The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2016, **119**, 311–328. With 6 figures.

Evidence for low-level hybridization between two allochronic populations of the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae)

CHRISTIAN BURBAN[1]*, MATHIEU GAUTIER[2], RAPHAEL LEBLOIS[2], JULIE LANDES[2†,‡], HELENA SANTOS[3], MARIA-ROSA PAIVA[4], MANUELA BRANCO[3] and CAROLE KERDELHUE[2]

[1]*BIOGECO, INRA, University Bordeaux, 33610, Cestas, France*
[2]*INRA, UMR CBGP, F-34988 Montferrier-sur-Lez, France*
[3]*CEF, Instituto Superior de Agronomia (ISA), Universidade de Lisboa (ULisboa), Lisboa, 1349-017, Portugal*
[4]*CENSE, Faculty of Sciences and Technology (FCT), Universidade Nova de Lisboa (UNL), 2829-516 Caparica, Portugal*

Received 10 February 2016; revised 8 April 2016; accepted for publication 8 April 2016

Divergence between populations sharing the same habitat can be initiated by different reproductive times, leading to allochronic differentiation. A spatially localized allochronic summer population (SP) of the pine processionary moth *Thaumetopoea pityocampa*, recently discovered in Portugal, occurs in sympatry with the local winter population (WP). We examined the level of genetic differentiation between the two populations and estimated the current gene flow within the spatial framework of their co-occurrence. Mitochondrial data indicated that the two sympatric populations were genetically closer than other WP populations. Conversely, microsatellite genotyping uncovered greater differentiation between the two sympatric populations than between allopatric ones. While male trapping confirmed that reproduction of SP and WP occurred at distinct times, clustering approaches demonstrated the presence of a few LateSP individuals emerging within the WP flight period, although genetically identified as SP. We also identified rare recent hybridization events apparently occurring mainly in the margins of the current SP range. The ongoing gene flow detected between the ancestral and the emerging allochronic populations revealed an incomplete reproductive isolation, which must therefore be taken into account and integrated with studies focussed on ecological drivers, so that a complete understanding of the ongoing speciation process might be achieved. © 2016 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2016, **119**, 311–328.

KEYWORDS: allochrony – gene flow – hybrid identification – incipient speciation – sympatry.

INTRODUCTION

Sympatric speciation is a concept that has remained controversial since its introduction (Sulloway, 1979; Mayr, 1982; Filchak, Roethele & Feder, 2000; Berlocher & Feder, 2002; Fitzpatrick *et al.*, 2007), although the need to identify the contexts and processes leading to sympatric divergence has been recognized (Fitzpatrick, Fordyce & Gavrilets, 2008). Furthermore, the dichotomous division between allopatric and sympatric speciation has been questioned, as non-exclusive processes can occur. As evolutionary forces and underlying mechanisms could act jointly and/or sequentially, a full understanding of evolution requires investigation of the order of occurrence of the different processes, their relative importance and interactions (Butlin *et al.*, 2012; Shaw & Mullen, 2014). Species may be envisaged as 'spatiotemporally

*Corresponding author. E-mail: christian.burban@pierroton.inra.fr
†Current address: Eco-Anthropologie et Ethnobiologie, UMR 7206 CNRS, MNHN, Univ. Paris Diderot, Sorbonne Paris Cité, F-75005 Paris, France
‡Current address: Mécanismes Adaptatifs et Evolution, UMR 7179, Sorbonne Universités, CNRS, MNHN, 91800 Brunoy, France

localised individuals, historical entities, rather than unrestricted classes' (Hull, 1978). However, the occurrence and role of gene flow during speciation is still contentious (Servedio *et al.*, 2011; Smadja & Butlin, 2011; Feder, Egan & Nosil, 2012; Via, 2012).

Differentiation of reproductive time is one of the mechanisms possibly preventing random mating and thus reducing gene flow. Such allochronic differentiation could occur gradually, with isolation by time acting in the same way as isolation by distance (Hendry & Day, 2005), or act as a barrier promoting discrete entities (Cooley *et al.*, 2001; Yamamoto & Sota, 2009, 2012). Separated reproductive time can evolve in sympatry under direct divergent selection pressures such as adaptation to local or changing climates giving rise to two optimal reproductive seasons (Fitter *et al.*, 1995; Yamamoto & Sota, 2009, 2012) or indirectly from the adaptation of other traits such as synchronization with mutualistic partners or hosts (Abbot & Withgott, 2004; Whipple *et al.*, 2009). Similarly to other mechanisms potentially involved in speciation, allochrony is influenced by the genetic variation and plasticity of the underlying traits (Barrett & Schluter, 2008; Pfennig *et al.*, 2010), spatial scaling (Kisel & Barraclough, 2010) and population size (Devaux & Lande, 2008).

A case of plausible incipient allochronic speciation was suggested to occur in Portugal for the winter pine processionary moth (PPM), *Thaumetopoea pityocampa* Denis and Schiffermüller (Lepidoptera: Notodontidae). In this Mediterranean moth species, adults mate in the summer and larval development occurs during the winter, representing the classical 'winter phenology'. After a gregarious larval development, the caterpillars leave the nest in typical head-to-tail processions and bury in the soil for pupal diapause and metamorphosis, until the adults emerge in the following summer. In contrast, a unique atypical summer population (SP) was discovered in 1997 in the Mata Nacional de Leiria (hereafter, MNL), a pure maritime pine forest where a 'winter population' (WP) co-occurs, following the classical cycle (Pimentel *et al.*, 2006). It was observed for the first time very locally over 200 ha in the south of the MNL during an outbreak (M.-R. Paiva, pers. observ.), and later expanded both northward and southward, as demonstrated by monitoring for several years (Santos *et al.*, 2007; Roques *et al.*, 2015). SP adults emerge and lay eggs from the spring to the beginning of the summer, thus enabling SP larvae to complete development until the autumn (Santos *et al.*, 2011a). Since the adults are short lived (3–4 days at most), the observed differences in phenology between the SP and local WP imply disjoint reproductive periods. SP is characterized by specific phenotypic traits, such as a lower fecundity than the WP, combined with larger eggs and a different egg cover (Santos *et al.*, 2013). For the early larval stages, an increased tolerance to high temperatures was demonstrated, consistent with the summer development of the larvae (Santos *et al.*, 2011b). At the genetic level, Gschloessl *et al.* (2014) found a mean identity of 99.5% between pairs of homologous transcripts from the two populations. In a previous study, nuclear Internal Transcribed Spacer 1 (ITS-1) and partial sequences of cytochrome oxidase I (COI) demonstrated that SP was phylogenetically closely related to the local WP (Santos *et al.*, 2007). While these markers depicted geographical differentiation between winter populations across the Iberian Peninsula, they showed only limited differentiation between the sympatric SP and WP in MNL. In contrast, using five and six microsatellite loci, respectively, Santos *et al.* (2007, 2011a) found a high differentiation between these two populations, possibly resulting from a founder effect.

Altogether, these findings showed sympatric divergence, probably resulting from differences in the phenology (Santos *et al.*, 2011a). Also noticeable was the presence of rare adults belonging to the same genetic cluster as SP individuals, but emerging within the reproductive period of the sympatric WP (hereafter referred as 'LateSP'). The occurrence of potential SP $\times$ WP hybrids was also suggested, but the limited number of loci used weakened the conclusion. Interestingly, successful experimental SP $\times$ WP crosses and viable offspring were obtained under laboratory conditions. High heritability was demonstrated for reproductive time in each population (Branco *et al.*, 2016).

The objectives of the present work were to address the genetic interactions effectively taking place between the two allochronic populations and, in particular, to question the occurrence of gene flow in this remarkable system of ongoing differentiation. We first used a longer polymorphic mitochondrial gene and a larger sampling to investigate the genetic diversity and structure at a regional scale, and confirm the proximity of the SP and WP populations from a mitochondrial perspective. Second, we focused on the region of Leiria in order to: (1) characterise the nuclear genetic differentiation at a local scale; (2) analyse in detail whether LateSP and introgressed, hybrid individuals actually occurred, or if their preliminary observation in previous works could be due to technical or sampling biases. To reach these goals, we significantly increased the number of microsatellite markers (from 5 to 17) and used both classical population genetic tools and dedicated Bayesian approaches to detect diverse hybrid classes. To test the effectiveness of the approach in our particular model system, we ran all analyses on

a simulated data set (based on the allelic distributions observed in the field) to determine the robustness of the obtained assignments for parental and hybrid categories. We then analysed the natural populations, using a conservative approach to interpret only the undoubtedly LateSP or hybrid individuals, that provided a first spatio-temporal insight into gene flow and plasticity in the studied diverging populations.

MATERIAL AND METHODS

Sampling and molecular markers

Sampling covered the known range of the SP, extending for 50 km along the coastal strip of Leiria District, cited hereafter as the Leiria zone, where SP and WP occur in sympatry. It included the plots where SP was first detected in the South of MNL, as well as the areas recently invaded, extending northwards to the limit of MNL and southwards to Nazaré and São Martinho, outside the MNL. Winter populations (WP) were also sampled in eight additional Iberian sites: Alcácer, Apostiça and Viseu in Portugal and Ciudad Rodrigo, Collado Mediano, Casa de don Pedro, Jaen, and Undiano in Spain (Figs 1B, 6A). In the Leiria zone, the following PPM stages were collected: (1) larvae and adults sampled from 2002 to 2008 and previously analysed with a subset of six microsatellite loci (Santos *et al.*, 2007, 2011a); (2) females collected as pupae, or L5 larvae, that emerged in the laboratory in 2009 and (3) males caught by pheromone trapping, from 2008 to 2010, following the method described in Santos *et al.* (2007, 2011a). Male sampling covered the entire period of adult flight, that is approximately from the end of April to the end of September, with small differences among years, regarding the date of trap installation and frequency of trap survey, which was conducted either weekly, or fortnightly. Phenology data, i.e. the date of collection of young larvae, the date of emergence in the laboratory of the females collected as pupae or as L5 larvae, and the date of male trapping, were used to allocate individuals *a priori*, according to their phenotype, that is either SP or WP, as described in Santos *et al.* (2011a). Information concerning samples used in the present study is summarised in Table 1, and detailed in Supporting Information (Table S1). For each individual, DNA was extracted as described in Santos *et al.* (2007).

