2 =1) were randomly eliminated and only one of them was retained in the final VCF matrix. The filtered VCF file with the genotypes for each individual was converted into the file formats needed for further analyses using PGDSpider v2.0.4.0 [50], fcGENE v1.0.7 [51] and/or using customised python scripts.

Association with dorsal colour phenotypes

For the SNPs dataset, single-SNP associations between allele frequencies and dorsal colour phenotypes were tested using a Fisher's exact test of allelic association in PLINK v1.07 [52]. Three pairwise analyses were performed: MAR vs. TRI, MAR vs. TYP and TRI vs. TYP. Allele frequencies in each pair, the odds ratio and *p*-values were obtained for each SNP and a false discovery rate (FDR) of 5% was applied [53] to each pairwise analysis to test for false positives.

To test for single and multi-SNP correlations between SNPs and colour morphs, a Bayesian

Variable Selection Regression (BVSR) model proposed by [54] was also performed in the same three pairs and carried out in piMASS v 0.9. Generally used for association studies with continuous response variables, piMASS is also appropriate for studies with binary phenotypes [54]. This method uses the phenotype as the response variable and genetic variants (SNPs) as covariates to evaluate SNPs that may be associated with a particular phenotype [54]. SNPs statistically associated with phenotypic variation are identified by the posterior distribution of γ, or the posterior inclusion probability (PIP). In our multi-locus analyses, markers with a PIP greater than 99% empirical quantile ($PIP_{0.99}$ SNPs) were considered as highly associated with colour morphs. For all $PIP_{0.99}$ SNPs we reported their PIP and the estimates of their phenotypic effect (β). A positive β in the pairwise morph1-morph2 (e.g. MAR-TRI) analysis means that the frequency of the minor allele (maf) is higher in morph2 (TRI in the example) and a negative β means that maf is higher in morph1 (MAR in the example). Thus, to investigate the phenotypic effect size of each $PIP_{0.99}$ SNP, the | β | was considered. The model contains additional parameters that are estimated from the data: proportion of variance explained by the SNPs (PVE), the number of SNPs in the regression model (nSNPs) and the average phenotypic effect of a SNP that is in the model (σSNP). For all pairwise analyses, we obtained four million Markov Chain Monte Carlo samples from the joint posterior probability distribution of model parameters (recording values every 400 iterations) and discarded the first 100,000 samples as burn-in. piMASS also outperforms a single-SNP approach to detect causal SNPs even in the absence of interactions between them [54]. For single-marker tests, SNPs above 95% empirical quantile for Bayes Factor (BF) ($BF_{0.95}$ SNPs) were considered to be strongly associated to the colour phenotypes. Those above 99% empirical quantile for BF ($BF_{0.99}$ SNPs) were considered to have the strongest associations. Imputation of the missing genotypes was performed in BIMBAM v1.0 [55].

Genetic differences among populations were tested using a *G*–test [56] and estimates of F_{ST} were obtained following the method of [57] implemented in GENEPOP v4.2.2 [58]. To better visualise and explore the correlation between significant SNPs, obtained in the several association analyses, and colour phenotypes, a Principal Component Analysis (PCA) was done using R Package SNPRelate (Bioconductor v3.2; R v3.2.3) implemented in the vcf2PCA.R script [59].

De novo sequencing and assembly of the meadow spittlebug genome

To attempt potential *de novo* assembly of the genome, genomic DNA of one *P. spumarius* individual from Quinta do Bom Sucesso, Lagoa de Óbidos (Portugal) was extracted using the

DNeasy Blood & Tissue Kit (Qiagen) and sequenced externally in GenoScreen. A whole-genome shotgun sequencing approach using one lane of Illumina HiSeq 2000 to generate a paired-end library of approximately 366 million 100 bp reads was carried out. After sequencing, the quality of the sequence reads was assessed in FastQC v0.10.1 [60] and low quality sequences were trimmed by using Trimmomatic v 0.35 [61] and the default parameters. *De novo* assembly of large genomes tends to be computationally demanding, requiring very large amounts of memory to facilitate successful assembly. Taking these conditions into account, the assembler SOAPdenovo2 [62,63] was chosen to assemble the sequenced *P. spumarius* genome. This assembler implements the *de Bruijn* graph algorithm tailored specifically to perform the assembly of short Illumina sequences and is optimised for large genomes. A k-mer parameter of 33 was used for this assembly. The quality of the assembly results was investigated through several metrics: N50, percentage of gaps, number of *contigs*, number of scaffolds and genome coverage (total number of base pairs).

De novo sequencing and assembly of the meadow spittlebug transcriptome

Fresh adult specimens of *P. spumarius* were obtained from Lexington, Fayette Co., Kentucky, USA in July 2013 and frozen at -80 C. Total RNA was extracted from 6 adult specimens by first grinding the entire body using a 1 mL glass tissue grinder with 1 mL Trizol (Invitrogen). This was followed by passing the homogenate over a Qiagen Qiashredder column. The eluate was extracted with 200 µL chloroform, and the RNA was precipitated with 500 µL isopropanol. Pellets were resuspended in RNAse-free water.

Paired-end RNA libraries were prepared using Illumina's TruSeq Stranded RNA sample preparation kit with an average cDNA size of 250 bp (range 80-550 bp). These libraries were sequenced using an Illumina HiSeq2500 machine with a TruSeq SBS sequencing kit version 1 analysed with Casava v1.8.2. Raw reads were filtered for duplicates using a custom script and trimmed for 5' bias and 3' quality using the FASTX-toolkit [64]. Transcriptome was assembled using SOAPdenovo-Trans v1.02 [65] with a k-mer of 49.

Genome size estimation by flow cytometry

Genome size estimates were obtained through flow cytometry [66]. A total of 22 individuals were analysed, seven females and six males of *P. spumarius,* and nine females of *P. maghresignus*, a closely related species of the same genus. A suspension of nuclei from both the *Philaenus* sample and a reference standard (*Solanum lycopersicum*, S.l., 'Stupické' with 2C = 1.96 pg; [67]) were prepared by chopping the thorax and the head of the insect together with

0.5 cm^2 of *S. lycopersicum* fresh leaf with a razor blade in a Petri dish containing 1 mL of WPB (0.2 M Tris.HCl, 4 mM $MgCl_2.6H_2O$, 1% Triton X-100, 2 mM EDTA $Na_2.2H_2O$, 86 mM NaCl, 10 mM metabisulfite, 1% PVP-10, pH adjusted to 7.5 and stored at 4 ºC; [68]). The nuclear suspension was filtered through a 30 µm nylon filter and 50 µg mL^{-1} of propidium iodide (PI, Fluka, Buchs, Switzerland) and 50 µg mL^{-1} of RNAse (Fluka, Buchs, Switzerland) were added to stained DNA and avoid staining of double stranded RNA, respectively. After 5 minutes of incubation, the nuclear suspension was analysed in a Partec CyFlow Space flow cytometer (532 nm green solid-state laser, operating at 30 mW; Partec GmbH., Görlitz, Germany). Data was acquired using the Partec FloMax software v 2.4d (Partec GmbH, Münster, Germany) in the form of four graphics: histogram of fluorescence pulse integral in linear scale (FL); forward light scatter (FS) vs. side light scatter (SS), both in logarithmic (log) scale; FL vs. time; and FL vs. SS in log scale. To remove debris, the FL histogram was gated using a polygonal region defined in the FL vs. SS histogram. At least 1,300 nuclei were analysed per *Philaenus*' G_1 peak [69]. Only CV values of 2C peak of *Philaenus* below 5% were accepted [70]. The homoploid genome size (2C in pg; [71]) was assessed through the formula: sample nuclear DNA content (pg) = (sample G_1 peak mean / *S. lycopersicum* G_1 peak mean) * genome size of *S. lycopersicum*. The obtained values were expressed in picograms (pg) and in giga base pairs (Gb), using the formula by [72] (1 pg = 0.978 Gb).

Differences in genome size between males and females were evaluated using a one-way ANOVA, followed by a Tukey test for multiple comparisons at $P < 0.05$. Statistical analyses were performed using SigmaPlot for Windows v. 12.5 (Systat Software).

Characterisation of RAD loci

A consensus sequence, with IUPAC ambiguity codes for variable sites, was generated for each RAD locus across individuals using the python script loci_consensus.py [73].

Homology to non-coding and coding regions was investigated for the inferred loci by locally querying consensus sequences against Arthropoda sequences available in the NCBI nucleotide database (RefSeq release 73, last modified 2 November 2015 and GenBank release 211, last modified 14 December 2015), using BLASTN 2.2.28+ [74]. A protein blast (RefSeq release 73, last modified 2 November 2015 and GenBank release 211, last modified 14 December 2015), using BLASTX 2.2.28+ [75], was also performed. An E-value threshold of $1e^{-5}$ was used.

RAD loci were also queried using BLASTN against the drafts of the *P. spumarius* genome and transcriptome assembled in this study. In this case, an E-value threshold of $1e^{-15}$ was

chosen as the cutoff for restricting the alignments to the most significant ones. The top five contigs and/or scaffolds were subsequently investigated by querying them using BLASTN against Arthropoda sequences available in nucleotide and protein databases of NCBI.

Availability of data and materials

Raw reads and assemblies of the transcriptome and genome are submitted to NCBI under BioProjects PRJNA272277 and PRJNA321110, respectively. RAD libraries used for association analyses are submitted to NCBI under BioProject PRJNA321110. Vcf file is available as additional file 3.

Ethics statement

This research does not involve any endangered or protected species and did not require any permits to obtain the spittlebug individuals.

Results

RAD sequencing and SNPs data matrix

The sequencing set produced a total of 341 million reads. After filtering reads based on quality scores, 269 million reads were retained, corresponding to an average of 7.4 million reads per individual. Before filtering, individuals yielded 335,767 to 12,711,816 sequenced reads of 90 bp each (Additional file 1: Fig. S1).

The average number of reads per locus per individual used to estimate a consensus sequence was 51.0 (Additional file 1: Fig. S2). For the clustering results, a total of 133,127 loci, consisting of 12,144,351 aligned nucleotides, inferred with a minimum of nine individuals (25%) per locus, and a total of 470,470 SNPs with a mean percentage of missing data per individual of 63.92%, were produced. Aligned loci, including gaps inserted in the course of the alignment, ranged from 90 to 109 bp in length (mean = 91 bp). When filtering by percentage of missing data, three individuals (TYP_5, TYP_13 and TRI_13; Additional file 1: Fig. S1, S2, S3) had more than 90% missing data and were excluded. After filtering, a set of 928 loci, 85,056 bases and 2,195 SNPs was retained. However, only 1,837 SNPs on 928 loci were considered for the analyses after those in the same locus sequence with a complete LD ($r^2 = 1$) were randomly excluded.

Single-SNP associations with colour phenotypes

The dataset was tested for allele frequency differences between pairs of dorsal colour phenotypes – MAR vs. TRI, MAR vs. TYP and TRI vs. TYP – using the Fisher's exact test and a Bayesian regression approach. Single-marker association analyses performed using the frequentist method found 205 SNPs with *p*-value < 0.05, corresponding to 11.16% of the analysed SNPs, but these were not significant after FDR correction (Additional file 2: Table S1). Single-SNP analyses using the Bayesian regression approach identified a total of 230 $BF_{0.95}$ SNPs (> 95% quantile Bayes Factor) associated with dorsal colour phenotypes, corresponding to 12.52% of the analysed markers. When a more strict, 99% quantile, threshold was applied 50 $BF_{0.99}$ SNPs (2.7%) showed the strongest associations to colour morphs, including eight shared among colour morph comparisons (Fig. 1) (Table 1) The number of $BF_{0.95}$ SNPs and $BF_{0.99}$ SNPs for each pairwise comparison were: 92 and 19, respectively, for MAR-TYP; 92 and 20, respectively, for TRI-TYP; 101 and 19, respectively, for MAR-TRI. Estimates of the phenotypic effects associated with $BF_{0.99}$ SNPs for each comparison were moderate with $0.10 < |\beta| < 0.15$ but much higher than the overall average for each pairwise analysis ($|\beta| = 0.0001$, MAR-TRI; $|\beta| = 0.0037$, MAR-TYP; $|\beta| = 0.0028$, TRI-TYP) (Table 1). Allele frequencies for the 50 SNPs involved in the differentiation of these colour morphs varied across the three colour phenotypes (Table 1). For the 50 $BF_{0.99}$ SNPs, F_{ST} estimates between pairs of colour morphs were highly significant (*p*-value < 0.000) (Additional file 2: Table S2), with the highest genetic differentiation between TRI and MAR (F_{ST} = 0.2145), intermediate between TRI and TYP (F_{ST} = 0.2125) and the lowest between MAR and TYP (F_{ST} = 0.1787) (Additional file 2: Table S3). Principal Component Analysis using the associated $BF_{0.99}$ SNPs showed a clear distinction among the three morphs when compared with the PCA using all 1,837 SNPs (Fig. 2a). Principal component 1 explained 13% of the total variation and indicated a differentiation between TRI and the other two colour morphs while PC2 explained 10% of the differences, separating TYP from MAR (Fig. 2b).

Multi-SNP Associations with colour phenotypes

The 1,837 SNPs dataset explained between 60 and 65% of the variance in dorsal colour phenotypes across all pairwise analyses of colour morphs. The highest proportions of variation explained by the investigated SNPs were detected in comparisons involving the TRI phenotype (Table 2). The highest proportion was observed in TRI-TYP analysis (*PVE* = 0.6515) while the lowest proportion was found in MAR-TYP analysis (*PVE* = 0.6018) (Table 2). Estimates of the mean number of SNPs (nSNPs) underlying dorsal colour variation ranged from 63 to 67 (Table

2). However, 95% credible intervals for these parameters estimates were typically large. The average effect of associated SNPs was high and similar among analyses but once again higher in comparisons involving TRI (σSNP = 1.1200, MAR-TRI; σSNP = 0.9776, TRI-TYP; σSNP = 0.9495, MAR-TYP) (Table 2). When considering models with the highest BFs ($\log_{10}$(BF) > 10) only, the mean number of SNPs included in the model (nSNPs_BF) for each comparison decreased up to values between nine and 12 while the mean effect size of the SNPs (σSNP_BF) increased ranging between 2.4 and 4.1 (Table 2). The posterior inclusion probabilities (PIPs) for the analysed SNPs were quite similar among all pairwise analyses but slightly higher in comparisons involving TRI (PIP = 0.0366, MAR-TRI; PIP = 0.0362, TRI-TYP and PIP = 0.0345, MAR-TYP) (Fig. 3) (Table 2). A subset of 19 SNPs with the highest inclusion probabilities ($PIP_{0.99}$ SNPs) were identified for each analysis and investigated (Table 3). This number was within the 95% credible intervals for the number of SNPs found to be associated with dorsal colour variation by the models with the highest BF (Additional file 1: Fig. S4) (Table 3). Estimates of the strength of association between genotypic variation at individual SNPs and phenotypic variation (| β |) varied among the analyses and all were greater than 0.5. We obtained SNPs with larger effect sizes for MAR-TRI analysis than for all other analyses. Seven $PIP_{0.99}$ SNPs were shared between two pairwise analyses (Table 3). In total, 50 different SNPs revealed a multi-association with colour morphs and, from those, 40 were also significant in the single-SNP analyses shown previously. For the 50 $PIP_{0.99}$ SNPs, population differentiation tests were also highly significant (p-value < 0.000) (Additional file 2: Table S2). Similarly, the highest genetic differentiation was observed between TRI and TYP (F_{ST} = 0.2159), intermediate between TRI and MAR (F_{ST} = 0.1907) and the lowest genetic differences were observed between MAR and TYP (F_{ST} = 0.1650) (Additional file 2: Table S3). Principal Component Analysis for all 50 $PIP_{0.99}$ SNPs of multi-association tests (Fig. 2c) and for the 40 intersected SNPs (Fig. 2d) showed the expected differentiation among dorsal colour morphs. Principal Component 1 explained 13% to 14% of the variance, differentiating TRI from the other morphs while PC2 explained 11% of the differences and revealed a differentiation between TYP and MAR.

Linkage patterns

The associated loci detected here had on average low levels of linkage disequilibrium for both analyses including all samples or analyses on each colour phenotype separately (Additional file 1: Fig. S5). However, strong allelic correlations (r^2 > 0.7) were found for five pairs of SNPs within MAR and for two pairs in TYP phenotypes (Additional file 2: Table S4).

Only two pairs, in MAR, consisted of SNPs present in the same RAD locus.

Genome size estimation

Philaenus spumarius and *P. maghresignus* estimates of genome size were 5.27 ± 0.25 pg (5.15 Gb) and 8.90 ± 0.20 pg (8.90 Gb), respectively. In *P. spumarius*, males and females differed significantly in genome size ($F_{1,11}$ = 14.292, *p*-value = 0.0030), with males presenting on average a lower genome size (5.07 ± 0.20 pg; 4.96 Gb) than females (5.44 ± 0.15 pg; 5.33 Gb) (Additional file 2: Table S5). Overall, the quality of the analyses was excellent, with a mean CV value of 2.97% being obtained for the sample's G_1 peak.

***De novo* sequencing and assembly of meadow spittlebug genome and transcriptome**

The genome sequencing set produced a total of 366 million reads. After filtering reads based on quality, 353 million reads (96.46%) were retained (Additional file 2: Table S6). SOAPdenovo2 produced 6,843,324 *contigs* and 4,010,521 scaffolds. The N50 was 686 bp and the percentage of gaps was 20.47%. In total, 1,218,749,078 bp were assembled which based on the total estimated genome size of 5.3 Gb, corresponds to approximately 24% of the *P. spumarius* genome.

For the transcriptome, the total number of 150 nt reads for each paired-end of the library was 17 million resulting in 5110.8 Mb of sequence (Additional file 2: Table S6). After quality filtering, 14 million (86.81%) read pairs were used in the assembly (Additional file 2: Table S6). The transcriptome assembly produced 173,691 contigs and 31,050 scaffolds. In this case, the observed N50 obtained was 803 bp and the percentage of gaps 0.39%. A total of 81,442,967 bp were assembled. Assembly statistics for the genome and transcriptome are summarised in Additional file 2: Table S6.

Characterisation of RAD loci

No significant hits were found when querying the 928 RAD loci against Arthropoda sequences of NCBI nt database and only 15 hits (E-value < $1e^{-11}$) were found against Arthropoda sequences of NCBI nr database (Additional file 2: Table S7). However, this was not unexpected considering RAD loci sequences are less than 100 bp and the most closely related insect species with an available genome is the pea aphid *Acyrthosiphon pisum*, which belongs to a separate hemipteran infraorder.

A total of 392 RAD loci (42.24%) aligned to the draft of *P. spumarius* genome (E-value threshold of $1e^{-15}$), 18 of which were associated with colour morphs (34.62% of the colour-

associated loci sequences) (Additional file 2: Table S8). On the other hand, 134 loci, corresponding to 14.44% of the total loci, aligned to *P. spumarius* transcriptome assembly. Five of those were colour-associated (9.62% of the colour-associated loci) (Additional file 2: Table S8). From the 18 colour-associated loci that aligned with the genome, four (22.22%) also aligned with the transcriptome. The proportion of colour-associated loci that aligned either with the genome or with the transcriptome was not significantly different from the proportions of the other loci (Fisher's exact test *p*-value = 0.8096). Some RAD loci had more than one contig/scaffold hit (Additional file 1: Fig. S6).

Transcriptome and genome scaffolds/contigs with RAD loci alignments, ranging from 100 to 12,325 bp (gaps included), were queried against Arthropoda nt and nr databases using BLASTN and BLASTX. Out of 210 transcriptome sequences, 22 (E-value < $1e^{-06}$) had homology with the nucleotide database (Additional file 2: Table S9) and 98 with the protein database (Additional file 2: Table S10). The majority of those sequences hits have E-values < $1e^{-12}$ in nucleotide (86.36%) and in protein (69.38%) blasts. On the other hand, one genome scaffold, out of 484 with RAD loci hits, matched with the nucleotide sequences (E-value < $1e^{-25}$) (Additional file 2: Table S11) and 90 with the protein database (Additional file 2: Table S12). The majority of those protein hits have E-values < $1e^{-12}$ (55.55%). Of the transcriptome and genome sequences with protein hits, five and three included associated loci, respectively (Additional file 2: Table S13). Four of these genome and transcriptome sequences matched with known proteins, the other four with uncharacterised ones. One of the identified proteins, to which the colour-associated locus 16628 aligned (genome scaffold 1372429 and transcriptome scaffolds 17697 and 17698), was a lysosomal-trafficking regulator, known to be involved in the trafficking of materials into lysosomes. Furthermore, a mutation of this protein in humans is associated with a pigmentation disorder [76]. The other identified protein, to which the colour-associated locus 22795 aligned (transcriptome scaffold 29739), was the nucleolar and coiled-body phosphoprotein 1.

Discussion

In this study, we aimed to identify candidate genomic regions associated with colour polymorphism in the meadow spittlebug *P. spumarius*, an insect species with a very large genome (5.3 Gb), as estimated here by flow cytometry. This large size is among the largest genomes reported in insects [77], making genomic analysis in this species particularly challenging. By using restriction site-associated DNA (RAD) sequencing in individuals of three dorsal colour phenotypes (*typicus*, *marginellus,* and *trilineatus*), we were able to detect

association with colour in 3% of the analysed SNPs (60 out of 1,837). These phenotypes did not reveal significant genome-wide differences but when considering only the associated SNPs, the three colour morphs were differentiated and the *trilineatus* phenotype showed the highest genetic differentiation. Interestingly, greater differences involving life-history traits such as longevity, number of eggs, and number of oviposition events are also known to occur in *trilineatus* [33]. It may be that the genetic differences detected in this morph also reflect some part of the genetic basis of these life-history differences among colour morphs. Therefore, we may not only be on the track to finding a colour gene but also perhaps an extensive region, or several regions of the genome, that links colour variation and other life-history or physiological traits, as previously suggested [38]. Our finding of several colour-associated SNPs, some of them mapped to different genome and transcriptome scaffolds, suggests to a complex genetic architecture involving this colour polymorphism.

In the single-SNP association analyses, the 50 individual SNPs found to be associated at 99% quantile ($BF_{0.99}$ SNPs) showed moderate phenotypic effects ($0.10 < |\beta| < 0.15$). In the multi-SNP association analyses, 50 SNPs with posterior inclusion probabilities at quantile 99% ($PIP_{0.99}$ SNPs) showed large effects for pairs of colour phenotypes ($\sigma SNP > 0.9$ and individual $PIP_{0.99}$ SNPs $|\beta| > 0.5$). From these, 40 were common to the SNPs identified in single SNP analyses ($BF_{0.99}$ SNPs), increasing the confidence for the detected associations. Although inferences about the genetic architecture are only tentative in this study, due to the relatively small proportion of the genome covered, our results suggest that differences among the three dorsal colour phenotypes are associated with several loci with large effects. However, it is still not entirely clear if these constitute the major loci, determining dorsal colour pattern revealed by Mendelian crosses in *P. spumarius* [31,36]. Large effect loci controlling colour pattern have been reported for *Heliconius* species [78,79], land snail *Cepaea nemoralis* [42], and more recently in *Timema cristinae* stick insects [80]. Several other examples [81–83] have shown that adaptive traits are affected by loci with large phenotypic effects and that this genetic architecture may be more common than initially thought. The majority of the colour-associated loci that we detected here did not show significant allelic correlations, being likely in independent genomic regions. However, a few loci were strongly correlated, indicating either physical linkage, random drift of rare alleles, or occurrence of recent mutations. The existence of tightly linked loci (a supergene) that can be maintained due to chromosomal rearrangements or selection of co-adapted loci with epistatic effects is also possible. In the mimetic butterfly *Heliconius melpomene*, a cluster of three tightly linked loci (HmN, HmYb and HmSb), lying just a few centimorgans apart, as well as other unlinked loci have been shown to control

distinct wing colour pattern elements in this species [84]. In a closely related species *Heliconius numata,* polymorphic colour variation is controlled by a single locus *P,* forming a supergene, resulting from chromosomal rearrangements [85]. A single gene, *doublesex*, with closely linked mutations, also controls supergene mimicry in *Papilio polytes* [86].

Various genes and pathways have been reported to be involved in insect coloration and pigmentation. These pathways comprise genes regulating the distribution of pigments in space and time, as well as genes that are involved in the synthesis of pigments [87]. Several colour genes have been described, mostly in *Drosophila* spp. (see [87] for a review) and are known to be involved in colour variation in the silk worm *Bombyx mori* [88] and *Papilio* spp. [86,89,90] as well. Novel and unexpected genes were found to be responsible for wing colour patterning in *Heliconius* species. Red wing elements are associated with expression of the transcription factor *optix* [91], which in turn is regulated by two distinct *cis*-regulatory loci [92]. Another gene, *cortex*, a member of a conserved cell cycle regulator family, appears to have adopted a novel function controlling colour pattern in *Heliconius* and probably across the Lepidoptera [93]. Regulatory regions are also known to control colour pattern in *Drosophila* flies [94,95]. However, none of the colour-associated loci that we found in our study matched these candidate genes and/or genomic regions of other insects. Approximately 10% of the loci with colour associations aligned with the *P. spumarius* transcriptome indicating that those loci are in coding regions that are expressed in adult stage. A similar proportion of alignment to the transcriptome was found between associated loci and all loci (Fisher's exact test *p*-value = 0.8096), suggesting that there is no enrichment/depletion of coding regions in the associated loci in relation to the total number of loci. Around 35% of the colour-associated loci aligned with the genome and 22% also aligned with the transcriptome. If we assume a good representation of the total transcriptome, this result point to the majority of the associated loci being in non-coding regions. Considering that our assembled genome represents only 24% of the total genome size, the low percentage of hits in the genome was expected. Also, the low number of nucleotide and protein matches of genome and transcriptome sequences is certainly due to the degree of similarity of *P. spumarius* to other available Arthropoda sequences being too low to allow significant matches. Increasing the genomic resources for this or related species will allow exploring the candidate loci here described and provide insight into some of the key questions that remain to be answered. What are the specific genes contributing to this balanced colour polymorphism? What mutations cause allelic differences in these genes and how do they contribute to the different colour phenotypes? Are there epistatic or additive effects among the alleles responsible for the polymorphism? Does this constitute a supergene?

Are coding or regulatory mutations involved? In the future, it would also be interesting to investigate the evolutionary history of the colour polymorphism within *Philaenus* since identical variation in dorsal colour/pattern can be observed in the other species of the genus, suggesting an ancestral polymorphism maintained through the speciation process.

Additional Files

Additional File 1: Figure S1 – S6. Histograms of the total number of raw reads, mean depth and proportion of missing data per individual and of the R^2 values for each colour-associated SNP comparison; scatterplots of the number of SNPs in the model as a function of the Bayes factor for each pairwise comparison in multi-SNP association tests; and number of RAD loci hits with genome and transcriptome. (pdf 1,433 kb)

Additional File 2: Table S1 – S13. Lists of colour-associated SNPs obtained for each pairwise comparison and association analyses; genic and genotypic differentiation tests; pairwise F_{st} estimates among dorsal colour phenotypes; SNP correlation value (r^2) in linkage disequilibrium analyses; Genome size estimates; Assembly statistics for genome and transcriptome; and lists of blast results. (xls 163.6 kb)

Additional File 3: Data file with 1,837 SNPs. (vcf 294.6 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ASBR, PAVB, JAQ, CDJ, OSP and SGS participated in the design of the study. ASBR, SES and SGS were responsible for sampling. ASBR was responsible for DNA extraction and RAD libraries preparation. KG was responsible for next generation sequencing of RAD libraries. ASBR performed the bioinformatics analyses, which included the genome and the RAD libraries assembly, the association analysis and blasts. FPM participated in the bioinformatics analysis. JL and MC were responsible for the genome size estimates by flow cytometry. KPJ and CHD provided sequencing and assembly of the transcriptome. ASBR wrote the manuscript with contributions from SGS, OSP, CDJ, JL, MC, KPJ, CHD. All authors read and approved

the final manuscript.

Acknowledgments

Financial support was provided from National Funds through FCT – Fundação para a Ciência e a Tecnologia, Portugal under projects PTDC/BIA-BEC/098783/2008 and UID/BIA/00329/2013, and grants SFRH/BD/73879/2010 (attributed to ASR); SFRH/BPD/26365/2006 (attributed to SGS); SFRH/BD/89617/2012 (attributed to MC). This work was also supported by US NSF grant DEB-1239788 to KPJ and CHD. We thank Kimberly Walden (University of Illinois) for assistance with RNA extraction and transcriptome assembly, Simon Baxter, from University of Adelaide, for assistance with RAD sequencing and Joana Fino and Diogo Silva for bioinformatics assistance.

References

1. Baird N A, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis Z a., et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3:e3376.

2. Davey JW, Blaxter ML. RADSeq: next-generation population genetics. Brief. Funct. Genomics. 2010;9:416–23.

3. Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome evolution and meiotic maps by massively parallel DNA sequencing: Spotted gar, an outgroup for the teleost genome duplication. Genetics. 2011;188:799–808.

4. Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, et al. Linkage mapping and comparative genomics using next-generation rad sequencing of a non-model organism. PLoS One. 2011;6:e19315.

5. Davey JW, Hohenlohe P a, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 2011;12:499–510.

6. Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. Special features of RAD Sequencing data: Implications for genotyping. Mol. Ecol. 2013;22:3151–64.

7. Rowe HC, Renaut S, Guggisberg A. RAD in the realm of next-generation sequencing technologies. Mol. Ecol. 2011;20:3499–3502.

