UNIVERSITA' DEGLI STUDI DI PADOVA

SCUOLA DI DOTTORATO DI RICERCA IN

SCIENZE DELLE PRODUZIONI VEGETALI

INDIRIZZO PROTEZIONE DELLE COLTURE - CICLO XXII

Dipartimento Di Agronomia Ambientale e Produzioni Vegetali

Genetics and genomics of pine processionary moths and their parasitoids

Direttore della Scuola : Ch.mo Prof. Andrea Battisti

Supervisore : Ch.mo Prof. Andrea Battisti

Dottorando : Mauro Simonato

DATA CONSEGNA TESI

01 febbraio 2010

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

February 1st, 2010 Mauro Simonato

A copy of the thesis will be available at *http://paduaresearch.cab.unipd.it/*

Dichiarazione

Con la presente affermo che questa tesi è frutto del mio lavoro e che, per quanto io ne sia a conoscenza, non contiene materiale precedentemente pubblicato o scritto da un'altra persona né materiale che è stato utilizzato per l'ottenimento di qualunque altro titolo o diploma dell'università o altro istituto di apprendimento, a eccezione del caso in cui ciò venga riconosciuto nel testo.

1 febbraio 2010 Mauro Simonato

Una copia della tesi sarà disponibile presso *http://paduaresearch.cab.unipd.it/*

Table of contents

Table of contents **...... 5**

Riassunto **...... 9**

Summary **...... 11**

Chapter 1 - Introduction 13

Objective and contents of the thesis 21

References **...... 22**

Chapter 2 - Phylogeography of the pine processionary moth *Thaumetopoea wilkinsoni* in the Near East

Introduction **...... 31**

Materials and methods **...... 32**

Sampling and DNA protocols 32

Data analysis 33

Results **...... 34**

Mitochondrial DNA phylogeography 34

Comparison between mitochondrial and nuclear markers 35

Discussion **...... 37**

Mitochondrial phylogeographic patterns and female colonization routes 37

Unexpected patterns of nuclear diversity, and sex-biased gene flow 38

References **...... 39**

Chapter 3 - Quaternary history and contemporary patterns in a currently expanding species

Background **...... 46**

Results **...... 48**

Phylogenetic inference and node datation 48

Haplotype distribution and haplotype network 51

Discussion **...... 52**

Overall phylogenetic patterns around the Mediterranean Basin 52
Phylogeographical patterns and within-clade structures 53
Evolution of insular populations 54
Contemporary patterns in a historical context 54
***Conclusion* 55**
***Methods* 55**
Moth sampling 55
DNA protocols 55
Data analyses 56
***References* 57**
***Additional material* 59**
***The role of topography in structuring the demographic history of the pine processionary moth, Thaumetopoea pityocampa (Lepidoptera: Notodontidae)* 66**
Abstract 66

Chapter 4 - The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae)
***Background* 72**
***Results and discussion* 72**
Genome organization, structure and composition 72
Protein-coding genes 74
Transfer and ribosomal RNA genes 79
Non coding regions 81
***Conclusion* 81**
***Methods* 83**
Sample origin and DNA extraction 83
PCR amplification and sequencing of *Ochrogaster lunifer* mtDNA 83
Sequence assembly and annotation 83
Genomic analysis 83
Abbreviations 83

References *84*

Chapter 5 - Testing host plant associated differentiation on two parthenogenetic parasitoid species feeding on the same insect host in a forest system

Introduction *89*

Materials and methods *91*

Sample collection 91

DNA extraction and fingerprint analysis 91

Assays for intracellular symbionts 92

Data analysis 93

Results *94*

Assays for intracellular symbionts 94

Genetic data analysis 95

Discussion *96*

Tables and figures *100*

References *109*

Conclusion **117**

Ecological aspects and applied results linked to population genetic analysis 117

Main phylogeographic events detected by mitochondrial markers with some insights into taxonomy 119

References *120*

Acknowledgments **121**

Riassunto

La processionaria del pino (*Thaumetopoea* spp.), un importante defogliatore dei pini in tutta l'area del Mediterraneo, ha mostrato, nel corso degli ultimi decenni, un'espansione del suo areale in risposta sia al cambiamento climatico che a fattori antropici. È quindi importante delineare le modalità con cui questa specie riesce a colonizzare nuove aree e per far questo i marcatori molecolari sembrano essere gli strumenti più utili. Nell'introduzione sono presentate alcuni dei marcatori molecolari usati negli ultimi anni per studiare problemi ecologici ed evolutivi in relazione agli insetti. L'obiettivo principale del mio lavoro è consistito nell'analizzare la variabilità genetica della processionaria del pino nel suo attuale areale e nel tentare di ricostruire la storia della sua colonizzazione recente e passata. In secondo luogo, un ulteriore obiettivo è stato quello di caratterizzare la struttura genetica di alcuni parassitoidi della processionaria del pino per comprendere meglio quali sono i fattori coinvolti nel loro differenziamento quale può essere la ricerca dell'ospite su diverse specie di pino.

Nel primo lavoro ho esteso a popolazioni del Vicino Oriente uno studio già iniziato sulla genetica di popolazione della processionaria del pino. Lo scopo principale di questo lavoro era quello di capire l'origine delle popolazioni che attaccano le piantagioni di pino in Israele e in secondo luogo di caratterizzare geneticamente tutte le popolazioni presenti nella regione. Questo studio ha inoltre mostrato l'utilità dell'impiego di marcatori mitocondriali e nucleari per il diverso tipo di informazione che possono produrre. In questo caso è stato possibile individuare una dispersione in relazione al sesso degli individui, processo che potrebbe essere importante per il mantenimento della variabilità genetica nelle aree di espansione.

Nel secondo lavoro, è stata delineata la struttura genetica della processionaria del pino in tutto il suo areale. Attraverso i marcatori mitocondriali utilizzati nello studio precedente è stato possibile definire per questa specie i principali eventi occorsi nel passato, identificando così i rifugi glaciali e i principali eventi di separazione tra le diverse popolazioni. Nel Nord Africa, è stato individuato in questo modo un nuovo clade geneticamente ben definito, analizzando popolazioni precedentemente considerate, su

base morfologica, appartenenti ad una sottospecie di una della due specie di processionaria già descritte.

Nel terzo lavoro, è stato descritto l'intero menoma mitocondriale di ochrogaster lunifer, una specie Australina appartenete alla stessa sottofamiglia della processionaria del pino. Questo studio, oltre a rappresentare il primo passo per un chiarimento della tassonomia di questa famiglia di Lepidotteri, ha prodotto informazioni utili riguardo ai marcatori che possono essere utilizzati negli studi di genetica di popolazione dei Lepidotteri.

Nel quarto lavoro, ho preso in considerazione la genetica di popolazione di due parassitoidi oofagi della processionaria del pino: lo specialista *Baryscapus servadeii* e il generalista *Ooencyrtus pityocampae*. L'obiettivo principale di questo studio era di testare la presenza di un'associazione con la pianta ospite in questi due parassitoidi e di comparare inoltre la struttura genetica di uno specialista con quella di un generalista. Nelle popolazioni del parassitoide specialista analizzate sembra essere presente un genotipo per lo più associato con una delle piante ospiti della processionaria. Inoltre, a differenza del generalista, il parassitoide specialista non presenta una variabilità genetica strutturata spazialmente. Questo potrebbe essere messo in relazione alla più alta mobilità dello specialista e quindi alla sua strategia per individuare l'ospite. Tali nuove informazioni su questi parassitoidi possono rivelarsi utili nel predire il loro comportamento nelle aree di espansione.

Nel complesso, i quattro contributi presentati qui forniscono suggerimenti per il controllo di questo insetto infestante su larga scala e una maggiore conoscenza della storia evolutiva del gruppo, includendo inoltre delle previsioni sul potenziale di adattamento di queste specie in relazione ai cambiamenti climatici in corso.

Summary

The pine processionary moth (*Thaumetopoea* spp.), an important defoliator of *Pinus* spp. in the Mediterranean area, is recently expanding its range in response to climate change and anthropogenic factors. Therefore it's important to outline the way in which this pest can colonize new areas, and to do this molecular markers seem to be the most suitable tools. In the introduction I present some of the molecular markers used in the last years to study ecological and evolutionary problems related to insects. The main aim of my work was to analyze the genetic variability of pine processionary moth in its present range and so to try to reconstruct the recent and past colonization history of this pest. Secondly, another goal was to characterize the genetic structure of pine processionary moth parasitoids to better understand the factors involved in their differentiation such as in tracking their host on different pine species.

In the first study I extended a work already begun on the population genetic of pine processionary moth to the populations of the Near East. The main issue of this study was to understand the origin of population attacking the Israel pine plantations and in second place to genetically characterize all the populations in the range. Moreover, this study shows the utility of the use of both nuclear and mitochondrial markers for the different information they can yield. In this case they permitted to track a gender-related dispersal, which could be important to maintain genetic variability in expansion areas.

In the second study, I contributed to outline the genetic structure of pine processionary moth in the whole range. Through the mitochondrial markers used in the previous studies it was possible to define the main events occurred to this species in the past, identifying glacial refugia and the main splitting events among the different lineages. In northern Africa, a new unexpected genetic clade was found analyzing populations that were previously considered, on a morphological base, to belong to a subspecies of the two already described pine processionary moth species.

In the third study, I contributed to the sequencing of the entire mitochondrial genome of *Ochrogaster lunifer*, an Australian species belonging to the same subfamily of the pine processionary moth. This study, besides to represent the first step to have

insights into the taxonomy of the family, gave useful information about the best markers to be used in population genetic studies on Lepidoptera.

In the fourth study, I dealt with the population genetic of two egg parasitoid species of the pine processionary moth: the specialist *Baryscapus servadeii* and the generalist *Ooencyrtus pityocampae*. The main goal of this study was to test the presence of a host plant association in these two parasitoids and to compare the genetic patterns of a specialist and a generalist. In the specialist parasitoid populations analyzed it seems to be present a genotype mostly associated with the host plant. Moreover, unlike the generalist, the specialist doesn't show to have a genetic pattern spatially structured.

This could be related to a higher mobility of the specialist, and hence to its strategy in finding hosts. These findings provide useful information to predict the behavior of parasitoids in expanding areas. On the whole, the four contributions provide suggestions for the range wide management of the pest, and insights into the evolutionary history of the group, including projections on the potential for adaptation to ongoing climate change.

Chapter 1

Introduction

Populations of almost all species are genetically structured across their range (Avise *et al.* 1987). These genetic patterns are influenced by ecological and evolutionary factors (e.g. migration, random genetic drift, natural selection) operating over a wide variety of spatial and temporal scales.

Molecular genetic markers have become a powerful tool for population studies in the last two decades. Development of new techniques such as polymerase chain reaction (PCR) and sequencing have extended the availability of molecular polymorphisms at affordable costs, thus providing wide datasets useful for answering questions about behavior, ecology and phylogeny of organisms. The study of polymorphisms of nucleic acids can be carried on without any previous knowledge about species and their habitats and with no need of classical genetic studies (e.g. controlled crosses and checking of mutants). Moreover, differences among individuals or species can be easily quantified through molecular information, avoiding problems related to the use of taxonomical traits that are often affected by environmental factors. Different molecular markers are now available (for insects see Behura 2006); their application is related to both geographical range of sampling and temporal scale of historical events involved (Avise 2004). Each type of molecular marker has its own characteristic level of genetic resolution that is appropriate for various methods and purposes depending on the corresponding resolution needed.

Mitochondrial DNA genes are particularly profitable in studies on conspecific populations and closely related species as animal mitochondrial DNA evolves rapidly at sequence level (Brown *et al.* 1979, Wilson *et al.* 1985) and is maternally inherited without recombination. Studies on mtDNA have defined thus an empirical and conceptual bridge between systematic and population genetics, a rather new discipline known as phylogeography (Avise *et al.* 1987). In the recent years the phylogeographical approach has been widely used in the study of populations of forest pests. Mitochondrial DNA markers have been used to track the postglacial colonizations and expansion routes of several bark beetles (Stauffer *et al.* 1999, Ritzerow *et al.* 2004, Sallé *et al.* 2007, Horn *et al.* 2006, Maroja *et al.* 2007, Mock *et al.* 2007); they have been used also to study the history of invasive species such as hemlock woolly adelgid (Havill *et al.* 2006) and asian longhorned beetle (Carter *et al.* 2009), suggesting the likely sources of introduction in the new areas.

Moreover, mtDNA markers have been useful in defining the taxonomic status of

several cryptic species (Sperling *et al.* 1999, Kerdelhue *et al.* 2002, Duan *et al.* 2004, Cognato *et al.* 2005), providing insight on the effect of both host plant and/or geographic location in structuring the pest populations (Kerdelhue *et al.* 2002, Cognato *et al.* 2005). Mitochondrial genetic markers have also been employed to evaluate gene flow among populations of phytophagous insects (Salvato *et al.* 2002, Schroeder & Degen 2008), thus yielding long-term, indirect dispersal estimates that can be helpful in understanding and predicting dynamics and consequences of pest expansions.

At higher phylogenetic levels standard mitochondrial genes (e.g cox1, cox2, cytB, 16S, 12S) cannot often resolve relationships among taxa. In these cases whole mithocondrial genomes are often preferred as they provide a better resolution for deep relationships at intra-family and intra-order level. Most animal mitochondrial genomes are of very similar size (about 15,000 bp in insects) with a similar set of genes (37 genes). In addition to the nucleotide data, other phylogenetically useful information can be obtained from mitochondrial genomes such as gene rearrangements (Boore & Brown 1998), gene insertion or deletion events (Rokas & Holland 2000), and genic or intergenic length variability (Schneider & Ebert 2004). The phylogenetic utility of mitochondrial genomes has been carefully studied in the past few years, especially for insects and related groups (Cameron *et al.* 2004, Cameron & Whiting 2007, Carapelli *et al.* 2007, Kjer & Honeycutt 2007). In particular, these genomes have been studied for a variety of purposes including divergences between sibling species (Yukuhiro *et al.* 2002), identifying gene variability between congeneric species (Coates *et al.* 2005), to facilitate population level studies (Kim *et al.* 2006) and to investigate relationships within the order Lepidoptera (Lee *et al.* 2006). A major outcome of the accumulation of insect mitochondrial genome data has been the capacity to investigate the utility of individual genes or regions commonly employed in phylogenetics, phylogeography, population genetics and molecular diagnostics and to identify novel genes which could be useful for future studies (Cameron & Whiting 2007, Nardi *et al.* 2003, Nardi *et al.* 2005, Simon *et al.* 2006).

Although gene genealogies based on mtDNA sequence variation have yielded valuable insights on population structure in several systems (Avise 2004), mtDNA often bears insufficient variation to reflect relatively recent evolution and to detect ongoing gene flow. Detecting individual movement among populations requires methods that use more

polymorphic markers. In this respect, fragment analysis of microsatellite regions (simple sequence repeat, SSR) and amplified length polymorphism (AFLP) analysis are effective in diagnosing parentage and studying the genetic structure of populations. As they can generate a large number of repeatable genomic polymorphic markers without the necessity for any prior research and development, AFLP markers (Vos *et al.* 1995) are often more attractive than microsatellites for studying genetic diversity and population structure. Moreover, since AFLP markers can yield a high number of loci scattered all over the genome, their analysis allow to disentangle locus-specific effects (selection, mutation, recombination, and assortative mating) that should affect one or a few genes at a time, from genome-wide effects like genetic drift, migration and inbreeding, which should affect all parts of a genome in the same way (Beaumont & Nichols 1996, Luikart *et al.* 2003). Two issues in the use of this method are the loss of information given by the dominant nature of AFLP markers (the presence of a band in a locus can indicate either the homozygous condition or the heterozygous condition) and size homoplasy (i.e. bands of the same length are not homologous and thus representing two or more different AFLP loci), that could be of particular concern in studies of genetic diversity and phylogenetic reconstructions (Vekemans *et al.* 2002). Within insects AFLP analysis has been used successfully to study closely related populations at fine taxonomic levels (Yan *et al.* 1999, Reineke *et al.* 1999, Parsons & Shaw 2001), addressing questions about dispersal and gene flow (Salvato *et al.* 2002, Grapputo *et al.* 2005, Conord *et al.* 2006, Timm *et al.* 2006, Ahern *et al.* 2009), insecticide resistence (Kazachkova *et al.* 2007, Thaler *et al.* 2008) and introgressive hybridization (Gompert *et al.* 2006, Gompert *et al.* 2008). Moreover, AFLPs have been used to test host specialization in both phytophagous (Althoff *et al.* 2006, Scheffer & Hawthorne 2007) and parasitoid insects (Kolaczan *et al.* 2009), in some cases attempting also to identify loci linked to genes undergoing selection for the host (Nosil *et al.* 2008, Egan *et al.* 2008, Manel *et al.* 2009).

In this thesis I outlined some aspects of population genetics and phylogeography of the processionary moths *Thaumetopoea* spp. (Lepidoptera, Notodontidae), that comprise 10 species distributed in the Mediterranean region and Europe (Tab.1). The larvae have gregarious behavior in all stages of their development and they produce urticating hairs that can cause an allergic reaction in mammals. Processionary moths can feed on various host

Tribe Genus Species subspecies	range	host plant
Thaumetopoeinae	Palaearctic, part of Oriental and Ethiopian region	
1 genus (12 species)		
Thaumetopoea		
processionea (Linnaeus 1758)	Europe	Quercus
pseudosolitaria (Daniel 1951)	Balkans, Near East	Quercus
solitaria (Freyer 1838)	Balkans, Near East	Pistacia
pityocampa (Denis & Schiffermüller 1775)	S Europe, NW Africa	Pinus Cedrus
orana (Staudinger 1901)	Morocco, Algeria	Pinus
ceballosi (Agenjo 1941)	Anatolia	Pinus
wilkinsoni (Tams 1925)	Cyprus, Near East	Pinus Cedrus
pinivora (Treitschke 1834)	Europe	Pinus
bonjeani (Powell 1922)	N Africa	Cedrus
herculeana (Rambur 1840)	Iberian pen., N Africa	Cistus Erodium Helianthemum
judea (Bang-Haas 1910)	Palestina	
jordana (Staudinger 1894)	Jordany, Israel	Rhus
libanotica (Kiriakoff & Talhouk 1975)	Lebanon	Cedrus
ispartaensis (Doganlar & Avci 2001)	Turkey	Cedrus
Anaphinae	Ethiopian-Malgascian region	
7 genera (52 species)		
Epicominae	Australian region	
8 genera (29 species) (Nielsen *et al.* 1996)		

Table 1. Taxonomy of pine processionary moths. For the genus *Thaumetopoea* only the palaearctic species are considered (from Lafontaine and Fibiger (2006), Kiriakoff (1970), and recent updates).

plants ranging from broadleaved species (e.g. *Pistacia, Quercus, Cistus*) to conifer trees (*Pinus* and *Cedrus*). The species feeding on conifers can be subdivided in two main groups, according to the seasonal period of defoliation of larvae: the summer species (*T. bonjeani, T. ispartaensis, T. libanotica, T. pinivora*) and the winter species (*T. pityocampa* and *T. wilkinsoni*) (Demolin & Frerot 1993). Further morphological variability within the range has supported the identification of subspecies within *T. processionea, T. pityocampa, T. herculeana* (Agenjo 1941, Kiriakoff 1970). To date relationships based on morphological data both inside the genus (see Freina & Witt (1987)) and among the upper taxa (e.g. subfamily Thaumetopoeinae, see Kiriakoff 1970, Miller 1991, Lafontaine & Fibiger 2006) are still to be resolved.

Most of my work has concerned *Thaumetopoea pityocampa* (Denis & Schiffermüller) and *Thaumetopoea wilkinsoni* (Tams), commonly defined as pine processionary moth (Plate 1). They were considered synonyms for a long time (Demolin & Frerot 1993, Demolin 1988), but the first genetic study on this taxon (Salvato *et al.* 2002a) provided evidence of species separation. They are both economically important defoliator of pines in southern Europe and Near East. In the last decades they are expanding their range for both the intense cultivation of conifer trees in exotic areas (Masutti & Battisti 1990) and the increasing winter temperatures (Battisti *et al.* 2005), associated with the climate change effect (Solomon *et al.* 2007).

There is evidence that climatic variability can change interactions between phytophagous insects and their parasitoids, impairing the ability of parasitoids to track host populations (Stireman *et al.* 2005, Menendez *et al.* 2008). Given the important role of parasitoids in regulating insect herbivore populations in natural and managed systems, an increase in the frequency and intensity of herbivore outbreaks as climates become more variable could be expected. Thus the study of the structure of parasitoids populations in areas experiencing climatic changes could help to understand how they respond to global climate changes. In this perspective, I therefore considered the genetic population study of two main egg parasitoids of *T. pityocampa*: the specialist *Baryscapus servadeii* Domenichini (Hymenoptera Eulophidae) and the generalist *Ooencyrtus pityocampae* Mercet (Hymenoptera Encyrtidae) that are found throughout the range of the pine processionary moth.

Plate 1

Ovipositing female on *P. nigra* (photo D. Zovi)

Winter nest on *P. nigra*

Third instar larvae feeding (photo D. Zovi)

Pupating larvae on soil

Objectives and content of the thesis

In this thesis, using both mitochondrial and nuclear markers, I developed further the population genetic study begun by (Salvato *et al.* 2002a) analyzing *T. wilkinsoni* populations from the Near East (Chapter 2), and trying to track the routes of the recently expanding populations in southern Israel and northern Turkey. Subsequently, I contributed to extend this analysis to the whole range of both species of pine processionary moths, comprising all the Mediterranean basin and southern Europe (Chapter 3). In this way it was possible to define all the genetic clades present in the area, and thus the colonization history and the occurrence of glacial refugia, as well as the origin of recently established populations.

To have preliminary sequence information for a phylogenetic study of the species inside the genus *Thaumetopoea* (work in progress not included in the present thesis), and to extend the taxonomic sampling of mitochondrial genomes inside Lepidoptera, I contributed to the sequencing of the whole mitochondrial genome of a member of Thaumetopoeinae, *Ochrogaster lunifer* (Lepidoptera, Notodontidae) the first complete sequence for the Superfamily Noctuoidea (Chapter 4). Besides to describe this genome, another aim of this study was to do a comparative genomics analysis to identify potential novel markers for phylogenetic studies inside Lepidoptera.

In the fourth study (Chapter 5) I analyzed, through the use of AFLP markers, the genetic structure of the two main egg parasitoids of *T. pityocampa* at local scale (north eastern Italy). The analysis had two objectives: firstly to assess the effect of host-plant species on the differentiation of parasitoid populations attacking the same insect (*T. pityocampa*) on two different host-plant species (*Pinus sylvestris* and *Pinus nigra*); secondly to compare the population structure of a specialist (*B.servadeii*) and a generalist (*O. pityocampae*) parasitoids that share the same host insect across a large geographic area.

References

Agenjo R. (1941) Monographia de la familia Thaumetopoeidae. *Eos Madrid,* **17**: 69-130.

Ahern R.G., Hawthorne D.J., Raupp M.J. (2009) Founder effects and phenotypic variation in *Adelges cooleyi*, an insect pest introduced to the eastern United States. *Biological Invasions,* **11**, 4: 959-971.

Althoff D.M., Segraves K.A., Leebens-Mack J., Pellmyr O. (2006) Patterns of speciation in the yucca moths: parallel species radiations within the *Tegeticula yuccasella* species complex. *Systematic Biology,* **55**, 3: 398-410.

Avise J. (2004) *Molecular Markers, Natural History and Evolution*, 2nd edn. Sinauer Associates, Sunderland, USA.

Avise J., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., Saunders N.C. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics,* **18**, 1: 489-522.

Battisti A., Stastny M., Netherer S., Robinet C., Schopf A., Roques A., Larsson S. (2005) Expansion of geographic range in the pine processionary moth caused by increased winter temperatures. *Ecological Applications,* **15**, 6: 2084-2096.

Beaumont M.A., Nichols R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences,* **263**, 1377: 1619-1626.

Behura S.K. (2006) Molecular marker systems in insects: current trends and future avenues. *Molecular Ecology,* **15**, 11: 3087-3114.

Boore J.L., Brown W.M. (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics & Development,* **8**, 6: 668-674.

Brown W, George M, Wilson A (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences,* **76**, 4: 1967-1971.

Cameron S.L., Miller K.B., D'Haese C.A., Whiting M.F., Barker S.C. (2004) Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). *Cladistics,* **20**, 6: 534-557.

Cameron S.L., Whiting M.F. (2007) Mitochondrial genomic comparisons of the subterranean termites from the genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). *Genome,* **50**, 2:

188-202.

Carapelli A., Liò P., Nardi F., van der Wath E., Frati F. (2007) Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evolutionary Biology,* **7**, Suppl 2, S8.

Carter M.E., Smith M.T., Harrison R.G. (2009) Patterns of Genetic Variation Among Populations of the Asian Longhorned Beetle (Coleoptera: Cerambycidae) in China and Korea. *Annals of the Entomological Society of America,* **102**, 5: 895-905.

Coates B.S., Sumerford D.V., Hellmich R.L., Lewis L.C. (2005) Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. *International Journal of Biological Sciences,* **1**, 1: 13-18.

Cognato A.I., Gillette N.E., Bolaños R.C., Sperling F.A.H. (2005) Mitochondrial phylogeny of pine cone beetles (Scolytinae, Conophthorus) and their affiliation with geographic area and host. *Molecular Phylogenetics and Evolution,* **36**, 3: 494-508.

Conord C., Lempérière G., Taberlet P., Després L. (2006) Genetic structure of the forest pest *Hylobius abietis* on conifer plantations at different spatial scales in Europe. *Heredity,* **97**, 1: 46-55.

Demolin G., Frerot B. (1993) Sex pheromone of the processionary moth and biosystematic consideration within the genus *Thaumetopoea* (Thaumetopoeidae, Thaumetopoeinae). *Boll Zool Agric Bach,* **25**: 33-40.

Demolin J. (1988) La processionnaire du Cedre: *Thaumetopoea bonjeani* Powel, Rapport Scientific et Rapport Iconographique–Intensification de la Protection Phytosanitaire des Forets, Algeria 1986-87. *FAO Report,* .

Duan Y., Kerdelhue C., Ye H., Lieutier F. (2004) Genetic study of the forest pest *Tomicus piniperda* (Col., Scolytinae) in Yunnan province (China) compared to Europe: new insights for the systematics and evolution of the genus *Tomicus*. *Heredity,* **93**, 5: 416-422.

Egan S.P., Nosil P., Funk D.J., Pfennig D. (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution,* **62**, 5: 1162-1181.

Freina J., Witt T. (1987) Die Bombyces und Sphinges der Westpalaearktis (Insecta, Lepidoptera), Bd. l. *Edition Forschung und Wissenschaft Verlag-GmbH, München,* .

Gompert Z., Fordyce J.A., Forister M.L., Shapiro A.M., Nice C.C. (2006) Homoploid hybrid

speciation in an extreme habitat. *Science,* **314**, 5807: 1923-1925.

Gompert Z., Forister M.L., Fordyce J.A., Nice C.C. (2008) Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology,* **17**, 24: 5231-5244.

Grapputo A., Boman S., Lindstrom L., Lyytinen A., Mappes J. (2005) The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. *Molecular Ecology,* **14**, 14: 4207-4219.

Havill N.P., Montgomery M.E., Yu G., Shiyake S., Caccone A. (2006) Mitochondrial DNA from hemlock woolly adelgid (Hemiptera: Adelgidae) suggests cryptic speciation and pinpoints the source of the introduction to eastern North America. *Annals of the Entomological Society of America,* **99**, 2: 195-203.

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**, 6: 1603-1615.

Kazachkova N., Meijer J., Ekbom B. (2007) Genetic diversity in pollen beetles (*Meligethes aeneus*) in Sweden: role of spatial, temporal and insecticide resistance factors. *Agricultural and Forest Entomology,* **9**, 4: 259-269.

Kerdelhue C., Roux-Morabito G., Forichon J., Chambon J.M., Robert A., Lieutier F. (2002) Population genetic structure of *Tomicus piniperda* L. (Curculionidae: Scolytinae) on different pine species and validation of *T. destruens* (Woll.). *Molecular Ecology,* **11**, 3: 483-494.

Kim I., Lee E.M., Seol K.Y., Yun E.Y., Lee Y.B., Hwang J.S., Jin B.R. (2006) The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). *Insect Molecular Biology,* **15**, 2: 217-226.

Kiriakoff S. (1970) Lepidoptera familia Thaumetopoeidae. *Genera Insectorum de P.Wystman Belgique,* **219**.

Kjer K.M., Honeycutt R.L. (2007) Site specific rates of mitochondrial genomes and the phylogeny of eutheria. *BMC Evolutionary Biology,* **7**: 8.

Kolaczan C.R., Heard S.B., Segraves K.A., Althoff D.M., Nason J.D. (2009) Spatial and genetic structure of host-associated differentiation in the parasitoid *Copidosoma gelechiae*. *Journal of Evolutionary Biology,* **22**, 6: 1275-1283.

Lafontaine J., Fibiger M. (2006) Revised higher phylogeny of the Noctuoidea (Lepidoptera). *Canadian Entomologist,* **138**: 610-635.

Lee E.S., Shin K.S., Kim M.S., Park H., Cho S., Kim C.B. (2006) The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene,* **373**: 52-57.

Luikart G., England P.R., Tallmon D., Jordan S., Taberlet P. (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics,* **4**, 12: 981-994.

Manel S., Conord C., Despres L. (2009) Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. *BMC Evolutionary Biology,* **9**: 1, 288.

Maroja L.S., Bogdanowicz S.M., Wallin K.F., Raffa K.F., Harrison R.G. (2007) Phylogeography of spruce beetles (*Dendroctonus rufipennis* kirby)(Curculionidae: Scolytinae) in North America. *Molecular Ecology,* **16**, 12: 2560-2573.

Masutti L., Battisti A. (1990) *Thaumetopoea pityocampa* (Den. & Schiff.) in Italy. *Journal of Applied Entomology,* **110**: 229-234.

Menendez R., Gonzalez-Megias A., Lewis O.T., Shaw M.R., Thomas C.D. (2008) Escape from natural enemies during climate-driven range expansion: a case study. *Ecological Entomology,* **33**, 3: 413-421.

Miller J. (1991) Cladistics and classification of the Notodontidae (Lepidoptera: Noctuoidea) based on larval and adult morphology. *Bulletin of the American Museum of Natural History (USA)*

Mock K.E., Bentz B.J., O'Neill E.M., Chong J.P., Orwin J., Pfrender M.E. (2007) Landscape-scale genetic variation in a forest outbreak species, the mountain pine beetle (*Dendroctonus ponderosae*). *Molecular Ecology,* **16**, 3: 553-568.

Nardi F., Carapelli A., Dallai R., Frati F. (2003) The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. *Insect Molecular Biology,* **12**, 6: 605-611.

Nardi F., Carapelli A., Dallai R., Roderick G.K., Frati F. (2005) Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). *Molecular Ecology,* **14**, 9: 2729-2738.

Nielsen E.S., Edwards E.D., Rangsi T.V. (1996) *Checklist of the Lepidoptera of Australia*,

CSIRO Australia, Collingwood, Vic.

Nosil P., Egan S.P., Funk D.J., Hoekstra H. (2008) Heterogeneous Genomic Differentiation Between Walking-Stick Ecotypes: "Isolation by Adaptation" and Multiple Roles for Divergent Selection. *Evolution,* **62**, 2: 316-336.

Parsons Y., Shaw K. (2001) Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Molecular Ecology,* **10**, 7: 1765-1772.

Reineke A., Karlovsky P., Zebitz C. (1999) Amplified fragment length polymorphism analysis of different geographic populations of the gypsy moth, *Lymantria dispar* (Lepidoptera: Lymantriidae). *Bulletin of Entomological Research,* **89**: 79-88.

Ritzerow S., Konrad H., Stauffer C. (2004) Phylogeography of the Eurasian pine shoot beetle *Tomicus piniperda* (Coleoptera: Scolytidae). *European Journal of Entomology,* **101**, 1: 13-20.

Rokas A., Holland P.W.H. (2000) Rare genomic changes as a tool for phylogenetics. *Trends in Ecology & Evolution,* **15**, 11: 454-459.

Sallé A., Arthofer W., Lieutier F., Stauffer C., Kerdelhué C. (2007) Phylogeography of a host-specific insect: genetic structure of *Ips typographus* in Europe does not reflect past fragmentation of its host. *Biological Journal of the Linnean Society,* **90**, 2: 239-246.

Salvato P., Battisti A., Concato S., Masutti L., Patarnello T., Zane L. (2002) Genetic differentiation in the winter pine processionary moth (*Thaumetopoea pityocampa-wilkinsoni* complex), inferred by AFLP and mitochondrial DNA markers. *Molecular Ecology,* **11**: 2435-2444.

Scheffer S.J., Hawthorne D.J. (2007) Molecular Evidence of Host-Associated Genetic Divergence in the Holly Leafminer *Phytomyza glabricola* (Diptera: Agromyzidae): Apparent Discordance among Marker Systems. *Molecular Ecology,* **16**: 2627-2637.

Schneider A., Ebert D. (2004) Covariation of mitochondrial genome size with gene lengths: evidence for gene length reduction during mitochondrial evolution. *Journal of Molecular Evolution,* **59**, 1: 90-96.

Schroeder H., Degen B. (2008) Genetic structure of the green oak leaf roller (*Tortrix viridana* L.) and one of its hosts, *Quercus robur* L. *Forest Ecology and Management,* **256**, 6: 1270-1279.

Simon C., Buckley T.R., Frati F., Stewart J.B., Beckenbach A.T. (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.

Solomon S., Qin D., Manning M., Alley R.B., Berntsen T., Bindoff N.L., Chen Z., Chidthaisong A., Gregory J.M., Hegerl G.C., Heimann M., Hewitson B., Hoskins B.J., Joos F., Jouzel J., Kattsov V., Lohmann U., Matsuno T., Molina M., Nicholls N., Overpeck J., Raga G., Ramaswamy V., Ren J., Rusticucci M., Somerville R., Stocker T.F., Whetton P., Wood R.A., Wratt D. (2007) Technical Summary. In: *Climate change 2007: The Physical Sciences Basis. Contribution of working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon S., Qin D., Manning M., Chen Z., Marquis M., Averyt K.B., Tignor M., Miller H.L. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Sperling F.A., Raske A.G., Otvos I.S. (1999) Mitochondrial DNA sequence variation among populations and host races of *Lambdina fiscellaria* (Gn.)(Lepidoptera: Geometridae). *Insect Molecular Biology,* **8**, 1: 97-106.

Stauffer C., Lakatos F., Hewitt G.M. (1999) Phylogeography and postglacial colonization routes of *Ips typographus* L.(Coleoptera, Scolytidae). *Molecular Ecology,* **8**, 5: 763-773.

Stireman J.O., Dyer L.A., Janzen D.H., Singer M.S., Lill J.T., Marquis R.J., Ricklefs R.E., Gentry G.L., Hallwachs W., Coley P.D., Barone J.A., Greeney H.F., Connahs H., Barbosa P., Morais H.C., Diniz I.R. (2005) Climatic unpredictability and parasitism of caterpillars: Implications of global warming. *Proceedings of the National Academy of Sciences,* **102**, 48, 17384-17387.

Thaler R., Brandstätter A., Meraner A., Chabicovski M., Parson W., Zelger R., Dalla Via J., Dallinger R. (2008) Molecular phylogeny and population structure of the codling moth (*Cydia pomonella*) in Central Europe: II. AFLP analysis reflects human-aided local adaptation of a global pest species. *Molecular Phylogenetics and Evolution,* **48**, 3: 838-849.

Timm A.E., Geertsema H., Warnich L. (2006) Gene flow among *Cydia pomonella* (Lepidoptera: Tortricidae) geographic and host populations in South Africa. *Journal of Economic Entomology,* **99**, 2: 341-348.

Vekemans X., Beauwens T., Lemaire M., Roldan-Ruiz I. (2002) Data from amplified fragment

length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology,* **11**, 1: 139-151.

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

Wilson A.C., Cann R.L., Carr S.M., George M., Gyllensten U.B., Helm-Bychowski K.M., Higuchi R.G., Palumbi S.R., Prager E.M., Sage R.D., Stoneking M. (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society,* **26**, 4: 375-400.

Yan G., Romero-Severson J., Walton M., Chadee D.D., Severson D.W. (1999) Population genetics of the yellow fever mosquito in Trinidad: comparisons of amplified fragment length polymorphism (AFLP) and restriction fragment length polymorphism (RFLP) markers. *Molecular Ecology,* **8**, 6: 951-963.

Yukuhiro K., Sezutsu H., Itoh M., Shimizu K., Banno Y. (2002) Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. *Molecular Biology and Evolution,* **19**, 8, 1385-1389.

Chapter 2

Phylogeography of the pine processionary moth *Thaumetopoea wilkinsoni* in the Near East

---

Published as:

Simonato M., Mendel Z., Kerdelhué C., Rousselet J., Magnoux E., Salvato P., Roques A., Battisti A., Zane, L. (2007) Phylogeography of the pine processionary moth *Thaumetopoea wilkinsoni* in the Near East. *Molecular Ecology* **16**: 2273-2283

I contributed to all parts of the experimental work and analysis, and to the paper writing.

Molecular Ecology (2007) **16**, 2273–2283 doi: 10.1111/j.1365-294X.2007.03302.x

Phylogeography of the pine processionary moth *Thaumetopoea wilkinsoni* in the Near East

M. SIMONATO,* Z. MENDEL,† C. KERDELHUÉ,‡ J. ROUSSELET,§ E. MAGNOUX,§ P. SALVATO,* A. ROQUES,§ A. BATTISTI* and L. ZANE¶

**Dipartimento di Agronomia Ambientale e Produzioni Vegetali Entomologia, Agripolis, Università di Padova, Via Romea 16, 35020 Legnaro PD, Italy, †Agricultural Research Organization, Volcani Center — Department of Entomology, PO Box 6, IL-50250 Bet Dagan, Israel, ‡INRA Centre de Bordeaux-Pierroton, UMR BIOGECO, Entomologie et Biodiversité, 69 route d'Arcachon, F-33612 Cestas Cedex, France, §INRA-Orléans, Zoologie Forestière, BP 20619, F-45166 Olivet Cedex, France, ¶Dipartimento di Biologia, Università di Padova, Via G. Colombo 3, 35121 Padova, Italy*

Abstract

Phylogeographic structure of the eastern pine processionary moth *Thaumetopoea wilkinsoni* was explored in this study by means of nested clade phylogeographic analyses of COI and COII sequences of mitochondrial DNA and Bayesian estimates of divergence times. Intraspecific relationships were inferred and hypotheses tested to understand historical spread patterns and spatial distribution of genetic variation. Analyses revealed that all *T. wilkinsoni* sequences were structured in three clades, which were associated with two major biogeographic events, the colonization of the island of Cyprus and the separation of southwestern and southeastern Anatolia during the Pleistocene. Genetic variation in populations of *T. wilkinsoni* was also investigated using amplified fragment length polymorphisms and four microsatellite loci. Contrasting nuclear with mitochondrial data revealed recurrent gene flow between Cyprus and the mainland, related to the long-distance male dispersal. In addition, a reduction in genetic variability was observed at both mitochondrial and nuclear markers at the expanding boundary of the range, consistent with a recent origin of these populations, founded by few individuals expanding from nearby localities. In contrast, several populations fixed for one single mitochondrial haplotype showed no reduction in nuclear variability, a pattern that can be explained by recurrent male gene flow or selective sweeps at the mitochondrial level. The use of both mitochondrial and nuclear markers was essential in understanding the spread patterns and the population genetic structure of *T. wilkinsoni*, and is recommended to study colonizing species characterized by sex-biased dispersal.

