

ORIGINAL ARTICLE

Multiple genetic divergences and population expansions of a Mediterranean sandfly, *Phlebotomus ariasi*, in Europe during the Pleistocene glacial cycles

SS Mahamdallie¹, B Pesson² and PD Ready¹¹Department of Entomology, Natural History Museum, London, UK and ²Faculté de Pharmacie, Université de Strasbourg, Strasbourg, France

Phlebotomus ariasi is one of the two sandflies transmitting the causative agent of zoonotic leishmaniasis, *Leishmania infantum*, in France and Iberia, and provides a rare case study of the postglacial re-colonization of France by a Mediterranean species. Four DNA sequences were analysed—mitochondrial cytochrome b (cyt b), nuclear elongation factor-1 α (EF-1 α) and two anonymous nuclear loci—for 14–15 French populations and single populations from northeast Spain, northwest Spain, Portugal and Morocco. The presence of cryptic sibling species was not revealed by phylogenetic analyses and testing for reproductive isolation between sympatric populations defined by the two most divergent cyt b haplogroups. No locus was shown to be under positive directional or balancing selection and, there-

fore, molecular variation was explained demographically. Each nuclear locus showed shallow isolation by distance from Portugal to the French Pyrenees, but for both cyt b and EF-1 α there was then a step change to the upland Massif Central, where leading-edge populations showed low diversity at all loci. Multiple genetic divergences and population expansions were detected by analyses of cyt b and dated to the Pleistocene. Endemicity of one cyt b sub-lineage suggested the presence of a refuge north of the Pyrenees during the last glacial period. Monopolization of the Massif Central by genetically differentiated populations of *P. ariasi* might possibly hinder the northwards spread of leishmaniasis.

Heredity (2011) **106**, 714–726; doi:10.1038/hdy.2010.111; published online 25 August 2010

Keywords: phylogeography; *Phlebotomus ariasi*; genomic DNA; Mediterranean Europe; Pleistocene refugia; Pyrenean barrier

Introduction

Phlebotomus (Larrousius) ariasi Tonnoir is one of the two sandfly species (Diptera, Phlebotominae) transmitting the causative agent of zoonotic leishmaniasis, *Leishmania infantum* Nicolle (Kinetoplastida, Trypanosomatidae), in the Mediterranean bioclimates of Iberia and France (Ready, 2010), and this report investigates whether the opportunities for its populations to spread northwards in France have been constrained by past climate change. In Italy, this sandfly's range includes only the cooler northwest (Maroli *et al.*, 2008), where its low population densities suggest re-colonization from France. From 20th C reports (Rioux and Golvan, 1969), it is far more abundant and widespread in southern France, where the adults are active on hillsides in wooded rural regions during the dry Mediterranean summer. Therefore, *P. ariasi* provides a rare case study of the role of the Pyrenees mountains as a barrier to the postglacial re-colonization of Europe (Hewitt, 1996; Schmitt, 2007) by a Mediterranean species.

Most reports on the four paradigms of postglacial range expansions in Europe (Taberlet *et al.*, 1998; Hewitt,

1999) have focused on the role of the Pyrenees, either for maintaining hybrid (or suture) zones of temperate species or as refugia for montane species (Hewitt, 2004). Glacial refugia for temperate species have been identified north of the Pyrenees and Alps (Schmitt, 2007), which often formed barriers for dispersals from the three Mediterranean peninsulas (Iberia, Italy, Balkans) where most diversity was conserved because of warm 'refugia within refugia' (Gómez and Lunt, 2007). The presence of refugia in southern France might have given *P. ariasi* a springboard for postglacial re-colonization northwards but, alternatively, such refugial populations might have blocked (*sensu* Hewitt, 2004) the dispersal of Spanish populations possibly better adapted to northern environments.

On the basis of isoenzyme variation among populations, *P. ariasi* shows greater local geographical variation than *P. (Larrousius) perniciosus* Newstead (Pesson *et al.*, 2004), the other regional vector of *L. infantum*, making it easier to investigate re-colonization events. Demographic analyses would be confounded by the presence of sibling species or adaptive selection, and so we considered these topics. We selected DNA markers developed for *Larrousius* species, because they identify and relate phylogenetic species—mitochondrial cytochrome b (cyt b), nuclear elongation factor-1 α (EF-1 α) (Esseghir *et al.*, 2000)—or can be sufficiently polymorphic for differentiating populations—cyt b, EF-1 α , and anonymous

Correspondence: Dr PD Ready, Department of Entomology, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

E-mail: P.Ready@nhm.ac.uk

Received 16 March 2010; revised 25 June 2010; accepted 2 July 2010; published online 25 August 2010

nuclear loci AAm20 and AAm24 (Aransay *et al.*, 2003). This report focuses on re-colonization in southwest France, and so *P. ariasi* was sampled with the objectives of testing for environmental barriers (a lowland corridor between eastern Pyrenees and upland Massif Central, France), leading-edge effects (Lot and Rhone valley, France), alternative immigration routes (via northwest or northeast Spain) and potential ancestral sources (Portugal, Morocco).

Materials and methods

Sandfly sampling

In all, 19 rural populations of adult *P. ariasi* were sampled (Figure 1; Supplementary Table A, Electronic Supplementary Information (ESI)), using Centers for Disease Control miniature light traps (Sudia and Chamberland, 1962) set in peri-domestic locations overnight or sticky traps (A4 papers soaked in castor oil) placed in road side walls (Rioux and Golvan, 1969) for four nights. Flies were stored in 80% ethanol at -20°C or dry at -80°C . Species were identified using the morphological characters of the head and genitalia (Gállego Berenguer *et al.*, 1992) and a dissection method that minimizes the risk of carryover to PCR (Testa *et al.*, 2002).

Molecular methods

Genomic DNA was extracted from each sandfly thorax and/or anterior abdomen, and PCR and sequencing were performed according to Testa *et al.* (2002) using the conditions first mentioned below unless stated otherwise.

PCR: A volume of 25 μl (cyt b, EF-1 α) or 20 μl (AAm20, AAm24) reactions were prepared with 1.5U GoTaq Flexi DNA polymerase and 1.5 mM MgCl_2 (Promega UK Ltd, Southampton, Hampshire, UK).

Cyt b: A 796-base pair (bp) fragment was amplified following Parvizi *et al.* (2010), except that the annealing temperature was 51°C .

EF-1 α : An 856-bp fragment was targeted by the primers of Essegir *et al.* (2000). Six allele-specific reverse primers were designed to discriminate genotypes using the PCR amplification of specific alleles (PASA) technique (Sommer *et al.*, 1992) with 1 mM MgCl_2 . The number in the primer name denotes the variable 3' nucleotide (underlined) that conferred specificity: EFRSM-817G 5'CTGAGCGGTAAAGTCAGAG (optimized annealing temperature 61°C), EFRSM-709C 5'ATTGTACACGGGAACGGCC (62°C), EFRSM-643T 5'GAGATTGGCCGGGCGAAT (64°C), EFRSM-631G 5'GGC GAAAGTACGACAGT (62°C), EFRSM-619C 5'GAC AGTTCCTGGCTTACAG (62°C) and EFSMR-496C 5'CA GAATGGCGTCCAGAGCC (62°C).

Anonymous loci: Of the six microsatellite loci developed for *P. perniciosus* (Aransay *et al.*, 2001, 2003), only two (AAm20, AAm24) could be amplified consistently from *P. ariasi*. They were characterized by direct sequencing, because of the insufficient size polymorphism. Two PASA forward primers were designed for locus AAm20: AAm20F-33G 5'AGTTGA GGCTTGCATATCCG (66°C) and AAm20F-51C 5'CCCA

GAGAGCGACGACTC (66°C). Two PASA primers were designed for AAm24: AAm24R-151C 5'TATTCCCGCC CCACTTGGC (66°C , 0.7 mM MgCl_2) paired with forward primer AAm24F1, and AAm24F-79G 5'AGTTC AGCCGTCGACAGCAG (64°C) paired with reverse primer AAm24R2.

Direct sequencing: Cycle sequencing of both strands used conserved or PASA primers and the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Warrington, Cheshire, UK) with one-eighth reactions. Sequences were read using a 3730 capillary sequencer (Applied Biosystems), manually edited in SEQUENCHER v4.6 for Macintosh (Gene Codes Corporation, Ann Arbor, MI, USA), and aligned based on open reading frames (ORFs).

Algorithm for allele inference: Most alleles in heterozygous genotypes were deduced directly (one dimorphic site) or by PASA. However, it was uneconomical to resolve by PASA genotypes present in only 1–3 specimens. Then alleles were inferred, based on estimating the number of alternatives at a single locus (r^n , where $r=2$ because all heterozygous sites were dimorphic, and n = number of sites) and the following steps: (1) For each population, the allele and genotype frequencies were calculated based on alleles read directly (homozygotes) or deduced. (2) These alleles were aligned with each ambiguous sequence to identify allele pairs that could constitute the latter. (3) The likelihood that each genotype formed the ambiguous sequence was then ranked, based on their alleles' frequencies in the population. Often, the selected genotype contained one allele with high local frequency (regional bias) and one unrecorded allele. (4) If alternative genotypes were equally likely, a TCS network was used to identify by statistical parsimony the least derived unrecorded allele.

