Journal of Biogeography (J. Biogeogr.) (2010) 37, 1478–1490

ORIGINAL ARTICLE

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] and 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 analysed 412 individuals from 61 locations and performed maximum likelihood and maximum parsimony phylogenetic analyses and hierarchical analysis of molecular variance, and we 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, yet 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.

Keywords

Glacial refugia, latitudinal shift, Mediterranean Basin, mitochondrial DNA, mountainous areas, *Pinus*, range expansion, *Thaumetopoea pityocampa*, vertical migration, western Europe.

*Correspondence: Jérôme Rousselet, INRA, Centre d'Orléans, Unité de Recherche de Zoologie Forestière, 2163 Avenue de la Pomme de Pin, CS 40001 Ardon, F-45075 Orléans Cedex 2, France.
E-mail: rousselet@orleans.inra.fr

www.blackwellpublishing.com/jbi
doi:10.1111/j.1365-2699.2010.02289.x

Journal of Biogeography

INTRODUCTION

Quaternary climatic oscillations have produced great changes in species ranges that have strongly influenced the present-day geographic distribution of genetic diversity (e.g. Hewitt, 1999, 2004; Schmitt, 2007). Ranges of most species shifted latitudinally and/or elevationally as a response to glacial/interglacial cycles, resulting in expansion–contraction phases (Hewitt, 2004; Habel *et al.*, 2005; Schmitt, 2007; Varga & Schmitt, 2008). In general, temperate species have expanded during warm periods and responded to cold phases by local extinctions in northern regions and by survival in southern glacial refugia (Hewitt, 2004). This has commonly resulted in a 'southern richness and northern purity' pattern, in which genetic diversity and divergence are higher at lower latitudes (Hewitt, 1999). Cold-tolerant arctic species exhibit opposite responses, as warm interglacials have caused fragmentation of habitat and range contraction into northernmost locations. Similarly, alpine species have tracked a suitable environment by upslope movements during the warmest periods, and survived the interglacials in limited refugia or 'sky islands' (DeChaine & Martin, 2005; Varga & Schmitt, 2008). More recently, accumulation of phylogeographical data has supported evidence of more complex patterns of response to Quaternary climatic oscillations, both because many species actually have intermediate ecological requirements (Varga & Schmitt, 2008) or habitat-generalist traits (Bhagwat & Willis, 2008) and because the palaeoenvironments were more complex than previously thought (Stewart & Lister, 2001; Hewitt, 2004; Willis & van Andel, 2004; Provan & Bennett, 2008; Médail & Diadema, 2009).

The winter pine processionary moth, *Thaumetopoea pityocampa* (Denis & Schiffermüller, 1776) (Lepidoptera: Notodontidae), is a phytophagous insect distributed from North Africa to the Balkans. It belongs to a species complex with a wide distribution around the Mediterranean Basin (Simonato *et al.*, 2007; Kerdelhué *et al.*, 2009). The moth's geographic range is constrained by sunshine requirements in winter and susceptibility to both cold winter and high summer temperatures (Huchon & Démolin, 1970; Battisti *et al.*, 2005; see Materials and Methods). *Thaumetopoea pityocampa* is more restricted geographically than the distribution area of its potential hosts, which include lowland Mediterranean as well as montane or boreal *Pinus* species. In southern Europe and North Africa, *T. pityocampa* occurs from thermo-mediterranean environments (with hot summers and mild winters) to oro-mediterranean environments (with milder summers and colder winters). However, the supra-mediterranean zone (with mild summers and relatively mild winters) could correspond to the optimal ecological niche of this species (Huchon & Démolin, 1970). *Thaumetopoea pityocampa* does not occur in areas under strong continental climates (with both hot summers and cold winters; Huchon & Démolin, 1970). Under Atlantic climates, this species can be found as far north as the 48th parallel (see Fig. 1).

In recent years, the range expansion of *T. pityocampa* to upper latitudes or elevations has been reported in several European countries (Rosenzweig *et al.*, 2007). This distributional change is primarily due to increased winter temperatures and is a consequence of climate warming (Battisti *et al.*, 2005). This rapid response to climatic changes suggests that the past distribution of this species is likely to have been strongly affected by Pleistocene climate changes during both glacial and interglacial episodes. Due to an obligate relationship with its pine hosts (*Pinus* spp.), *T. pityocampa* can have survived only in places where pines persisted. The locations of its refugial areas were thus constrained by those of its hosts, which exhibit different climatic requirements.

A preliminary genetic study in France using microsatellite markers showed that within-population genetic diversity was highest in the eastern Pyrenees (Kerdelhué *et al.*, 2006). This study also suggested that, in spite of its moderate elevation, the Massif Central was an effective barrier to gene flow. Moreover, using mitochondrial DNA and nuclear internal transcribed spacer 1 (ITS1) sequences, Santos *et al.* (2007) showed strong differentiation between Iberian and French populations, although with a limited sample size. Two hypotheses can be proposed to explain both the high genetic diversity observed within the Pyrenees and the strong genetic differentiation across this mountain range. In the first it is hypothesized that for such a cold-susceptible species with putatively limited dispersal abilities, the Pyrenean range could have acted as a barrier to post-glacial expansion routes from separated refugia. In this case, secondary contact zones should be found in favourable valleys and/or on western and eastern ends of this mountain range, where the elevation is lower. The high genetic diversity observed in the Pyrenees would then derive from admixture between two strongly differentiated lineages. Such a pattern has already been observed for various European species (Hewitt, 1999, 2004; Habel *et al.*, 2005; Schmitt, 2007). The second hypothesis is that the Pyrenees might have acted as a refugium rather than a barrier. The processionary moth could have survived locally by gradual elevational shifts. In this case, high genetic diversity would mirror ancestral polymorphism rather than being a sign of admixture. A similar scenario has been described for stenotopic montane species that were able to descend or ascend as the climate cooled or warmed, thus surviving glacial oscillations in the same region without major latitudinal shifts (Hewitt, 2004; Varga & Schmitt, 2008).

To test these hypotheses, we sampled *T. pityocampa* throughout western Europe, focusing on mountain ranges. We analysed the distribution of the genetic diversity based on mitochondrial cytochrome *c* oxidase subunit I (COI) partial sequences. Our objectives were: (1) to describe the phylogeographic population structure of *T. pityocampa* over western Europe and particularly to confirm the existence of two deeply divergent clades on both sides of the Pyrenees, and (2) to test if mountain ranges, especially the Pyrenees, Massif Central and Alps, have been effective barriers to gene flow during the Quaternary, and whether they played a strong role in structuring populations.