Mitochondrial sequencing was performed using 98 individuals (50 SP and 48 WP) collected in the Leiria zone, pulled together with 69 individuals from other eight Iberian populations (Fig. 1B), which represented the West Iberian PPM clade previously identified in a phylogeographic survey based on COI sequences (Rousselet *et al.*, 2010). The near full length mitochondrial COI gene, 1499 bp long, was amplified and sequenced using the primers LCO 1490 (5′-GGTCAACAAATCATAAAGATATTGG-3′) (Folmer *et al.*, 1994) and Pat TL2-N-3014 (5′-TCCAA TGCACTAATCTGCCATATTA-3′) (Simon *et al.*, 1994). PCR were performed in a final volume of 25 μL containing 2 U *Taq* polymerase and 1× buffer (Biolabs), 2.5 μg bovine serum albumin, 0.5 mM $MgCl_2$, 0.4 μM of each primer and 0.2 mM of each dNTP. Annealing temperature was set to 48 °C. Internal primers C1-J-2183 (Jerry) (5′-CAACATT-TATTTTGATTTTTTGG-3′) (Simon *et al.*, 1994) and the newly designed C1-N-2276 (5′-GCATAAAT-TATTCCYAAAC-3′) were used to ensure a complete double strand sequencing using the Big Dye terminator kit (Applied Biosystems). Microsatellite genotyping was realized for 321 individuals from the Leiria zone (163 SP and 158 WP), and 59 individuals from three Portuguese surrounding winter populations, namely Viseu in the north and Apostiça and Alcácer in the south. Nineteen microsatellite loci are now available for the PPM (Rousselet, Magnoux & Kerdelhué, 2004; Santos *et al.*, 2011a; Burban *et al.*, 2012), namely MS-*Thpit*01 to MS-*Thpit*19. We achieved a final data set for 17 loci, i.e. removing MS-*Thpit*02 because of the presence of null alleles and MS-*Thpit*06, further demonstrated to map on the Z chromosome (Santos *et al.*, 2011a; Gautier, 2014). Sequencing and genotyping were performed using an ABI 3730 automatic sequencer. Allele sizes were determined using the Genemapper v4.0 software (Applied Biosystems, Foster City, CA, USA).

Mitochondrial data analyses

Mitochondrial sequences were aligned with BIOEDIT 7.2.5 (Hall, 1999). Gene genealogy for the COI sequences was estimated by statistical parsimony using TCS 1.21 (Clement, Posada & Crandall, 2000). Diversity (*hs*, *ht*) and differentiation (*Gst*, *Nst*) indices were calculated with PERMUT, the phylogeographic signal being tested by comparison of *Gst* and *Nst* using 1000 permutations (Pons & Petit, 1996; Burban *et al.*, 1999). Pairwise *Fst* was assessed with ARLEQUIN 3.1 (Excoffier, Laval & Schneider, 2005), using 1000 permutations for significance inference.

Microsatellites data analyses

Diversity and population genetic structure

Concerning microsatellites data, diversity indices and tests for allelic associations were performed using ARLEQUIN. For each population and locus, departure from Hardy–Weinberg equilibrium (HWE) was tested using 1000 permutations and 100 000

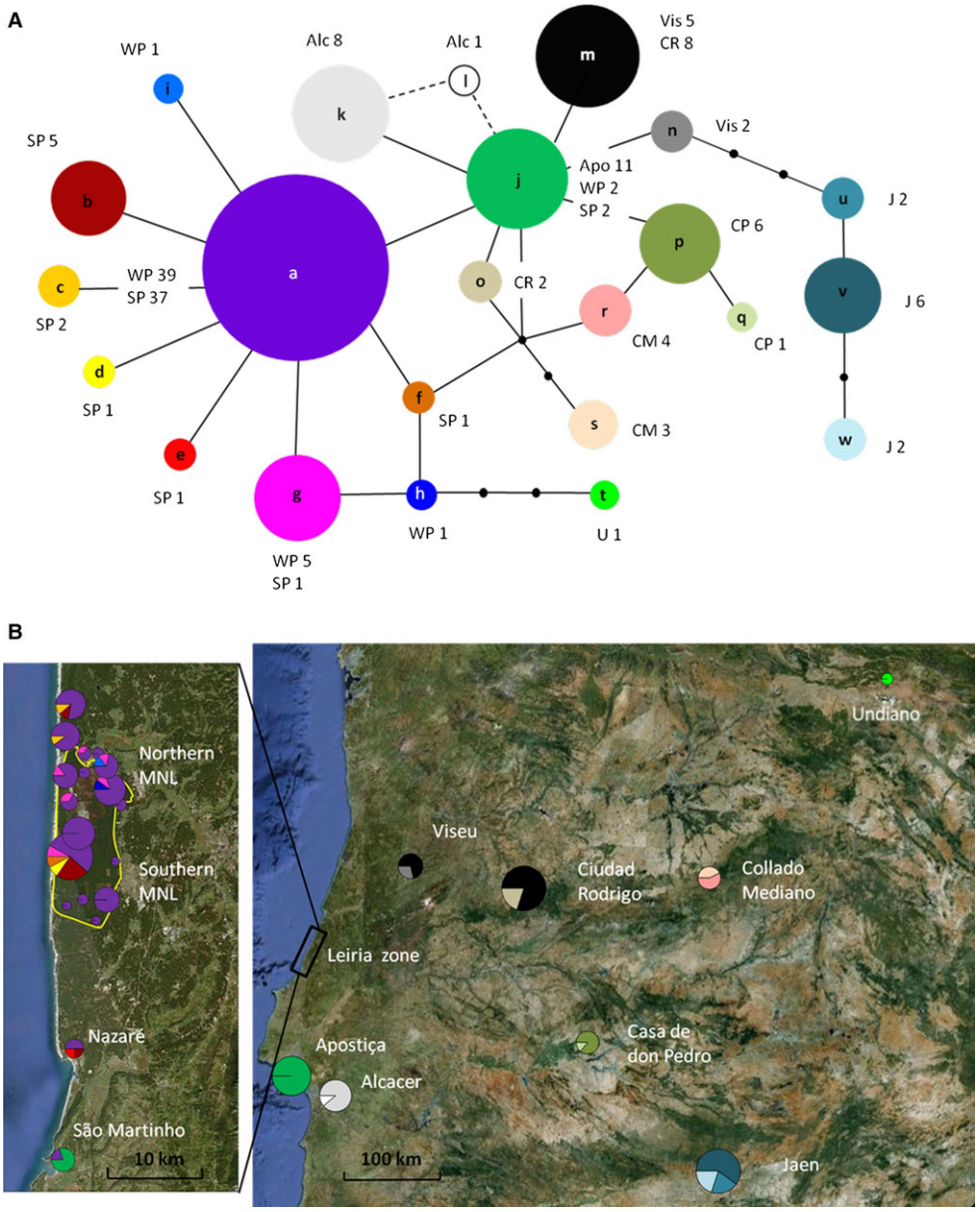

Figure 1. Mitochondrial haplotype network (A) and geographic distribution of the COI haplotypes (B). In the network, each haplotype is denominated by a letter. The corresponding number of individuals per population is shown; black dots correspond to missing haplotypes.

steps in the Markov chain, and linkage disequilibrium (LD) between all pairs of loci was assessed with 10 000 permutations. The occurrence of null alleles and pairwise *Fst* were estimated using FreeNA, with and without ENA correction (Chapuis & Estoup, 2007).

A principal component analysis (PCA) was first performed to explore the population genetic structure of the whole microsatellite dataset, using the adegenet 1.4-2 package (Jombart, 2008) implemented in R (R Core Team, 2015). Individual assignment to genetic clusters was realised using the Bayesian inference method implemented in STRUCTURE 2.3.4 (Pritchard, Stephens & Donnely, 2000). Using the whole microsatellite data set, and for a given number K of genetically homogeneous groups, it computed the posterior probabilities q_k that correspond to the ancestry proportion of each cluster k for each individual genome. For each value of K = 1–10, 10 independent runs were performed with 100 000 burn-in steps followed by 100 000 Markov Chain Monte Carlo (MCMC) simulation steps, inferring the

Table 1. Population code, sampling locality, phenology, developmental stage, years of collection and number of individuals used for microsatellite genotyping and COI sequencing

Population	Locality	Phenology	Stage	Years	*N* μsat	*N* CoI
SP	Leiria zone	Summer	Larvae	2003–2007	89	10
SP	Leiria zone	Summer	Adult males	2007–2010	55	40
SP	Leiria zone	Summer	Adult females	2008	19	–
WP	Leiria zone	Winter	Larvae	2002–2005	81	8
WP	Leiria zone	Winter	Adult males	2007–2010	51	40
WP	Leiria zone	Winter	Adult females	2008	27	–
Apo	Apostiça	Winter	Adult females	2008	17	–
Apo	Apostiça	Winter	Larvae	2003–2004	–	11
Alc	Alcácer	Winter	Larvae	2002	18	9
Vis	Viseu	Winter	Larvae	2002	24	7
CR	Ciudad Rodrigo	Winter	Larvae	2009	–	10
CM	Collado Mediano	Winter	Larvae	2003–2004	–	7
CP	Casa de don Pedro	Winter	Larvae	2009	–	7
J	Jaen	Winter	Larvae	2003–2004	–	10
U	Undiano	Winter	Larvae	2004	–	1

admixture prior alpha from an initial value of one and using a correlated allele frequency model with parameter lambda set to 1. Clustering solutions were evaluated examining the curve of Log P(X|K) and using the ΔK method from Evanno, Regnaut & Goudet (2005) implemented in STRUCTURE HARVESTER (Earl & von Holdt, 2012). We thereby identified the value of *K* that provided the best fit to the observed data.