Keywords: AFLP, microsatellites, mitochondrial DNA, *Pinus* pest, range expansion, *Thaumetopoea wilkinsoni*

Received 3 October 2006; revision received 17 December 2006; accepted 22 January 2007

Introduction

Geographic distributions of species are known to vary considerably in time, according to a number of factors including the geological and palaeoclimatic history of the habitat and the dispersal capacity of the organism (Gaston 2003). In particular, species' ranges have been strongly affected by Quaternary [2.4 million years ago (Ma) to present] climatic fluctuations and ice ages (Hewitt 2000), at least for European and North American temperate species. The organisms responded to climatic oscillations by local extinction in northern regions and survival in southern refugia during the glacial maxima, and by northward range expansions during interglacial, warmer periods. These events played a major role in promoting speciation through formation of isolating barriers allowing allopatric divergence, and in shaping species phylogeography (Hewitt 1996). Yet, species display different phylogeographic patterns, because their response to environmental changes

Correspondence: A. Battisti, Fax: +39 0498272810; E-mail: andrea.battisti@unipd.it

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

during the ice ages primarily depended on ecological, dispersal and life-history traits (Taberlet *et al*. 1998; Hewitt 1999, 2001). Some regions of the world, such as the Near East, were never covered with ice during the Pleistocene, but the occurring species may still have been influenced by climatic oscillations such as cycles of wet and dry periods (Horowitz 1988). Yet, very few studies have analysed the phylogeographic history of terrestrial organisms in the Near East (Tarkhnishvili *et al*. 2001; Veith *et al*. 2003), while more information is available for other regions (Soltis *et al*. 2006).

Moreover, the geographic distribution of phytophagous insects is necessarily embedded within the range of the host plants that provides the potentially exploitable habitat. Compared to the wealth of information about plants, for which fossil deposits and pollen series often allow to reconstruct the distribution over long periods (Klaus 1989; Willis *et al*. 1998), very little knowledge is available concerning the past distributions of phytophagous insects. Fossil remains are scarce (Wilf & Labandeira 1999), and it is rarely possible to directly compare host and associated insect past distributions (but see Koteja (1990) for scale insect–pine association since the Cretaceous). In this context, genetic markers are useful tools to reconstruct the evolution of insect herbivore lineages in relation to the history of their host plants (Hewitt 2001). Phylogeographic analyses of forest insect species have shown interesting patterns of lineage differentiation, partly driven by host plant distribution (Burban *et al*. 1999; Stauffer *et al*. 1999; Kerdelhué *et al*. 2002; Horn *et al*. 2006). These studies indicate a shared host–insect history of habitat colonization, eventually followed by low interpopulation gene flow. Different dispersal patterns may result either in low levels of genetic diversity in new portions of the insect species' range or in high diversity due to increased interpopulation gene flow (Bialozyt *et al*. 2006; Oliver 2006). Dispersal capacities can also affect spatial genetic structure via strong limitation of gene flow (Kerdelhué *et al*. 2006). Since dispersal strategies may differ between sexes (Greenwood & Swingland 1983), the use of sex-specific markers can then allow investigating the genetic effects and evolutionary implications of gender-biased dispersal (Burban & Petit 2003; Sallé *et al*. 2007). Adult females of phytophagous insects, especially among Lepidoptera laying eggs in large patches, are often constrained by heavy egg loads that reduce the flight distance (Thompson & Pellmyr 1991). The combination of powerful sexual pheromones emitted by the females and mobile males may counterbalance the negative effects on gene flow caused by a low female vagility (Salvato *et al*. 2005).

In this study, we explored the phylogeographic structure of a phytophagous insect endemic of the Near East, the eastern pine processionary moth *Thaumetopoea wilkinsoni* Tams (Lepidoptera: Notodontidae). It is a univoltine insect, oligophagous on *Pinus brutia, Pinus halepensis,* and *Pinus nigra* (Schimitschek 1944, Halperin 1990), damaging trees (Carus 2004; Kanat *et al*. 2005), and threatening public health by releasing toxic hairs (Turkmen & Oner 2004). The species was originally described from the island of Cyprus in 1925 (Tams 1925; Wilkinson 1927). Near East continental populations of pine processionary moths had long been considered to belong to its sibling species *Thaumetopoea pityocampa* (Denis et Schiffermüller), occurring on pine in southern Europe and northern Africa, until Salvato *et al*. (2002) provided evidence of species separation.

In particular, we tested the hypothesis that sex-biased dispersal affects genetic variability, by contrasting patterns of differentiation of mitochondrial and nuclear markers. Within this framework, we examined three major phylogeographic patterns of *T. wilkinsoni*, such as (i) the genetic divergence between the populations of the island of Cyprus, whose formation dates back to the Messinian period (5.3 Ma; Marra 2005) and Near East populations, (ii) the differentiation among continental populations, as a consequence of the climatic fluctuations associated with ice ages (Hewitt 2001), and (iii) the affinity between core continental populations and populations of recent origin, as those resulting from the invasion of the southernmost Israeli pine stands and of the Turkish coast of the Black Sea.

Materials and methods

Sampling and DNA protocols

Eggs and larvae of *Thaumetopoea wilkinsoni* were collected at 15 different locations in Turkey, Cyprus, Lebanon and Israel (Table 1). To reduce the risk of sampling siblings, each individual used in the analyses was collected from a different tree, either from an egg batch or from a nest. Eggs were maintained at room temperature until hatching, after which the first instar larvae were transferred to ethanol 70%. Alternatively, larvae were directly sampled from nests in the field and immediately transferred to ethanol 70%. All ethanol-preserved material was stored at −20 °C. DNA was extracted using a salting-out procedure (Patwary *et al*. 1994). The same individuals were generally used for all the analyses, different numbers resulted from limitations imposed by the analytical procedures.

Two mitochondrial DNA (mtDNA) fragments, corresponding to parts of the COI and COII genes, were amplified from 192 individuals and examined through single-strand conformation polymorphism (SSCP) analysis, as described in Salvato *et al*. (2002). For each mobility class, one to five individuals were sequenced directly using an ABI PRISM 3100 (Applied Biosystems) DNA sequencer and a Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems) to check for the accuracy of the SSCP analysis and to determine the corresponding haplotype. Sequences were aligned using

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 1 Location of *Thaumetopoea wilkinsoni* populations, according to geographic position from southeast to northwest and to the host plant on which samples were collected

Country	Region/district	Location	Latitude	Longitude	Altitude (m a.s.l.)	Host*	Collector
Israel	S Judean mountains	Yatir	31°20′N	35°03′E	550	PA	Authors
Israel	W Negev	Qisufim	31°22′N	34°24′E	50	PA	Authors
Israel	Judean foothills	Haruvit	31°45′N	34°50′E	150	PA	Authors
Israel	Lower Galilee	Segev	32°52′N	35°14′E	400	PA	Authors
Israel	Upper Galilee	Qiryat Shemona	33°11′N	35°33′E	350	PB	Authors
Lebanon	Beirut	Beirut	33°53′N	35°30′E	272	PB	American University Beirut
Turkey	Antakia	Seyhköy	36°04′N	36°10′E	450	PB	Authors
Turkey	Iskenderun	Iskenderun	36°34′N	36°10′E	210	PB	Authors
Turkey	Taurus mountains	Aladag	37°33′N	35°22′E	1100	PB	Authors
Turkey	Taurus mountains	Pozanti	37°17′N	34°51′E	970	PB, PN	Authors
Cyprus	E Cyprus	El Skopi	35°00′N	32°40′E	100–1000	PB, PN	Authors
Turkey	Antalya	Karaoz	36°54′N	30°43′E	200	PB	University of Isparta
Turkey	Isparta	Gunur	37°46′N	30°34′E	1050	PB, PN	University of Isparta
Turkey	Izmir	Aydin	37°51′N	27°50′E	600	PB	University of Izmir
Turkey	Samsun	Samsun	41°17′N	36°20′E	150	PN	Authors

*PA: *Pinus halepensis*, PB: *Pinus brutia*, PN: *Pinus nigra*.
m a.s.l., metres above sea level.

CLUSTAL X (Thompson *et al*. 1997). Sequences of COI (262 bp) and COII fragments (342 bp) were then concatenated, resulting in a 604 bp-long final alignment.

Four microsatellite loci (MS-Thpit1, MS-Thpit3, MS-Thpit4, MS-Thpit5) were characterized on 230 individuals. Microsatellite primers and amplification conditions are described in Rousselet *et al*. (2004). Fluorescent (polymerase chain reaction) PCR products were run and detected on an ABI PRISM 3100 automatic sequencer (Applied Biosystems) and product sizes were determined using the GENESCAN software (Applied Biosystems).

The amplified fragment length polymorphism (AFLP) protocol (Vos *et al*. 1995) was used with four primer combinations yielding 125 bands on 142 larvae analysed. Approximately 50 ng of DNA were digested with *Eco*RI and *Mse*I restriction enzymes and ligated to specific AFLP adapters. Each sample was subsequently diluted 10-fold and used as template for preselective and selective (*Eco*RI-AAC/*Mse*I-CAT, *Eco*RI-ACA/*Mse*I-CAG, *Eco*RI-AGC/*Mse*I-CAT, *Eco*RI-AAG/*Mse*I-CAC) PCR amplifications. AFLP products were run in an ABI PRISM 3700 DNA Analyser (Applied Biosystems). Band scoring was performed with GENOTYPER version 3.7 (Applied Biosystems) considering bands in the range 70–360 bp. AFLP profiles were checked by hand for accurate scoring. The intensity of each individual peak was normalized on the basis of the total signal intensity and the peak was considered only if its intensity exceeded a fixed threshold of 100 fluorescent units. AFLP profiles were recorded in a matrix as presence or absence of bands for each individual. Both polymorphic and monomorphic bands were scored.

Data analysis

Homologous mtDNA sequences of two related species, *Thaumetopoea pityocampa* (Salvato *et al*. 2002: GenBank Accession nos EF015538, EF015542) and *Thaumetopoea pinivora* (from Gotland, Sweden, accession number EF364032, EF364033), were included in mitochondrial data analysis. A partition homogeneity test was performed for the COI and COII fragments using PAUP* v4.0b10 (Swofford 2002). The test confirmed that these regions contained homogeneous signal ($P = 0.35$), allowing data to be pooled for further analyses.

Phylogenetic relationships between haplotypes were estimated by Bayesian Inference (BI) with MrBayes v3.1 (Huelsenbeck & Ronquist 2001); the analyses were performed without outgroup definition and best trees were rooted with *T. pityocampa* and *T. pinivora*. BI analysis was used because it implements codon position partitioned models (CP models), thus allowing the protein coding nature of the data to be considered. The best CP model was selected by comparing the exact likelihood under different models of a consensus maximum parsimony tree using the BASEML software of PAML package (Yang 1997). According to published suggestions (Shapiro *et al*. 2006), two CP models were tested, namely the Hasegawa, Kishino and Yano model (HKY, Hasegawa *et al*. 1985) and the general time reversible model (GTR, Lanave *et al*. 1984) with and without gamma distributed site heterogeneity. The sequences were partitioned according to codon position, and the chosen model (and alpha where appropriate) was assumed for all sites; different rates were allowed for each partition.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

The best CP model found was then used for Bayesian phylogenetic inference using MRBAYES, with and without enforcement of the molecular clock. Analyses were run for 1 million generations, and Markov chains were sampled every 10 generations. The length of the chain was chosen after that initial trials indicated approximate convergence after 30 000 generations. The 50% majority rule consensus tree and the Bayesian posterior probabilities were obtained from sampled trees, after burning first 25% of the chain.

Clades were approximately dated using BEAST (Drummond & Rambaut 2003), assuming a sequence divergence rate of 2–2.3% per million years (DeSalle *et al.* 1987; Brower 1994). Models of sequence evolution, data partitioning and clock assumptions followed the results obtained from previous analyses; Markov chain Monte Carlo (MCMC) was run for 10 million generations, results being logged every 1000 generations. After discarding the first 10% of the chain, convergence was checked by monitoring traces of sampled parameters and effective sample size following authors' suggestions.

A haplotype parsimony network was reconstructed using TCS 1.21 (Clement *et al.* 2000) as described by Templeton *et al.* (1992), with a probability cut-off set at 93%. The network was used to perform a nested clade phylogeographic analysis (NCPA) using GEODIS version 2.0 (Posada *et al.* 2000), to test the null hypothesis of lack of association between clades and geographic location. Significant values were used to discriminate the effects of recurrent gene flow and historical processes which may have affected the spatial genetic structure of populations (Templeton 2004) using the updated inference key (http://darwin.uvigo.es/download/geodisKey_11Nov05.pdf).

The genetic variability of each population was estimated for mitochondrial and microsatellite data using ARLEQUIN version 3.1 (Excoffier *et al.* 2005) and expressed as haplotype diversity and expected heterozygosity (H_E), respectively. For AFLP markers, the heterozygosity (H_S) was estimated by the Bayesian approach implemented in HICKORY version 1.0 (Holsinger & Lewis 2003), to overcome problems caused by dominance. In addition, for microsatellite data only, deviations from Hardy–Weinberg equilibrium were tested for each locus and population using ARLEQUIN, with 10 000 permutations. Comparisons of microsatellite nuclear diversity among population groups were carried out by FSTAT version 2.9.3.2 (Goudet 1995).

For all three markers, the partition of genetic variability among populations and among group of populations was defined by analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) using ARLEQUIN. Pairwise Φ_{ST} and F_{ST} between populations were also calculated. Distances used were Kimura 2-parameters distance for mitochondrial data, number of different alleles for microsatellites and pairwise differences (equivalent to simple matching in Apostol *et al.* 1993) for AFLP. The use of alternative genetic distances for mitochondrial data resulted in very similar results. Null hypothesis of genetic homogeneity was assessed by 10 000 replications, reshuffling individuals among populations, and, when needed, populations among groups.

Results

Mitochondrial DNA phylogeography

The SSCP analysis clearly distinguished 11 mobility classes for the COI fragment and 15 classes for the COII fragment. A total of 20 composite mobility classes (COI+COII) were found. Random sequencing of individuals confirmed the accuracy of the SSCP method, each mobility class corresponding to a single haplotype and vice-versa (GenBank Accession nos EF210075–EF210097). The uncorrected pairwise divergence between *Thaumetopoea wilkinsoni* haplotypes ranged from 0.0017 to 0.0348. When these haplotypes were aligned with the homologous sequence of the closely related *Thaumetopoea pityocampa* and *Thaumetopoea pinivora*, the divergence between the three species ranged from 0.0894 to 0.1159.

The best model of sequence evolution was the GTR with different rates for each codon position; this model was thus chosen for phylogenetic inference and for the Bayesian molecular clock analysis. BI consensus tree is showed in Fig. 1. All *T. wilkinsoni* sequences were clustered in a single monophyletic group (A) with 100% support. All haplotypes from Cyprus were grouped in a cluster (B) with 97% confidence, and appeared as the sister group of a well-supported clade (C, 85%) containing all the haplotypes

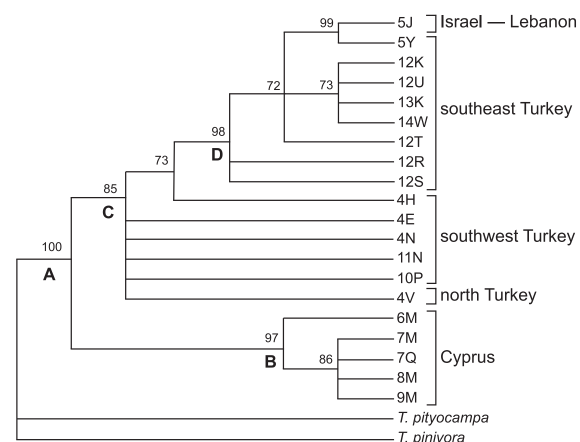


Fig. 1 Consensus tree obtained from Bayesian inference of COI and COII data. Numbers above branches indicate, when higher than 70%, the Bayesian posterior probability of support for the node. Clades discussed in the text are indicated by capital letters A-D.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 2 Descriptive statistics of mitochondrial and nuclear (microsatellite and AFLP) DNA markers, with the number of individuals analysed. The same individuals were generally used for all the analyses, different numbers resulted from limitations imposed by the analytical procedures. The symbol ± indicates the confidence interval (0.95) of each estimate

Country and location		mtDNA		Microsatellites							AFLP (HICKORY)	
				N	H_E (unbiased) per locus							
		N	Haplotype diversity		Thpit 1	Thpit 3	Thpit 4	Thpit 5	mean H_E	SD	N	H_S
Israel	Yatir	15	0.00 ± 0.00	14	0.20 ± 0.09	0.51 ± 0.04	0.20 ± 0.10	0.00 ± 0.00	0.23	0.21	15	0.15 ± 0.01
Israel	Qisufim	10	0.00 ± 0.00	9	0.00 ± 0.00	0.50 ± 0.06	0.29 ± 0.12	0.00 ± 0.00	0.20	0.25	10	0.18 ± 0.01
Israel	Haruvit	15	0.00 ± 0.00	14	0.25 ± 0.10	0.45 ± 0.07	0.14 ± 0.08	0.00 ± 0.00	0.21	0.19	15	0.21 ± 0.01
Israel	Segev	9	0.00 ± 0.00	10	0.28 ± 0.12	0.44 ± 0.09	0.10 ± 0.09	0.00 ± 0.00	0.21	0.19	9	0.20 ± 0.01
Israel	Qyriat Shemona	14	0.00 ± 0.00	13	0.50 ± 0.10	0.32 ± 0.10	0.76 ± 0.06	0.32 ± 0.10	0.48	0.21	14	0.19 ± 0.01
Lebanon	Beirut	24	0.00 ± 0.00	24	0.36 ± 0.07	0.47 ± 0.04	0.56 ± 0.08	0.19 ± 0.07	0.39	0.16	9	0.18 ± 0.01
Turkey	Seyhköy	11	0.00 ± 0.00	20	0.85 ± 0.02	0.43 ± 0.07	0.55 ± 0.09	0.00 ± 0.08	0.53	0.24	8	0.20 ± 0.01
Turkey	Iskenderun	10	0.71 ± 0.12	19	0.82 ± 0.04	0.60 ± 0.06	0.88 ± 0.04	0.10 ± 0.06	0.60	0.35	—	—
Turkey	Aladag	10	0.51 ± 0.16	20	0.73 ± 0.03	0.49 ± 0.04	0.67 ± 0.05	0.00 ± 0.00	0.47	0.33	9	0.18 ± 0.01
Turkey	Pozanti	11	0.51 ± 0.10	20	0.65 ± 0.06	0.36 ± 0.07	0.66 ± 0.05	0.00 ± 0.00	0.42	0.31	10	0.18 ± 0.01
Cyprus	El Skopi	18	0.74 ± 0.08	15	0.70 ± 0.05	0.52 ± 0.09	0.94 ± 0.02	0.58 ± 0.10	0.69	0.19	16	0.22 ± 0.01
Turkey	Karaoz	8	0.46 ± 0.20	8	0.52 ± 0.13	0.13 ± 0.11	0.88 ± 0.05	0.00 ± 0.00	0.38	0.40	8	0.20 ± 0.01
Turkey	Gunur	15	0.00 ± 0.00	13	0.31 ± 0.12	0.09 ± 0.08	0.89 ± 0.05	0.00 ± 0.00	0.32	0.40	13	0.25 ± 0.01
Turkey	Aydin	10	0.20 ± 0.15	11	0.00 ± 0.00	0.09 ± 0.08	0.82 ± 0.04	0.09 ± 0.08	0.25	0.38	—	—
Turkey	Samsun	12	0.00 ± 0.00	20	0.40 ± 0.08	0.00 ± 0.00	0.53 ± 0.07	0.00 ± 0.00	0.23	0.27	6	0.16 ± 0.01

H_E, expected heterozygosity.

from continental sites. Within this latter cluster, a highly supported group was identified (D, 98%) composed of haplotypes found in Israel, Lebanon and in southeast Turkey (Pozanti, Aladag, Iskenderun and Seyköy). The remaining haplotypes from north and southwest Turkey were not resolved inside the C group, except for a weak tendency of haplotype 4H to cluster a sister group of clade D (73%).

The same well-differentiated groups were found in the parsimony-based network (Fig. 2). It confirmed the strong divergence of Cyprus (clade B) that differed by at least 12 mutations from the closest continental haplotype, and identified two groups separated by at least 6 mutations, corresponding to the D clade previously identified (southeast Turkey) and a clade containing all haplotypes from north and southwest Turkey. NCPA further showed that the geographic distribution of Cypriot haplotypes (clade 3-3) was consistent with allopatric fragmentation, whereas for the Lebanese, Israeli and southeastern Turkish haplotypes (clade 3-1), it indicated a contiguous range expansion. No conclusive indications were obtained concerning the differentiation between the groups D and the remaining clades (clades 3-1 vs. 3-2, Fig. 2).

The age of the most recent ancestor of supported groups was estimated using BEAST, assuming a strict molecular clock because analyses conducted with MRBAYES showed no significant differences in likelihood when the clock was or was not enforced. Considering the 2–2.3% per million-year (Myr) divergence rate for arthropod mtDNA, and bearing in mind the large confidence intervals associated with these estimates, the split between Cyprus and continental haplotypes (clade A, Fig. 1), was tentatively dated to 1.90–1.27 Ma. The continental haplotypes (clade C) diverged 1.12–0.74 Ma, and those in Cyprus and southeast Turkey (clades B and D) diverged 0.30–0.20 Ma and 0.65–0.43 Ma, respectively.

Comparison between mitochondrial and nuclear markers

Population genetic variability was estimated for the three markers applied (Table 2). Most microsatellite loci and populations were at Hardy–Weinberg equilibrium, as only 8 tests were significant (locus MS-Thpit1 in Aladag, Iskenderun and Seyhköy; MS-Thpit4 in Karaoz, Aladag, Samsun and Iskenderun; MS-Thpit3 in Iskenderun). Haplotype diversity varied substantially between populations, ranging from 0 in several populations at the southern and northern edge of the species range, to 0.71–0.74 in the Iskenderun and Cyprus samples.

Several populations fixed for a single mitochondrial haplotype bore substantial microsatellite and AFLP variation. In particular, among the 9 populations fixed for a single mitochondrial haplotype, those at the boundary of the distribution (Samsun in northern Turkey, and the four southernmost Israeli populations of Segev, Haruvit, Yatir

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

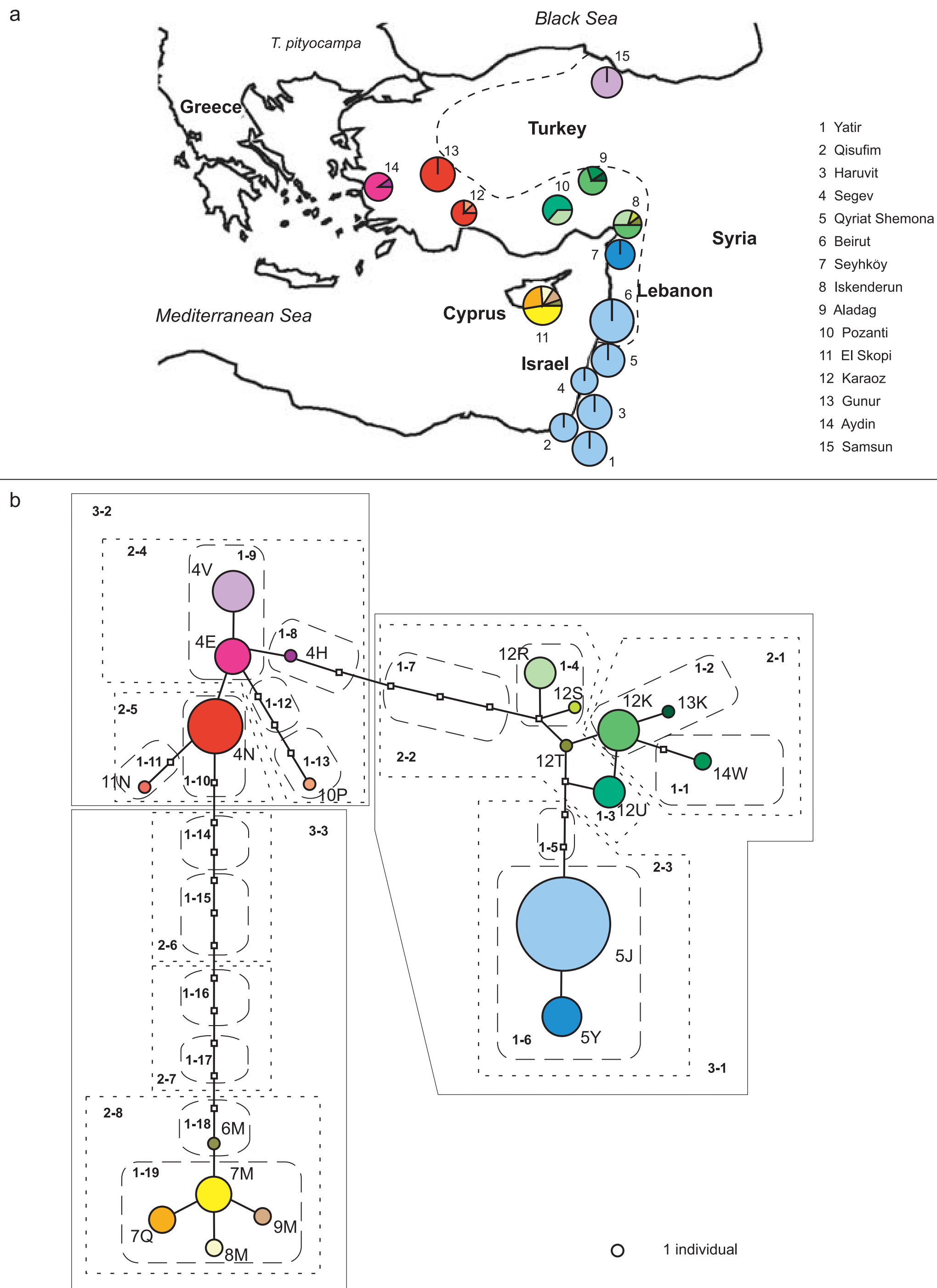


Fig. 2 Distribution of mitochondrial DNA haplotypes and range of *Thaumetopoea wilkinsoni* in the Near East (area between the dashed line and the coast), based on Schimitschek (1944) and Commonwealth Institute of Entomology (1977). (a) Haplotype network inferred by the criterion of parsimony with TCS 1.18 (Clement *et al*. 2000). (b) Each line in the network represents a single mutational change. Haplotype frequencies are represented by the area of the circles. Empty circles indicate intermediate, missing haplotypes. Boxes represent the n-step clades.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 3 Results of AMOVA tests on mitochondrial and nuclear (microsatellite and AFLP) DNA markers, divided according the phylogeographic hypotheses discussed in the text

	Source of variation	mtDNA Variance components	mtDNA Percentage of variation	Microsatellite Variance components	Microsatellite Percentage of variation	AFLP Variance components	AFLP Percentage of variation
Whole data set	Among populations (pops)	3.47621 Va	95.11%***	0.25420 Va	26.00%***	4.61137 Va	37.98%***
	Within populations	0.17868 Vb	4.89%***	0.72332 Vb	74.00%	7.53097 Vb	62.02%***
(a) two groups (Cyprus/continent)	Among groups	5.28635 Va	67.26%***	0.10831 Va	10.13%***	−0.19812 Va	−1.65% NS
	Among pops within groups	2.39403 Vb	30.46%***	0.23779 Vb	22.24%***	4.65459 Vb	38.83%***
	Within populations	0.17868 Vc	2.27%***	0.72332 Vc	67.64%***	7.53097 Vc	62.82%***
(b) three groups (Cyprus/Israel,	Among groups	5.21527 Va	84.14%***	0.25264 Va	22.75%***	2.05561 Va	15.63%**
Lebanon, east	Among pops within groups	0.80445 Vb	12.98%***	0.13446 Vb	12.11%***	3.56856 Vb	27.13%***
Turkey/north-west Turkey)	Within populations	0.17868 Vc	2.88%***	0.72332 Vc	65.14%***	7.53097 Vc	57.25%***
(c) 2 groups	Among groups	4.11001 Va	81.22%***	0.27190 Va	25.21%***	2.93633 Va	21.02%**
(Israel, Lebanon,	Among pops within groups	0.80727 Vb	15.95%***	0.13620 Vb	12.63%***	3.57583 Vb	25.60%***
east Turkey/northwest Turkey)	Within populations	0.14282 Vc	2.82%***	0.67046 Vc	62.16%***	7.45500 Vc	53.38%***

**$P < 0.01$; ***$P < 0.001$; NS; not significant. P values corrected according to Bonferroni's test.

and Qisufim) showed values of heterozygosity (0.20–0.23) for microsatellite loci lower than that of the other samples ($P = 0.0015$). In contrast, microsatellite variability of the remaining four populations (Gunur, Seyhköy, Beirut and Qyriat Shemona) showed level of variability not significantly different from that of populations not fixed for mitochondrial haplotypes ($P = 0.2724$). Finally, both microsatellite markers and mitochondrial sequences revealed the highest mean heterozygosity in the Iskenderun and Cyprus populations ($H_E = 0.60$ and 0.69, respectively).

Results of the AMOVA tests are shown in Table 3. When conducted on the whole sample of 13 populations, AMOVA showed that about 95% of mitochondrial variation was attributable to differences among populations. Highly significant values were also found using nuclear markers, though they explained a smaller proportion of the total variation, corresponding to *c.* 26% with microsatellites and 38% with AFLP markers. When populations were clustered in two groups according to geography, to test the separation of Cyprus vs. continental populations, among-group variation explained a significant proportion of mtDNA and microsatellites variation (67% and 10%, respectively), whereas it was not significant for AFLP markers (Table 3a). When splitting the continental populations into two groups separated by the Taurus mountains (i.e. Cyprus vs. northwest Turkey vs. southeast Turkey-Israel-Lebanon), a significant proportion of the genetic variation was found among groups for all markers used (16–84%), the AFLP markers yielding the smallest value (Table 3b). When considering only continental populations in relation to the climatic fluctuations associated with ice ages, the remaining two groups (northwest Turkey vs. southeast Turkey-Israel-Lebanon) significantly explained 21–81% of the genetic variation (Table 3c).

Discussion

Mitochondrial phylogeographic patterns and female colonization routes

Our results clearly show that all individuals sampled in Cyprus and the Near East belong to the same species, *Thaumetopoea wilkinsoni*, as all corresponding haplotypes cluster together in a well-supported monophyletic group. All the genetic distances between these haplotypes and the closely related *Thaumetopoea pityocampa* are over 8%, while all distances within *T. wilkinsoni* are comprised between 0.2 and 3.6%. It confirms the preliminary results of Salvato *et al.* (2002), showing that *T. pityocampa* is absent from the easternmost part of the Mediterranean Basin where its sibling *T. wilkinsoni* occurs.

Within *T. wilkinsoni*, mitochondrial data indicate three main phylogeographic events, namely: (i) the disjunction between Cypriot and Anatolian populations of the moth,

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

(ii) the split between western and eastern continental groups, and (iii) further divergence within the eastern clade between north and south populations. The Bayesian inference of divergence times indicates that the separation between Cyprus and Near East continental haplotypes occurred during the Pleistocene, in a period when land bridges between the island and the continent are excluded (Simmons 1999 and references therein). The formation of Cyprus is supposed to date back to 5.3 Ma (Marra 2005) during the early Pliocene. Moreover, during the Pleistocene minimum sea level, the distance between Cyprus probably never dropped below 30–40 km (Simmons 1999), a distance well beyond the known flight range of female moths (3–4 km, Halperin *et al.* 1981). Thus, the colonization of the island by the moth probably happened through a rare event of long-distance dispersal. This occasional long-range dispersal probably led to an extreme reduction of allelic richness in Cyprus due to a founder effect, and new alleles then arose, which could explain the typical star-shape topology of Cypriot haplotypes.

The split between the two continental groups (eastern vs. western clade) probably occurred about 1.5–0.5 Ma, concomitantly with the Quaternary transgressional cycle during which the Mediterranean sea level varied between –150 and +120 m when compared to the present, as a consequence of the glacial events which occurred in Europe (Horowitz 1988). Shoreline refugia of *T. wilkinsoni* associated with Mediterranean pines are thus unlikely for that period, whereas montane *Pinus nigra* forests close to the coast probably were favourable refugia for the moth, as shown by Ciesla (2004) for Cyprus. Furthermore, such potential refugial forests have a disconnected distribution in southern Anatolia, in the disjointed western and eastern Taurus (Vidakovic 1991). The split between the western and the eastern Anatolian lineages can thus be explained by the existence of two separate montane refugia of *P. nigra* and the subsequent isolation of the corresponding populations on this host during the Quaternary transgressional cycle. The northernmost population of Samsun, on the Black Sea, was colonized very recently, and our results show that the migrant individuals undoubtedly came from western Turkey. We expect that a more thorough regional sampling would reveal the Samsun haplotype (4V) in western Turkey, except if it arose locally from a fairly recent point mutation.

The eastern clade (D) includes populations from eastern Turkey, Lebanon and Israel. Network topology shows that it may be split into two subclades. As divergence time within the clade is estimated to range from 1 to 0.22 Ma, the two subclades may have originated from two isolated refugia areas on eastern Taurus mountains (*P. nigra* and *Pinus brutia*) and Lebanon mountains (*P. brutia*) during the Quaternary transgressional cycle. Genetic diversity was retained in the northern populations, in which effective population sizes probably never dropped below a critical threshold under which most alleles would have been lost (Young *et al.* 1996; Austerlitz *et al.* 2000). Instead, haplotype fixation was observed in southern populations, perhaps because the ecological features of the environment at the southern boundary of the host range. The occurrence of suitable host pines in southern Israel is recent, as it dates back to the afforestation conducted in the 1910s (Bonneh 2000), and the colonization of the southernmost localities by the moth was first detected in the 1930s (Anonymous 1939). Some relict, isolated stands of *Pinus halepensis* exist far south in Israel, but were probably exempt from the moth until recently, as *T. wilkinsoni* was not detected during an old survey of lepidopterans which detected other species of *Thaumetopoea* (Amsel 1933). The affinity between Israeli, Lebanese and southeastern Turkish populations indicates that the colonization of Israel was due to individuals from the southeastern part of the range, thus excluding the possibility of accidental introduction from Cyprus as previously hypothesized (Mendel 1990). As all the populations from Israel and southern Lebanon share the same single mtDNA haplotype, we are probably dealing with a single source of migrant females. The massive afforestation effort in Israel has created a suitable corridor that allowed the moth to reach some of the relic stands of *P. halepensis* in the south (Liphschitz & Biger 2001).

Unexpected patterns of nuclear diversity, and sex-biased gene flow

The information yielded by nuclear markers, both microsatellites and AFLP, provided a rather different estimate of gene flow between populations. The most striking result was that the separation of the Cypriot population from the continental ones explained much (67%) of the mitochondrial variation, but only a little proportion (10% to 0%) of microsatellite and AFLP nuclear variation. Even though homoplasy in nuclear markers (i.e. Cypriot and continental alleles being identical by state but not identical by descent) could account for this discrepancy, it is more plausible (given the high number of markers used) that the different histories reconstructed with nuclear and mitochondrial markers rather reflect sex-biased dispersal. In fact, a positive correlation between single-locus F_{ST} and average heterozygosity estimates was found for microsatellites (data not shown), in contrast to what expected in the case of homoplasy (O'Reilly *et al.* 2004). Moreover, no significant correlation between size and frequencies of AFLP fragments was found; a negative significant correlation could lead to underestimate genetic diversity and genetic divergence within and between populations (Vekemans *et al.* 2002). Thus, recurrent male gene flow possibly occurred between the island and the continent, although the female gene pool remained isolated for the past 1 or 2 Myr. Dispersal is

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

known to differ between sexes in *T. wilkinsoni*, as males can fly up to 20 km, whereas females can exceptionally reach 3–4 km (Halperin *et al*. 1981). This fivefold difference in maximal dispersal is probably an underestimation of the actual value, considering that the lifespan of the two sexes is few hours in female and up to 10 days in male moth (Halperin 1990). For instance, in the western sibling species *T. pityocampa* the mean female dispersal is 300 m (Demolin 1969) whereas males are attracted to pheromone traps located at about 20 km away from the nearest infested pine forest (Kerdelhué *et al*. 2006). Sex-biased gene flow has already been hypothesized to explain the incongruent results between mitochondrial and nuclear genes in the sibling *T. pityocampa* (Salvato *et al*. 2002) and in other forest insects (Sallé *et al*. 2007).

Our results show that both types of DNA markers are necessary to infer the genetic relatedness of populations accurately. This is evident also in comparison between continental populations: four populations which probably survived on relic natural stands and thus regarded as 'old origin' (Gunur, Seyhköy, Beirut and Qyriat Shemona) did not show any reduced nuclear diversity, although they were fixed for one single mitochondrial haplotype. This result may indicate that reduced mitochondrial diversity is due to a past reduction in population size, and that recurrent male gene flow allowed the nuclear variation to be recovered during the recent population history. Alternatively, in the light of the accumulating evidence that mtDNA is often not evolving neutrally (Ballard & Whitlock 2004), the observed pattern may be explained by a selective sweep at the mtDNA level. In particular, a low mitochondrial polymorphism could result from the linkage disequilibrium with maternally inherited symbiont microorganisms such as *Wolbachia (*reviewed in Hurst & Jiggins 2005). While the presence of such symbionts has not been reported so far in *Thametopoea* species, *Wolbachia* was found in one out of nine Noctuoidea species tested (West *et al*. 1998), leaving the selective sweep hypothesis open. If this is the case, we should hypothesize at least three independent selective sweeps, leading to the fixation of different haplotypes in distinct geographic areas (Gunur, Seyhköy, Beirut and Qyriat Shemona). At present, our data do not allow to discriminate between the two alternative hypotheses. On the contrary, populations from Samsun in northern Turkey, and the four southernmost Israeli populations of Segev, Haruvit, Yatir and Qisufim, show a reduction in both mitochondrial and microsatellite diversity, which is consistent with the hypothesis of recent origin of these populations, founded by individuals expanding from nearby localities into new afforestation areas (Oliver 2006).

In conclusion, our findings contribute to the amount of work recently devoted to study organism dispersal during range expansion, to describe the pattern of genetic variation at the species' range edge, in order to understand the effect of different dispersal strategies on the adaptation of new populations (e.g. Petit *et al*. 2004; Alleaume-Benharira *et al*. 2006; Bialozyt *et al*. 2006). In plants, these studies unveiled a much stronger structure at maternally than paternally or bi-parentally inherited loci due to different rates of seed and pollen dispersal (Petit *et al*. 2005). In this respect, our results indicate a remarkable analogy in the dispersal strategy between pine processionary females and seeds, and between male moths and pollen. However, our results add a further level of complexity to the picture, by showing that the current pattern of genetic variation can possibly result from processes so different as gene flow replenishment by migration or selective sweeps at the mitochondrial DNA level, and confirm the need for the use of different markers in phylogeographic studies.