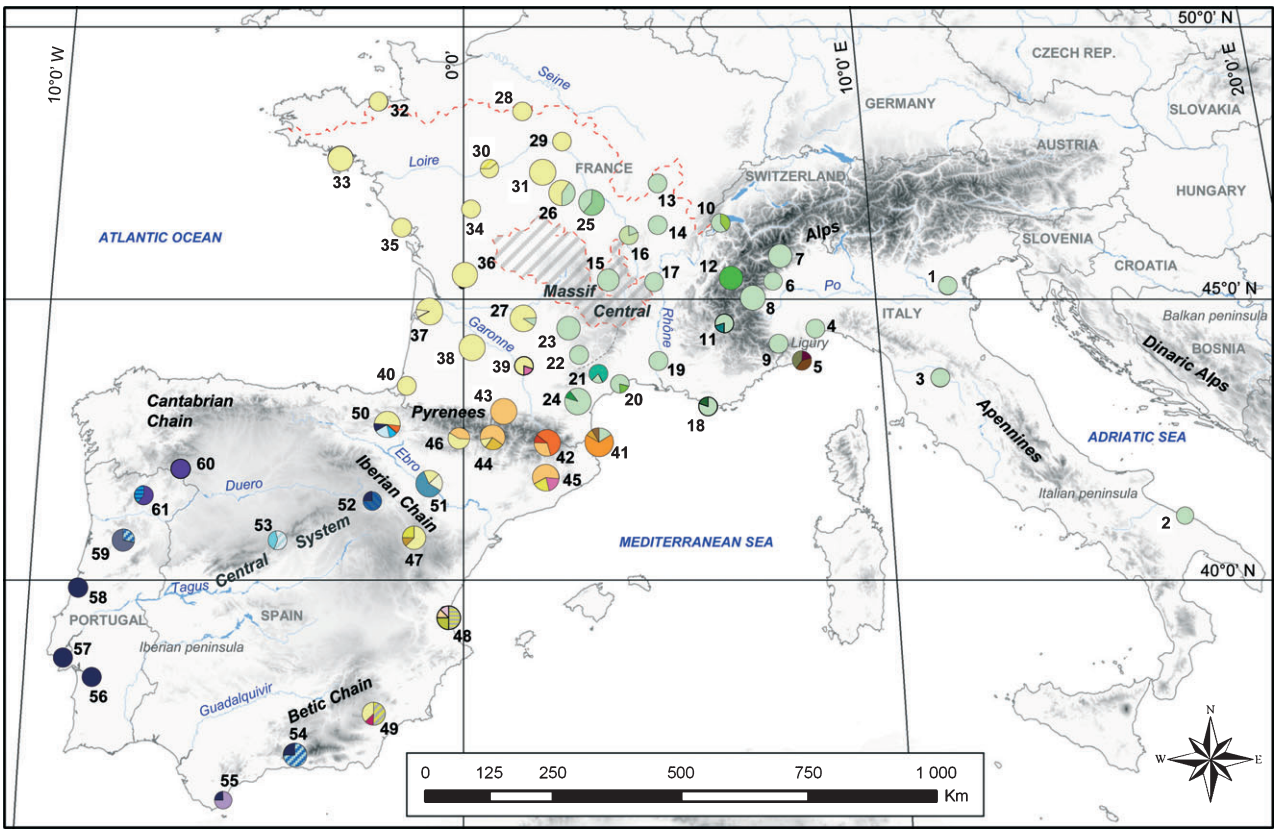

Figure 1 Geographic distribution of the 46 cytochrome *c* oxidase subunit I haplotypes of *Thaumetopoea pityocampa* among the 61 sites sampled in western Europe. The total area of each circle is proportional to the sample size and haplotype frequencies are represented by the area of the circle occupied. Colour codes refer to the colour used in the haplotype network (see Fig. 2). The black numbers correspond to the sampling sites (see Table 1). The red dotted line indicates the present-day northern limit of *T. pityocampa* in France and the hatched area indicates the uncolonized part of the Massif Central. The northern limit in Italy and the Balkans (not represented) corresponds to the southern side of the Alps and Dinaric Alps, respectively. The map was generated using ArcGIS software and a Mollweide projection.

MATERIALS AND METHODS

Study species – host and climate requirements

The pine processionary moth is a univoltine and semelparous species with very short-lived adults exhibiting sex-biased dispersal, as females may disperse a few kilometres while males may fly several tens of kilometres. The defoliating and urticating larvae develop in winter, feeding on various native pine and cedar species (*Pinus nigra* Arnold, *Pinus sylvestris* L., *Pinus uncinata* Ramond ex A. DC, *Pinus pinaster* Aiton, *P. pinea* L., *Pinus halepensis* Miller, *Cedrus atlantica* (Endl.) Manetti ex Carrière). The native ranges of these hosts are strongly spatially structured (Barbéro *et al.*, 1998; Kerdelhué *et al.*, 2009). This insect can also attack some exotic conifers (e.g. *Pinus radiata* D. Don, *Cedrus deodara* (Roxb.) G. Don, *Pseudotsuga menziesii* (Mirb.) Franco). The gregarious larvae spin a silk nest. Pupation takes place in the soil after the typical head-to-tail processions at the end of winter or early spring, and the subterranean survival rate depends on soil moisture (Huchon & Démolin, 1970). Adult emergence and subsequent oviposition take place in summer or autumn depending on latitude and elevation.

The life cycle of the pine processionary moth varies greatly according to climate and is controlled by two major temperature constraints, which also determine distribution area and population dynamics (Huchon & Démolin, 1970; Battisti *et al.*, 2005). The northward and upward limits of the species' range are determined by lower lethal temperatures in winter (−12 °C; Huchon & Démolin, 1970), by a minimal number of sunshine hours (isohele of 1800 h of annual sunshine; Huchon & Démolin, 1970) and by specific temperature requirements necessary for feeding (see Battisti *et al.*, 2005; Robinet *et al.*, 2007). The population dynamics of the species at the southern edge of its distribution are constrained by summer temperatures, as eggs and early instar larvae are susceptible to high summer temperatures (monthly mean of daily maximum temperatures above 25 °C, and maximum temperatures above 32 °C; Huchon & Démolin, 1970). Consequently, the highest population densities in France are usually located in sub-mediterranean mountains and in some areas under mild oceanic climate. Some plasticity in the timing of sexual reproduction allows the species to adapt to various environments, as the adults emerge later in the warmest regions and earlier in places where winters are coldest (Huchon & Démolin, 1970).

Sampling

Sixty-one locations were sampled from 1999 to 2008, and a total of 412 caterpillars were analysed. The number of individuals per site ranged from 4 to 12. They were collected on different native and non-native host tree species (six *Pinus* species and *Pseudotsuga menziesii*). The sampling sites, host tree and year of collection are summarized in Appendix S1 in Supporting Information, and sampling locations are shown in Fig. 1. The study area covers only the western European part of the distribution range, as populations from North Africa are known to form a distinct lineage (Kerdelhué *et al.*, 2009) and were not included in the present study. The study area includes both the recent expansion areas in northern France and the two southern peninsulas of western Europe (Iberia and Italy). The sampling effort was intentionally highest from north-eastern Spain to north-western Italy to test the hypothesized differentiation of Iberian populations compared with French ones (Santos *et al.*, 2007), to determine the role of the northerly mountainous ranges during post-glacial recolonizations and to locate possible contact zones. The main slopes of the European mountain ranges (French and Italian Alps, western and eastern Massif Central, northern and southern Pyrenees) were sampled. In order to avoid sampling related individuals, only one nest per tree was collected and only one larva per nest was sequenced. Larvae were immediately stored in absolute ethanol and then kept at −20 °C until DNA extraction.

DNA extraction and amplification

Genomic DNA extraction, polymerase chain reaction (PCR) amplifications and sequencing of part of the mitochondrial COI gene followed the protocol described in Santos *et al.* (2007). The primers used were C1-J-2183 (Jerry, 5′-CAAC ATTTATTTTGATTTTTTGG-3′) and TL2-N-3014 (Pat, 5′-TC CAATGCACTAATCTGCCATATTA-3′), respectively, located in the gene itself and in its flanking region (tRNA-leucine gene).

Data analysis

Sequences were aligned in BIOEDIT 7.05 (Hall, 1999). Haplotypes and their frequencies were calculated with DNASP 4.5 (Rozas *et al.*, 2003). Pairwise genetic distances between haplotypes were calculated using PAUP* 4.0 (Swofford, 2003).

To estimate gene genealogies a statistical parsimony network was constructed using TCS 1.21 (Clement *et al.*, 2000), allowing a connection between haplotypes of up to 12 steps, to fit the maximal divergence observed in our data set. Maximum likelihood and maximum parsimony inferences were also used to investigate the phylogenetic relationships among the mtDNA haplotypes. Maximum likelihood analyses were based on the best-fit model of sequence evolution estimated using Akaike information criterion (AIC) tests implemented in MODELTEST 3.7 (Posada & Crandall, 1998). For both methods, node support was estimated from 200 bootstrap replicates conducted heuristically using tree bisection–reconnection branch swapping on starting trees generated by five randomly derived stepwise addition sequences. The resulting trees were rooted with a sequence from the sibling species *Thaumetopoea wilkinsoni* Tams (GenBank accession number GU385952). Before following the bootstrapping procedure, maximum likelihood heuristic searches were also conducted with and without the molecular clock enforced. The molecular clock hypothesis was then tested with a likelihood ratio test (LRT; Felsenstein, 1988), computed in PAUP* 4.0, with a homogeneous rate of evolution as the null hypothesis.