Hybrid detection in the Leiria zone

Methods used: Recent hybridization between SP and WP in the Leiria zone can be implied for individuals exhibiting mixed assignments in STRUCTURE clustering. We thus tested whether hybridization currently occurs between the sympatric SP and WP through the estimation of the maximum-likelihood hybrid index (hereafter h-index) developed by Buerkle (2005) and implemented in the INTROGRESS R package (Gompert & Buerkle, 2010). It uses allele frequencies of parental reference populations that can be identified from preliminary Bayesian assignment tests. We used as reference a set of simulated parental SP and WP individuals (see below).

We also specifically addressed the question of current introgression using the dedicated software NEWHYBRID (NH, Anderson & Thompson, 2002), which estimates the posterior probabilities (*q*) that an individual falls into six different genotype frequency classes, i.e. two parental (named here pSP and pWP) and four hybrid categories (F1, backcross SP (bcSP), backcross WP (bcWP) and F2) through supervised clustering. The graphic interface allows the burn-in period to be stopped once the sampler ends up in a stable way (as advised in the NH User Guide), and 100 000 new sweeps are then performed to compute posterior probabilities. In all NH analyses, we used Jeffreys-like priors for allele frequencies θ and mixing proportions π. We tested the consistency of the results between ten runs from different seeds, and the use of alternative uniform priors for θ and π. We used three methods to interpret the results and assign the analysed individuals to a single genetic class. Two methods used a threshold on the q-values (either $Tq = 0.9$ or $Tq = 0.5$), which allowed only the individuals for which one genetic class reached the chosen *Tq* to be assigned; we also used a method named 'majority assignment', in which each individual was assigned to the category reaching the highest *q*, which allowed all individuals to be assigned.

Simulated data: The reliability of outputs from assignment methods can be questioned, and depends notably on the type and number of markers, the properties of the studied system such as differentiation or hybridization rate and the quality of the sampling. General recommendations can be found in the literature (Manel, Gaggiotti & Waples, 2005; Vähä & Primmer, 2006), but case-specific evaluations are needed. To estimate the performance of NH and test the reliability of the results obtained with both the h-index and NH, we used simulated genotypes of known ancestry.

We exploited STRUCTURE clustering of the Leiria zone data set alone with $K = 2$ to simulate parental populations that could be considered representative of the populations of interest (Marie, Bernatchez & Garant, 2011). Two procedures were compared.

The first one generated genotypes from the posterior estimates of allele frequencies (with a minimal value set to 0.001) in each cluster assuming HWE and no LD between markers. Partitioning of the information from mixed assigned genotypes relative to their membership to each cluster was thus achieved, so that the complete data set to be taken into account. These parental simulated populations were also used as reference populations to estimate the h-index using the INTROGRESS package. The second strategy consisted in taking into account only individuals with a STRUCTURE q-value > 0.9, thereby eliminating potential hybrids to simulate the parental populations (Coyner, Murphy & Matocq, 2015). From these simulated parental stocks, we simulated offsprings of different categories using Hybrid Lab (Nielsen, Bach & Kotlicki, 2006). The final simulated data sets corresponded to 1000 simulated genotypes for each one of the six genetic classes (pSPsim; pWPsim; F1sim, bcSPsim, bcWPsim and F2sim).

Analyses and interpretation of results: STRUCTURE, INTROGRESS and NH were used for both the 6000 simulated individuals of known ancestry and the Leiria zone data set (321 field individuals). For NH, we compared the known genetic class of each simulated individual to the corresponding assignment result and calculated efficiency, accuracy and overall performances (Vähä & Primmer, 2006) (see Supporting Information, Appendix S2 for definitions and details on calculations). Additionally, to further guide the interpretation of the results and characterise the potential resulting errors, we examined the type of mis-assignments obtained for each class of simulated genotype.

The rarity of hybrids is known to drastically reduce hybrid detection with NH (Pearse *et al.*, 2009; Neaves *et al.*, 2010). Hybrid detection could therefore be different in the Leiria zone data set (in which hybrids are expected to occur at low frequencies) compared to the simulated data set, for which we used a balanced size for all genetic classes. We tested this bias for the F1 genetic class (Supporting Information, Appendix S4). To avoid the bias due to unbalanced sizes of the different genetic classes, we finally performed a new NH analysis of the Leiria zone data set grouped with the simulated data set, using the same parameters (6321 individuals).

To interpret the results and identify signals of recent hybridization events, we compared the NH assignments obtained for the individuals collected in the Leiria zone with those of the simulated genotypes, aiming in particular to understand how mis-assignments might blur the results. We also made a similar evaluation for the ancestry proportions observed in the Leiria zone data set as estimated from STRUCTURE and INTROGRESS, compared with the range of assignments and h-index obtained for each simulated genetic class. Rather than considering and discussing all possible hybrids, we adopted a conservative approach to focus only on undoubtedly hybrid individuals (F1, F2 and backcrosses). To reach this goal, we finally considered only those individuals that did not fall into the assignments or h-index ranges of the simulated parental genotypes.

RESULTS

Population diversity and structure for all sampled individuals

Based on their phenology, individuals from the Leiria zone could *a priori* be separated into two populations (SP or WP). However, the examination of the microsatellite data set by clustering methods showed that seven phenotypically WP individuals were actually assigned to the SP genetic cluster and should thus rather be considered as LateSP individuals (see below). They were thus removed from the WP sample to estimate genetic descriptors at population level. Their individual characteristics (haplotype, localization and phenology) were specifically addressed.

Mitochondrial data

The sequencing of the mitochondrial COI marker allowed for the detection of 29 haplotypes in the whole sampling, due to 18 polymorphic sites (GenBank accession numbers KT768171-KT768193). Haplotypes were separated by a maximum of 13 mutations and revealed a highly structured phylogeographic pattern (Fig. 1). A relatively low intra-population diversity (hs = 0.37) was obtained, in contrast with a very high inter-population diversity (ht = 0.97), thus evidencing a high differentiation between populations, reinforced by a phylogenetic signal in the geographic distribution of the haplotypes ($Gst = 0.61 < Nst = 0.78$; $P < 0.001$). A significant pairwise mitochondrial differentiation between SP and WP (Fst = 0.04; $P < 0.05$) was also detected, yet much smaller than the differentiation between SP and all other sampled populations ($0.62 < Fst < 0.91$).

In the Leiria zone, SP and WP shared three haplotypes (a, g, j) and exhibited respectively five and two private haplotypes. A major haplotype (a) was shared by most individuals from the two phenological populations (37 out of 50 phenotypically SP and 39 out of 48 phenotypically WP individuals, the latter including six males that were considered as LateSP). All remaining haplotypes showed a distance of one mutational step from this major one, except haplotype (h), which was separated by two mutational

Table 2. Sample size (*N*), number of alleles (*Na*), observed (*Ho*) and expected (*He*) heterozygosity, and pairwise *Fst* between Portuguese populations estimated from microsatellites data

					Fst			
Population	*N*	*Na*	*Ho*	*He*	SP	WP	Apo	Alc
SP	163	110	0.522	0.520	–			
WP	151	167	0.585	0.596	0.26	–		
Apo	17	68	0.571	0.593	0.29	0.10	–	
Alc	18	133	0.716	0.730	0.23	0.12	0.08	–
Vis	24	150	0.745	0.745	0.22	0.11	0.10	0.06

steps. Haplotype j was the only one occurring both in the Leiria zone and in another locality, namely Apostiça.

Microsatellite data

Using a 1% cut-off for statistical significance, we observed departure from HWE for MS-*Thpit*13 in WP, and MS-*Thpit*12 in Alcácer. The percentage of null alleles was over 5% in two populations only for MS-*Thpit*08. Six pairs of loci exhibited LD for two populations, and one pair for three populations. Pairwise differentiation estimated with and without ENA correction gave similar results (Table 2). SP was highly differentiated compared with any other population ($0.22 < Fst < 0.30$), while all populations exhibiting the classical winter phenology were less differentiated ($0.06 < Fst < 0.12$). Leiria WP exhibited a higher genetic diversity than SP (Table 2), even though the sampling effort was similar for both sympatric populations. Comparing only these two populations, most alleles present in the SP were also present in the WP: only 20 alleles out of 110 present in the SP were specific to this population, while 77 alleles out of 167 were specific to the WP.

The two-first principal components (PC) of the PCA performed on standardized allele frequencies in the five Portuguese populations are represented in Figure 2. PC1 and PC2 explained respectively 7.2% and 2.9% of the total inertia. PC1 clearly separated most individuals according to their phenology, with the exception of seven WP individuals (six males and one female) projected among SP. Although these individuals were genetically grouped within SP, they emerged in summer rather than in spring, corresponding to LateSP. PC2 separated individuals according to their locality, without any clear geographical trend (Fig. 2). PC3 (2.2%, not shown) corresponded to a north–south gradient.