Acknowledgements

We thank Mustafa Avci, Mikdat Doganlar, Nabil Nemer, Stig Larsson for their help with the insect sampling, Gabriel Walton for linguistic revision, and three anonymous reviewers for valuable comments. Comments by Alberta Boato helped to improve earlier versions of the manuscript. We thank also three anonymous referees for comments and suggestions. The work was supported by the Commission of the European Communities 'Quality of Life and Management of Living Resources' specific RTD program, Contract QLK5-CT-2002-00852, 'PROMOTH: Global change and pine processionary moth: a new challenge for integrated pest management'. The study does not necessarily reflect the Commission's views and in no way anticipates the Commission's future policy on the matter.

References

Alleaume-Benharira M, Pen IR, Ronce O (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal of Evolutionary Biology*, **19**, 203–215.

Amsel HG (1933) Die Lepidopteren Palästinas. Eine zoogeographisch-ökologisch-faunistische Studie. *Zoogeographica*, **2**, 1–146.

Anonymous (1939) Forest protection. In: *Annual Report of Forest Department of Palestine, 1936–39*. Government Printer, Jerusalem.

Apostol BL, Black WCI, Miller BR, Reiter P, Beaty BJ (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. *Theoretical and Applied Genetics*, **86**, 991–1000.

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics*, **154**, 1309–1321.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Bialozyt R, Ziegenhagen B, Petit RJ (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, **19**, 12–20.

Bonneh O (2000) Management of planted pine forest in Israel: past, present and future. In: *Ecology, Biogeography and Management of Pinus Halepensis and P. brutia Forest Ecosystems in the Mediterranean Basin* (eds Ne'eman G, Trabaud L), pp. 377–390. Backhuys, Leiden, The Netherlands.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences, USA*, **91**, 6491–6495.

Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology*, **12**, 1487–1495.

Burban C, Petit RJ, Carcreff E, Jactel H (1999) Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Molecular Ecology*, **8**, 1593–1602.

Carus S (2004) Impact of defoliation by the pine processionary moth (*Thaumetopoea pityocampa*) on radial, height and volume growth of Calabrian pine (*Pinus brutia*) trees in Turkey. *Phytoparasitica*, **32**, 459–469.

Ciesla WM (2004) Forests and forest protection in Cyprus. *Forestry Chronicle*, **80**, 107–113.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Commonwealth Institute of Entomology (1977) Distribution maps of pests. *Thaumetopoea pityocampa* (Schiff.). Series A, Map No. 366, London.

Demolin G (1969) Comportement des adultes de *Thaumetopoea pityocampa* Schiff: dispersion spatiale, importance écologique. Annales des Sciences Forestières, **26**, 81–102.

DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *Journal of Molecular Evolution*, **26**, 157–164.

Drummond AJ, Rambaut A (2003) *BEAST*, Version 1.3. Oxford University Press, Oxford. http://evolve.zoc.ox.ac.uk/beast/).

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

Excoffier L, 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.

Gaston KJ (2003) *The Structure and Dynamics of Geographic Ranges*. Oxford University Press, New York.

Goudet J (1995) FSTAT, version 1.2: a program for IBM PC compatibles to calculate Weir and Cockerham's estimators of *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Greenwood PJ, Swingland IR (1983) *The Ecology of Animal Movement*. Clarendon Press, Oxford, UK.

Halperin J (1990) Life history of *Thaumetopoea* spp. (Lep., Thaumetopoeidae) in Israel. *Journal of Applied Entomology*, **110**, 1–6.

Halperin J, Golan Y, Descoins C *et al.* (1981) Preliminary trials on the attraction of *Thaumetopoea wilkinsoni* Tams to the synthesized sex pheromone of *Th. pityocampa* Schiff. *La Yaaran*, **31**, 24–31 (in Hebrew; English summary).

Hasegawa M, Kishino H, Yano T (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

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

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

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

Hewitt GM (2001) Speciation, hybrid zone and phylogeography – or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Holsinger KE, Lewis PO (2003) *HICKORY 1.0*. Department of Ecology & Evolutionary Biology. University of Connecticut, Storrs, Connecticut. (http://darwin.eeb.uconn.edu/hickory).

Horn A, Roux-Morabito G, Lieutier F, Kerdelhué C (2006) Phylogeographic structure and past history of the circum-Mediterranean species *Tomicus destruens* Woll. (Coleoptera: Scolytinae). *Molecular Ecology*, **15**, 1603–1615.

Horowitz A (1988) The Quaternary environments and paleogeography of Israel. In: *The Zoogeography of Israel. The Distribution and Abundance at a Zoogeographical Crossroad* (eds Yom-Tov Y, Tchernov E), pp. 35–58. Junk, Dordrecht, The Netherlands.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: The effects of inherited symbionts. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **272**, 1525–1534.

Kanat M, Alma MH, Sivrikaya F (2005) Effect of defoliation by *Thaumetopoea pityocampa* (Den. & Schiff.) (Lepidoptera: Thaumetopoeidae) on annual diameter increment of *Pinus brutia* Ten. in Turkey. *Annals of Forest Sciences*, **62**, 91–94.

Kerdelhué C, Magnoux E, Lieutier F, Roques A, Rousselet J (2006) Comparative population genetic study of two oligophagous insects associated with the same hosts. *Heredity*, **97**, 38–45.

Kerdelhué C, Roux-Morabito G, Forichon J *et al.* (2002) Population genetic structure of *Tomicus piniperda* L. (Curculionidae: Scolytinae) on different pine species and validation of *T. destruens* (Woll.). *Molecular Ecology*, **11**, 483–494.

Klaus W (1989) Mediterranean pines and their history. *Plants Systematics and Evolution*, **162**, 133–163.

Koteja J (1990) Paleontology. In: *World Crop Pest. Armored Scale Insects. Their Biology, Natural Enemies and Control* (ed. Rosen D), Vol. 4A, pp. 149–163. Elsevier, Amsterdam.

Lanave C, Preparata G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution*, **20**, 86–93.

Liphschitz N, Biger G (2001) Past distribution of Aleppo pine in the mountains of Israel (Palestine). *The Holocene*, **11**, 427–436.

Marra AC (2005) Pleistocene mammals of Mediterranean islands. *Quaternary International*, **129**, 5–14.

Mendel Z (1990) On the origin of the pine processionary caterpillar, *Thaumetopoea wilkinsoni* Tams (Lep., Thaumetopoeidae) in Israel. *Journal of Applied Entomology*, **109**, 311–314.

O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology*, **13**, 1799–1814.

Oliver JC (2006) Population genetic effects of human-mediated plant range expansions on native phytophagous insects. *Oikos*, **112**, 456–463.

Patwary MU, Kenchington EL, Bird CJ, Zouros E (1994) The use of random amplified polymorphic DNA markers in genetic studies of the sea scallop *Plactopecten magellanicus* (Gmellin, 1791). *Journal of Shellfish Research*, **13**, 547–553.

Petit RJ, Bialozyt R, Garnier-Gere P, Hampe A (2004) Ecology and genetics of tree invasions: from recent introductions to Quaternary migrations. *Forest Ecology and Management*, **197**, 117–137.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488. (http://darwin.uvigo.es/software/geodis.html).

Rousselet J, Magnoux E, Kerdelhué C (2004) Characterization of five microsatellite loci in the pine processionary moth *Thaumetopoea pityocampa* (Lepidoptera Notodontidae Thaumetopoeinae). *Molecular Ecology Notes*, **4**, 213–214.

Sallé A, Artofer W, Lieutier F, Stauffer C, Kerdelhué C (2007) Phylogeography of a host-specific insect: the genetic structure of *Ips typographus* in Europe does not reflect the past fragmentation of its host. *Biological Journal of the Linnean Society*, **90**, 239–246.

Salvato P, Battisti A, Concato S *et al*. (2002) Genetic differentiation in the winter pine processionary moth (*Thaumetopoea pityocampawilkinsoni* complex), inferred by AFLP and mitochondrial DNA markers. *Molecular Ecology*, **11**, 2435–2444.

Salvato P, Simonato M, Zane L *et al*. (2005) Do sexual pheromone traps provide biased information of the local gene pool in the pine processionary moth? *Agricultural and Forest Entomology*, **7**, 127–132.

Schimitschek E (1944) *Die Forstinsekten der Türkey und ihre Umwelt*. Volk und Reich, Prag.

Shapiro B, Rambaut A, Drummond A (2006) Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution*, **23**, 7–9.

Simmons AH (1999) *Faunal Extinction in an Island Society: Pygmy Hippopotamus Hunters of Cyprus*. Springer, Berlin.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Stauffer C, Lakatos F, Hewitt GM (1999) Phylogeography and postglacial colonization routes of *Ips typographus* L. (Coleoptera, Scolytidae). *Molecular Ecology*, **8**, 763–773.

Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.

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

Tams WHT (1925) A new processionary month (Notodontidae) injurious to pine trees in Cyprus. *Bulletin of Entomological Research*, **15**, 293–294.

Tarkhnishvili D, Hille A, Böhme W (2001) Humid forest refugia, speciation and secondary introgression between evolutionary lineages: differentiation in a Near Eastern brown frog, *Rana macrocnemis*. *Biological Journal of the Linnean Society*, **74**, 141–156.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789–809.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Thompson JN, Pellmyr O (1991) Evolution of oviposition behavior and host preference in Lepidoptera. *Annual Review of Entomology*, **36**, 65–89.

Turkmen H, Oner YA (2004) A human dermatitis caused by *Thaumetopoea pityocampa* (Denis and Schiffermuller, 1775) (Order: Lepidoptera) caterpillars in Istanbul, Turkey. *Allergy*, **59**, 232–233.

Veith M, Schmidtler JF, Kosuch J, Baran I, Seitz A (2003) Palaeoclimatic changes explain Anatolian mountain frog evolution: a test for alternating vicariance and dispersal events. *Molecular Ecology*, **12**, 185–199.

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

Vidakovic M (1991) *Conifers. Morphology and Variation*. Grafiècki zavod Hrvatske, Zagreb, Croatia.

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

West SA, Cook JM, Werren JH, Godfray HCJ (1998) *Wolbachia* in two insect host-parasitoid community. *Molecular Ecology*, **7**, 1457–1465.

Wilf P, Labandeira CC (1999) Response of plant–insect associations to Palaeocene-Eocene warming. *Science*, **284**, 2153–2156.

Wilkinson DS (1927) The Cyprus processionary caterpillar (*Thaumetopoea wilkinsoni*, Tams). *Bulletin of Entomological Research*, **18**, 173–182.

Willis KJ, Bennet KD, Birks HJB (1998) The late Quaternary dynamics of pines in Europe. In: *Ecology and Biogeography of Pinus* (ed. Richardson DM), pp. 107–121. Cambridge University Press, Cambridge, UK.

Yang Z (1997) PAML: a program for package for phylogenetic analysis by maximum likelihood. *Cabios*, **15**, 555–556. (http://abacus.gene.ucl.ac.uk/software/paml.html).

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, **11**, 413–418.

Mauro Simonato is a doctoral student working on population genetics and evolution of *Thaumetopoea* species, under the supervision of Andrea Battisti and Lorenzo Zane. Zvi Mendel is an entomologist specialized on forest insects of the Near East. Carole Kerdelhué, Jérôme Rousselet, Emmanuelle Magnoux and Paola Salvato work on phylogeography and molecular evolution of insects. Alain Roques is an entomologist specialized on forest insects.

Supplementary material

The following supplementary material is available for this article:

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/
10.1111/j.1365-294X.2007.03302.x
(This link will take you to the article abstract).

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

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Chapter 3

Quaternary history and contemporary patterns in a currently expanding species

---

Published as:

Kerdelhué C., Zane L., Simonato M., Salvato P., Rousselet J., Roques A., Battisti, A. (2009) Quaternary history and contemporary patterns in a currently expanding species. *BMC Evolutionary Biology* **9**: 220 doi:10.1186/1471-2148-9-220

I contributed to all parts of the experimental work and analysis and to some parts of the writing.

BMC Evolutionary Biology

BioMed Central

Research article

Open Access

Quaternary history and contemporary patterns in a currently expanding species

Carole Kerdelhué*†[1], Lorenzo Zane†[2], Mauro Simonato[3], Paola Salvato[3], Jérôme Rousselet[4], Alain Roques[4] and Andrea Battisti[3]

Address: [1]INRA, UMR1202 BIOGECO, F-33610 Cestas, France, [2]Dipartimento di Biologia, Università di Padova, 35121 Padova, Italy, [3]Dipartimento di Agronomia Ambientale e Produzioni Vegetali, Agripolis, Università di Padova, 35020 Legnaro PD, Italy and [4]INRA, UR633 Zoologie Forestière, F-45075 Orléans Cedex, France

Email: Carole Kerdelhué* - Carole.Kerdelhue@pierroton.inra.fr; Lorenzo Zane - lorenz@unipd.it; Mauro Simonato - mauro.simonato@unipd.it; Paola Salvato - paola.salvato@unipd.it; Jérôme Rousselet - Jerome.Rousselet@orleans.inra.fr; Alain Roques - Alain.Roques@orleans.inra.fr; Andrea Battisti - andrea.battisti@unipd.it

* Corresponding author †Equal contributors

Published: 4 September 2009

BMC Evolutionary Biology 2009, **9**:220 doi:10.1186/1471-2148-9-220

Received: 13 January 2009
Accepted: 4 September 2009

This article is available from: http://www.biomedcentral.com/1471-2148/9/220

© 2009 Kerdelhué et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Quaternary climatic oscillations had dramatic effects on species evolution. In northern latitudes, populations had to survive the coldest periods in refugial areas and recurrently colonized northern regions during interglacials. Such a history usually results in a loss of genetic diversity. Populations that did not experience glaciations, in contrast, probably maintained most of their ancestral genetic diversity. These characteristics dramatically affected the present-day distribution of genetic diversity and may influence the ability of species to cope with the current global changes. We conducted a range-wide study of mitochondrial genetic diversity in the pine processionary moth (*Thaumetopoea pityocampa/T. wilkinsoni* complex, Notodontidae), a forest pest occurring around the Mediterranean Basin and in southern Europe. This species is responding to the current climate change by rapid natural range expansion and can also be accidentally transported by humans. Our aim was to assess if Quaternary climatic oscillations had a different effect across the species' range and to determine if genetic footprints of contemporary processes can be identified in areas of recent introduction.

Results: We identified three main clades that were spatially structured. In most of Europe, the genetic diversity pattern was typical for species that experienced marked glaciation cycles. Except in refugia, European populations were characterized by the occurrence of one main haplotype and by a strong reduction in genetic diversity, which is expected in regions that were rapidly re-colonized when climatic conditions improved. In contrast, all other sub-clades around the Mediterranean Basin occurred in limited parts of the range and were strongly structured in space, as is expected in regions in which the impact of glaciations was limited. In such places, genetic diversity was retained in most populations, and almost all haplotypes were endemic. This pattern was extreme on remote Mediterranean islands (Crete, Cyprus, Corsica) where highly differentiated, endemic haplotypes were found. Recent introductions were typified by the existence of closely-related haplotypes in geographically distant populations, which is difficult to detect in most of Europe because of a lack of overall genetic structure.

Conclusion: In regions that were not prone to marked glaciations, recent moth introductions/expansions could be detected due to the existence of a strong spatial genetic structure. In contrast, in regions that experienced the most intense Quaternary climatic oscillations, the natural populations are not genetically structured, and contemporary patterns of population expansion remain undetected.

Background

Past climate changes have had dramatic impact on the geographic distribution, demography, and thus the evolution of species. The contemporary distribution of genetic diversity cannot be understood without studying how organisms responded to climate over geological times. Many terrestrial species are today responding to the contemporary global warming [1], and their future response will at least partially depend on their previous reactions to climatic oscillations. The 'genetic legacy of the Quaternary ice ages' [2], *i.e.* the genetic footprint of species' responses to glacial-interglacial successions, has been extensively studied on many species in Europe and North-America, that is, in the geographical regions where glaciations were most intense [3,4]. Forest insect herbivores, such as those associated with oaks and pines in Europe and the Mediterranean, for example, are known to have responded to post-glacial warming with rapid range expansion northwards and eventually westwards, and to have survived glaciations in southern refugia [5-10]. The intensity of the oscillations increased with latitude, which affected the impact they had on species occurring through a gradient in the so-called ORD (Orbitally forced species Range Dynamics: see [11]).

Following Pinho and collaborators [12], we can make two predictions. In northern latitudes, where the effects of glaciations were more severe, fewer and smaller patches of suitable habitat were left for the survival of populations across multiple glaciation cycles, which would have resulted in overall lower diversity, and a lower number of differentiated lineages in northern than in southern areas. Moreover, the effects of climatic changes on the effective population sizes were more dramatic in northern than in southern regions, meaning that northern populations should bear the signature of a rapid demographic expansion following the climate amelioration, whereas southern populations should evidence marks of more stable, long-term effective population sizes.

Going further, Dynesius and Jansson [11] have predicted differential evolutionary consequences depending on the intensity of the ORD, and these predictions were empirically demonstrated for some taxa. Species that survived a strong ORD during the Quaternary, *i.e.* species occurring at higher latitudes, were selected for increased vagility and generalism. Dispersal-related traits should have been optimized during the northward progression because high mobility provided an elevated fitness within populations that were tracking a moving habitat. In the same way, generalists (in terms of habitat, host, or diet) had a smaller risk of their niche disappearing. Over evolutionary times, the selective pressures are likely to have changed, with dispersion and generalism favoured during interglacials, and less so during glacial periods when the species were restricted to suitable refugia.

The effects of differential intensities of glaciations on the evolution of the species, described above, are expected for mainland species for which the tracking of acceptable environments through migration was possible. The situation was drastically different for species or populations on islands situated beyond dispersal range, for which any change had to be endured locally, either by altitudinal shifts or by the evolution of local adaptations. Moreover, smaller effective population sizes could have resulted in loss of genetic diversity due to genetic drift. In this case, evolution on islands may have been more rapid than the rate of change on continents [13], and island populations are thus expected to be highly differentiated from both a genetic and an ecological point of view.

Species or populations that experienced marked climatic oscillations in the past can be seen as a selected assemblage of geographically mobile and latitudinally-independent organisms that are likely to be best adapted for the future climate changes, unless human activity precludes such an option [13]. Yet, comparing the phylogeographic patterns of species occurring over a latitudinal gradient is not straightforward, as other important factors such as life-history traits, ecological requirements, and dispersal ability will probably differ among species. Moreover, data on current modifications to distribution ranges due to global changes are also required to link differential Quaternary histories to present-day evolution.

Here, we present a range-wide genetic study of a circum-Mediterranean insect taxon: the winter pine processionary moth (*Thaumetopoea pityocampa*/*wilkinsoni* species complex), which develops mainly on pine species (*Pinus* spp.). It is a serious forest pest as it can cause heavy defoliations of pines in Mediterranean countries. *T. pityocampa* has a typical winter larval development [14]. Adults lay eggs on pine leaves in summer, and larvae feed from needles during fall and winter. They pupate in the soil in late winter or early spring, and newly emerged adults disperse to reproduce during summer. Larvae are gregarious and develop in a typical silk shelter. Ecological and genetic data based on mitochondrial and nuclear markers suggest that the species exhibits clear sex-biased dispersal, as females are poorer fliers than males [15,16]. It is present on both the northern and southern rims of the Mediterranean Basin as well as in the Middle-East (Figure 1), that is, in regions where the impacts of glaciations varied in intensity. Glacial cycles were probably most intense in temperate Europe, while ice sheets are believed not to have occurred in southern Mediterranean countries, nor in the Near East. Populations of the pine processionary moth are currently believed to belong to a species com-

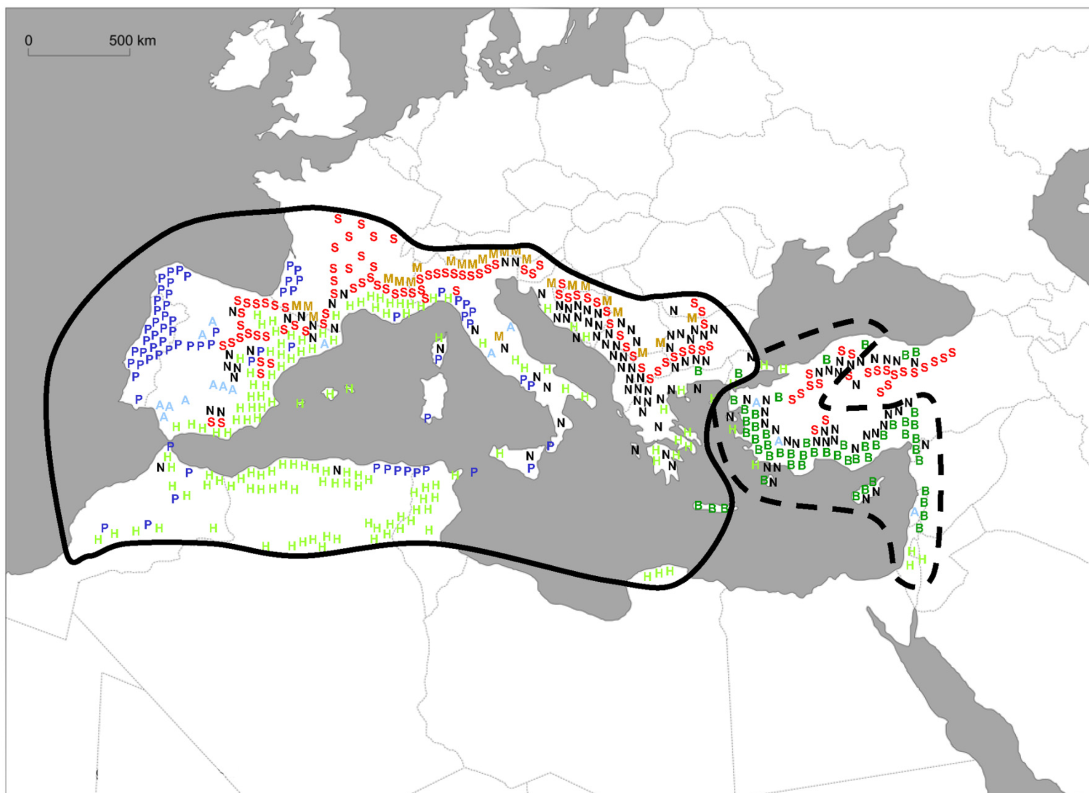


Figure 1
Ranges of the pine processionary moths indicating the occurrence of native *Pinus*. *Thaumetopoea pityocampa*, solid line; *Thaumetopoea wilkinsoni*, dashed line; A = *Pinus pinea*, B = *P. brutia*, H = *P. halepensis*, M = *P. mugo*, N = *P. nigra*, P = *P. pinaster*, S = *P. sylvestris*. Each letter refers to a land unit where the indicated pine species is dominant but not necessarily exclusive. Other pine species may occur in the same area. *Thaumetopoea* distribution was drawn from: Anonymous (1977) Pest: *Thaumetopoea pityocampa* (Schiff.) (Lep., Notodontidae) (Pine processionary moth). *Distribution Maps of Pests, CAB*, **366**, 1-2. and *Pinus* distribution from: Richardson DM (1998) *Ecology and Biogeography of Pinus*. Cambridge University Press, Cambridge, UK.

plex including two congeneric taxa: *Thaumetopoea pityocampa* and *T. wilkinsoni*. The differentiation between these two species was recently shown [17], and the monophyly of *T. wilkinsoni* populations in the near East was confirmed [16].

Current global changes can affect the genetic patterns of the pine processionary moth in different ways, and superimpose new signatures on existing natural phylogeographical structure. An increase in mean winter temperatures in Europe is known to drive moth expansion northward and to higher altitudes, in regions where hosts are available, by providing suitable conditions in places were larvae could not previously have survived [18]. And if environmental conditions are suitable for the insect's development, new pine plantations can also increase the potential range of the pest by offering hosts in places where they were not previously available. Contemporary changes in the moth's distribution range can proceed either from a natural, non-assisted expansion of insect populations into newly suitable habitats, or from long-range dispersal that is likely to be human-aided (accidental transportation of adults or larvae, or transplantation of buried pupae when mature trees are planted). In cases of natural expansion, we expect a gradual loss of diversity away from the native range (e.g., [10,19]), while long-distance, assisted introductions should result in a discontinuous distribution of genetic diversity. A recent study of the range-wide genetic structure of the oak gall-wasp *Andricus kollari* showed that the patterns observed in England were

consistent with the hypothesis of man-aided, long-distance introductions [7].

The aim of our study was to infer the Quaternary history of the species complex over its whole distribution range, to test if the effects of Quaternary climatic oscillations can be differentially detected in the different parts of the range, and if any impact of global change can be detected and interpreted in the light of the species' evolutionary history. Both mitochondrial and nuclear markers are useful to reconstruct the evolutionary history of a species complex. Although nuclear markers such as AFLPs and microsatellites were previously developed for this species [15-17], we were not able to use them in this range-wide study because of homoplasy and because of the occurrence of null alleles in divergent clades. We thus present data based on mitochondrial DNA alone. As female dispersion is the limiting factor for species expansion, inferring the history of female lineages provides a good indication of species dispersal. Yet, potential biases due to the use of mitochondrial markers alone, such as the selective sweep that can be caused by bacterial symbionts [20], as well as the limits inherent to single gene phylogenies, should be acknowledged.

Results

We obtained 34 COI and 51 COII haplotypes. Among these, 14 COI and 21 COII haplotypes were known from either Salvato et al. [17] or Simonato et al. [16] and were already available in GenBank (accession numbers EF015538-EF015549 and EF210075-EF210097). The new haplotypes found in the present study have been deposited in GenBank (accession numbers GQ507373 to GQ507422). A total of 67 combined (COI-COII) haplotypes (ht) were found. The selected model of evolution was the General Time Reversible model with gamma distributed heterogeneity of rates (GTR gamma). Interestingly, Bayes factors (BF) indicated a much stronger fit for this model when a clock was assumed than when branch lengths were unconstrained (BF = 142, computed as twice the difference in logarithm of harmonic means of likelihoods). This was confirmed when the performance of models was assessed with the Bollback approach [21]. The GTR gamma model was then used for all subsequent analyses. The specific rates were A-C: 0.144; A-G: 1.166; A-T: 0.068; C-G: 0.031; G-T: 0.019 and $\alpha = 0.152$.

Phylogenetic inference and node datation

The haplotype composition of each sampled population is given in Additional file 1 (Sampling sites, geographic coordinates, host pine, collector and haplotype composition of each locality). The phylogenetic analysis clearly showed that the *T. pityocampa - wilkinsoni* complex was structured in three strongly supported clades (Figure 2). A first group of 23 ht clustered all sequences corresponding to *T. wilkinsoni* [16] together with the ht found on the island of Crete. This '*wilkinsoni* clade' was the sister group of all other ht. A second clade of 13 ht was restricted to Libya, Tunisia (including the nearby Italian island of Pantelleria) and North Algeria ('Eastern North Africa clade', hereafter ENA clade). The third clade comprised 24 European ht, from Spain and Portugal to Greece (with the notable exception of Crete), together with the 7 ht found in Morocco and South Algeria. It will hereafter be referred to as the '*pityocampa* clade'. The main nodes were dated by Bayesian inference using a Yule prior and the estimates are given on the phylogenetic tree (Figure 2) and in Table 1, with 95% confidence intervals (CI). The split between the *wilkinsoni* clade and the 2 others was ca. 7.5 Million years ago (Myrs; 95% CI 5.8 - 9.3), while the separation of the *pityocampa* vs. ENA clades was dated back to 6.7 Myrs (4.9 - 8.6). The age of the most recent common ancestor (MRCA) of the *wilkinsoni* clade was estimated to 5.3 Myrs (3.7 - 7.1) while that of the ENA clade was ca. 3.1 Myrs (2.1 - 4.3) and that of the *pityocampa* clade was estimated to 2.3 Myrs (1.6 - 3.1).

Further geographic structure was found within the three main clades. In the *wilkinsoni* clade, 4 distinct sub-clades were found with very high support values (Figure 2). The Cretan sub-clade formed the sister group of all other ht. The Cypriot ht were the sister group of the North & West

Table 1: Age estimates of phylogenetic tree nodes and 95% confidence intervals.

Node code	Estimated age of the node (in Myrs)	95% confidence interval (in Myrs)
A	7.450	5.776 - 9.271
B	6.742	4.892 - 8.613
C	2.348	1.631 - 3.124
D	1.772	0.921 - 2.725
E	3.146	2.104 - 4.298
F	1.364	0.766- 2.025
G	5.332	3.688 - 7.067
H	1.846	1.210 - 2.545
I	1.259	0.742 - 1.060

Estimations were performed by analysing all the haplotypes and assuming a Yule prior. The node codes are given in Figure 2.

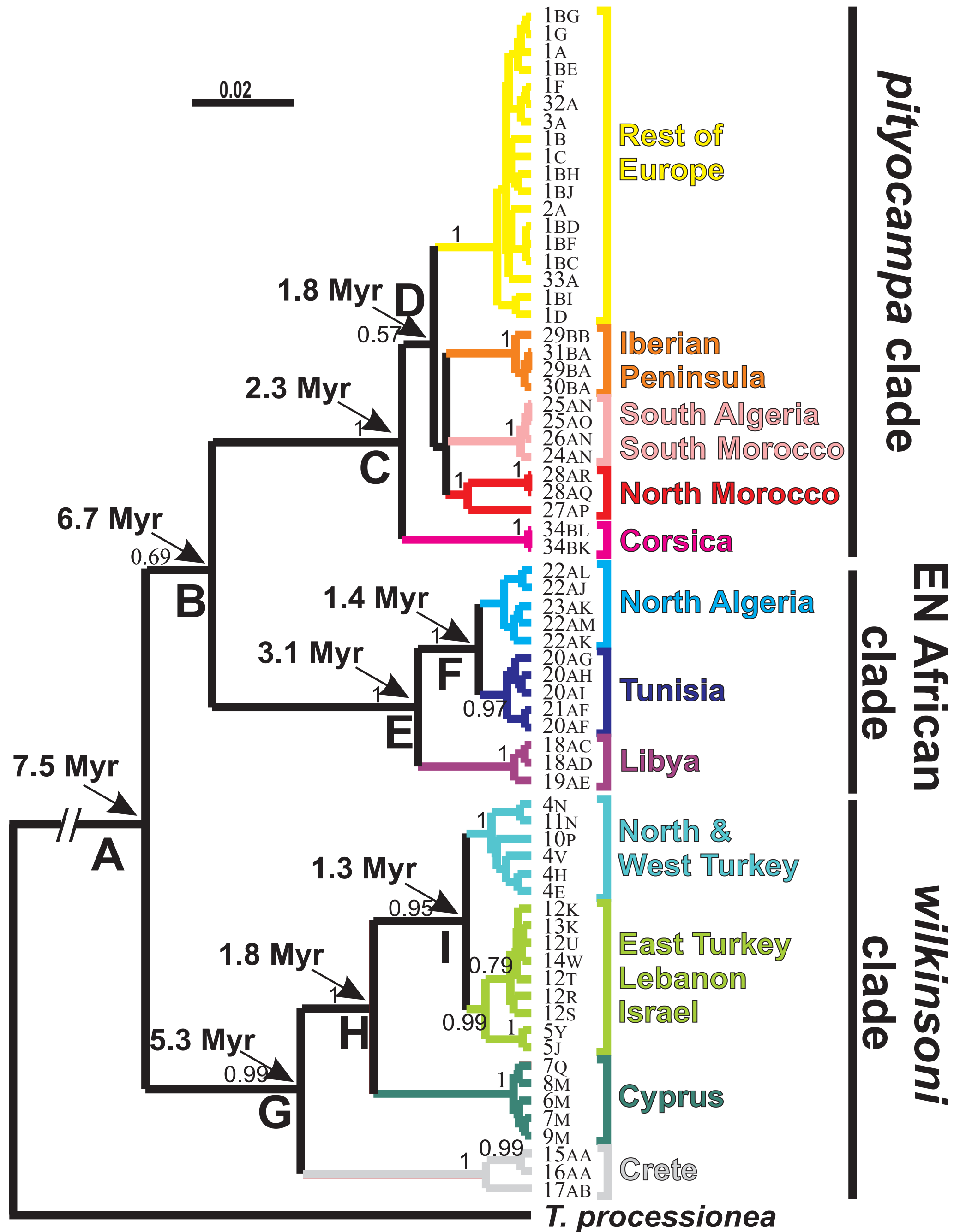


Figure 2

Bayesian consensus tree for all Mediterranean *Thaumetopoea pityocampa* and *T. wilkinsoni* haplotypes rooted on *T. processionea*. Bayesian supports over 0.5 are given. The arrows show the estimated age of the most recent common ancestors (in million years) of the deeper supported nodes. Age estimates and their corresponding 95% confidence intervals are given in Tables 1 & 2.

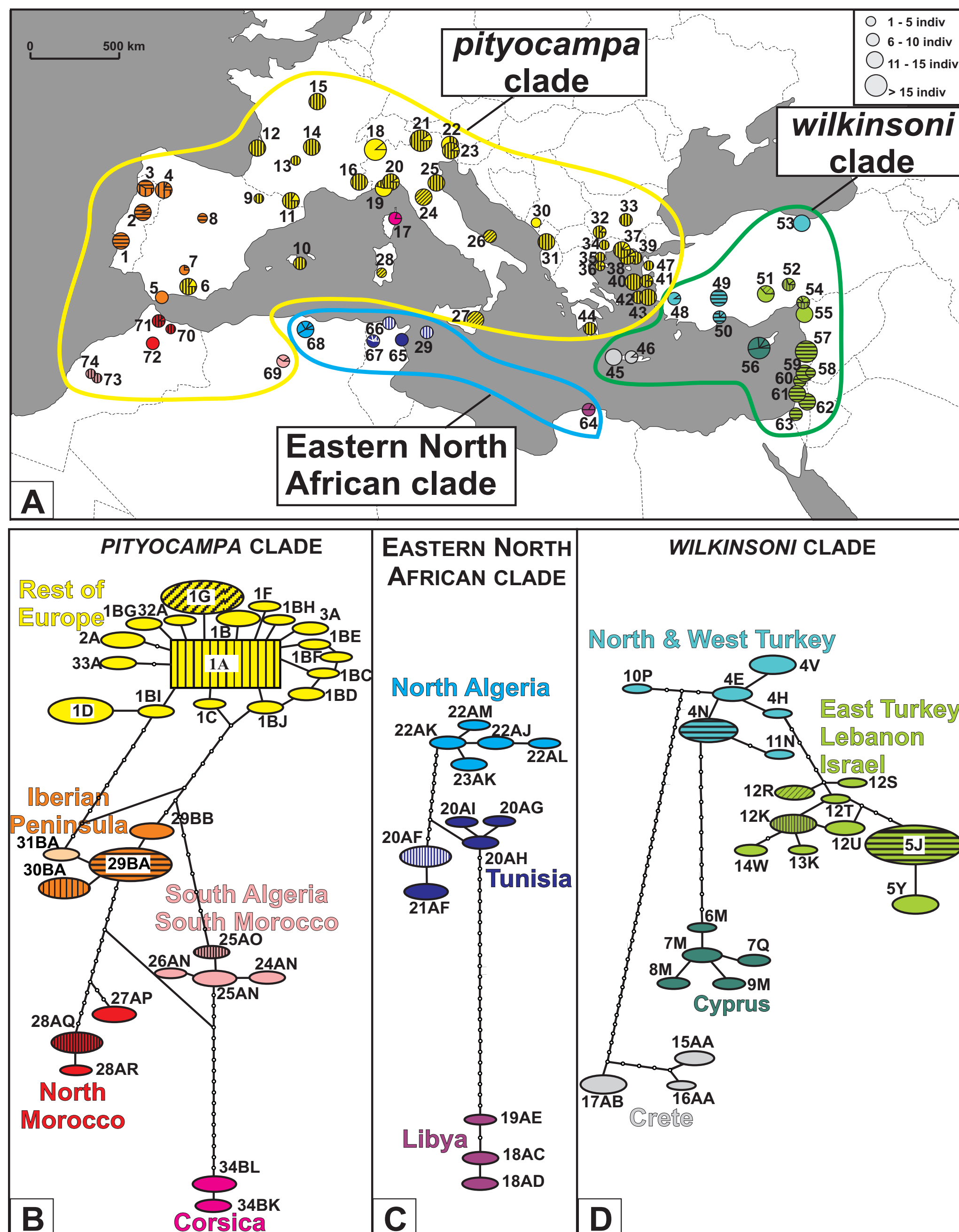


Figure 3

Geographical distribution of mitochondrial haplotypes of the *Thaumetopoea pityocampa/T. wilkinsoni* complex, and within-clade haplotype networks. A. Geographical mapping of haplotypes in the sampled populations. Circles are proportional to the number of individuals analyzed in each population and colors refer to the major clades identified in network analyses. Codes of populations are given in Additional file 1. B. Haplotype network of the '*pityocampa*' clade. Each line in the network represents a single mutational change. Empty circles indicate intermediate, missing haplotypes. C. Haplotype network of the 'Eastern North Africa' clade. D. Haplotype network of the '*wilkinsoni*' clade.

Table 2: Estimates of tMRCAs of the most recent nodes (main sub-clades) and 95% confidence intervals.

Sub-clade	tMRCA (in Myrs)	95% confidence interval (in Myrs)	Tree Prior
Rest of Europe	0.090	0.028 - 0.172	Exponential*
Iberian Peninsula	0.091	0.005 - 0.201	Constant
South Algeria - South Morocco	0.130	0.011 - 0.290	Constant
North Morocco	0.532	0.194 - 0.905	Constant
Corsica	--**		Constant
North Algeria	0.171	0.026 - 0.355	Constant
Tunisia	0.326	0.092 - 0.601	Constant
Libya	--**		Constant
N & W Turkey	0.332	0.114 - 0.608	Constant
E. Turkey, Lebanon, Israel	0.417	0.148 - 0.711	Constant
Cyprus	0.151	0.021 - 0.313	Constant
Crete	0.381	0.116 - 0.688	Constant

Estimates were obtained by assuming a coalescent prior of constant size or exponential growth and by including all the sequences of each given sub-clade. Names are the same as in Figure 2.
* Results based on an exponential prior are reported because the rate of exponential growth (g) was significantly higher than 0 for this group.
** MCMC did not converge due to small sample size (N = 10 and 6 for Corsica and Libya, respectively)

Turkey sub-clade and of the sub-clade grouping the ht from East Turkey, Lebanon and Israel. The ENA clade was divided in 3 sub-clades corresponding to the 3 countries in which the larvae were sampled. The Libyan ht formed a sister group relative to the North Algerian and to the Tunisian sub-clades. Finally, the *pityocampa* clade was comprised of five strongly supported geographical groups. Haplotypes from the island of Corsica appeared as the sister-group of the four remaining sub-clades: (i) the Iberian Peninsula, (ii) North Morocco, (iii) South Morocco & South Algeria and (iv) the Rest of Europe. Interestingly, 16 individuals sampled in Spain and 2 individuals from western Turkey had "Rest of Europe" haplotypes (Figure 3).