Phylogenetics

Phylogenies were reconstructed to assess the species status of *P. ariasi* and to identify appropriate outgroups for selection tests. Trees were constructed by Bayesian estimation using MRBAYES v3.1.2 (Ronquist and Huel- senbeck, 2003; <http://cbsuapps.tc.cornell.edu/mrbayes.aspx>). Nucleotide substitution models were selected using the Akaike information criterion approach in MRMODELTEST v2.3 (Nylander, 2004). Each analysis was run for 10 million generations with two parallel searches, using one cold and three heated Markov chains. Trees were sampled from each chain every thousandth generation and the first 5000 samples were discarded as burn-in. All other parameters of the Markov Chain Monte Carlo analysis were default. Mixing and convergence of the two Markov Chain Monte Carlo analysis onto a stationary distribution were assessed in the sump file. TRACER v1.4.1 (Rambaut and Drummond, 2007) was used to plot log likelihood values of the cold chain against generation sample, to determine the stationary-phase's burn-in point. From this, harmonic means were taken to calculate the Bayes factors for selecting the better of two models (Kass and Raftery, 1995). A probability value of 100 was applied when multiple taxa were constrained as the outgroup.

Maximum parsimony analyses (branch-and-bound search) were implemented in PAUP* v4.0b10 (Swofford,

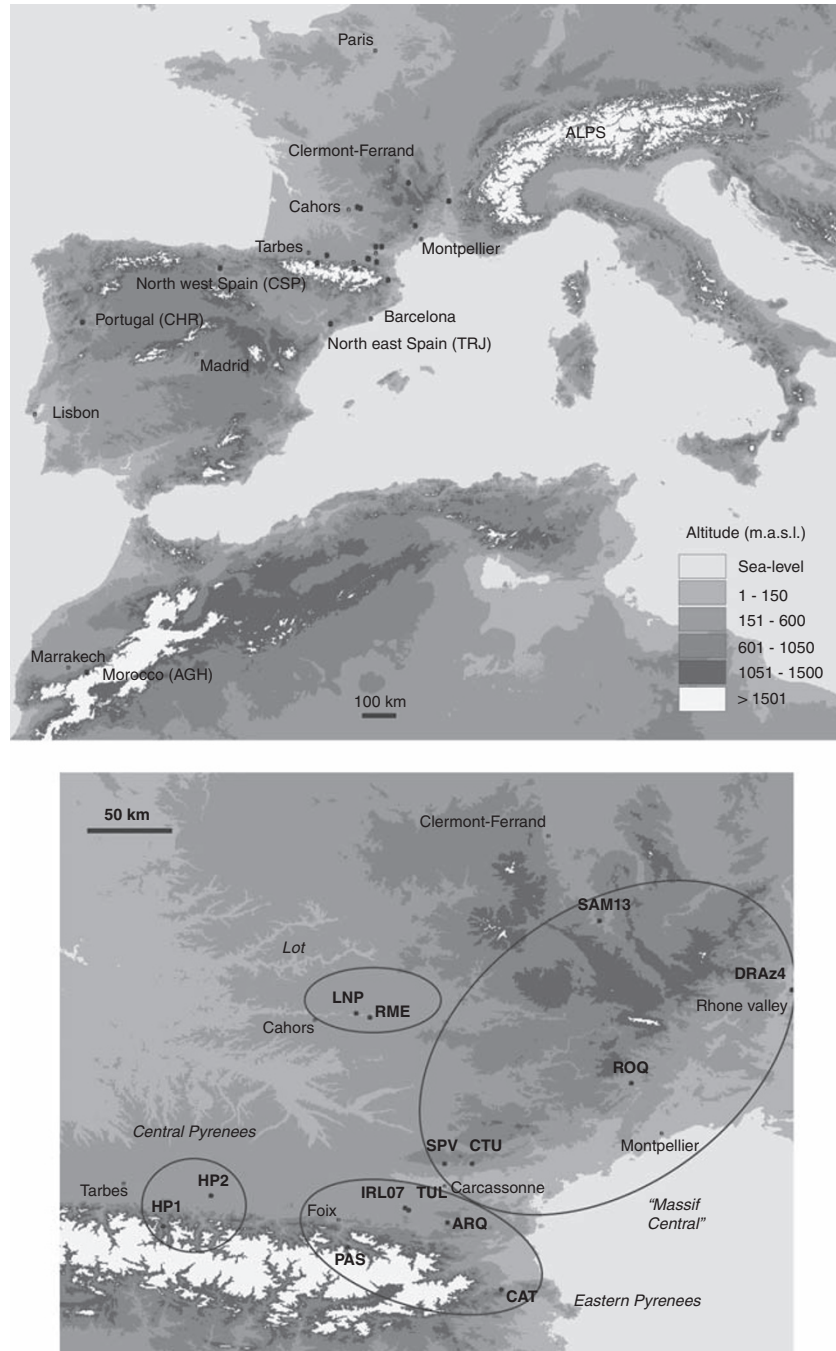


Figure 1 Digital elevation map of the western Mediterranean region, showing locations where 19 populations of *P. ariasi* were sampled for molecular characterization. Ellipses demarcate the labelled regions mentioned in the text. Additional information on locations is given in Supplementary Table A (ESI).

2002) for datasets considered to reconstruct the best Bayesian topologies. Each locus was analysed unweighted and then partitioned by codon positions weighted 2:5:1. Search parameters were default. Statistical support for trees was obtained by re-sampling using 1000 bootstrap replicates.

Genealogical network

Haplotype/allele networks were constructed in TCS v1.21 for Macintosh (Clement *et al.*, 2000), to account for population phenomena. A 95% parsimony connection

limit was set to test whether the haplotypes/alleles of *P. ariasi*, alone or with those of a putative sister species, formed a single network expected of a phylogenetic species (Hart and Sunday, 2007).

Testing for reproductive isolation, panmixia and independent gene assortment

The null hypothesis of random allele associations was investigated by testing for adherence to exact Hardy-Weinberg equilibrium at a single locus (ARLEQUIN v3.11; Excoffier *et al.*, 2005) and for linkage disequili-

brum across multiple unlinked loci (GENEPOP v4.0; Raymond and Rousset, 1995), making no assumptions about evolutionary factors influencing gamete frequencies or phase. For estimating the standard error and the probability of rejecting the null hypothesis, Markov chain parameters included: dememorization = 10 000, batches = 10 000, iterations per batch = 5000.

Testing for selection on molecular markers

The CODEML program of phylogenetic analysis by maximum likelihood PAML v4.2 (Yang, 2007) was run to test if the *P. ariasi* branch had an accelerated rate of amino acid substitution (d_N/d_S or $\omega > 1$) against the background rate in a *Phlebotomus* phylogeny, for which it estimated the ratio of non-synonymous mutations per non-synonymous site (d_N) to synonymous mutations per synonymous site (d_S). The branch lengths of input gene trees were re-estimated under ML (model = 0, NSsites = 0 for nucleotide substitutions per codon) and then used as initial values in further CODEML analyses. Alpha was fixed at a constant rate in the control file.

The MK population test for selection (McDonald and Kreitman, 1991) was implemented in DnaSP v4.90.1 (Rozas et al., 2003). A neutral model predicts that the proportions of non-synonymous to synonymous polymorphisms within a species are linearly related to the proportions of such divergences between it and an outgroup. The significance of departures from neutrality was estimated by a two-tailed Fisher's exact test. The direction and degree of selection was assessed using the neutrality index (Rand and Kann, 1996). Sensitivity depends on the choice of outgroup (Wayne and Simonsen, 1998), which was selected on the basis of d_N and d_S saturation levels estimated by the approximate per site model (Nei and Gojobori, 1986; Jukes-Cantor correction in DnaSP v4.90.1) and a substitution-based method in CODEML.

Departures from the assumption of neutrality, or their demographic alternatives, were also evaluated within *P. ariasi*, using tests not requiring outgroups and based on mutation frequency—Tajima's D (Tajima, 1989; ARLEQUIN v3.11) and R_2 (Ramos-Onsins and Rozas, 2002; DnaSP v4.90.1)—or haplotype distribution data, for Fu's F_s (Fu, 1997; ARLEQUIN v3.11). Statistical significance was estimated using 16 000 or 10 000 simulated replicates in ARLEQUIN and DnaSP, respectively. The minimum number of recombination events (R_m) was estimated in DnaSP v4.90.1.

Descriptive population statistics

The number of segregating sites (S), the number of haplotypes (h), haplotype diversity (H_d) and nucleotide diversity (π) were estimated in ARLEQUIN v3.11, partly to infer population bottlenecks and expansions.

Population structure

Pairwise F_{ST} : Estimates based on allele frequencies at each locus were obtained in FSTAT v2.9.3.2 using the exact test of Weir and Cockerham (Goudet, 2002), which is unaffected by the sampling scheme, with 1000 permutations and a sequential Bonferroni correction to derive significance levels for multiple comparisons. ARLEQUIN v3.11 was used for haploid data. F_{ST} estimates are relatively independent of mutation rate

and selection (Rousset, 1997), and thus reflect genetic drift and gene flow.

Analysis of molecular variance: Analysis of molecular variance in ARLEQUIN v3.11 was used to evaluate the amount of haplotype diversity (based on haplotype frequencies and pairwise genetic distances) correlated with different nested levels of hierarchical population subdivision. Non-parametric procedures (16 000 permutations) tested the significance of the Φ -statistics.