Individual clustering

The uppermost hierarchical level revealed by STRUCTURE in the whole data set clearly supported $K = 2$ clusters as the best fitting value (Supporting Information, Fig. S1). In all runs, it separated individuals according to their summer or winter phenology (Fig. 3), with the exception of the seven LateSP individuals from the Leiria zone already identified in the PCA. Consistently, these seven individuals, *a priori* identified as WP based on their phenology, were assigned to the SP genetic cluster with a high *q*-value (> 0.99). In each cluster, some individuals had intermediate assignments and in some cases overlapping credibility intervals. For $K = 3$, all runs consistently grouped most individuals from Apostiça

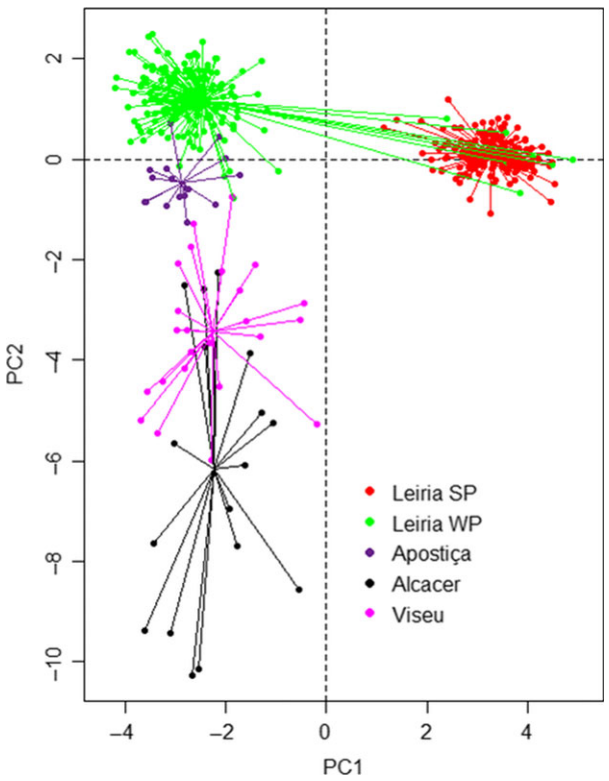

Figure 2. Graph of the first two axes from a principal component analysis of microsatellite genotypes of five Portuguese populations.

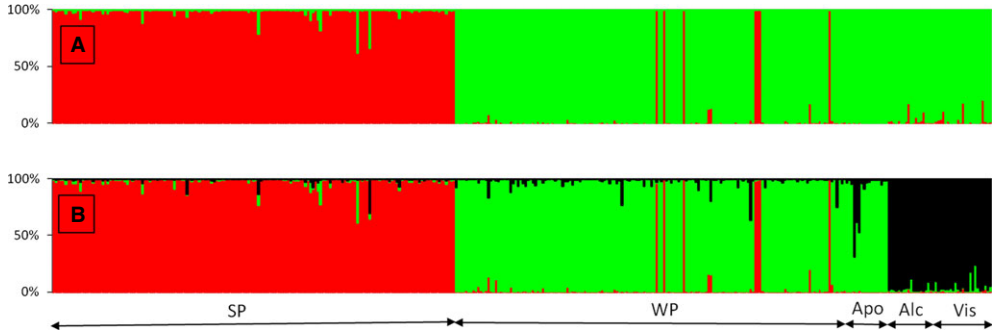

Figure 3. Bayesian clustering of five Portuguese populations inferred by STRUCTURE. Individuals are ordered as in Supporting Information (Table S1). A: individual assignments for $K = 2$; B: individual assignments for $K = 3$.

with WP, while individuals from Alcácer and Viseu formed a separate cluster. For upper values of K, separate runs gave inconstant results, clustering differently individuals from Alcácer, Apostiça and Viseu. Individuals sampled outside the Leiria zone were never assigned to the clusters containing the SP or WP individuals.

Hybrid detection in the Leiria zone

Analyses of simulated genotypes

STRUCTURE assignments and h-index were not fully discriminatory as the range of values obtained for each hybrid category overlapped with those of parental categories (Supporting Information, Appendix S1). NH assignments were scrutinized using the three described strategies (thresholds $Tq = 0.9$, $Tq = 0.5$ and majority assignment) to calculate accuracy, efficiency and overall performance (Table 3). Results were similar when using both methods to simulate the parental populations (Supporting Information, Appendix S2). Accuracy was above 90% for each genetic class at $Tq = 0.9$ but decreased around 80% for hybrid categories (F1sim, F2sim, bcSPsim and bcWPsim) at $Tq = 0.5$. Conversely, at $Tq = 0.9$, efficiency was under 80% for parental classes and drastically decreased under 30% for hybrids. At $Tq = 0.5$, efficiency was above 90% for parental classes but decreased below 80% for hybrids, the lowest score being obtained for F2sim. Efficiency and accuracy values obtained using majority assignment were in the same magnitude to those obtained at $Tq = 0.5$. This threshold allowed all individuals to be assigned, while 5% of individuals were not assigned to any genetic class for $Tq = 0.5$, and 59% for $Tq = 0.9$.

We also examined the rates and types of mis-assignments obtained with the majority assignment method (Table 4). Simulated parental individuals were mis-assigned mainly to the closest backcross category (5.7% pSPsim assigned to bcSP; 7.5% pWPsim to bcWP and 0.1% to F2). Simulated backcross individuals were never mis-assigned to the opposite parental category (bcSPsim never assigned to pWP, and bcWPsim never assigned to pSP). F1sim were never mis-assigned to parental categories, while F2sim exhibited all types of mis-assignment, but were only rarely classified in parental categories (0.3%). To summarise, these results show that individuals assigned as F1 or F2 by NH cannot correspond to a parental SP or WP individual, while individuals assigned as parental pSP or pWP can only be grouped into the corresponding parental or backcross category. Conversely, the exact category cannot be determined with certainty, as each category of the simulated genotype can be assigned to at

Table 3. Efficiency, accuracy, and overall performance of NH to detect parental and hybrid genetic classes across two threshold values (Tq) and majority assignment using simulated genotypes

	Majority assignment			$Tq = 0.5$			$Tq = 0.9$		
Class	Eff.	Acc.	Perf.	Eff.	Acc.	Perf.	Eff.	Acc.	Perf.
pSPsim	0.943	0.938	0.885	0.943	0.940	0.887	0.775	0.981	0.760
pWPsim	0.924	0.917	0.847	0.921	0.917	0.845	0.681	0.981	0.668
F1sim	0.845	0.767	0.648	0.787	0.801	0.631	0.277	0.942	0.261
F2sim	0.636	0.734	0.467	0.599	0.772	0.462	0.212	0.977	0.207
bcSPsim	0.77	0.758	0.584	0.752	0.778	0.585	0.28	0.921	0.258
bcWPsim	0.768	0.766	0.589	0.751	0.786	0.591	0.144	0.941	0.136

Table 4. Proportion of simulated genotypes (lines) from each genetic class assigned with NH to each class (columns) using the majority assignment method

Class	pSP	pWP	F1	F2	bcSP	bcWP
pSPsim	0.943	0	0	0	0.057	0
pWPsim	0	0.924	0	0.001	0	0.075
F1sim	0	0	0.845	0.057	0.057	0.041
F2sim	0.002	0.001	0.114	0.636	0.131	0.116
bcSPsim	0.060	0	0.078	0.090	0.770	0.002
bcWPsim	0	0.083	0.065	0.083	0.001	0.768

least two NH classes. The use of alternative uniform priors for θ and π had little influence on the assignments and mis-assignment scores (not shown).

Analyses of the Leiria zone data set

When the Leiria zone data set was analysed with STRUCTURE and INTROGRESS two individuals, namely 124SP and 129SP exhibited both assignments and h-index that were not included into the range of values obtained for the simulated parental categories (Supporting Information, Appendix S1).

When using NH on this data set only, most individuals exhibited majority assignments into parental categories except for seven SP individuals assigned as bcSP (Supporting Information, Appendix S3). Yet, since the rarity of hybrids affect their identification using NH (see Supporting Information, Appendix S4), the 6000 simulated genotypes were added to the Leiria zone data set to assign field individuals to the different genetic classes with a similar performance as the simulated data set.

We checked that accuracy, efficiency and overall performance of NH analyses of the simulated individuals were not modified (i.e., changes remained below 1%) when the simulated data set was merged with the Leiria zone data set. The majority of assignments for all individuals from the Leiria zone was the same in ten runs obtained from different seeds. NH analysis results were consistent with assignments from STRUCTURE (with $K = 2$) when comparing phenological identification to genetic clustering of the field samples (Fig. 4). Here, 147 SP were predominantly assigned to the pSP genetic cluster, 13 were assigned as bcSP, two as F1 and one as F2. Among the WP, 139 individuals were assigned as pWP, 12 as bcWP and seven as pSP. These seven individuals, assigned to the pSP with q-values over 0.9, except for one (q-value = 0.85), corresponded to the LateSP individuals already identified using STRUCTURE, INTROGRESS and PCA.

It is interesting to note that the patterns of NH assignments of the Leiria zone data set were similar to the ones observed for parental simulated genotypes (except for LateSP). In particular, the proportion of individuals predominantly assigned as backcrosses were similar in both cases: 8% of the SP were assigned as bcSP vs. 5.7% for the simulated individuals pSPsim; and 7.9% of the WP (after excluding the LateSP individuals) were assigned as bcWP vs. 7.5% for the simulated pWPsim (Fig. 5). It is therefore possible that most individuals assigned predominantly as backcrosses actually corresponded to parental individuals. Therefore, the inaccuracy of the methods did not allow the hybrid class of the 13 SP and 12 WP individuals assigned as backcrosses to be unambiguously assessed.

However, three individuals exhibited assignments that did not fall into the assignment ranges of the simulated parental categories as illustrated in Figure 5. It was notably the case for the two SP already recognized as hybrids from STRUCTURE and INTROGRESS: '124SP' (assigned as F1 using NH with 0.65 posterior probability while the maximum value reached among pSPsim was 0.02) and '129SP' (assigned as F2 with 0.52 posterior probability, whereas the highest assignment as F2 among pSPsim was 0.2). Moreover, another individual, namely '109SP' was assigned as F1 using NH with 0.53 posterior probability.

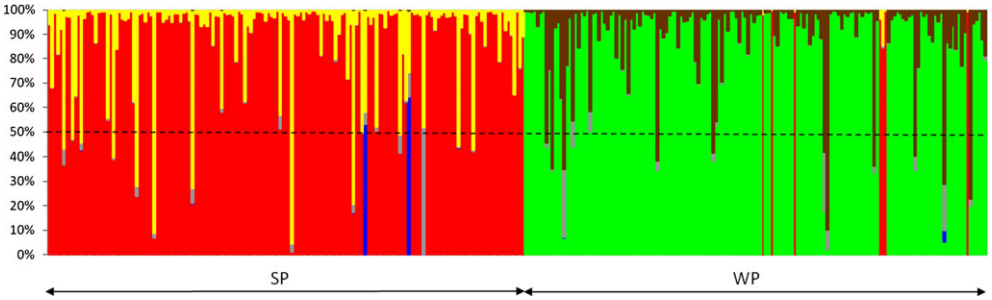

Figure 4. NH assignments of the 321 individuals from the Leiria zone, analysed together with 6000 simulated genotypes (not represented in the figure). Individuals are ordered as in Supporting Information (Table S1). Red: pSP, green: pWP, blue: F1, grey F2, yellow bcSP, brown bcWP.