The level of genetic polymorphism within sites was assessed by calculating haplotype and nucleotide diversity indices. Gene diversity (h) and within-population mean number of pairwise differences per sequence (k) were computed using ARLEQUIN 3.1 (Excoffier *et al.*, 2005). Correlations between population parameters (h and k) and latitude were assessed with linear regressions.

The occurrence of a significant phylogeographic structure was inferred by testing whether G_{ST} (the coefficient of genetic variation over all populations that only considers haplotype identity) was significantly smaller than N_{ST} (the equivalent coefficient taking into account haplotype divergence) by use of 1000 permutations implemented in PERMUT (Pons & Petit, 1996). Population genetic structure was examined by analysis of molecular variance (AMOVA) based on pairwise F_{ST} and computed using ARLEQUIN. This method was used to partition genetic variance within populations, among populations within groups, and among groups. The populations were grouped either by geographical location or by host species. Significance was determined by 5000 permutations. Geographical groups were defined on the basis of the distribution area of the lineages identified with phylogenetic and parsimony network analyses. Samples corresponding to putative secondary contact zones between these lineages (i.e. sampling sites containing haplotypes from different phylogenetic lineages) were treated using two options: (1) they were entirely attributed to one of the geographical groups (grouping by regions I); and (2) they were removed from the data set (grouping by regions II). Concerning grouping by hosts, sites where the insect was sampled from more than one *Pinus* species (see Table 1) were split so that each individual was attributed to its actual host group.

Two methods were used to infer the demographic history: mismatch distribution analyses (Rogers & Harpending, 1992) and neutrality tests. For the first approach, the distribution of pairwise nucleotide site differences between haplotypes was calculated and the observed values were compared with the expected values under a sudden expansion model. Demographic expansion parameters (θ_0, θ_1 and τ) were estimated with ARLEQUIN 3.1, and a test of goodness-of-fit based on the sum of square deviations between the observed and expected distributions was performed using 1000 bootstrap replicates. The parameters estimated with ARLEQUIN were used in DNASP

Table 1 Mitochondrial cytochrome *c* oxidase subunit I haplotypes (HT) found in each sample of *Thaumetopoea pityocampa* collected in western Europe and population parameters.

	Site of collection	Haplotype frequencies (according to host species)	*n*	N_{HT}	*h*	*k*
1	Calbarina	*Pinus nigra*: 5 HT1	5	1	0.00	0.00
2	Bari	*Pinus halepensis*: 5 HT1	5	1	0.00	0.00
3	Mt San Michele	*Pinus nigra*: 4 HT1	4	1	0.00	0.00
4	Massimino	*Pinus sylvestris*: 5 HT1	5	1	0.00	0.00
5	Rollo	*Pinus halepensis*: 2 HT2, 1 HT3, 2 HT4	5	3	0.80	3.40
6	Germagnano	*Pinus nigra*: 5 HT1	8	1	0.00	0.00
7	Ruines Verrès	*Pinus sylvestris*: 8 HT1	9	1	0.00	0.00
8	Susa, Oulx	*Pinus sylvestris*: 9 HT1	5	1	0.00	0.00
9	Tende	*Pinus sylvestris*: 5 HT1	5	1	0.00	0.00
10	Excenevex	*Pinus sylvestris*: 3 HT1, 2 HT5	5	2	0.60	0.60
11	Prunières	*Pinus nigra*: 2 HT1; *P. sylvestris*: 2 HT1, 1 HT6	5	2	0.40	0.40
12	Montagny	*Pinus sylvestris*: 8 HT7	8	1	0.00	0.00
13	Beaune	*Pinus nigra*: 5 HT1	5	1	0.00	0.00
14	Leynes	*Pinus nigra*: 5 HT1	5	1	0.00	0.00
15	Chaniat	*Pinus sylvestris*: 7 HT1	7	1	0.00	0.00
16	Briennon	*Pinus nigra*: 1 HT1, 4 HT8	5	2	0.40	0.40
17	Bourg-Argental	*Pinus sylvestris*: 5 HT1	5	1	0.00	0.00
18	La Seyne-sur-Mer	*Pinus halepensis*: 4 HT1, 1 HT9	5	2	0.40	0.40
19	Tarascon	*Pinus halepensis*: 5 HT1	5	1	0.00	0.00
20	Frontignan	*Pinus halepensis*: 4 HT1, 1 HT10	5	2	0.40	0.40
21	Bédarieux	*Pinus nigra*: 1 HT11; *P. sylvestris*: 1 HT1, 3 HT11	5	2	0.40	0.40
22	Saint-Affrique	*Pinus nigra*: 5 HT1	5	1	0.00	0.00
23	Marcillac-Vallon	*Pinus nigra*: 8 HT1	8	1	0.00	0.00
24	Fabrezan	*Pinus pinaster*: 4 HT1, 1 HT12; *P. halepensis*: 5 HT1	10	2	0.20	0.20
25	Toury-sur-Jour	*Pinus nigra*: 4 HT1, 6 HT13	10	2	0.53	0.53
26	Lapan	*Pinus nigra*: 4 HT1, 6 HT14	10	2	0.53	0.53
27	Lavercantière	*Pinus nigra*: 5 HT14; *Pseudotsuga menziesii*: 1 HT1, 4 HT14	10	2	0.20	0.20
28	Mainvilliers	*Pinus nigra*: 5 HT14	5	1	0.00	0.00
29	Lorris	*Pinus sylvestris*: 5 HT14	5	1	0.00	0.00
30	Fondettes	*Pinus nigra*: 3 HT14, 2 HT15	5	2	0.60	0.60
31	Vierzon	*Pinus nigra*: 10 HT14	10	1	0.00	0.00
32	Ploubalay	*Pinus nigra*: 5 HT14	5	1	0.00	0.00
33	Plouharnel	*Pinus nigra*: 9 HT14	9	1	0.00	0.00
34	Vouillé	*Pinus nigra*: 5 HT14	5	1	0.00	0.00
35	Les Portes-en-Ré	*Pinus nigra*: 5 HT14	5	1	0.00	0.00
36	Rioux-Martin	*Pinus nigra*: 5 HT14; *P. pinaster*: 4 HT14	9	1	0.00	0.00
37	Cestas	*Pinus pinaster*: 9 HT14, 1 HT16	10	2	0.20	0.20
38	Réaup-Lisse	*Pinus pinaster*: 10 HT14	10	1	0.00	0.00
39	Saint-Jory	*Pinus nigra*: 4 HT14, 1 HT17	5	2	0.40	1.20
40	Hasparren	*Pinus pinaster*: 5 HT14	5	1	0.00	0.00
41	Cerbère	*Pinus pinaster*: 2 HT1, 1 HT19, 1 HT18; *P. halepensis*: 4 HT19; *P. pinea*: 3 HT19, 1 HT20	12	4	0.56	1.55
42	Osséja	*Pinus sylvestris*: 3 HT22, 6 HT21, 1 HT23	10	3	0.60	0.93
43	Gajan	*Pinus sylvestris*: 10 HT22	10	1	0.00	0.00
44	Vilaller	*Pinus sylvestris*: 1 HT14, 4 HT22, 2 HT24; *P. uncinata*: 1 HT22	8	3	0.61	1.36
45	Santa Maria d'Oló	*Pinus nigra*: 2 HT17, 6 HT22, 2 HT25	10	3	0.62	2.49
46	Boltaña	*Pinus sylvestris*: 4 HT14, 3 HT22	7	2	0.57	1.14
47	Argente	*Pinus nigra*: 5 HT14, 1 HT31, 2 HT 32	8	3	0.61	0.68
48	Xeraco	*Pinus halepensis*: 2 HT33, 4 HT34, 1 HT35, 1 HT36	8	4	0.75	1.46
49	Vélez Blanco	*Pinus nigra*: 3 HT14, 1 HT37, 4 HT38	8	3	0.68	1.07
50	Undiano	*Pinus nigra*: 5 HT14, 1 HT21, 2 HT26, 1 HT27, 1 HT28	10	5	0.76	8.98
51	Zuera	*Pinus nigra*: 2 HT14, 6 HT29, 2 HT30	10	3	0.62	8.36
52	Ariza	*Pinus nigra*: 1 HT28, 3 HT39	4	2	0.50	0.50
53	Collado Mediano	*Pinus nigra*: 3 HT40, 2 HT41	5	2	0.60	1.80
54	Otívar	*Pinus pinaster*: 2 HT28, 6 HT42	8	2	0.43	0.43
55	Gibraltar	*Pinus pinea*, *P. halepensis*: 1 HT28, 3 HT43	4	2	0.50	0.50
56	Alcacer	*Pinus pinaster*: 5 HT28	5	1	0.00	0.00