8. Black WC, Baer CF, Antolin MF, DuTeau NM. Population genomics: genome-wide sampling of insect populations. Annu. Rev. Entomol. 2001;46:441–69.

9. Stinchcombe JR, Hoekstra HE. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity. 2008;100:158–70.

10. Hoekstra HE, Krenz JG, Nachman MW. Local adaptation in the rock pocket mouse (Chaetodipus intermedius): natural selection and phylogenetic history of populations. Heredity. 2005;94:217–28.

11. Munday PL, Eyre PJ, Jones GP. Ecological mechanisms for coexistence of colour polymorphism in a coral-reef fish: An experimental evaluation. Oecologia. 2003;137:519–26.

12. Hoffman EA, Blouin MS. A review of colour and pattern polymorphisms in anurans. Biol. J. Linn. Soc. 2000;70:633–65.

13. Nunes VL, Miraldo A, Beaumont MA, Butlin RK, Paulo OS. Association of Mc1r variants with ecologically relevant phenotypes in the European ocellated lizard, *Lacerta lepida*. J. Evol. Biol. 2011;24:2289–98.

14. Perez i de Lanuza G, Font E, Carazo P. Color-assortative mating in a color-polymorphic lacertid lizard. Behav. Ecol. 2013;24:273–9.

15. Galeotti P, Rubolini D, Dunn PO, Fasola M. Colour polymorphism in birds: Causes and functions. J. Evol. Biol. 2003;16:635–46.

16. Roulin A. The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. Biol. Rev. Camb. Philos. Soc. 2004;79:815–48.

17. Bond AB. The Evolution of Color Polymorphism: Crypticity, Searching Images, and Apostatic Selection. Annu. Rev. Ecol. Evol. Syst. 2007;38:489–514.

18. Forsman A, Ahnesjö J, Caesar S, Karlsson M. A model of ecological and evolutionary consequences of color polymorphism. Ecology. 2008;89:34–40.

19. Lozier JD, Jackson JM, Dillon ME, Strange JP. Population genomics of divergence among extreme and intermediate color forms in a polymorphic insect. Ecol. Evol. 2016;6:1075–91.

20. Gray SM, McKinnon JS. Linking color polymorphism maintenance and speciation. Trends Ecol. Evol. 2007;22:71–9.

21. Punzalan D, Rodd FH, Hughes KA. Perceptual Processes and the Maintenance of

Polymorphism Through Frequency-dependent Predation. Evol. Ecol. 2005;19:303–20.

22. Svensson EI, Abbott J, Härdling R. Female Polymorphism, Frequency Dependence, and Rapid Evolutionary Dynamics in Natural Populations. Am. Nat. 2005;165:567–76.

23. Kusche H, Meyer A. One cost of being gold: selective predation and implications for the maintenance of the Midas cichlid colour polymorphism (Perciformes: Cichlidae). Biol. J. Linn. Soc. 2014;111:350–8.

24. Vercken E, Clobert J, Sinervo B. Frequency-dependent reproductive success in female common lizards: a real-life hawk–dove–bully game? Oecologia. 2010;162:49–58.

25. Jones JS, Leith BH, Rawlings P. Polymorphism in Cepaea: A Problem with Too Many Solutions? Annu. Rev. Ecol. Syst. 1977;8:109–43.

26. Oxford GS. A countrywide survey of colour morph frequencies in the spider *Enoplognatha ovata* (Clerck) (Araneae: Theridiidae): evidence for natural selection. Biol. J. Linn. Soc. 1985. p. 103–42.

27. Roulin A, Ducret B, Ravussin PA, Altwegg R. Female colour polymorphism covaries with reproductive strategies in the tawny owl *Strix aluco*. J. Avian Biol. 2003;34:393–401.

28. Halkka O, Halkka L. Population-Genetics Of the Polymorphic Meadow Spittlebug, *Philaenus spumarius* (L). Evol. Biol. 1990;24:149–91.

29. Rodrigues ASB, Silva SE, Marabuto E, Silva DN, Wilson MR, Thompson V, et al. New mitochondrial and nuclear evidences support recent demographic expansion and an atypical phylogeographic pattern in the spittlebug *Philaenus spumarius* (Hemiptera, Aphrophoridae). PLoS One. 2014;9:1–12.

30. Yurtsever S. On the polymorphic meadow spittlebug, *Philaenus spumarius* (L.) (Homoptera: Cercopidae). Turkish J. Zool. 2000;24:447–59.

31. Stewart a. J a., Lees DR. The Colour/Pattern Polymorphism of *Philaenus spumarius* (L.) (Homoptera: Cercopidae) in England and Wales. Philos. Trans. R. Soc. B Biol. Sci. 1996;351:69–89.

32. Quartau JA, Borges PA V. On the colour polymorphism of *Philaenus spumarius* (L.) in Portugal. Misc. Zool. 1997;2:19–30.

33. Silva SE, Rodrigues ASB, Marabuto E, Yurtserver S, Borges PA V., Quartau JA, et al. Differential survival and reproduction in colour forms of *Philaenus spumarius* give new

insights to the study of its balanced polymorphism. Ecol. Entomol. 2015;40:759–66.

34. Berry AJ, Willmer PG. Temperature and the colour polymorphism of *Philaenus spumarius* (Homoptera: Aphrophoridae). Ecol. Entomol. 1986;11:251–9.

35. Harper G, Whittaker JB. The Role of Natural Enemies in the Colour Polymorphism of *Philaenus spumarius* (L.). J. Anim. Ecol. 1976;45:91.

36. Halkka O. The genetic basis of balanced polymorphism in *Philaenus* (Homoptera). Hereditas. 1973;74:69–80.

37. Stewart AJA, Lees DR. Genetic control of colour/pattern polymorphism in British populations of the spittlebug *Philaenus spumarius* (L.) (Homoptera: Aphrophoridae). Biol. J. Linn. Soc. 1988;34:57–79.

38. Hallka O, Lallukka R. The origin of balanced polymorphism in the spittlebugs (*Philaenus*, Homoptera). Ann. Zool. Fennici. 1969;6:431–4.

39. Thompson MJ, Jiggins CD. Supergenes and their role in evolution. Heredity. 2014;113:1–8.

40. Joron M, Wynne IR, Lamas G, Mallet J. Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. Evol. Ecol. 1999;13:721–54.

41. Nijhout HF. Polymorphic mimicry in *Papilio dardanus*: mosaic dominance, big effects, and origins. Evol. Dev. 2003;5:579–92.

42. Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML, Davison A. RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. Mol. Ecol. 2013;22:3077–89.

43. Tuttle EM, Bergland AO, Korody ML, Brewer MS, Newhouse DJ, Minx P, et al. Divergence and Functional Degradation of a Sex Chromosome-like Supergene. Curr. Biol. 2016;1–7.

44. Eaton DAR. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics. 2014;30:1844–9.

45. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping Loci de novo from short-read sequences. G3 (Bethesda). 2011;1:171–82.

46. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

47. VSEARCH GitHub repository [https://github.com/torognes/vsearch/commit/55b8c583e5e20652c511a6bd40c4456b6778fc02]

48. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.

49. Pritchard JK, Przeworski M. Linkage Disequilibrium in Humans: Models and Data. Am. J. Hum. Genet. 2001;69:1–14.

50. Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28:298–9.

51. Roshyara NR, Scholz M. fcGENE: A Versatile Tool for Processing and Transforming SNP Datasets. PLoS One. 2014;9:e97589.

52. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007;81:559–75.

53. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B. 1995. p. 289–300.

54. Guan Y, Stephens M. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. Ann. Appl. Stat. 2011;5:1780–815.

55. Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS Genet. Public Library of Science. 2007;3:e114.

56. Raymond M, Rousset F. An Exact Test for Population Differentiation. Evolution. 1995;49:1280.

57. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984;38:1358.

58. Rousset F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 2008;8:103–6.

59. R_little_scripts repository [https://github.com/Nymeria8/R_little_scripts/commit/0d91d1b89219c27ebf51a3074a32e5f191b19990].

60. Babraham Bioinformatics webpage

[http://www.bioinformatics.babraham.ac.uk/projects/fastqc/].

61. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

62. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1:18.

63. SOAPdenovo2 repository [https://github.com/aquaskyline/SOAPdenovo2/commit/dd6a98ba19bb21c3513a46ad5047d08e57583ab0].

64. FASTX-Toolkit webpage

(http://hannonlab.cshl.edu/fastx_toolkit/index.html).

65. SOAPdenovo-Trans repository [https://sourceforge.net/projects/soapdenovotrans/files/SOAPdenovo-Trans].

66. Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E. Rapid Flow Cytometric Analysis of the Cell Cycle in Intact Plant Tissues. Science. 1983; 220:1049–1051.

67. Doležel J, Číhalíková J, Lucretti S. A high-yield procedure for isolation of metaphase chromosomes from root tips of *Vicia faba* L. Planta. 1992;188:93–98.

68. Loureiro J, Rodriguez E, Dolezel J, Santos C. Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species. Annals of Botany. 2007;100:875-888.

69. Suda J, Kron P, Husband BC, Trávníček P: Flow Cytometry and Ploidy: Applications in Plant Systematics, Ecology and Evolutionary Biology. *In Flow Cytometry with Plant Cells: Analysis of Genes, Chromosomes and Genomes.* Edited by Doležel J, Greilhuber J and Suda J. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany; 2007:103–130.

70. Greilhuber J, Temsch E, Loureiro J: Nuclear DNA content measurement. *In Flow cytometry with plant cells: analysis of genes, chromosomes and genomes*. Edited by Doležel J, Greilhuber J and Suda J. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany; 2007:67–101.

71. Greilhuber J, Dolezel J, Lysak MA, Bennett MD. The origin, evolution and proposed stabilization of the terms "genome size' and 'C-value' to describe nuclear DNA contents. Annals of Botany. 2005; 95:255-260.

72. Dolezel J, Bartos J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. Cytometry. A. 2003;51:127–8.

73. RAD_Tools repository [https://github.com/CoBiG2/RAD_Tools/commit/425ab4feca895430d30e102d03dcfaa8cb629523].

74. Zhang Z, Schwartz S, Wagner L, Miller W. A Greedy Algorithm for Aligning DNA Sequences. J. Comput. Biol. 2000;7:203–14.

75. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997. 25:3389–402.

76. Introne W, Boissy RE, Gahl WA. Clinical, Molecular, and Cell Biological Aspects of Chediak–Higashi Syndrome. Mol. Genet. Metab. 1999;68:283–303.

77. Gregory TR: Genome Size Evolution in Animals. *In The Evolution of Genome*. Edited by Gregory T. Elsevier, Burlington, Massachusetts; 2005:3–87.

78. Jiggins CD, McMillan WO. The genetic basis of an adaptive radiation: warning colour in two Heliconius species. Proc. R. Soc. B Biol. Sci. 1997;264:1167–75.

79. Joron M, Jiggins CD, Papanicolaou A, McMillan WO. Heliconius wing patterns: an evo-devo model for understanding phenotypic diversity. Heredity. 2006;97:157–67.

80. Comeault A a, Soria-Carrasco V, Gompert Z, Farkas TE, Buerkle CA, Parchman TL, et al. Genome-Wide Association Mapping of Phenotypic Traits Subject to a Range of Intensities of Natural Selection in *Timema cristinae* *. Am. Nat. 2014;183:711–27.

81. Bradshaw HD, Schemske DW. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature. 2003;426:176–8.

82. Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, et al. The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. PLoS Biol. 2004;2:e109.

83. Steiner CC, Weber JN, Hoekstra HE. Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes. PLoS Biol. 2007;5:e219.

84. Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, Baxter S, et al. A Conserved Supergene Locus Controls Colour Pattern Diversity in Heliconius Butterflies. PLoS Biol.

2006;4:e303.

85. Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature. 2011;477:203–6.

86. Kunte K, Zhang W, Tenger-Trolander A, Palmer DH, Martin A, Reed RD, et al. doublesex is a mimicry supergene. Nature. 2014;507:229–32.

87. Wittkopp PJ, Beldade P. Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy. Semin. Cell Dev. Biol. 2009;20:65–71.

88. Futahashi R, Sato J, Meng Y, Okamoto S, Daimon T, Yamamoto K, et al. yellow and ebony are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*. Genetics. 2008;180:1995–2005.

89. Futahashi R, Fujiwara H. Melanin-synthesis enzymes coregulate stage-specific larval cuticular markings in the swallowtail butterfly, *Papilio xuthus*. Dev. Genes Evol. 2005;215:519–29.

90. Clark R, Brown SM, Collins SC, Jiggins CD, Heckel DG, Vogler AP. Colour pattern specification in the Mocker swallowtail *Papilio dardanus*: the transcription factor invected is a candidate for the mimicry locus H. Proc. Biol. Sci. 2008;275:1181–8.

91. Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, et al. optix Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. Science. 2011;333:1137–41.

92. Wallbank RWR, Baxter SW, Pardo-Diaz C, Hanly JJ, Martin SH, Mallet J, et al. Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. PLoS Biol. 2016;14:e1002353.

93. Nadeau NJ, Pardo-Diaz C, Whibley A, Supple MA, Saenko S V., Wallbank RWR, et al. The gene cortex controls mimicry and crypsis in butterflies and moths. Nature. 2016;534:106–10.

94. Gompel N, Prud'homme B, Wittkopp PJ, Kassner V a, Carroll SB. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature. 2005;433:481–7.

95. Koshikawa S, Giorgianni MW, Vaccaro K, Kassner V a, Yoder JH, Werner T, et al. Gain of

cis -regulatory activities underlies novel domains of wingless gene expression in Drosophila. Proc. Natl. Acad. Sci. 2015;112:7524–9.

Table 1 – SNPs associated with dorsal colour morphs for each pairwise comparison and obtained through Single-SNP association tests using Bayesian regression approach.

Bayes factor values above 0.99 quantile ($BF_{0.99}$); Effect size of an individual SNP on the phenotype (β); Minor allele frequency for each locus and morph (maf); Mean effect size of $BF_{0.99}$ SNPs (Mean $BF_{0.99}$ SNPs); Mean effect size of all 1837 SNPs. SNPs common to comparisons are underlined.

MAR-TRI								MAR-TYP							
SNP_ID	**Minor Allele**	**Major Allele**	**$BF_{0.99}$**	**β**	**maf_TRI**	**maf_MAR**	**maf_TYP**	**SNP_ID**	**Minor Allele**	**Major Allele**	**$BF_{0.99}$**	**β**	**maf_TRI**	**maf_MAR**	**maf_TYP**
3950:1	G	A	0.432	-0.1198	0.1667	0.7727	0.3750	**3950:10**	T	A	0.264	-0.1010	0.1667	0.7273	0.3125
3950:10	T	A	0.337	-0.1099	0.1667	0.7273	0.3125	**7095:50**	C	T	0.319	0.1123	0.0556	0.0417	0.2000
22795:88	G	A	0.457	-0.1246	0.0500	0.5000	0.4286	**11381:9**	A	G	0.329	0.1131	0.1667	0.0000	0.2857
40633:18	A	G	0.365	0.1159	0.4500	0.0000	0.0714	**16628:65**	C	A	0.502	-0.1287	0.0000	0.5000	0.0000
41239:75	G	T	0.373	0.1170	0.3333	0.0000	0.1000	**20734:39**	T	C	0.463	0.1275	0.4286	0.3000	0.7778
43069:10	C	G	0.492	-0.1305	0.0000	0.4167	0.5714	**24668:63**	C	G	0.369	0.1173	0.3333	0.3182	0.7500
43069:34	A	T	0.492	-0.1305	0.0000	0.4167	0.5000	**35205:6**	G	C	0.444	0.1255	0.0625	0.2273	0.0000
50515:83	A	T	0.349	0.1165	0.2857	0.0000	0.0000	**45009:87**	T	G	0.365	0.1096	0.1500	0.1818	0.6500
54226:66	A	T	0.418	-0.1245	0.0000	0.3500	0.0000	**54049:70**	G	A	0.424	0.1231	0.5000	0.2727	0.7500
55187:79	G	T	0.490	0.1303	0.3889	0.0000	0.5000	**55187:46**	A	G	0.719	0.1555	0.2778	0.0000	0.5000
63439:28	C	A	0.403	-0.1203	0.1818	0.6000	0.3889	**55187:79**	G	T	0.784	0.1603	0.3889	0.0000	0.5000
63439:8	A	C	0.430	-0.1224	0.1364	0.5500	0.3500	**56842:83**	A	G	0.265	0.1053	0.2000	0.0455	0.2500
75897:50	A	G	0.339	0.1158	0.2500	0.0000	0.0000	**64204:16**	T	G	0.303	-0.1080	0.1250	0.5000	0.1111
75897:7	C	T	0.785	0.1581	0.5833	0.0000	0.0000	**64204:46**	G	T	0.296	-0.1072	0.2778	0.5000	0.1111
83460:19	C	T	0.376	-0.1173	0.0000	0.3889	0.0000	**64258:61**	G	A	0.376	0.1207	0.1500	0.0455	0.4286
87932:85	C	T	0.410	-0.1238	0.0000	0.5000	0.5556	**82682:38**	T	G	0.258	0.1019	0.4000	0.1111	0.5833
106126:52	C	T	0.337	0.1161	0.2500	0.0000	0.0000	**92187:65**	A	C	0.265	-0.1028	0.5000	0.6111	0.1875
124817:20	A	G	0.338	0.1159	0.3500	0.0000	0.0625	**102702:13**	T	A	0.314	0.1120	0.0000	0.0000	0.2143
126355:29	T	C	0.579	0.1369	0.5500	0.0455	0.1667	**104139:11**	T	A	0.267	-0.1040	0.3500	0.7000	0.3333
Mean BF0.99 SNPs				|0.1235|							|0.1177|			
Mean all SNPs				|0.0001|							|0.0037|			

TRI-TYP							
SNP_ID	**Minor Allele**	**Major Allele**	**$BF_{0.99}$**	**β**	**maf_TRI**	**maf_MAR**	**maf_TYP**
6535:26	T	C	0.270	-0.1035	0.4444	0.2222	0.0000
6535:35	G	A	0.270	0.1023	0.1667	0.5556	0.6250
22795:88	G	A	0.279	0.1075	0.0500	0.5000	0.4286
24031:66	C	T	0.292	0.1099	0.0000	0.1111	0.2000
24031:81	T	G	0.370	0.1202	0.0000	0.0556	0.2000
27816:86	G	A	0.352	-0.1155	0.6364	0.4375	0.1875
37095:26	T	G	0.313	-0.1110	0.4444	0.1000	0.0556
41742:86	C	G	0.316	0.1133	0.0000	0.4167	0.5714
43069:10	C	G	0.778	0.1577	0.0000	0.4167	0.5714
43069:34	A	T	0.717	0.1551	0.0000	0.4167	0.5000
43143:5	T	C	0.408	-0.1195	0.7222	0.4000	0.2500
45009:87	T	G	0.429	0.1180	0.1500	0.1818	0.6500
56752:20	G	A	0.431	-0.1223	0.6875	0.4444	0.1250
59359:24	G	A	0.447	0.1285	0.1000	0.2000	0.6000
75897:7	C	T	0.528	-0.1342	0.5833	0.0000	0.0000
87932:85	C	T	0.355	0.1135	0.0000	0.5000	0.5556
103746:74	T	A	0.496	0.1342	0.0909	0.4000	0.5833
118051:49	G	C	0.335	-0.1170	0.1000	0.1500	0.0000
118835:54	C	A	0.322	-0.1122	0.0455	0.3636	0.0000
123202:88	T	A	0.330	0.1130	0.0909	0.1000	0.4286
Mean BF0.99 SNPs				|0.1204|			
Mean all SNPs				|0.0028|			

Table 2 – Parameter estimates from Bayesian variable selection regression for each pairwise analysis.

Analyses	PVE	σSNP	σSNP_BF	nSNP	nSNP_BF	PIP SNP
MAR-TRI	0.6429 (0.031 – 0.998)	1.1200 (0.0570 – 5.559)	3.4300 (0.8475 – 11.8320)	67 (1 –268)	12 (2 – 31)	0.0366 (0.0320 – 0.0465)
MAR-TYP	0.6018 (0.027 – 0.995)	0.9495 (0.0520 – 4.0220)	2.4070 (0.8531 – 7.2788)	63 (1 – 264)	9 (2 – 26)	0.0345 (0.0303 – 0.0418)
TRI-TYP	0.6515 (0.035– 0.996)	0.9776 (0.0570– 4.4040)	4.1420 (1.6660 – 8.7020)	66 (1 – 263)	10 (2 – 25)	0.0361 (0.0320 – 0.0448)

Proportion of variance explained (PVE); mean phenotypic effect associated with a SNP in the regression model including all models (σSNP) and models with a $\log_{10}(BF) > 10$ (σSNP_BF); mean number of SNPs in the model considering all models (nSNP) and models with a $\log_{10}(BF) > 10$ (nSNP_BF) and; mean posterior inclusion probability associated to SNPs in the model (PIP). 95% empirical quantiles are reported in parenthesis.

Table 3 - SNPs $PIP_{0.99}$ associated with dorsal colour morphs obtained through Multi-SNP association tests using Bayesian regression approach.

Posterior inclusion probability associated to SNP (PIP); Effect size of an individual SNP on the phenotype (β) and minor allele frequency for each locus and morph (maf).

MARTRI

SNP_ID	PIP	β	Minor Allele	Major Allele	MAF TRI	MAF MAR	MAF TYP
41239:75	0.05228	0.73688	G	T	0.3333	0.0000	0.1000
50515:83	0.05294	0.70499	A	T	0.2857	0.0000	0.0000
55187:46	0.05326	0.67185	A	G	0.2778	0.0000	0.5000
55187:79	0.07371	1.18427	G	T	0.3889	0.0000	0.5000
69098:53	0.0545	4.0856	C	A	0.1818	0.0000	0.2778
75897:7	0.13035	1.89186	C	T	0.5833	0.0000	0.0000
94147:30	0.05298	0.7147	G	A	0.6818	0.3182	0.3889
106126:52	0.05356	1.10651	C	T	0.2500	0.0000	0.0000
126355:29	0.08618	3.89385	T	C	0.5500	0.0455	0.1667
3950:1	0.05946	-0.86693	G	A	0.1667	0.7727	0.3750
22795:88	0.06048	-0.98089	G	A	0.0500	0.5000	0.4286
25027:11	0.05188	-1.28702	T	A	0.0000	0.1667	0.1667
43069:10	0.07528	-2.3343	C	G	0.0000	0.4167	0.5714
43069:34	0.06763	-1.50904	A	T	0.0000	0.4167	0.5000
54226:66	0.05956	-1.00561	A	T	0.0000	0.3500	0.0000
63439:8	0.05754	-0.73234	A	C	0.1364	0.5500	0.3500
83460:19	0.05406	-0.72601	C	T	0.0000	0.3889	0.0000
87932:85	0.05747	-0.72742	C	T	0.0000	0.5000	0.5556
103246:16	0.05827	-1.14876	T	C	0.0000	0.3571	0.1000

MARTYP

SNP_ID	PIP	β	Minor Allele	Major Allele	MAF TRI	MAF MAR	MAF TYP
7095:50	0.04707	0.65574	C	T	0.0556	0.0417	0.2000
11381:9	0.04714	0.61586	A	G	0.1667	0.0000	0.2857
20734:39	0.0621	1.06957	T	C	0.4286	0.3000	0.7778
23155:83	0.04559	0.58512	T	G	0.1111	0.1000	0.5000
24668:63	0.05372	0.84822	C	G	0.3333	0.3182	0.7500
27059:59	0.04681	1.00875	T	C	0.2222	0.2083	0.5000
35205:6	0.06037	0.89098	G	C	0.0625	0.2273	0.0000
54049:70	0.0568	0.76067	G	A	0.5000	0.2727	0.7500
55187:46	0.11972	1.88694	A	G	0.2778	0.0000	0.5000
55187:79	0.13314	1.66567	G	T	0.3889	0.0000	0.5000
64258:61	0.05429	0.96162	G	A	0.1500	0.0455	0.4286
102702:13	0.04718	1.03275	T	A	0.0000	0.0000	0.2143
104623:88	0.04711	0.70632	G	A	0.1875	0.0000	0.1429
108304:78	0.04626	0.48985	C	A	0.3571	0.2917	0.6250
16628:65	0.06209	-0.89022	C	A	0.0000	0.5000	0.0000
51349:15	0.04563	-0.59847	T	C	0.1667	0.2500	0.0000
64204:16	0.04703	-0.60862	T	G	0.1250	0.5000	0.1111
64204:46	0.04626	-0.59146	G	T	0.2778	0.5000	0.1111
66105:38	0.04521	-0.53359	A	C	0.3000	0.2143	0.0500

TRITYP

SNP_ID	PIP	β	Minor Allele	Major Allele	MAF TRI	MAF MAR	MAF TYP
6535:35	0.0534	0.92979	G	A	0.1667	0.5556	0.6250
22795:88	0.0519	0.71932	G	A	0.0500	0.5000	0.4286
24031:66	0.05241	1.12531	C	T	0.0000	0.1111	0.2000
24031:81	0.05745	0.84617	T	G	0.0000	0.0556	0.2000
41742:86	0.05111	0.74532	C	G	0.0000	0.4167	0.5714
43069:10	0.09056	1.35007	A	T	0.0000	0.4167	0.5000
43069:34	0.09699	1.79618	A	T	0.0000	0.4167	0.5000
45009:87	0.05206	0.64932	T	G	0.1500	0.1818	0.6500
59359:24	0.05896	1.0716	G	A	0.1000	0.2000	0.6000
87932:85	0.05394	0.98257	C	T	0.0000	0.5000	0.5556
103746:74	0.07094	1.25641	T	A	0.0909	0.4000	0.5833
123202:88	0.05166	0.6811	T	A	0.0909	0.1000	0.4286
27816:86	0.05363	-0.79657	G	A	0.6364	0.4375	0.1875
37095:26	0.05663	-0.84607	T	G	0.4444	0.1000	0.0556
43143:5	0.05286	-0.67184	T	C	0.7222	0.4000	0.2500
56752:20	0.05638	-0.73623	G	A	0.6875	0.4444	0.1250
75897:7	0.07396	-1.31379	C	T	0.5833	0.0000	0.0000
118051:49	0.05605	-0.82887	G	C	0.1000	0.1500	0.0000
118835:54	0.05156	-0.68197	C	A	0.0455	0.3636	0.0000

List of Figure Legends

Figure 1 – Bayes factor for each SNP in each pairwise comparison in single-SNP association tests. (a) MAR vs. TRI; (b) MAR vs. TYP; and (c) TRI vs. TYP. The horizontal straight lines correspond to the Bayes factor 95% empirical quantile threshold and the dash lines to the 99% empirical quantile.

Figure 2 – Genetic variation of the 33 individuals summarized on principal component axis 1 (PC1) and 2 (PC2) from a Principal Component Analysis using SNPs identified through Bayesian regression analyses. (a) All 1,837 SNPs; (b) 50 SNPs $BF_{0.99}$ identified in Single-SNP association tests; (c) 50 SNPs $PIP_{0.99}$ identified in Multi-SNP Association tests; and (d) 40 SNPs shared between both association analyses.

Figure 3 – Posterior inclusion probabilities (PIPs) for each SNP in each pairwise comparison in multi-SNP association tests. (a) MAR vs. TRI; (b) MAR vs. TYP; and (c) TRI vs. TYP. The horizontal straight lines correspond to the PIP 95% empirical quantile threshold and the dash lines to the 99% empirical quantile.