Time of most recent ancestor (tMRCA) of each sub-clade was estimated by Bayesian inference including all the individuals of a given group and assuming a priori either a constant population size or an exponential growth. However, the rate of exponential growth resulted to be positive, with an associated 95% confidence interval excluding 0, for the Rest of Europe only ($g = 6.2\ 10^{-5}$ yrs^{-1}; 95% CI: $6.2\ 10^{-6}$ - $1.4\ 10^{-1}$ yrs^{-1}); tMRCAs obtained with an exponential prior are reported only for this group (Table 2). Keeping in mind that our use of a rate from phylogenetics studies will bias estimates upwards, estimated ages for tMRCA resulted to range from 532 000 years ago for North Morocco to 90 000 years ago for Rest of Europe.

Haplotype distribution and haplotype network

Haplotype networks were reconstructed for each of the 3 main clades, and haplotype distributions were mapped (Figure 3). The haplotype networks recovered the same strong geographical patterns as the phylogenetic tree. Within the *wilkinsoni* clade, most ht were found in a single population, except ht 5J (found throughout Lebanon and Israel and shared by 91 individuals), and ht 4N, 12R and 12K that each occurred in two populations (see Addi-

Table 3: Indices of genetic diversity per identified sub-clade, Tajima's D and Fu's Fs statistics

Sub-clade	N	Hd	π	Tajima's D	Fu's Fs
Corsica	10	0.36	0.06%	0.01 NS	0.42 NS
South Algeria - South Morocco	13	0.73	0.15%	- 0.14 NS	- 0.69 NS
North Morocco	24	0.55	0.69%	2.35 NS	7.22 NS
Iberian Peninsula	61	0.60	0.12%	0.18 NS	0.10 NS
Rest of Europe	358	0.44	0.11%	- 1.82 **	- 15.82 **
North Algeria	12	0.79	0.19%	- 0.54 NS	- 1.61 NS
Tunisia	30	0.62	0.30%	0.10 NS	1.04 NS
Libya	6	-	-	-	-
Crete	21	0.55	0.52%	2.03 NS	5.26 NS
Cyprus	19	0.72	0.15%	-0.60 NS	- 1.42 NS
North & West Turkey	45	0.68	0.20%	- 0.93 NS	- 0.56 NS
East Turkey, Lebanon, Israel	133	0.51	0.35%	- 0.10 NS	0.44 NS

N: # individuals; Hd: gene diversity; π: nucleotide diversity per site. NS: non significant; **: $p < 0.001$. The names of the sub-clades are the same as in Figure 2.

tional file 1, sampling sites, geographic coordinates, host pine, collector and haplotype composition of each locality). All but one ht (20AF, found in Tunisia and Pantelleria) in the ENA clade were endemic to one population. Finally, the *pityocampa* clade was divided into the 5 sub-clades found on the phylogenetic tree. The network of the Rest of Europe sub-clade was star-shaped (which is typical for expanding populations), with one main ht shared by ca. 74% of sampled individuals and all other ht diverging from it by only one or two mutations. Haplotype 1G was restricted to central and southern Italy. Interestingly, all other haplotypes in Europe were rare, shared by 6 individuals at most and usually endemic to one population. In the Iberian Peninsula, ht 29BA was found in 57% of individuals and in all populations but Gibraltar and two southern sites. None of the other sub-clades in the *pityocampa* group showed a star shape.

Within population, gene diversity H and nucleotide diversity per site π are given in Additional file 1. Fu's Fs and Tajima's D were estimated and tested within each of the 12 sub-clades except for the Libyan group as it was composed of only 6 individuals. Both indices were significantly negative only in the Rest of Europe sub-clade (see Figures 2 &3 and Table 3). Mismatch analyses were consistent with the sudden expansion model for this sub-clade (SSD = 0.00298, P = 0.746) and showed a unimodal distribution that closely fit the expected distribution. In this sub-clade, τ was estimated to 1.77 (95% CI: 0 - 4.20), and the corresponding expansion was thus estimated to date back ca. 147 000 years (95% CI: 0 - 348 261 years).

Discussion

Overall phylogenetic patterns around the Mediterranean Basin

The pine processionary moth is currently understood to consist of a species complex containing two taxa, namely *Thaumetopoea pityocampa* and *T. wilkinsoni* [16,17]. Surprisingly, the thorough sampling we obtained clearly proved that the species complex is composed of three rather than two main clades, as the populations from ENA appeared as the monophyletic sister group of the *pityocampa* clade (Figure 2), and the *wilkinsoni* clade (including populations from Crete) as the sister group of the ENA and *pityocampa* clades. Determining the taxonomic status of the clusters identified here is beyond the scope of the present study, and would need complementary data such as nuclear markers and morphological data. For this reason, we will hereafter mention three clades (the *pityocampa* clade, the ENA clade and the *wilkinsoni* clade) without further discussion of their taxonomic level.

Another striking result is that the species complex is ancient, and predates the Quaternary by a few million years. In a previous study, the divergence between *Thaumetopoea pityocampa* and *T. wilkinsoni* was estimated to 4.5 - 5.2 Myrs [17]. That result was obtained from a limited sampling, in which only the European clade of *T. pityocampa* (as the population from Spain contained only European haplotypes rather than Iberian ones) and one Turkish population of *T. wilkinsoni* were analyzed. In the study presented here, a thorough sampling of populations (including the Cretan lineage of *wilkinsoni*, as well as most sub-clades of *pityocampa* and the previously unknown ENA clade) and a Bayesian approach taking into account the gamma-distributed heterogeneity of rates, allowed us to obtain a different estimate for the age of the main evolutionary events. In particular, the split between the *wilkinsoni* and the *pityocampa*-ENA clades was dated on average to 7.5 Myrs, with a confidence interval of 5.8 - 9.3 Myrs, which could correspond to the full opening of the Aegean Trench ca. 9 Myrs ago [4,22]. Interestingly, within the *wilkinsoni* clade, the estimates of node ages we obtained were very similar to estimates obtained previously using codon-partitioned models [16]. While we did not have enough a priori evidence to calibrate our own molecular clock, it should be noted that, by using the universal rate, the divergence of Crete from all the other *wilkinsoni* haplotypes was dated back to about 5.3 Myrs, which corresponds to the Messinian salinity crisis and the time when the Mediterranean Sea was at its lowest level, thus making the colonization of islands easier [23]. Node ages should, however, always be interpreted with caution, given that a single mitochondrial locus was used [20].

The differentiation between the *pityocampa* and the ENA clades was unexpected, and cannot be explained by classical barriers to gene flow such as mountain ranges or fragmentation of suitable habitats. Similar patterns of East-West genetic differentiation have occasionally been found in North Africa for other organisms [24-27], but were estimated to date back to various times, from 1.6 to 12 Myrs. A range of hypotheses have been proposed by the authors to explain the abrupt genetic differentiation within species in this region. They invoked either climatic scenarios, with the rapid alternations of arid and humid periods acting as a spatially structuring force in this region during the Quaternary; or biogeographical scenarios such as the formation of the Straits of Gibraltar after the Messinian salinity crisis, the split of the Tellian (Tell) Atlas at the Sicilian Channel, or the more ancient formation of the Neo-Pyrenees. Indeed, the pine processionary moth depends on the presence of pine hosts for development, and it is known to be susceptible to summer aridity and excessive heat [28]. Moreover, it was recently shown that barriers of moderate altitude can hamper gene flow in this species [29]. Finally, the species also exhibits large among-population variation in term of reproductive phenology [28] that permits the adaptation of populations to the local climatic conditions and may also limit gene flow. Thus, the

conjunction of major biogeographical events (the rise of the Tellian Atlas) and late Tertiary climatic change (with a possible gap in host availability during more arid phases) could explain the split that occurred between the *pityocampa* and the ENA clades some 6-7 Myrs ago.

If the main divergences within the *T. pityocampa/wilkinsoni* complex date from the end of the Miocene, all clades also predate the Quaternary. Each of the identified clades thus experienced the Quaternary climatic oscillations after they split from a common ancestor, and the impact of ice ages can easily be compared between these closely-related clusters.

Phylogeographical patterns and within-clade structures

Each of the three identified clades showed a strong phylogeographical structure, and was composed of 3, 4 or 5 well-differentiated sub-clades. With the notable exception of the Rest of Europe (see below), each sub-clade was restricted to a rather narrow geographical region. Interestingly, a vast majority of haplotypes (54 out of 67) were endemic to one single population, and only five were found in three or more populations. Thus, the pine processionary moth exhibits an extreme spatial structure and a highly reduced mitochondrial female gene flow even on a regional scale, even though results based solely on a mitochondrial marker should be interpreted with caution. Over most of the distribution range, the actual dispersal of the females is thus highly limited. The main barriers to gene flow are sea straits, mountain ranges (the Pyrenees, Taurus Mountains, High Atlas, Saharian Atlas), or desert regions where hosts are lacking (Libya).

Within-clade structures were all dated back to at least 1.3 Myrs (Figure 2), *i.e.* to the Early- or Mid- Pleistocene. One could suggest that local ecological pressures recurrently acted to reinforce and maintain the genetic structures whenever gene flow had been interrupted. As migration is very limited and cannot counteract the effects of drift, genetic differentiation then simply increases with time, leading to divergent lineages in different regions. Ecological factors involved in differentiation include reproductive phenology, which can prevent mating by shifting adult emergence periods in different populations, or local adaptation to host characteristics, which, it has been proved, can lead to complete mortality in translocated larvae [30]. A more precise sampling in North Africa would allow the delimitation of the exact distribution ranges of each sub-clade, and the determination of whether contact zones do exist between them.

Once again with the exception of the Rest of Europe sub-clade, a majority of the sampled populations in the natural area of the species show more than one haplotype, even when only 5-10 individuals were sampled, and even at the edge of the distribution or in very isolated places such as Libya or on remote islands. Like many insects, the processionary moth has evolved the capacity of prolonged diapause, which allows the emergence of adults of the same generation over several years, thus limiting the risk of local extinction and increasing the probability of retaining local genetic diversity. A high genetic diversity in the southernmost populations has also been observed for other Mediterranean insect species (e.g., [8,9,31,32]). Interestingly, no sign of demographic expansion could be detected in these regions, as is expected in regions where glaciations were less intense [12]. However, one region in the Near East is characterized by an extreme genetic depauperization as one single haplotype is present in Lebanon and Israel. This is probably linked to the very recent origin of moth populations in this region, where pine trees were not present before the beginning of the XXth century except for remote relictual stands (see Simonato et al. [16] for a detailed discussion). The moth has expanded slowly following afforestation. Recent expansions due to global changes are discussed below.

Europe (except the Iberian Peninsula that harbours a specific sub-clade) is characterized by a major haplotype that occurs from the Atlantic coast to the Greek islands and even along the Turkish border. Moreover, the Rest of Europe sub-clade had the star-shape that is typical for populations expanding after a demographic bottleneck [33], and the Bayesian analyses indicated for this group a positive exponential growth supporting a past demographic expansion [34,35]. Tajima's D and Fu's Fs statistics revealed an excess of rare haplotypes and allowed us to reject mutation-drift equilibrium. As similar results can be obtained from different processes (see for instance [36,37]), we conducted a mismatch analysis that also indicated that European populations underwent bottleneck events due to the recurrent glaciation periods and then recurrently expanded after the retreat of the ice. Such results are classically found for temperate and cold-sensitive species in this region [4,9,10]. The spatial distribution of the rare haplotypes gives insights into the existence and locality of refugial areas where the moths survived the glaciations, and possibly also the interglacials as this Mediterranean species is susceptible to both winter cold and summer heat and aridity [28]. As for most of the European temperate species, these moth refugia are located in the Balkans and in Italy, as well as in the western part of the Iberian Peninsula [4]. Our results also show that the Alps and the North of Italy form a region with a high proportion of endemic haplotypes, thus differing from all other regions in Europe. This could indicate that this area also was a Quaternary refugium where part of the ancestral polymorphism was locally retained. Interestingly, the Alpine Arc was recently proved to be a refugial area for *Pinus sylvestris* [38], which suggests that the refugial moth

populations could have survived the glacial maxima in this region on that particular host.

With the exception of Lebanon and Israel where the moth settled and expanded only recently (see below), our results show contrasting patterns of evolution during the Quaternary in the different regions of the moth's distribution range, corresponding to our expectations. In particular, populations occurring in the highest latitudes exhibit a radically different genetic footprint to that of all other sub-clades. If moth populations in the vast majority of the distribution range are characterized by a strong spatial genetic structure, a high number of endemic haplotypes and a restricted geographical range for each identified sub-clade, the patterns in the Rest of Europe are completely the opposite. In this European region, overall genetic diversity is low; spatial genetic structure is limited as a consequence of the large distribution of the major haplotype 1A; and this single sub-clade is distributed over one half of the total distribution area of the species complex. Moreover, signs of recent expansion were detected only in the European sub-clade, that is, in the region where glacial cycles were probably most intense. As for most European species, endemic haplotypes and some genetic variability can still be detected in plausible refugial areas near the Pyrenees, in Italy and in the Balkans [4,8,13]. In the rest of the area, the recurrent northward expansions that followed climate warming after glacial maxima were probably rapid, pioneer-like [39], and lead to a genetic homogenization of populations. In other temperate forest insect species, genetic diversity was also mostly retained either in the southernmost populations [9,31], or in the eastern regions where the impact of the Quaternary cycles was less pronounced (as for *Andricus* gall wasps developing on oaks, see [5,6,32]).

Evolution of insular populations

In each of the three main clades, the most divergent sub-clade corresponds to an island, or to an island-like continental region. The Corsican ht are the most differentiated within the *pityocampa* clade, the Cretan ht form the sister-group of all other sub-clades within the *wilkinsoni* clade, and the highly isolated moths of Cyrenaica (Libya) are most divergent in the ENA clade. Moreover, the second most differentiated group in the *wilkinsoni* clade is the Cypriot cluster. Each of the island lineages thus diverged from the corresponding sub-clade a long time ago (from 5.3 Myrs for the Cretan haplotypes to 1.8 Myrs for Cyprus). On the other hand, the most recent common ancestors for each island are much more recent (0.38 Myrs in Crete and 0.15 Myrs in Cyprus for example). Hence, it is not possible at this point to determine when exactly the colonization of each island (or isolated place) occurred, and for how long the moths have been isolated from the continent. However, even if we consider only the estimated age of the MRCA (which could be overestimated because we used a rate from phylogenetic studies, see [40], though the use of a Bayesian coalescent prior should in part address this problem), we can suggest that the pine processionary moths survived locally on these remote islands without female exchanges from the continent during few glacial cycles. As a consequence, they had to evolve locally to cope with at least some Quaternary oscillations and environmental changes [13]. The quite recent estimate for the age of MRCA for each island could be due to a founder effect followed by the effect of genetic drift in small populations [5], as well as by fixation of selected variation. We have evidence, in the pine processionary moth, that male gene flow have occurred between Cyprus and the continent [16], as was suggested by the strong genetic similarity between Cypriot and Turkish populations found with both AFLPs and microsatellite markers. This could also be true for islands situated at moderate distance from the continent.

Contemporary patterns in a historical context

In recent years, the distribution range of the processionary moth has been affected by global changes, mainly through winter warming [18] and pine afforestation. Moreover, it is suspected that human-aided dispersal occurs over various distances, either via 'hitch-hiking' (passive transportation of individuals) or accidental transplantation of pupae with grown trees moved with a substantial amount of soil. The genetic signatures of these contemporary events will be different, and may not be easy to detect in all regions. In most regions around the Mediterranean Basin, apart from Europe, the natural phylogeographic pattern consists of genetically diverse and spatially structured populations. Regions with surprisingly low levels of genetic diversity (e.g. Lebanon and Israel), or sampling sites that are genetically closely related to geographically distant populations (e.g. site 53 in Turkey, or 69 in Algeria) can be easily identified. These sites actually correspond to zones of recent moth expansion either following anthropogenic pine expansion, such as in Israel or Algeria where pines were planted both in the beginning and at the end of the XXth century, or following the ongoing climate warming that allows insects to survive winter in places where they could not some decades ago (site 53 near the Black Sea). Given the natural spatial genetic structure in these regions, the recent modifications in moth distributions due to global changes are actually easy to track. The populations discussed above all likely originated from the closest natural stand, and could be the result of non-assisted moth expansion (but a better sampling in Algeria is needed to confirm this). The mitochondrial marker we used here would also be useful to identify between-subclades female gene flow, but a nuclear marker is necessary to track male exchanges. In most of Europe, however, where the populations are not genetically struc-

tured in space and where overall genetic diversity is low, probably as a consequence of Quaternary history, one cannot distinguish recent and historical events, as contemporary expansions (proved at both higher latitudes and altitudes, see [18]) result in the loss of genetic diversity, as in the case of rapid, leptokurtic dispersal northwards that allowed re-colonization of northern habitats during interglacials [10,19].

The patterns are somewhat different for islands. Some harbour populations of moths that are genetically very close, or even similar, to their closest continental neighbours. This is not surprising for islands that are located very close to the continent, like most Greek islands or Sicily, that can probably be recurrently colonized from mainland sources. A similar result was found, for example, for rodents [41]. In contrast, one would expect the populations of Sardinia, Pantelleria, or the Balearic Islands, that are beyond the natural dispersal range, to be highly differentiated, as are the moths from Corsica, Cyprus or Crete. In Sardinia, pines are still very rare and, until recently, no pine processionary moths were found on the island. In 2004-2005, pines were transplanted from Tuscany and a population of the moth was detected the following year [42]. Not surprisingly, the moths sampled in Sardinia all bore the haplotype found in Tuscany, showing that the pests were accidentally introduced with their hosts. A similar hypothesis could be invoked to explain the occurrence of moths bearing the major haplotype 1A in the Balearic Islands, where the moth was first detected in the 1950s (G. Sanchez, pers. com.). The situation on the island of Pantelleria is different as genetic data show that pine trees (*Pinus pinaster*) occur naturally and exhibit a high degree of local genetic diversity [43]. In contrast to its pine host, the local moth population has low genetic diversity and bears the main Tunisian haplotype, suggesting that it was recently introduced.

Conclusion

We conducted a range-wide study of genetic diversity in a species complex occurring across regions in which Quaternary oscillations differed in intensity - or were absent. We have clearly shown that the sub-clade distributed over Europe had a phylogeographical pattern typical for species that experienced marked glaciation cycles. Refugial areas, where genetic diversity was retained and where endemic haplotypes were found, were identified in Italy, in the Alps and in the Balkans. All other populations were characterized by the occurrence of one main haplotype and by a strong reduction in genetic diversity, as is expected in regions that were rapidly re-colonized by a limited number of migrants when climatic conditions improved. We have ecological evidence that the moth populations are currently experiencing an expansion due to global change (both climate warming and host plantations). However, in the temperate regions of Europe, the natural populations are not genetically structured in space. The contemporary patterns are thus indistinguishable from historical ones as they also consist in progressions of the most widely distributed haplotypes. In contrast, all other sub-clades occur in limited ranges and are strongly structured in space, as is expected in regions that did not experience Quaternary cycles of glaciations. In these areas, genetic diversity has been retained in most populations, and each haplotype is usually found in only one population. The genetic signatures of recent moth introductions/expansions in these regions can be easily detected: recent expansions are characterized by the loss of genetic diversity across whole regions (e.g. Lebanon and Israel), and recent introductions are typified by the existence of closely related haplotypes in geographically distant populations. A strong differentiation is also expected for island populations if the island colonization occurred naturally in geological times. Thus, the occurrence (or not) of a significant 'natural' genetic structure of populations will determine whether or not recent expansions or introductions can be detected in the genetic data.

Complementary data based on polymorphic nuclear sequences would now be useful to compare biparental and maternally inherited markers, and to detect how male dispersal may have influenced the global evolutionary history of the species. Finally, our findings could be interesting for pest control as individuals present in different clades or sub-clades may have evolved different ecological characteristics (dispersal ability, host adaptation, egg size, resistance to parasitoids or pathogens), which can affect pest management strategies. Phenotypic traits should now be measured within each phylogenetic clade and sub-clade and compared between regions to test this hypothesis.

Methods

Moth sampling

Eggs and larvae of *Thaumetopoea pityocampa* and *T. wilkinsoni* were collected in 51 different locations from 16 countries in Europe and around the Mediterranean Basin. In addition, data from the 9 populations studied in Salvato et al. [17] and the 14 populations from Simonato et al. [16] were updated with newly sampled individuals and used here. The complete data set thus consisted of 74 populations (see Additional file 1 and Figure 3). Two to 26 individuals were sampled per population following a protocol described elsewhere [16], except in one locality in Morocco where only one individual could be found.

DNA protocols

DNA was extracted using a salting-out procedure [44]. Two mitochondrial DNA (mtDNA) fragments, corresponding to parts of the COI and COII genes, were ampli-

fied from 732 individuals and analyzed by SSCP, as described in Salvato et al. [17]. For each mobility class, 1-5 individuals were sequenced to check for the accuracy of SSCP analysis and to determine the corresponding haplotype. Sequences were aligned using ClustalX [45]. Sequences of COI (263 bp) and COII (341 bp) fragments were then concatenated, resulting in a 604 bp-long final alignment.

Data analyses

A partition homogeneity test was performed for the COI and COII fragments using Paup*4b10 [46]. The test confirmed that these regions contained homogeneous signals ($p = 0.15$), allowing data to be pooled for further analyses.

Model selection was performed using a Bayesian framework, through comparison of Bayes factors [47]. In addition, model performance was assessed using a posterior predictive test [48]. Models tested were selected using a modified version of Hierarchy 1 in MrModeltest 2.2 [49], enforcing or not a molecular clock. Given the limited length of the fragment analyzed and the correlation between proportion of invariant sites and the parameter alpha of the gamma distribution [47], we decided not to consider the invariant+gamma models.

For Bayes factors calculation, likelihoods for a given model were estimated using MrBayes v3.1.2 [50], and harmonic means were used as estimators of the overall marginal likelihood of the model. Each MrBayes analysis was the result of two independent chains of 2.10^6 generations, incrementally heated with $T = 0.15$. Convergence was assessed by computing the potential scale reduction factor with *sump* in MrBayes. Differences between Bayes factors obtained from the different models tested, calculated as twice the difference in the logarithm of harmonic means of likelihoods, were compared with reference values from Kass and Raftery [51].

For model performance assessment we chose as discrepancy variable the multinomial test statistics [52]. Posterior predictive distribution was evaluated through Monte-Carlo simulations of 1,000 datasets for each model using posterior densities of model parameters (tree topology, branch lengths and substitution parameters) inferred by MrBayes. MAPPS software [21] was used for simulations. The discrepancy between observed test statistics and simulated predictive distributions in the various models was quantified using Bayesian p-values [48] and the L-criterion proposed by Laud and Ibrahim [53], both computed with MAPPS.

Relationships between haplotypes and molecular dating were estimated by Bayesian inference of phylogeny using Beast v1.4.8 [54]. The model of sequence evolution and clock assumptions followed the results obtained from previous analyses and a Yule prior on the tree was assumed [55,56]; Markov chain Monte Carlo (MCMC) was run for 10 million generations, results being logged every 1,000 generations. After discarding the first 10% of the chain, convergence was checked by monitoring traces of sampled parameters and effective sample size following authors' suggestions. Analyses were cross-checked with MrBayes and the time of the most recent common ancestor (tMRCA) of selected clades was determined, assuming a sequence divergence rate of 2% per million years [57], and reported as a mean value with 95% highest posterior density interval (HPD).

For the most recent nodes, demographic Bayesian analyses were performed separately for each of the identified sub-clades using Beast and including all the sequences of a given group. Assumptions and settings were the same as above, except that coalescent priors of constant size and of exponential growth were used instead of Yule priors, and that two MCMC runs of 100 million steps were performed. tMRCAs of recent sub-clades were estimated assuming a 2% divergence, and must therefore be interpreted as the maximum age for a given sub-clade [40].

The phylogenetic reconstructions allowed us to identify three highly supported monophyletic clades within which a statistical parsimony network was computed using TCS v1.21 [58]. Such a network estimates genes genealogies from DNA sequences following the method described in Templeton et al. [59].

Gene diversity H and nucleotide diversity per site π were calculated within populations and within previously identified sub-clades. To infer whether each sub-clade has experienced recent population expansions, Tajima's D and Fu's Fs statistics were calculated and tested with DnaSP 4.10 [60]. Mismatch distributions of the pairwise genetic differences [61] were then performed using Arlequin 3.1 [62] and their goodness-of-fit to a sudden expansion model was tested using parametric bootstrap approaches (1000 replicates). The sum of squared deviations (SSD) between the observed and expected mismatch distributions was used to assess the significance of the test. Mismatch analyses were also used to estimate the approximate timing of expansion in the sub-clades where mutation-drift equilibrium was rejected. We used the relationship $\tau = 2ut$ [61], τ being the age of expansion measured in units of mutational time, t the expansion time in number of generations, and u the mutation rate per sequence and per generation. This last value was calculated using the relationship $u = 2\mu k$, with μ the mutation rate per nucleotide and k the length of the sequence in nucleotides. The 2% pairwise sequence divergence defined by DeSalle [57] was used to approximate μ.

Authors' contributions

CK, LZ and MS analyzed the data, AB, JR and AR planned the research, MS and PS performed the research, CK, LZ and AB wrote the paper and revised the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Sampling sites, geographic coordinates, host pine, collector and haplotype composition of each locality. *The number in brackets after each haplotype name is the number of individuals with that haplotype. Codes refer to the localities shown in Figure 3. Codes for hosts are as follows: PB:* Pinus brutia*; PH:* P. halepensis*; PM:* Pinus mugo/uncinata*; PN:* Pinus nigra*; PP:* P. pinaster*; PR:* Pinus radiata*; PS:* P. sylvestris*; CA:* Cedrus atlantica*, CD:* Cedrus deodara.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-220-S1.pdf]

Acknowledgements

We thank E. Magnoux, H. Santos, E. Petrucco Toffolo and D. Zovi for help with the lab work. J. Garcia, F. Goussard, P. Pineau and P. Menassieu, as well as all the people cited in Additional file 1 are gratefully acknowledged for sampling work. We thank C. Burban for valuable comments on a previous version of the manuscript. The work was supported by URTICLIM, a French project funded by the 'Agence Nationale de la Recherche' (ANR 07BDIV 013), and by EU projects PROMOTH QLK5-CT-2002-00852 and PRATIQUE 7FP-KBBE-212459. The study does not necessarily reflect the Commission's views and in no way anticipates the Commission's future policy.

References

1. Parmesan C: **Ecological and evolutionary responses to recent climate change.** *Annual Review of Ecology, Evolution, and Systematics* 2006, **37:**637-669.
2. Hewitt GM: **The genetic legacy of the Quaternary ice ages.** *Nature* 2000, **405:**907-913.
3. Hewitt GM: **Genetic consequences of climatic oscillations in the Quaternary.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2004, **359:**183-195.
4. Schmitt T: **Molecular biogeography of Europe: Pleistocene cycles and postglacial trends.** *Frontiers in Zoology* 2007, **4:**11.
5. Challis RJ, Mutun S, Nieves-Aldrey J-L, Preuss S, Rokas A, Aebi A, Sadeghi E, Tavakoli M, Stone GN: **Longitudinal range expansion and cryptic eastern species in the western Palaearctic oak gallwasp, *Andricus coriarius*.** *Molecular Ecology* 2007, **16:**2103-2114.
6. Rokas A, Atkinson R, Webster L, Csoka G, Stone GN: **Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp *Andricus quercustozae*.** *Molecular Ecology* 2003, **12:**2153-2174.
7. Stone GN, Challis RJ, Atkinson RJ, Csóka G, Hayward A, Melika G, Mutun S, Preuss S, Rokas A, Sadeghi E, Schönrogge K: **The phylogeographical clade trade: tracing the impact of human-mediated dispersal on the colonization of northern Europe by the oak gallwasp *Andricus kollari*.** *Molecular Ecology* 2007, **16:**2768-2781.
8. Hayward A, Stone GN: **Comparative phylogeography across two trophic levels: the oak gall wasp *Andricus kollari* and its chalcid parasitoid *Megastigmus stigmatizans*.** *Molecular Ecology* 2006, **15:**479-489.
9. Horn A, Roux-Morabito G, Lieutier F, Kerdelhué C: **Phylogeographic structure and past history of the circum-Mediterranean species *Tomicus destruens* Woll. (Coleoptera: Scolytinae).** *Molecular Ecology* 2006, **15:**1603-1615.
10. Horn A, Stauffer C, Lieutier F, Kerdelhué C: **Complex post-glacial history of the bark beetle *Tomicus piniperda* L. (Coleoptera, Scolytinae).** *Heredity* 2009, **103:**238-247.
11. Dynesius M, Jansson R: **Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations.** *Proceedings of the National Academy of Sciences, USA* 2000, **97:**9115-9120.
12. Pinho C, Harris DJ, Ferrand N: **Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation.** *Molecular Ecology* 2007, **16:**1191-1205.
13. Coope GR: **Several million years of stability among insect species because of, or in spite of, Ice Age climatic instability?** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2004, **359:**209-214.
14. Démolin G: **Comportement des adultes de *Thaumetopoea pityocampa* Schiff. Dispersion spatiale, importance écologique.** *Annales des Sciences Forestières* 1969, **26:**81-102.
15. Salvato P, Simonato M, Zane L, Patarnello T, Masutti L, Battisti A: **Do sexual pheromone traps provide biased information of the local gene pool in the pine processionary moth?** *Agricultural and Forest Entomology* 2005, **7:**127-132.
16. Simonato M, Mendel Z, Kerdelhué C, Rousselet J, Magnoux E, Salvato P, Roques A, Battisti A, Zane L: **Phylogeography of the pine processionary moth *Thaumetopoea wilkinsoni* in the Near East.** *Molecular Ecology* 2007, **16:**2273-2283.
17. Salvato P, Battisti A, Concato S, Masutti L, Patarnello T, Zane L: **Genetic differentiation in the winter pine processionary moth (*Thaumetopoea pityocampa-wilkinsoni* complex), inferred by AFLP and mitochondrial DNA markers.** *Molecular Ecology* 2002, **11:**2435-2444.
18. Battisti A, Stastny M, Netherer S, Robinet C, Schopf A, Roques A, Larsson S: **Expansion of geographic range in the pine processionary moth caused by increased winter temperatures.** *Ecological Applications* 2005, **15:**2084-2096.
19. Stone GN, Sunnucks P: **Genetic consequences of an invasion through a patchy environment: the cynipid gallwasp *Andricus quercuscalicis* (Hymenoptera: Cynipidae).** *Molecular Ecology* 1993, **2(4):**251-268.
20. Hurst GD, Jiggins FM: **Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts.** *Proceedings of the Royal Society B: Biological Sciences* 2005, **272:**1525-1534.
21. Bollback JP: **Bayesian model adequacy and choice in phylogenetics.** *Molecular Biology and Evolution* 2002, **19:**1171-1180.
22. Dermitzakis DM, Papanicolaou DJ: **Paleogeography and geodynamics of the Aegean region during the Neogene.** *Annales Géologiques des Pays Helléniques* 1981, **30:**245-289.
23. Krijgsman W, Hilgen FJ, Raffi I, Sierro FJ, Wilson DS: **Chronology, causes and progression of the Messinian salinity crisis.** *Nature* 1999, **400:**652-655.
24. Recuero E, Iraola A, Rubio X, Machordom A, Garcia-Paris M: **Mitochondrial differentiation and biogeography of *Hyla meridionalis* (Anura: Hylidae): an unusual phylogeographical pattern.** *Journal of Biogeography* 2007, **34:**1207-1219.
25. Modolo L, Salzburger W, Martin RD: **Phylogeography of Barbary macaques (*Macaca sylvanus*) and the origin of the Gibraltar colony.** *Proceedings of the National Academy of Sciences, USA* 2005, **102:**7392-7397.
26. Fromhage L, Vences M, Veith M: **Testing alternative vicariance scenarios in Western Mediterranean discoglossid frogs.** *Molecular Phylogenetics and Evolution* 2004, **31:**308-322.
27. Cosson JF, Hutterer R, Libois R, Sara M, Taberlet P, Vogel P: **Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae).** *Molecular Ecology* 2005, **14:**1151-1162.
28. Huchon H, Démolin G: **La bioécologie de la processionnaire du pin. Dispersion potentielle - Dispersion actuelle.** *Revue Forestière Française* 1970, **22:**220-234.
29. Kerdelhué C, Magnoux E, Lieutier F, Roques A, Rousselet J: **Comparative population genetic study of two oligophagous**

insects associated with the same hosts. *Heredity* 2006, **97:**38-45.
30. Zovi D, Stastny M, Battisti A, Larsson S: **Ecological costs on local adaptation of an insect herbivore imposed by host plants and enemies.** *Ecology* 2008, **89:**1388-1398.
31. Burban C, Petit RJ, Carcreff E, Jactel H: **Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host.** *Molecular Ecology* 1999, **8:**1593-1602.
32. Stone G, Atkinson R, Rokas A, Csoka G, Nieves-Aldrey J-L: **Differential success in northwards range expansion between ecotypes of the marble gallwasp *Andricus kollari*: a tale of two lifecycles.** *Molecular Ecology* 2001, **10:**761-778.
33. Slatkin M, Hudson RR: **Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations.** *Genetics* 1991, **129:**555-562.
34. Korsten M, Ho SY, Davison J, Pähn B, Vulla E, Roht M, Tumanov IL, Kojola I, Andersone-Lilley Z, Ozolins J, *et al.*: **Sudden expansion of a single brown bear maternal lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals?** *Molecular Ecology* 2009, **18:**1963-1979.
35. Ho SYW, Larson G, Edwards CJ, Heupink TH, Lakin KE, Holland PWH, Shapiro B: **Correlating Bayesian date estimates with climatic events and domestication using a bovine case study.** *Biology Letters* 2008, **4:**370-374.
36. Ford MJ: **Applications of selective neutrality tests to molecular ecology.** *Molecular Ecology* 2002, **11:**1245-1262.
37. Johnson JA, Dunn PO, Bouzat JL: **Effects of recent population bottlenecks on reconstructing the demographic history of prairie-chickens.** *Molecular Ecology* 2007, **16:**2203-2222.
38. Cheddadi R, Vendramin GG, Litt T, François L, Kageyama M, Lorentz S, Laurent J-M, de Beaulieu J-L, Sadori L, Jost A, Lunt D: **Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*.** *Global Ecology and Biogeography* 2006, **15:**271-282.
39. Ibrahim KM, Nichols RA, Hewitt GM: **Spatial patterns of genetic variation generated by different forms of dispersal during range expansion.** *Heredity* 1996, **77:**282-291.
40. Ho SY, Phillips MJ, Cooper A, Drummont AJ: **Time dependency of molecular rate estimates and systematic overestimation of recent divergence times.** *Molecular Biology and Evolution* 2005, **22:**1561-1568.
41. Dubey S, Cosson J-F, Magnanou E, Vohralik V, Benda P, Frynta D, Hutterer R, Vogel V, Vogel P: **Mediterranean populations of the lesser white-toothed shrew (*Crocidura suaveolens* group): an unexpected puzzle of Pleistocene survivors and prehistoric introductions.** *Molecular Ecology* 2007, **16:**3438-3452.
42. Luciano P, Lentini A, Battisti A: **First record of *Thaumetopoea pityocampa* in Sardinia.** *Proceedings of the Italian Congress of Entomology* 2007:273.
43. Vendramin GG, Anzidei M, Madaghiele A, Bucci G: **Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites.** *Theoretical and Applied Genetics* 1998, **97:**456-463.
44. Patwary MU, Kenchington EL, Bird CJ, Zouros E: **The use of random amplified polymorphic DNA markers in genetic studies of the sea scallop *Plactopecten magellanicus* (Gmellin, 1791).** *Journal of Shellfish Research* 1994, **13:**547-553.
45. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The Clustal_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Research* 1997, **25:**4876-4882.
46. Swofford DL: *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)* 4th edition. Sunderland, Massachussetts: Sinauer Associates; 2003.
47. Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL: **Bayesian phylogenetic analysis of combined data.** *Systematic Biology* 2004, **53:**47-67.
48. Gelman A, Meng XL, Stern H: **Posterior predictive assessment of model fitness via realized discrepancies.** *Statistica Sinica* 1996, **6:**733-760.
49. Nylander JAA: *MrModeltest v2. Program distributed by the author* Evolutionary Biology Centre, Uppsala University; 2004.
50. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572-1574.
51. Kass RE, Raftery AE: **Bayes factors.** *Journal of the American Statistical Association* 1995, **90:**773-795.
52. Goldman N: **Statistical tests of models of DNA substitution.** *Journal of Molecular Evolution* 1993, **36:**182-198.
53. Laud PW, Ibrahim JG: **Predictive model selection.** *Journal of the Royal Statistical Society Series B - Methodological* 1995, **57:**247-262.
54. Drummont AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evolutionary Biology* 2007, **7:**214.
55. Villacorta C, Jaume D, Oromí P, Juan C: **Under the volcano: phylogeography and evolution of the cave-dwelling *Palmorchestia hypogaea* (Amphipoda, Crustacea) at La Palma (Canary Islands).** *BMC Biology* 2008, **6:**7.
56. Zinner D, Groeneveld LF, Keller C, Roos C: **Mitochondrial phylogeography of baboons (*Papio spp.*) - Indication for introgressive hybridization?** *BMC Evolutionary Biology* 2009, **9:**83.
57. DeSalle R, Freedman T, Prager EM, Wilson AC: **Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*.** *Journal of Molecular Evolution* 1987, **26:**157-164.
58. Clement M, Posada D, Crandall KA: **TCS: a computer program to estimate gene genealogies.** *Molecular Ecology* 2000, **9:**1657-1659.
59. Templeton AR, Crandall KA, Sing CF: **A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data. III. Cladogram estimation.** *Genetics* 1992, **132:**619-633.
60. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R: **DnaSP, DNA polymorphism analysis by the coalescent and other methods.** *Bioinformatics* 2003, **19:**2496-2497.
61. Rogers AR, Harpending H: **Population growth makes waves in the distribution of pairwise genetic differences.** *Molecular Biology and Evolution* 1992, **9:**552-569.
62. Excoffier L, Laval G, Schneider S: **Arlequin (version 3.0): an integrated software package for population genetics data analysis.** *Evolutionary Bioinformatics Online* 2005, **1:**47-50.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