Isolation by distance (IBD): Association was sought between genetic differentiation and geographical proximity of population pairs (Rousset, 1997). F_{ST} values based on allele/haplotype frequencies were converted to $F_{ST}/(1-F_{ST})$ in GENEPOP v4.0 and geographical distances were measured from a digital map using the Distance and Azimuth Matrix v2.1 extension within ARCVIEW v3.2 (Jenness, 2005). Regression and non-parametric significance of association were implemented within the ISOLDE sub-option of GENEPOP v4.0. (1000 permutations for the Mantel test). No assumptions were made on the dispersal dimension.

Dating demographic events

MDIV analysis of cyt b haplogroups: The basic isolation with migration coalescence model in MDIV (Nielsen and Wakely, 2001; <http://cbsuapps.tc.cornell.edu/mdiv.aspx>), implementing likelihood and Bayesian methods, was used to estimate posterior distributions of θ (scaled parameter for nucleotide heterozygosity, $2N_{fe}\mu$), M (scaled migration rate), T (scaled gene divergence time, t/N_{fe}), estimated time to the most recent common ancestor (T_{MRCA}) or gene coalescence time (t_{MRCA}/N_{fe}). A Markov chain length of 3×10^6 was used with a burn-in time of 10%. Model priors included nucleotide substitution according to the finite site model (HKY) to account for multiple hits (Hasegawa et al., 1985). Preliminary M_{max} and T_{max} were varied to select the final optimal priors for each pairwise comparison. A minimum of three replicate Markov chains using different random seeds were run with the optimal priors, to check for convergence and consistency of the parameter estimates. The estimators of T and T_{MRCA} were converted to years for a haploid genome according to Nielsen and Wakely (2001): $t = (T\theta/2\mu)$; and $t_{MRCA} = T_{MRCA}\theta/2\mu$, where μ is the mutation rate for the whole sequence per year per generation.

Mismatch distributions of cyt b haplogroups: These were calculated in ARLEQUIN v3.11, to test for sudden demographic expansion. Pairwise differences among haplotypes are unimodal for recent demographic expansions but multimodal for populations at demographic equilibrium (Rogers and Harpending, 1992). The raggedness index (Harpending, 1994) was used to quantify the smoothness of the observed distribution against the simulation, and statistical support was provided by 10 000 bootstrap replicates under the null model of expansion. Time since the beginning of an expansion event was estimated using the mode of an observed Poisson distribution expressed as the parameter $\tau = 2\mu t$, where μ is the mutation rate per generation for the DNA sequence, and t is the number of generations since the event's start.

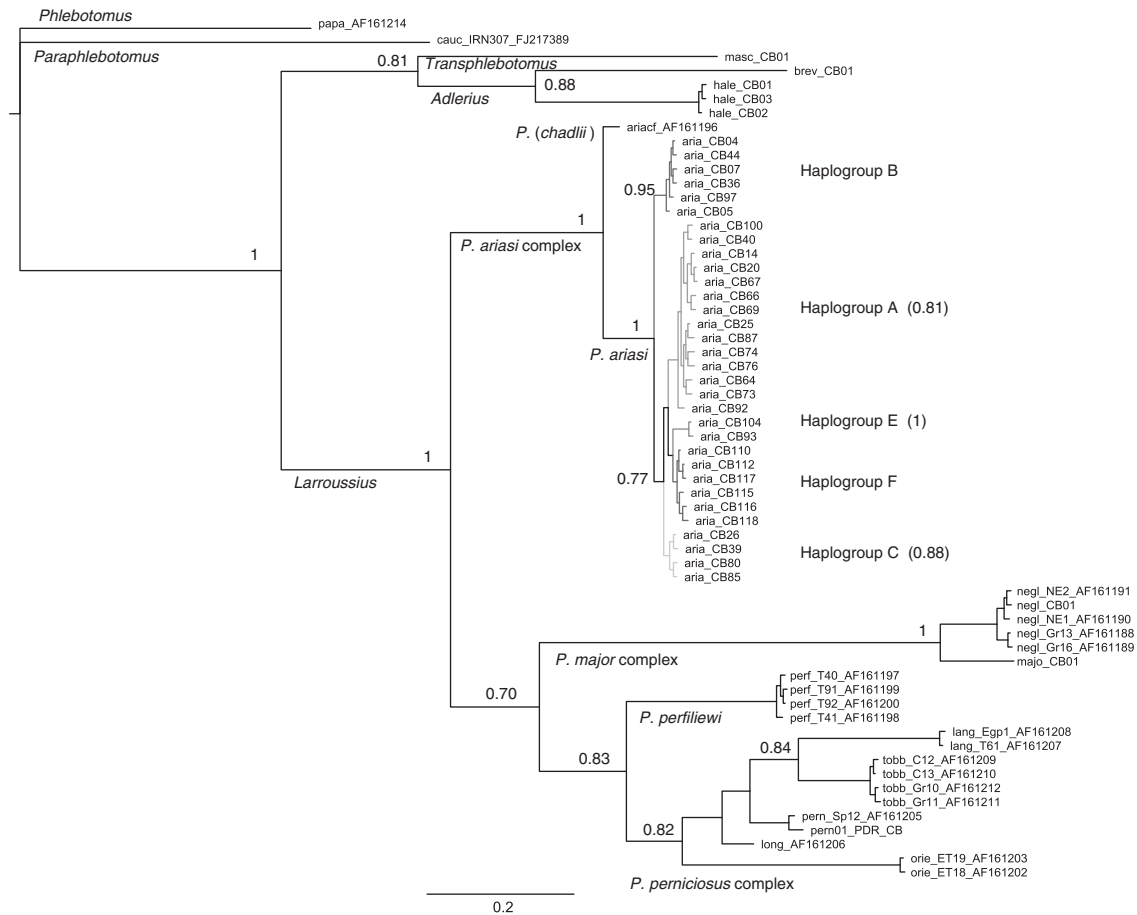


Figure 2 Bayesian phylogeny of the haplotypes of the 3' end of cyt b (714 bp) from *Phlebotomus* species. Branches for subgenera, species complexes, some species, and the haplogroups of *P. ariasi* are labelled. Haplotypes obtained from GenBank include the accession number in their code. Codes for species—papa: *P. papatasi*; cauc: *P. caucasicus*; masc: *P. mascittii*; brev: *P. brevis*; hale: *P. halepensis*; chad: *P. (chadlii)*; aria: *P. ariasi*; negl: *P. neglectus*; majo: *P. major*; perf: *P. perfiliewi*; lang: *P. langeroni*; tobb: *P. tobbi*; pern: *P. perniciosus*; long: *P. longicuspis*; orie: *P. orientalis*. Cyt b was partitioned by codon positions, each with an independent substitution model selected by MRMODELTEST v2.3. Posterior probability values on branches (and to the right of haplogroups A, C, E) indicate the support for nodes (significant for posterior probability > 0.7).

Rates of cyt b sequence divergence

Dating used two rates of pairwise divergence of cyt b: 1% per million years upper limit (Esseghir et al., 2000) and 2.3% per million years lower limit (Brower, 1994). Two generation times for *P. ariasi* were used: one per annum (p.a.) upper limit and three p.a. lower limit (Ready and Croset, 1980).

Results

Mitochondrial cyt b phylogenetics: The cyt b fragment was sequenced for 452 out of 464 specimens of *P. ariasi* from all 19 populations. Pairwise-distance analysis identified 94 haplotypes (Supplementary Table B (ESI)) defined by 86 polymorphic nucleotide sites. Haplotypes were characterized from other sandflies, including sympatric *Phlebotomus* (*Transphlebotomus*) *mascittii* (Supplementary Table A (ESI)). The uninterrupted ORF and the lack of heteroplasmy indicated the absence of pseudogenes (GenBank accessions HQ023281–HQ023380).

With single species of the subgenera *Phlebotomus* and *Paraphlebotomus* as the outgroup, the Bayesian phylogeny of new and GenBank haplotypes (714 bp, no indels) gave strong support (posterior probability = 0.7–1) for: the

monophyly of all taxonomic species; subgenera *Transphlebotomus* and *Adlerius* being sister to subgenus *Larrousius*; and, the latter with a branching order of *P. ariasi* complex, *P. major* complex and others (Figure 2). GenBank accession AF161196 came from a female we refer to as *P. (chadlii)*, because it is probably this species based on location and the large genetic distances between its cyt b haplotype and those of males of *P. ariasi* (Esseghir et al., 2000). Only the males of the two species are morphologically distinct. The haplotype of *P. (chadlii)* was sister to a monophyletic pruned set of haplotypes of *P. ariasi*, which showed a well-supported primary bifurcation of a European haplogroup B from a macro-haplogroup A. The latter contained three well-supported European haplogroups (C, A, E) but a poorly supported branching order and one unsupported Moroccan haplogroup (F). This phylogeny was the statistically favoured model (Bayes factor > 6). There was greater statistical support for: partitioning by the three codon positions independently, rather than not partitioning or partitioning first + second apart from the third position; and, model selection by MRMODELTEST, rather than over-parameterizing with the GTR + I + G model. With or without weighting, maximum parsimony

analyses was concordant with the Bayesian model except for not resolving the *P. perniciosus* complex.