Table 1 Continued

	Site of collection	Haplotype frequencies (according to host species)	n	N_{HT}	h	k
57	Apostiça	*Pinus pinaster*: 5 HT28	5	1	0.00	0.00
58	Leiria	*Pinus pinaster*: 5 HT28	5	1	0.00	0.00
59	Viseu	*Pinus pinaster*: 2 HT42, 5 HT44	7	2	0.48	0.95
60	Varges	*Pinus pinaster*: 5 HT45	5	1	0.00	0.00
61	Sevivas	*Pinus pinaster*: 3 HT45, 2 HT46	5	2	0.60	1.20

n, sample size; N_{HT}, total number of haplotypes for each sampling location; h, gene diversity; k, mean number of pairwise differences per sequence.

to generate mismatch distributions. Unimodal distributions can be related to sudden demographic expansions while multimodal distributions are consistent with stability (Slatkin & Hudson, 1991). We performed Fu's F_S (Fu, 1997) and R_2 tests (Ramos-Onsins & Rozas, 2002) to examine the neutrality of genetic variation. F_S tends to be negative under an excess of recent mutations, and a significantly negative value can be taken as an evidence of population growth and/or selection. The R_2 measure is based on the difference between the number of singleton mutations and the average number of nucleotide differences among sequences within a population sample. The significance of both tests was assessed with 10,000 coalescent simulations implemented in DnaSP. These tests were conducted on the whole data set and within each haplogroup.

RESULTS

Haplotype distribution and gene genealogy

The final alignment contains 412 sequences of 802 bp, corresponding to the second half of the COI gene. Fifty polymorphic sites were detected and 46 haplotypes were identified (Appendix S2). Pairwise uncorrected p-distances among haplotypes ranged from 0.125 to 2.618 (Appendix S3). Observed haplotype frequencies for each sampled location are given in Table 1. The geographic distribution of the haplotypes is shown in Fig. 1. Haplotype sequences were deposited in GenBank and are available under accession numbers GU385906–GU385951.

The best-fit model of sequence evolution is the transitional model (variable base frequencies and variable transition frequencies; Posada, 2003) with invariant sites and equal substitution rates among sites (TIM+I). The proportion of invariable sites (I) is 80.10%, the base frequencies are $\pi_A = 0.3250$, $\pi_C = 0.1874$, $\pi_G = 0.1191$, $\pi_T = 0.3684$, and the substitution rate parameters are 95.9003 for A ↔ G and 33.9135 for T ↔ C transitions, 1 for A ↔ C and G ↔ T transversions, and 0 for A ↔ T and C ↔ G transversions. A LRT for COI of the TIM+I model with and without the molecular clock enforced does not reject overall rate homogeneity. Consequently, the molecular clock hypothesis was accepted.

Both the maximum likelihood and maximum parsimony phylogenetic trees (Appendix S4) show the existence of two major clades, respectively composed of the haplotypes 1–25, 30–38 (clade A) and the haplotypes 26–29, 39–46 (clade B). Clade A is distributed from eastern Spain to Italy, while clade B is found in Portugal and western Spain. These clades are very well supported by bootstrap values (Appendix S4).

The haplotype network shows the existence of four haplogroups (Fig. 2). Three of these (namely A1, A2 and A3) are subdivisions of the previously identified clade A, while the fourth corresponds to clade B. The two clades are separated by 12 mutational steps. Haplogroup A1 (haplotypes 1–13) is distributed from eastern France to Italy (Fig. 1 and Table 1). Haplogroup A2 (haplotypes 14–17, 25, 30–38) is found in eastern Spain and western France, more or less along the Greenwich Meridian. Haplogroup A3 (haplotypes 18–24) is

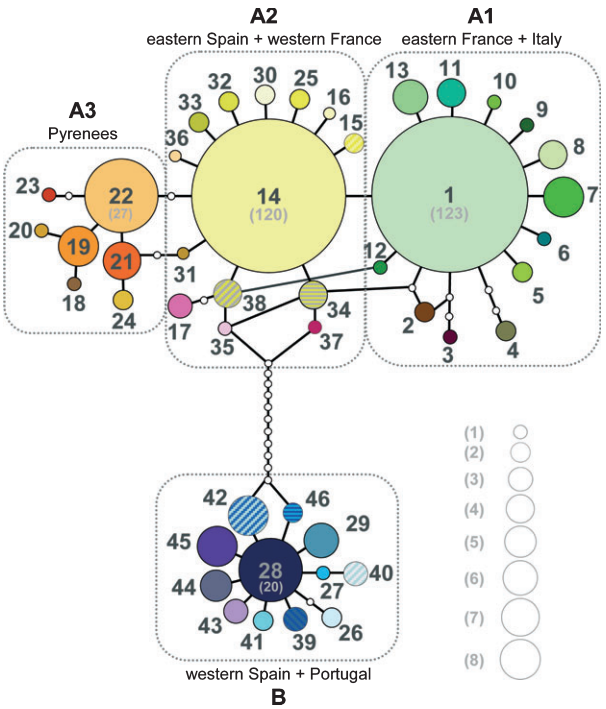

Figure 2 Haplotype network of the 46 cytochrome *c* oxidase subunit I haplotypes of *Thaumetopoea pityocampa* found in the study. Each circle represents a different haplotype (identified by a different colour and numbered from 1 to 46). Haplotype frequencies are represented by the area of the circle (see scale and grey number in brackets). Each line between circles corresponds to a mutational step and each small empty circle to a missing intermediate haplotype.

restricted to the Pyrenean range and corresponds to a supported subclade in maximum likelihood and parsimony phylogenetic analyses (Appendix S4). Each of the four haplogroups has a star-shaped topology with one central common haplotype surrounded by rarer but closely allied haplotypes (Fig. 2). The most common haplotype is haplotype 1 for A1 (78.85% of individuals), 14 for A2 (82.19%), 22 for A3 (57.45%) and 28 for B (31.75%). These four common and widely distributed haplotypes are found on several host plants (Table 1, Appendix S5).

Population parameters and genetic diversity

For each sampling location, gene diversity (h) and mean number of pairwise differences (k) are given in Table 1. Gene diversity ranges from 0 to 0.80 and k is between 0 and 8.98. In most sampling locations, we found haplotypes belonging to only one haplogroup (Table 1 and Fig. 1). Yet two populations contain haplotypes from groups A1 and A2 (sites 26 and 27), one from groups A1 and A3 (41), three from groups A2 and A3 (44–46), one from groups A2 and B (51), and one from groups A2, A3 and B (50). These two latter populations (50 and 51) exhibit the highest values of k. All these samples were also divided into subsamples, for which h and k were calculated separately (Appendix S6). Within the haplogroup A2, gene diversity (h) and mean number of pairwise differences (k) exhibit a significant negative relationship with latitude ($P < 0.01$ and $P < 0.001$, respectively). The relationship between h or k and latitude is not significant in any other haplogroup.