Figure 1

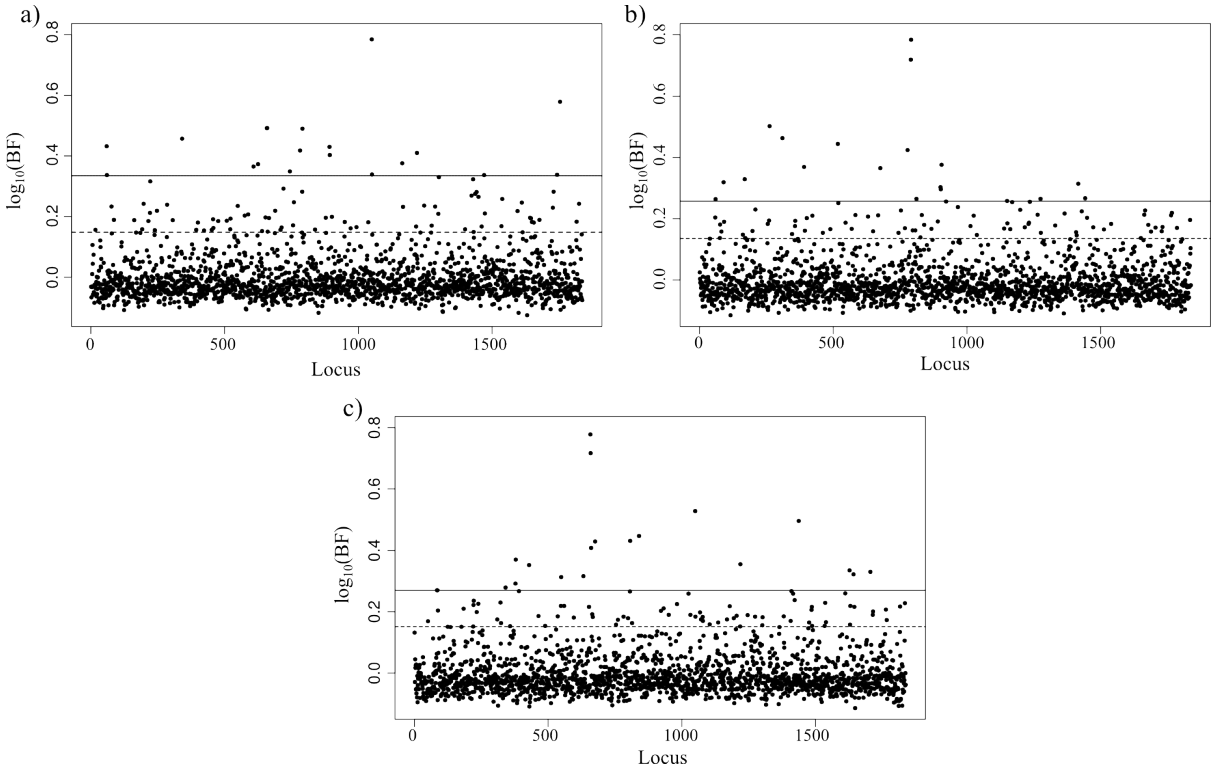


Figure 2

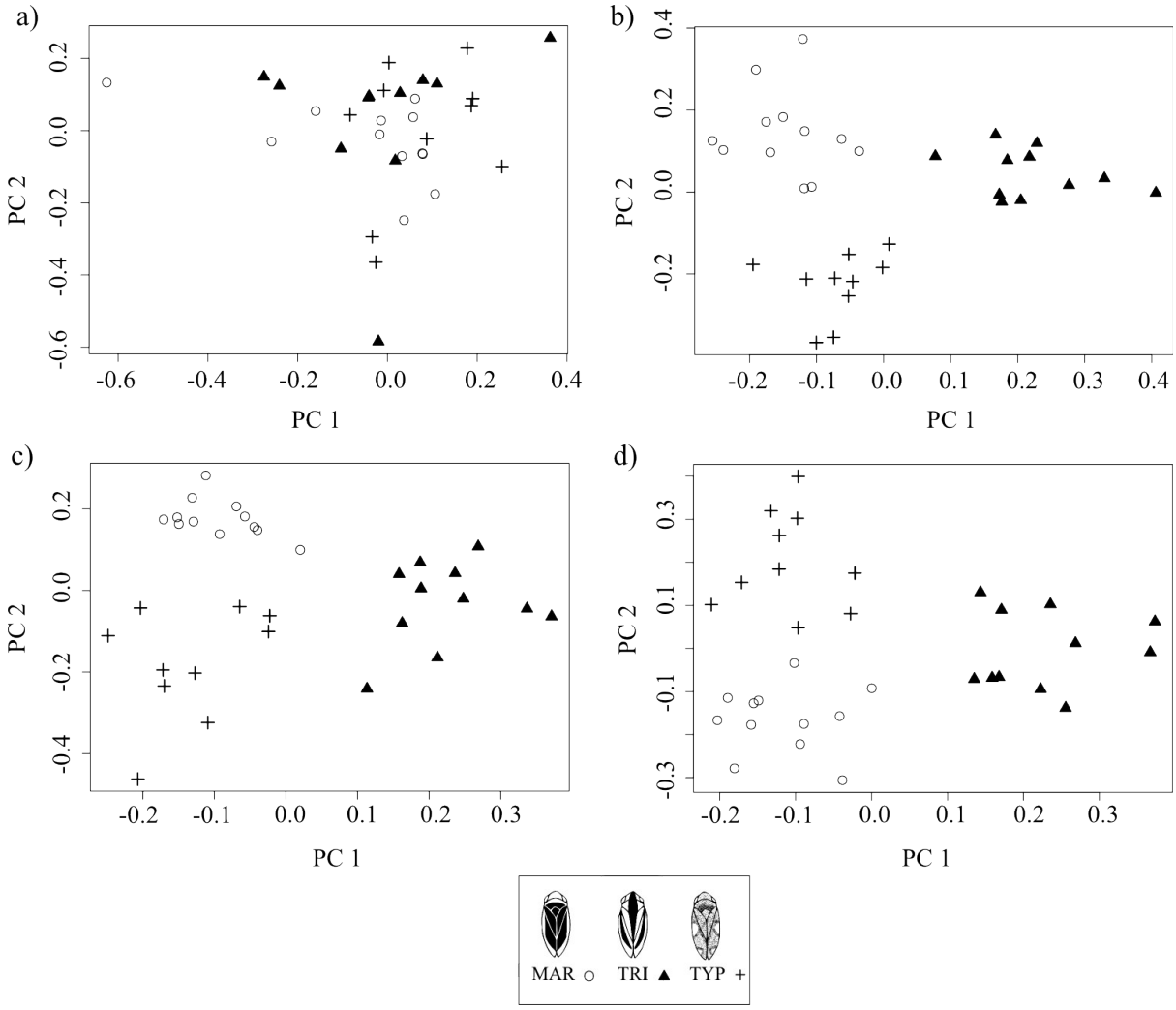
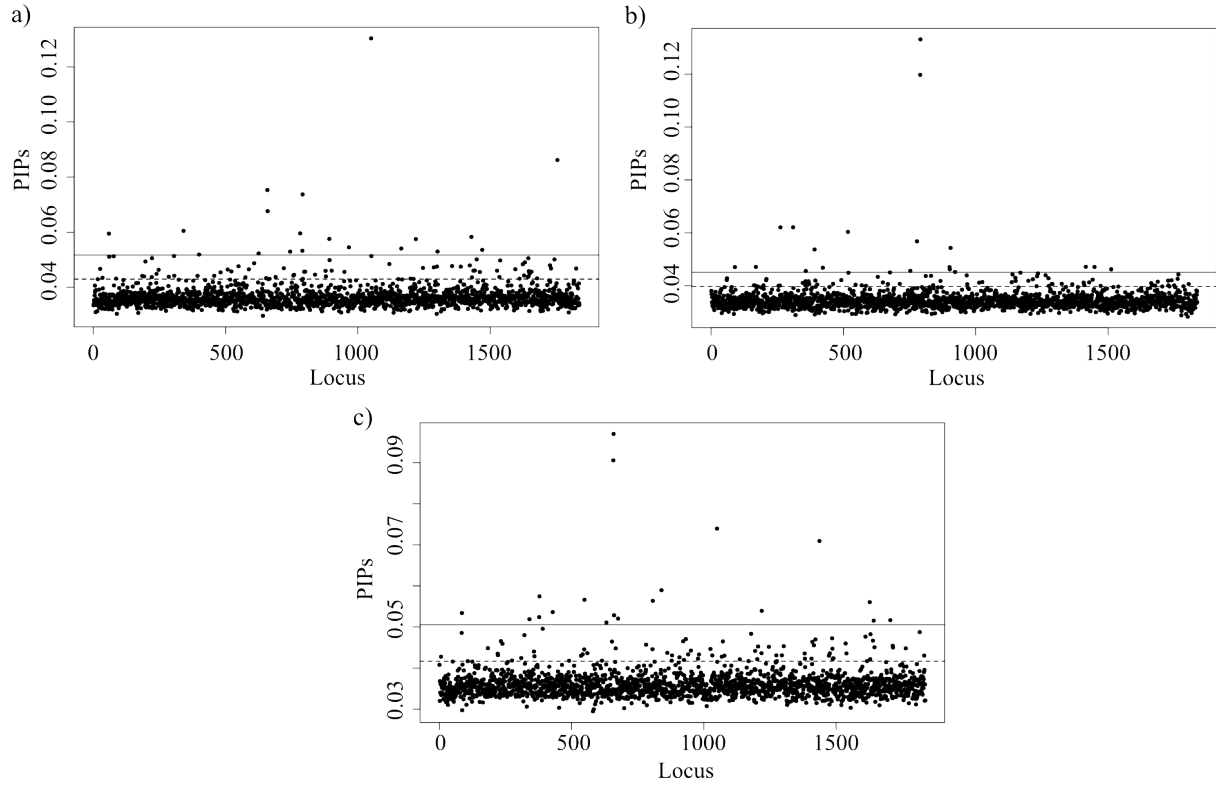


Figure 3


a)
PIPs
0.04
0.06
0.08
0.10
0.12
0
500
1000
1500
Locus
b)
PIPs
0.04
0.06
0.08
0.10
0.12
0
500
1000
1500
Locus
c)
PIPs
0.03
0.05
0.07
0.09
0
500
1000
1500
Locus

CHAPTER 4

Supporting Information

Additional File 1

Caption for entire file:

Histograms of the total number of raw reads, mean depth and proportion of missing data per individual and of the R^2 values for each colour-associated SNP comparison; scatterplots of the number of SNPs in the model as a function of the Bayes factor for each pairwise comparison in multi-SNP association tests; and number of RAD loci hits with genome and transcriptome.

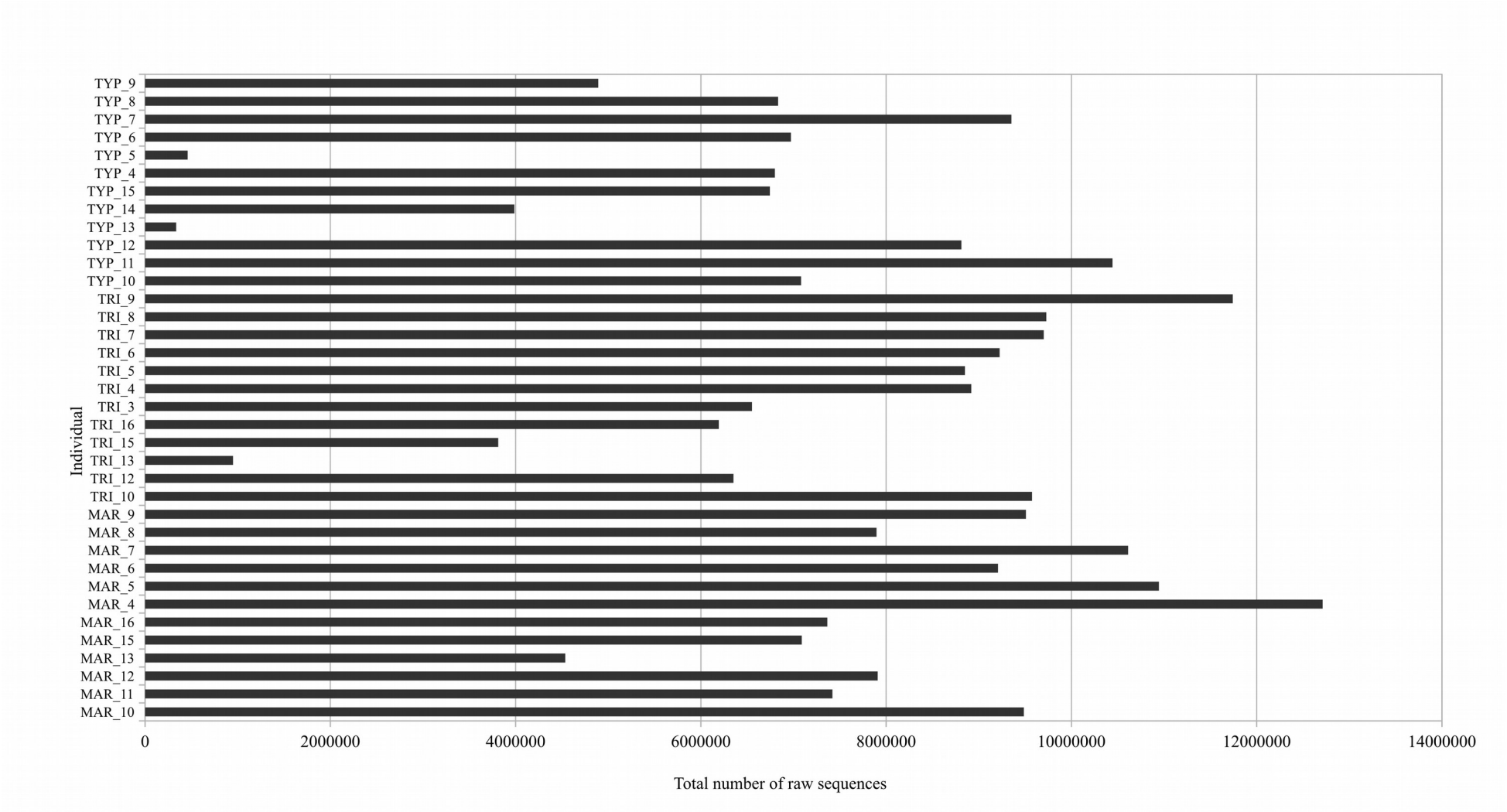


Figure S1 – Total number of raw reads per individual.

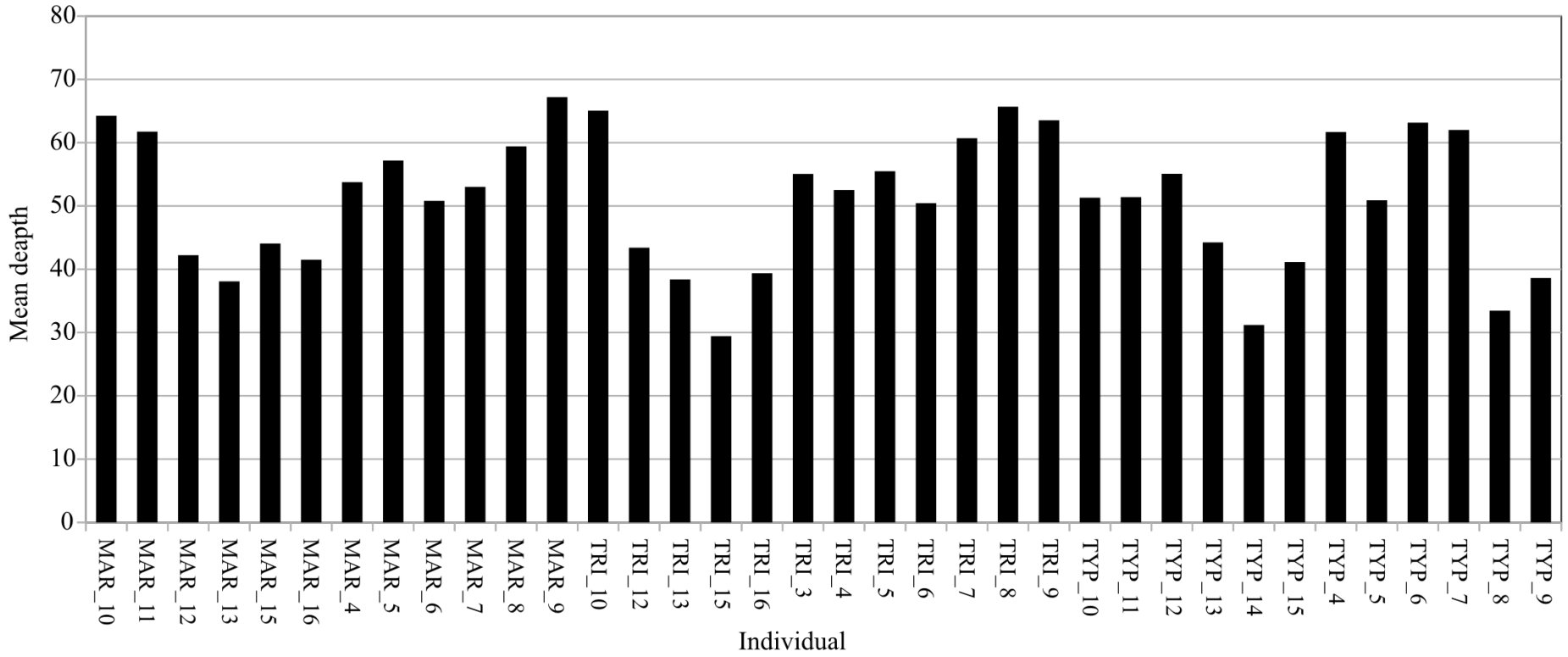


Figure S2 – Mean depth per individual.

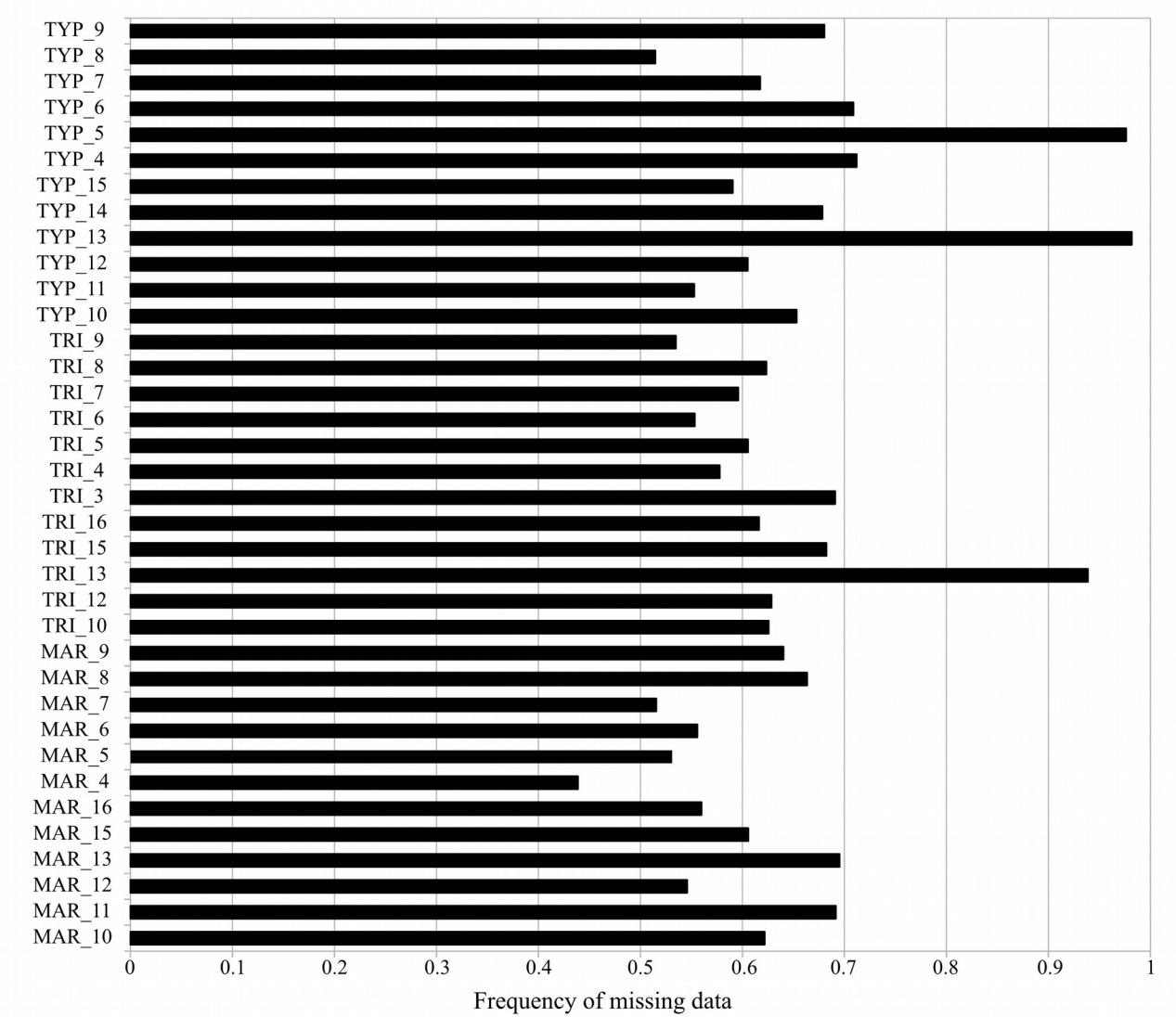


Figure S3 – Proportion of missing data per individual

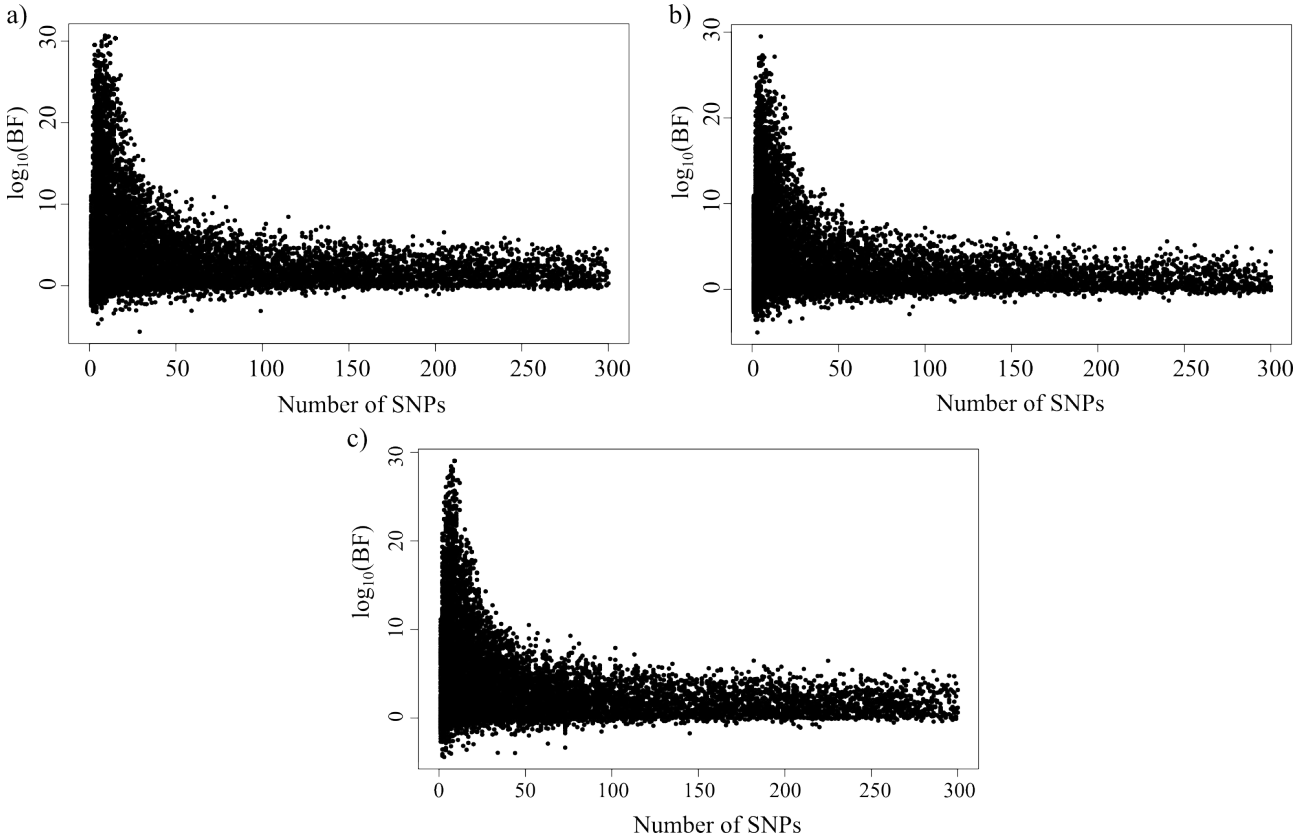


Figure S4 – Number of SNPs in the model as a function of the Bayes factor for each pairwise comparison in multi-SNP association tests. (a) MAR vs TRI; (b) MAR vs TYP; and (c) TRI vs TYP.

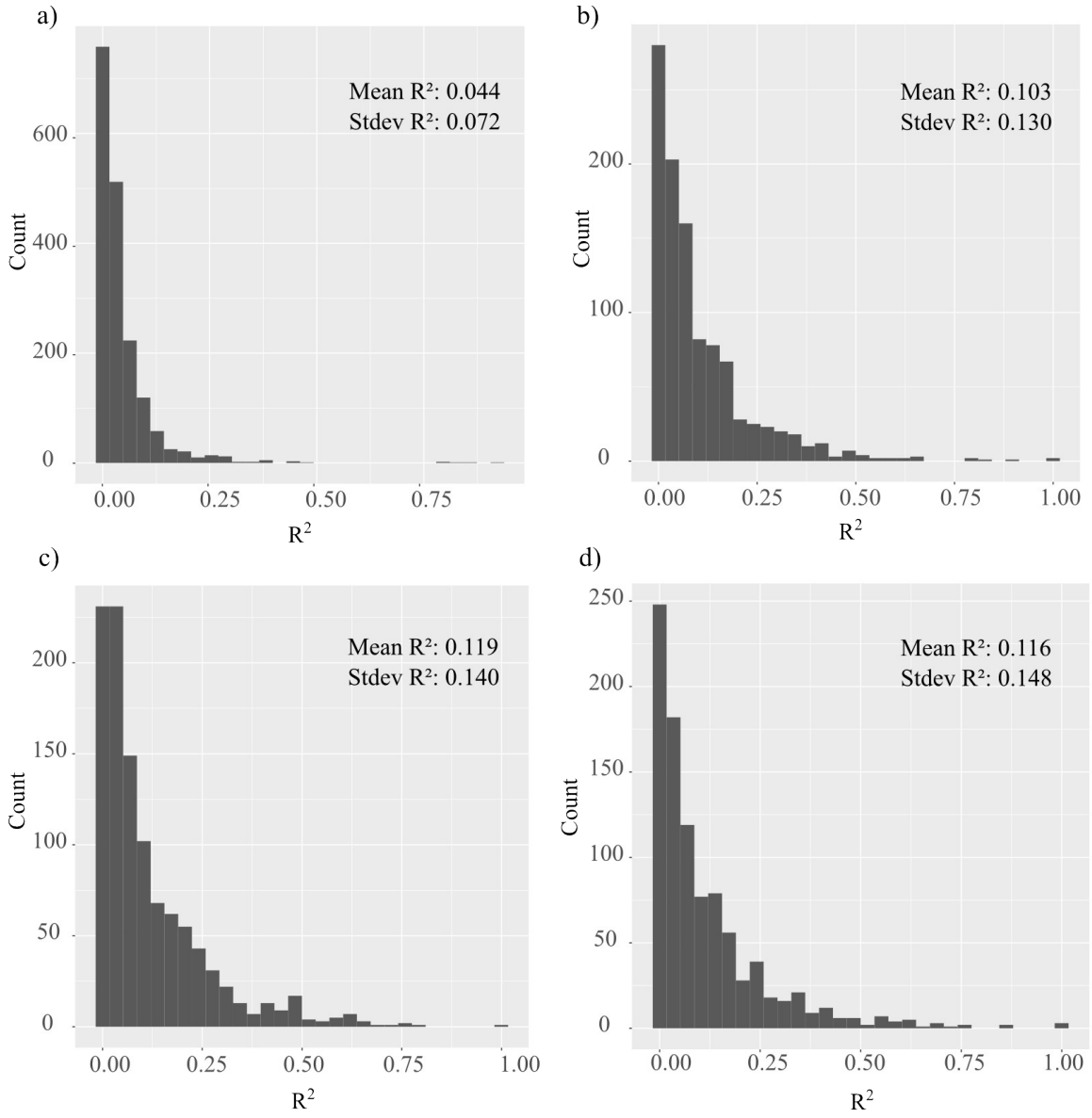


Figure S5 – Histograms with R^2 values for each SNP comparison. Values for the 60 SNPs associated with dorsal colour phenotypes in single and multi-association tests using (a) all individuals; (b) only *marginellus* individuals; (c) only *trilineatus* individuals; and (d) only *typicus* individuals.

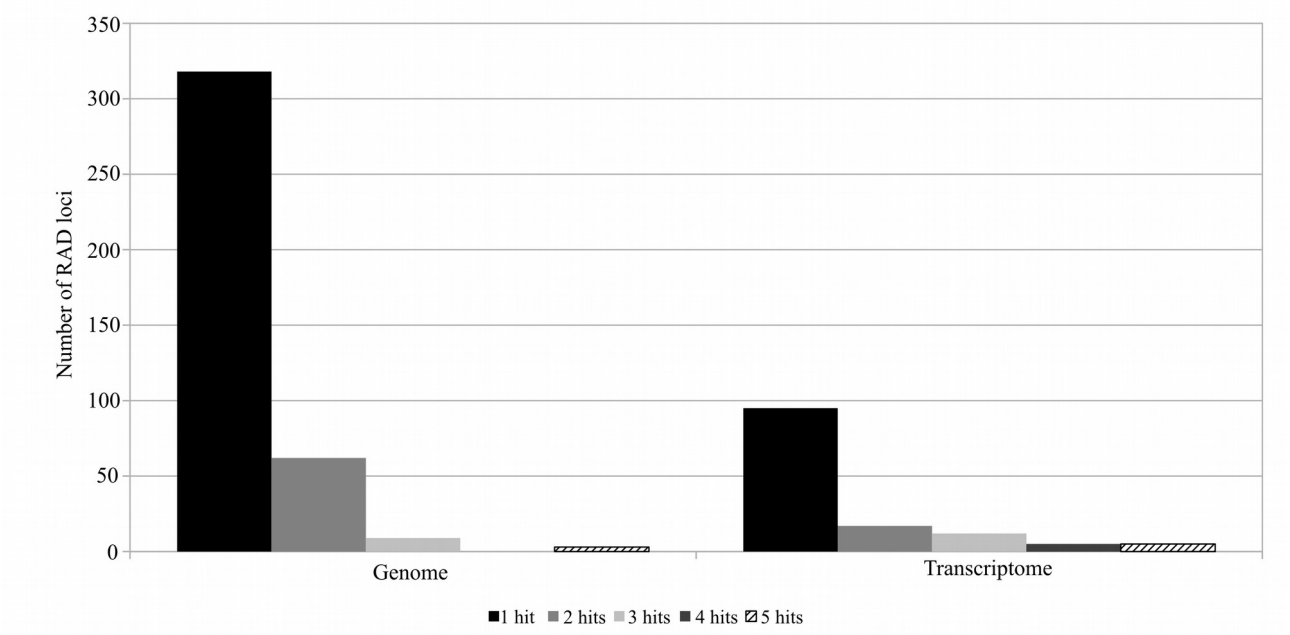


Figure S6 – Number of RAD loci hits with genome and transcriptome contigs/scaffolds.