Locality code	Country	Region / District	Location	# Indiv.	Latitude	Longitude	Altitude (m)	Host	Collector	Haplotype composition	π	H
1	Portugal	Setubal	Alcácer do Sal	12	38º 23' N	08º 31' W	40	PP	Teresa Vasconcelos & Manuela Branco	29BA(12)	0	0
2	Portugal	Viseu	Viseu	11	40º 40' N	07º 54' W	500	PP	Teresa Vasconcelos	29BA(10) 31BA(1)	0.03%	0.18
3	Portugal	Vila Real	Sevivas	12	41º 31' N	07º 30' W	50	PP	Paola Arnaldo	29BA(5) 30BA(7)	0.088%	0.53
4	Portugal	Braganca	Varges	11	41° 52' N	06° 40' W	400	PP	Paola Arnaldo	29BA(2) 30BA(9)	0.054%	0.33
5	Spain	Southern Iberia	Gibraltar	6	36° 08' N	05° 21' W	400	PH	John Cortes	29BB(6)	0	0
6[1]	Spain	Andalusia	Sierra Nevada	12	37° 05' N	03° 27' W	1800	PS	José Hodar	1A(10) 3A(2)	0.050%	0.30
7	Spain	Andalusia	Sierra de la Cabra Montés	4	37º 47' N	03º 46' W	1200	PP	Ramon Gonzalez Ruiz	29BA(1) 31BA(3)	0.083%	0.50
8	Spain	Guadarrama	Collado Mediano	5	40° 41' N	04° 02' W	1100	PN PP	INRA	29BA(5)	0	0
9	Spain	Aragón	Boltaña	4	42° 26' N	00° 02' E	650	PS	INRA	1A(4)	0	0
10	Spain	Balearic islands	Sant Llorençt Cardassar	10	39° 37' N	03° 17' E	300	PH	INRA	1A(10)	0	0

11	France	Pyrénées Orientales	Osseja	12	42° 23’ N	02° 00’ E	1400	PS PM	INRA	1A(10) 1BG(2)	0.050%	0.30
12	France	Gironde	Pierroton	12	44° 44’ N	00° 46’ W	60	PP	INRA	1A(12)	0	0
13	France	Aveyron	Jouas	3	44° 30’ N	02° 24’ E	430	CD	INRA	1A(3)	0	0
14	France	Haute-Loire	Brioude	12	45° 18’ N	03° 29’ E	600	PS	INRA	1A(12)	0	0
15	France	Loiret	Lorris	12	47° 49’ N	02° 29’ E	150	PS	INRA	1A(12)	0	0
16	France	Alpes de Haute Provence	Thorame	12	44° 04’ N	06° 34’ E	1200	PS	INRA	1A(12)	0	0
17	France	Corsica	Barchetta	10	42° 30’ N	09° 22’ E	100	PP PR	INRA	34BK(2) 34BL(8)	0.059%	0.36
18[1]	Italy	Aosta	Ruines Verrès	26	45° 39’ N	07° 41’ E	1000	PS	University of Padova	1D(23) 1BI(3)	0.035%	0.21
19[1]	Italy	Liguria	Rollo	11	43° 57’ N	08° 08’ E	250	PH	University of Torino	1A(5) 2A(6)	0.18%	0.55
20[1]	Italy	Liguria	Massimino	12	44° 25’ N	08° 26’ E	600	PS	University of Torino	1A(11) 1C(1)	0.028%	0.17
21	Italy	Alto Adige	Silandro	19	46° 38’ N	10° 47’ E	1100	PN PS	University of Padova	1A(17) 1BJ(2)	0.033%	0.20

22[1]	Italy	Friuli	Tugliezzo	12	46° 22’ N	13° 11’ E	450	PN	University of Padova	1A(6) 1B(6)	0.090%	0.55
23[1]	Italy	Friuli	Venzone	11	46° 20’ N	13° 08’ E	350	PN	University of Padova	1A(10) 1F(1)	0.030%	0.18
24[1]	Italy	Toscana	M. S. Michele	12	43° 31’ N	11° 24’ E	890	PN	Univ. of Firenze	1G(12)	0	0
25[1]	Italy	Veneto	Calbarina	12	45° 16’ N	11° 43’ E	136	PN	University of Padova	1A(12)	0	0
26	Italy	Puglia	Gargano	9	41° 46’ N	16° 11’ E	200	PH	University of Padova	1G(9)	0	0
27	Italy	Sicily	Nicolosi	12	37° 37’ N	15° 01’ E	700	PN	University of Catania	1G(12)	0	0
28	Italy	Sardegna	Sanluri	4	39° 33’ N	08° 54’ E	633	PH	University of Sassari	1G(4)	0	0
29	Italy	Pantelleria	Pantelleria	8	36° 48’ N	11° 59’ E	190	PP	University of Catania	20AF(8)	0	0
30	Montenegro	Podgorica	Cemovskopolje	4	42° 28’ N	19° 17’ E	700	PN	M. Glavendekic	33A(4)	0	0
31	Albania	Tirana	Tirana	12	41° 19’ N	19° 49’ E	90	PN	Fatmir Laceja	1A(12)	0	0

32	Bulgaria	Pirin	Sandanski	8	41° 34’ N	23° 17’ E	220	PN	Daniela Pilarska	1A(7) 1BH(1)	0.041%	0.25
33	Bulgaria	Asenovgrad	Javrovo	10	42° 01’ N	24° 52’ E	700-1000	PN	Plovdiv Station	1A(10)	0	0
34	Greece	Serres	Lailia	2	41º 12´ N	23º 36´ E	900	PN	Maria Kalapanida	1A(2)	0	0
35	Greece	Thessaloniki	Agia Anastasia	3	40º 28´ N	23º 23´ E	50	PB	Maria Kalapanida	1A(3)	0	0
36	Greece	Halkidiki	Skioni	5	39º 58´ N	23º 23´ E	50	PH	Maria Kalapanida	1A(4) 1BE(1)	0.066%	0.40
37	Greece	Thasos	Thasos	11	40° 46’ N	24° 42’ E	na	PH	Maria Kalapanida	1A(8) 1BC(1) 1BD(1) 1BF(1)	0.13%	0.49
38	Greece	Thasos	Potamia	11	40° 43’ N	24° 45’ E	na	PH	Maria Kalapanida	1A(10) 1BD(1)	0.060%	0.18
39	Greece	Samothraki	Samothraki	10	40° 29’ N	25° 31’ E	na	PH	Maria Kalapanida	1A(10)	0	0
40	Greece	Levos	West-Levos	11	39° 10’ N	25° 56’ E	na	PH	Maria Kalapanida	1A(11)	0	0
41	Greece	Levos	East-Levos	10	39° 06’ N	26° 25’ E	na	PH	Maria Kalapanida	1A(9) 32A(1)	0.033%	0.20
42	Greece	Chios	West-Chios	10	38° 22’ N	26° 01’ E	na	PH	Maria Kalapanida	1A(10)	0	0
43	Greece	Chios	East-Chios	11	38° 22’ N	26° 08’ E	na	PH	Maria Kalapanida	1A(11)	0	0
44	Greece	Kardamili	Vassiliki Ft	9	36° 53’ N	22° 19’ E	1400	PN	INRA	1A(9)	0	0
45	Greece	Crete	Chania	12	35° 30’ N	24° 01’ E	na	PH	Maria Kalapanida	17AB(12)	0	0

46	Greece	Crete	Heraklion	9	35° 18’ N	25° 07’ E	na	PH	Maria Kalapanida	15AA(8) 16AA(1)	0.074%	0.22
47	Turkey	Canakkale	Bayramic	2	39° 49’ N	26° 36’ E	300	PB	Mustafa Avci	1A(2)	0	0
48[1]	Turkey	Izmir	Aydin	10	37° 51’ N	27° 50’ E	600	PB	University of Izmir	4E(9) 4H(1)	0.033%	0.20
49[2]	Turkey	Isparta	Gunur	15	37° 46’ N	30° 34’ E	1050	PB PN	University of Isparta	4N(15)	0	0
50[2]	Turkey	Antalya	Karaoz	8	36° 54’ N	30° 43’ E	200	PB	University of Isparta	4N(6) 10P(1) 11N(1)	0.30%	0.46
51[2]	Turkey	Taurus mountains	Pozanti	11	37° 17’ N	34° 51’ E	970	PB PN	University of Padova	12U(7) 12R(4)	0.34%	0.51
52[2]	Turkey	Taurus mountains	Aladag	10	37° 33’ N	35° 22’ E	1100	PB	University of Padova	12K(7) 13K(1) 14W(2)	0.15%	0.51
53[2]	Turkey	Samsun	Samsun	12	41° 17’ N	36° 20’ E	150	PN	Zvi Mendel	4V(12)	0	0
54[2]	Turkey	Iskenderun	Iskenderun	10	36° 34’ N	36° 10’ E	210	PB	University of Padova	12K(5) 12R(3) 12S(1) 12T(1)	0.30%	0.71
55[2]	Turkey	Antakia	Seyhköy	11	36° 04’ N	36° 10’ E	450	PB	University of Padova	5Y (11)	0	0
56[2]	Cyprus	East Cyprus	El Skopi	19	35° 00’ N	32° 40’ E	100-1000	PB PN	Zvi Mendel	6M(1) 7M(9) 7Q(5) 8M(2) 9M(2)	0.15%	0.72

57[2]	Lebanon	Beirut	Beirut	24	33° 53’ N	35° 30’ E	272	PB	American University Beirut	5J(24)	0	0
58	Israel	Golan	Golan Heights	4	32° 58’ N	35° 44’ E	1000	PN	Zvi Mendel	5J(4)	0	0
59[2]	Israel	Upper Galilee	Qiryat Shemona	14	33° 11’ N	35° 33’ E	350	PB	Zvi Mendel	5J(14)	0	0
60[2]	Israel	Lower Galilee	Segev	9	32° 52’ N	35° 14’ E	400	PH	Zvi Mendel	5J(9)	0	0
61[2]	Israel	Judean foothills	Haruvit	15	31° 45’ N	34° 50’ E	150	PH	Zvi Mendel	5J(15)	0	0
62[2]	Israel	S Judean Mts.	Yatir	15	31° 20’ N	35° 03’ E	550	PH	Zvi Mendel	5J(15)	0	0
63[2]	Israel	W Negev	Qisufim	10	31° 22’ N	34° 24’ E	50	PH	Zvi Mendel	5J(10)	0	0
64	Libya	Cyrenaica	Al Bayda	6	32° 45’ N	21° 37’ E	470	PH	University Omar Almukhtar	18AC(3) 18AD(2) 19AE(1)	0.20%	0.73
65	Tunisia	Nabeul	Nabeul	7	36° 27’ N	10° 44’ E	40	PH	M. El Habib Ben Jamâa	21AF(7)	0	0
66	Tunisia	Bizerte	Bizerte	8	37° 02' N	09° 42' E	15	PH	M. El Habib Ben Jamâa	20AF(8)	0	0
67	Tunisia	Dir El Kef	Dir El Kef	7	36° 11’ N	08° 43’ E	370	PH	M. El Habib Ben	20AI(1) 20AG(1) 20AH(4)	0.29%	0.71

									Jamâa	20AF(1)		
68	Algeria	Tellien Atlas	Tikjda	12	36° 00' N	04° 17' E	800-1000	PH	Mohamed Zamoum	22AJ(3) 22AK(5) 22AL(1) 22AM(1) 23AK(2)	0.19%	0.78
69	Algeria	Saharian Atlas	Djelfa Moudjbara	10	34° 30’ N	03° 28’ E	1100	PH	Mohamed Zamoum	26AN(1) 24AN(3) 25AN(6)	0.11%	0.60
70	Morocco	Eastern Middle Atlas	Aknoul	4	34° 40’ N	03° 52’ W	1250	CA	Driss Ghaioule	28AQ(4)	0	0
71	Morocco	Rif	Bab Barred	10	34° 59’ N	04° 50’ W	1300	CA	Driss Ghaioule	28AQ(9) 28AR(1)	0.033%	0.2
72	Morocco	Middle Atlas	Boutrouba	10	33° 27’ N	05° 03’ W	1900	CA	Driss Ghaioule	27AP(10)	0	0
73	Morocco	High Atlas	Oukaimeden	2	31° 17' N	07° 48' W	2300	PP	INRA	25AO(2)	0	0
74	Morocco	High Atlas	Lalla Takerkoust	1	31° 22' N	08° 08' W	750	PH	INRA	25AO(1)	0	0

1. Populations from Salvato et al. *Molecular Ecology* 2002, **11:**2435-2444.

2. Populations from Simonato et al. *Molecular Ecology* 2007, **16:**2273-2283.

The previous paper has been completed by a further manuscript to which I contributed marginally, accepted for publication in Journal of Biogeography in December 2009, as from the abtract:

The role of topography in structuring the demographic history of the pine processionary moth, ***Thaumetopoea pityocampa*** **(Lepidoptera: Notodontidae)**

Jérôme Rousselet[1*], Ruixing Zhao[1,2], Dallal Argal[1], Mauro Simonato[3], Andrea Battisti[3], Alain Roques[1] & Carole Kerdelhué[4]

[1] INRA, UR633 Unité de Recherche de Zoologie Forestière, F-45075 Orléans, France

[2] Liaoning Forest Pest and Disease Control and Quarantine Station, Changjiang Street, Huanggu District, 110036 Shenyang, China

[3] Dipartimento di Agronomia Ambientale e Produzioni Vegetali, Agripolis, Università di Padova, 35020 Legnaro PD, Italy

[4] INRA, UMR1202 BIOGECO, F-33610 Cestas, France

ABSTRACT

Aim We investigated the Quaternary history of the pine processionary moth, *Thaumetopoea pityocampa,* an oligophagous insect currently expanding its range. We tested the potential role played by mountain ranges during the post-glacial recolonization of western Europe.

Location Western Europe, with a focus on the Pyrenees, Massif Central and western Alps.

Methods Maternal genetic structure was investigated using a fragment of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene. We performed phylogenetic analyses, hierarchical analysis of molecular variance, and investigated signs of past expansion.

Results A strong phylogeographic pattern was found, with two deeply divergent clades. Surprisingly, these clades were not separated by the Pyrenees but rather were distributed from western to central Iberia and from eastern Iberia to the Italian peninsula, respectively. This latter group consisted of three shallowly divergent lineages that exhibited strong geographic structure and independent population expansions. The three identified lineages

occurred: (1) on both sides of the Pyrenean range, with more genetically diverse populations in the east, (2) from eastern Iberia to western France, with a higher genetic diversity in the south, and (3) from the western Massif Central to Italy. Admixture areas were found at the foot of the Pyrenees and Massif Central.

Main conclusions The identified genetic lineages were geographically structured, but surprisingly the unsuitable high elevation areas of the main mountainous ranges were not responsible for the spatial separation of genetic groups. Rather than acting as barriers to dispersal, mountains appear to have served as refugia during the Pleistocene glaciations, and current distributions largely reflect expansion from these bottlenecked refugial populations. The western and central Iberian clade did not contribute to the northward post-glacial recolonization of Europe, but its northern limit does not correspond to the Pyrenees. The different contributions of the identified refugia to post-glacial expansion might be explained by differences in host plant species richness. For example, the Pyrenean lineage could have been trapped elevationally by tracking montane pines, while the eastern Iberian lineage could have expanded latitudinally by tracking thermophilic lowland pine species.

Chapter 4

The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae)

Published as:

Salvato P., Simonato M., Battisti A. and Negrisolo E. (2008) The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* **9**: 331 doi:10.1186/1471-2164-9-331

I contributed to all parts of the experimental work and some parts of writing.

BMC Genomics

BioMed Central

Research article

Open Access

The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae)

Paola Salvato†[1], Mauro Simonato†[1], Andrea Battisti[1] and Enrico Negrisolo*[2]

Address: [1]Department of Environmental Agronomy and Vegetal Productions-Entomology, University of Padova, Agripolis, Viale dell'Università 16, 35020 Legnaro, Italy and [2]Department of Public Health, Comparative Pathology and Veterinary Hygiene, University of Padova, Agripolis, Viale dell'Università 16, 35020 Legnaro, Italy

Email: Paola Salvato - paola.salvato@unipd.it; Mauro Simonato - mauro.simonato@unipd.it; Andrea Battisti - andrea.battisti@unipd.it; Enrico Negrisolo* - enrico.negrisolo@unipd.it

* Corresponding author †Equal contributors

Published: 15 July 2008

BMC Genomics 2008, **9**:331 doi:10.1186/1471-2164-9-331

Received: 21 May 2008
Accepted: 15 July 2008

This article is available from: http://www.biomedcentral.com/1471-2164/9/331

© 2008 Salvato et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Knowledge of animal mitochondrial genomes is very important to understand their molecular evolution as well as for phylogenetic and population genetic studies. The Lepidoptera encompasses more than 160,000 described species and is one of the largest insect orders. To date only nine lepidopteran mitochondrial DNAs have been fully and two others partly sequenced. Furthermore the taxon sampling is very scant. Thus advance of lepidopteran mitogenomics deeply requires new genomes derived from a broad taxon sampling. In present work we describe the mitochondrial genome of the moth *Ochrogaster lunifer*.

Results: The mitochondrial genome of *O. lunifer* is a circular molecule 15593 bp long. It includes the entire set of 37 genes usually present in animal mitochondrial genomes. It contains also 7 intergenic spacers. The gene order of the newly sequenced genome is that typical for Lepidoptera and differs from the insect ancestral type for the placement of *trnM*. The 77.84% A+T content of its α strand is the lowest among known lepidopteran genomes. The mitochondrial genome of *O. lunifer* exhibits one of the most marked C-skew among available insect Pterygota genomes. The protein-coding genes have typical mitochondrial start codons except for *cox1* that present an unusual CGA. The *O. lunifer* genome exhibits the less biased synonymous codon usage among lepidopterans. Comparative genomics analysis study identified *atp6*, *cox1*, *cox2* as *cox3*, *cob*, *nad1*, *nad2*, *nad4*, and *nad5* as potential markers for population genetics/phylogenetics studies. A peculiar feature of *O. lunifer* mitochondrial genome it that the intergenic spacers are mostly made by repetitive sequences.

Conclusion: The mitochondrial genome of *O. lunifer* is the first representative of superfamily Noctuoidea that account for about 40% of all described Lepidoptera. New genome shares many features with other known lepidopteran genomes. It differs however for its low A+T content and marked C-skew. Compared to other lepidopteran genomes it is less biased in synonymous codon usage. Comparative evolutionary analysis of lepidopteran mitochondrial genomes allowed the identification of previously neglected coding genes as potential phylogenetic markers. Presence of repetitive elements in intergenic spacers of *O. lunifer* genome supports the role of DNA slippage as possible mechanism to produce spacers during replication.

Background

Animal mitochondrial genomes (mtDNAs) are usually circular molecules spanning 16–20 kbp that contain 13 protein-coding genes (PCGs), 2 ribosomal RNA and 22 transfer (tRNA) genes [1]. Non-coding control elements, that regulate the transcription and replication of the genome, are also present in mtDNAs [1,2]. Mitochondrial genomes are very important subject for different scientific disciplines including animal health, comparative and evolutionary genomics, molecular evolution, phylogenetics and population genetics. However, current knowledge on mtDNAs is very uneven as well exemplified by sequences available in GenBank that were obtained mostly from vertebrate taxa. Insects constitute the most species-rich class among animals with almost a million of taxa described to date [3]. Within the insects, the Lepidoptera (butterflies plus moths) order accounts for more than 160,000 species [4]. Despite this huge taxonomic diversity the existing information on lepidopteran mtDNA is very limited. Complete sequences have been determined for the two butterflies *Coreana raphaelis* and *Artogeia melete*, and for the seven moths *Adoxophyes honmai*, *Antheraea pernyi*, *Bombyx mori*, *Bombyx mandarina* and *Manduca sexta*, *Phthonandria atrilineata* and *Saturnia boisduvalii* [5-9] while near complete sequences exist for *Ostrinia furnacalis* and *Ostrinia nubilalis* [10] (Table 1). Current genomic knowledge of Lepidoptera is very scanty and the covered taxon-sampling is extremely poor and limited to six superfamiles among the 45–48 known, and to 9 families of the recognized 120 [4]. A better understanding of the lepidopteran mtDNA requires an expansion of taxon and genome samplings. We were able to fully sequence the mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer*. The newly determined mtDNA is the first complete sequence for the Superfamily Noctuoidea, a very large assemblage that accounts for about 40% of all described Lepidoptera. [4]. In the present paper the *Ochrogaster* genome is described and compared with mtDNAs of other lepidopterans as well as pterygote Insecta.

Results and discussion

Genome organization, structure and composition

The mtDNA genome of *O. lunifer* is a circular molecule 15593 bp long. It includes the entire set of 37 genes usually present in animal mtDNAs [1], i.e., 13 PCGs, 22 tRNA genes, and 2 ribosomal genes (Figure 1). The mtDNA genome of *O. lunifer* contains also 7 intergenic spacers (s1–s7), spanning at least 15 bp, described in a paragraph below. Genes on the same strand are overlapped (e.g. *trnM* vs. *trnI*; *atp8* vs. *atp6*), contiguous, separated by few nucleotides or by intergenic spacers (e.g. *nad3* vs. *trnA*; *trnC* vs. *trnY*). Genes on opposite strands exhibit a similar behavior (Figure 1).

The *O. lunifer* mtDNA has the typical lepidopteran gene order [8,9] that differs from the ancestral gene order of insects [1] for the placement of *trnM*. In the ancestral type (e.g. *Drosophila yakuba* mtDNA) the order in the α strand is: A+T region, *trnI*, *trnQ*, *trnM*, *nad2*. In all lepidopteran mtDNAs, sequenced to date, the order is: A+T region, *trnM*, *trnI*, *trnQ*, *nad2* which implies the translocation of *trnM* [5-11]. This placement of *trnM* is a molecular feature exclusive to lepidopteran mtDNAs. Further genome sequencing is necessary to establish if this feature is a mitochondrial signature of the whole order Lepidoptera.

The composition of the α strand of *O. lunifer* mtDNA is A = 6252 (40.09%), T = 5886 (37.75%), G = 1179 (7.56%) and C = 2276 (14.60%).

The A+T% and G+C% values for the α strand as well as the A- and G-skews [12] were calculated for all available complete mtDNA genomes of Pterygota and are presented in the scatter plots of Figure 2.

The average A+T% value for the analyzed mtDNAs set is 76. 63 ± 4.84. The highest A+T% values are shared by the mtDNAs of three bees (*Apis mellifera*, *Bombus ignitus* and *Melipona bicolor*) and two bugs (*Aleurodicus dugesii* and *Schizaphis graminum*). All lepidopteran mtDNAs but *O. lunifer* exhibit high A+T% values. The A+T content of *O. lunifer* mtDNA is 77.84% that represents the lowest value for lepidopteran complete mtDNAs [5-8,10]. The lowest A+T contents are found in the termite mtDNAs (*Reticulitermes* spp.). Extreme A+T values are also shared by species having highly re-arranged gene order [13]. However the possession of a re-arranged genome is not sufficient *per se* to have an A+T content drastically departing from the average (e.g. *Aleurochiton aceris* and *Bemisia tabaci*). The A+T values appear to be linked to taxonomic relatedness at low rank (i.e. genus, family) (e.g. species of *Drosophila*, species of *Bactrocera*, members of family *Apidae*). The relation is not true at higher ranks (i.e. superfamily; order) where patterns become inconsistent and the A+T content can be very different among species as exemplified by Hemiptera (*A. dugesii* vs.*Triatoma dimidiata)*.

The average A-skew is 0.04214 ± 0.11350 and most of pterygote mtDNAs are slightly to moderately A-skewed with values ranging from 0.00287 (*B. ignitus*) to 0.18247 (*Locusta migratoria*). The lepidopteran A-skews vary from -0.04748 (*C. raphaelis*) to 0.05872 (*B. mori*) with the *O. lunifer* mtDNA exhibiting a slight A-skew (0.03015). The *Reticulitermes* mtDNA genomes, having the lowest A+T% content, exhibit a very pronounced A-skew. Most marked T-skews are observed in the mtDNA genomes of *Campanulotes bidentatus* and *Trialeuroides vaporarium* that have low A+T% content and gene-orders different than insect ancestral gene order [1,14,15]. Gene order re-arrangement is

Table 1: List of taxa analyzed in present paper

Order	Family	Species	Acc. number	Reference
Orthoptera	Acrididae	*Locusta migratoria*	NC_001712	[42]
	Acrididae	*Oxya chinensis*	NC_010219	Hang and Zhang, unpublished
	Tettigonidae	*Anabrus simplex*	NC_009967	[43]
	Tettigonidae	*Ruspolia dubia*	NC_009876	[44]
	Gryllotalpidae	*Gryllotalpa orientalis*	NC_006678	[45]
Isoptera	Rhinotermitidae	*Reticulitermes flavipes*	NC_009498	[13]
	Rhinotermitidae	*Reticulitermes hageni*	NC_009501	[13]
	Rhinotermitidae	*Reticulitermes virginicus*	NC_009500	[13]
	Rhinotermitidae	*Reticulitermes santonensis*	NC_009499	[13]
Manthophasmatodea	Mantophasmatidae	*Sclerophasma paresisense*	NC_007701	[46]
Mantodea	Mantidae	*Tamolanica tamolana*	NC_007702	[46]
Blattaria	Blattidae	*Periplaneta fuliginosa*	NC_006076	[47]
Plecoptera	Pteronarcyidae	*Pteronarcys princeps*	NC_006133	[48]
Phthiraptera	Boopidae	*Heterodoxus macropus*	NC_002651	[16]
	Philopteridae	*Bothriometopus macrocnemis*	NC_009983	[49]
	Philopteridae	*Campanulotes bidentatus*	NC_007884	[15]
Hemiptera	Aleyrodidae	*Aleurochiton aceris*	NC_006160	[14]
	Aleyrodidae	*Aleurodicus dugesii*	NC_005939	[14]
	Aleyrodidae	*Bemisia tabaci*	NC_006279	[14]
	Aleyrodidae	*Neomaskellia andropogonis*	NC_006159	[14]
	Aleyrodidae	*Tetraleurodes acaciae*	NC_006292	[14]
	Aleyrodidae	*Trialeurodes vaporariorum*	NC_006280	[14]
	Aphididae	*Schizaphis graminum*	NC_006158	[14]
	Cicadellidae	*Homalodisca coagulata*	NC_006899	Baumann and Baumann, unpublished
	Psyllidae	*Pachypsylla venusta*	NC_006157	[14]
	Aphrophoridae	*Philaenus spumarius*	NC_005944	[50]
	Reduviidae	*Triatoma dimidiata*	NC_002609	[51]
Psocoptera	Lepidopsocidae	Lepidopsocid sp. RS-2001	NC_004816	[52]
Thysanoptera	Thripidae	*Thrips imaginis*	NC_004371	[53]
Coleoptera	Cerambycidae	*Anoplophora glabripennis*	NC_008221	An et al., unpublished
	Chrysomelidae	*Crioceris duodecimpunctata*	NC_003372	[54]
	Elateridae	*Pyrophorus divergens*	NC_009964	[55]
	Lampyridae	*Pyrocoelia rufa*	NC_003970	[56]
	Tenebrionidae	*Tribolium castaneum*	NC_003081	[57]
Lepidoptera				
Tortricoidea	Tortricidae	*Adoxophyes honmai*	NC_008141	[7]
Bombycoidea	Saturniidae	*Antheraea pernyi*	NC_004622	Liu et al., unpublished
Bombycoidea	Bombycidae	*Bombyx mandarina*	NC_003395	[5]
Bombycoidea	Bombycidae	*Bombyx mori*	NC_002355	Lee et al., unpublished
Bombycoidea	Saturniidae	*Saturnia boisduvalii*	NC_010613	[9]
Geometroidea	Geometridae	*Phthonandria atrilineata*	NC_010522	Yang et al., unpublished
Papilionoidea	Pieridae	*Artogeia melete*	NC_010568	Hong et al., unpublished
Papilionoidea	Lycaenidae	*Coreana raphaelis*	NC_007976	[6]
Sphingoidea	Sphingidae	*Manduca sexta*	EU286785	[8]
Noctuoidea	**Notodontidae**	***Ochrogaster lunifer***	AM946601	**This paper**
Pyraloidea	Crambidae	*Ostrinia furnacalis*	NC_003368	[10]
Pyraloidea	Crambidae	*Ostrinia nubilalis*	NC_003367	[10]
Diptera	Ceratopogonidae	*Culicoides arakawae*	NC_009809	Matsumoto, unpublished
	Culicidae	*Aedes albopictus*	NC_006817	Ho et al., unpublished
	Culicidae	*Aedes aegypti*	NC_010241	Lobo et al., unpublished
	Culicidae	*Anopheles gambiae*	NC_002084	[58]
	Culicidae	*Anopheles quadrimaculatus* A	NC_000875	[59]
	Calliphoridae	*Cochliomyia hominivorax*	NC_002660	[60]
	Calliphoridae	*Lucilia sericata*	NC_009733	Cibrario et al., unpublished
	Calliphoridae	*Chrysomya putoria*	NC_002697	[61]
	Drosophilidae	*Drosophila melanogaster*	NC_001709	[62]
	Drosophilidae	*Drosophila mauritiana*	NC_005779	[63]
	Drosophilidae	*Drosophila sechellia*	NC_005780	[63]
	Drosophilidae	*Drosophila simulans*	NC_005781	[63]

Table 1: List of taxa analyzed in present paper *(Continued)*

	Drosophilidae	*Drosophila yakuba*	NC_001322	[64]
	Oestridae	*Dermatobia hominis*	NC_006378	Azeredo-Espin et al., unpublished
	Muscidae	*Haematobia irritans*	NC_007102	Lessinger et al., unpublished
	Nemestrinidae	*Trichophthalma punctata*	NC_008755	[65]
	Syrphidae	*Simosyrphus grandicornis*	NC_008754	[65]
	Tabanidae	*Cydistomyia duplonotata*	NC_008756	[65]
	Tephritidae	*Ceratitis capitata*	NC_000857	[66]
	Tephritidae	*Bactrocera carambolae*	NC_009772	Ye et al., unpublished
	Tephritidae	*Bactrocera dorsalis*	NC_008748	Yu et al., unpublished
	Tephritidae	*Bactrocera oleae*	NC_005333	[67]
	Tephritidae	*Bactrocera papayae*	NC_009770	Ye et al., unpublished
	Tephritidae	*Bactrocera philippinensis*	NC_009771	Ye et al., unpublished
Hymenoptera	Apidae	*Apis mellifera ligustica*	NC_001566	[68]
	Apidae	*Bombus ignitus*	DQ870926	[69]
	Apidae	*Melipona bicolor*	NC_004529	Silvestre and Arias, unpublished
	Vanhornidae	*Vanhornia eucnemidarum*	NC_008323	[70]

not necessarily linked to strong A/T-skew as proved by the highly rearranged, but low skew, genome of *Heterodoxus macropus* [16].

The average G+C% content is 23.37 ± 4.84. The G+C% pattern among various species is obviously opposite to the A+T% thus it does not require further comments. More composite is the G/C-skew distribution. The average G-skew is -0.16006 ± 0.138235. Most of pterygota mtDNAs are C-skewed with G-skew values ranging from -0.32827 (*Vanhornia eucnemidarum*) to -0.01250 (*Heterodoxus macropus*). The main exception is represented by the mtDNA of bugs, while the highest G-skewed genome is that of *C. bidentatus*. Most of lepidopteran mtDNAs share very similar G-skew values that are included within the bulk of mtDNAs. The notable exception is represented by the newly determined mtDNA of *O. lunifer* that exhibits the second most pronounced C-skew (G-skew = -0.31751) among analyzed genomes.

G-skew can be markedly different even in species belonging to the same genus and having a very similar G+C content as well exemplified by *Reticulitermes santonensis* and *Reticulitermes virginicus* mtDNAs. The same reasoning applies at high taxonomic rank to the Hemiptera. The mtDNA of *C. bidentatus* exhibits very high A-skew and G-skew. However, this feature is not a general rule and extreme A-skew and G-skew are not necessarily reciprocally linked, as proved by species of genus *Reticulitermes* that exhibit very strong A-skews but not G-skews.

The list of currently available mtDNAs reveals that there is a strong bias in term of taxon sampling both at low and high taxonomic ranks within Pterygota. A direct consequence is that present knowledge of base composition and A/G skews reflects such biases and addition of a single taxon can change our view on these features. This point is well exemplified by the *O. lunifer* mtDNA that exhibits a A+T percentage different than other lepidopteran mtDNAs that share high A+T contents [8,9]. Thus a broad and more balanced taxon sampling appears to be a mandatory goal to investigate and identify general patterns for the parameters considered above.

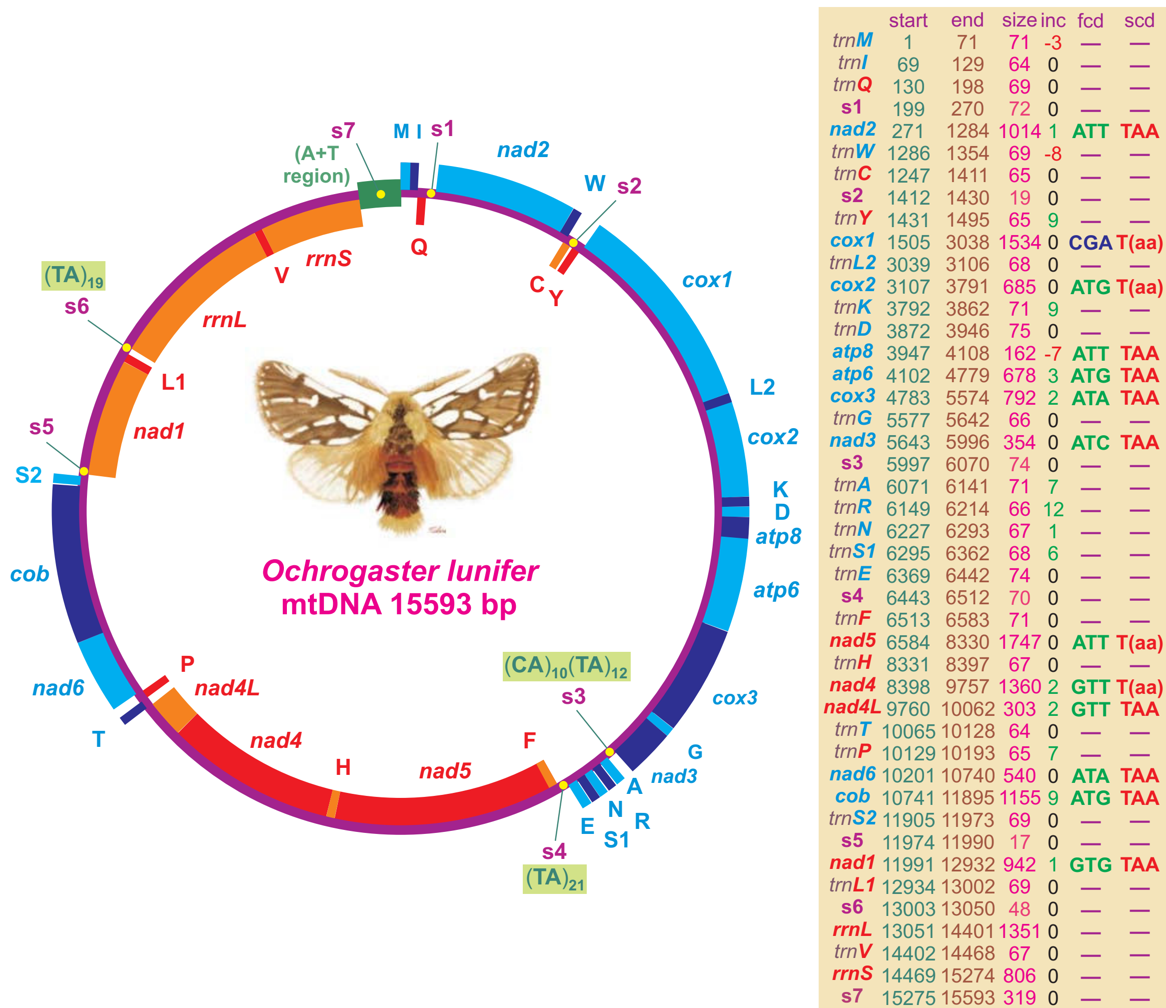
Protein-coding genes

The mtDNA of *O. lunifer* contains the full set of PCGs usually present in animal mtDNA. PCGs are arranged along the genome according to the standard order of Insects [1] (Figure 1). The putative start codons of PCGs are those previously known for animal mtDNA i.e. ATN, GTG, TTG, GTT [17] with the only exception represented by the CGA start codon of *cox1* gene. This non-canonical putative start codon is found also in the butterfly *A. melete* and in the moths *A. honmai*, *B. mori*, *B. mandarina*, *M. sexta* and *P. atrilineata* [5-8]. In the butterfly *C. raphaelis* the tetranucleotide TTAG is the putative start codon [6] and the six nucleotide TATTAG has been suggested as putative start codon for the moths *O. nubilalis* and *O. furnicalis* [10]. An unusual start codon for *cox1* gene is known in various arthropod mtDNA [e.g. [18]].

The *cox1*, *cox2*, *nad5*, and *nad4* genes of *O. lunifer* mtDNA have incomplete stop codons. The presence of incomplete stop codons is a feature shared with all lepidopteran mtDNAs sequenced to date [5-10] and more in general with many arthropod mtDNAs [1].

The *atp8* and a *atp6* of *O. lunifer* are the only PCGs having a seven nucleotides overlap (Figure 1). This feature is common to all lepidopteran mtDNA genomes known [5-10] and is found in many animal mtDNAs [1].

The abundance of codon families and Relative Synonymous Codon Usage (RSCU) [19] in PCGs were investigated for all available lepidopteran mtDNAs and the results are summarized in Figures 3 and 4. All first codons as well as stop codons, complete and incomplete, were



	start	end	size	inc	fcd	scd
trnM	1	71	71	-3	—	—
trnI	69	129	64	0	—	—
trnQ	130	198	69	0	—	—
s1	199	270	72	0	—	—
nad2	271	1284	1014	1	ATT	TAA
trnW	1286	1354	69	-8	—	—
trnC	1247	1411	65	0	—	—
s2	1412	1430	19	0	—	—
trnY	1431	1495	65	9	—	—
cox1	1505	3038	1534	0	CGA	T(aa)
trnL2	3039	3106	68	0	—	—
cox2	3107	3791	685	0	ATG	T(aa)
trnK	3792	3862	71	9	—	—
trnD	3872	3946	75	0	—	—
atp8	3947	4108	162	-7	ATT	TAA
atp6	4102	4779	678	3	ATG	TAA
cox3	4783	5574	792	2	ATA	TAA
trnG	5577	5642	66	0	—	—
nad3	5643	5996	354	0	ATC	TAA
s3	5997	6070	74	0	—	—
trnA	6071	6141	71	7	—	—
trnR	6149	6214	66	12	—	—
trnN	6227	6293	67	1	—	—
trnS1	6295	6362	68	6	—	—
trnE	6369	6442	74	0	—	—
s4	6443	6512	70	0	—	—
trnF	6513	6583	71	0	—	—
nad5	6584	8330	1747	0	ATT	T(aa)
trnH	8331	8397	67	0	—	—
nad4	8398	9757	1360	2	GTT	T(aa)
nad4L	9760	10062	303	2	GTT	TAA
trnT	10065	10128	64	0	—	—
trnP	10129	10193	65	7	—	—
nad6	10201	10740	540	0	ATA	TAA
cob	10741	11895	1155	9	ATG	TAA
trnS2	11905	11973	69	0	—	—
s5	11974	11990	17	0	—	—
nad1	11991	12932	942	1	GTG	TAA
trnL1	12934	13002	69	0	—	—
s6	13003	13050	48	0	—	—
rrnL	13051	14401	1351	0	—	—
trnV	14402	14468	67	0	—	—
rrnS	14469	15274	806	0	—	—
s7	15275	15593	319	0	—	—

Figure 1
Map of the mitochondrial genome of *O. lunifer*. Genes coded in the α strand (clockwise orientation) are blue or cyan colored. Genes coded in the β strand (anti-clockwise orientation) are red or orange colored. Alternation of colors was applied for clarity. Start, first position along α strand; end, last position along α strand; size, size of the sequence; inc, intergenic nucleotides; fcd, first codon; scd, stop codon. Incomplete stop codons are presented with parentheses. Negative inc values refer to overlapping nucleotides for genes located in the same or different strands. Gene names are the standard abbreviations used in this paper; tRNA genes are indicated by the single letter IUPAC-IUB abbreviation for their corresponding amino acid in the draw. s1–s7, intergenic spacers.

excluded from the analysis to avoid biases due to unusual putative start codons and incomplete stop codons.