Phylogeographic pattern and population structure

Total gene diversity (H_T) is 0.818 (± 0.032), while the average within-population diversity (H_S) is 0.255 (± 0.036). The indices of population structure G_{ST} and N_{ST} are 0.689 (± 0.038) and 0.880 (± 0.036), respectively. The permutation test shows that N_{ST} is significantly greater than G_{ST} ($P < 0.001$) when considering the whole data set. Within clade A, G_{ST} and N_{ST} values are 0.679 (± 0.043) and 0.697 (± 0.043), respectively, and N_{ST} is not significantly greater than G_{ST}.

Four geographical regions were defined on the basis of the distribution of the four haplogroups for AMOVA: (1) Italy and eastern France, (2) western France and eastern Iberia, (3) the Pyrenees, and (4) central and western Iberia (Appendix S7). When individuals were grouped by geographical regions, the results always showed that a large and significant proportion of the variance was found among groups (Table 2). Similar results were found when considering only clade A (Table 2). Populations were then grouped by host species. Most of the genetic diversity was then found among populations within groups (Table 2). Nevertheless, a significant part of the variance was found among groups for the whole data set (21.36% of the total variance, $P < 0.001$), but not within clade A (4.58%, $P = 0.1085$).

Demographic history

The mismatch distribution curves are presented in Appendix S8. The parameters estimated under the sudden expansion

Table 2 Analyses of molecular variance (AMOVA) among populations of *Thaumetopoea pityocampa* in western Europe based on mitochondrial cytochrome *c* oxidase subunit I data. Results for groupings by geographical regions or by hosts are shown for the whole data set and for clade A only.

		Whole data set (Clade A + B)			Clade A		
Structure	Source of variation	Variance (%)	Fixation indices	*P*-value	Variance (%)	Fixation indices	*P*-value
Grouping by geographical regions I*	Among groups	78.52	$\Phi_{CT} = 0.78524$	< 0.001	53.00	$\Phi_{CT} = 0.53002$	< 0.001
	Among populations within groups	8.80	$\Phi_{SC} = 0.40995$	< 0.001	15.68	$\Phi_{SC} = 0.33362$	< 0.001
	Within populations	12.67	$\Phi_{ST} = 0.87328$	< 0.001	31.32	$\Phi_{ST} = 0.68681$	< 0.001
Grouping by geographical regions II†	Among groups	91.54	$\Phi_{CT} = 0.91540$	< 0.001	72.90	$\Phi_{CT} = 0.72904$	< 0.001
	Among populations within groups	4.06	$\Phi_{SC} = 0.48043$	< 0.001	11.49	$\Phi_{SC} = 0.42416$	< 0.001
	Within populations	4.40	$\Phi_{ST} = 0.95604$	< 0.001	15.60	$\Phi_{ST} = 0.84397$	< 0.001
Grouping by hosts‡	Among groups	21.36	$\Phi_{CT} = 0.2136$	< 0.001	4.58	$\Phi_{CT} = 0.04581$	0.1085
	Among populations within groups	63.26	$\Phi_{SC} = 0.8045$	< 0.001	66.29	$\Phi_{SC} = 0.69478$	< 0.001
	Within populations	15.38	$\Phi_{ST} = 0.8462$	< 0.001	29.12	$\Phi_{ST} = 0.70876$	< 0.001

*Group 1: Italy and eastern France (samples 1–24); group 2: western France and eastern Spain, including the Ebro Valley (samples 26–40, 47–51; including contact zones 26, 27 and 50, 51); group 3: Pyrenees (samples 41–46; including contact zones 41 and 44–46); group 4: western Spain and Portugal (samples 52–61); clade A: the same three first groupings but without samples 50, 51 and 52–61; see Appendix S7.

†Same regional grouping as I but samples 26, 27, 41, 44–46, 50 and 51 (all the putative contact zones) were removed from the data set.

‡Group 1: *Pinus halepensis* and *P. pinea* (samples 1, 5, 18–20, 24b, 41b, 48, 55); group 2: *P. pinaster* (samples 24a, 36b, 37–38, 40, 41a, 54, 56–61); group 3: *P. nigra* (samples 1, 3, 6, 11a, 13–14, 16, 21–23, 25, 26, 27a, 28, 30–35, 36a, 39, 45, 47, 49–53); group: 4: *P. sylvestris*, *P. uncinata*, *Pseudotsuga menziesii* (4, 7–10, 12, 15, 17, 27b, 29, 42–44, 46); samples collected on several host trees were divided into subsamples (a, b) attributed to the corresponding groups; clade A : the same four groups without individuals from clade B.

Table 3 Results of mismatch distribution and neutrality tests against population growth for each cytochrome *c* oxidase subunit I haplogroup of *Thaumetopoea pityocampa* and for the whole data set.

	Haplogroups*				
	A_1	A_2	A_3	B	Global
Parameters estimated under the sudden expansion model					
θ_0	0.000	0.000	0.000	0.000	0.000
θ_1	3.512	0.479	99999.00	99999.00	99999.00
τ	0.500	3.000	0.973	1.607	0.455
Goodness-of-fit test					
SSD	0.00114	0.00145	0.01566	0.02378	0.23466
P-value	0.63800	0.62100	0.67400	0.01300	0.01000
Expansion	OK	OK	OK	NO	NO
Tests of selective neutrality					
Fu's F_S	−12.39260	−15.83490	−2.38929	−5.13665	−13.06930
P-value	0.00000	0.00000	0.07621	0.01003	0.01610
Expansion	OK	OK	NO	OK	OK
R_2	0.01907	0.0199	0.0726	0.0592	0.0532
P-value	0.04266	0.02612	0.13155	0.09548	0.21100
Expansion	OK	OK	NO	NO	NO

*Haplogroups as defined in Fig. 2: A_1, haplotypes 1–13 (Italy and eastern France); A_2, haplotypes 14–17, 25, 30–38 (western France and eastern Spain); A_3, haplotypes 18–24 (Pyrenees); B, haplotypes 26–29, 39–46 (western Spain and Portugal).
θ_0 pre-expansion and θ_1 post-expansion population sizes; τ, time in number of generations since the sudden expansion episode; SSD, sum of squared deviations; R_2, Ramos-Onsins and Rozas' R_2.

model and the results of goodness-of-fit and selective neutrality tests are presented in Table 3. For the whole data set, the mismatch distribution exhibits a bimodal curve and the expansion model is rejected ($P = 0.01$). Consistently, the R_2 test does not reject neutrality ($R_2 = 0.053$, $P = 0.211$), and only Fu's F_S shows a significant negative value ($F_S = -13.069$, $P = 0.016$). Conversely, haplogroups A1 and A2 both exhibit a unimodal curve and all tests detect a departure from neutrality. The Pyrenean haplogroup A3 also exhibits a unimodal curve, but only the goodness-of-fit test suggests population expansion. Nevertheless, the F_S value is negative but not significant. For the western Iberian haplogroup (B), only Fu's F_S test indicates a departure from neutrality ($F_S = -5.137$; $P = 0.01$).