Additional File 2

Table S1 – Associated SNPs obtained for each pairwise comparison through Fisher's exact test, no significant after a 5% FDR correction.

Estimated odds ratio for minor allele (OR); Standard error for odds ration (SE); Lower bound of 99% confidence interval for odds ratio (L99); Upper bound of 99% confidence interval for odds ratio (U99).

SNP_ID	Minor Allele	Major Allele	P-value	Q-value	OR	SE	L99	U99	MAR-TRI
3950:1	G	A	0.001	0.777	17.000	0.927	2.765	104.500	
3950:10	T	A	0.003	0.990	13.330	0.911	2.238	79.440	
3950:42	G	T	0.036	1.000	NA	NA	NA	NA	
5407:80	C	T	0.035	1.000	5.667	0.784	1.219	26.330	
5480:14	A	G	0.036	1.000	0.000	inf	0.000	nan	
6535:35	G	A	0.035	1.000	6.250	0.791	1.327	29.430	
11205:42	T	C	0.020	1.000	0.000	inf	0.000	nan	
11244:17	C	T	0.025	1.000	0.191	0.728	0.046	0.793	
12131:31	G	C	0.020	1.000	6.000	0.764	1.343	26.810	
12421:26	A	G	0.044	1.000	10.230	1.128	1.121	93.340	
14632:5	A	C	0.017	1.000	NA	NA	NA	NA	
14632:66	C	T	0.007	1.000	NA	NA	NA	NA	
15962:90	C	T	0.037	1.000	0.245	0.659	0.067	0.891	
16195:63	T	C	0.048	1.000	0.000	inf	0.000	nan	
16628:65	C	A	0.004	0.998	NA	NA	NA	NA	
20573:80	A	T	0.031	1.000	0.000	inf	0.000	nan	
22795:88	G	A	0.003	0.971	19.000	1.129	2.078	173.700	
23056:51	A	C	0.047	1.000	0.000	inf	0.000	nan	
27421:1	C	T	0.042	1.000	0.200	0.733	0.048	0.842	
27421:54	G	A	0.041	1.000	0.000	inf	0.000	nan	
30435:35	T	A	0.049	1.000	0.143	0.900	0.024	0.833	
36583:13	G	A	0.005	0.998	NA	NA	NA	NA	
36583:91	G	A	0.005	0.998	NA	NA	NA	NA	
37095:26	T	G	0.027	1.000	0.139	0.884	0.025	0.785	
38603:51	A	C	0.013	1.000	NA	NA	NA	NA	
39061:36	A	G	0.047	1.000	0.000	inf	0.000	nan	
40633:18	A	G	0.001	0.819	0.000	inf	0.000	nan	
41239:75	G	T	0.004	0.993	0.000	inf	0.000	nan	
41889:55	T	C	0.029	1.000	NA	NA	NA	NA	
42568:80	C	T	0.022	1.000	0.000	inf	0.000	nan	
43069:10	C	G	0.001	0.591	NA	NA	NA	NA	
43069:34	A	T	0.001	0.591	NA	NA	NA	NA	
43143:1	A	T	0.048	1.000	NA	NA	NA	NA	
46573:72	G	A	0.048	1.000	NA	NA	NA	NA	
47224:67	A	G	0.018	1.000	NA	NA	NA	NA	
47824:70	A	C	0.045	1.000	4.714	0.753	1.077	20.630	
48889:5	G	C	0.004	0.997	0.000	inf	0.000	nan	
50515:83	A	T	0.014	1.000	0.000	inf	0.000	nan	
51251:64	A	G	0.041	1.000	0.000	inf	0.000	nan	
51349:87	G	T	0.031	1.000	NA	NA	NA	NA	
54226:66	A	T	0.003	0.982	NA	NA	NA	NA	
54456:51	A	T	0.040	1.000	0.257	0.645	0.073	0.910	
55187:46	A	G	0.010	1.000	0.000	inf	0.000	nan	
55187:79	G	T	0.001	0.806	0.000	inf	0.000	nan	
55519:56	T	A	0.035	1.000	6.667	0.864	1.227	36.230	
60545:46	G	A	0.045	1.000	9.154	1.131	0.998	83.970	
63439:28	C	A	0.010	1.000	6.750	0.717	1.656	27.510	
63439:8	A	C	0.008	1.000	7.741	0.767	1.722	34.790	
63685:86	T	C	0.009	1.000	0.000	inf	0.000	nan	
64204:16	T	G	0.030	1.000	7.000	0.891	1.221	40.120	
66474:79	T	G	0.030	1.000	NA	NA	NA	NA	
69098:53	C	A	0.045	1.000	0.000	inf	0.000	nan	
70991:59	A	T	0.008	1.000	NA	NA	NA	NA	
74826:17	T	C	0.045	1.000	0.000	inf	0.000	nan	
75133:13	A	G	0.020	1.000	6.111	0.776	1.336	27.960	
75897:50	A	G	0.037	1.000	0.000	inf	0.000	nan	
75897:7	C	T	0.000	0.186	0.000	inf	0.000	nan	
77054:5	G	T	0.041	1.000	0.000	inf	0.000	nan	
77054:74	A	T	0.026	1.000	10.000	1.129	1.094	91.440	

Table S1 – Associated SNPs obtained for each pairwise comparison through Fisher's exact test, no significant after a 5% FDR correction (*cont.*).

Estimated odds ratio for minor allele (OR); Standard error for odds ration (SE); Lower bound of 99% confidence interval for odds ratio (L99); Upper bound of 99% confidence interval for odds ratio (U99).

SNP_ID	Minor Allele	Major Allele	P-value	Q-value	OR	SE	L99	U99	MAR-TRI
80830:68	T	C	0.040	1.000	5.833	0.869	1.063	32.020	
81088:18	A	T	0.027	1.000	9.714	1.120	1.081	87.310	
83460:19	C	T	0.003	0.970	NA	NA	NA	NA	
86470:57	T	G	0.017	1.000	0.000	inf	0.000	nan	
87569:17	A	C	0.009	1.000	NA	NA	NA	NA	
87932:85	C	T	0.002	0.967	NA	NA	NA	NA	
88696:40	C	A	0.036	1.000	0.000	inf	0.000	nan	
89787:21	T	A	0.020	1.000	12.670	1.123	1.402	114.400	
89892:81	G	T	0.035	1.000	0.098	1.151	0.010	0.936	
90363:3	A	T	0.033	1.000	0.000	inf	0.000	nan	
92099:34	T	C	0.041	1.000	NA	NA	NA	NA	
92238:67	A	T	0.033	1.000	NA	NA	NA	NA	
93686:64	A	G	0.041	1.000	5.000	0.730	1.195	20.920	
94147:30	G	A	0.034	1.000	0.218	0.647	0.061	0.775	
103246:16	T	C	0.010	1.000	NA	NA	NA	NA	
103746:74	T	A	0.030	1.000	6.667	0.871	1.210	36.740	
105384:90	T	C	0.041	1.000	10.820	1.137	1.165	100.400	
106126:52	C	T	0.020	1.000	0.000	inf	0.000	nan	
106126:64	C	T	0.001	0.785	NA	NA	NA	NA	
111157:13	A	G	0.019	1.000	NA	NA	NA	NA	
111157:86	A	C	0.012	1.000	10.000	0.887	1.756	56.930	
115255:45	G	A	0.020	1.000	NA	NA	NA	NA	
116514:82	T	C	0.020	1.000	NA	NA	NA	NA	
118835:54	C	A	0.021	1.000	0.083	1.115	0.009	0.742	
118835:73	C	T	0.021	1.000	0.083	1.115	0.009	0.742	
120070:36	A	T	0.003	0.992	0.000	inf	0.000	nan	
120286:48	G	A	0.033	1.000	0.000	inf	0.000	nan	
124817:20	A	G	0.003	0.982	0.000	inf	0.000	nan	
126355:29	T	C	0.000	0.431	0.039	1.118	0.004	0.349	
130757:44	T	C	0.025	1.000	10.380	1.120	1.156	93.290	
130757:61	C	A	0.025	1.000	0.000	inf	0.000	nan	

Table S1 – Associated SNPs obtained for each pairwise comparison through Fisher's exact test, no significant after a 5% FDR correction (*cont.*).

Estimated odds ratio for minor allele (OR); Standard error for odds ration (SE); Lower bound of 99% confidence interval for odds ratio (L99); Upper bound of 99% confidence interval for odds ratio (U99).

SNP_ID	Minor Allele	Major Allele	P-value	Q-value	OR	SE	L99	U99	MAR-TYP
3950:1	G	A	0.020	1.000	5.667	0.725	1.369	23.460	
3950:10	T	A	0.020	1.000	5.867	0.721	1.427	24.110	
5407:44	T	C	0.048	1.000	NA	NA	NA	NA	
5407:80	C	T	0.027	1.000	8.000	0.886	1.409	45.410	
6195:66	T	C	0.044	1.000	0.000	inf	0.000	nan	
10593:75	T	G	0.043	1.000	0.000	inf	0.000	nan	
11381:70	T	G	0.042	1.000	0.200	0.733	0.048	0.842	
11381:9	A	G	0.014	1.000	0.000	inf	0.000	nan	
11396:42	A	T	0.031	1.000	0.000	inf	0.000	nan	
11425:84	T	C	0.013	1.000	NA	NA	NA	NA	
13813:20	T	A	0.035	1.000	5.833	0.767	1.298	26.220	
16628:2	A	G	0.021	1.000	NA	NA	NA	NA	
16628:65	C	A	0.000	0.486	NA	NA	NA	NA	
20573:80	A	T	0.041	1.000	0.000	inf	0.000	nan	
20734:39	T	C	0.004	0.998	0.122	0.748	0.028	0.531	
22879:47	A	T	0.049	1.000	NA	NA	NA	NA	
23155:23	T	A	0.048	1.000	NA	NA	NA	NA	
23155:79	A	G	0.003	0.969	0.000	inf	0.000	nan	
23155:83	T	G	0.011	1.000	0.111	0.882	0.020	0.626	
24155:20	G	A	0.031	1.000	0.000	inf	0.000	nan	
24668:63	C	G	0.020	1.000	0.156	0.737	0.037	0.659	
25027:36	A	G	0.044	1.000	5.200	0.726	1.253	21.570	
31166:54	G	T	0.028	1.000	0.000	inf	0.000	nan	
33164:47	T	C	0.049	1.000	0.161	0.840	0.031	0.834	
34535:49	A	G	0.016	1.000	NA	NA	NA	NA	
35205:28	C	T	0.020	1.000	0.177	0.725	0.043	0.731	
35205:6	A	T	0.008	1.000	0.134	0.741	0.031	0.572	
35580:77	C	T	0.048	1.000	NA	NA	NA	NA	
38512:7	C	G	0.027	1.000	0.150	0.832	0.029	0.766	
38603:51	A	C	0.012	1.000	NA	NA	NA	NA	
41742:47	A	G	0.018	1.000	13.750	1.147	1.452	130.200	
41742:86	T	C	0.028	1.000	0.000	inf	0.000	nan	
42253:58	C	T	0.035	1.000	0.171	0.767	0.038	0.771	
43343:68	A	G	0.029	1.000	7.222	0.823	1.440	36.220	
45009:87	T	G	0.004	0.996	0.120	0.725	0.029	0.495	
47799:36	T	C	0.045	1.000	0.000	inf	0.000	nan	
51349:15	T	C	0.039	1.000	NA	NA	NA	NA	
54049:70	G	A	0.008	1.000	0.125	0.750	0.029	0.544	
55187:46	A	G	0.001	0.712	0.000	inf	0.000	nan	
55187:79	G	T	0.001	0.712	0.000	inf	0.000	nan	
55519:56	T	A	0.012	1.000	12.500	1.118	1.397	111.800	
59359:24	G	A	0.045	1.000	0.167	0.854	0.031	0.889	
59359:5	A	G	0.011	1.000	NA	NA	NA	NA	
59615:31	T	A	0.039	1.000	0.000	inf	0.000	nan	
59756:39	A	G	0.036	1.000	NA	NA	NA	NA	
64204:16	T	G	0.027	1.000	8.000	0.886	1.409	45.410	
64204:46	G	T	0.027	1.000	8.000	0.886	1.409	45.410	
64258:61	G	A	0.008	1.000	0.063	1.157	0.007	0.614	
69098:53	C	A	0.010	1.000	0.000	inf	0.000	nan	
69168:75	C	T	0.020	1.000	NA	NA	NA	NA	
73849:2	G	T	0.017	1.000	0.152	0.747	0.035	0.655	
79183:11	T	A	0.031	1.000	0.000	inf	0.000	nan	

Table S1 – Associated SNPs obtained for each pairwise comparison through Fisher's exact test, no significant after a 5% FDR correction (*cont.*).

Estimated odds ratio for minor allele (OR); Standard error for odds ration (SE); Lower bound of 99% confidence interval for odds ratio (L99); Upper bound of 99% confidence interval for odds ratio (U99).

SNP_ID	Minor Allele	Major Allele	P-value	Q-value	OR	SE	L99	U99	MAR-TYP
81653:72	G	A	0.049	1.000	NA	NA	NA	NA	
82682:38	T	G	0.013	1.000	0.089	0.952	0.014	0.576	
82682:56	T	C	0.018	1.000	0.000	inf	0.000	nan	
83460:19	C	T	0.024	1.000	NA	NA	NA	NA	
86470:57	T	G	0.004	0.997	0.000	inf	0.000	nan	
87569:17	A	C	0.047	1.000	8.867	1.123	0.981	80.180	
88684:50	T	A	0.045	1.000	9.154	1.131	0.998	83.970	
89026:32	G	T	0.003	0.990	0.000	inf	0.000	nan	
90805:28	T	C	0.018	1.000	0.088	1.126	0.010	0.803	
92187:65	A	C	0.017	1.000	6.810	0.803	1.413	32.830	
96962:7	A	C	0.022	1.000	0.000	inf	0.000	nan	
101939:54	T	C	0.030	1.000	0.143	0.891	0.025	0.819	
102198:61	A	G	0.008	1.000	0.062	1.141	0.007	0.576	
102702:13	T	A	0.043	1.000	0.000	inf	0.000	nan	
103460:91	C	A	0.039	1.000	0.000	inf	0.000	nan	
104139:11	T	A	0.050	1.000	4.667	0.699	1.187	18.350	
105240:56	A	T	0.031	1.000	0.000	inf	0.000	nan	
105384:90	T	C	0.010	1.000	NA	NA	NA	NA	
119443:36	G	C	0.025	1.000	NA	NA	NA	NA	
122410:33	A	G	0.025	1.000	0.000	inf	0.000	nan	
122441:74	T	A	0.030	1.000	NA	NA	NA	NA	
123202:88	T	A	0.042	1.000	0.148	0.920	0.024	0.900	
124278:24	T	C	0.008	1.000	0.000	inf	0.000	nan	
124812:88	T	G	0.031	1.000	0.000	inf	0.000	nan	
127190:89	A	C	0.028	1.000	0.000	inf	0.000	nan	
127431:25	G	C	0.043	1.000	0.000	inf	0.000	nan	
130525:56	T	C	0.039	1.000	0.000	inf	0.000	nan	
131716:37	G	A	0.039	1.000	0.000	inf	0.000	nan	
132574:59	G	A	0.008	1.000	0.094	0.924	0.015	0.574	

Table S1 – Associated SNPs obtained for each pairwise comparison through Fisher's exact test, no significant after a 5% FDR correction (*cont.*).

Estimated odds ratio for minor allele (OR); Standard error for odds ration (SE); Lower bound of 99% confidence interval for odds ratio (L99); Upper bound of 99% confidence interval for odds ratio (U99).

SNP_ID	Minor Allele	Major Allele	P-value	Q-value	OR	SE	L99	U99	TRI-TYP
6535:26	T	C	0.003	0.985	NA	NA	NA	NA	
6535:35	G	A	0.012	1.000	0.120	0.817	0.024	0.595	
7111:45	G	A	0.018	1.000	0.000	inf	0.000	nan	
9382:30	G	A	0.046	1.000	NA	NA	NA	NA	
9495:57	T	C	0.046	1.000	0.143	0.926	0.023	0.877	
10593:75	T	G	0.043	1.000	0.000	inf	0.000	nan	
11425:84	T	C	0.041	1.000	NA	NA	NA	NA	
11717:84	T	A	0.025	1.000	0.125	0.880	0.022	0.702	
14153:54	A	G	0.025	1.000	10.830	1.123	1.200	97.800	
14632:5	A	C	0.044	1.000	0.000	inf	0.000	nan	
14632:66	C	T	0.044	1.000	0.000	inf	0.000	nan	
22795:88	G	A	0.012	1.000	0.070	1.159	0.007	0.681	
23155:83	T	G	0.027	1.000	0.125	0.886	0.022	0.710	
23577:46	A	C	0.039	1.000	NA	NA	NA	NA	
23772:67	T	A	0.033	1.000	0.000	inf	0.000	nan	
24668:63	C	G	0.020	1.000	0.167	0.764	0.037	0.745	
27816:86	G	A	0.009	1.000	7.583	0.779	1.648	34.900	
31166:54	G	T	0.022	1.000	0.000	inf	0.000	nan	
33164:47	T	C	0.050	1.000	0.165	0.868	0.030	0.903	
33497:69	T	G	0.041	1.000	0.000	inf	0.000	nan	
34753:54	A	G	0.043	1.000	NA	NA	NA	NA	
36283:34	G	C	0.047	1.000	0.225	0.704	0.057	0.895	
36583:13	G	A	0.005	0.999	0.000	inf	0.000	nan	
36583:91	G	A	0.037	1.000	0.000	inf	0.000	nan	
37095:11	T	C	0.041	1.000	10.820	1.137	1.165	100.400	
37095:26	T	G	0.018	1.000	13.600	1.133	1.476	125.300	
38067:37	T	C	0.028	1.000	NA	NA	NA	NA	
38305:87	T	C	0.041	1.000	0.000	inf	0.000	nan	
40633:18	A	G	0.024	1.000	10.640	1.131	1.159	97.590	
41742:86	T	C	0.017	1.000	0.000	inf	0.000	nan	
42137:43	A	G	0.047	1.000	0.000	inf	0.000	nan	
42568:80	C	T	0.028	1.000	NA	NA	NA	NA	
43069:10	C	G	0.000	0.126	0.000	inf	0.000	nan	
43069:34	A	T	0.000	0.427	0.000	inf	0.000	nan	
43143:1	A	T	0.021	1.000	0.000	inf	0.000	nan	
43143:5	T	C	0.008	1.000	7.800	0.737	1.839	33.090	
43343:4	G	A	0.023	1.000	5.571	0.744	1.297	23.930	
45009:87	T	G	0.003	0.984	0.095	0.782	0.021	0.440	
51349:23	A	C	0.045	1.000	9.154	1.131	0.998	83.970	
56372:77	A	C	0.027	1.000	0.150	0.832	0.029	0.766	
56752:20	G	A	0.003	0.987	15.400	0.929	2.495	95.050	
56752:7	T	C	0.016	1.000	15.000	1.147	1.583	142.200	
57440:74	C	T	0.024	1.000	0.000	inf	0.000	nan	
57896:39	A	G	0.005	0.998	0.000	inf	0.000	nan	
57969:64	A	G	0.046	1.000	NA	NA	NA	NA	
59359:24	G	A	0.007	1.000	0.074	0.986	0.011	0.512	
59359:5	A	G	0.011	1.000	NA	NA	NA	NA	
59756:39	A	G	0.010	1.000	NA	NA	NA	NA	
66474:79	T	G	0.018	1.000	0.000	inf	0.000	nan	
70991:59	A	T	0.018	1.000	0.000	inf	0.000	nan	
73849:2	G	T	0.014	1.000	0.130	0.783	0.028	0.602	
74826:17	T	C	0.046	1.000	NA	NA	NA	NA	

Table S1 – Associated SNPs obtained for each pairwise comparison through Fisher's exact test, no significant after a 5% FDR correction (*cont.*).

Estimated odds ratio for minor allele (OR); Standard error for odds ration (SE); Lower bound of 99% confidence interval for odds ratio (L99); Upper bound of 99% confidence interval for odds ratio (U99).

SNP_ID	Minor Allele	Major Allele	P-value	Q-value	OR	SE	L99	U99	TRI-TYP
75897:7	C	T	0.001	0.595	NA	NA	NA	NA	
77054:74	A	T	0.043	1.000	0.105	1.140	0.011	0.979	
81058:20	A	G	0.031	1.000	0.000	inf	0.000	nan	
81058:3	A	T	0.031	1.000	0.000	inf	0.000	nan	
81653:10	A	G	0.047	1.000	NA	NA	NA	NA	
83243:8	G	A	0.019	1.000	0.083	1.137	0.009	0.773	
83606:76	C	G	0.038	1.000	9.500	1.141	1.014	88.970	
84043:12	C	T	0.021	1.000	9.167	0.880	1.634	51.430	
87087:62	A	C	0.022	1.000	0.000	inf	0.000	nan	
87932:59	A	G	0.020	1.000	6.000	0.764	1.343	26.810	
87932:85	C	T	0.000	0.414	0.000	inf	0.000	nan	
91056:66	A	G	0.033	1.000	0.000	inf	0.000	nan	
96962:7	A	C	0.037	1.000	0.000	inf	0.000	nan	
99654:23	T	A	0.050	1.000	0.165	0.868	0.030	0.903	
102198:31	T	A	0.045	1.000	NA	NA	NA	NA	
102198:61	A	G	0.012	1.000	0.100	0.887	0.018	0.569	
102870:54	A	C	0.018	1.000	0.093	0.946	0.015	0.591	
103246:89	T	C	0.045	1.000	0.109	1.131	0.012	1.002	
103746:74	T	A	0.004	0.993	0.071	0.945	0.011	0.455	
105240:56	A	T	0.031	1.000	0.000	inf	0.000	nan	
105681:85	T	A	0.045	1.000	0.000	inf	0.000	nan	
106126:64	C	T	0.016	1.000	0.000	inf	0.000	nan	
108159:78	T	A	0.026	1.000	NA	NA	NA	NA	
111157:42	T	G	0.017	1.000	NA	NA	NA	NA	
111157:7	A	G	0.041	1.000	NA	NA	NA	NA	
111750:29	T	A	0.050	1.000	9.000	1.137	0.969	83.580	
116514:82	T	C	0.008	1.000	0.000	inf	0.000	nan	
117209:38	A	C	0.031	1.000	0.000	inf	0.000	nan	
118835:54	C	A	0.004	0.994	NA	NA	NA	NA	
118835:73	C	T	0.022	1.000	10.860	1.118	1.215	97.060	
123202:88	T	A	0.036	1.000	0.133	0.917	0.022	0.805	
125972:31	T	C	0.043	1.000	9.545	1.140	1.021	89.220	
126393:2	A	C	0.024	1.000	11.000	1.146	1.164	103.900	
126742:83	A	G	0.018	1.000	0.000	inf	0.000	nan	
126991:11	G	A	0.017	1.000	0.147	0.803	0.030	0.708	
126991:84	T	A	0.039	1.000	0.185	0.797	0.039	0.880	
130757:44	T	C	0.019	1.000	0.083	1.137	0.009	0.773	
130757:61	C	A	0.031	1.000	NA	NA	NA	NA	
132574:59	G	A	0.008	1.000	0.094	0.924	0.015	0.574	

Table S2 – Genic and genotypic differentiation among dorsal colour phenotypes using single-association SNPs with Bayes factor above 0.99 quantile (BF0.99 SNPs), multi-association SNPs with a posterior inclusion probability above 0.99 quantile (PIP0.99 SNPs) and all 1,837 SNPs.

		Genic differentiation			Genotypic differentiation		
	Pairwise analysis	**χ^2**	**df**	**P-Value**	**χ^2**	**df**	**P-Value**
50 $BF_{0.99}$ SNPs	MAR vs TRI	336.88622	96	Highly sign.	237.695648	96	Highly sign.
	MAR vs TYP	245.03116	92	Highly sign.	179.46547	92	0.00000
	TRI vs TYP	290.11909	94	Highly sign.	210.880335	94	Highly sign.
	Pairwise analysis	**χ^2**	**df**	**P-Value**	**χ^2**	**df**	**P-Value**
50 $PIP_{0.99}$ SNPs	MAR vs TRI	301.39129	96	Highly sign.	217.427269	96	Highly sign.
	MAR vs TYP	248.37097	94	Highly sign.	183.95215	94	0.00000
	TRI vs TYP	283.1366	94	Highly sign.	208.907152	94	Highly sign.
	Pairwise analysis	**χ^2**	**df**	**P-Value**	**χ^2**	**df**	**P-Value**
All 1,837 SNPs	MAR vs TRI	3277.69821	32	1.000	2408.80810	32	1.000
	MAR vs TYP	3168.21556	96	1.000	2347.58297	96	1.000
	TRI vs TYP	3240.57252	84	1.000	2321.40547	84	1.000

Table S3 – Pairwise Fst estimates among dorsal colour phenotypes using single-association SNPs with Bayes factor above 0.99 quantile (BF0.99 SNPs), multi-association SNPs with a posterior inclusion probability above 0.99 quantile (PIP0.99 SNPs) and all 1,837 SNPs.

		TYP	**MAR**
50 $BF_{0.99}$ SNPs	**MAR**	0.1787	--
	TRI	0.2125	0.2145
		TYP	**MAR**
50 $PIP_{0.99}$ SNPs	**MAR**	0.1650	--
	TRI	0.2159	0.1907
		TYP	**MAR**
All 1,837 SNPs	**MAR**	-0.0017	--
	TRI	0.0042	0.0051

Table S4 – SNP correlation value (r^2) in linkage disequilibrium analyses. A FDR of 5% was applied.

	SNP_a	SNP_b	r^2
All phenotypes	43069:10	43069:34	0.922
	3950:1	3950:10	0.851
	63439:8	63439:28	0.824
	55187:46	55187:79	0.793
	64204:16	64204:46	0.791
	27816:86	43143:5	0.490
	40633:18	104623:88	0.448
	75897:7	106126:52	0.435
	41239:75	75897:50	0.430
	75897:7	75897:50	0.391
	41742:86	104623:88	0.387
	16628:65	54226:66	0.384
	11381:9	41742:86	0.373
	59359:24	123202:88	0.371
	3950:1	82682:38	0.342
	41742:86	55187:46	0.335
	118835:54	126355:29	0.319
	20734:39	59359:24	0.315
	6535:26	20734:39	0.300
	3950:1	55187:79	0.300
	24668:63	55187:46	0.295
	23155:83	64258:61	0.293
	3950:10	82682:38	0.291
	92187:65	108304:78	0.285
	20734:39	27816:86	0.283
	22795:88	43069:34	0.282
	11381:9	55187:46	0.281
	43143:5	87932:85	0.276
	43069:10	103746:74	0.272
	11381:9	55187:79	0.267
	41742:86	55187:79	0.266
	23155:83	41742:86	0.264
	41239:75	50515:83	0.263
	50515:83	118835:54	0.262
	24668:63	103746:74	0.261
	3950:10	55187:79	0.255
	25027:11	103746:74	0.251
	35205:6	64204:16	0.244
	102702:13	123202:88	0.241
	22795:88	43069:10	0.239
	6535:26	6535:35	0.238
	43143:5	118835:54	0.201
	43069:34	63439:28	0.187
MAR phenotype	55187:46	55187:79	1.000
	37095:26	40633:18	0.999
	7095:50	50515:83	0.909
	54049:70	54226:66	0.814
	3950:1	3950:10	0.784
TYP phenotype	51349:15	54049:70	0.999
	45009:87	50515:83	0.846
TRI phenotype	No significant LD comparisons		

Table S5 – Genome size estimated by flow cytometry for each individual tested of *P. spumarius* and *P. maghresignus*.

Florescence (FL); Coefficient of variation (CV)

Species	Sex	Individual	FL Sample	FL Standard	DI	2C genome size (pg)	CV Sample (%)	CV Standard (%)	Standard
P. spumarius	M	2	277.75	105.31	2.64	5.17	2.12	2.48	S.l.
P. spumarius	M	3	264.53	104.04	2.54	4.98	2.73	2.78	S.l.
P. spumarius	M	4	246.14	102.23	2.41	4.72	2.77	2.74	S.l.
P. spumarius	M	5	281.04	106.58	2.64	5.17	2.98	3.72	S.l.
P. spumarius	M	6	272.27	100.93	2.70	5.29	3.01	3.31	S.l.
P. spumarius	M	7	276.82	106.86	2.59	5.08	2.96	3.07	S.l.
P. spumarius	F	1	279.18	103.47	2.70	5.29	3.09	3.26	S.l.
P. spumarius	F	2	279.72	102.46	2.73	5.35	3.14	3.13	S.l.
P. spumarius	F	3	276.62	103.66	2.67	5.23	2.77	3.11	S.l.
P. spumarius	F	4	292.74	102.11	2.87	5.62	2.70	3.00	S.l.
P. spumarius	F	5	290.21	102.96	2.82	5.52	2.85	2.75	S.l.
P. spumarius	F	8	287.97	103.49	2.78	5.45	3.39	4.38	S.l.
P. spumarius	F	9	301.03	105.28	2.86	5.60	2.44	2.92	S.l.
P. maghresignus	F	1	451.81	103.18	4.38	8.58	2.94	3.26	S.l.
P. maghresignus	F	2	482.33	104.36	4.62	9.06	3.25	3.9	S.l.
P. maghresignus	F	3	480.16	107.15	4.48	8.78	3.64	4.38	S.l.
P. maghresignus	F	4	490.48	104.07	4.71	9.24	3.15	3.65	S.l.
P. maghresignus	F	5	489.65	105.28	4.65	9.12	2.9	2.92	S.l.
P. maghresignus	F	6	471.38	104.76	4.50	8.82	2.73	3.96	S.l.
P. maghresignus	F	7	471.1	104.38	4.51	8.85	3.39	2.89	S.l.
P. maghresignus	F	8	480.27	105.47	4.55	8.93	3.69	3.58	S.l.
P. maghresignus	F	9	487.09	109.04	4.47	8.76	3.48	3.63	S.l.