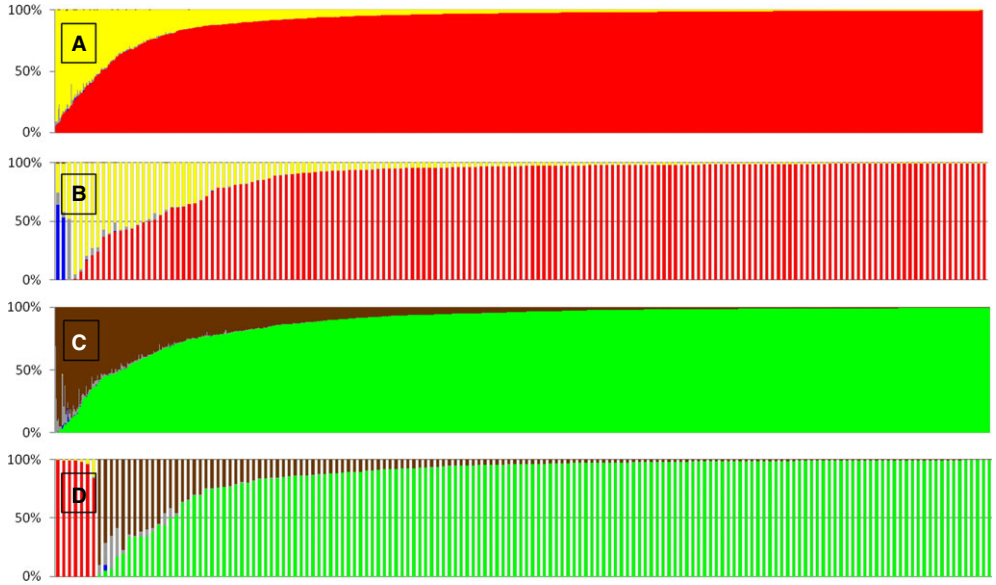

Figure 5. Graphical comparison of NH assignments of parental simulated genotypes (pSPsim and pWPsim) and Leiria zone data set (SP and WP). Simulated and Leiria zone data sets were analysed together (N = 6321). Individuals are sorted by increasing parental assignment. A: 1000 pSPsim; B: 163 SP; C: 1000 pWPsim; D:158 WP. Colours as in Figure 4.

Given the range of mis-assignments obtained from the simulated data set for the F1 and F2 categories (Table 4), these three individuals probably do not correspond to a parental category and can be considered as actual hybrids. To investigate their genetic class into more details, we compared their NH assignments to the range of posterior probabilities estimated for simulated hybrids in their own genetic classes (Supporting Information, Appendix S5). This tended to favour F1 assignment for 124SP and F2 for 129SP, while 109SP could be considered F1 or bcSP. However, for this latter individual, h-index fell in the extreme limits of F1sim, STRUCTURE assignment was out of the range of F1sim (Supporting Information, Appendix S1), but it was fully in the range in bcSP for both analyses.

To sum up, even if some uncertainty remains concerning their precise category, among the 28 individuals predominantly assigned to hybrid genetic classes by NH, three SP males were considered as recent hybrids, while all others could also be mis-assigned as pSP or pWP. The three identified hybrids therefore presumably corresponded to a F1, a F2, and a bcSP (Table 5).

Spatial and temporal distribution and genetic assignments

Male trapping allowed for the obtention of data both on geo-location and precise phenology of adult emergence. Surveys conducted from 2007 to 2010 consistently confirmed the presence of a *c.* 2 week gap in mid July, between the emergence of SP and WP individuals in the Leiria zone. However, this interval had a variable duration between years, lasting for at least 1 week in 2009 and 4 weeks in 2010, for a comparable sampling effort. In total, 839 SP males, but only 54 WP, were trapped in the Leiria zone between 2007 and 2010, even if all traps were active during the whole period of flight for both populations. SP individuals were particularly abundant in Nazaré and in Southern Leiria, while the number of individuals caught was lower both southward (São Martinho) and northward (Northern Leiria) (H. Santos, pers. observ.). No WP males were trapped in Nazaré and very few males in São Martinho (as low as one male per year in each trap), most of the individuals being caught within the MNL.

All LateSP corresponded to adult samples and were never detected as larvae. Among WP males, 11.5% were assigned as LateSP, i.e. three individuals in 2007 and three in 2010. It was also the case for one female emerged in the laboratory in 2009. The six LateSP males were all trapped in the southern MNL, representing 50% of the WP caught there (Fig. 6A). Their emergence occurred from the beginning to the end of the winter phenology, half of them actually flying after all other WP (Fig. 6B).

From the three individuals identified as hybrids, two males represented one-third of the sampling in São Martinho, the southernmost locality in the SP range, one of which was identified as F1 (124SP) and

Table 5. Individual code, geographical location, developmental stage, date of collection, phenology and COI haplotype of the three adult males considered as recent hybrids

Code	Location	Date	Phenology	COI	Genetic class: STRUCTURE/h index/NH
109SP	NMNL	27/05/2008	Summer	a	**bcSP**: 0.776/0.778/0.53F1 – 0.42bcSP – 0.05 F2
124SP	São Marthino	16/07/2009	Median	j	**F1**: 0.615/0.632/0.65 F1 – 0.25 bcSP – 0.10 F2
129SP	São Marthino	24/06/2009	Summer	j	**F2**: 0.681/0.748/0.48 bcSP – 0.52 F2

STRUCTURE assignment corresponds to assignment into the SP genetic cluster and h-index estimates the proportion of genome inherited from SP. NH assignments could fall into the six genetic class studied (only assignments > 1% are shown). The genetic class (in bold) is the most plausible by comparison with the results obtained for each simulated populations taking into account the three methods (Supporting Information, Appendix S4).

the other one as F2 (129SP). Interestingly, these two individuals were the only ones among SP to exhibit the mitochondrial haplotype (j), mostly found in Apostiça. No hybrids were found either in Nazaré or in Southern MNL. The third hybrid male (109SP) was trapped in the Northern MNL (Fig. 6A) and considered as bcSP.

Individual 124SP, here identified as F1, emerged at the end of the SP flight period and was alone responsible for the reduced flight gap between the SP and WP, observed in 2009. The two other hybrids emerged within the flight period of SP (Fig. 6B). Individuals assigned as backcross using NH majority assignment (which could correspond to mis-assigned parental categories) were found in all stages, and were present in all parts of the Leiria zone (not identified in Fig. 6; individual data provided in Supporting Information, Table S1). Among them, males emerged in the same period as their respective parental category.

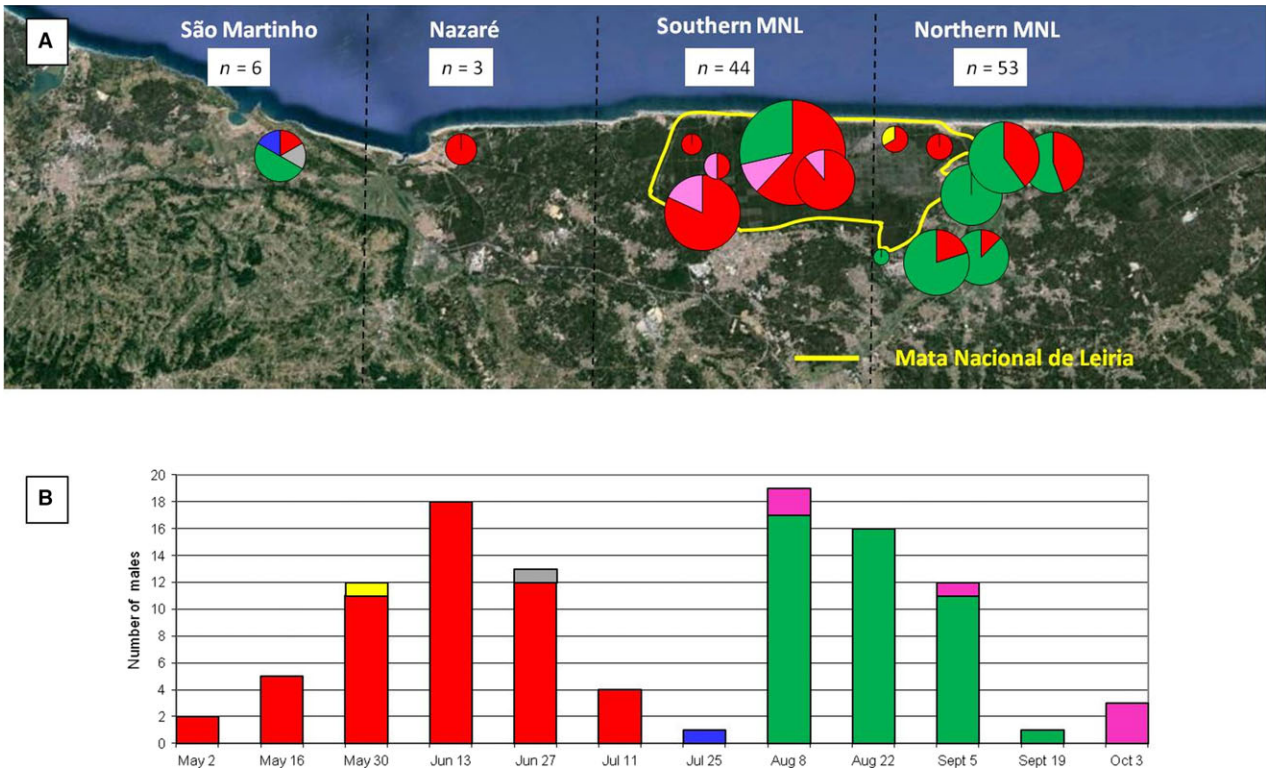

Figure 6. Genetic assignments of 106 males trapped in the Leiria zone. Only the recent hybrids ascertained from STRUCTURE, INTROGRESS and NH analyses appeared in hybrid categories. Red: pSP, Pink: pSP (for LateSP), green: pWP, blue: F1, grey F2, yellow bcSP. (A) Cartography. (B) Temporal distribution: due to slight differences in the days of collection between years and localities, the dates indicate the last possible day of trapping when grouping all years and localities.