Nuclear gene phylogenetics

The EF-1 α fragment was sequenced for 403 out of 424 specimens of *P. ariasi* from 18 populations (not ROQ). On the basis of 29 polymorphic nucleotide sites, 45 alleles (Supplementary Table C (ESI)) and 65 genotypes were scored (GenBank accessions HQ023381-HQ023425). Only four alleles and seven individuals' genotypes (1.7%) had to be inferred. There were no non-synonymous mutations in a single intronless ORF. The new alleles aligned with the 720-bp alleles of *Larroussius* in GenBank (Esseghir et al., 2000) without introducing indels. The Bayesian phylogeny (Supplementary Figure A (ESI)) gave strong support for: the *P. major* complex as the outgroup; the monophyly of all taxonomic species or complexes; and a branching order of the *P. ariasi* complex, and others. There were no intraspecific lineages of *P. ariasi*. Maximum parsimony analyses gave a congruent tree.

The c. 146-bp fragment of anonymous locus AAm20 was sequenced for 396 out of 398 specimens of *P. ariasi* from 18 populations (not ROQ) and also for *P. mascittii*. Allele size variation was recorded for only 12 *P. ariasi*, and always within or flanking the microsatellite region identified in *P. perniciosus* (Aransay et al., 2001). BLAST searches found only GenBank sequence AJ303377 (*P. perniciosus*) with significant homology to the alleles of *P. ariasi*, and an interspecific alignment gave a 149-bp ORF shared by both species and *P. mascittii*. Bayesian phylogenetic reconstruction supported only the monophyly of *P. ariasi* (posterior probability = 0.82). For all population genetics analyses of *P. ariasi*, the 5' indel region and 3' missing data were deleted from the alignment, and the remaining 90 bp contained 13 polymorphic sites, from which 14 alleles (Supplementary Table C (ESI)) and 19 genotypes were scored (GenBank accessions HQ026001-HQ026018). Only four alleles and five genotypes had to be inferred for seven individuals (1.8%).

The 130-bp fragment of anonymous locus AAm24 was sequenced for all 398 specimens of *P. ariasi* from 18 populations (not ROQ). Allele size variation was not discovered. The aligned sequences contained five polymorphic sites, from which 11 alleles (Supplementary Table C (ESI)) and 21 genotypes were scored (GenBank accessions HQ025989-HQ026000). No alleles and only one individual's genotype (0.3%) had to be inferred. Manual gap insertion permitted the alignment of the alleles of *P. ariasi* with a new one of *P. mascittii* and that of *P. perniciosus* (GenBank accession AJ303378), to produce a 175-bp ORF containing 13 contiguous amino acids that matched 100% the 'Jumonji domain containing 1B' of *Nasonia vitripennis* (GenBank accession LOC100120387), a transcription factor with homologues in mouse and *Drosophila* (Sasai et al., 2007). Phylogenetic reconstruction by maximum parsimony analysis provided strong support (80% bootstrap) for the monophyly of *P. ariasi*.

Genealogical networks and geographical variation

On the basis of a 95% parsimony connection limit, TCS rejected the inclusion of *P. chadlii* in the networks of cyt b (714 bp) and EF-1 α (720 bp) of *P. ariasi*. All four loci

supported the monophyly of *P. ariasi* and the absence of cryptic sibling species. TCS collapsed 452 *P. ariasi* into 92 cyt b haplotypes connected by 11 steps and identified the same haplogroups as the Bayesian phylogeny, with neighbouring haplogroups separated by three to seven mutations (Figure 3). Only haplogroup B had no reticulate loops with any other haplogroup.

Geographical variation (Supplementary Table B and C (ESI)) was mapped on the gene networks of *P. ariasi*. Each moderate-high frequency cyt b haplogroup had one to two modal haplotypes with a shallow radiation of less frequent haplotypes derived from them by one to four mutations (Figure 3), signalling recent expansion from a small or modest number of founders. The distribution of haplogroup F was disjunct, found only in Morocco, where all specimens contained it. Haplogroup A predominated in numbers (80.0% in Europe), haplotype diversity (54), and geographical range. Except in north-west Spain, its modal haplotype (CB25) predominated in each Iberian and French population (69% overall). Haplogroup B was less abundant (15.9% in Europe), absent from Portugal and Morocco, present at low-moderate frequency in northern Spain and the southern Massif Central (4.2–12.5%), but omnipresent at moderate frequency (11.1–48.1%) in the French Pyrenees. Haplogroup C was uncommon (3.4% in Europe) but widespread in Iberia and France, whilst haplogroup E was rare (0.7% in Europe) and found only in the Central Pyrenees and northeast Spain.

Each nuclear network (Supplementary Figure B (ESI)) showed geographical structure, with all alleles from Morocco and most from Portugal being associated with just one of the two modal alleles from Europe. The EF-1 α network had numerous reticulate loops, but those of AAm20 and AAm24 were simpler (0–1 loops).

Near the north of its range, *P. ariasi* had populations lacking genetic diversity, consistent with them being leading edge. For cyt b, haplogroup A or its modal haplotype (CB25) were near to fixation in the Massif Central with Lot, where haplogroups B and C occurred only near the Pyrenees. Each nuclear locus also showed low diversity in the Lot (one to three alleles, none private), and the distinction between the French Pyrenees and the northern Massif Central with Lot was observed for EF-1 α (Predominant alleles 01 and 03, respectively), AAm20 (Allele 02 with frequencies of 22.7–35.4% and 28.3–61.4%, respectively) and AAm24 (Alleles 08 and 09 found only in the Pyrenees).

The unique demographic history of *P. ariasi* in the French Eastern Pyrenees was indicated by the presence of two cyt b sub-haplogroups B with nearly disjunct distributions (Figure 3). Modal haplotype CB04 and its radiating haplotypes were restricted to the Eastern Pyrenees, east from the Ariège valley, where they were omnipresent at low-moderate frequency (6.3–31.8%) and constituted 61.4% of haplogroup B. In contrast, modal haplotype CB05 had a contiguous distribution only in the bordering populations, in the Central Pyrenees and northern Spain.

No reproductive isolation between populations defined by cyt b haplogroups

Two populations of *P. ariasi* in the Eastern Pyrenees (PAS, ARQ) contained sufficient flies with cyt b haplogroups A

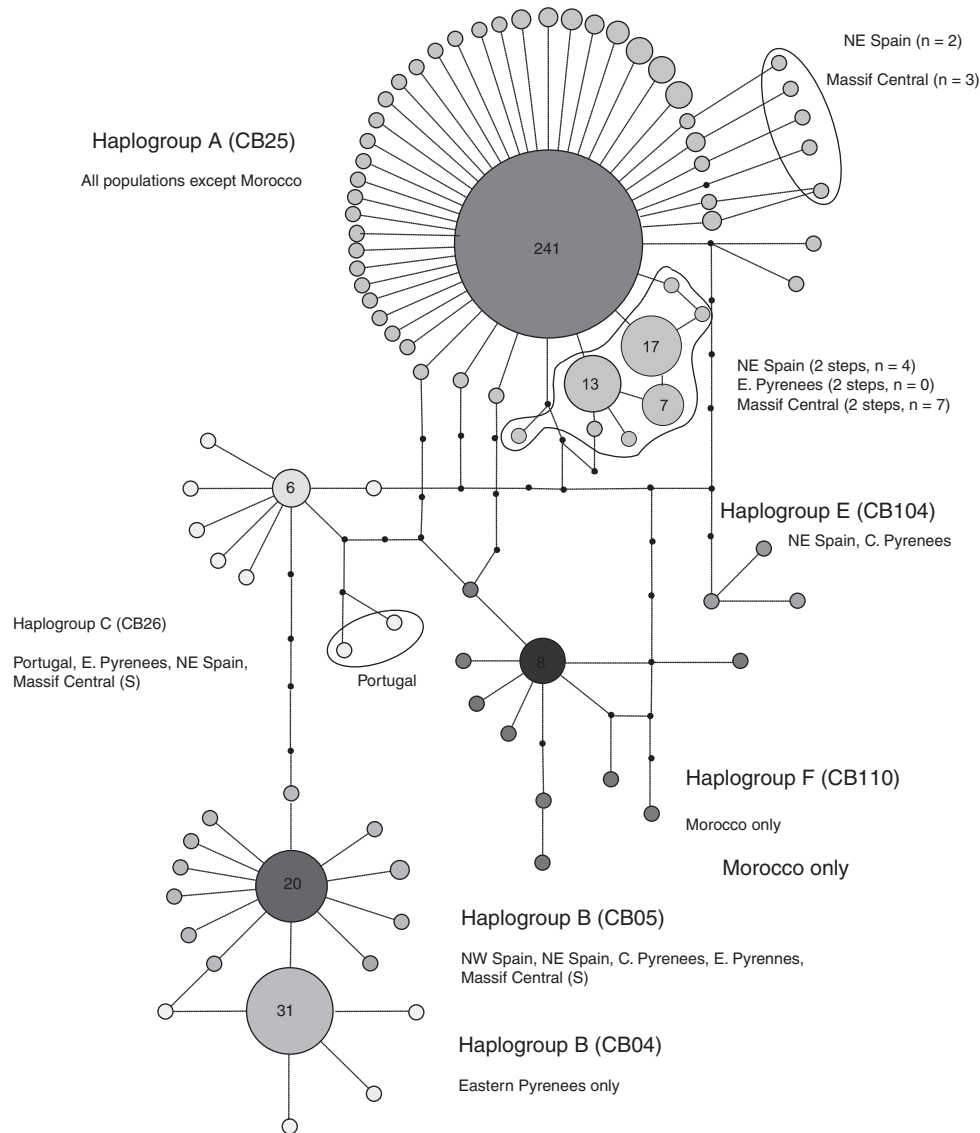


Figure 3 Parsimony network (TCS v1.21) showing the genealogical relationships between the 92 cyt b haplotypes from 452 *P. ariasi*. These haplotypes are shown as shaded or open circles with sizes proportional to their frequency of occurrence, which is given if >5 . Smallest filled circles denote missing haplotypes. The lettered haplogroups or sub-haplogroups (B) are followed by the code of their modal haplotype (CBNN) along with their geographical distributions. All most parsimonious pathways are shown. C, central; E, eastern, NE, northeast; NW, northwest.

or B to test for such reproductive isolation. The population size for ARQ was increased by pooling 2 years' collections, which was considered valid because each nuclear locus remained in Hardy–Weinberg equilibrium. No evidence of biological speciation was found (Supplementary Table D (ESI)). Observed genotype frequencies did not differ from those expected in a single randomly mating population in each location: no significant deviation from Hardy–Weinberg equilibrium was found ($P > 0.05$). No linkage between haplotypes or alleles at different loci (cytonuclear and nuclear–nuclear) was found: pairwise comparisons of linkage disequilibrium showed no significant difference from the null hypothesis of independent haplotype/allele association between loci ($P > 0.05$).