Table S6 – Results of the filtering process and assembly statistics for genome and transcriptome.

	Genome	Transcriptome
Total sequences analyzed	366,658,698	17,035,046
Total sequences trashed (%)	12,966,076 (3.54%)	2,247,494 (13.19%)
Total sequences retained (%)	353,692,622 (96.46%)	14,787,552 (86.81%)
Number of assembled contigs	6,843,324	173,691
Maximum contig length	2,6118 bp	9,381 bp
Minimum contig length	100 bp	100 bp
Average contig length	226 bp	235 bp
Number of scaffolds	4,010,521	31,050
Maximum scaffold length with gaps	34,575 bp	17,088 bp
Average scaffold length with gaps	382 bp	1,318 bp
N50	686 bp	803 bp
% of gaps	20.47	0.39
Total assembled length with gaps	1,532,459,407 bp	81,759,018 bp
Total assembled length without gaps	1,218,749,078 bp	81,442,967 bp

Table S7 – Blast results of RAD loci against Arthropoda protein sequences of NCBI.

Full table available at https://dl.dropboxusercontent.com/u/5639287/Additional_file_2_edited_link.xlsx

query id	subject id	Description	% identity	alignment length	mismatches	gap opens	q. start	q. end	s. start	s. end	E-value	bit score
locus19718	gi|646705714|gb|KDR13273.1|	Protein BTG3, partial [Zootermopsis nevadensis]	76.67	30	7	0	3	92	9	38	2.00E-008	52.4
	gi|501293050|dbj|BAN20554.1|	b-cell translocation protein [Riptortus pedestris]	80	30	6	0	3	92	9	38	3.00E-007	48.5
	gi|939680495|ref|XP_014285444.1|	PREDICTED: protein BTG3-like [Halyomorpha halys]	76.67	30	7	0	3	92	9	38	4.00E-007	48.9
	gi|939278830|ref|XP_014259974.1|	PREDICTED: protein BTG3-like [Cimex lectularius]	73.33	30	8	0	3	92	9	38	3.00E-006	47
locus34634	gi|861615211|gb|KMQ86051.1|	retrovirus-related pol polyprotein from transposon tnt 1-94 [Lasius niger]	65.52	29	10	0	2	88	131	159	4.00E-005	44.3
locus58509	gi|827552693|ref|XP_012548103.1|	PREDICTED: uncharacterized protein LOC105842033 [Bombyx mori]	66.67	30	10	0	1	90	213	242	1.00E-008	54.7
	gi|641650328|ref|XP_008189178.1|	PREDICTED: uncharacterized protein LOC103311361 [Acyrthosiphon Pisum]	66.67	30	10	0	1	90	219	248	1.00E-007	52
	gi|913295464|ref|XP_013194825.1|	PREDICTED: uncharacterized protein LOC106138234 [Amyelois transitella]	63.33	30	11	0	1	90	229	258	8.00E-007	49.7
	gi|930673761|gb|KPJ14194.1|	Pogo transposable element with KRAB domain [Papilio machaon]	63.33	30	11	0	1	90	104	133	1.00E-006	49.3
	gi|641652957|ref|XP_008178605.1|	PREDICTED: uncharacterized protein LOC103307917 [Acyrthosiphon Pisum]	65.38	26	9	0	1	78	185	210	2.00E-006	48.1
locus59359	gi|913295464|ref|XP_013194825.1|	PREDICTED: uncharacterized protein LOC106138234 [Amyelois transitella]	62.07	29	11	0	95	9	270	298	6.00E-005	44.3
locus66967	gi|795015422|ref|XP_011858146.1|	PREDICTED: tigger transposable element-derived protein 6-like [Vollenhovia emeryi]	67.86	28	9	0	3	86	406	433	5.00E-006	47
	gi|795018711|ref|XP_011859321.1|	PREDICTED: uncharacterized protein LOC105556820 [Vollenhovia Emeryi]	67.86	28	9	0	3	86	377	404	1.00E-005	46.2
	gi|861625729|gb|KMQ88629.1|	tigger transposable element-derived protein 6-like protein [Lasius niger]	67.86	28	9	0	3	86	377	404	1.00E-005	45.8
locus67578	gi|478256910|gb|ENN77079.1|	hypothetical protein YQE_06414, partial [Dendroctonus ponderosae]	75.86	29	7	0	89	3	37	65	8.00E-008	49.7
	gi|861625729|gb|KMQ88629.1|	tigger transposable element-derived protein 6-like protein [Lasius niger]	68.97	29	9	0	89	3	308	336	2.00E-006	48.1
	gi|321454537|gb|EFX65704.1|	PREDICTED: uncharacterized protein LOC105202865 [Solenopsis invicta]	62.07	29	11	0	89	3	222	250	9.00E-006	45.4
	gi|751232467|ref|XP_011169850.1|	PREDICTED: uncharacterized protein LOC105202865 [Solenopsis invicta]	68.97	29	9	0	89	3	320	348	2.00E-005	45.8
	gi|780616041|ref|XP_011696449.1|	PREDICTED: uncharacterized protein LOC105455090 [Wasmannia auropunctata]	65.52	29	10	0	89	3	315	343	3.00E-005	44.7
locus68031	gi|805822205|ref|XP_012151115.1|	PREDICTED: fibrillin-2-like isoform X3 [Megachile rotundata]	90.91	22	2	0	90	25	715	736	4.00E-006	48.1
	gi|805822203|ref|XP_012151113.1|	PREDICTED: fibrillin-2-like isoform X2 [Megachile rotundata]	90.91	22	2	0	90	25	1170	1191	4.00E-006	48.1
	gi|815916428|ref|XP_003491105.2|	PREDICTED: fibrillin-2-like [Bombus impatiens]	90.91	22	2	0	90	25	1170	1191	4.00E-006	48.1
locus69741	gi|939660472|ref|XP_014278153.1|	PREDICTED: tetratricopeptide repeat protein 28 [Halyomorpha halys]	81.48	27	5	0	1	81	68	94	4.00E-007	50.8
	gi|939241274|ref|XP_014239890.1|	PREDICTED: tetratricopeptide repeat protein 28 [Cimex lectularius]	81.48	27	5	0	1	81	68	94	4.00E-007	50.8
	gi|242011457|ref|XP_002426466.1|	rapsynoid, putative [Pediculus humanus corporis]	74.07	27	7	0	1	81	63	89	3.00E-005	45.4
	gi|195588763|ref|XP_002084127.1|	GD12984 [Drosophila simulans]	70.37	27	8	0	1	81	66	92	8.00E-005	43.9
locus80449	gi|827542484|ref|XP_012544315.1|	PREDICTED: uncharacterized protein LOC105841373 [Bombyx mori]	75.86	29	7	0	87	1	429	457	1.00E-007	52.4
	gi|913328928|ref|XP_013196001.1|	PREDICTED: uncharacterized protein LOC106139163 isoform X1 [Amyelois transitella]	62.07	29	11	0	87	1	499	527	3.00E-006	47.8
locus81867	gi|641657266|ref|XP_003242632.2|	PREDICTED: uncharacterized protein LOC100575723 [Acyrthosiphon pisum]	79.17	24	5	0	15	86	236	259	5.00E-005	43.9
	gi|170068745|ref|XP_001868982.1|	conserved hypothetical protein [Culex quinquefasciatus]	59.26	27	11	0	6	86	215	241	9.00E-005	43.5

Table S8– Summary of the alignment results of RAD loci against drafts of *P. spumarius* genome and transcriptome.

	Genome		**Transcriptome**	
	hits	**no hits**	**hits**	**no hits**
Number of RAD loci	392	536	134	795
% of RAD loci	*42.24*	*57.76*	*14.44*	*85.67*
Number of associated RAD loci	18	34	5	47
% of associated RAD loci	*34.62*	*65.38*	*9.62*	*90.38*

Table S9 – Blast results of *P. spumarius* transcriptome contigs/scaffolds with RAD loci alignments against Arthropoda nucleotide sequences of NCBI. Full table available at https://dl.dropboxusercontent.com/u/5639287/Additional_file_2_edited_link.xlsx

query id	subject id	% identity	alignment length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
C56875	gi|780644672|ref|XM_011690559.1|	100	35	0	0	9	43	1,071	1,037	1.00E-007	66
scaffold18605	gi|242012820|ref|XM_002427080.1|	74.04	1,256	254	65	4,902	6,118	2,246	1,024	5.00E-122	448
	gi|805769431|ref|XM_012280733.1|	78.91	147	25	6	2,558	2,701	4,976	4,833	8.00E-016	95
	gi|805769428|ref|XM_003701235.2|	78.91	147	25	6	2,558	2,701	5,009	4,866	8.00E-016	95
C42981	gi|571568204|ref|XM_006562746.1|	74.49	294	52	16	232	509	2,166	2,452	4.00E-020	106
	gi|571568202|ref|XM_003250115.2|	74.49	294	52	16	232	509	2,163	2,449	4.00E-020	106
	gi|571568197|ref|XM_006562745.1|	74.49	294	52	16	232	509	2,163	2,449	4.00E-020	106
scaffold1670	gi|755990511|ref|XM_011314251.1|	72.57	689	141	48	368	1,032	951	1,615	2.00E-042	182
scaffold1669	gi|755990511|ref|XM_011314251.1|	72.57	689	141	48	368	1,032	951	1,615	2.00E-042	182
C68167	gi|805766218|ref|XM_003700821.2|	74.57	2,273	425	139	80	2,280	160	2,351	0	854
	gi|751236556|ref|XM_011173816.1|	73.69	2,258	456	127	85	2,278	103	2,286	0	752
	gi|936572716|ref|XM_014363177.1|	75.94	1,168	219	58	94	1,230	136	1,272	6.00E-151	544
scaffold29880	gi|964098332|ref|XM_001964584.2|	83.19	119	18	2	634	751	702	585	2.00E-020	108
scaffold21260	gi|630575147|gb|KF383591.1|	78.86	946	168	31	5,553	6,482	939	10	1.00E-170	610
	gi|630574795|gb|KF383583.1|	78.82	949	163	34	5,553	6,482	939	10	5.00E-169	604
scaffold21259	gi|630575147|gb|KF383591.1|	78.86	946	168	31	5,553	6,482	939	10	6.00E-171	610
	gi|630574795|gb|KF383583.1|	78.82	949	163	34	5,553	6,482	939	10	3.00E-169	604
scaffold21258	gi|630575147|gb|KF383591.1|	78.86	946	168	31	5,553	6,482	939	10	6.00E-171	610
	gi|630574795|gb|KF383583.1|	78.82	949	163	34	5,553	6,482	939	10	3.00E-169	604
C64171	gi|780644672|ref|XM_011690559.1|	100	34	0	0	10	43	1,071	1,038	6.00E-007	64

Table S10 – Blast results of *P. spumarius* transcriptome contigs/scaffolds with RAD loci alignments against Arthropoda protein sequences of NCBI. Full table available at https://dl.dropboxusercontent.com/u/5639287/Additional_file_2_edited_link.xlsx

query id	subject id	% identity	alignment length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
scaffold17698	gi|795040272|ref|XP_011866546.1|	40	251	113	4	504	1,142	1,479	1,729	1.00E-058	183
	gi|795040272|ref|XP_011866546.1|	28	159	96	5	11	463	1,313	1,461	1.00E-058	69
	gi|795040277|ref|XP_011866548.1|	40	251	113	4	504	1,142	1,462	1,712	1.00E-058	183
	gi|795040277|ref|XP_011866548.1|	28	159	96	5	11	463	1,296	1,444	1.00E-058	69
	gi|795040284|ref|XP_011866550.1|	40	251	113	4	504	1,142	1,479	1,729	1.00E-058	183
	gi|795040284|ref|XP_011866550.1|	28	159	96	5	11	463	1,313	1,461	1.00E-058	69
	gi|795040287|ref|XP_011866551.1|	40	251	113	4	504	1,142	1,479	1,729	1.00E-058	183
	gi|795040287|ref|XP_011866551.1|	28	159	96	5	11	463	1,313	1,461	1.00E-058	69
	gi|915667197|gb|KOC69033.1|	39	251	115	4	504	1,142	1,348	1,598	1.00E-054	174
scaffold17697	gi|939653348|ref|XP_014275474.1|	40	382	183	9	2	1,138	1,148	1,484	2.00E-069	265
	gi|939653344|ref|XP_014275473.1|	40	382	183	9	2	1,138	1,148	1,484	2.00E-069	265
	gi|939263801|ref|XP_014251876.1|	37	383	190	7	11	1,138	1,148	1,487	3.00E-066	255
	gi|939263799|ref|XP_014251875.1|	37	383	190	7	11	1,138	1,148	1,487	3.00E-066	255
	gi|939263797|ref|XP_014251874.1|	37	383	190	7	11	1,138	1,148	1,487	3.00E-066	255
scaffold29739	gi|746865007|ref|XP_011063523.1|	26	265	176	6	756	7	241	499	2.00E-007	61
	gi|746865005|ref|XP_011063522.1|	26	265	176	6	756	7	241	499	2.00E-007	61
	gi|332019345|gb|EGI59851.1|	26	265	176	6	756	7	277	535	2.00E-007	61
	gi|808130549|ref|XP_012168000.1|	23	251	179	6	744	7	344	585	4.00E-006	56
	gi|662198534|ref|XP_008472393.1|	23	435	296	12	1,221	34	63	496	6.00E-006	56
scaffold13661	gi|646715032|gb|KDR18788.1|	48	367	176	5	1,082	3	11	369	2.00E-088	328
	gi|907617340|ref|XP_013119188.1|	40	359	196	8	1,073	6	70	413	3.00E-072	275
	gi|749789658|ref|XP_011148137.1|	39	362	203	7	1,070	6	25	374	6.00E-072	274
	gi|906467269|gb|KNC28514.1|	40	368	204	8	1,100	6	535	887	5.00E-071	271
	gi|906467269|gb|KNC28514.1|	37	358	209	7	1,070	6	89	433	1.00E-063	246
	gi|195440672|ref|XP_002068164.1|	40	379	211	9	1,130	3	39	402	3.00E-070	268
scaffold13660	gi|646715032|gb|KDR18788.1|	48	367	176	5	1,082	3	11	369	2.00E-088	328
	gi|907617340|ref|XP_013119188.1|	40	359	196	8	1,073	6	70	413	3.00E-072	275
	gi|749789658|ref|XP_011148137.1|	39	362	203	7	1,070	6	25	374	6.00E-072	274
	gi|906467269|gb|KNC28514.1|	40	368	204	8	1,100	6	535	887	5.00E-071	271
	gi|906467269|gb|KNC28514.1|	37	358	209	7	1,070	6	89	433	1.00E-063	246
	gi|195440672|ref|XP_002068164.1|	40	379	211	9	1,130	3	39	402	3.00E-070	268

Table S11 – Blast results of *P. spumarius* genome contigs/scaffolds with RAD loci alignments against Arthropoda nucleotide sequences of NCBI.

query id	subject id	% identity	alignment length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
	gi|795024761|ref|XM_012005991.1|	87	1,932	218	28	3,821	5,741	3	1,914	0	2,159
	gi|795024758|ref|XM_012005990.1|	88	1,073	124	10	11,088	12,155	1	1,068	0	1,230
	gi|936711822|ref|XM_014378388.1|	82	151	23	4	1,779	1,927	7,708	7,856	9E-026	126
C111684269	gi|936711819|ref|XM_014378387.1|	82	151	23	4	1,779	1,927	7,815	7,963	9E-026	126
	gi|936711817|ref|XM_014378385.1|	82	151	23	4	1,779	1,927	7,881	8,029	9E-026	126
	gi|936711815|ref|XM_014378384.1|	82	151	23	4	1,779	1,927	7,935	8,083	9E-026	126
	gi|936711812|ref|XM_014378383.1|	82	151	23	4	1,779	1,927	7,947	8,095	9E-026	126

Table S12 – Blast results of *P. spumarius* genome contigs/scaffolds with RAD loci alignments against Arthropoda protein sequences of NCBI.

Full table available at https://dl.dropboxusercontent.com/u/5639287/Additional_file_2_edited_link.xlsx

query id	subject id	% identity	alignment length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
scaffold1372429	gi|939653348|ref|XP_014275474.1|	43	76	36	3	1,576	1,797	1,188	1,258	6.00E-009	67.4
	gi|939653348|ref|XP_014275474.1|	50	52	26	0	4,702	4,857	1,286	1,337	7.00E-007	60.5
	gi|939653344|ref|XP_014275473.1|	43	76	36	3	1,576	1,797	1,188	1,258	6.00E-009	67.4
	gi|939653344|ref|XP_014275473.1|	50	52	26	0	4,702	4,857	1,286	1,337	7.00E-007	60.5
	gi|939263801|ref|XP_014251876.1|	40	70	41	1	1,567	1,773	1,182	1,251	5.00E-008	64.3
	gi|939263801|ref|XP_014251876.1|	44	52	29	0	4,690	4,845	1,279	1,330	1.00E-006	59.7
	gi|939263799|ref|XP_014251875.1|	40	70	41	1	1,567	1,773	1,182	1,251	5.00E-008	64.3
	gi|939263799|ref|XP_014251875.1|	44	52	29	0	4,690	4,845	1,279	1,330	1.00E-006	59.7
	gi|939263797|ref|XP_014251874.1|	40	70	41	1	1,567	1,773	1,182	1,251	5.00E-008	64.3
C106506244	gi|646715032|gb|KDR18788.1|	63	59	22	0	179	3	295	353	3.00E-016	86.3
	gi|817068467|ref|XP_012256173.1|	49	61	31	0	185	3	107	167	7.00E-013	75.1
	gi|751781541|ref|XP_011199940.1|	54	63	27	1	185	3	372	434	9.00E-013	74.7
	gi|195496390|ref|XP_002095674.1|	55	62	26	1	182	3	323	384	2.00E-012	73.6
	gi|195128499|ref|XP_002008700.1|	52	62	28	1	182	3	313	374	3.00E-012	73.2
scaffold1264404	gi|827539751|ref|XP_012553005.1|	45	158	85	1	1,288	821	61	218	2.00E-031	140
	gi|827539751|ref|XP_012553005.1|	49	102	45	3	349	50	307	403	4.00E-019	99.4
	gi|827539751|ref|XP_012553005.1|	54	50	23	0	663	514	235	284	4.00E-009	66.2
	gi|641653467|ref|XP_003241645.2|	46	157	84	1	1,288	821	61	217	4.00E-031	139
	gi|641672247|ref|XP_008185868.1|	47	157	82	1	1,288	821	25	181	5.00E-031	139
	gi|641672247|ref|XP_008185868.1|	56	50	22	0	663	514	198	247	2.00E-009	67
	gi|913332317|ref|XP_013197778.1|	44	156	85	1	1,282	821	61	216	1.00E-030	137
	gi|913332317|ref|XP_013197778.1|	54	50	23	0	663	514	233	282	3.00E-009	66.6
	gi|913332317|ref|XP_013197778.1|	57	54	20	2	349	194	305	357	4.00E-007	59.7
	gi|913326429|ref|XP_013194622.1|	44	156	85	1	1,282	821	61	216	1.00E-030	137

Table S13 – Genome and transcriptome sequences which aligned with associated RAD loci and had protein hits.

Final column indicates if scaffold/contig region that had a protein hit comprise the region of the RAD loci.

TRANSCRIPTOME

RAD loci id	Position in scaffold	Scaffold id	Protein id	Protein Description	Scaffold alignment region	Protein alignment region	RAD loci region
16628	From 123 to 36	scaffold17698	gi|795040272|ref|XP_011866546.1|	PREDICTED: lysosomal-trafficking regulator isoform X1 [Vollenhovia emeryi]	504 to 1142	1479 to 1729	no
			gi|795040272|ref|XP_011866546.1|	PREDICTED: lysosomal-trafficking regulator isoform X1 [Vollenhovia emeryi]	11 to 463	1313 to 1461	yes
			gi|795040277|ref|XP_011866548.1|	PREDICTED: lysosomal-trafficking regulator isoform X3 [Vollenhovia emeryi]	504 to 1142	1462 to 1712	no
			gi|795040277|ref|XP_011866548.1|	PREDICTED: lysosomal-trafficking regulator isoform X3 [Vollenhovia emeryi]	11 to 463	1296 to 1444	yes
			gi|795040284|ref|XP_011866550.1|	PREDICTED: lysosomal-trafficking regulator isoform X5 [Vollenhovia emeryi]	504 to 1142	1479 to 1729	no
			gi|795040284|ref|XP_011866550.1|	PREDICTED: lysosomal-trafficking regulator isoform X5 [Vollenhovia emeryi]	11 to 463	1313 to 1461	yes
			gi|795040287|ref|XP_011866551.1|	PREDICTED: lysosomal-trafficking regulator isoform X6 [Vollenhovia emeryi]	504 to 1142	1479 to 1729	no
			gi|795040287|ref|XP_011866551.1|	PREDICTED: lysosomal-trafficking regulator isoform X6 [Vollenhovia emeryi]	11 to 463	1313 to 1461	yes
			gi|915667197|gb|KOC69033.1|	Lysosomal-trafficking regulator [Habropoda laboriosa]	504 to 1142	1348 to 1598	no
		scaffold17697	gi|939653348|ref|XP_014275474.1|	PREDICTED: lysosomal-trafficking regulator isoform X2 [Halyomorpha halys]	2 to1138	1148 to 1484	yes
			gi|939653344|ref|XP_014275473.1|	PREDICTED: lysosomal-trafficking regulator isoform X1 [Halyomorpha halys]	2 to1138	1148 to 1484	yes
			gi|939263801|ref|XP_014251876.1|	PREDICTED: lysosomal-trafficking regulator isoform X3 [Cimex lectularius]	11 to 1138	1148 to 1487	yes
			gi|939263799|ref|XP_014251875.1|	PREDICTED: lysosomal-trafficking regulator isoform X2 [Cimex lectularius]	11 to 1138	1148 to 1487	yes
			gi|939263797|ref|XP_014251874.1|	PREDICTED: lysosomal-trafficking regulator isoform X1 [Cimex lectularius]	11 to 1138	1148 to 1487	yes
22795	From 593 to 507	scaffold29739	gi|746865007|ref|XP_011063523.1|	PREDICTED: nucleolar and coiled-body phosphoprotein 1 isoform X2 [Acromyrmex echinatior]	756 to 7	241 to 499	yes
			gi|746865005|ref|XP_011063522.1|	PREDICTED: nucleolin 1 isoform X1 [Acromyrmex echinatior]	756 to 7	241 to 499	yes
			gi|332019345|gb|EGI59851.1|	Protein unc-79-like protein [Acromyrmex echinatior]	756 to 7	277 to 535	yes
			gi|808130549|ref|XP_012168000.1|	PREDICTED: LOW QUALITY PROTEIN: nucleolar and coiled-body phosphoprotein 1 [Bombus terrestris]	756 to 7	344 to 585	yes
			gi|662198534|ref|XP_008472393.1|	PREDICTED: nucleolar and coiled-body phosphoprotein 1-like [Diaphorina citri]	1221 to 34	63 to 496	yes
56752	From 128 to 218	scaffold13661	gi|646715032|gb|KDR18788.1|	hypothetical protein L798_06469, partial [Zootermopsis nevadensis]	1082 to 3	11 to 369	yes
			gi|907617340|ref|XP_013119188.1|	PREDICTED: uncharacterized protein LOC106096144 [Stomoxys calcitrans]	1073 to 6	70 to 413	yes
			gi|749789658|ref|XP_011148137.1|	hypothetical protein FF38_05739 [Lucilia cuprina]	1070 to 6	25 to 374	yes
			gi|906467269|gb|KNC28514.1|	hypothetical protein FF38_05739 [Lucilia cuprina]	1100 to 6	535 to 887	yes
			gi|906467269|gb|KNC28514.1|	hypothetical protein FF38_05739 [Lucilia cuprina]	1070 to 6	89 to 433	yes
			gi|195440672|ref|XP_002068164.1|	uncharacterized protein Dwil_GK10387 [Drosophila willistoni]	1130 to 3	39 to 402	yes
		scaffold13660	gi|646715032|gb|KDR18788.1|	hypothetical protein L798_06469, partial [Zootermopsis nevadensis]	1082 to 3	11 to 369	yes
			gi|907617340|ref|XP_013119188.1|	PREDICTED: uncharacterized protein LOC106096144 [Stomoxys calcitrans]	1073 to 6	70 to 413	yes
			gi|749789658|ref|XP_011148137.1|	hypothetical protein FF38_05739 [Lucilia cuprina]	1070 to 6	25 to 374	yes
			gi|906467269|gb|KNC28514.1|	hypothetical protein FF38_05739 [Lucilia cuprina]	1100 to 6	535 to 887	yes
			gi|906467269|gb|KNC28514.1|	hypothetical protein FF38_05739 [Lucilia cuprina]	1070 to 6	89 to 433	yes
			gi|195440672|ref|XP_002068164.1|	uncharacterized protein Dwil_GK10387 [Drosophila willistoni]	1130 to 3	39 to 402	yes

Table S13 – Genome and transcriptome sequences which aligned with associated RAD loci and had protein hits (*cont.*).

Final column indicates if scaffold/contig region that had a protein hit comprise the region of the RAD loci.

GENOME

RAD loci id	Position in scaffold	Scaffold id	Protein id	Protein Description	Scaffold alignment region	Protein alignment region	RAD loci region
16628	From 582 to 495	scaffold1372429	gi|939653348|ref|XP_014275474.1|	PREDICTED: lysosomal-trafficking regulator isoform X2 [Halyomorpha halys]	1576 to 1797	1188 to 1258	no
			gi|939653348|ref|XP_014275474.1|	PREDICTED: lysosomal-trafficking regulator isoform X2 [Halyomorpha halys]	4702 to 4857	1286 to 1337	no
			gi|939653344|ref|XP_014275473.1|	PREDICTED: lysosomal-trafficking regulator isoform X1 [Halyomorpha halys]	1576 to 1797	1188 to 1258	no
			gi|939653344|ref|XP_014275473.1|	PREDICTED: lysosomal-trafficking regulator isoform X1 [Halyomorpha halys]	4702 to 4857	1286 to 1337	no
			gi|939263801|ref|XP_014251876.1|	PREDICTED: lysosomal-trafficking regulator isoform X3 [Cimex lectularius]	1567 to 1773	1182 to 1251	no
			gi|939263801|ref|XP_014251876.1|	PREDICTED: lysosomal-trafficking regulator isoform X3 [Cimex lectularius]	4690 to 4845	1279 to 1330	no
			gi|939263799|ref|XP_014251875.1|	PREDICTED: lysosomal-trafficking regulator isoform X2 [Cimex lectularius]	1567 to 1773	1182 to 1251	no
			gi|939263799|ref|XP_014251875.1|	PREDICTED: lysosomal-trafficking regulator isoform X2 [Cimex lectularius]	4690 to 4845	1279 to 1330	no
			gi|939263797|ref|XP_014251874.1|	PREDICTED: lysosomal-trafficking regulator isoform X1 [Cimex lectularius]	1567 to 1773	1182 to 1251	no
56752	From 80 to 170	C106506244	gi|646715032|gb|KDR18788.1|	hypothetical protein L798_06469, partial [Zootermopsis nevadensis]	179 to 3	295 to 353	yes
			gi|817068467|ref|XP_012256173.1|	PREDICTED: uncharacterized protein LOC105686127 [Athalia rosae]	185 to 3	107 to 167	yes
			gi|751781541|ref|XP_011199940.1|	PREDICTED: uncharacterized protein LOC105223789 [Bactrocera dorsalis]	185 to 3	372 to 434	yes
			gi|195496390|ref|XP_002095674.1|	uncharacterized protein Dyak_GE19581 [Drosophila yakuba]	182 to 3	323 to 384	yes
			gi|195128499|ref|XP_002008700.1|	uncharacterized protein Dmoj_GI13641 [Drosophila mojavensis]	182 to 3	313 to 374	yes
59359	From 515 to 586	scaffold1264404	gi|827539751|ref|XP_012553005.1|	PREDICTED: uncharacterized protein LOC105842971 [Bombyx mori]	1288 to 821	61 to 218	no
			gi|827539751|ref|XP_012553005.1|	PREDICTED: uncharacterized protein LOC105842971 [Bombyx mori]	349 to 50	307 to 403	no
			gi|827539751|ref|XP_012553005.1|	PREDICTED: uncharacterized protein LOC105842971 [Bombyx mori]	663 to 514	235 to 284	yes
			gi|641653467|ref|XP_003241645.2|	PREDICTED: uncharacterized protein LOC100571097 [Acyrthosiphon pisum]	1288 to 821	61 to 217	no
			gi|641672247|ref|XP_008185868.1|	PREDICTED: uncharacterized protein LOC103310201 [Acyrthosiphon pisum]	1288 to 821	25 to 181	no
			gi|641672247|ref|XP_008185868.1|	PREDICTED: uncharacterized protein LOC103310201 [Acyrthosiphon pisum]	663 to 514	198 to 247	yes
			gi|913332317|ref|XP_013197778.1|	PREDICTED: uncharacterized protein LOC106140695 [Amyelois transitella]	1282 to 821	61 to 216	no
			gi|913332317|ref|XP_013197778.1|	PREDICTED: uncharacterized protein LOC106140695 [Amyelois transitella]	663 to 514	233 to 282	yes
			gi|913332317|ref|XP_013197778.1|	PREDICTED: uncharacterized protein LOC106140695 [Amyelois transitella]	349 to 194	305 to 357	no
			gi|913326429|ref|XP_013194622.1|	PREDICTED: uncharacterized protein LOC106138118 [Amyelois transitella]	1282 to 821	61 to 216	no

CHAPTER 5

Genetic variation of the *yellow* gene in spittlebugs of genus *Philaenus* (Hemiptera, Aphrophoridae)

Ana S.B. Rodrigues[1*], Sofia G. Seabra[1L], Octávio S. Paulo[1L]

[1] Computational Biology and Population Genomics Group, cE3c – Centre for Ecology, Evolution and Environmental Changes, DBA/FCUL, Campo Grande, P-1749-016 Lisboa, Portugal.