DISCUSSION

Phylogeographic population structure

Two deeply divergent clades in western Europe

Our results demonstrate that the western European populations consist of two deeply divergent clades with a strong geographical structure. One of these clades (A) is widely distributed from eastern Iberia to the Italian Peninsula, whereas the second one (B) only occurs in central and western Iberia. A strong phylogeographic pattern is found in the present-day populations as shown by N_{ST} being significantly greater than G_{ST}. This demonstrates that the most related haplotypes tend to co-occur in the same geographic area. The allopatric separation between the two major western European lineages was likely to have been maintained through the Quaternary climatic oscillations, suggesting that even during the most favourable periods, the gene pools remained isolated. The genetic distances found between clades (Appendix S3) are compatible with a recent phylogenetic study in which this divergence was estimated to date back to *c.* 1.8 Ma (Kerdelhué *et al.*, 2009), i.e. the divergence is much older than the last few glacial cycles. Whatever the cause, the present delimitation between the two clades is south–north oriented, and definitely does not correspond to the Pyrenees. These results suggest the existence of a barrier to gene flow between eastern and western Iberia. Interestingly, the distributions of several Iberian endemic plant and animal species suggest a similar east to west polarity, with a trend for the areas of endemism to coincide with the largest mountain ranges (García-Barros *et al.*, 2002). In our case, the separation between western and eastern Iberia could be due either to the existence of a region where environmental conditions remained unsuitable, or to a gap in host availability. Recent studies suggest that pine hosts were present in at least some parts of the distribution areas of clades A and B even during the Last Glacial Maximum (LGM) (Willis *et al.*, 1998; Cheddadi *et al.*, 2006; Gómez & Lunt, 2006; Benito Garzón *et al.*, 2007), but pines also repeatedly experienced population fragmentation when the terrain was dominated either by other tree species or by steppe vegetation during the driest (cold or warm) phases (Willis *et al.*, 1998; Suc & Popescu, 2005; Carrión *et al.*, 2009).

Within clade A, both the haplotype network and the AMOVA show the existence of three groups of haplotypes that are spatially structured (Figs 1 & 2). In many species of phytophagous insects, the host plant is expected to play a role in population structure. However, no significant host effect

was observed within clade A (Table 2). The three haplogroups had a star-shaped topology, which could be a genetic signature of population growth (Slatkin & Hudson, 1991; Rogers & Harpending, 1992) consistent with post-glacial expansion. Plausible scenarios for each of the three groups are discussed below.

Haplogroup A3, a mountain lineage originating from the eastern Pyrenees

Haplogroup A3 is a monophyletic lineage that occurs only in the Pyrenees. It probably differentiated in this area over the most recent glacial cycles. Within this strictly Pyrenean lineage, most of the private haplotypes and the highest diversity parameters are observed in the samples from the eastern Pyrenees (sites 41, 42 and 44; Table 1, Appendix S6), suggesting the existence of a refugial area. This region, known as a biodiversity hotspot (Médail & Diadema, 2009), probably satisfied both temperature and host requirements during the LGM, in spite of its close proximity to the Pyrenean ice sheet. It could have benefited both from the adiabatic warming of downward air masses (Brown & Lomolino, 1998) and from the sea buffer effect. Moreover, the ice sheets were restricted to mountain systems over 1500 m and to some adjacent valleys (Jalut *et al.*, 1982). Pollen and fossil records support the local continuous occurrence of pine species despite strong variation in abundance (González-Sampériz *et al.*, 2005; Cheddadi *et al.*, 2006). Molecular data indicate the persistence of montane pine species (Gómez & Lunt, 2006; Afzal-Rafii & Dodd, 2007) and suggest the possible occurrence of relictual and rather coastal populations of the Aleppo pine (Gómez *et al.*, 2005). Continuous host availability and favourable climatic conditions could thus have allowed the pine processionary moth to survive the glaciation in the eastern Pyrenees. Interestingly, a similar hotspot of genetic diversity was found in the same region for other species (Horn *et al.*, 2009).

Haplogroup A2, a lineage occurring from Spain to France and showing a phylogeographical pattern of 'southern richness and northern purity'

Within the A2 group, haplotypic and nucleotidic diversities are significantly and negatively correlated with latitude. The highest values of these parameters are found in eastern Iberia, while most of the populations north of the Pyrenees are monomorphic (Table 1, Appendix S6). This is consistent with the 'southern richness and northern purity' pattern, well known for numerous temperate taxa (Hewitt, 1999). The southern areas where these species persisted through glaciations would have accumulated and maintained a high genetic diversity that mirrors ancestral diversity, while founder effects during northward post-glacial expansion led to the loss of genetic variation in the recolonized areas (Hewitt, 1999, 2004; Canestrelli *et al.*, 2006). Even during the LGM, eastern Iberia offered spatial and elevational climatic gradients thanks to mountainous and coastal areas. The persistence of the pine processionary moth along the Mediterranean coast of Spain is thus supported by the putative past distribution of several hosts, including Mediterranean native pines (Carrión *et al.*, 2000; Carrión, 2002; Gómez *et al.*, 2005; Gómez & Lunt, 2006).

Haplogroup A1, a non-Iberian lineage possibly with more northerly refugial areas

Haplogroup A1 was the only one of the four major lineages that did not occur in the Iberian Peninsula. More extensive sampling, especially in the Italian and Balkan peninsulas, is needed to elucidate the origin of this lineage and to know whether distinct lineages occur in the unsampled eastern regions. Nevertheless, in the present data set, most of the diversity was found in south-eastern France, from the Massif Central to the Alps. It is worth noting that few diverging private haplotypes were found in one location along the north-western Italian coast, in Ligury (site 5 in Italy), which could reflect the existence of a localized coastal refugium south of the western Alps. This suggests that refugial areas were not confined to the southernmost parts of the peninsula during the LGM. Moreover, several rare and private haplotypes closely related to the most common one were found in eastern France (sites 10–12, 16, 18, 20, 21, 24, 25; Table 1, Fig. 1). They could have independently appeared from point mutations during or following range expansion, but, for this shallowly divergent lineage, some of them might also originate from a more northerly and diffuse refugial area, as was hypothesized for other temperate species (Provan & Bennett, 2008; Horn *et al.*, 2009; Médail & Diadema, 2009). Two of the private haplotypes occurred in very recent expansion areas where pine afforestation dates back to the 19th century (sites 16, 25), but the southernmost ones occurred in areas where some pine species (*P. nigra* for instance) probably occurred throughout the glacial ages (Afzal-Rafii & Dodd, 2007; Beaudoin *et al.*, 2007), allowing the persistence of associated insect species. It was recently suggested that palaeoenvironments in southern France were more complex than previously thought (Blondel & Aronson, 1999; Médail & Diadema, 2009) and might have permitted the local survival of populations of the pine processionary moth.

To summarize, clade A exhibits at least three main refugial areas located along the Mediterranean coast: (1) along the Spanish shore from the Betic to the Iberian Chain, (2) in the eastern Pyrenees, and (3) probably near the Massif Central and the Alps and possibly in the unsampled eastern range of the pine processionary moth. *Pinus nigra* and/or *P. sylvestris* probably persisted in all the glacial refugia identified for clade A (Cheddadi *et al.*, 2006; Gómez & Lunt, 2006; Afzal-Rafii & Dodd, 2007). These pines occur at present mainly from the meso- to the mountain-mediterranean belt, and from the supra- to the oro-mediterranean belt, respectively, and probably largely predominated in the Pyrenean refugial areas of the pine processionary moth. On the other hand, eastern and south-eastern Iberia were major refugia for *P. halepensis*

and/or *P. pinaster* (Gómez *et al.*, 2005), which occur at present from the thermo- to the meso-mediterranean belts. We can thus hypothesize that refugial populations of *T. pityocampa* mostly survived the ice ages on *P. nigra*, which is nowadays the preferred host for egg-laying (Huchon & Démolin, 1970; Montoya, 1981). Yet haplogroup A2 may also have survived the glaciations on Mediterranean pines, and could exhibit different adaptation to pine hosts. Concerning clade B, the available sampling did not permit us to clearly describe the patterns of distribution of genetic diversity or to identify the regions of endemism. A better sampling all over the Iberian Peninsula will probably allow the identification of additional refugial areas.