Effectively neutral evolution of loci: population expansions
Positive directional or balancing selection was not detected at any locus using the appropriate tests.

CODEML and MK investigate polymorphism in codons relative to divergence in other species and, therefore, were used only to analyse the coding fragments of cyt b and EF-1 α .

CODEML provided no support for positive directional or diversifying selection along the *P. ariasi* branch compared with the other branches in phylogenies of cyt b (Figure 2) and EF-1 α (Supplementary Figure A (ESI)). Only cyt b showed a significant difference (LRT: $2\Delta l = 4.4$; $\chi^2 df = 1$; $P < 0.05$) in ω on the *P. ariasi* branch (0.0084) compared with the background branches (0.0202), indicating that cyt b is under greater purifying selection in *P. ariasi* ($\omega < 1$). It is often assumed that mitochondrial genes are under strong purifying selection, but this should always be tested (Dowling *et al.*, 2008).

Phylogenetic analysis by maximum likelihood has low power for detecting intra-specific selection (Anisimova *et al.*, 2002). The more sensitive MK test showed no significant departures from neutral expectation for cyt b

and EF-1 α of *P. ariasi* (Supplementary Table E (ESI)). The associated neutrality index indicated that cyt b tended towards purifying selection (not significant), there being few fixed non-synonymous substitutions, but it could not perform well for EF-1 α because there were no fixed differences.

Tajima's *D* is more sensitive than the former tests for detecting recent selection, but significant results can arise from demography (Schmidt and Pool, 2002). No significant *D* value was positive for *P. ariasi*, indicating that balancing selection or some demographic process (population decrease, subdivision or recent bottleneck) was not maintaining some haplotypes/alleles at intermediate frequencies. All significant *D* values were negative, indicating that directional selection (positive or purifying), a selective sweep/hitchhiking or an alternative demographic process (population expansion or older bottleneck) had produced an excess of low-frequency haplotypes/alleles.

Negative values of Fu's F_s also arise from an excess of low-frequency haplotypes/alleles and, compared with *D*, are a more sensitive indicator of recent population expansions (or selective sweep/hitchhiking) (Ramirez-Soriano et al., 2008). For cyt b (Supplementary Table F (ESI)), 8 out of 14 French populations (haplogroup A), or 3 out of 14 (all haplogroups), had expansion signals based on significant negative values for both F_s and *D*. By these criteria, only two populations (one French) were expanding based on EF-1 α , and none was expanding based on AAm20 or AAm24 (Supplementary Table G (ESI)).

With high recombination rates, however, F_s loses power or becomes significant in the absence of population expansion, and then the use of R_2 with *D* can be more informative (Ramirez-Soriano et al., 2008). Considering the longer gene fragments, the estimated minimum number of recombination events (R_m) was low for cyt b (0 for 14 populations, and 1–3 for four populations only when considering all haplogroups), as expected for mitochondrial DNA, and for EF-1 α (0 for nine populations, 1–2 for nine populations). Nevertheless, the significant results for R_2 were compared with those for F_s and *D*. For cyt b (Supplementary Table F (ESI)), R_2 failed to support the expansion indicated by F_s and *D* in 4 out of 10 populations (haplogroup A only), or 2 out of 5 (all haplogroups), and this could be explained by its lower power when recombination and *S* are low or expansions are not sudden (Ramirez-Soriano et al., 2008). Within the eastern Pyrenees, expansion signals were detected only for haplogroup A. The detection of expansion in one population by F_s alone and in two populations by R_2 alone suggests that these statistics need further modelling.

For the three nuclear loci, no population had a significant R_2 , although values approached significance for EF-1 α in the two populations (Central Pyrenees, northeast Spain) with significant and negative values of F_s and *D* (Supplementary Table G (ESI)), making population expansion the most likely explanation. The three significant values given for EF-1 α by F_s alone might have arisen from recombination not population expansion.

Descriptive population diversity of *P. ariasi*

Low cyt b diversity was associated with northern populations (Supplementary Table F (ESI)), likely to be

leading edge: the populations from Lot (LNP, RME) and the Rhone valley (DRAZ4) had significantly lower haplotype diversities (H_d : 0–0.363) than the others (0.541–0.825); and those from the Massif Central, Lot and Rhone valley had lower nucleotide diversities (π : 0.00011–0.00248) than the others (0.00263–0.00839). Populations from the eastern Pyrenees had the highest π values (0.00524–0.00839), because of the diversity of haplogroups (A–C). Population HP2 was an exception to the Pyrenean trend: low H_d (0.314) can be explained by its leading-edge location, on a plain distant from the preferred forested foothills.

EF-1 α was the only nuclear gene (Supplementary Table G (ESI)) to show geographical variation in diversity, from highest in the south to lowest in the north: Morocco, Portugal and northwest Spain (H_d : 0.761–0.881; π : 0.00174–0.0271); northeast Spain, French Pyrenees and southern Massif Central (H_d : 0.427–0.659; π : 0.00060–0.00138); and, northern Massif Central, Lot and Rhone valley (H_d : 0.212–0.552; π : 0.00027–0.00077).

Populations: genetic differentiation and IBD

Many demographic tests require individual populations to be panmictic. Fortunately, all populations of *P. ariasi* were found to be in Hardy–Weinberg equilibrium at each nuclear locus (Supplementary Table G (ESI)), except the possibly mixed population from northeast Spain at EF-1 α .

Very great genetic differentiation between populations (Wright, 1978), caused by drift and limited gene flow, was signalled by very high F_{ST} values (>0.25 , $P < 0.05$) for cyt b haplogroup A (northwest Spain versus most populations except northeast Spain; Lot versus nearby Massif Central), EF-1 α (northern Massif Central or Lot versus most except northwest Spain), AAm20 (Portugal versus all populations; northwest Spain versus 7 out of 13 locations), and AAm24 (Portugal versus all populations except northwest Spain; northwest Spain versus 9 out of 14 French populations; Lot versus many populations). Thus, populations from the north Massif Central and Lot showed a leading-edge effect, with F_{ST} values often being significantly high in comparisons with the French Pyrenees and northeast Spain.

Hierarchical analysis of molecular variance of all cyt b haplotypes supported three regional population subdivisions (Table 1). Among region variance was significant ($P < 0.05$) between eastern Pyrenees, 'Massif Central' (including Lot and Rhone Valley) and Central Pyrenees/northeast Spain. The *a priori* regional groupings were supported by the homogeneity of within-region variances, except for the comparison between eastern Pyrenees and the 'Massif Central'. This can be explained by gene flow between eastern Pyrenees and Southwest Massif Central, because excluding the latter eliminated significant within-region variance.

For cyt b, IBD was not supported by a Mantel test fitting $F_{ST}/(1-F_{ST})$ against geographical distance ($P > 0.05$), using European haplogroup A only to eliminate the effects of multiple haplogroup histories. Noticeably, genetic differentiation did not increase with distance from Portugal. In contrast, IBD was supported by Mantel tests ($P < 0.05$) for each nuclear locus by including flies associated with all cyt b haplogroups from all countries (Figure 4). Some inter-regional comparisons did not

Table 1 Hierarchical analysis of molecular variance tests of cytochrome b haplotype frequencies for four different *a priori* regional divisions of the spatial populations of *P. ariasi*

Division tested	Degrees of freedom	Φ -Statistic	% Variation	P-value
<i>Eastern Pyrenees versus Massif Central versus Central Pyrenees and northeast Spain</i>				
Among regions (Φ_{CT})	2	0.18494	18.49	<0.001*
Within regions (Φ_{SC})	13	0.01829	1.49	0.07218 ± 0.00261
Within pops (Φ_{ST})	371	0.19984	80.02	<0.001*
<i>Eastern Pyrenees versus Massif Central</i>				
Among regions (Φ_{CT})	1	0.25422	25.42	<0.001*
Within regions (Φ_{SC})	11	0.02525	1.88	0.04970 ± 0.00233*
Within pops (Φ_{ST})	306	0.27305	72.69	<0.001*
<i>Eastern Pyrenees versus central Pyrenees and northeast Spain</i>				
Among regions (Φ_{CT})	1	0.08231	8.23	<0.05*
Within regions (Φ_{SC})	7	0.01963	1.8	0.11673 ± 0.00341
Within pops (Φ_{ST})	234	0.10033	89.97	<0.05*
<i>Massif Central versus central Pyrenees and northeast Spain</i>				
Among regions (Φ_{CT})	1	0.09306	9.01	<0.001*
Within regions (Φ_{SC})	8	0.0039	0.38	0.22307 ± 0.00395
Within pops (Φ_{ST})	202	0.93562	90.61	<0.05*

*Significant $P < 0.05$.

support the overall result: for cyt b haplogroup A or EF-1 α , there was a wide range of pairwise genetic distances between the Pyrenees and the 'Massif Central'; and, for AAm24 genetic distances were disproportionately high between the Lot populations and most outside France.

