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
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 East, were never covered with ice during the Pleistocene, 1999, 2001). Some regions of the world, such as the Near

dispersal and life-history traits (Taberlet et al. 1998). Environmental conditions in particular may have imposed by the analytical procedures.

cs, 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 (Patrway 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
and monomorphic bands were scored. Both polymorphic units. AFLP profiles were recorded in a matrix as presence or absence of bands for each individual. The intensity of bands in the range 70–360 bp. AFLP profiles 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 PARS 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.

**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

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

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_e) 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 \( \Phi_{ST} \) 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 pityocampa 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.
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

<table>
<thead>
<tr>
<th>Country and location</th>
<th>Haplotype diversity</th>
<th>N</th>
<th>Thpit 1</th>
<th>Thpit 3</th>
<th>Thpit 4</th>
<th>Thpit 5</th>
<th>mean Ĥ</th>
<th>SD</th>
<th>N</th>
<th>Hs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Israel Yatir</td>
<td>15 0.00 ± 0.00</td>
<td>14</td>
<td>0.20 ± 0.09</td>
<td>0.51 ± 0.04</td>
<td>0.20 ± 0.10</td>
<td>0.00 ± 0.00</td>
<td>0.23</td>
<td>0.21</td>
<td>15</td>
<td>0.15 ± 0.01</td>
</tr>
<tr>
<td>Israel Qissufim</td>
<td>10 0.00 ± 0.00</td>
<td>9</td>
<td>0.00 ± 0.00</td>
<td>0.50 ± 0.06</td>
<td>0.29 ± 0.12</td>
<td>0.00 ± 0.00</td>
<td>0.20</td>
<td>0.25</td>
<td>10</td>
<td>0.18 ± 0.01</td>
</tr>
<tr>
<td>Israel Haruvit</td>
<td>15 0.00 ± 0.00</td>
<td>14</td>
<td>0.25 ± 0.10</td>
<td>0.45 ± 0.07</td>
<td>0.14 ± 0.08</td>
<td>0.00 ± 0.00</td>
<td>0.21</td>
<td>0.19</td>
<td>15</td>
<td>0.21 ± 0.01</td>
</tr>
<tr>
<td>Israel Segev</td>
<td>9 0.00 ± 0.00</td>
<td>10</td>
<td>0.28 ± 0.12</td>
<td>0.44 ± 0.09</td>
<td>0.10 ± 0.09</td>
<td>0.00 ± 0.00</td>
<td>0.21</td>
<td>0.19</td>
<td>9</td>
<td>0.20 ± 0.01</td>
</tr>
<tr>
<td>Israel Qyrat Shemona</td>
<td>14 0.00 ± 0.00</td>
<td>13</td>
<td>0.50 ± 0.10</td>
<td>0.32 ± 0.10</td>
<td>0.76 ± 0.06</td>
<td>0.32 ± 0.10</td>
<td>0.48</td>
<td>0.21</td>
<td>14</td>
<td>0.19 ± 0.01</td>
</tr>
<tr>
<td>Lebanon Beirut</td>
<td>24 0.00 ± 0.00</td>
<td>24</td>
<td>0.36 ± 0.07</td>
<td>0.47 ± 0.04</td>
<td>0.56 ± 0.08</td>
<td>0.19 ± 0.07</td>
<td>0.39</td>
<td>0.16</td>
<td>9</td>
<td>0.18 ± 0.01</td>
</tr>
<tr>
<td>Turkey Seyhköy</td>
<td>11 0.00 ± 0.00</td>
<td>20</td>
<td>0.85 ± 0.02</td>
<td>0.43 ± 0.07</td>
<td>0.55 ± 0.09</td>
<td>0.00 ± 0.08</td>
<td>0.53</td>
<td>0.24</td>
<td>8</td>
<td>0.20 ± 0.01</td>
</tr>
<tr>
<td>Turkey Iskanderun</td>
<td>10 0.71 ± 0.12</td>
<td>19</td>
<td>0.82 ± 0.04</td>
<td>0.60 ± 0.06</td>
<td>0.88 ± 0.04</td>
<td>0.10 ± 0.06</td>
<td>0.60</td>
<td>0.35</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Turkey Aladag</td>
<td>10 0.51 ± 0.16</td>
<td>20</td>
<td>0.73 ± 0.03</td>
<td>0.49 ± 0.04</td>
<td>0.67 ± 0.05</td>
<td>0.00 ± 0.00</td>
<td>0.47</td>
<td>0.33</td>
<td>9</td>
<td>0.18 ± 0.01</td>
</tr>
<tr>
<td>Turkey Pozanti</td>
<td>11 0.51 ± 0.10</td>
<td>20</td>
<td>0.65 ± 0.06</td>
<td>0.56 ± 0.07</td>
<td>0.66 ± 0.05</td>
<td>0.00 ± 0.00</td>
<td>0.42</td>
<td>0.31</td>
<td>10</td>
<td>0.18 ± 0.01</td>
</tr>
<tr>
<td>Cyprus El Skopi</td>
<td>18 0.74 ± 0.08</td>
<td>17</td>
<td>0.70 ± 0.05</td>
<td>0.52 ± 0.09</td>
<td>0.94 ± 0.02</td>
<td>0.58 ± 0.10</td>
<td>0.69</td>
<td>0.19</td>
<td>16</td>
<td>0.22 ± 0.01</td>
</tr>
<tr>
<td>Turkey Karaoz</td>
<td>8 0.46 ± 0.20</td>
<td>8</td>
<td>0.52 ± 0.13</td>
<td>0.13 ± 0.11</td>
<td>0.88 ± 0.05</td>
<td>0.00 ± 0.00</td>
<td>0.38</td>
<td>0.40</td>
<td>8</td>
<td>0.20 ± 0.01</td>
</tr>
<tr>
<td>Turkey Gunur</td>
<td>15 0.00 ± 0.00</td>
<td>13</td>
<td>0.31 ± 0.12</td>
<td>0.09 ± 0.08</td>
<td>0.89 ± 0.05</td>
<td>0.00 ± 0.00</td>
<td>0.32</td>
<td>0.40</td>
<td>13</td>
<td>0.25 ± 0.01</td>
</tr>
<tr>
<td>Turkey Aydin</td>
<td>10 0.20 ± 0.15</td>
<td>11</td>
<td>0.00 ± 0.00</td>
<td>0.09 ± 0.08</td>
<td>0.82 ± 0.04</td>
<td>0.09 ± 0.08</td>
<td>0.25</td>
<td>0.38</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Turkey Samsun</td>
<td>12 0.00 ± 0.00</td>
<td>20</td>
<td>0.40 ± 0.08</td>
<td>0.00 ± 0.00</td>
<td>0.53 ± 0.07</td>
<td>0.00 ± 0.00</td>
<td>0.23</td>
<td>0.27</td>
<td>6</td>
<td>0.16 ± 0.01</td>
</tr>
</tbody>
</table>

Ĥ, 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, Iskanderun and Seyhköy). 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, Iskanderun and Seyhköy; MS-Thpit4 in Karaoz, Aladag, Samsun and Iskanderun; MS-Thpit3 in Iskanderun). 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 Iskanderun 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
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.
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,
(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 transgressive 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 *Thaumetopoea* 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 transgressive 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. 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 transgressive 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 relic, isolated stands of *Pinus halepensis* exist far south in Israel, but were probably exempt from the moth until recently, as *T. wilkinsonii* 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
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 Qsufim, 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; Białozyt 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


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 Rouesselet, 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.