[L] Both authors contributed equally (senior authors)

In prep.

Abstract

The study of the colour polymorphisms in many animal species have greatly contributed to understand the molecular mechanisms underlying phenotypic variation and its role in the process of adaptation. The widespread meadow spittlebug *Philaenus spumarius* shows a dorsal colour/pattern polymorphism, mainly characterised by melanic and non-melanic forms, with a relatively simple mendelian inheritance, but unknown molecular genetic basis. Similar dorsal colour variation is observed in the other species of the genus, suggesting the maintenance of an old polymorphism. In this work, a candidate gene approach was used, for the first time, to investigate the genetic basis of *P. spumarius* dorsal colour polymorphism. A fragment of 482 bp of a candidate gene known to be involved in melanic colouration in *Drosophila*, *yellow* was isolated and analysed. No association between *yellow* variation (at nucleotide and haplotype levels) was found, but a possible effect of this gene in colour variation cannot be totally excluded since we analysed only a fraction of its coding region (29.64%). The phylogenetic relationships at the *yellow* fragment in *P. spumarius* and in other closely related species were examined. *yellow* pattern contrasted with the neutral pattern of the gene Internal Transcribed Spacer 2 (*ITS2*). For *yellow*, the same haplotype, probably the ancestral one, was found in almost all *Philaenus* species, except for *P. maghresignus* and *P. arslani*. The inter-specific differentiation in two main groups, observed for *ITS2*, and already described in previous works, was not found for *yellow*. These results indicate that, contrary to *ITS2, yellow* is conserved within *Philaenus*, probably, a functional gene whose evolution is constrained with new deleterious mutations being eliminated by natural selection. A higher genetic diversity was found for *P. spumarius*. Probably, due to its different ecology this species was less affected by factors such as drift and population declines. Although applying a candidate gene approach in a model species like *P. spumarius* has been challenging, it would be important to investigate other candidates. Ultimately, novel and unexpected genes, including regulatory genes, could revealed to be implicated in this species' colour polymorphism.

Keywords: candidate genes, colour variation, molecular evolution; population genetics

Introduction

Investigating the colour polymorphisms in vertebrate and invertebrate species have greatly contributed to understand the genetic mechanisms underlying phenotypic variation, and its role in the process of adaptation. Many colour patterns are known to have an adaptive function (for a review see Forsman *et al.*, 2008) and several mechanisms have been proposed to explain the observed polymorphisms (for a review see Gray & McKinnon, 2007).

The widespread meadow spittlebug *Philaenus spumarius* (Linnaeus, 1758) (Hemiptera, Aphrophoridae) shows a dorsal colour/pattern variation with thirteen main phenotypes described and other three rare colour morphs reported (Halkka & Halkka, 1990; Stewart & Lees, 1996; Quartau & Borges, 1997; Yurtsever, 2000). These are divided into non-melanic and melanic forms (Halkka & Halkka, 1990; Yurtsever, 2000). Different selective pressures such as habitat composition, climatic conditions (including altitudinal and latitudinal gradients), industrial melanism and predation (reviewed in Halkka & Halkka, 1990; Yurtsever, 2000) may explain the differences observed in the occurrence and frequency of the colour phenotypes among populations. The idea that complex mechanisms are involved in the maintenance of this polymorphism was also supported by Silva *et al.* (2015) that detected a higher longevity and fertility of the *trilineatus* phenotype in laboratory conditions.

A similar dorsal colour polymorphism is observed in the other species of the genus (Drosopoulos, 2003). Currently, ten species are recognised within the *Philaenus*: *P. spumarius* (Linnaeus, 1758) in the Holarctic region, *P. tesselatus* (Melichar, 1889) in southern Iberia and North Africa, *P. signatus* (Melichar, 1896) in eastern Mediterranean and western Asia, *P. loukasi* in Greece (Drosopoulos & Asche, 1991), *P. arslani* in Lebanon (Abdul-Nour & Lahoud, 1996), *P. maghresignus* in southern Iberia and North Africa (Drosopoulos & Remane, 2000), *P. italosignus* in southern Italy and Sicily (Drosopoulos & Remane, 2000), *P. tarifa* in southern Iberia (Remane & Drosopoulos, 2001), and *P. elbursianus* (Tishechkin, 2013) and *P. iranicus* (Tishechkin, 2013) in Iran. Phylogenetic analyses using the mitochondrial genes cytochrome oxidase I (*COI*), cytochrome oxidase II (*COII*) and cytochrome b (*Cyt b*), and the nuclear gene Internal Transcribed Spacer 2 (*ITS2*), and that did not include the Iranian species, support the division of the group into two main groups: a group including *P. spumarius, P. tesselatus, P. arslani* and *P. loukasi* and another group comprising *P. signatus*, *P. maghresignus*, *P. tarifa* and *P. italosignus* (Maryańska-Nadachowska *et al*., 2010; Rodrigues, 2010; Seabra *et al.*, 2010). A closer relationship between *P. spumarius* and *P. tesselatus* and also between *P. arslani* and *P. loukasi* was found but the relationship between the other Mediterranean species

is not clear (Maryańska-Nadachowska *et al*., 2010; Rodrigues, 2010; Seabra *et al.*, 2010). The dorsal colour morphs in *P. spumarius, P. tesselatus, P. signatus, P. italosignus, P. maghresignus* and *P. tarifa* are similar in pattern, although, differing in frequencies. In the remaining species the number of morphs is lower. The Mediterranean species *P. arslani* and *P. loukasi* are monomorphic and trimorphic, respectively (Drosopoulos *et al.,* 2010) and in the Iranian species only two colour morphs are described (Tishechkin, 2013). The presence of similar colour morphs suggests that this colour polymorphism may be old, having appeared in the ancestral form of the genus and maintained through the speciation process, probably due to balancing selection (Halkka & Halkka, 1990) keeping genetic variability within species (Reininga *et al.*, 2009).

Crossing experiments showed that dorsal colour variation in *P. spumarius* has a Mendelian inheritance and is mainly controlled by an autosomal locus (*p*) with seven alleles (Halkka *et al.,* 1973). These alleles have complex dominance and co-dominance relationships and are probably interacting with other loci (Halkka *et al.* 1973; Stewart & Lees, 1988). The molecular basis of this colour polymorphism is unknown. Notwithstanding, Halkka & Lallukka (1969) suggested the colour genes may be epistatically linked to genes responsible for the response to the physical environment, constituting a *supergene*, and selection may not be directly related to colour. In mimetic butterflies (Joron *et al.,* 1999; Nijhout, 2003), land snails (Richards *et al.,* 2013) and birds (Tuttle *et al.,* 2016), for example, there is evidence that balanced polymorphisms can result from tight genetic linkage between multiple functional loci, known as *supergenes* (Thompson & Jiggins, 2014).

In *P. spumarius* and remaining species, dorsal colour patterns, which range from completely dark to pale/brown with lighter spots or stripes, are likely melanin based (Halkka & Halkka, 1990; Yurtsever, 2000). Several genes and pathways involved in insect colouration and pigmentation have been identified, mainly in *Drosophila* and in some Lepidoptera (Wittkopp & Beldade, 2009). In the melanin synthesis pathway of *Drosophila*, the *yellow* gene promotes the formation of the black pigment and mutants for this gene are unable to produce melanin (Wittkopp *et al.*, 2002a). This gene is also involved in larval colour pattern in the silk worm *Bombyx mori* (Futahashi *et al.,* 2008). Therefore, *yellow* can be a potential candidate to control melanisation pattern within *Philaenus* species as well. Here, association between *yellow* variation and three *P. spumarius* dorsal colour morphs (*typicus*, *trilineatus* and *marginellus*) was tested. The evolutionary relationships at the *yellow* gene fragment in *P. spumarius* and in other closely related species was also examined. The pattern of *yellow* was posteriorly compared with the neutral pattern of the nuclear gene Internal Transcribed Spacer 2 (*ITS2*).

This gene is a molecular marker very suited for phylogenetic analysis at the species level (Schultz & Wolf, 2009) and was already used in previous studies on the phylogeny of the *Philaenus* species (Maryańska-Nadachowska *et al.*, 2010; Rodrigues, 2010).

Material & Methods

Sampling and DNA extraction

A total of 61 individuals of *P. spumarius* from Portugal, Finland, Turkey, Greece, and Morocco, 15 individuals of *P. arslani* from Lebanon, 15 specimens of *P. italosignus* from Italy, 15 samples of *P. signatus* from Turkey, 14 individuals of *P. tarifa* from Gibraltar and 13 samples of *P. maghresignus* from Portugal were collected or sent by collaborators between 2007 and 2012 (Table S1). Additional individuals were collected for *ITS2* analysis (Table S1). One individual of *Mesoptyelus impictifrons* from Lebanon was also captured and sent by a collaborator in 2011 (Table S1). Adult insects were captured using a sweep net suitable for low-growing vegetation and an entomological aspirator. In some cases, larval stages were collected by hand. Specimens were preserved in absolute ethanol and stored at 4 ºC. Entire larval stage specimens were used for DNA extraction while in the adults the wings and abdomen were removed and only the thorax and the head were used. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen).

yellow gene isolation, amplification and sequencing

Aiming to look for conserved regions in the *yellow* gene to design primers for its amplification in the *P. spumarius* and other related species, *Drosophila melanogaster* gene sequence of *yellow* (FlyBase Acession number: FBgn0004034) was initially aligned against the partial draft genome and transcriptome of *P. spumarius* (see Chapter 4), using a local BLASTN 2.2.28+ (Altschul *et al.* 1997). However, no homology was found. Alternatively, this sequence was queried against the online database of Expressed Sequenced Tags (*ESTs*) from the pea aphid (*Acyrthosiphon pisum*), using the Aphidbase blast server (http://tools.genouest.org/tools/aphidblast/, last modified 21/10/2011). Although belonging to a different infraorder, this species is the most closely related of *P. spumarius* (mtDNA genetic distance of 0.48 (Song *et al.*, 2012)) with available genomic sequences that could be used to isolate the *yellow* gene and to look for conserved regions. A significant match (E-value of $9e^{-11}$) of the *Drosophila* sequence to a 608 bp pea aphid *EST* (GenBank Acession number:

DY226299.1) was obtained, and the sequence confirmed to correspond to the *A. pisum yellow-y* reference mRNA sequence (NCBI Reference Sequence: NM_001172377.1). This sequence was, posteriorly, queried against the partial draft genome of *P. spumarius* and also against the transcriptome to look for an homologous region of the pea aphid *yellow-y* mRNA sequence. The sequence aligned to the genome scaffold number 776671 only. Primers were designed in a conserved region, targeting a 530 bp fragment of the *yellow* gene in *P. spumarius*.

A fragment of 482 bp of the *yellow* gene was amplified by polymerase chain reaction (PCR) using the designed primers PSP22460074_y_fw (5'– CTGATGAATTAGGCTACGG – 3') and PSP22460074_y_rv_comp (5' – GTATACTCTAAAGTTGACATCCC – 3'). PCR was performed in a 20 µL reaction volume containing: 1 µM of each primer, 0.1 mM dNTPs, 1 mM $MgCl_2$, 4 µL 5x Colorless GoTaq Flexi Buffer, 0.0375U GoTaq DNA Polimerase (Promega) and approximately 30 ng of DNA. The PCR conditions were: an initial denaturation step at 95 ºC for 5 min, followed by 36 cycles of denaturation at 95 ºC for 45 sec, annealing at 59 ºC for 35 sec and extension at 72 ºC for 2 min, with a final extension period at 72 ºC for 10 min. All PCR products were purified with SureClean (Bioline) following the manufacturer's protocol and sequenced at Beckman Coulter Genomics (United Kingdom) and Macrogen (Netherlands). Sequences were verified and edited using the software SEQUENCHER v. 4.0.5 (Gene Codes Corporation) and BIOEDIT v. 7.0.9 (Hall, 1999). They were then aligned using MAFFT v. 7.029b (Katoh & Standley, 2013) and converted in the appropriate format with CONCATENATOR v. 1.1.0 (Pina-Martins & Paulo, 2008). Haplotype phase for individuals that were heterozygous for base positions was determined using PHASE (Stephens *et al.*, 2001; Stephens & Donelly 2003), implemented in DNASP v. 5 (Librado & Rozas, 2009).

In *Drosophila* this gene encodes a simple transcription unit of two exons, encoding a 541 amino acid protein (http://flybase.org). The region amplified in *Philaenus* is homologous to part of exon 2 of the *D. melanogaster yellow* gene, more specifically to the region encompassing a conserved domain commonly found in proteins belonging to the MRJP (Major Royal Jelly Proteins) family and from which *yellow* protein is a member.

Internal Transcribed Spacer 2 (*ITS2*) amplification and sequencing

A fragment of 445 bp of the nuclear region *ITS2* was also amplified using the primers CAS5p8sFt (5'–TGAACATCGACATTTCGAACGCACAT–3') and CAS28sB1d (5'–TTGTTTTCCTCCGCTTATTAATATGCTTAA–3') (Kim & Lee, 2008). PCR concentrations and conditions were the same used for the *yellow* gene except for the annealing temperature (55 ºC). All PCR products were purified with SureClean (Bioline) and sequenced at Beckman

Coulter Genomics (France) and Macrogen (Netherlands).

Genetic analyses

yellow diversity

For the *yellow* gene fragment and each *Philaenus* species, number of homozygotes and heterozygotes, number of polymorphic sites (S), number of haplotypes (H), haplotype diversity (*h*), nucleotide diversity (π) and number of synonymous and non-synonymous substitutions were calculated using the program DNASP v. 5. Amino acid replacements were assessed through the translation of the putative ORFs (Open Reading Frame) in MEGA v. 5.0 (Tamura *et al.*, 2011). Alignments with *yellow* sequences from *M. impictifrons* and other insects available in GenBank (*Drosophila melanogaster* NM_057444; *Acyrthosiphon pisum* NM_001172377; *Apis melifera* NM_001098223.1) and in European Nucleotide Archive (*Lygus hesperus* GBHO01038061.1) were performed in MAFFT. Sequences were translated into amino acids and sequence variation in the *yellow* protein investigated.

Neutrality tests

The neutrality test Tajima's *D* (Tajima, 1989) was performed in DNASP v. 5 to detect signals of natural selection in the *yellow* gene fragment. For analysis of selective pressures acting on site-specific *yellow* sequences, a maximum likelihood approach was implemented using CODEML from PAML v4.6 (Yang, 2007). To test for positive selection on different sites across the protein sequence, two site models were tested: M1, which corresponds to neutrality and assumes two values for ω ($\omega=1$ and $\omega<1$), and M2, that estimates three values of ω ($\omega=1$, $\omega<1$ and $\omega>1$) and accounts for positive selection. ω corresponds to the ratio between the non-synonymous rate (d_N) and the synonymous rate (d_S). The branch-site model was also tested to estimate different dN/dS among sites and among branches. Two branches, where the non-synonymous mutations occurred, were tested. The null model assumes that sites have different values of ω, but always $\omega<1$ (neutral selection) and the alternative model assumes that sites have different values of ω, but always $\omega>1$ (positive selection) (Yang, 2007). Likelihood ratio tests (LRT) were performed to compare the models, and a χ^2 distribution was used to check for significant differences. The analyses were conducted using *M. impictifrons* as outgroup.

yellow structure analyses

To infer the evolutionary dynamics at the *yellow* gene fragment in *P. spumarius* and in

other close related species, a median-joining network was constructed for *yellow* using NETWORK V. 4.5.0.1 (FLUXUS TECHNOLOGY LTD. 2004), and its pattern was compared with the pattern of the nuclear *ITS2*. To this analysis *ITS2* sequences of *Philaenus* species from Maryańska-Nadachowska *et al.* (2010) and available at NCBI Genbank were added to our *ITS2* matrix and included in the phylogenetic analysis (GenBank accession numbers: FJ560708 – FJ560716). For *ITS2* and each *Philaenus* species, number of homozygotes and heterozygotes, number of polymorphic sites (S), number of haplotypes (H), haplotype diversity (*h*) and nucleotide diversity (π) were calculated using the program DNASP v. 5.

Association analyses between yellow variation and dorsal colour patterns

Association between Single Nucleotide Polymorphisms (SNPs), distributed along the *yellow* fragment, and dorsal colour phenotypes was tested using a Fisher's exact test of allelic association in PLINK v 1.07 (Purcell *et al*., 2007) (http://pngu.mgh.harvard.edu/purcell/plink/). Three pairwise analyses were performed: MAR vs. TRI, MAR vs. TYP and TRI vs. TYP. Allele frequencies in each pair, the odds ratio and *p*-values were obtained for each SNP and a false discovery rate (FDR) of 5% was applied (Benjamini & Hochberg, 1995) to each pairwise analysis to test for false positives. A total of 14 MAR, 14 TRI and 13 TYP *P. spumarius* females from Portugal were analysed. To analyse the relationship among *yellow* haplotypes, and a possible association with colour variation, a median-joining haplotypre network was obtained for *P. spumarius* colour phenotypes in NETWORK.

Results

yellow sequence diversity

A total of 12 different haplotypes was obtained. Seven haplotypes were unique to *P. spumarius*, two to *P. maghresignus* and one to *P. arslani* (Fig. 1). Values of nucleotide and haplotype diversity were higher in *P. spumarius* (π = 0.0046 and *h* = 0.803). In *P. tarifa*, *P. signatus* and *P. arslani*, no genetic variation was found (one single haplotype) among all 15 individuals of each species sequenced and in *P. italosignus* (15 individuals) and *P. maghresignus* (13 individuals) only 2 haplotypes were detected. (Fig. 1 and Table 1). A higher number of heterozygotes was observed for *P. spumarius*, compared with the other *Philaenus* species. From the 53 *P. spumarius* sequences, 38 were heterozygous. Individuals of *P. signatus, P. tarifa* and *P. arslani* were all homozygous. Homozygosity was also observed for all *P.*

italosignus and *P. maghresignus* individuals, except for one and two, respectively. In *P. spumarius*, from a total of nine haplotypes, seven were unique to this species (Fig. 1). One most common haplotype (H7) is shared among some individuals of *P. spumarius*, all individuals of *P. signatus* and *P. tarifa*, and all individuals of *P. italosignus*, except for one heterozygous individual, that also have the haplotype H9 (Fig. 1). *Philaenus maghresignus* and *P. arslani* are the only species that do not share this haplotype.

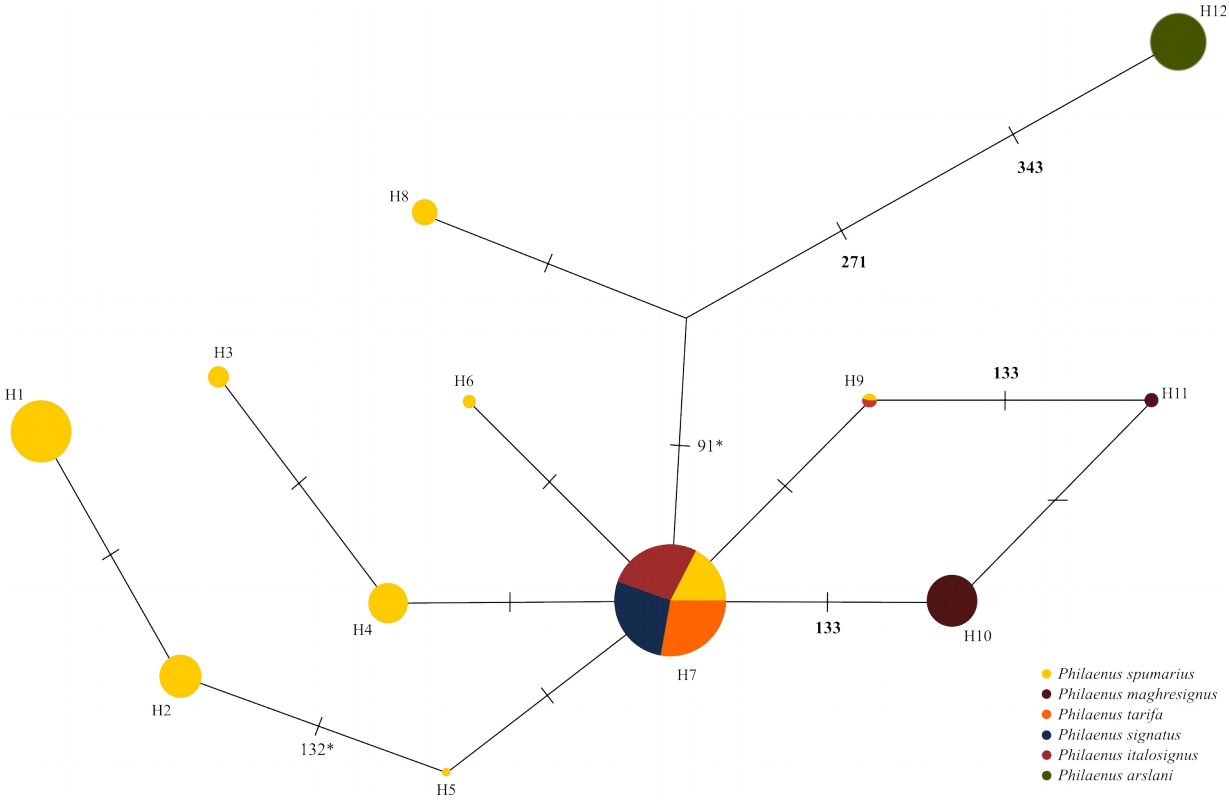


Fig. 1: Median-joining haplotype network for the *yellow* gene (482bp) and *Philaenus* species. A mutational step between haplotypes is represented by a dash. Site number and bold indicates mutations that are fixed for *P. arslani* and for *P. maghresignus* species. Site number and asterisks denote nonsynonymous changes.

A total of 12 segregating sites (S) were observed in the *yellow* fragment, two of them being non-synonymous at positions 91 and 132. At site 91 the mutation corresponds to a G-T transversion in the third position of the codon 30 that led to the replacement of a glutamic acid residue (E) for an aspartic acid (D). A glutamic acid residue (E) was found in *P. spumarius*, *P. tarifa*, *P. maghresignus*, *P. italosignus* and *P. signatus*. For *P. spumarius* an aspartic residue (D) was also found. Curiously, only an aspartic residue (D) was detected in *P. arslani*. In the other analysed species, *M. impictifrons* and *A. pisum* (Hemiptera), *Apis melifera* (Hymenoptera) and *D. melanogaster* (Diptera), a glutamic acid residue (E) at this position was also found. On the other hand, an alanine residue (A) was found at this position in *L. hesperus* (Hemiptera)

indicating that this position is not conserved across insects. The mutation at site 132 was an A-G transition in the second position of codon 44, where an asparagine residue (N) was replaced by a serine (S). All species within *Philaenus* showed an asparagine (N) at this position. The exception was found for *P. spumarius* that have both amino acid residues, an asparagine residue (N) and a serine (S). *Mesoptyeflus impictifrons* and *A. melifera* have an asparagine (N) while a serine (S) was found in *A. pisum. Lygus hesperus* and *D. melanogaster* showed different amino acid residues at this codon position, an alanine (A) and an aspartic acid (D), respectively. Both replacements were conservative mutations within *Philaenus* (Fig. 2).

From the 10 silent sites (nine were synonymous mutations) observed in *Philaenus*, two mutations at positions 271 and 343 were fixed in *P. arslani*, separating it from the other species. *Philaenus maghresignus* has also a fixed mutation at position 133. By contrast, the remaining species can not be distinguished from each other by any fixed mutation.

Signatures of selection in yellow sequences

No signature of selection was detected with Tajima's *D* test. For *P. spumarius* the value was positive but not significant (Table 1). Two non-synonymous mutations were found in the *yellow* fragment. Therefore, PAML was used for detecting signal of positive selection in this fragment and a possible protein evolution. The site model M2 was not significantly more adjusted to the data than the site model M1 for the investigated region. However, the M2 detected one position potentially under positive selection, SNP position 91 (amino acid position 30) but no significant *p*-value. For the branch-site model selection, two branches were tested. The results indicated that the alternative model, was not significantly different from the null model for the tested branches. However, the alternative model for one of the branches detected one position potentially under positive selection, SNP position 132 (amino acid position 44), with a significant *p*-value (p-value < 0.05). This indicates that in some *P. spumarius* individuals from Morocco and Iberian Peninsula, *yellow* may be under positive selection.

Table 1: Number of alleles, number of haplotypes, haplotype diversity (*h*), number of segregating sites (S), number of silent sites, synonymous (Syn. Sub) and non-synonymous (Nonsyn. Sub) mutations, nucleotide diversity (π), Tajima's D value and number of homozygotes and heterozygotes calculated for *yellow* fragment and for *Philaenus* species and *P. spumarius* populations.

Species	Number of alleles	Number of haplotypes	*h*	Segregating Sites (S)	Silent Sites	Syn. Sub	Nonsyn. Sub	π	Tajima's *D*	Number of Homozygotes	Number of Heterozygotes
Philaenus	252	12	0.766	12	10	9	2	0.0043	0.14236	85	41
P. spumarius	106	9	0.803	9 (4 Singleton)	7	6	2	0.0046	0.71663	15	38
Portugal	82	6	0.759	7 (1 Singleton)	5	4	2	0.0047	1.5028	9	32
Morocco	14	5	0.769	4 (1 Singleton)	3	2	1	0.0031	0.56234	3	4
Finland	2	2	1.000	2	2	2	0	0.0042	NA	0	1
Balkans	4	3	0.833	2	2	2	0	0.0021	-0.7099	1	1
Anatolia	4	2	0.667	1	1	1	0	0.0014	1.63299	2	0
P. italosignus	30	2	0.067	1	1	1	0	0.0001	-1.147	14	1
P. maghresignus	26	2	0.148	1	1	1	0	0.0003	-0.71385	11	2
P. tarifa	30	1	0.000	0	0	0	0	0.0000	NA	15	0
P. signatus	30	1	0.000	0	0	0	0	0.0000	NA	15	0
P. arslani	30	1	0.000	0	0	0	0	0.0000	NA	15	0

	9					**14**					**19**					**24**					**29**					**34**					**39**					**44**					**49**					**54**					**59**
	*	.	.	.	.	*	.	.	.	.	*	.	.	.	.	*	.	.	.	.	*	.	.	.	.	*	.	.	.	.	*	.	.	.	.	*	.	.	.	.	*	.	.	.	.	*	.	.	.	.	*
PSP_scaffold	Y	F	M	P	D	P	L	A	G	D	F	N	I	A	G	L	N	F	Q	W	G	D	E	G	I	F	G	L	A	L	T	P	L	Q	S	S	G	F	R	T	L	L	F	H	P	L	A	S	N	R	E
Philaenus spumarius_V2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Philaenus spumarius_V3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Philaenus spumarius_V4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Philaenus maghresignus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Philaenus tarifa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Philaenus italosignus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Philaenus signatus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Philaenus arslani	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mesoptyeflus impictifrons	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Acyrthosiphon pisum	F	L	.	.	.	.	.	.	.	.	Y	.	.	G	.	I	.	.	.	.	.	E	.	.	.	.	.	I	T	.	S	.	I	T	E	.	.	.	.	L	.	F	.	.	.	.	.	.	.	.
Lygus hesperus	F	.	H	.	.	.	.	K	.	.	.	.	.	G	.	.	.	.	.	.	E	A	.	.	.	.	.	M	.	.	S	.	-	R	Q	A	A	D	.	F	.	F	.	.	.	.	.	.	.	.
Drosophila melanogaster	.	.	F	.	.	.	.	R	.	.	.	.	V	.	.	I	.	.	.	.	.	E	.	.	.	.	.	M	S	.	S	.	I	R	.	D	.	Y	.	.	.	Y	.	S	.	.	.	H	.	Q
Apis melifera	.	.	.	.	.	.	.	.	.	.	Y	.	.	G	.	.	.	.	.	.	.	E	.	.	.	.	.	M	S	.	S	.	I	A	V	N	.	Y	.	.	.	F	.	.	.	S	.	R	.	.

Fig. 2: Partial alignment of *yellow* amino acid chain from *Philaenus* species with other insect species. The positions where amino acid changes were detected within *Philaenus* are highlighted to white (positions 30 and 44).