Total number of non-stop codons (CDs) used by the 12 analyzed mtDNAs is very similar ranging from 3695 of *C. raphaelis* to 3732 of *O. lunifer*. The codon families exhibit a very similar behavior among considered species. The eight codon families with at least 50 CDs per thousand CDs (Leu2, Ile, Phe, Met, Asn, Ser2, Gly, Tyr) encompass an average 65.82% ± 1.20% of all CDs. The three families with at least 100 CDs per thousand CDs (Leu2, Ile, Phe) account for an average 35.36% ± 0.98% of all CDs (Figure 3). The A+T rich CDs are favored over synonymous CDs with lower A+T content as proved by RSCU results (Figure 4). This point is well exemplified by the Leu2 family where the TTA codon accounts for the large majority of CDs in the family (see below). Invertebrate mitochondrial code includes 62 amino-acid encoding codons [1].

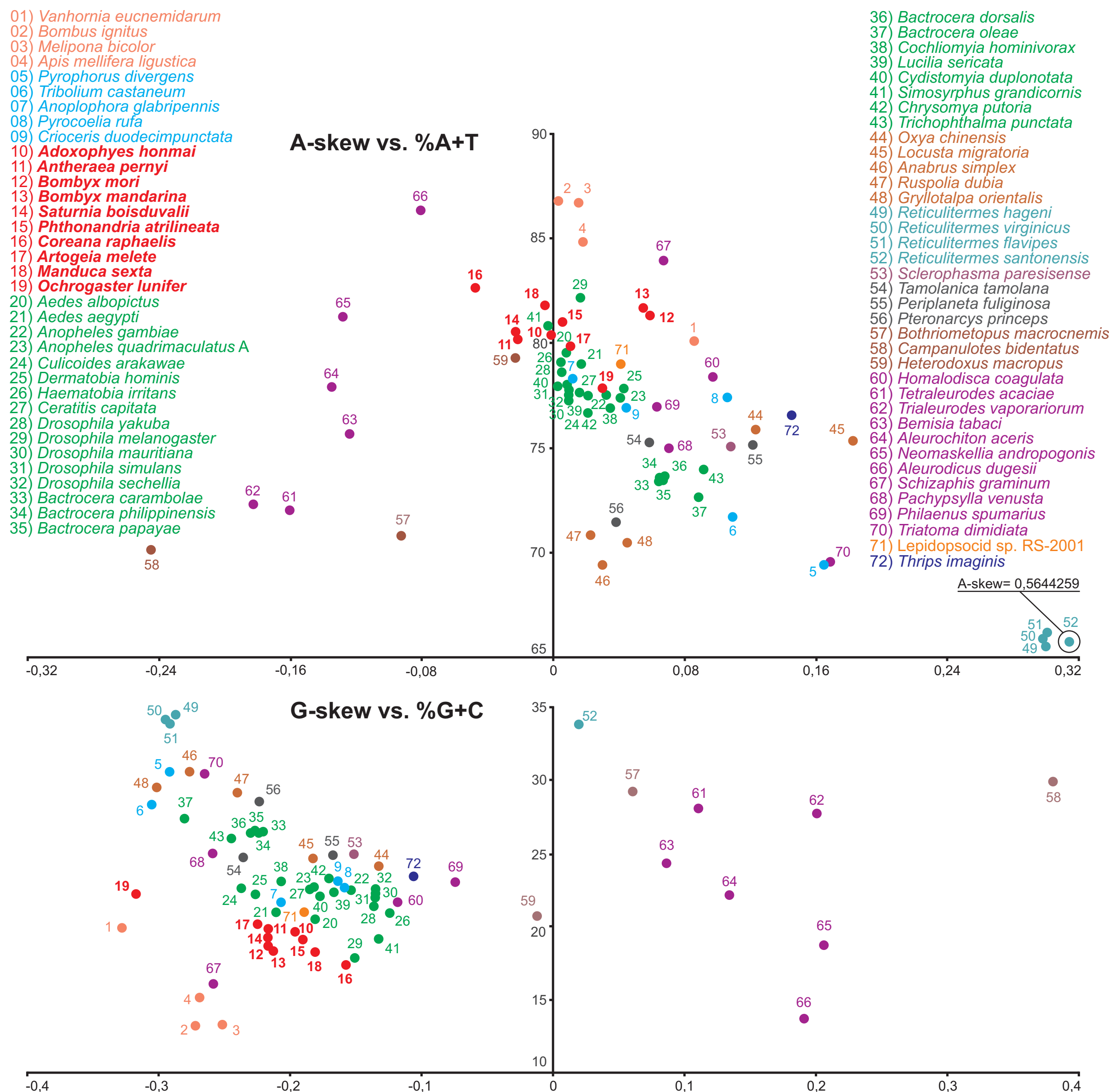


Figure 2
A-skew vs. A+T% and G-skew vs. G+C% in the Pteryogota mtDNAs. Values were calculated on α strands for full lengh mtDNA genomes. The X axis provides the skews values, while the Y axis provides the A+T/G+C values. Named of species are colored according to their taxonomic placement at Order level (see Table 1).

Among the 12 analyzed genomes the total number of used codons results to be directly linked to the A+T content. The *C. raphaelis* mtDNA, having the highest A+T% content (see Figure 2) uses 52 codons, and never utilized the 10 G+C rich codons listed in Figure 2. Conversely, *O. lunifer* mtDNA, characterized by the lowest A+T% among considered lepidopteran genomes, uses all 62 codons. Differences in the number of used CDs are present between species of the same genus (e.g. *B. mandarina* vs. *B. mori*) even if the discrepancies appear circumscribed to

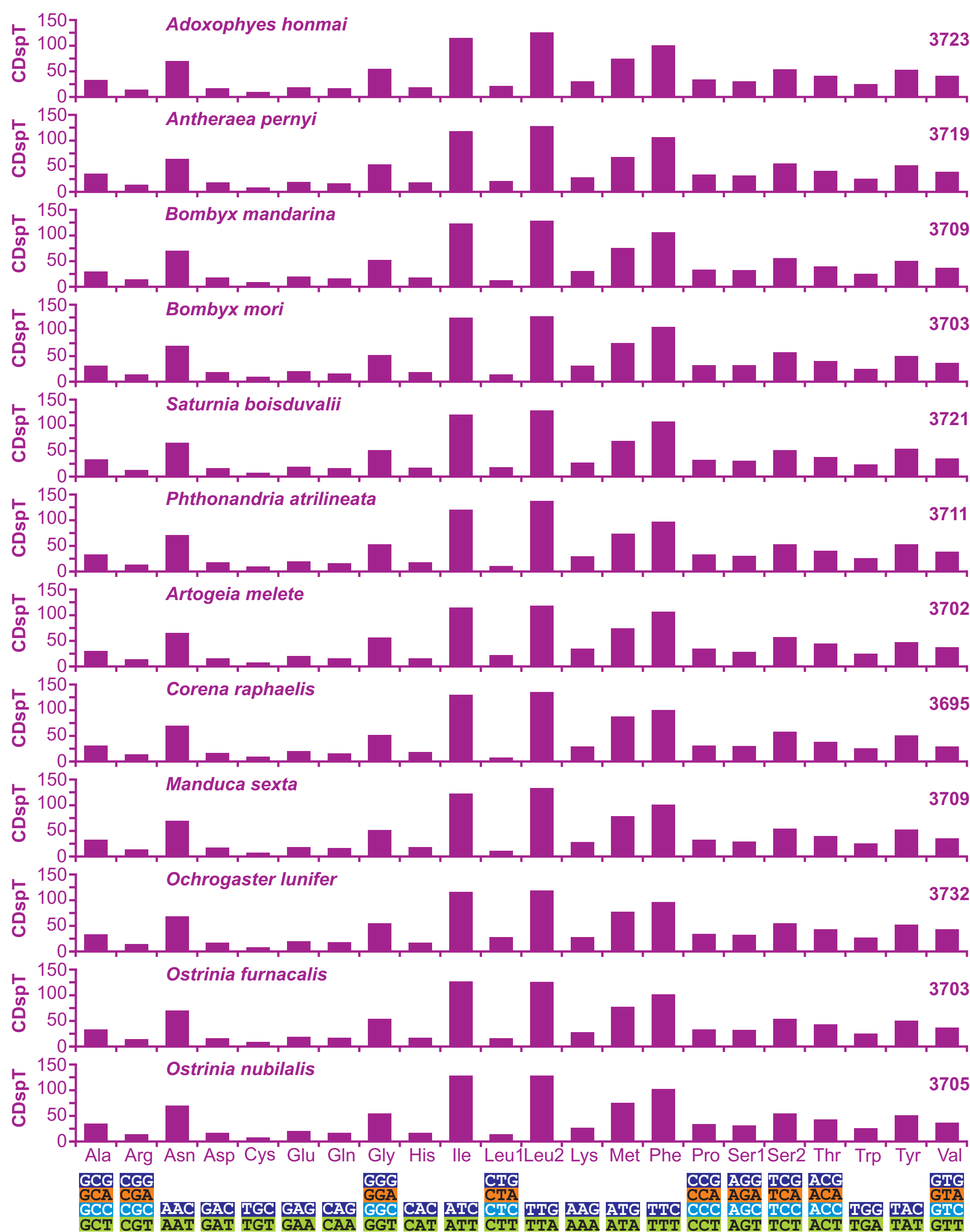


Figure 3
Codon distribution in lepidopteran mtDNAs. Numbers to the left refer to the total number of codon. CDspT, codons per thousands codons. Codon Families are provided on the x axis.

Figure 4
Relative Synonymous Codon Usage (RSCU) in lepidopteran mtDNAs. Codon Families are provided on the x axis. Red-colored codon, codon not present in the genome. Codon Families are provided on the x axis.

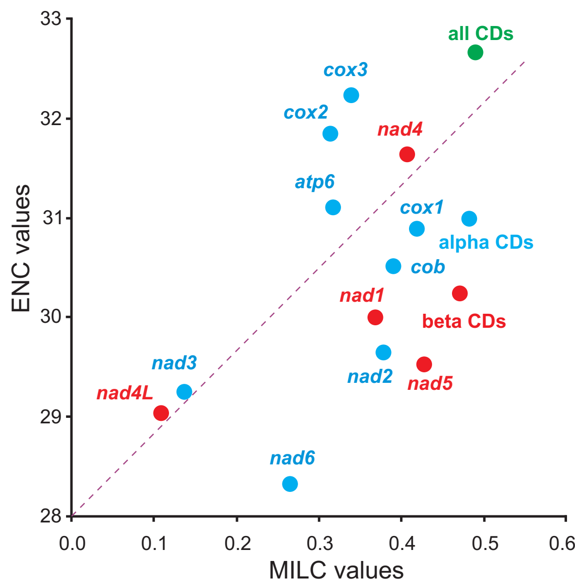


Figure 5
Scatter plot graphic of MILC vs. ENC calculated for PCGs of lepidopteran mtDNAs. Dots correspond to average values calculated for different genes. PCGs on α strand are blue-colored, PCGs on β strand are red colored. All pooled PCGs are presented as a green dot plot. Genes nomenclature as in main text.

G+C rich CDs with very limited use (e.g. GCG and CGC). The Leu1 (average = 11.73 ± 3. 82%) and Leu2 (average = 88.44 ± 3.89%) codon families are very differently represented in lepidopteran PCGs while Ser1 (average = 34.95 ± 3.67%) and Ser2 (average = 64.05 ± 1.09%) exhibit a more balanced composition.

Four amino acid residues (Leu, Ile, Phe and Ser) account for more than 44.50% (average = 45.68 ± 0.58%) of all residues forming the 13 mitochondrial proteins. The Leu and Ile amino acids share hydrophobic lateral chains, Phe is also hydrophobic and Ser exhibits an aliphatic behavior [20] thus their massive presence is striking but not surprising for membrane proteins.

Codon usage by single PCGs was investigated by calculating the two indices ENC (Effective number of codon used) [21] and MILC (Measure Independent of Length and Composition) [22]. Both indices, based on different approaches [21,22] provide a measure of codon variability of PCGs. The ENC and MILC estimate the codon variability in a way that allows comparison among sequences having different lengths as is the case of various PCGs. Genes exhibiting a higher diversity in codon usage have generally a higher number of variable sites, a prerequisite to be potential phylogenetic markers. Thus the use of ENC and MILC scores, according to the new approach presented in this paper, is a way to study PCG sequences variability on a codon perspective. The best scores of both indices should allow to identify the more diverse PCGs in a approach complementary to the usual method based on evolutionary distances among orthologous sequences (e.g. [8]). The assessment of genetic variability is an interesting point. Indeed some PCGs are standard marker for species recognition [23] or have been extensively used as phylogenetic markers in Lepidoptera while others have received so far limited or no attention. Understanding the genetic diversity of each PCG is a prerequisite to determine its phylogenetic usefulness. The ENC and MILC values were calculated for all PCGs but *atp8* that contains too a few codons to get reliable ENC/MILC estimations [22]. Calculations were extended also to all 13 PCGs pooled as well as to the pooled PCGs belonging to α and β strands respectively. The scatter plot analysis is provided in Figure 5. As expected the greatest diversity in codon usage is found when all codons are considered. Good codon diversity is found also when all PCGs of α or β strands are considered. More interesting is the behavior of single genes. In this latter case sequences well established as phylogenetic markers (i.e. *cox1*, *cob*, *nad5*, and *cox2*) are intermixed with PCGs poorly or not considered by researchers (e.g. *cox3*, *nad4*, *nad1*, *nad2*). Our results suggest that the neglected PCGs should be considered as potential markers thus extending the number of mtDNA PCGs sampled for population as well as phylogenetic markers. Findings, based on codon diversity, must be integrated with direct comparisons of sequences [8] that allow to better define the optimal task that each gene can perform i.e. to be used at low taxonomic level or at high taxonomic level.

Transfer and ribosomal RNA genes

Ochrogaster genome has the characteristic 22 tRNAs set (Figure 6) present in most of animal mtDNAs [1]. All tRNAs present the typical clover leaf secondary structure but *trnS1* lacks the DHU stem. This feature is shared with the *C. raphaelis* mtDNA [6] but is not a general feature of lepidopteran mtDNA as proved by *A. honmai* that has all tRNAs with a complete clover leaf structure [7]. In general, the lack of DHU arm in *trnS1* is a common condition in metazoan mtDNAs [24].

The *trnA*, *trnD*, *trnG*, *trnK*, *trnL1*, *trnL2*, *trnQ*, and *trnS2* of *Ochrogaster* mtDNA show mismatches in their stems. Mismatches are located mostly in the acceptor and anticodon stems with a single exception represented by *trnD* that exhibits the mismatch on the TΨC stem. Mismatches on tRNA stems are known also for the *trnA*, *trnL1*, *trnL2*, and *trnQ*, of *C. raphaelis* [6]. Mismatches observed in tRNAs

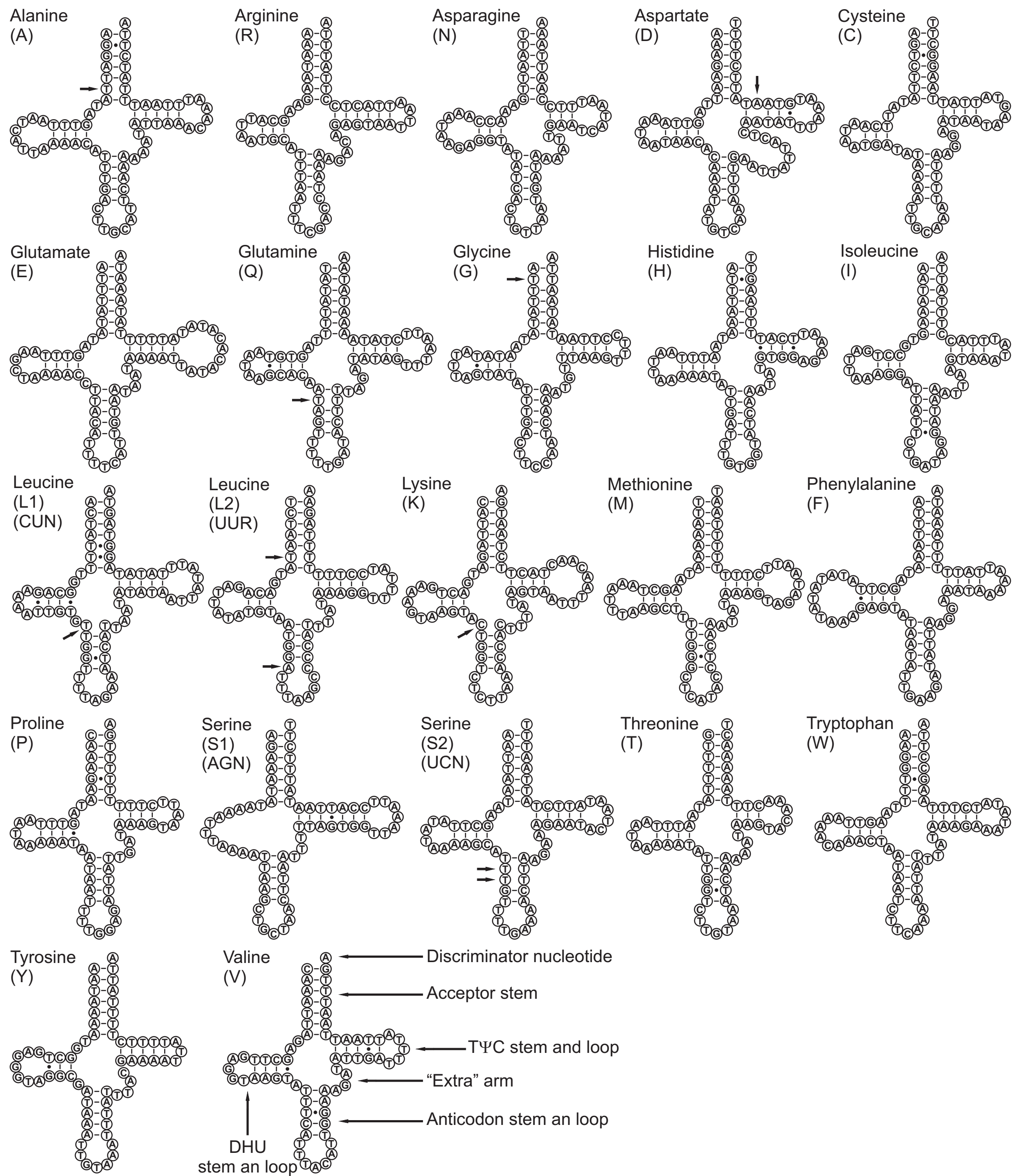


Figure 6
Secondary structures of transfer tRNAs in *O. lunifer* mtDNA.

are corrected through RNA-editing mechanisms that are well known for arthropod mtDNA [e.g. [24]].

Preliminary analysis performed on *rrnL* and *rrnS* of *O. lunifer* revealed that these genes are capable of folding into structures (data not shown) similar to those already produced for lepidopteran mitochondrial ribosomal subunits [8,25,26]. Further studies, that extend the taxon sampling, are currently in progress in our lab to better define *rrnL* and *rrnS* structures within the Thaumetopoeinae subfamily that includes also *O. lunifer*.

Non coding regions

The mtDNA genome of *O. lunifer* contains 7 intergenic spacers (s1–s7) spanning at least 15 bp (Figures 1 and 7). The features of s1–s7 spacers are presented below with reference to the α strand for orientation and sequence motifs description.

The s1 spacer, located between *trnQ* and *nad2*, appears to be the result of a duplicated segment (Figure 7). The s1 spacer is present in all 12 lepidopteran mtDNAs so far sequenced while it is absent in other insects [8]. While the genomic location is constant the sequence divergence is high among species [8]. Further investigation with a broad taxon sampling within the Lepidoptera is necessary to assess if the s1 spacer is a constant molecular signature of lepidopteran mtDNA.

The s2 spacer, placed between *trnC* and *trnY*, derives from the triplication of a six nucleotides motif with minor changes (Figure 7). An 11 bp spacer between *trnC* and *trnY* is found also in the mtDNA of *A. melete* and shares the ACAATT motif with the s2 spacer of *O. lunifer*. Because no other known lepidopteran mtDNA exhibits such a spacer its presence in *A. melete* and *O. lunifer* has to be interpreted as the result of independent events.

Spacer s3, located between *nad3* and *trnA*, exhibits a partial duplicated segment and a poly-T motif within the first 30 nt. The second half of s3 spacer is characterized by two microsatellite repeats $(CA)_{10}(TA)_{12}$. Spacers having the same genomic location, and containing TA microsatellites are found also in *B. mori* and *B. mandarina* mtDNA genomes.

Spacer s4, inserted between *trnE* and *trnF*, contains a 5' microsatellite $(TA)_{23}$, while the 3' half seems to be the triplication of a 10 nucleotides motif with some changes (Figure 7). A spacer characterized by a different motif $(TATTA)_{31}$, but having the same genomic placement, is found in the *A. honmai* mtDNA genome.

The spacer s5, located between *trnS2* and *nad1*, contains the ATACTAA motif which is conserved across the Lepidoptera order [8]. This motif is possibly fundamental to site recognition by the transcription termination peptide (mtTERM protein) [2]. Spacer s5 is present in most insect mtDNAs even if the nucleotide sequence can be quite divergent [8].

The s6 spacer is located between *trnS2* and *-rrnL* and exhibits a di-nucleotide microsatellite $(TA)_{19}$ directly in contact with the 3' end of *rrnL* gene. To date spacer s6 is known only for the mtDNA of *O. lunifer*.

The s7 spacer coincides with the A+T region. Several features common to the Lepidoptera A+T region [8] are present in the s7 spacer. The $O_R\beta$ (origin of the β strand replication) is located 21 bp downstream from *rrnS* gene in *B. mori* [27]. It contains the motif ATAGA followed by an 18 bp poly-T stretch. A very similar pattern occurs in *O. lunifer* where the ATAGA motif is located 17 bp downstream from *rnnS* gene and is followed by a 20 bp poly-T stretch (Figure 7). A microsatellite-like $(AT)_7(TA)_3$ element preceded by the ATTTA motif is present in the 3' third of *O. lunifer* s7 spacer. The presence of a microsatellite preceded by the ATTA motif is also a feature found in the A+T regions of other Lepidoptera [8]. Finally a 10 bp poly-A is present immediately upstream *trnM*. This poly-T (in the β strand) element is still a common feature of the A+T region in Lepidoptera [8,28]. No large repeated segments were detected in the A+T region of *O. lunifer*. This arrangement is consistent with other lepidopteran A+T regions while markedly contrasts with patterns observed in other insect orders [8,29].

Intergenic spacers containing repeated elements are scattered all over the lepidopteran mtDNAs while repeated elements are restricted mostly to the A+T region in other insects [8]. Most parts of spacers of *O. lunifer* are made by repeated motifs. Predominance of repeated elements suggest that mtDNA expansion can be achieved through a miss-pairing duplication mechanism, i.e. DNA slippage, during genome replication. Several intergenic spacers are restricted to a single butterfly/moth species and have not counterparts even within Lepidoptera. Thus it is plausible to suggest that spacers production occurs independently and recursively within Lepidoptera. It remains unknown while this feature is so prominent in moths and butterflies and apparently limited, reduced or absent in other insect mtDNAs sequenced to date. This behavior requires further investigation provided that mtDNA intergenic spacers are found in non-insect Arthropoda as well as other animal phyla [e.g. [18,30]].

Conclusion

The mitochondrial genome of *O. lunifer* is the first sequenced mtDNA for a representative of the Noctuoidea a superfamily that includes about 40% of all described

Spacer s1

TTATTAATAAAAATATAAATTAATTATTTAGATTAAATTCTAATAAAAATTATTAAATTATAATTATTTTTA

TTATTAATAAAAAT-A-TAAATTA--ATTATTTAGATTAAA
TT-CTAATAAAAATTATTAAATTATAATTATTT---TTA

Spacer s2

AATTACAATTATAATTATT

AATTAC
7 AATTAT 12
AATTATT

Spacer s3

ACATCTCATTAAATTTTTATAAACATCCTCACACACACACACACACACACATATATATATATATATATATATATAT

ACATCTCATTAAATTTTTATAA
ACATC-C-T

CACACACACACACACACACATATATATATATATATATATATATAT

Spacer s4

TATATATATATATATATATATATATATATATATATATATATATTACCTATCTTATCCTTCTTCCCTTTAT

TATATATATATATATATATATATATATATATATATATATATA

53 TTACCTATCT 61
TATCCT-TCT
T-CCCTTTAT

Spacer s5

ATACTAAAAATAATTAA

ATACTAA

Spacer s6

ATTAACTAATTATATATATATATATATATATATATATATATATATATA

1 ATTAA-
-CTAAT 10

TATATATATATATATATATATATATATATATATATATA

Spacer s7 (A+T region)

AATAATAAATTAACTAAATAGATTTTTTTTTTTTTTTTTTTTATATTAAAATATTTAATAATCAATAATAAATTTT
TAATAATTTCTTTTCCTTTCTTTTTATAATATTAATATTAAATACATAATTGCATATTAAATTTTTATAATTCATA
TATTTTATATTATATTTATATACCATTATATATTTGTTAATAAATTAATAAATTATTAATATTAATAATAAATTAA
ATATTTAATATATATATATATTATATAAAGATAAATAAAACACTTAAATGTTTTATTTACCATTTTTAATAATATT
TACATAAAAAAAAAA

Figure 7
Genomic spacers in the mtDNA of O. lunifer. The sequences of spacers are those present in the α strand.

lepidopteran species. The newly determined genome shares the gene order, the presence of intergenic spacers, and other features with previously known lepidopteran genomes. The placement of *trnM* immediately after the A+T region results to be an exclusive molecular signature of all lepidopteran mtDNAs sequenced to date. Further

genome sequencing will establish if this feature characterizes the whole order Lepidoptera. The mtDNA of *O. lunifer* exhibits a peculiar low A+T content and marked C-skew. Compared to other lepidopteran genomes it is less biased in synonymous codon usage. Comparative analysis on codon usage among lepidopteran mitochondrial genomes identified *atp6*, *cox1*, *cox2*, *cox3*, *cob*, *nad1*, *nad2*, *nad4*, and *nad5* as potential markers for phylogenetic and population genetic studies. Most of the genes listed above have been previously neglected for the tasks suggested here. The massive presence of repetitive elements in intergenic spacers of *O. lunifer* genome lead us to suggest an important role of DNA slippage as possible mechanism to produce spacers during replication.

Methods

Sample origin and DNA extraction

An ethanol-preserved larva specimen of *Ochrogaster lunifer* collected in Australia (Suburb of Kenmore, Queensland, 25th February 2005) by Myron P. Zalucki (University of Queensland) was used as starting material for this study. Total DNA was extracted by applying a salting-out protocol [31]. Quality of DNA was assessed through electrophoresis in a 1% agarose gel and staining with ethidium bromide.

PCR amplification and sequencing of Ochogaster lunifer mtDNA

PCR amplification was performed using a mix of insect universal primers [32,33] and primers specifically designed on the *O. lunifer* sequences. For a full list of successful primers as well as PCR conditions see Additional file 1. The PCR products were visualized in electrophoresis in a 1% agarose gel and staining with ethidium bromide. Each PCR product represented by a single electrophoretic band was purified with the ExoSAP-IT kit (Amersham Biosciences) and directly sequenced. Sequencing of both strands was performed at the BMR Genomics service (Padova, Italy) on automated DNA sequencers mostly employing the primers used for PCR amplification.

Sequence assembly and annotation

The mtDNA final consensus sequence was assembled using the SeqMan II program from the Lasergene software package (DNAStar, Madison, WI). Genes and strands nomenclature used in this paper follows Negrisolo et al. [18].

Sequence analysis was performed as follows. Initially the mtDNA sequence was translated into putative proteins using the Transeq program available at the EBI web site. The true identity of these polypeptides was established using the BLAST program [34,35] available at the NCBI web site. Gene boundaries were determined as follows. The 5' ends of PEGs were inferred to be at the first legitimate in-frame start codon (ATN, GTG, TTG, GTT; [17]) in the open reading frame (ORF) that was not located within the upstream gene encoded on the same strand. The only exception was *atp6*, which has been previously demonstrated to overlap with its upstream gene *atp8* in many mtDNAs [17]. The PCG terminus was inferred to be at the first in-frame stop codon encountered. When the stop codon was located within the sequence of a downstream gene encoded on the same strand, a truncated stop codon (T or TA) adjacent to the beginning of the downstream gene was designated as the termination codon. This codon was thought to be completed by polyadenylation to a complete TAA stop codon after transcript processing. Finally pair-wise comparisons with orthologous proteins were performed with ClustalW program [36] to better define the limits of PCGs.

Irrespectively of the real initiation codon, a formyl-Met was assumed to be the starting amino acid for all the proteins as previously proved for other mitochondrial genomes [37,38].

The transfer RNA genes were identified using the tRNAscan-SE program [39] or recognized manually as sequences having the appropriate anticodon and capable of folding into the typical cloverleaf secondary structure [17].

The boundaries of the ribosomal *rrnL* gene were assumed to be delimited by the ends of the *trnV*-s6 pair. The 3' end of *rrnS* gene was assumed to be delimited by the start of *trnV* while the 5'end was determined through comparison with orthologous genes of other Lepidoptera so far sequenced.

Genomic analysis

Nucleotide composition was calculated with the EditSeq program included in the Lasergene software package. The GC-skew = (G-C)/(G+C) and AT-skew = (A-T)/(A+T) were used [12] to measure the base compositional difference between the different strands or between genes coded on the alternative strands. The Relative Synonymous Codon Usage (RSCU) values were calculated with MEGA 4 program [40].

The codon usage by analyzed genomes was investigated by calculating the two indices ENC (Effective Number of Codon used) [21] and MILC (Measure Independent of Length and Composition [22]. ENC and MILC values were calculated with the INCA 2.1 program [41].

Abbreviations

mtDNA: mitochondrial DNA; *atp6* and *atp8*: ATP synthase subunits 6 and 8; *cob*: apocytochrome b; *cox1-3*: cytochrome c oxidase subunits 1–3; *nad1-6* and *nad4L*: NADH dehydrogenase subunits 1–6 and 4L; *rrnS* and *rrnL*: small

and large subunit ribosomal RNA (rRNA) genes; *trnX*: transfer RNA (tRNA) genes, where X is the one-letter abbreviation of the corresponding amino acid; s1–s7: mitochondrial genomic spacers; A+T region: the putative control region; PCG: protein coding gene; RSCU: Relative Synonymous Codon Usage; ENC, MILC: Measure Independent of Length and Composition; aa: amino acids; nt: nucleotides; bp: base pairs.

Authors' contributions

PS and MS carried out the molecular experiments. AB and EN designed and coordinated all experiments. EN performed the genomic analyses. All authors contributed to the manuscript and then read and approved the final version.

Additional material

Additional file 1
Additional file 1. List of primers and PCR conditions used in the sequencing of Ochogaster lunifer *mtDNA.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-331-S1.pdf]

Acknowledgements

We express our sincere thanks to Myron P. Zalucki (School of Integrative Biology, University of Queensland, Brisbane, Australia) who kindly provided the specimen of *Ochrogaster lunifer* used in present study. We thank Filippo Calore (Albignasego, Padova, Italy) who painted the icon of *O. lunifer* included in Figure 1, using as template a picture publically available at the CSIRO web site. Finally we thank two anonymous referees that provided very useful suggestions that helped to improve the manuscript.

References

1. Boore JL: **Animal mitochondrial genomes.** *Nucleic Acids Res* 1999, **27:**1767-1780.
2. Taanman JW: **The mitochondrial genome: structure, transcription, translation and replication.** *Biochim Biophys Acta* 1999, **1410:**103-123.
3. Resh VH, Cardé RG: **Insecta, Overview.** In *Encyclopedia of Insects* Edited by: Resh VH, CArdé RG. Academic Press, Burlington MA, USA; 2003:564-566. 1266pp.
4. Powell JA: **Lepidoptera (Moths, Butterflies).** In *Encyclopedia of Insects* Edited by: Resh VH, CArdé RG. Academic Press, Burlington MA, USA; 2003:631-663. 1266pp.
5. Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y: **Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina* and its close relative, the domesticated silkmoth, *Bombyx mori*.** *Mol Biol Evol* 2002, **19:**1385-1389.
6. Kim I, Lee EM, Seol KY, Yun EY, Lee YB, Hwang JS, Jin BR: **The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae).** *Insect Mol Biol* 2006, **15(2):**217-225.
7. Lee E-S, Shin KS, Kim M-S, Park H, Cho S, Kim C-B: **The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae).** *Gene* 2006, **373:**52-57.
8. Cameron SL, Whiting MF: **The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths.** *Gene* 2008, **408:**112-123.
9. Hong MY, Lee EM, Jo YH, Park HC, Kim SR, Hwang JS, Jin BR, Kang PD, Kim KG, Han YS, Kim I: **Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects.** *Gene* 2008, **413:**49-57.
10. Coates BS, Sumerford DV, Hellmich RL, Lewis LC: **Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*.** *Int J Biol Sci* 2005, **1:**13-18.
11. Taylor MFJ, McKechnie SW, Pierce N, Kreitman M: **The lepidopteran mitochondrial control region: structure and evolution.** *Mol Biol Evol* 1993, **10:**1259-1272.
12. Perna NT, Kocher TD: **Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes.** *J Mol Evol* 1995, **41:**353-358.
13. Cameron SL, Whiting MF: **Mitochondrial genomic comparisons of the subterranean termites from the genus Reticulitermes (Insecta: Isoptera: Rhinotermitidae).** *Genome* 2007, **50:**188-202.
14. Thao ML, Baumann L, Baumann P: **Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha).** *BMC Evolutionary Biology* 2004, **4:**25.
15. Covacin C, Shao R, Cameron S, Barker SC: **Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta).** *Insect Mol Biol* 2006, **15:**63-68.
16. Shao R, Campbell NJ, Barker SC: **Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera).** *Mol Biol Evol* 2001, **18:**858-865.
17. Wolstenholme DR: **Animal mitochondrial DNA: structure and evolution.** *Int Rev Cytol* 1992, **141:**173-216.
18. Negrisolo E, Minelli A, Valle G: **Extensive gene order rearrangement in the mitochondrial genome of the centipede *Scutigera coleoptrata*.** *J Mol Evol* 2004, **58:**413-423.
19. Sharp PM, Tuohy TMF, Mosurski KR: **Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes.** *Nucleic Acids Research* 1986, **14:**5125-5143.
20. Patthy L: **Protein evolution.** 2nd edition. Blackwell, London; 2008:374.
21. Wright F: **The 'effective number of codons' used in a gene.** *Gene* 1990, **87:**23-29.
22. Supek F, Vlahoviček K: **Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity.** *BMC Bioinformatics* 2005, **6:**182.
23. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN: **DNA barcodes distinguish species of tropical Lepidoptera.** *Proc Natl Acad Sci USA* 2006, **103:**968-971.
24. Lavrov DV, Brown WM, Boore JL: **A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*.** *Proc Natl Acad Sci USA* 2000, **97:**13738-13742.
25. Niehuis O, Naumann CM, Misof B: **Identification of evolutionary conserved structural elements in the mt SSU Rrna of Zygaenoidea (Lepidoptera): a comparative sequence analysis.** *Org Divers Evol* 2006, **6:**17-32.
26. Niehuis O, Yen S-H, Naumann CM, Misof B: **Higher phylogeny of zygaenid moths (Insecta: Lepidoptera) inferred from nuclear and mitochondrial sequence data and the evolution of larval cuticular cavities for chemical defence.** *Mol Phylogenet Evol* 2006, **39:**812-829.
27. Saito S, Tamura K, Aotsuka T: **Replication origin of mitochondrial DNA in insects.** *Genetics* 2005, **171:**1695-1705.
28. Vila M, Björklund M: **The utility of the neglected mitochondrial controlregion for evolutionary studies in Lepidoptera (Insecta).** *J Mol Evol* 2004, **58:**280-290.
29. Zhang DX, Hweitt GM: **Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies.** *Biochem Syst Ecol* 1997, **25:**99-120.
30. Boore JL: **The complete sequence of the mitochondrial genome of *Nautilus macromphalus* (Mollusca: Cephalopoda).** *BMC Genomics* 2006, **7:**182.
31. Patwary MU, Kenchington EL, Bird CJ, Zouros E: **The use of random amplified polymorphic DNA markers in genetic studies of the sea scallop *Placopecten magellanicus* (Gmelin, 1791).** *J Shellfish Res* 1994, **13:**547-553.

32. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P: **Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers.** *Ann Entomol Soc Am* 1994, **87:**651-704.
33. Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT: **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* 2006, **37:**545-579.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
35. Tatusova TA, Madden TL: **BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences.** *FEMS Microbiol Lett* 1999, **174(2):**247-250.
36. Thompson JD, Higgins DG, Gibson TJ: **Clustal-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
37. Smith AE, Marcker KA: **N-formylmethionyl transfer RNA in mitochondria from yeast and rat liver.** *J Mol Biol* 1968, **38:**241-243.
38. Fearnley IM, Walker JE: **Initiation codons in mammalian mitochondria: differences in genetic code in the organelle.** *Biochemistry* 1987, **26:**8247-8251.
39. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25:**955-964.
40. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0.** *Mol Biol Evol* 2007, **24:**1596-1599.
41. Supek F, Vlahovičekl K: **INCA: synonymous codon usage analysis and clustering by means of self-organizing map.** *Bioinformatics* 2004, **20:**2329-2330.
42. Flook PK, Rowell CH, Gellissen G: **The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome.** *J Mol Evol* 1995, **41:**928-941.
43. Fenn JD, Cameron SL, Whiting MF: **The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability.** *Insect Mol Biol* 2007, **16:**239-252.
44. Zhou Z, Huang Y, Shi F: **The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A+T-rich region of 70 bp in length.** *Genome* 2007, **50:**855-866.
45. Kim I, Cha SY, Yoon MH, Hwang JS, Lee SM, Sohn HD, Jin BR: **The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae).** *Gene* 2005, **353:**155-168.
46. Cameron SL, Barker SC, Whiting MF: **Mitochondrial genomics and the new insect order Mantophasmatodea.** *Mol Phylogenet Evol* 2006, **38(1):**274-279.
47. Yamauchi MM, Miya MU, Nishida M: **Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans.** *Insect Mol Biol* 2004, **13:**435-442.
48. Stewart JB, Beckenbach AT: **Insect mitochondrial genomics 2: The complete mitochondrial genome sequence of a giant stonefly, *Pteronarcys princeps*, asymmetric directional mutation bias, and conserved plecopteran A+T-region elements.** *Genome* 2006, **49:**815-824.
49. Cameron SL, Johnson KP, Whiting MF: **The mitochondrial genome of the screamer louse *Bothriometopus* (phthiraptera: ischnocera): effects of extensive gene rearrangements on the evolution of the genome.** *J Mol Evol* 2007, **65:**589-604.
50. Stewart JB, Beckenbach AT: **Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae).** *Genome* 2005, **48:**46-54.
51. Dotson EM, Beard CB: **Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*.** *Insect Mol Biol* 2001, **10:**205-215.
52. Shao R, Dowton M, Murrell A, Barker SC: **Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of Insects.** *Mol Biol Evol* 2003, **20:**1612-1619.
53. Shao R, Barker SC: **The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes.** *Mol Biol Evol* 2003, **20:**362-370.
54. Stewart JB, Beckenbach AT: **Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata*.** *Mol Phylogenet Evol* 2003, **26(3):**513-526.
55. Arnoldi FG, Ogoh K, Ohmiya Y, Viviani VR: **Mitochondrial genome sequence of the Brazilian luminescent click beetle *Pyrophorus divergens* (Coleoptera: Elateridae): mitochondrial genes utility to investigate the evolutionary history of Coleoptera and its bioluminescence.** *Gene* 2007, **405:**1-9.
56. Bae JS, Kim I, Sohn HD, Jin BR: **The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects.** *Mol Phylogenet Evol* 2004, **32(3):**978-985.
57. Friedrich M, Muqim N: **Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castanaeum*.** *Mol Phylogenet Evol* 2003, **26(3):**502-512.
58. Beard CB, Hamm DM, Collins FH: **The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects.** *Insect Mol Biol* 1993, **2:**103-124.
59. Mitchell SE, Cockburn AF, Seawright JA: **The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization.** *Genome* 1993, **36:**1058-1073.
60. Lessinger AC, Martins Junqueira AC, Lemos TA, Kemper EL, da Silva FR, Vettore AL, Arruda P, Azeredo-Espin AM: **The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae).** *Insect Mol Biol* 2000, **9:**521-529.
61. Junqueira AC, Lessinger AC, Torres TT, Da Silva FR, Vettore AL, Arruda P, Azeredo Espin AM: **The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae).** *Gene* 2004, **339:**7-15.
62. Lewis DL, Farr CL, Kaguni LS: ***Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons.** *Insect Mol Biol* 1995, **4:**263-278.
63. Ballard JW: **Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup.** *J Mol Evol* 2000, **51:**48-63.
64. Clary DO, Wolstenholme DR: **The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code.** *J Mol Evol* 1985, **22:**252-271.
65. Cameron SL, Lambkin CL, Barker SC, Whiting MF: **A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision.** *Syst Entomol* 2007, **32:**40-59.
66. Spanos L, Koutroumbas G, Kotsyfakis M, Louis C: **The mitochondrial genome of the mediterranean fruit fly, *Ceratitis capitata*.** *Insect Mol Biol* 2000, **9:**139-144.
67. Nardi F, Carapelli A, Dallai R, Frati F: **The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations.** *Insect Mol Biol* 2003, **12:**605-611.
68. Crozier RH, Crozier YC: **The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization.** *Genetics* 1993, **133:**97-117.
69. Cha SY, Yoon HJ, Lee EM, Yoon MH, Hwang JS, Jin BR, Han YS, Kim I: **The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae).** *Gene* 2007, **392:**206-220.
70. Castro LR, Ruberu K, Dowton M: **Mitochondrial genomes of *Vanhornia eucnemidarum* (Apocrita: Vanhorniidae) and *Primeuchroeus* spp. (Aculeata: Chrysididae): Evidence of rearranged mitochondrial genomes within the Apocrita (Insecta: Hymenoptera).** *Genome* 2006, **49:**752-766.

Chapter 5

Testing host plant associated differentiation on two parthenogenetic parasitoid species feeding on the same insect host in a forest system

---

Manuscript ready for submission as:

Testing host plant associated differentiation on two parthenogenetic parasitoid species feeding on the same insect host in a forest system.

I contributed to all parts of the experimental work, data analysis and paper writing.

Introduction

The actual mechanisms underlying speciation still stir discussion (Schluter 2001, Barton 2001, Coyne & Orr 2004). Gene flow and differential selection among populations play opposite roles in many of the models proposed. While in allopatric speciation geographic barriers prevent any gene flow allowing populations to diverge, in parapatric and sympatric speciation, local adaptation and disruptive selection can drive population differentiation in spite of the presence of gene flow (Coyne 2007). Sympatric speciation has been studied in particular in phytophagous insects. In fact the high species diversity of these organisms may be explained, in part, by their specialization on certain host plant species that can lead phytophagous insects to undergo differential selection and then sympatric speciation through host associated differentiation or HAD (Jaenike 1981, Diehl & Bush 1984, Berlocher & Feder 2002, Dres & Mallet 2002, Funk *et al.* 2002). Most phytophagous insects are strictly associated to one or few host plant species often phylogenetically related.

Several plant-insect systems have been studied in which genetic differentiation and ecological speciation have been found to occur in sympatry in relation to the host plant, both in agricultural (Bush 1969, Via 1999, Bethenod *et al.* 2004, Malausa *et al.* 2005) and unmanaged environments (Carroll *et al.* 1997, Stireman 2005). However, the required conditions for HAD to be found are still not well understood. Strong host preference, internal feeding mode (i.e., endophagy), short insect life cycles, univoltinism, reduced dispersal ability of adults and host fidelity (Berlocher & Feder 2002, Malausa *et al.* 2005, Price *et al.* 1980, Mopper & Strauss 1998, Abrahamson *et al.* 2001, Funk *et al.* 2002) are all factors that are thought to lead to ecological and/or allochronic separation and to consequent reproductive isolation among populations associated with different host plant species. As predicted by ecological models (Mopper & Strauss 1998, Futuyma & Moreno 1988, Jaenike 1990) and as suggested by some studies (Vialatte *et al.* 2005, Frantz *et al.* 2006) specialization of phytophagous insects to different host plant species could be favored especially on plants with long vegetative cycles such as perennials as these plants provide a more stable and predictable environment. In more variable environments like, for example, in agricultural systems, HAD is often associated with parthenogenesis (Miller *et*

al. 2005, Lozier *et al.* 2007, Charaabi *et al.* 2008, Carletto *et al.* 2009, Peccoud 2009), a process that allow to rapidly amplify adaptive responses to selection in ephemeral environments (Lynch 1984). Parthenogenesis can strongly affect the genetic structure of populations; more or less specific predictions about parthenogenesis effects are provided by several mutational and ecological models (Normark & Moran 2000, Normark *et al.* 2003, Simon *et al.* 2003). However parthenogenesis, if strictly clonal, may promote adaptive gene complexes, thus facilitating ecological specialization (Sunnucks *et al.* 1997).

Diversification of phytophagous insects on their host plants can produce a 'sequential radiation' on taxa that use these herbivores as a resource (Abrahamson W.G. & Blair C.P. 2007) like predators and parasitoids. Like phytophagous insects, insect parasitoids are another group of parasites that show an intimate association with their hosts. However, unlike phytophagous insects, very few studies have been done to date on HAD at this trophic level. HAD cascading up trophic levels have been found in parasitoids of *Eurosta solidaginis* and of *Gnorimoschema gallaesolidaginis* associated with *Solidago altissima* and *Solidago gigantea* (Cronin & Abrahamson 2001, Stireman *et al.* 2006). Similarly, HAD have been found to cascade up in *Diachasma alloeum* attacking *Rhagoletis pomonella* on hawthornes, blueberries and apples (Forbes *et al.* 2009). Host-plants are an important component of parasitoids ecology. For instance, physical and chemical traits of host-plants influence host searching and oviposition behavior in several parasitoid species (Vinson 1998, Fatouros *et al.* 2008). In addition, differences in herbivore hosts' nutritional quality when feeding on different host-plant species might produce phenotypic differences in parasitoid traits related to fitness such as adult mass, adult longevity, fecundity, and developmental time (Medina & Barbosa 2008). In this way host plants may indirectly affect the survival and fitness of parasitoids through their influence on these traits (Price *et al.* 1980, Vinson 1999). In parasitoids HAD could thus also occur without requiring the differentiation of the herbivore host.

In this paper, we explore HAD in a system involving two egg parasitoid species feeding on the same herbivore species on two different host plant species. The herbivore is the pine processionary moth, *Thaumetopoea pityocampa* Denis & Schiffermüller (Lepidoptera: Notodontidae), one of the main pests of pine forests in southern Europe and in the Mediterranean region (Masutti & Battisti 1990). This herbivore has recently shown

an expansion of its range as a consequence of climate change (Battisti *et al.* 2005). The egg parasitoids are *Baryscapus servadeii* Domenichini (Hymenoptera: Eulophidae), a specialist of *T. pityocampa* and of few other sibling species feeding on conifers (Graham 1991), and *Ooencyrtus pityocampae* Mercet (Hymenoptera: Encyrtidae), associated with conifer-feeding *Thaumetopoea* but also able to develop in eggs of other Lepidoptera and Hemiptera occurring in pine forests (Battisti *et al.* 1988). Both parasitoid species reproduce by thelytokous parthenogenesis and males are very rarely observed (Graham 1991, Battisti *et al.* 1990). Nothing is known about the frequency of sexual reproduction in any of these two parasitoid species. The two host plants of *T. pityocampa* are *Pinus nigra* Arnold and *Pinus sylvestris* L., co-occurring in both native and introduced stands all over the range of the pine processionary moth in southern Europe (Richardson 1998). There is no evidence that moth populations associated with either one of these two hosts differ in ecology or genetics (Stastny *et al.* October 2006, Kerdelhue *et al.* 2006). By sampling populations where the two host-plants co-occur, we want to address the following question: 1. Is there evidence of HAD in each of the two parthenogenetic egg parasitoid species? 2. in which way host-range influences HAD in parasitoid species?

Materials and Methods

Sample collection

Egg batches of *T. pityocampa* were collected from both *P. nigra* and *P. sylvestris* at 6 localities in North-Eastern Italy (Table 1, Fig. 1) during August-November 2008. Egg batches were then stored in vials under outdoor conditions and taken indoor in February to get parasitoid emergence. Adult parasitoids were identified, counted and stored in 70% ethanol until DNA extraction.

DNA extraction and fingerprint analysis

Genomic DNA was extracted from 62 individuals of *B. servadeii* and 118 individuals of *O. pityocampae* (Table 2) using Qiagen ® DNeasy kit (Valencia, CA) following the manufacturer recommended protocol for animal tissue (Qiagen 2002). DNA

was eluted in 100 μl of AE buffer and its concentration and quality were then assessed using a NanoDrop-1000 spectrophotometer (NanoDrop, Wilmington, DE).

AFLP protocol (Vos *et al.* 1995) was followed with some modifications concerning the use of fluorescent dye labeled selective primers. Each of the parasitoid species was processed independently. Samples for each of the species were processed in a randomized order. Restriction/ligation reactions were carried out with EcoRI/MseI endonucleases and their respective adaptors. Five primers combinations were used in the selective amplification: EcoRI-ACT/MseI-CAT, EcoRI-AG/MseI-CAT, EcoRI-AG/MseI-CAC, EcoRI-AC/MseI-CAA, EcoRI-AG/MseI-CAA. These primers were used for each individual of both parasitoids species; for unknown reasons EcoRI-AC/MseI-CAA combination didn't yield any AFLP markers from *B. servadeii* and so was only used in *O. pityocampae.* Fragment separation and detection were carried out using a 3130xl ABI DNA sequencer. Presence or absence of each band was scored using GeneMapper v4.0. Reproducibility of AFLP patterns was assessed by a positive control repeated twice for each plate, whereas reliability of results was confirmed by checking bands by eye. AFLP profiles were converted in a 0-1 matrix and all monophorphic loci removed from each dataset.

Assays for intracellular symbionts

To test for presence of intracellular symbionts a PCR was performed in all samples of both species using specific primers for *Wolbachia* (99F and 994R from O'Neill *et al.* 1992)), *Cardinium* (Ch-F and Ch-R from Zchori-Fein & Perlman (2004)) and *Rickettsia* (rct1f 5'-CCGCGTCAGATTAGGTAGTT-3' and rct1r 5'-TCAGTTGTAGCCCAGATGAC-3'). Positive samples for *Wolbachia* were then re-checked by PCR using specific primers for *Wolbachia* A and B strains (wsp81F and wsp691R from (Zhou *et al.* 1998)). Positive results of all bacteria species were confirmed by sequencing of few samples for each type. Sequences were identified by a BLAST search in the GenBank database (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Data analysis

The mean number of emerged parasitoids per egg batch was compared between parasitoids from *P. nigra* and *P. sylvestris* and among collection sites using multifactorial ANOVA. The populations of Tregnago and Calbarina were excluded from the analysis for a dissimilar sampling period in the former site and lack of one host plant (*P. sylvestris*) in the latter site.

To confirm that both samples and AFLP loci were enough to statistically support our results, SESim (Medina *et al.* 2006) was used for each dataset of the two parasitoid species. In addition, as size homoplasy of AFLP bands could lead to underestimates of genetic diversity and genetic divergence, correlation between marker size and frequencies of AFLP fragments was also tested using the software AFLPsurv (Vekemans *et al.* 2002). The number of genetic clusters (K) present in each of our two datasets (i.e., the two parasitoid species studied) was obtained using Structure 2.3.2. (Falush *et al.* 2007). Structure performs the Bayesian assignment analysis of Pritchard *et al.* 2000 to assign individuals to genetically similar clusters. We used an admixture model and no prior information about populations. All running conditions were set as suggested by (Evanno *et al.* 2005): we ran 20 replicates of each simulation with a range of K between 1 and 10, and for each K value a burn-in of 10000 and 10000 post burn-in Markov chain Monte Carlo (MCMC) iterations. The number of K in each dataset was then chosen by the method described in Evanno *et al.* 2005. The Structure output was visualized using the Distruct software (www-hto.usc.edu/_noahr/distruct.html). Association of the different clusters obtained by Structure with host plant species was tested with a Fisher exact test.

To evaluate differences in genetic variability between the specialist and the generalist parasitoids, estimates of genetic diversity (i.e., band richness, percent of polymorphic loci), inbreeding coefficients, multilocus linkage disequilibrium and clonal diversity for each main cluster identified by Structure were obtained. Band richness (Br) was calculated using the AFLPDIV program (Coart *et al.* 2005). Band richness refers to the number of phenotypes expected at each locus when correcting for different sample size (scored AFLP fragment); it can be considered as an analogue of allelic richness (Coart *et al.* 2005). Inbreeding coefficient (F_{is}) was calculated using ABC program (Foll *et al.* 2008) a

software that, through a Bayesian approach, takes into account the dominant nature of AFLP markers. This software calculates the probability of sampling an individual inbred for a particular locus, and the resulting F_{is} can therefore range from 0 to 1. Measure of linkage disequilibrium over all loci (i.e., multilocus linkage disequilibium) was obtained by the standardised index of association (rd) (Agapow & Burt 2001). To measure the clonal diversity of each cluster and sub-cluster, we used the evenness of Simpson's index (Grapputo *et al.* 2005). Further, we performed a comparison of the distributions of pairwise genetic similarities (Jaccard's index) among AFLP profiles inside the main clusters of Structure using the non parametric test of Kolmogorov-Smirnov. In order to compare the partition of genetic variability among sampled populations and the partition among Structure clusters, we performed an analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) for dominant data using GenAlEx 6.0.

Results

Parasitoid distribution on *T. pityocampa* egg batches on different host plant species. The mean number of emerging *B. servadeii* adults from *T. pityocampa* egg batches on *P. nigra* and *P. sylvestris* was significantly different ($F_{1,86}$ = 9.53, P = 0.003). Significant difference in the number of emerging *B. servadeii* adults were also observed among localities ($F_{2,86}$ = 42.75, P = 0.000). Differences in adult *O. pityocampae* emergence from *T. pityocampa* egg batches was significantly different only among localities ($F_{2,86}$ = 5.97, P = 0.004). For both parasitoid species the interaction between localities and host plants was not significant.

Assays for intracellular symbionts

All *B. servadeii* samples were infected with a *Rickettsia* strain confirmed by sequencing. The strain we found show a 98% similarity to the strain infecting *Neochrysocharis formosa* (Hymenoptera: Eulophidae) (Hagimori *et al.* 2006). All *O. pityocampae* samples were infected with *Wolbachia*. Sequences of a subset of samples showed that *Wolbachia* in *O. pityocampae* belongs to the B strain.

Genetic data analysis

Five primer combinations used in AFLP analysis produced a total of 52 polymorphic bands in *B. servadei* and 118 polymorphic bands in *O. pityocampae*. SESim simulations confirmed that the number of both samples and bands was enough to have a statistical support (SESim <0.03) for all the following analyses. In addition, AFLPsurv found no correlation between fragment size and fragment frequency, thus excluding homoplasy for the AFLP fragment sizes considered (50-500 bp) in this study. The analyses carried out with Structure on *B. servadeii* and *O. pityocampae* datasets revealed the presence of two main genotypes groups or clusters (K=2) in both species (Fig. 3a,b). Diversity indices for these two clusters obtained with Structure for both parasitoid species are summarized in Table 3. In *B. servadeii*, the two main clusters present a slight difference in their genetic variability, with cluster A showing a higher band richness than cluster B. Moreover estimates of inbreeding coefficient (F_{is}) for the two clusters show a non overlapping 95% highest posterior density interval and thus a significant difference between the two values, with a higher level of heterozygosity in cluster B compared to cluster A. Multilocus linkage disequilibrium is found with values significantly different from zero in both clusters, with a higher value in cluster B. The latter cluster (B) is also the one with lower clonal diversity (0.90 compared with 1 for cluster A). In contrast, in *O. pityocampae*, band richness seems to be quite similar between the two main clusters. Moreover no significant differences from a value expected at Hardy-Weinberg equilibrium (F_{is} = 0.5) were found for the inbreeding coefficient values although cluster A presents a slightly higher level of heterozigosity than clusters B. Both clusters A and B show values of multilocus linkage disequilibrium significantly different from zero with a low clonal diversity (0.846 and 0.940 respectively).

In both *B. servadeii* and *O, pityocampae*, the comparison between the distributions of pairwise genetic similarities inside each of the main clusters confirmed a significant difference between the variability of the two clusters (p<0.0001, Kolmogorov Smirnov test) (Fig. 4). In *B. servadeii* the re-run of Structure performed for each of the main clusters (i.e., A and B) showed a further grouping of samples: two sub-clusters for cluster A (A1 and A2)(Fig. 5a), and three sub-clusters for cluster B (B1, B2, B3) (Fig. 5b). A significant

association with different host plant species was found for cluster B, where sub-cluster B2 was found with a higher frequency in *P. nigra* populations (Fisher exact test, $p<0.05$). In *O. pityocampae* the re-run of each of the main clusters (A and B) produced three sub-clusters (K=3) for the cluster A (A1, A2, A3) (Fig. 6a) and three sub-clusters (K=3) for cluster B (B1, B2, B3) (Fig. 6b). No significant association with host plant species was found in any of the clusters or sub-clusters. It seems there is instead a geographic pattern comprising a north-east and an eastern region: inside cluster A (Fig. 6a), sub-cluster A1 is present in southern (Tregnago and Calbarina) and north-western (Venosta) populations, whereas A2 and A3 are mainly found in eastern populations (Longarone and Carnia) and in Rovereto.

In the B cluster (Fig. 6b), sub-cluster B1 is mostly associated to northern populations (Venosta, Rovereto, Longarone and Carnia), whereas B2 and B3 are mostly present in the southern populations of Tregnago and Calbarina.

AMOVA on *B. servadeii* revealed that, when considering individuals grouped per site of collection and/or per host plant, almost all genetic variation is within populations (from 97% to 99% for the different groupings, Table 4). A significant component of variance is also explained when individuals are grouped according to Structure sub-clusters (39%); when sub-clusters are grouped into the two main clusters A and B this component increases (distributed between clusters, 25%, and among sub-clusters, 19%, Table 4). In *O. pityocampae*, AMOVA (Tab. 4) revealed a significant percentage of variation explained by groupings individuals only per site of collection (16%) while grouping individuals per host plant has no influence on variance components. Variation among groups remarkably increases when considering only sub-clusters (41%) and both clusters and sub-clusters (distributed between clusters, 25%, and among sub-clusters, 22%, Table 4).

Discussion

In this study we analyzed two egg parasitoids *B. servadei* and *O. pityocampae* feeding on the same host, *T. pityocampa*, a phytophagous insect which does not show genetic differentiation when in association with different host plant species. We focused only on parasitoids emerged from egg batches found on two pine species, *P. nigra* and *P.*

sylvestris, to test whether there is host plant associated differentiation in parasitoids in a system whereas a corresponding differentiation is absent in their insect host. As both species of parasitoids studied reproduce by thelytokous parthenogenesis we first investigated the presence of endosymbionts in our samples. We found that all *B. servadeii* samples are infected by *Rickettsia* while all *O. pityocampae* samples are associated with a B strain of *Wolbachia*. Although more studies (e.g. antibiotic treatments) would be needed to determine the association between the presence of these endosymbionts and parthenogenesis, we could hypothesize that parthenogenesis in both species is due to the presence of endosymbionts.

Analyzing the genetic structure of parasitoids we found a subdivision of individuals in two main clusters in both species. In *B. servadeii* these two main clusters show a significant difference in their values of heterozygosity. The high level of heterozigosity (F_{is}<0.5) associated with low clonal diversity and the significant linkage disequilibrium found in cluster B, is in agreement with values expected in strictly apomictic populations (Vialatte *et al.* 2005, Simon *et al.* 1999, Delmotte *et al.* 2002, Vorburger *et al.* 2003). Interestingly, apomixes is the mechanism hypothesized (Adachi-Hagimori 2008) for the *Rickettsia* induced parthenogenesis in another Eulophidae, *Neochrysocharis formosa,* to date the only other Hymenoptera known to be infected by *Rickettsia* (Hagimori *et al.* 2006). In *O. pityocampae* the estimated level of heterozigosity in both the main clusters A and B does not show to be significantly different from Hardy-Weinberg equilibrium (F_{is}=0.5). In Hymenoptera with *Wolbachia* induced parthenogenesis complete homozygosis should be expected, due to automictic parthenogenesis with gamete duplication (Stouthamer and Kazmer, 1994). However, this mechanism has been demonstrated only in few hymenopteran species (van Wilgenburg *et al.* 2006), and other mechanisms may occur in which heterozygosity is preserved to some degree (Stouthamer 1997, Weeks *et al.* 2002).

In both *B. servadei* and *O. pityocampae* the distribution pattern of pairwise genetic similarities is significantly different between the two main clusters A and B (Fig. 4). This result is similar in other parthenogenetic species such as aphids (Sunnucks *et al.* 1997, Delmotte *et al.* 2002, Simon *et al.* 1999, Papura *et al.* 2003, Vorburger *et al.* 2003) in which lineages undergoing cyclical parthenogenesis (i.e., several parthenogenetic generations alternating with one sexual generation) show a significantly higher genetic

diversity than sympatric lineages with obligate parthenogenesis. We might thus infer that the B cluster in *B. servadeii* and the A cluster in *O. pityocampae* could include strictly parthenogenetic individuals while the A cluster in *B. servadeii* and the B cluster in *O. pityo*campae could be composed by individuals that can sexually reproduce. Presence of endosymbionts inducing parthenogenesis in all samples analyzed could seem in contrast with the existence of sexual lineages.Nevertheless, for *Wolbachia* it is known that endosymbiont density, that is an important factor in parthenogenesis induction (Stouthamer 1997), can be regulated by host genotype with significant differences among host strains also at the intra-specific level (Clark *et al.* 2003, Kondo *et al.* 2005, Mouton *et al.* 2007).

Inside each of the main clusters of both parasitoid species there is a further structure with clear differences between the two species. *B. servadeii* samples are grouped in five sub-clusters (two inside cluster A and three inside cluster B) showing no geographical distribution (AMOVA, Tab. 4). One of the sub-clusters identified in cluster B (B2) seems to be associated with one of the host plant, *P. nigra*. This result is also partially confirmed by the analysis of the attack frequency of egg-batches by *B. servadeii* on the two host plants: in the three populations in which the comparison was possible, *B. servadeii* showed a significant preference for *P. nigra* (Fig. 2).

Factors that could be involved in host plant choice by parasitoids are many. *P. nigra* and *P. sylvestris* present different chemical compositions and morphological differences that could influence their possibility to be found by parasitoid species searching for hosts (Vinson, 1998). Plants can also signal to parasitoids the presence of phytophagous insects in the early stages of the attack: it has been shown that *P. sylvestris* can emit volatiles to attract the eulophid egg parasitoid *Chrysonotomyia ruforum* after induction by egg deposition of the pine sawfly *Diprion pini* (Hilker *et al.* 2002). Interestingly *B. servadeii* is known to search its host through long-range orientation with visual and chemical cues (Battisti 1989) and for this reason an association of *B. servadeii* with the host plant has already been hypothesized (Battisti 1989). On the other hand *O. pityocampae* can detect egg batches by sex pheromones traces (Battisti, 1989) and so no influence by the host plant should be expected. The six sub-clusters identified in *O. pityocampae* do not actually show any association with the host-plant. They present instead a clear geographic pattern for each of the main clusters (Fig 6a,b). This pattern may be related to the west-east gradient in *T.*

pityocampa egg size that is thought to be due to an adaptation to needle toughness of the main host *P. nigra* (Zovi *et al.* 2008).

The different genetic structure in these two species living in the same environment could be related to the two different strategies that these parasitoids follow to find their host. The insect host, *T. pityocampa,* undergoes regular variation of its population densities (Hodar *et al*. 2004) and its presence in a locality is ensured by prolonged diapause of its pupae. To cope with the scarcity of its main host insect the generalist *O. pityocampae* can 'wait' for its prefer host population densities to increase while ovipositing on alternative host (Battisti 1989); on the other hand, the specialist *B. servadeii,* which has no alternative hosts, has to actively search *T. pityocampa* egg batches at several localities, being in this way more prone to local extinction and recolonization. The association with the host plant of one of *B.servadeii* lineages could be thus the result of one of the diverse strategies adopted by the parasitoid to better find its host in a long range research.

Site	Coordinates	Elevation (m)
1) Venosta	10°46’ E 46°38’ N	1050
2) Rovereto	11° 03’ E 45°53’N	550
3) Tregnago	11° 09’ E 45°30’N	510
4) Calbarina	11° 43’ E 45°16’N	200
5) Longarone	12° 19’ E 46°16’N	475
6) Carnia	13° 08’ E 46°22’N	320

Table 1. Sampling sites.

	Pinus nigra			*Pinus sylvestris*		
Site	N egg-batches	*B.s.*	*O.p.*	N egg-batches	*B.s.*	*O.p.*
Venosta	48	13	12	19	13	10
Rovereto	31	0	12	25	0	12
Tregnago	35	0	13	20	7	13
Calbarina	30	0	11	-	-	-
Longarone	17	11	5	20	8	8
Carnia	57	10	14	27	0	8

Table 2. Egg batches from which parasitoids had emerged and number of parasitoids used in the genetic analysis for each host plant species. *B.s.*: *Baryscapus servadeii*; *O.p.*: *Ooencyrtus pityocampae*.

	N	Br	PPL 1%	F_{is} [95% HPDI]	r_d	E(k)
Baryscapus servadeii						
Cluster A	27	1.961	0.961	0.772 [0.405-1]	0.010**	1
Cluster B	35	1.717	0.745	0.088 [0-0.265]	0.039**	0.90
Ooencyrtus pityocampae						
Cluster A	55	1.852	0.875	0.224 [0-0.581]	0.042**	0.846
Cluster B	63	1.875	0.875	0.627 [0.155-1]	0.033**	0.940

Table 3. Diversity statistics within main clusters found in Structure: band richness (Br), calculated in each species with the lowest N for each cluster, percent of polymorphic loci (PPL), inbreeding coefficient (F_{is}) with its highest probability distribution interval (HPDI), standardized index of association among loci (r_d), and evenness of Simpson's index k, E(k).

	Source of variation	*Baryscapus servadeii*		*Ooencyrtus pityocampae*	
		Variance components	Percentage of variation	Variance components	Percentage of variation
Populations per site and host plant	Among populations	0.130	1	1.540**	14**
	Within populations	8.571	99	9.329	86
Groups per site	Among groups	0.257	3	1.807**	16**
	Among pops within groups	0.000	0	0.000	0
	Within populations	8.571	97	9.329	84
Groups per host-plant	Among groups	0.000	0	0	0
	Among pops within groups	0.194	2	1.757**	16**
	Within populations	8.571	98	9.329	84
Populations per sub-clusters of Structure	Among populations	3.660**	39**	4.792**	41**
	Within populations	5.826	61	6.810	59
Groups per main clusters of Structure	Among groups	2.559**	25**	3.153**	25**
	Among populations	2.021**	19**	2.859**	22**
	Within populations	5.826	56	6.810	53

Table 4. Analysis of molecular variance (AMOVA) considering individuals grouped per site/host and per Structure clustering.

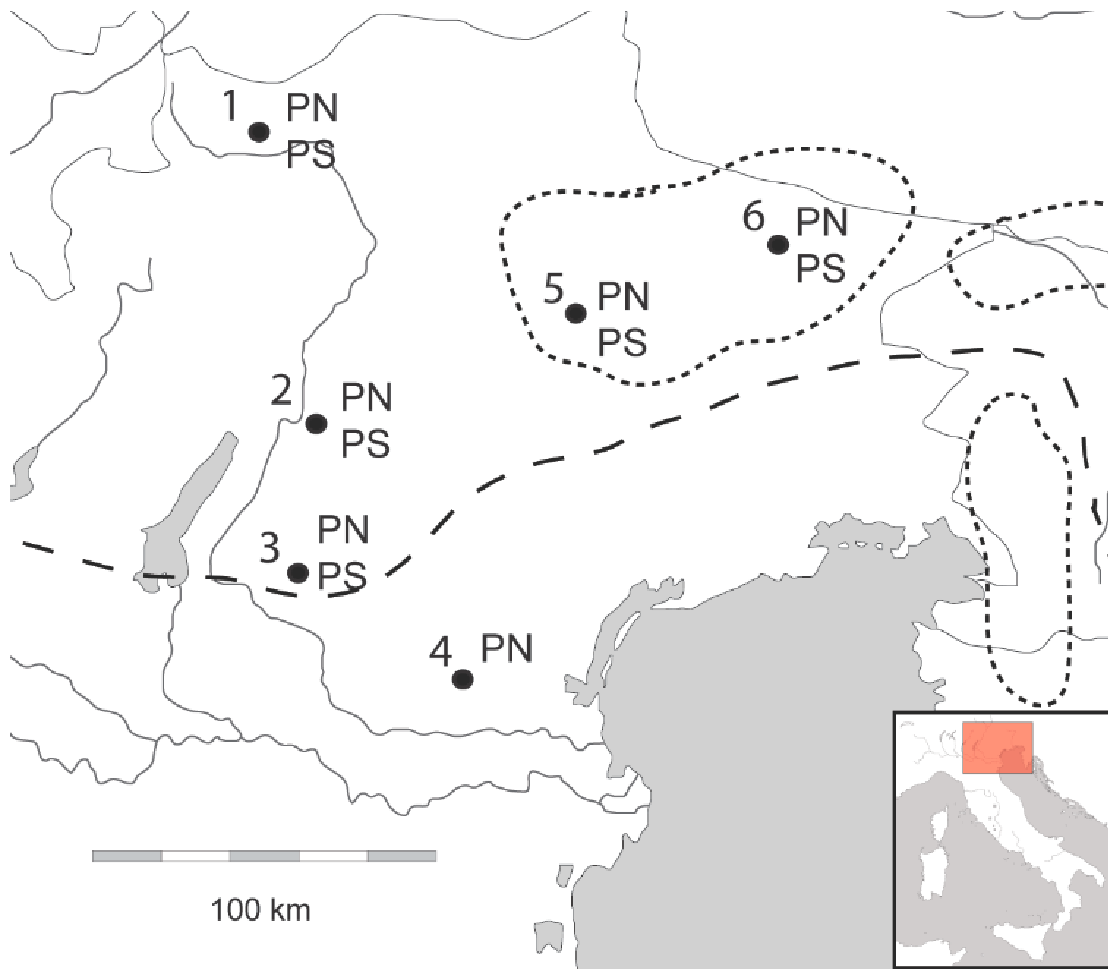


Fig. 1. Geographic location of sampling sites, with the indication of the native range of *Pinus nigra* (short dashed line) and the southern edge of the native range of *Pinus sylvestris* (long dashed line): Venosta (1), Rovereto (2), Tregnago (3), Calbarina(4), Longarone(5), Carnia (6). For each locality the pine species present are indicated with PN (*Pinus nigra*) and PS (*Pinus sylvestris*).

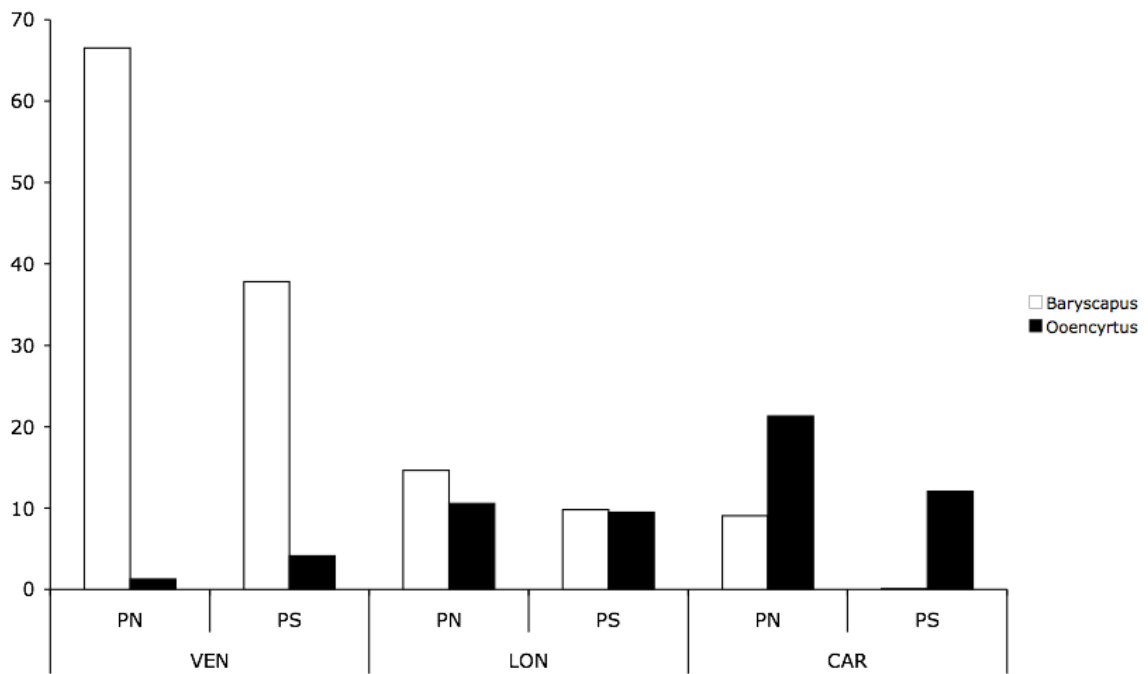
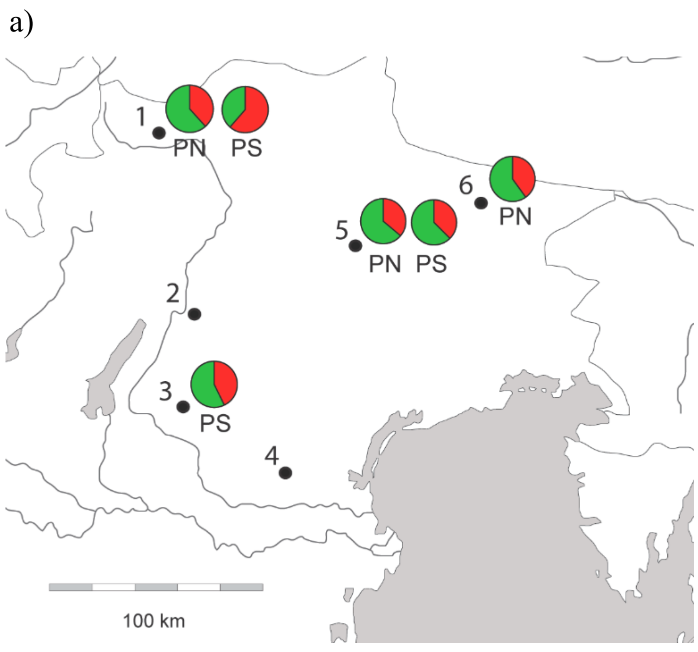
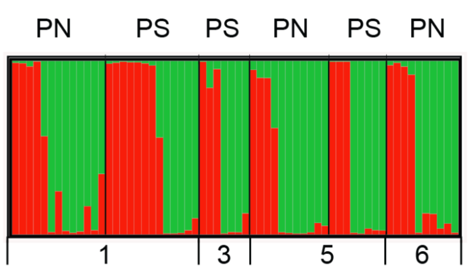


Fig. 2 . Number of parasitoids emerged from the total of egg batches collected for each pine species (PN: *Pinus nigra*, PS: *Pinus sylvestris*) and each locality (Ven: Venosta, Lon: Longarone, Car: Carnia).