Geographical distances were initially calculated as straight-line measurements. Within the 'Massif Central', IBD was supported for cyt b, and each nuclear locus showed positive but statistically non-significant associations with distance (16–244 km). However, for no locus was there support for IBD within either the French Pyrenees (2.85–202.57 km) or a region containing the eastern Pyrenees and the 'Massif Central'. IBD was also tested using alternative dispersal pathways, but there was no greater statistical support for migrations from Spain to France through the western or eastern coastal foothills (800 m a.s.l.) compared with straight-line crossings of the Pyrenees.

Multiple genetic divergences and population expansions of *P. ariasi* in the Pleistocene

MDIV analyses were used to estimate gene coalescence (t_{MRCA}) and divergence times (t) between all pairs of cyt b haplogroups (Table 2). On the basis of two mutation rates ($\mu = 2.3\%$ or 1%) and a generation time of 1 year, all divergence times (1 254 492–376 757 years ago) were dated to the second half of the Pleistocene epoch (2 588 000–10 000 years ago). The estimates would be reduced by two-thirds for three generations p.a. but, even using $\mu = 2.3\%$, all divergences would still have been well before the last interglacial (Eemian, 125 000–110 000 years ago) (Gibbard and van Kolfschoten, 2004).

Population expansions were investigated by mismatch distribution analyses (Supplementary Figure C (ESI)). Cyt b haplogroups A and F showed unimodal distributions of pairwise nucleotide differences among haplotypes of individual flies and, along with haplogroup C, had distributions not significantly differing from those expected under the null hypothesis of expansion (raggedness index $P > 0.05$). On the basis of τ statistic, the 95% confidence intervals for the estimated dates of

expansion overlapped for populations defined by haplogroup (Table 3). All regional groupings of populations defined by haplogroup A (eastern Pyrenees, Central Pyrenees plus northeast Spain, 'Massif Central') showed unimodal mismatch distributions, with significant signals of expansion (raggedness index $P > 0.05$), suggesting that they underwent haplogroup expansions shortly before or during the last glacial period (c. 110 000–12 000 years ago). Treating haplogroup B as a single population gave a unimodal mismatch distribution, but the hypothesis of expansion was rejected (raggedness index $P < 0.05$). Expansion was supported for sub-haplogroup B CB04 (420 168–25 452 years ago.), which occurred only in eastern Pyrenees, but not for sub-haplogroup B CB05 or its haplotypes from eastern Pyrenees.

Discussion

On the basis of mitochondrial cyt b and three nuclear loci, *P. ariasi* was shown to be a single phylogenetic species, and there was no reproductive isolation between sympatric European populations defined by its two most abundant and divergent cyt b haplogroups (A and B). In addition, no locus was shown to be under strong directional or balancing selection and, therefore, it was appropriate to apply demographic tests to all populations of this sandfly using all loci.

Pleistocene genetic divergences and population expansions of *P. ariasi*

The MDIV datings placed all the divergences of the cyt b haplogroups in the Pleistocene, 2 588 000–10 000 years ago (Gibbard and van Kolfschoten, 2004), long after the final opening of the Gibraltar Straits, 5.5–4.9 m.y.a. (Steininger and Rogl, 1984). Therefore, any vicariant evolution was caused by other barriers, probably related to the climate oscillations that intensified in the early Pleistocene (Hewitt, 2004). Only more widespread sampling might indicate the origins of the Pleistocene 'Eve' of *P. ariasi* and if haplogroup F is restricted to the Atlas region where it was found. However, the current

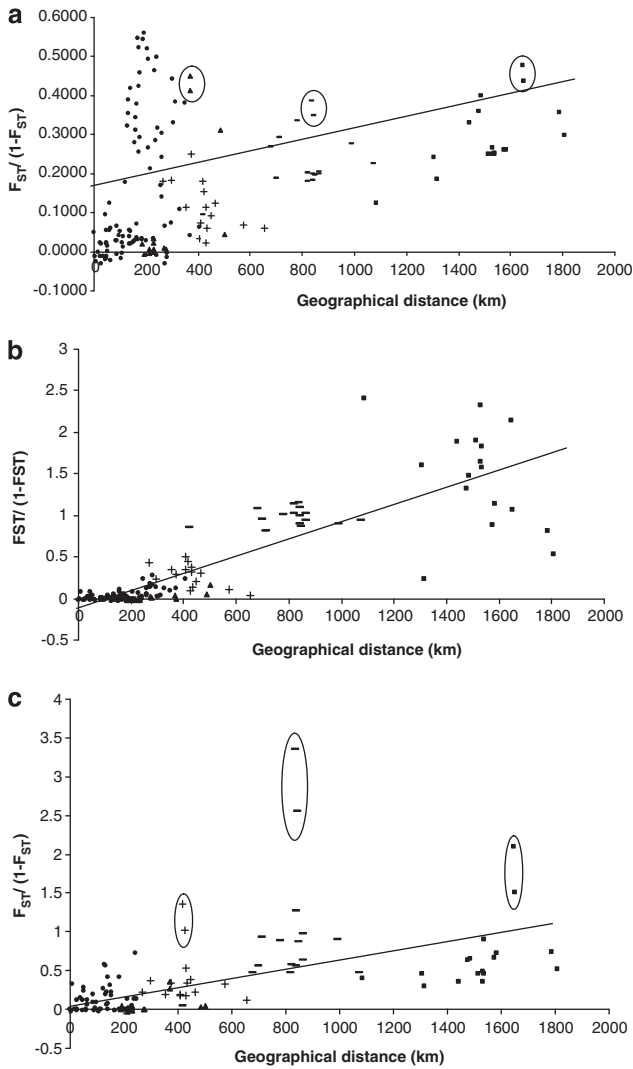


Figure 4 Plots and regression of genetic distance ($F_{ST}/(1-F_{ST})$) on geographical distance (km) between pairs of *P. ariasi* populations, for nuclear loci: (a) EF-1 α , (b) AAm20, and (c) AAm24. Symbols of pairwise population comparisons denote: ■, Morocco versus all others; ■■■, Portugal versus all others except Morocco (that is, France, Spain); +, NW (northwest) Spain versus all others except Morocco and Portugal (that is, France, northeast Spain); ▲, northeast Spain versus all others except Morocco, Portugal and northwest Spain (that is, France); ●, France only. Ringed data points indicate from left to right: Lot (LNP and RME) with northeast Spain, Portugal and Morocco for EF-1 α ; Lot (LNP and RME) with northwest Spain, Portugal and Morocco for AAm24.

distribution of haplogroups is only possible if at least one made an intercontinental crossing. On the basis of Bayesian phylogeny, cyt b haplogroup B is a marker for a phylogenetic species (Figure 2). There is similar evidence for the speciation of other taxa in Spain during the Pleistocene and Holocene, including water beetles (Ribera and Vogler, 2004). However, many species might only be phylogeographical variants, associated with refugia, unless reproductive isolation or strong divergent selection (Hewitt, 2004) has been demonstrated.

Each *P. ariasi* contained just one cyt b haplotype, and in most European flies this belonged to one of the two most divergent haplogroups, A (80.0%) and B (15.9%). These haplogroups did not correspond directly to geographical populations, indicating a distinct and separate history for each. Therefore, we followed O’Loughlin *et al.* (2007) by constructing separate mismatch distributions for each haplogroup, to infer whether there had been expansions experienced by each lineage rather than the mixed-ancestry geographical populations. Our analyses showed that cyt b lineages had experienced expansions much later than their divergences, consistent with population contractions during glacial periods. The expansion of haplogroup A was the more recent (Table 3) and, despite the large confidence intervals for the dating obtained by using two extreme rates of change of cyt b, it is likely to have occurred in the late Pleistocene before the last glacial maximum, 23–18 k.y.a. (Hewitt, 2004). This is consistent with the expansion signal detected in many French populations by comparing Fu’s F_s with Tajima’s D and R_2 statistics. On the basis of current predominance of haplogroup A in Lot and the northern Massif Central, where the species is unlikely to have survived the last glacial maximum, there must have been a post-glacial spatial expansion northwards in the Holocene. Immigration into France was probably from northeast Spain, because of the regional similarities at all loci and the broad littoral region that has often been warmer than northwest Spain (Delmas *et al.*, 2008), but IBD analyses did not demonstrate this.

Refugial populations of *P. ariasi* north of the Pyrenees?