Role of mountainous areas in structuring populations

No detectable role of physical barriers to dispersal…

Based on previous studies (Kerdelhué *et al.*, 2006; Santos *et al.*, 2007), it had been hypothesized that the Pyrenees, the Massif Central and maybe all mountain ranges could have posed a barrier to dispersal during the post-glacial expansion of this cold-sensitive species with short-range dispersal. It was thus expected that the favourable low-elevation habitats on each slope of the main ranges were colonized by different lineages, still separated by unsuitable high-elevation areas (Italian versus French Alps, eastern versus western Massif Central, and southern versus northern Pyrenees). Secondary contact zones with higher genetic diversity were expected to occur where favourable habitats connect the two sides. The present study, based on a much more extensive sampling, now rules out this hypothesis, as we show that the same haplogroup occurs on all slopes of any given mountain range.

Lineage A1 occurs from southern Italy to eastern France, showing that the Alps do not separate lineages originating from different refugial areas as known for several other taxa (Hewitt, 1999, 2004; Schmitt, 2007). In France, we hypothesized that the higher-elevation areas of the Massif Central, which separate a wide western and a more abrupt eastern side, contributed to strongly structure the populations, as was suggested in a preliminary study using microsatellite markers (Kerdelhué *et al.*, 2006). Our results using mitochondrial sequences confirmed this east–west differentiation, but showed that the two lineages are not separated by the high-elevation areas of the south-eastern Massif Central. On the contrary, lineage A1 occupied all suitable areas from eastern France up to the western side of the Massif Central, and the A1/A2 contact zone is located there in lowlands at the foot of this mountain range (Fig. 1, site 27), where there is no obvious physical barrier to dispersal. However, this secondary contact zone might be of very recent origin, because the native forests of *P. pinaster* (in the west) and *P. sylvestris* (in the Massif Central) have been connected by artificial plantations.

One of our major questions was to determine whether the high genetic diversity observed in the eastern Pyrenees resulted from admixture (secondary contact of diverged lineages) or from retention of ancestral variation (reflecting a glacial refugium imprint). We identified that one of the genetic lineages, namely haplogroup A3, managed to survive the glaciations *in situ*. Nevertheless, the four identified lineages A1, A2, A3 and B appeared to be in contact near the southern rim of the Pyrenees. North-eastern Spain was thus both an admixture area and a centre of differentiation, which means that the role of the Pyrenees in structuring populations was more complex than merely posing a physical barrier to dispersal. Lineages did not all respond in the same way to the climatic oscillations. While lineage A3 probably colonized both sides of the Pyrenean range in an upward and westward movement after the ice sheet retreat, but did not contribute to northward recolonization of newly suitable environments, lineage A2 expanded mainly latitudinally and colonized the south-western French lowlands, bypassing the Pyrenees to the west. Artificial plantations and, more recently, climate change further allowed colonization of northern France. A more limited spatial expansion and a gradual upslope movement could thus account for the contradictory results of the expansion tests for the haplogroup A3, contrary to A2 (Table 2).

…but a possible role via the elevational distribution of the host species

Maternal lineages A2 and A3 show very different responses to past climatic oscillations that might be explained by contrasting responses of their host species to post-glacial warming. Haplogroup A3 probably originated from a glacial refugium located in the eastern Pyrenees and did not extend much geographically. Our results rather suggest that it responded to glacial/interglacial cycles by limited upslope movements and was 'trapped' within a mountainous zone. In this region, the montane pine species were probably the main continuously available hosts (González-Sampériz *et al.*, 2005), but these species did not contribute to post-glacial recolonization of northern Europe because of *in situ* persistence and vertical migrations throughout climatic pulses (Robledo-Arnuncio *et al.*, 2005; Cheddadi *et al.*, 2006; Afzal-Rafii & Dodd, 2007). We hypothesize that lineage A3 tracked the early recolonization of the Pyrenean range by the largely dominant pine species and was consequently trapped by vertical migration. On the contrary, lineage A2 most probably survived the ice ages in refugia located along the eastern coast of Spain, where the Mediterranean pines *P. pinaster* and *P. halepensis* could also have persisted (Gómez *et al.*, 2005). During warming periods, these thermophilic lowland pine species could have made possible the expansion to the north, and thus the moth could have reached the lowlands of western France from the eastern Iberian Chain. Interestingly, this expansion pathway corresponds to one of the migration routes suggested for *P. pinaster* (Salvador *et al.*, 2000), which would be consistent with the moth following the migration route of one of its main hosts.

Sampling the entire range would allow one to test whether all the Mediterranean refugia of montane pines, especially *P. nigra*, correspond to differentiation centres of the moth, and if the major dispersal centres are associated with expansions of the lowland pines. Rather than showing that mountains acted as physical barriers to dispersal, our results suggest that topography played a major role in shaping the distribution of maternal lineages through the demographic history of its main host plants. Most mid-elevation regions served as glacial refugia, and the moth later expanded into lowlands from these bottlenecked populations, following its relatively thermophilic pine hosts. Mountains offered suitable environmental conditions along the slopes that permitted the persistence of this oligophagous insect during the glacial and interglacial periods. The rest of the species' range could be recurrently recolonized by spatial expansions from these refugia.

ACKNOWLEDGEMENTS

The authors are indebted to John Cortes, Ramón González Ruiz, Enrique Martín Bernal, José A. Hódar, Carlos Lopez-Vaamonde, Manuela Branco, Helena Santos, Eustachio Tarasco, Miguel Ángel Gómez de Dios, Neus Marí Mena, Zvi Mendel, Jean-Claude Martin, René Mazet, Pierre Ménassieu, Patrick Pineau, Jacques Garcia and Francis Goussard for their help in collecting samples. We are also very grateful to Emmanuelle Magnoux, Claudine Courtin, Yeliz Yildiz, Nese Kaplan, Magalie Devavry and Antoine Daunay for technical laboratory assistance. We gratefully acknowledge Olivier Denux for editing the map and Marie-Anne Auger-Rozenberg, Thierry Lamant, Carlos Lopez-Vaamonde and David Lees for useful comments and language revision. This work was supported by grants from the Agence Nationale de la Recherche (ANR 07 BDIV 013-07-URTICLIM), from the French Ministry of Agriculture (DGFAR N°59-02 E13/05) and from the European Union (PROMOTH QLK5-CT-2002-00852).

REFERENCES

Afzal-Rafii, Z. & Dodd, R.S. (2007) Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in western Europe. *Molecular Ecology*, **16,** 723–736.

Barbéro, M., Loisel, R., Quézel, P., Richardson, D.M. & Romane, F. (1998) Pines of the Mediterranean Basin. *Ecology and biogeography of* Pinus (ed. by D.M. Richardson), pp. 153–170. Cambridge University Press, Cambridge.

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,** 2084–2096.

Beaudoin, C., Jouet, G., Suc, J.P., Berné, S. & Escarguel, G. (2007) Vegetation dynamics in southern France during the last 30 ky BP in the light of marine palynology. *Quaternary Science Reviews*, **26,** 1037–1054.

Benito Garzón, M.B., Sánchez de Dios, R. & Sáinz Ollero, H. (2007) Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and mid-Holocene. *Ecography*, **30,** 120–134.

Bhagwat, S.A. & Willis, K.J. (2008) Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography*, **35,** 464–482.

Blondel, J. & Aronson, J. (1999) *Biology and wildlife of the Mediterranean region*. Oxford University Press, Oxford.

Brown, J.H. & Lomolino, M.V. (1998) *Biogeography*, 2nd edn. Sinauer Associates, Sunderland, MA.

Canestrelli, D., Cimmaruta, R., Costantini, V. & Nascetti, G. (2006) Genetic diversity and phylogeography of the Apennine yellow-bellied toad *Bombina pachypus*, with implications for conservation. *Molecular Ecology*, **15,** 3741–3754.

Carrión, J.S. (2002) Patterns and processes of Late Quaternary environmental change in a montane region of southwestern Europe. *Quaternary Science Reviews*, **21,** 2047–2066.

Carrión, J.S., Navarro, C., Navarro, J. & Munuera, M. (2000) The distribution of cluster pine (*Pinus pinaster*) in Spain as derived from palaeoecological data: relationships with phytosociological classification. *The Holocene*, **10,** 243–252.