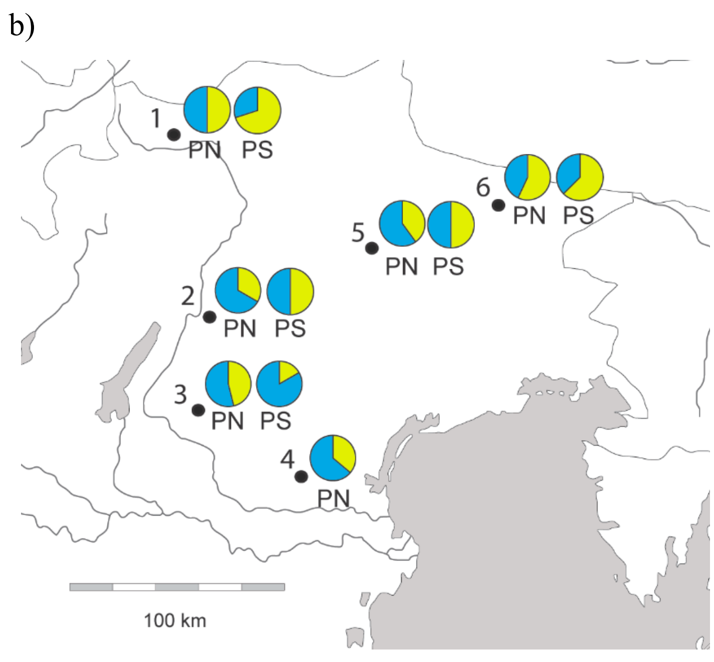
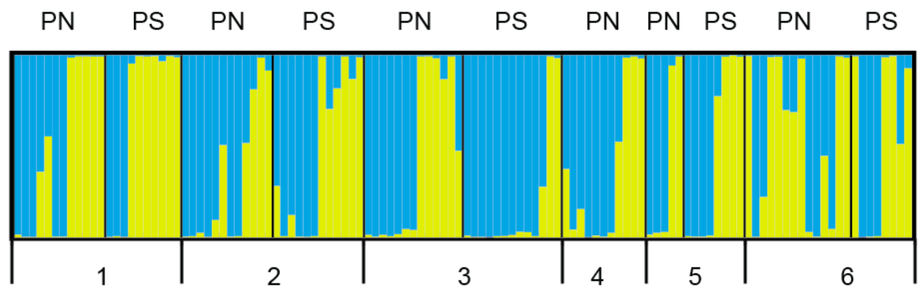
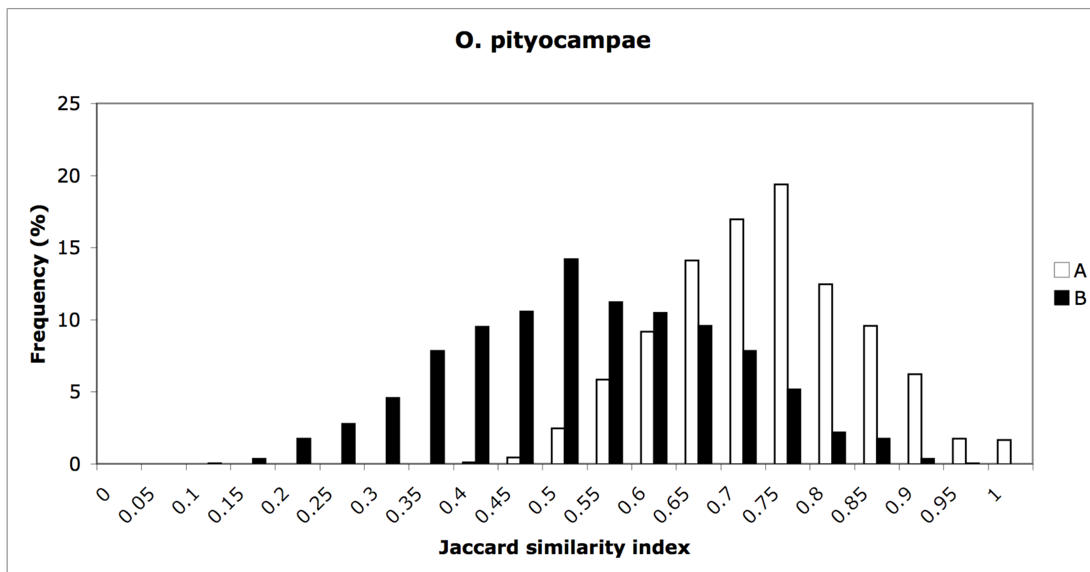


Fig 3a,b. Graphical presentation of population structure analyses for *B. servadeii* (a) and *O. pityocampae* (b). A single vertical line represent each individual. Colored segments lenghts within each line are proportional to the estimated membership to inferred clusters. Each cluster is represented by a single color. The distribution of the clusters is reported on the map. Population and pine species are indicated as in Fig 1.





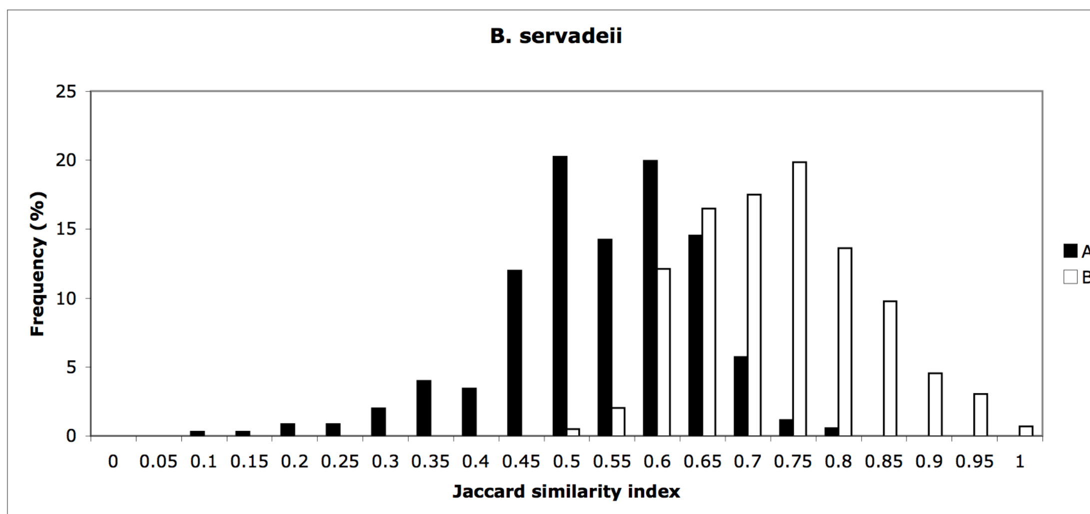
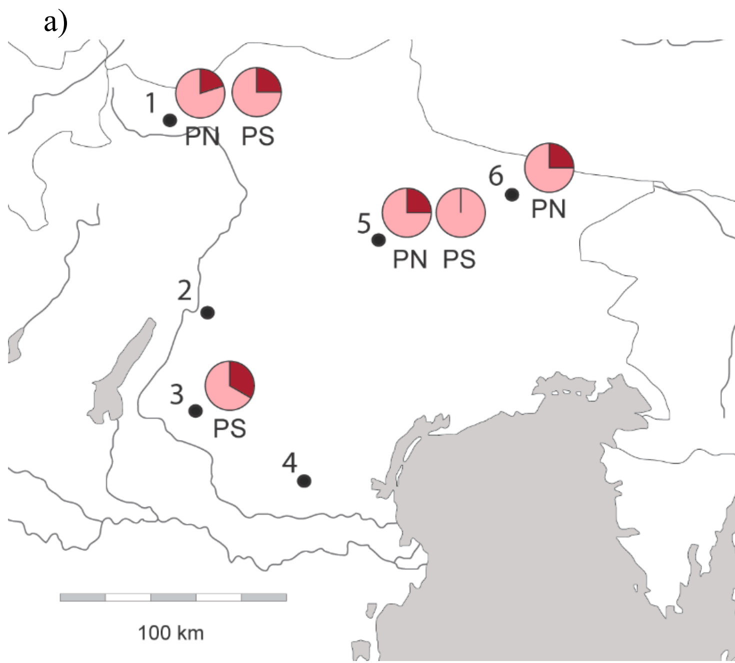
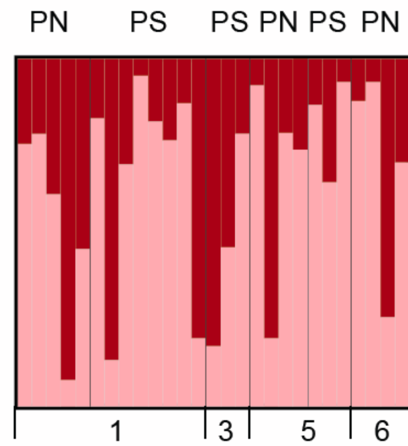


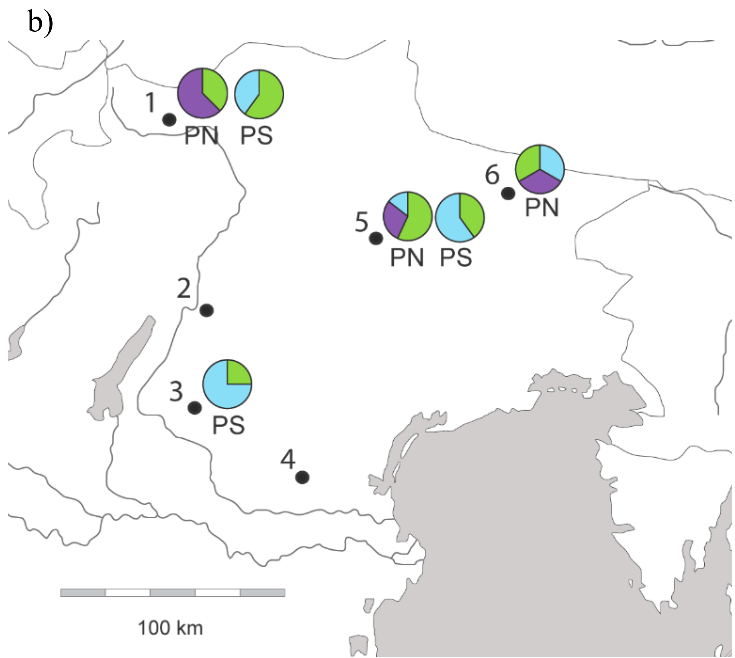
Fig. 4 Distribution of pairwise genetic similarities (Jaccard's index) calculated between genotypes inside the two main Structure clusters (indicated with A and B) in *B. servadeii* and *O. pityocampae*.

a)





b)



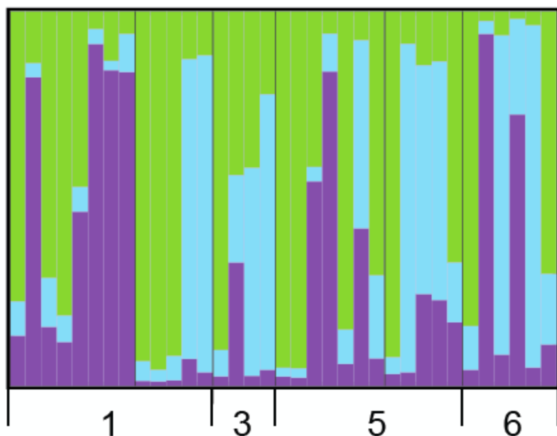
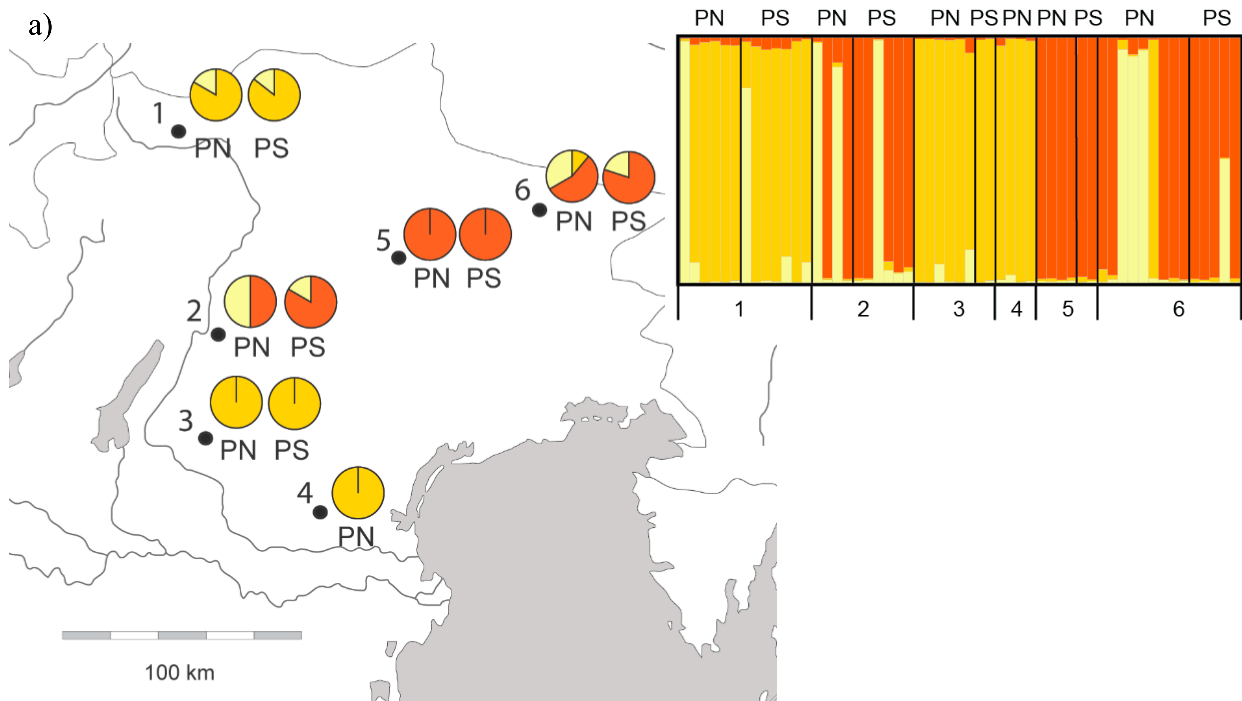
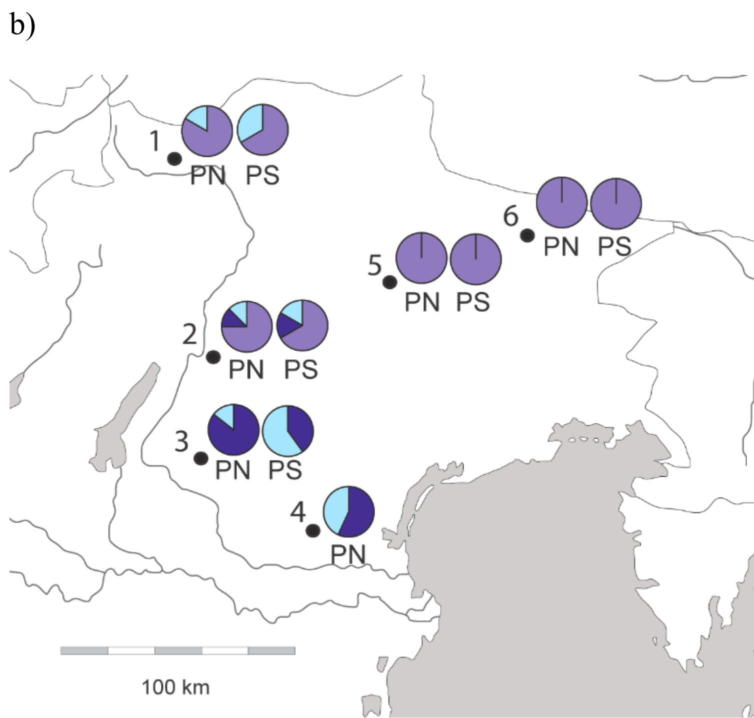


Fig 5a,b. Graphical presentation of population structure analyses for *B. servadeii* for cluster A (a) and cluster B (b). A single vertical line represent each individual. Colored segments lenghts within each line are proportional to the estimated membership to inferred clusters. Each cluster is represented by a single color. The distribution of the clusters is reported on the map. Population and pine species are indicated as in Fig 1.

a)



b)



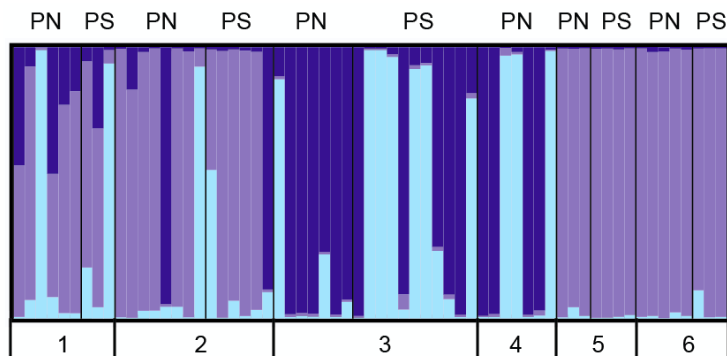


Fig 6a,b. Graphical presentation of population structure analyses for *O. pityocampae* for cluster A (a) and cluster B (b). A single vertical line represent each individual. Colored segments lenghts within each line are proportional to the estimated membership to inferred clusters. Each cluster is represented by a single color. The distribution of the clusters is reported on the map. Population and pine species are indicated as in Fig 1.

References

Abrahamson W.G., Blair C.P. (2007) Sequential radiation through host-race formation: herbivore diversity leads to diversity in natural enemies. In: *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects.* (ed. Tilmon K.J.) pp. 188-202 University of California Press, Berkeley, CA.

Abrahamson W.G., Eubanks M.D., Blair C.P., Whipple A.V. (2001) Gall flies, inquilines, and goldenrods: a model for host-race formation and sympatric speciation. *Integrative and Comparative Biology,* **41**, 4: 928-938.

Adachi-Hagimori T. (2008) Development of a Multiplex Method to Discriminate Between *Neochrysocharis formosa* (Hymenoptera: Eulophidae) Reproductive Modes. *Journal of Economic Entomology.,* **101**, 4: 1510-1514.

Agapow P., Burt A. (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes,* **1,** 1-2: 101-102.

Barton N.H. (2001) Speciation. *Trends in Ecology & Evolution,* **16**, 7: 325.

Battisti A. (1989) Field studies on the behaviour of two egg parasitoids of the pine processionary moth *Thaumetopoea pityocampa*. *Biocontrol,* **34**, 1: 29-38.

Battisti A., Colazza S., Roversi P., Tiberi R. (1988) Alternative hosts of *Ooencyrtus pityocampae* (Mercet)(Hymenoptera, Encyrtidae) in Italy. *Redia,* **71**, 2: 321-328.

Battisti A., Ianne P., Milani N., Zanata M. (1990) Preliminary accounts on the rearing of *Ooencyrtus pityocampae* (Mercet)(Hym., Encyrtidae). *Journal of Applied Entomology,* 110, 1-5: 121-127.

Battisti A., Stastny M., Netherer S., Robinet C., Schopf A., Roques A., Larsson S. (2005) Expansion of geographic range in the pine processionary moth caused by increased winter temperatures. *Ecological Applications*, **15**, 6: 2084-2096.

Berlocher S., Feder J. (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology,* **47**, 1: 773-815.

Bethenod M., Thomas Y., Rousset F., Frérot B., Pélozuelo L., Genestier G., Bourguet D. (2004) Genetic isolation between two sympatric host plant races of the European corn borer, *Ostrinia nubilalis* Hubner. II: assortative mating and host-plant preferences for oviposition. *Heredity,* **94**, 2: 264-270.

Bush G.L. (1969) Sympatric Host Race Formation and Speciation in Frugivorous Flies of the Genus *Rhagoletis* (Diptera, Tephritidae). *Evolution,* **23**, 2: 237-251.

Carletto J., Lombaert E., Chavigny P., Brévault T., Lapchin L., Vanlerberghe-Masutti F. (2009) Ecological specialization of the aphid *Aphis gossypii* Glover on cultivated host plants. *Molecular Ecology,* **18**, 10: 2198-2212.

Carroll S.P., Dingle H., Klassen S.P. (1997) Genetic Differentiation of Fitness-Associated Traits Among Rapidly Evolving Populations of the Soapberry Bug. *Evolution,* **51**, 4: 1182-1188.

Charaabi K., Carletto J., Chavigny P., Marrakchi M., Makni M., Vanlerberghe-Masutti F. (2008) Genotypic diversity of the cotton-melon aphid *Aphis gossypii* (Glover) in Tunisia is structured by host plants. *Bulletin of Entomological Research,* **98**, 4: 333-341.

Clark M.E., Veneti Z., Bourtzis K., Karr T.L. (2003) *Wolbachia* distribution and cytoplasmic incompatibility during sperm development: the cyst as the basic cellular unit of CI expression. *Mechanisms of Development,* **120**, 2: 185-198.

Coart E., Glabeke S.V., Petit R., Bockstaele E.V., Roldàn-Ruiz I. (2005) Range wide versus local patterns of genetic diversity in hornbeam (*Carpinus betulus* L.). *Conservation Genetics,* **6**, 2: 259-273.

Coyne J.A. (2007) Sympatric speciation. *Current Biology,* **17**, 18: R787-R788.

Coyne J.A., Orr H.A. (2004) *Speciation,* Sinauer Associates, Sunderland, MA, USA.

Cronin J.T., Abrahamson W.G. (2001) Do parasitoids diversify in response to host-plant shifts by herbivorous insects? *Ecological Entomology,* **26**, 4: 347-355.

Delmotte F., Leterme N., Gauthier J., Rispe C., Simon J. (2002) Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology,* **11**, 4: 711-723.

Diehl S.R., Bush G.L. (1984) An Evolutionary and Applied Perspective of Insect Biotypes. *Annual Review of Entomology,* **29**, 1: 471-504.

Dres M., Mallet J. (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences,* **357**, 1420: 471-492.

Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology,* **14**, 8: 2611-2620.

Excoffier L., Smouse P.E., Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,* **131**, 2: 479-491.

Falush D., Stephens M., Pritchard J.K. (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes,* **7**, 4: 574-578.

Fatouros N.E., Dicke M., Mumm R., Meiners T., Hilker M. (2008) Foraging behavior of egg parasitoids exploiting chemical information. *Behavioral Ecology,* **19**, 3: 677-689.

Foll M., Beaumont M.A., Gaggiotti O. (2008) An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure. *Genetics,* **179**, 2: 927-939.

Forbes A.A., Powell T.H.Q., Stelinski L.L., Smith J.J., Feder J.L. (2009) Sequential Sympatric Speciation Across Trophic Levels. *Science,* **323**, 5915: 776-779.

Frantz A., Plantegenest M., Mieuzet L., Simon J. (2006) Ecological specialization correlates with genotypic differentiation in sympatric host-populations of the pea aphid. *Journal of Evolutionary Biology,* **19**, 2: 392-401.

Funk D.J., Filchak K.E., Feder J.L. (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica,* **116**, 2: 251-267.

Futuyma D.J., Moreno G. (1988) The Evolution of Ecological Specialization. *Annual Review of Ecology and Systematics,* **19**, 1: 207-233.

Graham M.W.R. de V. (1991) A reclassification of the European Tetrastichinae (Hymenoptera: Eulophidae): revision of the remaining genera. *Memoirs of the American Entomological Institute (USA),* **49**: 113-115.

Grapputo A., Kumpulainen T., Mappes J. (2005) Phylogeny and evolution of parthenogenesis in Finnish bagworm moth species (Lepidoptera: Psychidae: Naryciinae) based on mtDNA-markers. *Annales Zoologici Fennici,* **42**, 2: 141-160.

Hagimori T., Abe Y., Date S., Miura K. (2006) The First Finding of a *Rickettsia* Bacterium Associated with Parthenogenesis Induction Among Insects. *Current microbiology,* **52**, 2: 97-101.

Hilker M., Kobs C., Varama M., Schrank K. (2002) Insect egg deposition induces *Pinus sylvestris* to attract egg parasitoids. *Journal of Experimental Biology,* **205**: 4, 455.

Hodar J. H., Zamora R., Castro J., Baraza E. (2004) Feast and famine: previous defoliation limiting survival of pine processionary caterpillar *Thaumetopoea pityocampa* in Scots pine *Pinus sylvestris*. *Acta Oecologica*, **26**: 203-210.

Jaenike J. (1990) Host Specialization in Phytophagous Insects. *Annual Review of Ecology and Systematics,* **21**, 1: 243-273.

Jaenike J. (1981) Criteria for Ascertaining the Existence of Host Races. *The American Naturalist,* **117**, 5: 830-834.

Kerdelhue C., Magnoux E., Lieutier F., Roques A., Rousselet J. (2006) Comparative population genetic study of two oligophagous insects associated with the same hosts. *Heredity,* **97**, 1: 38-45.

Kondo N., Shimada M., Fukatsu T. (2005) Infection density of *Wolbachia* endosymbiont affected by co-infection and host genotype. *Biology Letters,* **1**, 4: 488-491.

Lozier J.D., Roderick G.K., Mills N.J., Peterson M. (2007) Genetic evidence from mitochondrial, nuclear, and endosymbiont markers for the evolution of host plant associated species in the aphid genus *Hyalopterus* (Hemiptera: Aphididae). *Evolution,* **61**, 6: 1353-1367.

Lynch M. (1984) The Genetic Structure of a Cyclical Parthenogen. *Evolution,* **38**: 1, 186-203.

Malausa T., Bethenod M., Bontemps A., Bourguet D., Cornuet J., Ponsard S. (2005) Assortative Mating in Sympatric Host Races of the European Corn Borer. *Science,* **308**, 5719: 258-260.

Masutti L., Battisti A. (1990) *Thaumetopoea pityocampa* (Den. & Schiff.) in Italy Bionomics and perspectives of integrated control. *Journal of Applied Entomology,* **110**, 1-5: 229-234.

Medina R.F., Barbosa P., Christman M., Battisti A (2006) Number of individuals and molecular markers to use in genetic differentiation studies. *Molecular Ecology Notes,* **6**, 4: 1010-1013.

Medina, R.F. and Barbosa, P. (2008) The role of host plant species in the phenotypic differentiation of sympatric populations of *Aleiodes nolophanae* and *Cotesia marginiventris*. *Entomologia Experimentalis et Applicata,* **128:** 14-26

Miller N.J., Kift N.B., Tatchell G.M. (2005) Host-associated populations in the lettuce root aphid, *Pemphigus bursarius* (L.). *Heredity,* **94**, 5: 556-564.

Mopper S., Strauss S.Y. (1998) *Genetic structure and local adaptation in natural insect populations : effects of ecology, life history, and behavior* Chapman & Hall, New York.

Mouton L., Henri H., Charif D., Boulétreau M., Vavre F. (2007) Interaction between host genotype and environmental conditions affects bacterial density in *Wolbachia* symbiosis. *Biology Letters,* **3**, 2: 210-213.

Normark B.B., Judson O.P., Moran N.A. (2003) Genomic signatures of ancient asexual lineages. *Botanical Journal of the Linnean Society,* **79**, 1: 69-84.

Normark B.B., Moran N.A. (2000) Testing for the accumulation of deleterious mutations in asexual eukaryote genomes using molecular sequences. *Journal of Natural History,* **34**, 9: 1719-1729.

O'Neill S.L., Giordano R., Colbert A.M., Karr T.L., Robertson H.M. (1992) 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proceedings of the National Academy of Sciences of the United States of America,* **89**, 7: 2699-2702.

Papura D., Simon J., Halkett F., Delmotte F., Le Gallic J., Dedryver C. (2003) Predominance of sexual reproduction in Romanian populations of the aphid *Sitobion avenae* inferred from phenotypic and genetic structure. *Heredity,* **90**, 5: 397-404.

Peccoud J. (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the United States of America,* **106**, 18: 7495-7500.

Price P.W., Bouton C.E., Gross P., McPheron B.A., Thompson J.N., Weis A.E. (1980) Interactions among three trophic levels: influence of plants on interactions between insect herbivores and natural enemies. *Annual Review of Ecology and Systematics,* **11**, 1: 41-65.

Pritchard J.K., Stephens M., Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155**, 2: 945-959.

Richardson D.M. (1998) *Ecology and biogeography of Pinus.* Cambridge University Press, Cambridge; New York, NY, USA.

Schluter D. (2001) Ecology and the origin of species. *Trends in Ecology & Evolution,* **16**, 7, 372-380.

Simon J.C., Baumann S., Sunnucks P., Hebert P.D., Pierre J.S., Le Gallic J.F., Dedryver C.A. (1999) Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology,* **8**, 4: 531-545.

Simon J.C., Delmotte F., Rispe C., Crease T. (2003) Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. *Journal of the Linnean Society,* **79**: 151-163.

Simon J.C., Leterme N., Latorre A. (1999) Molecular markers linked to breeding system differences in segregating and natural populations of the cereal aphid *Rhopalosiphum padi* L. *Molecular Ecology,* **8**, 6: 965-973.

Stastny M., Battisti A., Petrucco Toffolo E., Schlyter F., Larsson S. (2006) Host-plant use in the range expansion of the pine processionary moth, *Thaumetopoea pityocampa*. *Ecological Entomology,* **31**, 481-490.

Stireman J.O. (2005) Host-associated genetic differentiation in phytophagous insects: General phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution,* **59**, 12: 2573-2587.

Stireman J.O., Nason J.D., Heard S.B., Seehawer J.M. (2006) Cascading host-associated genetic differentiation in parasitoids of phytophagous insects. *Proceedings of The Royal Society B: Biological,* **273**, 1586: 523-530.

Stouthamer R., Kazmer, D.J. (1994) Cytogenetics of microbe-associated parthenogenesis and its consequences for gene flow in *Trichogramma* wasps. *Heredity* **73**: 317-327.

Stouthamer R. (1997) *Wolbachia* induced parthenogenesis. In: *Influencial passengers, inherited microorganisms and arthropod reproduction.* (eds. O'Neill SL, Hoffman AA, Werren JH) pp. 102-124, Oxford University Press, Oxford.

Sunnucks P., De Barro P.J., Lushai G., Maclean N., Hales D. (1997) Genetic structure of an aphid studied using microsatellites: cyclic parthenogenesis, differentiated lineages and host specialization. *Molecular Ecology,* **6**, 11: 1059-1073.

van Wilgenburg E., Driessen G., Beukeboom L.W. (2006) Single locus complementary sex determination in Hymenoptera: an "unintelligent" design? *Frontiers in Zoology,* **3**: 1.

Vekemans X., Beauwens T., Lemaire M., Roldán-Ruiz I. (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology,* **11**, 1: 139-151.

Via S. (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution,* **53**, 5: 1446-1457.

Vialatte A., Dedryver C., Simon J., Galman M., Plantegenest M. (2005) Limited genetic exchanges between populations of an insect pest living on uncultivated and related cultivated host plants. *Proceedings of the Royal Society B: Biological Sciences,* **272**, 1567: 1075-1082.

Vinson S.B. (1999) Parasitoid manipulation as a plant defense strategy. *Annals of the Entomological Society of America,* **92**, 6: 812-828.

Vinson S.B. (1998) The general host selection behavior of parasitoid Hymenoptera and a comparison of initial strategies utilized by larvaphagous and oophagous species. *Biological Control,* **11**, 2: 79-96.

Vorburger C., Lancaster M., Sunnucks P. (2003) Environmentally related patterns of reproductive modes in the aphid *Myzus persicae* and the predominance of two 'superclones' in Victoria, Australia. *Molecular Ecology,* **12**, 12: 3493-3504.

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

Weeks A.R., Tracy Reynolds K., Hoffmann A.A. (2002) *Wolbachia* dynamics and host effects: what has (and has not) been demonstrated? *Trends in Ecology & Evolution,* **17**, 6: 257-262.

Zchori-Fein E., Perlman S.J. (2004) Distribution of the bacterial symbiont *Cardinium* in arthropods. *Molecular Ecology,* **13**, 7: 2009-2016.

Zhou W., Rousset F., O'Neil S. (1998) Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. *Proceedings The Royal Society B: Biological sciences,* **265**, 1395: 509-515.

Zovi D., Stastny M., Battisti A., Larsson S. (2008) Ecological costs on local adaptation of an insect herbivore imposed by host plants and enemies. *Ecology,* **89**, 5: 1388-1398.

Conclusion

Although the genetic study of the population system of the processionary moth will require further development, including the genomic analysis of genes with specific adaptive functions, there are two general conclusions that can be taken with the results obtained so far. The first concerns a practical application of the knowledge while the second is oriented to clarify the taxonomic status of the group and to understand its adaptation potential.

Ecological aspects and applied results linked to population genetic analysis.

Winter pine processionary moths are currently experiencing an expansion due to both climate warming and host plant plantations in new areas. Genetic analysis through mitochondrial markers of populations in the whole range of *T. pityocampa* and *T. wilkinsoni* permitted to trace introduction/expansion patterns in the Mediterranean area. Whereas only few cases seem to be ascribed to human transportations (e.g. Sardinia and Baleari islands), in all the other situations (Israel, northern Turkey, Algeria) the colonizers come from geographically closer populations. In temperate regions of Europe instead (e.g. northern France) it was not possible to disentangle historical events from contemporary traces, due to the lack of genetic structure of the populations in the area explained by a low variability at the mitochondrial level.

The comparison of mitochondrial markers with nuclear markers confirmed also in *T. wilkinsoni* the gender related dispersal found in *T. pityocampa* (Salvato *et al.* 2002), with males showing a higher gene flow due to their higher vagility. In the expanding populations, the risk of a reduced genetic diversity due to the low females dispersal ability is traded off by male allelic contribution. However, this does not seem to fully counteract the drift effects, as it has been proved by some cases of local adaptation (Zovi 2008).

The genetic population study of two egg parasitoids of pine processionary moth gave an interesting insight into the different strategies adopted by generalist and specialist parasitoids to find their host. The higher mobility of the specialist (*B. servadeii*), inferred by the lack of a spatial genetic pattern, in association with a host plant linkage of some genotypes, could be the result of a long-range host research strategy. On the other hand, the geographical structure found in the generalist (*O. pityocampae*) populations could be the effect of a lower vagility of this parasitoid which is not so strictly dependent on its main host. This different skill to find their host could thus affect the ability of these parasitoids to follow pine processionary moth during its expansion.

The take-home message of this part is that measures to be taken to contain the pest in the forest and urban habitats may have different outcomes according to the genetic structure of the populations. In the expansion areas contiguous to the native range we should expect a similar response behavior, indicating that any treatment will be likely have the same effect. In areas where the species has been introduced by man, the limited genetic variability could affect the performance of the pest and indirectly the countermeasures. This is typically the case of the Sardinia island, where the population settled in 2004 from a Tuscany genotype has spread around rather quickly, although with a very low performance and damage to trees and people. In the new expansion areas, either natural or anthropogenic, a different impact of natural enemies can be expected. The structured populations of the polyphagous *Ooencyrtus* may respond in a different way to the sudden availability of the new host than the monophagous, non-structured *Baryscapus*. A bright example of such a situation comes from the natural expansion area of the pine processionary moth in the Venosta valley of Central Alps. Here the outbreak started in 1999 and the local *Ooencyrtus* was present, although unable to cause significant mortality to the new host. Until now its population stays at very low density, likely because maladapted to track the population dynamics of the host. Conversely, the monophagous *Baryscapus* was detected only five years after the outbreak but achieved immediately a great success, with parasitism rate as high as 50% (Zovi *et al.* 2006). Evidently, there were no limitations in the host exploitation because the parasitoids probably arrived from a nearby native area where they were already

adapted to it. Thus the generally most common and efficient egg parasitoid *Ooencyrtus* has strong limitations in regulating host populations in the expansion areas, where the insect seems to be present but probably adapted to other hosts. All these issues add on the unpredictability of the regulation mechanisms and of the efficiency of control measures in the expansion areas of this forest pest.

Main phylogeographic events detected by mitochondrial markers with some insights into taxonomy.

The study on the genetic diversity of pine processionary moths has shown how this species complex responded to Quaternary glaciation cycles. Through the analysis of genetic variability it has been possible to identify glacial refugial areas (Italy, Alps, Balkans), and areas recently colonized (most part of Europe). Populations in Iberian peninsula, northern Africa and Near East, does not seem to have experienced glaciations effects. The most striking result of the phylogeographic study on pine processionary moths is the identification of one more clade (ENA clade, northern Africa) besides the two we expected to find (*T. pityocampa* and *T. wilkinsoni*). Although determining the taxonomic status of the clusters identified here is beyond the scope of the present study, it is noteworthy that the range of this clade overlaps the one described by Agenjo (1941) for the variety *T. pityocampa orana* (Tab.1.1). This is one of the two subspecies considered valid by Kiriakoff (1970) among the many described by Agenjo (1941) and other authors, based on slightly divergent morphological traits. The second subspecies, *ceballosi* from Anatolia, is likely is a synonym of *wilkinsoni* as the locus typicus is a locality (Aydin) of western Turkey, from where I obtained individuals that clearly belong to the latter species. The question of the status of *T. pityocampa orana* is still open to a more detailed genetic and morphological analysis of individuals coming from the contact zone with in central and western Algeria. Another interesting result is given by the molecular dating, which show that the three main clades diverged between early and late Messinian. Intriguingly the Crete island population divergence dates at the end of this age, characterized by the conclusion of the salinity crisis of the Mediterranean sea. If this dating pattern will be confirmed by the

ongoing work on the phylogeny of the genus in the West Palaearctic, it will result that the origin of the group is rather ancient, mirroring at some extent the history of vegetation in the area. The complex phenology of the pine processionary moths, with species and forms shifted chronologically among all the seasons, is a spectacular example of how an herbivore may track the host plant in the midst of changing habitat and climate. The quick response of these species to current climate change is a further evidence of the high adaptation potential to new conditions, making them at the same time a threat to forest productivity and to human health.

References

Agenjo R. (1941) Monographia de la familia Thaumetopoeidae. *Eos Madrid,* **17**: 69-130.

Kiriakoff S. (1970) Lepidoptera familia Thaumetopoeidae. *Genera Insectorum de P.Wystman Belgique,* **219**.

Salvato P., Battisti A., Concato S., Masutti L., Patarnello T., Zane L. (2002) Genetic differentiation in the winter pine processionary moth (*Thaumetopoea pityocampa-wilkinsoni* complex), inferred by AFLP and mitochondrial DNA markers. *Molecular Ecology,* **11**: 2435-2444.

Zovi D., Battisti A., Hellrigl K., Minerbi S. (2006) Egg parasitoids of the pine processionary moth and their occurrence in Venosta/Vinschgau. *forest observer,* **2-3**: 81-88.

Zovi D., Stastny M., Battisti A., Larsson S. (2008) Ecological costs on local adaptation of an insect herbivore imposed by host plants and enemies. *Ecology,* **89**, 5: 1388-1398.

Acknowledgments

Firstly I would like to thank Andrea Battisti, my supervisor, Lorenzo Zane, molecular ecologist at the University of Padova, and Enrico Negrisolo, molecular biologist at the University of Padova, who assisted and supported me during the three years of my experience. A special thank to Raul F. Medina who hosted and supported me during my stay at Texas A&M University, College Station, TX, between June and December 2008 and all his staff: Apurba Barman, Aaron Dickey, Kyle Harrison, Emilie Hartfield, Andrea Joyce, Jorge San Juan. I thank all the people who helped in the experimental phases at Padova University: in particular Paola Salvato, who assisted me in the lab work, and Edoardo Petrucco Toffolo and Daniel Zovi for their contributions to the chapter of egg parasitoids.

The study was supported by the University of Padova for the three years fellowship, the EU project PROMOTH, the PRIN project 2007 Massimo Faccoli "Insetti forestali e cambiamento climatico", and the grant from the Agence Nationale de la Recherche, France (ANR Urticlim).

I thank particularly all my friends and colleagues in the department, in particular Teshale Assefa, Fernanda Colombari, Massimo Faccoli, Gabriella Frigimelica, Lorenzo Marini, Luca Mazzon, Isabel Martinez-Sanudo, Claudia Savio, Paola Tirello, Edoardo Petrucco Toffolo, Caterina Villari and Daniel Zovi. Thanks also to the entomology staff at the department (Professors Carlo Duso, Vincenzo Girolami, Luigi Masutti, Giuseppina Pelizzari and technicians Patrizia dall'Ara, Gabriella Fenici, Paolo Fontana, Paolo Paolucci).

And finally I have to be grateful also to my family and friends who gave me encouragement and moral support throughout this period.