Cyt b haplogroups A–C were found both north and south of the Pyrenees, indicating that *P. ariasi* has not undergone any long-term bottleneck in France. The nuclear loci indicate shallow IBD, except sometimes for northern leading-edge populations (Massif Central, Lot and Rhone valley), which is also consistent with the paradigm for some temperate species (Hewitt, 2004) of a

Table 2 Isolation with migration coalescence model in MDIV: to estimate gene coalescence and divergence times for pairs of cytochrome b haplogroups found in all populations of *P. ariasi*, using two mutation rates (μ)

Cytochrome b haplogroup	θ	T	T_{MRCA}	t_{MRCA} (y.a. for $\mu = 2.3\%$)	t_{MRCA} (y.a. for $\mu = 1\%$)	t (y.a. for $\mu = 2.3\%$)	t (y.a. for $\mu = 1\%$)
B+A	5.444	1.644	2.862	949 771	2 182 175	545 997	1 254 492
B+C	2.079	2.976	5.312	672 491	1 546 729	376 757	866 541
B+F	2.507	3.312	5.648	862 230	1 983 128	505 613	1 162 911
A+C	5.787	1.272	2.074	730 989	1 681 276	448 321	1 031 139
A+F	6.479	1.288	2.026	799 321	1 838 439	508 156	1 168 761
C+F	2.408	2.72	4.49	658 380	1 514 275	399 840	917 333

Abbreviations: θ , scaled parameter for nucleotide heterozygosity; T, scaled gene divergence time; T_{MRCA} , expected time to the most recent common ancestor, t_{MRCA} , gene coalescence time; t, gene divergence time; y.a., years ago.

Table 3 Mismatch distribution statistics for *P. ariasi* cytochrome b haplogroups (A–F) and sub-haplogroups (B_CB0N)

Statistics/haplogroup	A	B	C	F	B_CB05	B_CB04
Raggedness index	0.07941	0.11236	0.09959	0.0157	0.1656	0.36789
Raggedness <i>P</i> -value	0.4784	0.0229	0.3379	0.996	0.0292	0.6405
Tau, τ (95% CI)	0.73 (0.60–0.92)	1.199 (0.86–1.69)	2.826 (0.79–5.01)	2.168 (0.53–4.68)	1.020 (0.36–1.65)	3.000 (0.42–3.00)
<i>t</i> y.a. (95% CI) for $\mu = 2.3\%$	44 453 (36 869–55 780)	NA	172 086 (48 406–305 302)	132 081 (32 350–284 965)	NA	182 682 (25 452–182 682)
<i>t</i> y.a. (95% CI) for $\mu = 1.0\%$	102 240 (84 800–128 294)	NA	395 798 (111 333–702 195)	303 641 (74 405–655 419)	NA	420 168 (58 539–420 168)

Abbreviations: CI, confidence interval; NA, not applicable; y.a., years ago.

Sudden demographic expansion detected when Raggedness *P*-value > 0.05. CI: 95%. Time elapsed since beginning of expansion event (*t*) calculated from $\tau = 2 \mu t$, where μ is mutation rate. Generation time = 1 per annum.

spread northwards out of Iberia in the Holocene. However, we found cyt b sub-haplogroup B CB04 to be omnipresent in the French eastern Pyrenees, but absent from all other populations in France, Spain and Portugal, and with its latest expansion date (420 168–25 452 years ago for one generation p.a.) before the last glacial maximum. Its current endemicity can be explained by a range spread from Spain into France, either in the Holocene or before the last glacial period, 110–12 k.y.a. (Gibbard and van Kolfschoten, 2004). The Holocene scenario requires sub-haplogroup B CB04 to have arrived in a phalanx-like mass displacement (see next section) of all haplogroups (A–C), whereas the Pleistocene scenario requires it to be a marker for a population that survived north of the Pyrenees during one or two glacial periods, before its refuge was invaded by inter-glacial dispersers with haplogroup A. The evidence supports the latter: the combination of F_s , D , and R_2 statistics detected expansion signals only for haplogroup A of populations within the putative refuge; and, mismatch analysis demonstrated sub-haplogroup B CB04 undergoing an expansion event separate from that of sub-haplogroup B CB05, which is now widespread both sides of the Pyrenees. However, a higher density of sampling is needed to confirm the absence of sub-haplogroup B CB04 in northern Spain.

The ability of *P. ariasi* to survive north of the Pyrenees during glacial periods can only be inferred from its current bioclimate envelope. Its hibernating larvae can survive for weeks at 2–7 °C (Ready and Croset, 1980), and the oaks characteristic of its favoured humid and sub-humid Mediterranean bioclimates (Rioux *et al.*, 1984) left pollen traces of their survival in southern France during the last glacial (Beaudouin *et al.*, 2007). However, these oaks flourish in colder (Deciduous *Quercus pubescens*) and drier climes (Evergreen *Q. ilex*) (Jalut *et al.*, 2009) than *P. ariasi*, and the interpretation of the pollen record is questionable (Calvet, 2004). The snow line on the northern face of the eastern Pyrenees is now higher (2700–2800 m a.s.l.) than it was at the last glacial maximum, 1400–1500 m a.s.l. (Calvet, 2004), when the altitudinal limit of *P. ariasi* could have dropped from the current c. 1500 m a.s.l. (Rioux and Golvan, 1969) to near sea level. The saddlebacked bushcricket is a temperate species also proposed to have had glacial refugia in the French eastern Pyrenees (Spooner and Ritchie, 2006), in the valleys where *P. ariasi* was sampled.

Recent post-glacial re-colonizations of *P. ariasi* not blocked by refugial populations north of the Pyrenees Hewitt (1996, 1999) highlighted the effects of different modes of dispersal on re-colonization. Our study indicates that the dispersal mode of *P. ariasi* has often been phalanx—on a broad front, typified by IBD and mixing of cyt b haplogroups. Long-range pioneer dispersal (leptokurtic) often produces genetically homogeneous populations, which was not typical of *P. ariasi* and would not be expected because of a flight range limited to 0.1–2 km (Killick-Kendrick *et al.*, 1984). Phalanx-like dispersal is least likely to have been blocked by small refugial or relict leading-edge populations that had survived glacial periods, and this fits finding cyt b haplogroup A predominating over haplogroups B and C in the eastern Pyrenees. The re-colonization of *P. ariasi* could have kept pace with that of its woodland habitat,

which spread at a rate of 50 + m p.a. (Hewitt, 1999), and for *Q. ilex* produced an Iberia–Italy hybrid zone in the Rhone valley (Lumaret et al., 2002).

Monopolization currently blocking the northward spread of Pyrenean *P. ariasi* and potentially of leishmaniasis

There was a step change in F_{ST} values between the eastern Pyrenees and the Massif Central for EF-1 α and cyt b, and this regionality was supported by the analysis of molecular variance of cyt b. We conclude that cyt b haplogroup A is a marker for the most recent population expansion of *P. ariasi*, and this probably started south of the Pyrenees (c. 128 294–36 869 years ago) because of the high frequency of this haplogroup and its modal haplotype (CB25) in northern Portugal and Spain. It predominates north of the Pyrenees, to the exclusion of all other haplogroups in bottlenecked, leading-edge populations in the Massif Central and nearby. Currently, stepping-stone dispersal across a lowland corridor from the eastern Pyrenees to the Massif Central might be being blocked by the monopolization (Loeille and Leibold, 2008) of the latter by sandfly populations characterized by cyt b haplogroup A and relatively large discontinuous genetic distances for both cyt b and EF-1 α . This would hinder the spread of zoonotic leishmaniasis to the northern France if the Massif Central populations were found to be relatively poor dispersers or vectors. Reforestation of the lowland corridor between the two southern uplands might increase gene flow and alter population characteristics. These working hypotheses should be tested by using more polymorphic markers (such as true microsatellites) to estimate migration rates.

Actually, leishmaniasis foci in the Massif Central are distinctive, characterized by low diversity of regional *L. infantum* strains, high disease prevalences in domestic dogs (the reservoir), frequent cutaneous lesions but low prevalences of visceral leishmaniasis in humans, and a preponderance of *P. ariasi* (Pratlong et al., 2004). The population differentiation of *P. ariasi* is unlikely to match that of *P. perniciosus*, because this alternative vector peaks at lower altitudes (Rioux et al., 1984), and has two cyt b lineages (Iberia, Italy–northern Africa) mixing in France (Perrotey et al., 2005).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Ana Aransay, Samia Boussaa, Alison Cownie, Robert Farkas, Montserrat Gállego, Parviz Parvizi and Carlos Pires for specimens, Julia Llewellyn-Hughes and Claire Griffin for DNA sequencing, and Anna Papadopoulou for advice on phylogenetic analysis. This research was funded by EU grant GOCE-2003–010284 EDEN and is catalogued as publication EDEN0234 (<http://www.eden-fp6project.net/>), which does not necessarily reflect the views of the European Commission.

References

Anisimova M, Bielawski JP, Yang Z (2002). Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol Biol Evol* **19**: 950–958.