DISCUSSION

Genetic structure

PPM populations are known to exhibit a high mtDNA geographical structure, in relation to the low mobility of the females (Kerdelhué *et al.*, 2009). Yet, mitochondrial data revealed here a very low mitochondrial differentiation between SP and WP, most individuals from the Leiria zone sharing the same private haplotype (haplotype a; see Fig. 1). This finding could be explained by a very local origin of the SP, from the local WP population (i.e., the SP being founded by individuals having WP haplotypes), or to a prime role of hybridization when the SP arose, causing mitochondrial introgression. All SP-specific rare haplotypes were distant from only one mutational step from this common haplotype. Such a star-like haplotype network is commonly observed for recently expanding populations (Avise, 2000; Petit, Hampe & Cheddadi, 2005). The likely recent origin of the SP probably explains why highly differentiated haplotypes did not appear by mutation in the diverging populations. Apart from the frequent haplotype a, only two rare haplotypes were shared between SP and WP. The epidemic demography of PPM, and the eventual decrease of local WP in relation to SP outbreaks, could explain the loss of some haplotypes in this population.

Conversely, based on microsatellite data, we observed a relatively low differentiation between all distant WPs ($0.06 < Fst < 0.12$), suggesting some nuclear gene flow at this spatial scale. Flight capacities of males, recently determined in flight mill experiments, demonstrated that they were better dispersers than females (Battisti *et al.*, 2015) and thus could be responsible for migration between WPs. Sampling PPM over a larger spatial scale comprising Portugal and Spain would improve the description and understanding of both the mitochondrial and microsatellite geographic structure of WPs. In particular, it could help to clarify whether the SP-specific alleles and haplotypes could correspond to rare variants also occurring in the WP but not sampled in the present study. Microsatellite data also showed that SP was highly differentiated from all winter populations, including the sympatric Leiria WP, which is consistent with the conclusions of previous studies using fewer markers (Santos *et al.*, 2007, 2011a). This result could be due to divergent demographic histories of the two populations in the recent past, after reproductive isolation was effective, causing genetic drift and distorted allelic frequencies. The bimodal period of flight recurrently observed indicates that phenology could be a main factor responsible for this reproductive isolation. However, it could be modulated by two kinds of remarkable individuals identified in the present study, namely the LateSP and recent hybrids.

LateSP individuals

LateSP individuals correspond to phenologically winter samples that were strongly assigned to the SP genetic cluster using PCA, STRUCTURE, INTROGRESS and NH. Interestingly, all analyses based on simulations showed that pWPsim were never assigned to the SP category, nor to bcSP or F1. The simulation of WP genotypes therefore never produced any 'SP-like' combination of alleles, which means that the LateSP individuals undoubtedly originated from an SP × SP cross and were not a by-product of an incomplete lineage sorting.

The mechanisms responsible for the LateSP phenotype are still unknown and the large variance observed raises questions about the possible origin of this group of individuals. The timing of insect development is a plastic phenotypic trait influenced by environmental conditions (Tauber & Tauber, 1981; Nylin & Gotthard, 1998; Danks, 2006) as exemplified by the timing of PPM adult emergence (Battisti *et al.*, 2015). Most LateSP (six out of seven) individuals were detected among the adult males trapped in summer, during the WP period of flight rather than in spring, which is the usual season of flight for SP individuals. No LateSP individuals were found among the sampled WP larvae. Therefore, the LateSP phenotype could be due to either a longer larval development or a longer period of pupal diapause. It is possible that, despite the gap recurrently observed between SP and WP catches, some SP pupae could have been subjected to particular environmental conditions such as site microclimate or burying depth, causing the delayed emergences. Such individuals should be considered as incidentally having a LateSP phenology due to late diapause termination, by comparison to other SP individuals. This mechanism could explain the cases in which genetically determined SP individuals emerge in the early period of WP flight. Conversely, from thousands of individuals reared under laboratory conditions, Branco *et al.* (2016) observed two SP adults that emerged shortly after larval burial in the soil that is without undergoing the obligate winter pupal diapause. These individuals emerged in the laboratory in November of year N rather than in May of year N + 1. Some of the last LateSP adults found in the field and caught as late as the end of September or beginning of October could similarly be individuals lacking pupal diapause. Such a phenotype is expected to be counter-selected, given the very low probability of finding a mate and because potential offspring would hatch in a colder season, possibly

encountering lethal conditions. An in-depth study of the fate and behaviour of LateSP individuals was not yet conducted because their identification preliminary requires thorough molecular analyses.

Seeking evidence of ongoing gene flow

Technical consideration and hybrid identification: We took advantage of the development of dedicated analytical approaches to detect recent gene flow between the sympatric SP and WP individuals. As the performance of these methods can be data set-dependent, results were estimated using different criteria (Vähä & Primmer, 2006). We first analysed a set of simulated genotypes to test the usability of STRUCTURE, INTROGRESS and NH and to understand potential errors in the assignments obtained for the individuals sampled in the field. This procedure helped us to interpret the results obtained for the Leiria zone, and determine whether some individuals could undoubtedly be identified as recent hybrids, based on the type and the rate of mis-assignments observed for the simulated genotypes. In the specific model system studied here, NH proved to be the most powerful approach to identify hybrids. Yet, some simulated parental individuals were always assigned as backcrosses in NH analyses, whatever the assignation criterion used. We therefore may have under-estimated the actual hybrid frequency because all individuals assigned as backcross were considered as potentially belonging to parental categories, even though some of them may correspond to true backcrosses or individuals originating from earlier hybridization events. Furthermore, it should be noted that we only looked for signs of fairly recent gene flow, i.e. F1, F2 or backcrosses, as the methods did not aim to detect introgressed individuals resulting from backcrosses of the umpteenth generation, or hybridization events that would have occurred more than two generations ago.

Nevertheless, the type of mis-assignment observed for each simulated category allowed for some individuals to be identified as undoubtedly resulting from recent hybridization events. We could consider that their assignments were not due to shared ancestral polymorphism. We could thereby identify with NH three individuals for which assignments did not fall in the range of simulated parental genotypes, while it was the case for only two individuals using STRUCTURE assignments and h-index estimates. Accordingly, the mis-assignments observed in the F1 and F2 show that they cannot belong to a parental category, and that they should be considered hybrids. Moreover, comparison with assignment ranges of simulated genotypes gave some insight into their exact genetic class. This approach allowed us to hypothesise the presence of one F1, one F2 and one bcSP individual.

Hybrid spatial and temporal characteristics: SP and WP proved to have similar sexual pheromones (Paiva *et al.*, 2011; Kerdelhué *et al.*, 2015), and produced viable offspring when crossed under experimentally manipulated conditions (Branco *et al.*, 2016). LateSP correspond to SP individuals emerging when the WP individuals reproduce. We could thus have expected the LateSP to be involved in relatively high hybrid production, as they represented < 1% of the SP males trapped, but 9% of the genotyped WP adults. We could therefore expect a correspondingly high proportion of F1 hybrids (16%) resulting from panmictic matings within WP, which is not the case. Actually, all LateSP adults came from southern MNL, where very few WP and none of the detected hybrid individuals were trapped. Moreover, LateSP phenology was quite variable, half of them emerging even later than all WP males (Fig. 6). Therefore, the possibility of crosses presently occurring between LateSP and WP could be limited, due to both spatial and temporal isolation. A reduced fitness of the hybrids could also be hypothesised to explain the rarity of F1 hybrids found in our sampling.

Our results show that the identified hybrids were found in the southern and northern limits of the SP distribution, i.e. in recently colonised regions (Roques *et al.*, 2015). In particular, in the southern locality of São Martinho, hybrids represented one-third of the sampling. Considering the mitochondrial data, this locality was remarkable because four out of five individuals had a mitochondrial haplotype otherwise found in Apostiça. The relative geographical proximity of the two localities could easily explain this pattern for WP individuals, but it is more surprising to find this haplotype (which is not present in the core area of the SP) in two SP individuals in a newly colonised site. As genetic assignments revealed that both were indeed hybrids, it appears that they originated from crosses between colonising SP males and local WP females. No hybrids were found in Nazaré, where the SP exhibited a very high population size. However, relatively few individuals from Nazaré were genotyped compared to the high number of SP nests observed there (H. Santos, pers. observ.). Most of them were larvae, and information about their time of emergence was not available. Genotyping and sampling in this locality should be improved in future studies. More, if hybridization is rare over the SP range nowadays, focus on the recently colonised range would be necessary to measure its local frequency and potential impact on the expansion processes.

The intermediate date of emergence of the individual identified as F1 was noticeable (i.e., after all SP males and before all WP males, see Fig. 6B). This date

is consistent with the intermediate emergence period observed for F1 adults resulting from SP × WP crosses obtained under experimental conditions in the laboratory (Branco *et al.*, 2016). It could imply a reduced possibility for hybrids to find a partner to mate with, thus limiting introgression between the two populations. Such an intermediate phenology implies that an F1 could further mate with either of the two parental populations (late-emerging SP as well as early-emerging WP), the probability of such mating possibly depending on the local frequencies of the two populations. Once an F1 individual reproduced with one of the parental populations, the direction of introgression would be fixed if backcross individuals exhibited phenology of their respective parental population. This is indeed suggested by the observed phenology of all individuals exhibiting backcross NH majority assignments, even though it has to be kept in mind that these individuals could correspond to mis-assigned parental SP or WP.