Phylogenetic patterns comparisons between yellow and ITS2

A phylogenetic pattern was observed for *Philaenus* genus using the fragment *ITS2* (Fig. 3). Species are divided in two main lineages: a lineage *A* which includes the species *P. signatus*, *P. maghresignus*, *P. tarifa* and *P. italosignus*, and a lineage *B* that contains the *P. spumarius*, the *P. tesselatus*, *P. loukasi* and *P. arslani*. The lineage *B* is divided in two sub-lineages diverging by more than one mutation: a sub-lineage including individuals from species *P. spumarius* and from *P. tesselatus*; and a lineage that clustered the high-altitude habitat species *P. arslani* and *P. loukasi*. Species within the group *A* cannot be distinguished by any fixed mutation. This pattern of inter-specific differentiation contrasted with the pattern observed for the *yellow* gene (Fig. 1 and 3). Nucleotide and haplotype diversities were slightly higher for *ITS2* ($\pi = 0.0093$ and $h = 0.779$) than for *yellow* (Table 2).

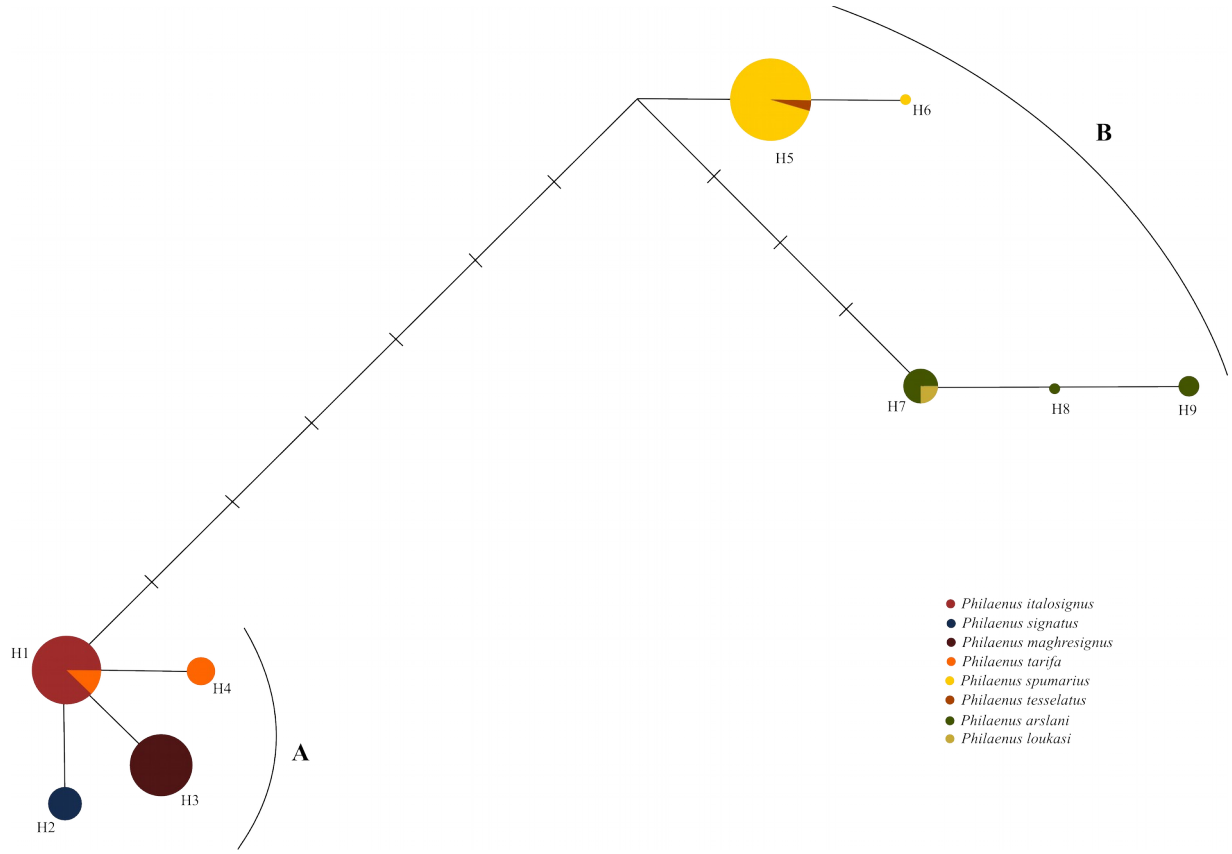


Fig. 3: Median-joining haplotype network for the nuclear *ITS2* gene (445bp) and for *Philaenus* species. A mutational step between haplotypes is represented by a dash.

Table 2: Number of alleles, number of haplotypes, haplotype diversity (h), number of segregating sites (S), nucleotide diversity (π) and number of homozygotes and heterozygotes calculated for *ITS2* fragment and for *Philaenus* species.

Species	Number of alleles	Number of haplotypes	*h*	Segregating Sites (S)	π	Number of Homozygotes	Number of Heterozygotes
Philaenus	128	9	0.779	15	0.0093	56	8
P. spumarius	42	2	0.048	0	0.0001	20	1
P. italosignus	28	1	0.000	0	0.0000	14	0
P. maghresignus	26	1	0.000	0	0.0000	13	0
P. tarifa	10	2	0.533	1	0.0012	1	4
P. signatus	8	1	0.000	0	0.0000	4	0
P. arslani	10	3	0.600	2	0.0023	2	3
P. loukasi	2	1	0.000	0	0.0000	1	0
P. tesselatus	2	1	0.000	0	0.0000	1	0

Association analyses

Association analyses performed using Fisher's exact test did not found any statistical association between *yellow* SNPs and the three dorsal colour phenotypes (*p*-value > 0.05). The median-joining network did not show any association between *yellow* haplotypes and colour morphs. The haplotypes were all uniformly distributed across the three colour morphs, except H4 that only exists in *trilineatus* individuals (Fig. 4). There were genotypes (combination of two haplotypes) that were exclusive to a colour morph. Genotype 2 (H1/H2) and genotype 12(H3/H3) were only found in *marginellus* while genotype 1 (H1/H1), genotype 8 (H2/H3) and genotype 19 (H5/H5) were observed only in *typicus* individuals (Table S2 and S3). *trilineatus*

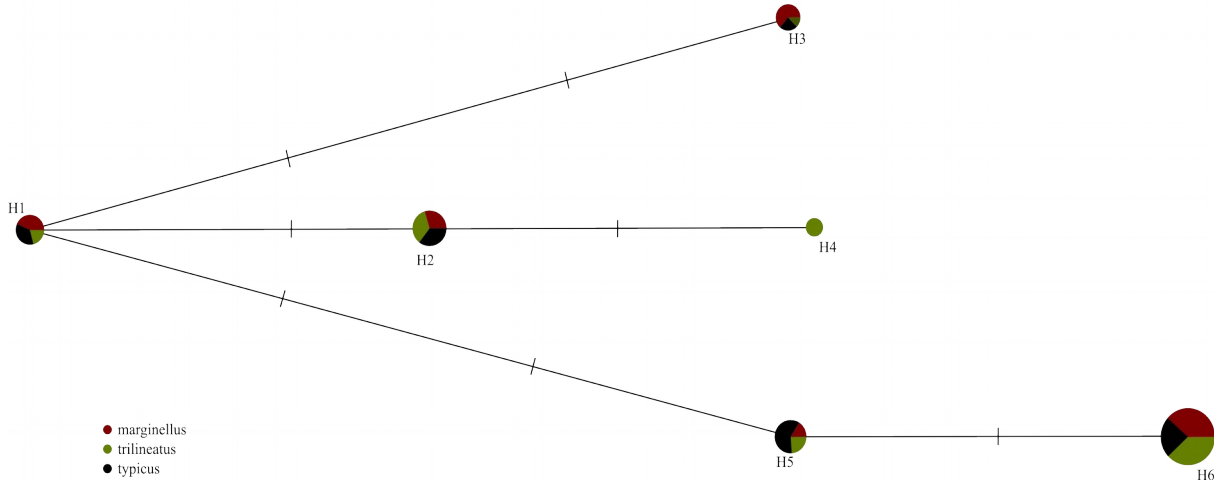


Fig. 4: Median-joining haplotype network for the putative *yellow* gene (482bp) and for the three dorsal colour phenotypes of *P. spumarius*. A mutational step between haplotypes is represented by a dash.

presented the higher number of unique genotypes (genotype 3 (H1/H3); genotype 7 (H2/H2); genotype 16 (H4/H4); genotype 17 (H4/H5) and genotype 18 (H4/H6) (Table S2 and S3). On the contrary, the genotypes more common were the genotypes 11 (n=8), 20 (n=5) and 21 (n=6) (Table S2 and S3). The majority of the individuals were heterozygous. From 14 individuals *marginellus* and *trilineatus*, 10 and 11 were heterozygous, respectively. Eleven out of 13 *typicus* individuals were heterozygous.

Discussion

In this work, we chose to investigate the *yellow* gene, a gene known to be involved in melanin production. This is because it seemed more relevant in *P. spumarius* where melanin based pigments are, most likely, responsible for its different colour patterns. Because sequence similarity to *Drosophila* was low, to isolate the coding sequence of *yellow* in this species revealed challenging. Moreover, only a small fraction (25%) of the *P. spumarius* genome (5.3 Gb), previously sequenced and assembled (see Chapter 4), was available to be examined. Therefore, it is possible the entire coding sequence of the *yellow* gene was not represented in that part of *P. spumarius* genome. The attempting to isolate *yellow* also failed in the transcriptome, probably, because the individual used to obtained the transcripts was in a stage of its life cycle when the expression of *yellow* is limited or inexistent.

Nevertheless, we successfully amplified a fragment of 482 bp, corresponding to part of exon 2 of *Drosophila*'s *yellow* gene. The association analyses did not find any association between *yellow* SNPs and the three dorsal colour morphs. No correlation between *yellow* variants (haplotypes) and these morphs was found either. It is possible that, as this fragment corresponds to a coding region, more exactly to a conserved domain, it is under purifying selection. In fact all mutations detected were conservative, and did not implicate a change in the protein structure. The number of heterozygotes and homozygotes did not differ among the three dorsal colour morphs analysed. However, a higher number of heterozygotes than homozygotes was observed for all morphs. Although only part of the coding sequence (29.64%, *Drosophila* gene size used as reference) has been investigated, these results may indicate that *yellow* is not involved in colour variation in this species. However, we can not totally exclude a possible effect of this gene until its entire coding sequence be analysed.

The phylogenetic pattern of *yellow* was also investigated. Low differentiation was observed within *Philaenus* indicating that this gene is conserved across the genus. In *Drosophila,* this gene promotes the formation of black melanin (Wittkopp *et al.*, 2002b). Assuming that *yellow* has also a function in these species, new mutations that may arise and

alter the protein are expected to be eliminated by natural selection. This would explain the lack of genetic divergence found for this fragment within *Philaenus*. Although we cannot exclude the possibility of *yellow* be involved in the melanin pathway in *Philaenus*, it is also possible that colour variation is related to differences in the expression of *yellow* and/or other candidates, and is not directly related to sequence variation. Mutations in regulatory regions are more likely to contribute to adaptive divergence because, contrary to changes in protein-coding sequences, this changes can produce tissue-specific expression patterns while avoiding deleterious pleiotropic effects (Stern, 2000; Wittkopp & Kalay, 2011). Therefore, regulatory elements may be the responsible as they have been shown to control colour in *Drosophila* (Wittkopp *et al.*, 2002b; Cooley *et al*., 2012). The haplotype H7 is shared among the studied species, except for *P. maghresignus* and *P. arslani*, indicating that this may be the ancestral haplotype. This observed pattern contrasted with the neutral pattern obtained for *ITS2* with two main groups, already described in previous works (Maryanska-Nadachowska, *et al*., 2010; Rodrigues, 2010): a group including *P. spumarius, P. tesselatus, P. arslani* and *P. loukasi* and another group comprising *P. signatus*, *P. maghresignus*, *P. tarifa* and *P. italosignus*, and that corroborates the established division based on the morphology of male genitalia (Drosopoulos & Remane, 2000) and chromosomes (Maryańska-Nadachowska *et al*., 2008, 2012, 2013). These contrasting patterns found between *yellow* and *ITS2* indicate that the evolution of *yellow* is constrained, limiting the number of mutations that can be fixed without the protein function be affected. Curiously, a higher haplotype and nucleotide diversity for *yellow* was found for *P. spumarius*. This could be related to a higher sample size, but for the 13 or 15 individuals of species *P. maghresignus*, *P. italosignus*, *P. tarifa*, *P. signatus* and *P. arslani*, we only detected 1 or 2 haplotypes, while in some populations of *P. spumarius* with 1, 2 or 7 individuals, 2, 3 and 5 haplotypes were detected. These species have different ecologies. *Philaenus spumarius* is widely distributed, occupying a higher number of terrestrial habitats (Halkka & Halkka, 1990; Quartau & Borges, 1997; Yurtsever, 2000), thus having probably been less affected by drift or population declines (bottlenecks) during less favourable periods. On the contrary, the other Mediterranean species are mostly specialists regarding food and oviposition plants (Drosopoulos, 2003), and, in this way, more susceptible to these factors.

Although the entire sequence of *yellow* is not included in the available draft genome, and it has not been found in the transcriptome of *P. spumarius* that we assembled (see Chapter 4), the candidate gene approach still remains valid and other colour genes, already described in insects, should be investigated. Ultimately, the analysis of the candidates could reveal no relation to colour in this species. Novel and uncharacterised genes may be implicated as in

Heliconius group, where a new gene was found to control wing colour pattern (Nadeau *et al*., 2016). On the other hand, in *Philaenus*, the pattern of colouration, with some body regions darker and others lighter, suggests that variation may be in the differential expression of pigment genes across those body regions and not due to mutations on coding genes. Therefore, further gene expression analyses in *P. spumarius* individuals from different colour morphs, using techniques like tiling arrays (Bertone *et al.*, 2005; Mockler & Ecker, 2005) or RNA sequencing (Wang *et al*., 2009), would be important to uncover the genes involved in this species' colour variation. In some Lepidoptera species, gene expression analyses proved to be very useful (Nadeau *et al.*, 2016). Since major loci, consisting of multi-genomic regions, are suggested to control dorsal colour variation in this species (see Chapter 4), identifying the independent contribution of a single genetic variant for this polymorphism may be challenging. Individuals from different colour morphs could be used for transcriptome sequencing and differences in the transcripts investigated. However, in the 5thinstar, the stage when many genes controlling adult color pattern are presumed to be expressed, the different colour morphs are undistinguished. Genome wide association mapping analyses can be, therefore, a complementary approach to highlight unsuspected targets of selection in the genome and possible candidates.

Acknowledgments

Financial support was provided from National Funds through FCT – Fundação para a Ciência e a Tecnologia, Portugal under projects PTDC/BIA-BEC/098783/2008 and UID/BIA/00329/2013, and grants SFRH/BD/73879/2010 (attributed to ASR); SFRH/BPD/26365/2006 (attributed to SGS). We thank for samples to Eduardo Marabuto, Sara E. Silva, José Quartau, Luis Ferreira, Selçuk Yurtserver, Andreia Penado, Vera Nunes, Carla Pereira, Filipe Ribeiro, Daniel Pires, Rui Nunes, Paulo Borges, Paulo Fonseca, Olli Halkka, Liisa Halkka and Vera d' Urso.

References

Abdul-Nour H, Lahoud L (1996) Révision du genre *Philaenus* Stål, 1964 au Liban avec la description d'une nouvelle espèce: *P. arslani* n. sp. (Homoptera, Cercopidae). *Nouv. Rev. Entomol.*, **12**, 297–303.

Altschul SF, Madden TL, Schäffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289 – 300.

Bertone P, Gerstein M, Snyder M (2005) Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, **13**, 259–74.

Cooley AM, Shefner L, McLaughlin WN, Stewart EE, Wittkopp PJ (2012) The ontogeny of color: Developmental origins of divergent pigmentation in *Drosophila americana* and *D. novamexicana*. *Evolution and Development*, **14**, 317–325.

Drosopoulos S (2003) New data on the nature and origin of colour polymorphism in the spittlebug genus *Philaenus* (Hemiptera : Aphorophoridae). *Annales De La Societe Entomologique De France*, **39**, 31–42.

Drosopoulos S, Asche M (1991) Biosystematic studies on the spittlebug genus *Philaenus* with the description of a new species. *Zoological Journal of the Linnean Society*, **101**, 169–177.

Drosopoulos S, Maryańska-Nadachowska A, Kuznetsova VG (2010) The Mediterranean: Area of origin of polymorphism and speciation in the spittlebug *Philaenus* (Hemiptera, Aphrophoridae). *Zoosystematics and Evolution*, **86**, 125–128.

Drosopoulos S, Remane R (2000) Biogeographic studies on the spittlebug *Philaenus signatus* melichar, 1986 species group (hemiptera: aphrophoridae) with the description of two new allopatric species. *Ann. Soc. Entomol. Fr.*, **36**, 269–277.

Forsman A, Ahnesjö J, Caesar S, Karlsson M (2008) A model of ecological and evolutionary consequences of color polymorphism. *Ecology*, **89**, 34–40.

Futahashi R, Sato J, Meng Y *et al.* (2008) yellow and ebony are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*. *Genetics*, **180**, 1995–2005.

Gray SM, McKinnon JS (2007) Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution*, **22**, 71–79.

Halkka O, Halkka L (1990) Population-Genetics Of the Polymorphic Meadow Spittlebug, *Philaenus spumarius* (L). *Evolutionary Biology*, **24**, 149–191.

Halkka O, Halkka L, Raatikainen M, Hovinen R (1973) The genetic basis of balanced polymorphism in *Philaenus* (Homoptera). *Hereditas*, **74**, 69–80.

Hallka O, Lallukka R (1969) The origin of balanced polymorphism in the spittlebugs (*Philaenus*, Homoptera). *Annales Zoologici Fennici*, **6**, 431–434.

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

Joron M, Wynne IR, Lamas G, Mallet J (1999) Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. *Evolutionary Ecology*, **13**, 721–754.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, **30**, 772–80.

Kim HJ, Lee SH (2008) Molecular systematics of the genus *Megoura* (Hemiptera: Aphididae) using mitochondrial and nuclear DNA sequences. *Molecules and Cells*, **25**, 510–522.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)*, **25**, 1451–2.

Maryańska-Nadachowska A, Drosopoulos S, Lachowska D, Kajtoch Ł, Kuznetsova VG (2010) Molecular phylogeny of the Mediterranean species of *Philaenus* (Hemiptera: Auchenorrhyncha: Aphrophoridae) using mitochondrial and nuclear DNA sequences. *Systematic Entomology*, **35**, 318–328.

Maryańska-Nadachowska A, Kuznetsova VG, Drosopoulos S, Lachowska-Cierlik D (2008) A chromosomal analysis of eight Mediterranean species of *Philaenus*. *Bulletin of Insectology*, **61**, 133–134.

Maryańska-Nadachowska A, Kuznetsova VG, Karamysheva TV. (2013) Chromosomal location of rDNA clusters and TTAGG telomeric repeats in eight species of the spittlebug genus *Philaenus* (Hemiptera: Auchenorrhyncha: Aphrophoridae). *European Journal of Entomology*, **110**, 411–418.

Maryańska-Nadachowska A, Kuznetsova VG, Lachowska D, Drosopoulos S (2012) Mediterranean species of the spittlebug genus *Philaenus*: modes of chromosome

evolution. *Journal of insect science*, **12**, 54.

Mockler TC, Ecker JR (2005) Applications of DNA tiling arrays for whole-genome analysis. *Genomics*, **85**, 1–15.

Nadeau NJ, Pardo-Diaz C, Whibley A *et al.* (2016) The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, **534**, 106–110.

Nijhout HF (2003) Polymorphic mimicry in *Papilio dardanus*: mosaic dominance, big effects, and origins. *Evolution and Development*, **5**, 579–592.

Pina-Martins F, Paulo OS (2008) concatenator: sequence data matrices handling made easy. *Molecular ecology resources*, **8**, 1254–5.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, **81**, 559–75.

Quartau JA, Borges PAV (1997) On the colour polymorphism of *Philaenus spumarius* (L .) in Portugal. *Miscel-lánia Zoológica*, **2**, 19–30.

Reininga JM, Nielsen D, Purugganan MD (2009) Functional and geographical differentiation of candidate balanced polymorphisms in *Arabidopsis thaliana*. *Molecular Ecology*, **18**, 2844–2855.

Remane R, Drosopoulos S (2001) *Philaenus tarifa* sp. n.: An additional spittlebug species from southern Spain (Homoptera, Auchenorrhyncha Cercopidae). *Mitteilungen aus dem Museum fur Naturkunde in Berlin - Deutsche Entomologische Zeitschrift*, **48**, 277–279.

Richards PM, Liu MM, Lowe N *et al.* (2013) RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology*, **22**, 3077–3089.

Rodrigues ASB (2010) Estudo dos padrões filogenéticos e filogeográficos em *Philaenus spumarius* (Hemiptera, Aphrophoridae) e espécies próximas.

Schultz J, Wolf M (2009) ITS2 sequence-structure analysis in phylogenetics: A how-to manual for molecular systematics. *Molecular Phylogenetics and Evolution*, **52**, 520–523.

Seabra SG, Pina-Martins F, Marabuto E *et al.* (2010) Molecular phylogeny and DNA barcoding in the meadow-spittlebug *Philaenus spumarius* (Hemiptera, Cercopidae) and its related species. *Molecular Phylogenetics and Evolution*, **56**, 462–467.

Silva SE, Rodrigues ASB, Marabuto E *et al.* (2015) Differential survival and reproduction in colour forms of *Philaenus spumarius* give new insights to the study of its balanced polymorphism. *Ecological Entomology*, **40**, 759–766.

Song N, Liang A-P, Bu C-P (2012) A molecular phylogeny of Hemiptera inferred from mitochondrial genome sequences. *PloS one*, **7**, e48778.

Stephens M, Donnelly P (2003) Report A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. *Am. J. Hum. Genet*, **73**, 1162–1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American journal of human genetics*, **68**, 978–989.

Stern DL (2000) Perspective: Evolutionary Developmental Biology and the Problem of Variation. *Evolution*, **54**, 1079.

Stewart AJA, Lees DR (1988) Genetic control of colour/pattern polymorphism in British populations of the spittlebug *Philaenus spumarius* (L.) (Homoptera: Aphrophoridae). *Biological Journal of the Linnean Society*, **34**, 57–79.

Stewart a. J a., Lees DR (1996) The Colour/Pattern Polymorphism of *Philaenus spumarius* (L.) (Homoptera: Cercopidae) in England and Wales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **351**, 69–89.

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

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Thompson MJ, Jiggins CD (2014) Supergenes and their role in evolution. *Heredity*, **113**, 1–8.

Tishechkin DY (2013) Two new species of the genus *Philaenus* (Homoptera, Aphrophoridae) from Iran. *Entomological Review*, **93**, 73–76.

Tuttle EM, Bergland AO, Korody ML *et al.* (2016) Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Current Biology*, 1–7.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, **10**, 57–63.

Wittkopp PJ, Beldade P (2009) Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy. *Seminars in Cell and Developmental Biology*, **20**, 65–71.

Wittkopp PJ, Kalay G (2011) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, **13**, 59–69.

Wittkopp PJ, True JR, Carroll SB (2002a) Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development*, **129**, 1849–58.

Wittkopp PJ, Vaccaro K, Carroll SB (2002b) Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Current Biology*, **12**, 1547–1556.

Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yurtsever S (2000) On the polymorphic meadow spittlebug, *Philaenus spumarius* (L.) (Homoptera: Cercopidae). *Turkish Journal of Zoology*, **24**, 447–459.

CHAPTER 5

Supporting Information

Table S1: Analysed samples of *Philaenus* and of the outgroup with description of the sampling locations. Indication of the samples that were sequenced for *yellow* and/or *ITS2*.

		Location (Collector)	GPS Coordinates	*yellow*	*ITS2*
Philaenus					
P. spumarius					
	Portugal	Bom Sucesso, Foz do Arelho (Quartau, Seabra & Penado)	39°25'2.95"N; 9°13'39.18"W	yes	
		Fontanelas, Sintra (Rodrigues, Seabra & Pereira)	38°50'15.75"N; 9°25'20.77"W	yes	yes
		Ribeira do Torgal, Odemira, Alentejo (Ribeiro & Pires)	37°39'35.14"N; 8°37'40.86"W		yes
		Santa Clara a Velha, Rio Mira, Alentejo (Rodrigues & Marabuto)	37°30'38.22"N; 8°28'27.23"W		yes
		Silves, Algarve (Seabra)	37°11'40.31"N; 8°27'55.28"W		yes
	Morocco	Near Rabat (Rodrigues, Silva, Marabuto & Ferreira)	33°46'37.56"N; 7°13'58.92"W	yes	yes
		Rabat (Rodrigues, Silva, Marabuto & Ferreira)	34°0'41.70"N; 6°42'32.94"W	yes	yes
		Rif (Rodrigues, Silva, Marabuto & Ferreira)	35°51'58.14"N; 5°24'30.30"W	yes	yes
		Ifrane (Rodrigues, Silva, Marabuto & Ferreira)	33°30'41.82"N; 5°5'34.08"W	yes	yes
		Ifrane Centre (Rodrigues, Silva, Marabuto & Ferreira)	33°31'58.86"N; 5°6'7.86"W	yes	yes
		Azrou (Rodrigues, Silva, Marabuto & Ferreira)	33°26'57.78"N; 5°13'55.14"W	yes	yes
		Azrou (Rodrigues, Silva, Marabuto & Ferreira)	33°29'39.48"N; 5°15'47.46"W	yes	yes
	Finland	Brannskar (Halkka)		yes	yes
	Greece	Mt Parnassus (Rodrigues, Silva, Marabuto)	38°34'42.24"N; 22°34'30.30"E	yes	
	Turkey	Azdavay (Yurtserver)	41°38'29.26"N; 33°17'52.54"E	yes	
		Keçan (North) (Yurtserver)	40°53'29.60"N; 26°38'42.81"E	yes	
		Cerkes (Yurtserver)	40°48'59.24"N; 32°54'9.50"E	yes	yes
	UK	Cambridge, England (Borges)	52°12'22.80"N; 0°7'29.41"E		yes
	France	Lautrec (Rodrigues, Silva, Marabuto, Nunes & Ferreira)	43°42'47.14"N; 2°7'6.04"E		yes
	USA	Michigan (Fonseca)	44°18'49.13"N; 85°35'6.26"W		yes
P. italosignus					
	Italy	Tardaria, Etna, Sicily (d'Urso)	37°43'53.05"N; 14°59'8.12"E	yes	
		Puglia (Marabuto)	40°41'26.60"N; 16°57'22.90"E	yes	yes
		Cosenza (Marabuto, Silva)	39°25'49.19"N; 16°36'19.55"E	yes	yes
P. maghresignus					
	Portugal	Ribeira do Torgal, Odemira, Alentejo (Ribeiro & Pires)	37°39'35.14"N; 8°37'40.86"W	yes	yes
		Amadora (Marabuto)	38°44'10.27"N; 9°14'32.57"W	yes	
		Negrais – Sintra (Rodrigues, Silva & Seabra)	38°52'19.50"N; 9°17'11.82"W	yes	yes
P. tarifa					
	Gibraltar	Gibraltar (Rodrigues, Marabuto)	36° 8'12.17"N; 5°20'51.36"W	yes	yes
P. signatus					
	Turkey	Gokçeada (Yurtserver)	40°12'44.04"N; 25°54'25.47"E	yes	yes
P. arslani					
	Lebanon	Chabrouh (Quartau)	34°1'42.12"N; 35°50'17.36"E	yes	yes
		Faqra (Quartau)	34°0'28.09"N; 35°50'16.90"E	yes	yes
Mesoptyelus impictifrons					
	Lebanon	Laqlouk (Quartau)	34° 8'2.13"N; 35°51'40.61"E	yes	

Table S2: *yellow* haplotypes obtained for the three dorsal colour morphs of *P. spumarius* and frequency.

***yellow* haplotypes**	***marginellus* Freq.**	***trilineatus* Freq.**	***typicus* Freq.**	**Total Freq.**
H1 : T C T A G A C	4	2	3	9
H2 : T C C A G A C	4	5	5	14
H3 : T T T A T A C	5	1	2	8
H4 : T C C C G A C	0	4	0	4
H5 : C C T A G G C	2	3	8	13
H6 : C C T A G G T	13	13	8	34

Table S3: *yellow* genotypes for the three dorsal colour morphs of *P. spumarius* and frequency.

Genotypes	**Haplotype Combination**		***marginellus* Freq.**	***trilineatus* Freq.**	***typicus* Freq.**	**Total Freq.**
G1	H1	H1	0	0	1	1
G2	H1	H2	1	0	0	1
G3	H1	H3	0	1	0	1
G4	H1	H4	0	0	0	0
G5	H1	H5	1	0	1	2
G6	H1	H6	2	1	0	3
G7	H2	H2	0	1	0	1
G8	H2	H3	0	0	1	1
G9	H2	H4	0	0	0	0
G10	H2	H5	1	1	1	3
G11	H2	H6	2	3	3	8
G12	H3	H3	1	0	0	1
G13	H3	H4	0	0	0	0
G14	H3	H5	0	0	0	0
G15	H3	H6	3	0	1	4
G16	H4	H4	0	1	0	1
G17	H4	H5	0	1	0	1
G18	H4	H6	0	1	0	1
G19	H5	H5	0	0	1	1
G20	H5	H6	0	1	4	5
G21	H6	H6	3	3	0	6

CHAPTER 6

General Discussion

6.1 GENERAL DISCUSSION

The present study aimed to give insights on the genetic basis underlying phenotypic variation involved in evolutionary change and adaptation. In order to do that, the molecular basis and the adaptive significance of *Philaenus spumarius*' dorsal/pattern colour polymorphism was explored, and possible genomic regions linked to this polymorphism identified. The evolutionary history of this species was also investigated in more detail. Patterns of neutral variation were analysed and time estimates provided for the main evolutionary events. The following chapter presents a brief discussion of the main results reported in chapters 2 to 5.