- Aransay AM, Malarky G, Ready PD (2001). Isolation (with enrichment) and characterization of trinucleotide microsatellites from *Phlebotomus perniciosus*, a vector of *Leishmania infantum*. *Mol Ecol Notes* **1**: 176–178.
- Aransay AM, Ready PD, Morillas-Marquez F (2003). Population differentiation of *Phlebotomus perniciosus* in Spain following postglacial dispersal. *Heredity* **90**: 316–325.
- Beaudouin C, Jouet G, Suc J-P, Berné S, Escarguel G (2007). Vegetation dynamics in southern France during the last 30 ky BP in the light of marine palynology. *Quaternary Sci Rev* **26**: 1037–1054.
- Brower AVZ (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proc Natl Acad Sci USA* **91**: 6491–6495.
- Calvet M (2004). The quaternary glaciation of the pyrenees. In: Ehlers J, Gibbard PL (eds). *Quaternary Glaciations; Extent and Chronology: Europe Pt.1 (developments in Quaternary science)*. Elsevier Science: London, pp 119–128.
- Clement M, Posada D, Crandall KA (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1659.
- Delmas M, Gunnell Y, Braucher R, Calvet M, Bourlès D (2008). Exposure age chronology of the last glaciation in the eastern Pyrenees. *Quaternary Res* **69**: 231–241.
- Dowling DK, Friberg U, Lindell J (2008). Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol Evol* **23**: 546–554.
- Esseghir S, Ready PD, Ben-Ismaïl R (2000). Speciation of *Phlebotomus* sandflies of the subgenus *Larrousius* coincided with the late Miocene-Pliocene aridification of the Mediterranean subregion. *Biol J Linn Soc* **70**: 189–219.
- Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**: 47–50.
- Fu Y-X (1997). Statistical tests of neutrality against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- Gállego Berenguer J, Botet Fregola J, Gállego Culleré M, Portús Vinyeta M (1992). Memoriam Professor Doc D Francisco de Paula Martínez Gomez. Los phlebotomos de la España peninsular e Ilos Baleares Identificación y corología. In: Hernandez-Rodríguez S (ed). *Comentarios sobre los métodos de captura*. Universidad de Córdoba. pp 579–600.
- Gibbard P, van Kolfschoten T (2004). The Pleistocene and Holocene Epochs. In: Gradstein FM, Ogg JG, Smith AG (eds). *A Geologic Time Scale 2004*. Cambridge University Press: Cambridge.
- Gómez A, Lunt DH (2007). Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N (eds). *Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity*. Kluwer Academic Publishers Group: The Netherlands.
- Goudet J (2002) FSTAT version 2.9.3.2. *A Programme to Estimate and Test Gene Diversities and Fixation Indices*. <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Harpending RC (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* **66**: 591–600.
- Hart MW, Sunday J (2007). Things fall apart: biological species form unconnected parsimony networks. *Biol Lett* **3**: 509–512.
- Hasegawa M, Kishino H, Yano T (1985). Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**: 160–174.
- Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc* **58**: 247–276.
- Hewitt GM (1999). Post-glacial re-colonization of European biota. *Biol J Linn Soc* **68**: 87–112.
- Hewitt GM (2004). Genetic consequences of climatic oscillations in the Quaternary. *Phil Trans R Soc Lond B* **359**: 183–195.

- Jalut G, Dedoubat JJ, Fontugne M, Otto T (2009). Holocene circum-Mediterranean vegetation changes: Climate forcing and human impact. *Quaternary Intl* **200**: 4–18.
- Jenness J (2005). *Distance Matrix (dist_mat_jen.avx) extension for ArcView 3.x, v. 2*. Jenness Enterprises, Flagstaff, AZ, USA. Available http://www.jennessent.com/arcview/dist_matrix.htm.
- Kass RE, Raftery AE (1995). Bayes factors. *J Amer Stat Assoc* **90**: 773–795.
- Killick-Kendrick R, Rioux JA, Bailly M, Guy MW, Wilkes TJ, Guy FM et al (1984). Ecology of leishmaniasis in the south of France. 20. Dispersal of *Phlebotomus ariasi* Tonnoir, 1921 as a factor in the spread of visceral leishmaniasis in the Cévennes. *Ann Parasitol Hum Comp* **59**: 555–572.
- Loeille N, Leibold MA (2008). Evolution in metacommunities: On the relative importance of species sorting and monopolization in structuring communities. *Amer Naturalist* **171**: 788–799.
- Lumaret R, Mir C, Michaud H, Raynal V (2002). Phylogeographical variation of chloroplast DNA in holm oak (*Quercus ilex* L.). *Mol Ecol* **11**: 2327–2336.
- Maroli M, Rossi L, Baldelli R, Capelli G, Ferroglio E, Genchi C et al. (2008). The northward spread of leishmaniasis in Italy: evidence from retrospective and ongoing studies on the canine reservoir and phlebotomine vectors. *Trop Med Int Health* **13**: 256–264.
- McDonald JH, Kreitman M (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- Nei M, Gojobori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.
- Nielsen R, Wakely J (2001). Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.
- Nylander JAA (2004). *MrModeltest v2*, Program Distributed by the author, Evolutionary Biology Centre, Uppsala University.
- O' Loughlin SM, Somboon P, Walton C (2007). High levels of population structure caused by habitat islands in the malarial vector *Anopheles scanloni*. *Heredity* **99**: 31–40.
- Parvizi P, Taherkhani H, Ready PD (2010). *Phlebotomus caucasicus* and *Phlebotomus mongolensis* (Diptera: Psychodidae): indistinguishable by the mitochondrial cytochrome b gene in Iran. *Bull Entom Res* **100**: 415–420.
- Perrotey S, Mahamdallie S, Pesson B, Richardson KJ, Gallego M, Ready PD (2005). Postglacial origins and dispersal of *Phlebotomus perniciosus* in France. *Parasite* **12**: 283–291.
- Pesson B, Ready JS, Benabdenni I, Martín-Sánchez J, Esseghir S, Cadi-Soussi M et al. (2004). Sandflies of the *Phlebotomus perniciosus* complex: mitochondrial introgression and a new sibling species of *P. longicuspis* in the Moroccan Rif. *Med Vet Entomol* **18**: 25–73.
- Pratlong F, Rioux JA, Marty P, Faraut-Gambarelli FF, Dereure J, Lanotte G et al. (2004). Isoenzymatic analysis of 712 strains of *Leishmania infantum* in the south of France and relationship of enzymatic polymorphism to clinical and epidemiological features. *J Clin Microbiol* **42**: 4077–4082.
- Rambaut A, Drummond AJ (2007). *Tracer v1.4*. Available from <http://beast.bio.ac.uk/Tracer>.
- Ramirez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008). Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* **179**: 555–567.
- Ramos-Onsins SE, Rozas J (2002). Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* **19**: 2092–2100.
- Rand DM, Kann LM (1996). Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol* **13**: 735–748.
- Raymond M, Rousset F (1995). GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity* **86**: 248–249.
- Ready PD (2010). Leishmaniasis emergence in Europe. *Euro Surveill* **15**: pii = 19505.
- Ready PD, Croset H (1980). Diapause and laboratory breeding of *Phlebotomus perniciosus* Newstead and *Phlebotomus ariasi* Tonnoir (Diptera: Psychodidae) from southern France. *Bull Entom Res* **70**: 511–523.
- Ribera I, Vogler AP (2004). Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Mol Ecol* **13**: 179–193.
- Rioux JA, Golvan YJ (1969). Epidémiologie des leishmanioses dans le sud de la France. *Monographie de l'INSERM (Paris)* **37**: 1–223.
- Rioux JA, Rispaill P, Lanotte G, Lepart J (1984). Relations Phlébotomes-Bioclimats en écologie des leishmanioses. Corollaires épidémiologiques. *L'exemple du Maroc Bull Soc Bot Fr* **131**: 549–557.
- Rogers AR, Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* **9**: 552–569.
- Ronquist F, Huelsenbeck JP (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Rousset F (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Rozas J, Sánchez-Delbarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- Sasai N, Kato Y, Kimura G, Takeuchi T, Yamaguchi M (2007). The *Drosophila jumonji* gene encodes a JmjC-containing nuclear protein that is required for metamorphosis. *FEBS Journal* **274**: 6139–6151.
- Schmidt D, Pool J (2002). The effect of population history on the distribution of the Tajima's *D* statistic. <http://www.cam.cornell.edu/~deena/TajimasD.pdf>.
- Schmitt T (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers Zool* **4**: 11.
- Sommer SS, Groszback AR, Bottema CD (1992). PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. *Biotechniques* **12**: 82–87.
- Spooner LJ, Ritchie MG (2006). An unusual phylogeography in the bushcricket *Ephippiger ephippiger* from Southern France. *Heredity* **97**: 398–408.
- Steininger FF, Rogl F (1984). Palaeogeography and palinspastic reconstruction of the Neogene of the Mediterranean and Paratethys. In: Dixon JE, Robertson AFH (eds). *The Geological Evolution of the Eastern Mediterranean*. Blackwell Scientific Publications: Oxford.
- Sudia WD, Chamberland RW (1962). Battery operated light trap, an improved model. *Mosquito News* **22**: 126–129.
- Swofford DL (2002). *PAUP Phylogenetic Analysis Using Parsimony, version 4.0b10*. Smithsonian Institution Press: Washington DC.
- Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* **7**: 453–464.
- Testa JM, Montoya-Lerma J, Cadena H, Oviedo M, Ready PD (2002). Molecular identification of vectors of *Leishmania* in Colombia: mitochondrial introgression in the *Lutzomyia townsendi* series. *Acta Trop* **84**: 205–218.
- Wayne ML, Simonsen KL (1998). Statistical tests of neutrality in the age of weak selection. *TREE* **13**: 236–240.
- Wright S (1978). *Evolution and the Genetics of Populations, Vol. IV. Variability within and among natural populations*. University of Chicago Press: Chicago.
- Yang Z (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)