Whatever the role gene flow played when SP arose, it seems that it is now rare and does not counter-balance the high genetic differentiation between SP and WP maintained by allochrony. Our results suggest that it occurs mainly at the expansion edge of the SP. As males are better dispersers than females, hybridization is more likely to be asymmetric during the colonization process, because colonizing SP males are more susceptible to mate with early-emerging WP females. However, expansion of the SP range was observed only recently, and was certainly linked to the SP outbreak detected in 1997. Possible past hybridization events that could have occurred in the core area of SP were more likely not sex biased. This gene flow could benefit the SP, by introgression of locally adapted alleles as expansion is in progress and by enhancing genetic diversity. This process could be similar to the one observed for invasive populations despite original bottlenecks (Verhoeven *et al.*, 2011). In contrast, hybridization seems to be rare in the core part of the SP range. Interestingly, the size of the WP is apparently low in that particular area. Whether this possible decline is linked to competition with the SP will need to be addressed in future studies. However, as SP is characterised by particular features, such as a higher larval survival rate at elevated temperatures (Santos *et al.*, 2011b), introgression of SP genes into WP could be favoured in the context of global warming, whenever high temperatures will occur in early fall and affect L1 and L2 WP larvae.

Concluding remarks and perspectives

The results obtained confirmed that an allochronic speciation process of the PPM is ongoing in the Leiria area, in Portugal. PPM range has been subjected to intensive survey through international research networks over its whole range (Roques *et al.*, 2015); the singleness of the SP population is therefore remarkable, even if an early spring emergence concerning a very low percentage of adults was documented in a Greek population (Athanassiou *et al.*, 2007). References addressing the relative importance of allopatric vs. sympatric speciation processes often concede that even in a sympatric context, it is generally difficult to exclude the existence of an initial period of divergence occurring under allopatry (Yukilevich, 2014). PPM has a Mediterranean origin, and the Iberian Peninsula probably corresponded to glacial refugia (Kerdelhué *et al.*, 2009; Rousselet *et al.*, 2010). Plausibly, it remained continuously present in the coastal pine forests of Portugal, where it exhibited the classical winter phenology. Therefore, the initial SP differentiation as well as the sequential ongoing processes, must have occurred in sympatry with WP(s).

While most evolutionary studies attempt to deduce the past history and demo-genetic processes from the genetic structure of contemporary populations, the peculiar system studied here enables us to follow in real time the evolution of a newly expanding population. Still, the microsatellite panel used was not powerful enough to draw and compare demo-temporal scenarios at the origin of the SP (e.g. Approximate Bayesian Computation analyses, ABC, were run using notably DIYABC 2 software (Cornuet *et al.*, 2014) but the results were inconclusive – not shown here). Other markers such as pan-genomic SNPs will be necessary to investigate in detail the history of the SP. High heritability in the timing of emergence and intermediate phenology of F1 offspring could be consistent with the hypothesis of a sudden mutational change at the origin of the SP (Branco *et al.*, 2016). More informative genomic markers and transcriptomic data are now required to better characterise the evolutionary scenario and targeted genetic changes and to understand the late phenology of some individuals.

This work demonstrated that peculiar phenomena, such as the presence of LateSP or the density dependent spatial pattern of gene flow between ancestral and emerging differentiated populations, could play an important role in modulating the evolution of the two sympatric PPM populations, WP and SP. Studies adopting a standardized temporal and spatial framework are needed, so that a full integration of the ecological and demographic parameters can be achieved. Incomplete reproductive isolation, at least at the SP geographical margins, must therefore be analysed considering the ecological drivers of this ongoing

speciation process. In particular, the potential expansion of SP and the responses of both SP and WP to global changes could be impacted by gene flow between these two interacting populations.

ACKNOWLEDGEMENTS

This study was financed by Fundação para a Ciência e Tecnologia, FCT-MCES, Portugal, project PTDC/AGR-CFL/73107/2006. It was also partly supported by the French National Agency for Research, through the ANR JCJC GenoPheno project (2010-JCJC-1705-01). Helena Santos received a Ph.D. scholarship from FCT-MCES, reference SFRH/BD/30518/2006. Genotyping and sequencing were performed at the Genomic and Sequencing Facility of Bordeaux (grants from the Conseil Regional d'Aquitaine n°20030304002FA and 20040305003FA and from the European Union, FEDER no. 2003227 and from Investissements d'avenir, Convention attributive d'aide No. ANR-10-EQPX-16-01). We thank I. Pivotto for her participation in the laboratory work for the acquisition of the mitochondrial data, and S. Rocha, who helped with formatting and clarifying the results of pheromone trapping data. We thank four anonymous reviewers for their helpful comments.

REFERENCES

Abbot P, Withgott JH. 2004. Phylogenetic and molecular evidence for allochronic speciation in gall-forming aphids (*Pemphigus*). *Evolution* **58:** 539–553.

Anderson EC, Thompson EA. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160:** 1217–1229.

Athanassiou CG, Kavallieratos NG, Gakis SF, Kyrtsa LA, Mazomenos BE, Gravanis FT. 2007. Influence of trap type, trap colour, and trapping location on the capture of the pine moth, *Thaumetopoea pityocampa*. *Entomologia Experimentalis et Applicata* **122:** 117–123.

Avise JC. 2000. *Phylogeography: the history and formation of species*. Cambridge, MA: Harvard University Press.

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

Battisti A, Avci M, Avtzis DN, Ben Jamaa ML, Berardi L, Berretima W, Branco M, Chakali G, Fels MAEAE, Frérot B, Hódar JA, Ionescu-Malancus I, Ipekdal K, Larsson S, Manole T, Mendel Z, Meurisse N, Mirchev P, Nemer N, Paiva M-R, Pino J, Protasov A, Rahim N, Rousselet J, Santos H, Sauvard D, Schopf A, Simonato M, Yart A, Zamoum M. 2015. Natural history of the processionary moths (*Thaumetopoea* spp.): new insights in relation to climate change. In: Roques A, ed. *Processionary moths and climate change: an update*. Dordrecht: Springer. 15–80.

Berlocher SH, Feder JL. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology* **47:** 773–815.

Branco M, Paiva M-R, Santos H, Burban C, Kerdelhué C. 2016. Experimental evidence for heritable reproductive time in two allochronic populations of pine processionary moth. *Insect Science*, early view, doi:10.1111/1744-7917.12287

Buerkle CA. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* **5:** 684–687.

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

Burban C, Magnoux E, Rousselet J, Kerdelhué C. 2012. Development and characterization of 13 new microsatellite markers in the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae), in Molecular Ecology Resources Primer Development Consortium **A'Hara SW, Amouroux P, Argo EE, Avand-Faghih A, Barat A, Barbieri L, Bert TM, Blatrix R, Blin A, Bouktila D, Broome A, Burban C, Capdevielle-Dulac C, Casse N, Chandra S, Cho KJ, Cottrell JE, Crawford CR, Davis MC, Delatte H, Desneux N, Djieto-Lordon C, Dubois MP, El-Mergawy R, Gallardo-Escarate C, Garcia M, Gardiner MM, Guillemaud T, Haye PA, Hellemans B, Hinrichsen P, Jeon JH, Kerdelhue C, Kharrat I, Kim KH, Kim YY, Kwan YS, Labbe EM, Lahood E, Lee KM, Lee WO, Lee YH, Legoff I, Li H, Lin CP, Liu SS, Liu YG, Long D, Maes GE, Magnoux E, Mahanta PC, Makni H, Makni M, Malausa T, Matura R, McKey D, McMillen-Jackson AL, Mendez MA, Mezghani-Khemakhem M, Michel AP, Paul M, Muriel-Cunha J, Nibouche S, Normand F, Palkovacs EP, Pande V, Parmentier K, Peccoud J, Piats-Check F, Puchulutegui C, Ramos R, Ravest G, Richner H, Robbens J, Rochat D, Rousselet J, Saladin V, Sauve M, Schlei O, Schultz TF, Scobie AR, Segovia NI, Seyoum S, Silvain JF, Tabone E, Van Houdt JKJ, Vandamme SG, Volckaert FAM, Wenburg J, Willis TV, Won YJ, Ye NH, Zhang W, Zhang YX,** Permanent genetic resources added to Molecular Ecology Resources Database 1 August 2011–30 September 2011. *Molecular Ecology Resources* **12:** 185–189.

Burban C, Gautier M, Leblois R, Landes J, Santos H, Paiva MR, Branco M, Kerdelhue C. 2016. Data from: Evidence for rare ongoing hybridisation events between two allochronic populations of the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *Dryad Digital Repository*. doi:10.5061/dryad.6p2v0.

Butlin RK, Debelle A, Kerth C, Snook RR, Beukeboom LW, Castillo Cajas RF, Diao W, Maan ME, Paolucci S, Weissing FJ, de van Zande L, Hoikkala A, Geuverink E, Jennings J, Kankare M, Knott KE, Tyukmaeva VI, Zoumadakis C, Ritchie MG, Barker D, Immonen E, Kirkpatrick M, Noor M, Macias Garcia C, Schmitt T, Schilthuizen M. 2012. What do we need to know about speciation? *Trends in Ecology & Evolution* **27:** 27–39.

Chapuis M-P, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24:** 621–631.

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

Cooley JR, Simon C, Marshall DC, Slon K, Ehrhardt C. 2001. Allochronic speciation, secondary contact, and reproductive character displacement in periodical cicadas (Hemiptera: *Magicicada* spp): genetic, morphological and behavioural evidence. *Molecular Ecology* **10:** 661–671.

Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A. 2014. DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30:** 1187–1189.

Coyner B, Murphy P, Matocq M. 2015. Ongoing hybridization and asymmetric introgression in a narrow zone of contact between *Neotoma fuscipes* and *N. macrotis*. *Biological Journal of the Linnean Society* **115:** 162–172.