On the evolutionary history of *P. spumarius*

The biogeographic pattern of the meadow spittlebug *P. spumarius* has been investigated in the last few years, using mitochondrial and nuclear genes (Rodrigues, 2010; Seabra *et al*., 2010; Lis *et al.*, 2014; Maryańska-Nadachowska *et al.*, 2012, 2015). Mitochondrial DNA markers showed an earlier split of *P. spumarius* in two main mitochondrial lineages, the ‘‘western’’ in the Mediterranean region and the ‘‘eastern’’ in Anatolia/Caucasus region. Within the ‘‘western’’ lineage a posterior differentiation in two sub-lineages was observed: the ‘‘western-Mediterranean’’, centred in the Iberian Peninsula, and the ‘‘eastern-Mediterranean’’, centred in the Balkans. Mitochondrial haplogroups did not correspond to the nuclear clades, suggesting the existence of admixture and/or incomplete lineage sorting. In this work, we do not only corroborated this general pattern as also contributed with new findings for the knowledge of *P. spumarius* evolutionary history (Chapter 2). Time estimates indicated that divergence within *P. spumarius* is recent, having occurred most probably in the Middle/Late Pleistocene (no more than 0.5 Ma), and is probably related to the climate changes of the Quaternary period (~2.588–0.0117 Ma ago (Gibbard *et al.*, 2010)). There are evidence that divergence events took place in southern Mediterranean refugia during Pleistocene glacial periods, and that they were followed by northwards population expansions from the Iberian Peninsula to the central and north-western Europe and, from the Anatolia/Caucasus (and eventually from western Asia) to east, north and central Europe, during the warmer interglacials periods. This study also detected gene-flow among the main southern Mediterranean peninsulas. Therefore, the current biogeographic pattern of the species do not follow none of the common four paradigms described for the European species (Hewitt, 1999; Habel *et al.*, 2005).

The colonisation patterns out of Eurasia, namely those of north-western Africa, North America, and the islands of Azores and New Zealand, were investigated here for the first time. The analysed individuals from the Azores archipelago and New Zealand were found to have originated most likely from British populations. Apart from the geographic origin from western Europe, already reported in previous studies (Rodrigues, 2010; Seabra *et al*., 2010; Maryańska-Nadachowska *et al.*, 2012; Lis *et al.*, 2014), a new origin, from the "eastern" lineage, was identified for North American populations. This study also indicates that, as observed for many species (Franck *et al.*, 2001; Schimitt *et al.*, 2005; Horn *et al.*, 2006; Rodriguez-Sanchez *et al.*, 2008), the Strait of Gibraltar appears have not been an effective barrier to the dispersal of *P. spumarius* between Europe and Africa during glacial periods. This is suggested by the sharing of mitochondrial haplotypes from the ''western-Mediterranean'' sub-lineage, between Morocco and the Iberian Peninsula populations, and corroborated by the nuclear data. Here, once again, it was evident that the taxonomic status of *P. tesselatus* species is puzzling and demands further investigation. Although the genetic data indicate individuals from Morocco are *P. spumarius*, morphological analyses of male genitalia revealed similarities with the genitalia described for *P. tesselatus* (Drosopoulos & Quartau, 2002).

No long ago, the majority of phylogeographic analyses were based on the traditional mitochondrial DNA and sometimes complemented with a few nuclear markers (e.g. Lunt *et al.*, 1998; Kawahara *et al.*, 2009; Bihari *et al.*, 2011). Despite their advantages, information provided by mitochondrial genes may not always be concordant with information from nuclear markers. Because they are maternally inherited, they may have an unique evolutionary history. Furthermore, mitochondrial and nuclear genes may not represent the whole genome diversity and tell an incomplete history. With the new sequencing methods it is now possible to address the evolutionary history of organisms in a genome wide scale by analysing thousands of independent loci (Emerson *et al.*, 2010; McCormack *et al.*, 2013). For *P. spumarius*, a phylogeographic study based on genome wide markers would potentially improve the robustness and complement the results obtained by the traditional genetic markers. This was true for the montane caddisfly *Thremma gallicum* (Macher *et al.,* 2015). A population genomic analysis using genome wide markers would also be useful to disentangled the taxonomic status of *P. tesselatus* and to understand if we are in the presence of two distinct taxonomic units/species or in the presence of intra-specific morphological variation within *P. spumarius.* An effective method to generate genome wide loci for *P. spumarius*, a species with a genome size larger than 5.0 Gb and lacking a complete reference genome, would be the restriction-site-associated DNA sequencing (RAD) technique (Baird *et al.*, 2008).

The investigation of the evolutionary history of *P. spumarius* and its main demographic events are important contributions to our understanding of the process of adaptation in this species, and, particularly, to help us to understand how the dorsal colour/pattern polymorphism is maintained in natural populations. This is because, by comparing genes with a neutral pattern with genes evidencing an adaptive pattern, it may be possible to distinguish the effects resulting from evolutionary history and demography from the effects of natural selection.

Assessing the adaptive significance of *P. spumarius* dorsal colour polymorphism

Colouration is known to have an adaptive function (Forsman *et al.*, 2008). Although several mechanisms, have been proposed to explain the observed polymorphisms (Gray & McKinnon, 2007), the maintenance of different colour morphs in natural populations is still puzzling. As already refered, *P. spumarius* shows a dorsal colour/pattern polymorphism extensively studied (Halkka & Halkka, 1990) but whose adaptive significance and evolutionary mechanisms are still poorly understood. Aiming to uncover this polymorphism's adaptive function, an experimental approach, conducted in captivity under semi-natural conditions, was performed and the differential survival, reproductive success, and duration of egg maturation in three dorsal colour phenotypes (*typicus*, *trilineatus* and *marginellus*) tested (Chapter 3). Interestingly, a higher longevity, a higher number of oviposition events, and a higher number of eggs laid was found for *trilineatus* phenotype females than for *typicus* and *marginellus*. This supports previous reports of an increase in *trilineatus* frequency during the season (Owen & Wiegert, 1962; Halkka, 1964). The higher values observed for *trilineatus* may be a way to counteract the higher rate of attack by parasitoids (Harper & Whittaker, 1976) and/or the higher solar radiation reflectance (Berry & Willmer, 1986) reported for this phenotype in previous studies, thus constituting a possible mechanism for the maintenance of this polymorphism in *P. spumarius* populations. Contrary to what was found for some ladybird species (Brakefield & Willmer, 1985; Rhamhalinghan, 1990), the duration of egg maturation did not differ among the three *P. spumarius* colour phenotypes analysed. This indicates that, maybe there is no relation between melanism and egg maturation efficiency in this species. However, only the melanic form *marginellus* was investigated. It would be important to study the remaining melanic phenotypes and even the different levels of melanism in *typicus*. In the tested conditions, no differences were observed between the emergence times of males and females, or among different phenotypes. This was curious since an early emergence of males, as well as *typicus*, in comparison with other phenotypes was observed by Halkka *et al.* (1967). They suggested that males and *typicus* emerged earlier and were favoured by natural selection as a form to prevent

predation (apostatic selection) and increase protection of females until egg-laying. This was not supported by our results. It is possible that the very beginning of the emergence has not be recorded, however, it is very unlikely.

Laboratory studies have been widely used to study polymorphisms in many animal species (e.g ladybirds, Brakefield & Willmer (1985); coral-reef fishes, Munday *et al.* (2003)). In *P. spumarius*, laboratory experiments revealed extremely challenging mainly due to the long life cycle of the species (annual) and its sensibility to factors such as humidity and temperature, specially during the earlier stages of its life cycle. Despite the difficulties, it was possible to evaluate the effect of the colour phenotypes in the survival and reproduction of *P. spumarius*, and therefore, to contribute to the knowledge of the adaptive function of this colour polymorphism. In this experiment, the same semi-natural conditions were applied to all colour phenotypes. It would be useful to carry out experiments under different environmental conditions (different plant species and (micro)climates) and to test differential fitness among phenotypes. Besides, the potential effects of climate change on the distribution and frequency of phenotypes in this species should be addressed as well. Climate change was already implicated in the northward range shift detected for *P. spumarius* in California (Karban & Strauss, 2004). There are also predictions that the increased temperature, UV radiation, humidity, and pathogens have a selective effect in melanin-based colouration polymorphisms (Roulin, 2014). Therefore, by analysing the effects of climate change on the distribution and frequency of phenotypes in *P. spumarius* may constitute an opportunity to test some of this predictions. During this work, it was also observed a range of colour variation of the *typicus* phenotype, from pale to almost melanic. This variation should be explored in future studies as this may be influencing variation in the studied traits within *typicus*. Other colour phenotypes should also be considered in the future since only three of the sixteen described phenotypes have been tested.

Having been a first step toward the comprehension of the adaptive significance of *P. spumarius*' dorsal colour polymorphism, this work indicated that several factors, acting together, are most likely responsible for the maintenance of this polymorphism. However, to fully understand the process, is essential to know the biogeographic patterns of the species and, at same time, to investigate possible interactions among selection, migration and drift.

On the genetic basis of dorsal colour polymorphism

Although experimental crosses have revealed the Mendelian inheritance of dorsal colour polymorphism in *P. spumarius,* its molecular basis remains unknown. In the present

work we tried to identify a genomic region or genomic regions that could be associated with this species colour variation. Two approaches were chosen: a genome-wide association approach using Single Nucleotide Polymorphisms (SNPs) obtained through RAD sequencing (Chapter 4); and a candidate gene approach to investigate genes involved in melanin-based colouration in other insect species (Chapter 5).

RAD sequencing (Baird *et al.* 2008) has proved to be extremely useful, namely for genome wide association studies (e.g Parchman *et al.*, 2010; Hecht *et al.*, 2013), in species lacking a reference genome or in organisms in which a complete sequence genome is not always possible or suitable. This reduced complexity method of sampling allows to increase the number and type of molecular markers, reduce the costs and increase the speed analyses. By using this method we were able to obtain a set of 1,837 markers across 33 individuals to test for associations with three dorsal colour phenotypes (*typicus*, *marginellus*, and *trilineatus*) in *P. spumarius*. A total of 60 SNPs associated with the dorsal colour pattern were identified by single and multi-association analyses. The associated loci showed stronger differentiation of the *trilineatus* colour phenotype. Interestingly, in the experimental study (Chapter 3), *trilineatus* also revealed to be more differentiated in life-history and physiological traits, showing a higher longevity, a higher number of oviposition events, and a higher number of eggs laid, compared with the *marginellus* and *typicus*. The RAD data suggested that major loci, consisting of multi-genomic regions, may be involved in dorsal colour variation among the three dorsal colour morphs investigated. Taking these findings into account, it can be hypothesised that a complex molecular basis is likely controlling not only colour variation but also life-history and/or physiological traits, and that, as suggested by Halkka & Lallukka (1969), natural selection may not be acting directly in colour variation. In this RAD study, no homology was found between the associated loci and genes known to be responsible for colouration pattern in other insect species. It is possible that these loci are in linkage with the real colour gene/region or that they belong to a regulatory region. The lack of homology can also be due to the absence of a genome sequence belonging to a species close enough to get significant matches, and/or due to the small size of the RAD sequences (less than 100 bp).

In this genomic study, additional genomic and transcriptomic resources were developed to help in the characterisation of the genomic regions associated with colour variation. These consisted of a partial genome assembly, representing 24% of the total size (5.3 Gb), and an 81.4 Mb transcriptome assembly. From the loci found to be associated with colour, 35% aligned to the genome and 10% to the transcriptome. The fact that the RNA sequences used to

assembly the transcriptome were obtained from a fully sclerotised adult *P. spumarius* may partly accounts for the low recovery of mRNA transcripts associated with colour polymorphism. It would be interesting to use mature 5th instar nymphs for transcriptome sequencing since this is probably the stage when many genes controlling adult colour pattern are being expressed.

This work was a first attempt to investigate the genetic architecture of *P. spumarius* dorsal colour polymorphism. It allowed the detection of loci associated with colour, and contributed to the development of genomic and transcriptomic resources that can be very useful for further research on the genetic basis of *P. spumarius*' dorsal colour pattern. The study also showed that working with a non-model species, with a limited genetic knowledge and a large genome size, may be a challenge. In the case of *P. spumarius*, this was evidenced in the RAD sequencing by the high percentage of missing data per individual (63.92%) that we obtained. After filtering for missing data, the number of RAD markers, with a coverage good enough to be used in the association study, decreased from 470,470 to 1,837 SNPs. As consequence, the chances of finding the genomic regions responsible for colour pattern differences were limited, and, in the future, it would be important re-sequencing the individuals in order to improve coverage. The challenge of working with large genomes was also clear by the low fraction of the genome covered in the whole-genome shotgun sequencing performed for *P. spumarius*.

A candidate gene approach can be a powerful tool for studying the genetic architecture of complex traits, being and effective and economical method for direct gene discovery. However, in a non-model species like *P. spumarius*, the difficulties of using this approach, for identifying genes involved in colour variation, were clear (Chapter 5). The choice of candidate genes was one of them. Despite the association study (Chapter 4) has detected loci linked to colour variation, it was not able to characterise those loci. Its identification could, somehow, have given some insights on the genes involved in this species' colour polymorphism. Several regulatory regions are known to control colour pattern in insects. However, they are more variable than coding regions, thus being more difficult to detect and study. In *P. spumarius*, as colour patterns are mainly melanic/non-melanic and, probably, melanin based, three genes (*yellow*, *ebony* and *tan*), known to be involved in the melanin synthesis pathway in *Drosophila* (Wittkopp *et al.*, 2002a; True *et al.*, 2005; Wittkopp *et al.*, 2009), and that seemed more relevant, were selected. To isolate those genes and to look for conserved regions to design primers for their amplification in *P. spumarius,* was another challenge. Initially, the three

Drosophila's gene sequences were queried against the partial genome draft of *P. spumarius* and also against its transcriptome. No homology was found, probably, because these two species were not close enough to get significant matches. Besides, as the available draft genome of the *P. spumarius* was incomplete, and the transcriptome was obtained using an adult specimen, sequences of these genes may not have been represented. Alternatively, sequences of the pea aphid *A. pisum*, the closest species to *P. spumarius* with available genome, were used to look for homologous of *Drosophila*'s genes. Unfortunately, also between those species, homology was difficult to obtain. By querying the *Drosophila*'s gene sequences against pea aphid ESTs database, only an homologous coding sequence of the *yellow* gene was found. It is possible that the other genes are not expressed in the stage of the individuals that were used to obtain the EST database.

The *yellow* sequence was queried against *P. spumarius* draft genome and an homologous region was found. Primers were designed in conserved regions and a fragment of 482 bp, corresponding to part of exon 2 in *Drosophila*'s *yellow* gene, was successfully amplified in three *P. spumarius* colour morphs (*typicus*, *marginellus*, and *trilineatus*) and in some of its close related species. The sequence variation and the amino acid changes observed did not show any association with the three colour morphs (p-value > 0.05). All mutations detected were conservative and did not implicate a change in the protein. These results may indicate that *yellow* is not involved in colour variation in this species but, a possible effect of this gene can not be totally excluded until its entire coding sequence be investigated. The phylogenetic pattern of *yellow* was also investigated. Low genetic differentiation was observed within *Philaenus*, indicating that this gene is conserved across the genus. This pattern contrasted with the neutral pattern of *ITS2*, with two main lineages, already described in previous works (Maryanska-Nadachowska, *et al*., 2010; Rodrigues, 2010). *yellow* is probably a functional gene under a relatively high evolutionary pressure that prevents the accumulation of deleterious mutations and variation. In *Drosophila,* this gene promotes the formation of black melanin (Wittkopp *et al.*, 2002b). It is possible that it is also involved in the melanin pathway in *Philaenus* but that colour variation is related to differences in the expression of *yellow* and/or other candidates, and is not directly related to sequence variation. A higher haplotype and nucleotide diversity, that could be related to a higher sample size or to the different ecology of the species, was observed for *P. spumarius*. Being a widely distributed species, that occupies a higher number of terrestrial habitats (Halkka & Halkka, 1990; Quartau & Borges, 1997; Yurtsever, 2000), *P. spumarius* could have been less affected by the effects of genetic drift or population declines (bottlenecks).

In this study, we tried to explore a set of colour candidates and investigate their possible involvement in colour variation in *P. spumarius,* and also in its close related species. Although only a fragment of the *yellow* gene have been isolated and analysed with success (Chapter 5), this work showed that it is possible to apply a candidate gene approach to a species like *P. spumarius*. Therefore, other candidate genes for colouration should be investigated, including regulatory elements since they are showed to control wing pattern in other insect species (Wittkopp *et al.*, 2002b; Gompel *et al.*, 2005; Wallbank *et al.,* 2016).

6.2 Final Remarks

This thesis provided new insights on the biogeographic pattern of the meadow spittlebug *P. spumarius*, and explored the adaptive significance and the molecular genetic basis underlying this species dorsal colour polymorphism. Therefore, it contributed to our understanding of the evolutionary history of *P. spumarius* and the importance of dorsal colour variation in the process of adaptation. This study was the first attempt to identify and characterise regions of the genome associated with adaptive colour variation in this species. The many challenges faced to accomplish this task, in a non-model species with a large genome size, were evidenced.

In more detail, the present work allowed:

- To provide time estimates for the main demographic and evolutionary events occurred for European populations during the climate oscillations of the Quaternary period, and to analyse colonisation patterns for populations out of Eurasia. We found evidence of recent divergence and expansion events at less than 0.5 Ma ago indicating that Quaternary climate fluctuations were important in shaping the current phylogeographic pattern of the species. A pattern of recent gene-flow between Mediterranean peninsulas, a close relationship between Iberia and North Africa, as well as a probable British origin for the populations of the Azores and New Zealand, and the colonisation of North America from both western and northern Europe was found;

- To study the effect of the colour phenotypes in the survival and reproduction of *P. spumarius* and to explore the adaptive function of its dorsal colour polymorphism. *trilineatus* showed a higher longevity, a higher number of oviposition events, and a higher number of eggs laid, compared with the *marginellus* and *typicus*.

- To obtain hundreds of genomic markers (SNPs) through the RAD sequencing

technique and use them in a genome wide association study to identify regions related to dorsal colour variation. Sixty SNPs associated with the three colour morphs studied were detected, and should be further investigated and validated. None of these SNPs showed homology with candidate genes known to be responsible for colouration pattern in other insect species.

- To develop genomic and transcriptomic resources that can be used for further identification and characterisation of the genomic regions involved in *P. spumarius* colour variation. The proportion of the genome here obtained corresponded to only 25% of the total genome, estimated to be among the largest in insects (5.3 Gb). Although it has not been possible to obtain a complete genome sequence of *P. spumarius*, the partial genome and transcriptome drafts constitute important tools and a basis for future research in this species.

- To be aware of the difficulties of carrying out a candidate gene approach in a non-model species like *P. spumarius*, lacking information about the genetic basis of dorsal colour polymorphism and without a complete reference genome to properly isolate candidate genes.

6.3 Future directions

Understanding the genetic basis underlying adaptive traits is a central problem in evolutionary biology. However, its investigation in non-model species could be very challenging. In studies exploring the genetic basis of adaptive traits, is still difficult to detect not only the genes/loci that are involved in a trait, but also their relative contribution (major/minor effect), specially, when the trait has a polygenic basis as seems to be the case of colour variation in *P. spumarius*. With the rapid development of new methods of sequencing and analysis, to get information about the genetic architecture of adaptive traits is expected to be much easier in the future.

The present work contributed with new and important findings in *P. spumarius* but many questions remain unanswered and, future research demands more detailed phenotypic and genotypic data. Regarding the adaptive function of *P. spumarius'* colour polymorphism, it would be important to carry out experiments to test differential fitness among different colour phenotypes under different environmental conditions (different plant host species and (micro)climates). Many species around the world are currently suffering the effects of climate change, namely in their distribution range and survival. Therefore, to address the potential effects of climate change on the distribution and frequency of phenotypes in this species should also be considered.

To disentangle the genetic basis of this polymorphism, it is necessary to validate the

SNPs found to be associated with colour, to genotype these markers in other *P. spumarius* populations, and finally to characterise them. Additional RAD sequencing, covering a higher fraction of the genome, is needed to increase the number of SNPs, thus increasing the changes of identifying genomic regions linked to colour. A genome-wide scan using RAD markers should also be performed in other dorsal colour phenotypes.

To do RNA sequencing using different colour phenotypes at the phase of higher expression of colour genes, could also be useful in future studies. In this species, the 5thinstar is presumably the stage when many genes controlling adult colour pattern are highly expressed. Therefore, using mature 5thinstar nymphs for obtaining the transcripts would be prefered. However, as dorsal colour morphs are only discernible in the adult stage, this may not be possible. The alternative would be the use of the adult forms. Since this species small size can limit the quantity of RNA material to be recovered, using pools of individuals of the same colour morphs would be a effective way of overcoming this problem.

A complete reference genome for *P. spumarius* will help in further studies related to this species' polymorphism as well. The rapid evolution of sequencing technologies are expected to provide very soon the technical resources to be able to fully sequencing a genome of this size. Adding to this are the fast reducing costs. Moreover, the concomitant development of the bioinformatics resources will also allow not only to deal with the huge amounts of data, but also the fully assemblage, annotation and structural understanding of large genomes.

Although a still under-studied organism, this work showed that *P. spumarius* constitutes a potential good model system to study the process of adaptation, specially the evolutionary mechanisms maintaining polymorphisms.

6.4 References

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, **3**, e3376.

Berry AJ, Willmer PG (1986) Temperature and the colour polymorphism of *Philaenus spumarius* (Homoptera: Aphrophoridae). *Ecological Entomology*, **11**, 251–259.

Bihari P, Sipos B, Melika G *et al.* (2011) Western Palaearctic phylogeography of an inquiline oak gall wasp, *Synergus umbraculus*. *Biological Journal of the Linnean Society*, **102**, 750–764.

Brakefield PM, Willmer PG (1985) The basis of thermal melanism in the ladybird *Adalia*

bipunctata: Differences in reflectance and thermal properties between the morphs. *Heredity*, **54**, 9–14.

Drosopoulos S, Quartau J a (2002) The spittle bug *Philaenus tesselatus* Melichar, 1899 (Hemiptera, Auchenorrhyncha, Cercopidae) is a distinct species. *Zootaxa*, **68**, 1–8.

Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16196–16200.

Forsman A, Ahnesjö J, Caesar S, Karlsson M (2008) A model of ecological and evolutionary consequences of color polymorphism. *Ecology*, **89**, 34–40.

Franck P, Garnery L, Loiseau a. *et al.* (2001) Genetic diversity of the honeybee in Africa: Microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Gibbard PL, Head MJ, Walker MJC (2010) Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *Journal of Quaternary Science*, **25**, 96–102.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature*, **433**, 481–487.

Gray SM, McKinnon JS (2007) Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution*, **22**, 71–79.

Habel JC, Schmitt T, Müller P (2005) The fourth paradigm pattern of post-glacial range expansion of European terrestrial species: The phylogeography of the Marbled White butterfly (Satyrinae, Lepidoptera). *Journal of Biogeography*, **32**, 1489–1497.

Halkka O (1964) Geographical, spatial and temporal variability in the balanced polymorphism of *Philænus spumarius*. *Heredity*, **19**, 383–401.

Halkka O, Halkka L (1990) Population-Genetics Of the Polymorphic Meadow Spittlebug, *Philaenus spumarius* (L). *Evolutionary Biology*, **24**, 149–191.

Halkka O, Raatikainen M, Vasarainen A, Heinonen L (1967) Ecology and ecological genetics of *Philaenus spumarius* (L.) (Homoptera). *Ann. Zool. Fenn.*, 1–18.

Hallka O, Lallukka R (1969) The origin of balanced polymorphism in the spittlebugs (Philaenus, Homoptera). *Annales Zoologici Fennici*, **6**, 431–434.

Harper G, Whittaker JB (1976) The Role of Natural Enemies in the Colour Polymorphism of *Philaenus spumarius* (L.). *The Journal of Animal Ecology*, **45**, 91.

Hecht BC, Campbell NR, Holecek DE, Narum SR (2013) Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, **22**, 3061–3076.

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

Horn A, Roux-Morabito G, Lieutier F, Kerdelhue C (2006) Phylogeographic structure and past history of the circum-Mediterranean species *Tomicus destruens* Woll. (Coleoptera: Scolytinae). *Molecular Ecology*, **15**, 1603–1615.

Karban R, Strauss SY (2004) Physiological tolerance, climate change, and a northward range shift in the spittlebug, *Philaenus spumarius*. *Ecological Entomology*, **29**, 251–254.

Kawahara AY, Mignault AA, Regier JC, Kitching IJ, Mitter C (2009) Phylogeny and biogeography of hawkmoths (Lepidoptera: Sphingidae): Evidence from five nuclear genes. *PLoS ONE*, **4,** e5719**.**

Lis A, Maryańska-Nadachowska A, Lachowska-Cierlik D, Kajtoch U (2014) The Secondary Contact Zone of Phylogenetic Lineages of the *Philaenus spumarius* (Hemiptera: Aphrophoridae): An Example of Incomplete Allopatric Speciation. *Journal of Insect Science*, **14**, 227–227.

Lunt DH, Ibrahim KM, Hewitt GM (1998) mtDNA phylogeography and postglacial patterns of subdivision in the meadow grasshopper Chorthippus parallelus. *Heredity*, **80**, 633–641.

Macher J-N, Rozenberg A, Pauls SU *et al.* (2015) Assessing the phylogeographic history of the montane caddisfly *Thremma gallicum* using mitochondrial and restriction-site-associated DNA (RAD) markers. *Ecology and Evolution*, **5**, 648–662.

Maryańska-Nadachowska A, Drosopoulos S, Lachowska D, Kajtoch Ł, Kuznetsova VG (2010) Molecular phylogeny of the Mediterranean species of *Philaenus* (Hemiptera: Auchenorrhyncha: Aphrophoridae) using mitochondrial and nuclear DNA sequences. *Systematic Entomology*, **35**, 318–328.

Maryańska-Nadachowska A, Kajtoch Ł, Lachowska D (2012) Genetic diversity of *Philaenus spumarius* and *P. tesselatus* (Hemiptera, Aphrophoridae): implications for evolution and

taxonomy. *Systematic Entomology*, **37**, 55–64.

Maryańska-Nadachowskaa A, Sanaieb E, Kajtocha Ł (2015) High genetic diversity in southwest Asian populations of *Philaenus spumarius* (Hemiptera: Auchenorrhyncha). *Zoology in the Middle East*, **61**, 264–272.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

Munday PL, Eyre PJ, Jones GP (2003) Ecological mechanisms for coexistence of colour polymorphism in a coral-reef fish: An experimental evaluation. *Oecologia*, **137**, 519–526.

Owen DF, Wiegert RG (1962) Balanced Polymorphism in the Meadow Spittlebug, *Philaenus spumarius*. *The American Naturalist*, **96**, pp. 353–359.

Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC genomics*, **11**, 180.

Quartau JA, Borges PA V (1997) On the colour polymorphism of *Philaenus spumarius* (L .) in Portugal. *Miscel-lánia Zoológica*, **2**, 19–30.

Rhamhalinghan M (1990) Variations in the internal temperatures of melanics and typicals of *Coccinella septempunctata* L. (Coleoptera, Coccinellidae). *Journal of Advanced Zoology*, **10**, 31–36.

Rodrigues ASB (2010) Estudo dos padrões filogenéticos e filogeográficos em *Philaenus spumarius* (Hemiptera, Aphrophoridae) e espécies próximas.

Rodriguez-Sanchez F, Perez-Barrales R, Ojeda F, Vargas P, Arroyo J (2008) The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews*, **27**, 2100–2117.

Roulin A (2014) Melanin-based colour polymorphism responding to climate change. *Global Change Biology*, **20**, 3344–3350.

Schmitt T, Röber S, Seitz A (2005) Is the last glaciation the only relevant event for the present genetic population structure of the meadow brown butterfly *Maniola jurtina* (Lepidoptera: Nymphalidae)? *Biological Journal of the Linnean Society*, **85**, 419–431.

Seabra SG, Pina-Martins F, Marabuto E *et al.* (2010) Molecular phylogeny and DNA barcoding

in the meadow-spittlebug *Philaenus spumarius* (Hemiptera, Cercopidae) and its related species. *Molecular Phylogenetics and Evolution*, **56**, 462–467.

True JR, Yeh SD, Hovemann BT *et al.* (2005) *Drosophila* tan encodes a novel hydrolase required in pigmentation and vision. *PLoS genetics*, **1**.

Wallbank RWR, Baxter SW, Pardo-Diaz C *et al.* (2016) Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. *PLoS biology*, **14**, e1002353.

Wittkopp PJ, Beldade P (2009) Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy. *Seminars in Cell and Developmental Biology*, **20**, 65–71.

Wittkopp PJ, True JR, Carroll SB (2002a) Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development*, **129**, 1849–58.

Wittkopp PJ, Vaccaro K, Carroll SB (2002b) Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Current Biology*, **12**, 1547–1556.

Yurtsever S (2000) On the polymorphic meadow spittlebug, *Philaenus spumarius* (L.) (Homoptera: Cercopidae). *Turkish Journal of Zoology*, **24**, 447–459.