Carrión, J.S., Fernández, S., Jiménez-Moreno, G., Fauquette, S., Gil-Romera, G., González-Sampériz, P. & Finlayson, C. (2009) The historical origins of aridity and vegetation degradation in southeastern Spain. *Journal of Arid Environments*, doi:10.1016/j.jaridenv.2008.11.014.

Cheddadi, R., Vendramin, G., Litt, T., François, L., Kageyama, M., Lorentz, S., Laurent, J.M., de Beaulieu, J.L., Sadori, L., Jost, A. & Lunt, D. (2006) Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography*, **15,** 271–282.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9,** 1657–1659.

DeChaine, E.G. & Martin, A.P. (2005) Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance. *Journal of Biogeography*, **32,** 1943–1956.

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

Felsenstein, J. (1988) Phylogenies from molecular sequences: inferences and reliability. *Annual Review of Genetics*, **22,** 521–565.

Fu, Y.-X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147,** 915–925.

García-Barros, E., Gurrea, P., Luciáñez, M.J., Cano, J.M., Munguira, M.L., Moreno, J.C., Sainz, H., Sanz, M.J. & Simón, J.C. (2002) Parsimony analysis of endemicity and its application to animal and plant geographical distributions in the Ibero-Balearic region (western Mediterranean). *Journal of Biogeography*, **29,** 109–124.

Gómez, A. & Lunt, D.H. (2006) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia – evolutionary perspectives on the origins and conservation of European biodiversity* (ed. by S. Weiss and N. Ferrand), pp. 155–188. Kluwer Academic Publishers, Dordrecht.

Gómez, A., Vendramin, G.G., González-Martínez, S.C. & Alía, R. (2005) Genetic diversity and differentiation of two Mediterranean pines (*Pinus halepensis* Mill. and *Pinus pinaster* Ait.) along a latitudinal cline using chloroplast microsatellite markers. *Diversity and Distributions*, **11,** 257–263.

González-Sampériz, P., Valero, B., Carrión, J.S., Peña, J.L., García Ruíz, J.M., Sancho, C. & Martí Bono, C. (2005) Glacial and lateglacial vegetation in northeastern Spain: new data and a review. *Quaternary International*, **140–141,** 4–20.

Habel, J.C., Schmitt, T. & Müller, P. (2005) The fourth paradigm pattern of post-glacial range expansion of European terrestrial species: the phylogeography of the marbled white butterfly (Satyrinae, Lepidoptera). *Journal of Biogeography*, **32,** 1489–1497.

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

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

Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359,** 183–195.

Horn, A., Stauffer, C., Lieutier, F. & Kerdelhué, C. (2009) Complex postglacial history of the temperate bark beetle *Tomicus piniperda* L. (Coleoptera, Scolytinae). *Heredity*, **103,** 238–247.

Huchon, H. & Démolin, G. (1970) La bioécologie de la processionnaire du pin. Dispersion potentielle – dispersion actuelle. *Revue Forestière Française, no. spécial 'La lutte biologique en forêt'*, **22,** 220–234.

Jalut, G., Delibrias, G., Dagnac, J., Mardones, M. & Bouhours, M. (1982) A palaeoecological approach to the last 21 000 years in the Pyrenees: the peat bog of Freychinede (alt. 1350 m, Ariège, south France). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **40,** 321–359.

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., 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.

Médail, F. & Diadema, K. (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, **36,** 1333–1345.

Montoya, R. (1981) La procesionaria del pino. *Plagas de Insectos en las Masas Forestales Españolas*. Ministerio de Agricultura Pesca y Alimentación, Madrid.

Pons, O. & Petit, R.J. (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144,** 1237–1245.

Posada, D. (2003) Using Modeltest and PAUP* to select a model of nucleotide substitution. *Current protocols in bioinformatics* (ed. by A.D. Baxevanis, D.B. Davison, R.D.M. Page, G.A. Petsko, L.D. Stein and G.D. Stormo), pp. 6.5.1–6.5.14. John Wiley and Sons, New York.

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

Provan, J. & Bennett, K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, **23,** 564–571.

Ramos-Onsins, S.E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19,** 2092–2100.

Robinet, C., Baier, P., Pennerstorfer, J., Schopf, A. & Roques, A. (2007) Modelling the effects of climate change on the potential feeding activity of *Thaumetopoea pityocampa* (Den & Schiff.) (Lep., Notodontidae) in France. *Global Ecology and Biogeography*, **16,** 460–471.

Robledo-Arnuncio, J.J., Collada, C., Alía, R. & Gil, L. (2005) Genetic structure of montane isolates of *Pinus sylvestris* L. in a Mediterranean refugial area. *Journal of Biogeography*, **32,** 595–605.

Rogers, A.R. & Harpending, H.C. (1992) Population growth makes waves in the distribution of pairwise genetic distances. *Molecular Biology and Evolution*, **9,** 552–569.

Rosenzweig, C., Casassa, G., Karoly, D.J., Imeson, A., Liu, C., Menzel, A., Rawlins, S., Root, T.L., Seguin, B. & Tryjanowski, P. (2007) Assessment of observed changes and responses in natural and managed systems. *Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson), pp. 79–131. Cambridge University Press, Cambridge.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. & Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19,** 2496–2497.

Salvador, L., Alía, R., Agúndez, D. & Gil, L. (2000) Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait) in the Iberian peninsula. *Theoretical and Applied Genetics*, **100,** 89–95.

Santos, H., Rousselet, J., Magnoux, E., Paiva, M.R., Branco, M. & Kerdelhué, C. (2007) Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society B: Biological Sciences*, **274,** 935–941.

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

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.

Slatkin, M. & Hudson, R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129,** 555–562.

Stewart, J.R. & Lister, A.M. (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, **16,** 609–613.

Suc, J.P. & Popescu, S.M. (2005) Pollen records and climatic cycles in the North Mediterranean region since 2.7 Ma. *Early–Middle Pleistocene transitions: the land–ocean evidence* (ed. by M.J. Head and P.L. Gibbard), pp. 147–158. Geological Society of London Special Publication No. 247. Geological Society of London, London.

Swofford, D.L. (2003) *PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4*. Sinauer Associates, Sunderland, MA.

Varga, Z.S. & Schmitt, T. (2008) Types of oreal and oreotundral disjunctions in the western Palearctic. *Biological Journal of the Linnean Society*, **93,** 415–430.

Willis, K.J. & van Andel, T.H. (2004) Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews*, **23,** 2369–2387.

Willis, K.J., Bennett, K.D. & Birks, H.J.B. (1998) The late Quaternary dynamics of pines in Europe. *Ecology and biogeography of* Pinus (ed. by D.M. Richardson), pp. 107–121. Cambridge University Press, Cambridge.

SUPPORTING INFORMATION

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

Appendix S1 Geographical coordinates of sampling locations and year of collection.

Appendix S2 Polymorphic sites of the 46 mitochondrial cytochrome *c* oxidase subunit I (COI) haplotypes.

Appendix S3 Pairwise genetic distances between haplotypes: (a) matrices and (b) histograms.

Appendix S4 Maximum likelihood and maximum parsimony estimates of haplotype phylogeny.

Appendix S5 Host plant species mapped onto the haplotype network.

Appendix S6 Population parameters for each haplogroup considered separately.

Appendix S7 Map of sampling locations with the geographical groupings used in the analyses of molecular variance (AMOVA).

Appendix S8 Mismatch distribution curves.

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

BIOSKETCH

Jérôme Rousselet works on the phylogeography and molecular evolution of forest insects, with a special focus on expansion processes. The focus of the research team is on the ecology and evolution of native and invasive forest insects in the context of global change, with a particular interest in molecular ecology.

Author contributions: J.R., A.R., A.B., M.S. and C.K. formulated the questions and planned the research, R.Z. and D.A. performed the research, and J.R. and C.K. analysed the data and led the writing, to which A.R. and A.B. contributed.

Editor: Melodie McGeoch