Danks HV. 2006. Key themes in the study of seasonal adaptations in insects II. Life-cycle patterns. *Applied Entomology and Zoology* **41:** 1–13.

Devaux C, Lande R. 2008. Incipient allochronic speciation due to non-selective assortative mating by flowering time, mutation and genetic drift. *Proceedings of the Royal Society of London B: Biological Sciences* **275:** 2723–2732.

Earl DA, von Holdt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359–361.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

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

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28:** 342–350.

Filchak KE, Roethele JB, Feder JL. 2000. Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* **407:** 739–742.

Fitter AH, Fitter RSR, Harris ITB, Williamson MH. 1995. Relationships between first flowering date and temperature in the flora of a locality in central England. *Functional Ecology* **9:** 55–60.

Fitzpatrick MJ, Feder E, Rowe L, Sokolowski MB. 2007. Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. *Nature* **447:** 210–212.

Fitzpatrick BM, Fordyce JA, Gavrilets S. 2008. What, if anything, is sympatric speciation? *Journal of Evolutionary Biology* **21:** 1452–1459.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3:** 294–299.

Gautier M. 2014. Using genotyping data to assign markers to their chromosome type and to infer the sex of individuals: a Bayesian model-based classifier. *Molecular Ecology Resources* **14:** 1141–1159.

Gompert Z, Buerkle CA. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10:** 378–384.

Gschloessl B, Vogel H, Burban C, Heckel D, Streiff R, Kerdelhué C. 2014. Comparative analysis of two phenologically divergent populations of the pine processionary moth (*Thaumetopoea pityocampa*) by de novo transcriptome sequencing. *Insect Biochemistry and Molecular Biology* **46:** 31–42.

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

Hendry AP, Day T. 2005. Population structure attributable to reproductive time: isolation by time and adaptation by time. *Molecular Ecology* **14:** 901–916.

Hull DL. 1978. A matter of individuality. *Philosophy of Science* **45:** 335–360.

Jombart T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405.

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.

Kerdelhué C, Battisti A, Burban C, Branco M, Cassel-Lundhagen A, Ipekdal K, Larsson S, Lopez-Vaamonde C, Magnoux E, Mateus E, Mendel Z, Negrisolo E, Paiva M-R, Pivotto I, Rocha S, Ronnås C, Roques A, Rossi J-P, Rousselet J, Salvato P, Santos H, Simonato M, Zane L. 2015. Genetic diversity and structure at different spatial scales in the processionary moths. In: Roques A, ed. *Processionary moths and climate change: an update*. Dordrecht: Springer. 163–226.

Kisel Y, Barraclough TG. 2010. Speciation has a spatial scale that depends on levels of gene flow. *The American Naturalist* **175:** 316–334.

Manel S, Gaggiotti O, Waples RS. 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution* **20:** 136–142.

Marie AD, Bernatchez L, Garant D. 2011. Empirical assessment of software efficiency and accuracy to detect introgression under variable stocking scenarios in brook charr (*Salvelinus fontinalis*). *Conservation Genetics* **12:** 1215–1227.

Mayr E. 1982. *The growth of biological thought: diversity, evolution, and inheritance*. Cambridge, MA: Belknap Press.

Neaves LE, Zenger KR, Cooper DW, Eldridge MDB. 2010. Molecular detection of hybridization between sympatric kangaroo species in south-eastern Australia. *Heredity* **104:** 502–512.

Nielsen EE, Bach LA, Kotlicki P. 2006. Hybridlab (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes* **6:** 971–973.

Nylin S, Gotthard K. 1998. Plasticity in life-history traits. *Annual Review of Entomology* **43:** 63–83.

Paiva M-R, Santos H, Kerdelhué C, Mateus EP, Branco M. 2011. Can climate change drive speciation ? In Silva AM, Reis H, Bortoli D, Carneiro J, Costa MJ, Bezzeghoud M, Salgado R, eds. *Proceedings of the global conference on global warming-2011*. Lisbon, Portugal, 779–783.

Pearse DE, Hayes SA, Bond MH, Hanson CV, Anderson EC, Macfarlane RB, Garza JC. 2009. Over the falls? Rapid evolution of ecotypic differentiation in steelhead/rainbow trout (*Oncorhynchus mykiss*). *Journal of Heredity* **100:** 515–525.

Petit RJ, Hampe A, Cheddadi R. 2005. Climate changes and tree phylogeography in the Mediterranean. *Taxon* **54:** 877–885.

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP. 2010. Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution* **25:** 459–467.

Pimentel C, Calvão T, Santos M, Ferreira C, Neves M, Nilsson J-Å. 2006. Establishment and expansion of a *Thaumetopoea pityocampa* (Den. & Schiff.) (Lep. Notodontidae) population with a shifted life cycle in a production pine forest, Central-Coastal Portugal. *Forest Ecology and Management* **233:** 108–115.

Pons O, Petit RJ. 1996. Measuring and testing genetic differentiation with ordered *versus* unordered alleles. *Genetics* **144:** 1237–1245.

Pritchard JK, Stephens M, Donnely P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

R Core Team. 2015. R: a language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Roques A, Rousselet J, Avci M, Avtzis DN, Basso A, Battisti A, Ben Jamaa M, Bensidi A, Berardi L, Berretima W, Branco M, Chakali G, Çota E, Dautbašic M, Delb H, El Alaoui, El Fels MA, El Mercht S, El Mokhefi M, Forster B, Garcia J, Georgiev G, Glavendekic MM, Goussard F, Halbig P, Henke L, Hernandez R, Hodar JA, Ipekdal K, Jurc M, Klimetzek D, Laparie M, Larsson S, Mateus E, Matoševic D, Meier F, Mendel Z, Meurisse N, Mihajlovic L, Mirchev P, Nasceski S, Nussbaumer C, Paiva M-R, Papazova I, Pino J, Podlesnik J, Poirot J, Protasov A, Rahim N, Sanchez Peña G, Santos H, Sauvard D, Schopf A, Simonato M, Tsankov G, Wagenhoff E, Yart A, Zamora R, Zamoum M, Robinet C. 2015. Climate warming and past and present distribution of the processionary moths (*Thaumetopoea* spp.) in Europe, Asia Minor and North Africa. In: Roques A, ed. *Processionary moths and climate change: an update*. Dordrecht: Springer. 81–162.

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

Rousselet J, Zhao R, Argal D, Simonato M, Battisti A, Roques A, Kerdelhué C. 2010. The role of topography in structuring the demographic history of the pine processionary moth *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *Journal of Biogeography* **37:** 1478–1490.

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 of London Series B* **274:** 935–941.

Santos H, Burban C, Rousselet J, Rossi J-P, Branco M, Kerdelhué C. 2011a. Incipient allochronic speciation in the pine processionary moth *Thaumetopoea pityocampa* (Lepidoptera, Notodontidae). *Journal of Evolutionary Biology* **24:** 146–158.

Santos H, Paiva M-R, Tavares C, Kerdelhué C, Branco M. 2011b. Temperature niche shift observed in a Lepidoptera population under allochronic divergence. *Journal of Evolutionary Biology* **24:** 1897–1905.

Santos H, Paiva M-R, Rocha S, Kerdelhué C, Branco M. 2013. Phenotypic divergence in reproductive traits of a moth population experiencing a phenological shift. *Ecology and Evolution* **3:** 5098–5108.

Servedio MR, Doorn GSV, Kopp M, Frame AM, Nosil P. 2011. Magic traits in speciation: magic but not rare? *Trends in Ecology & Evolution* **26:** 389–397.

Shaw KL, Mullen SP. 2014. Speciation continuum. *Journal of Heredity* **105:** 741–742.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87:** 651–701.

Smadja CM, Butlin RK. 2011. A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology* **20:** 5123–5140.

Sulloway FJ. 1979. Geographic isolation in Darwin's thinking: the vicissitudes of a crucial idea. *Studies in History of Biology* **3:** 23–65.

Tauber CA, Tauber MJ. 1981. Insect seasonal cycles: genetics and evolution. *Annual Review of Ecology and Systematics* **12:** 281–308.

Vähä J-P, Primmer CR. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology* **15:** 63–72.

Verhoeven KJF, Macel M, Wolfe LM, Biere A. 2011. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proceedings of the Royal Society of London B: Biological Sciences* **278:** 2–8.

Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London B* **367:** 451–460.

Whipple AV, Abrahamson WG, Khamiss MA, Heinrich PL, Urian AG, Northridge EM. 2009. Host-race formation: promoted by phenology, constrained by heritability. *Journal of Evolutionary Biology* **22:** 793–804.

Yamamoto S, Sota T. 2009. Incipient allochronic speciation by climatic disruption of the reproductive period. *Proceedings of the Royal Society of London B: Biological Sciences* **276:** 2711–2719.

Yamamoto S, Sota T. 2012. Parallel allochronic divergence in a winter moth due to disruption of reproductive period by winter harshness. *Molecular Ecology* **21:** 174–183.

Yukilevich R. 2014. The rate test of speciation: estimating the likelihood of non-allopatric speciation from reproductive isolation rates in *Drosophila*. *Evolution* **68:** 1150–1162.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. STRUCTURE assignment of Portuguese populations: Log P(X|K) and ΔK.
Table S1. Sampling individual information and NH majority assignment.
Appendix S1. Comparison of STRUCTURE assignments and h-index for the simulated genotypes and the Leiria zone data sets.
Appendix S2. Performances of NH analyses inferred from a simulated data set generated from individuals exhibiting a STRUCTURE assignment over a 0.9 threshold.
Appendix S3. NH assignments of the Leiria zone data set analysed alone.
Appendix S4. Influence of the rarity of hybrids on NH assignments.
Appendix S5. Comparison of the NH assignments of hybrid individuals with the range of assignments obtained for hybrid simulated genotypes.

SHARED DATA

Data available from the Dryad Digital Repository: doi:10.5061/dryad.6p2v0 (Burban *et al.*, 2016).