Population differentiation of *Phlebotomus perniciosus* in Spain following postglacial dispersal

AM Aransay[1], PD Ready[1] and F Morillas-Marquez[2]

[1]*Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK;* [2]*Departamento de Parasitologia, Facultad de Farmacia, Universidad de Granada, Granada, Spain*

Comparative sequencing of mitochondrial cytochrome *b* (Cyt *b*) and isoenzyme analyses have not resolved the population structure of the Iberian lineage of the sandfly *Phlebotomus perniciosus*, the most widespread vector of *Leishmania infantum* (Protozoa, Trypanosomatidae) to humans and dogs in the western Mediterranean subregion. Allelic variation at trinucleotide microsatellite loci was investigated in 13 Spanish populations of *P. perniciosus*. Four out of five loci showed significant differentiation between (pairwise $F_{ST} > 0.23$), but not within (pairwise $F_{ST} < 0.05$), two regional groups of populations (southern and northeastern). All Cyt *b* sequences belonged to the Iberian lineage, which differs by six fixed nucleotide differences from the typical lineage found in northwest Africa, Malta and Italy. The northeastern group of Spanish populations had a reduced number of micro-satellite alleles (16 out of the 29 present in the southern populations), indicating its derivation as a peripheral isolate following the species' dispersal from a southern Ice Age refuge 8000–12 000 years ago. Pairwise F_{ST} values did not increase with geographical distance between populations, over distances of 246–850 km (between regions) and 16–491 km (within regions). This suggests that the two regional groups of populations remain isolated, but that within each region there are no significant permanent barriers to gene flow between contiguous populations. These findings will help to predict the capacity of this sandfly to disperse, and originate new foci of leishmaniasis, in response to climate warming.
Heredity (2003) **90,** 316–325. doi:10.1038/sj.hdy.6800246

Keywords: *Phlebotomus perniciosus*; sandfly; microsatellites; population genetics; postglacial dispersal; Spain

Introduction

In southwest Europe and northwest Africa, the blood-sucking females of *Phlebotomus (Larroussius) perniciosus* Newstead, 1911 (Diptera: Psychodidae) are the most widespread sandfly vectors of *Leishmania infantum* (Kinetoplastida: Trypanosomatidae), the parasitic proto-zoan that causes most visceral and cutaneous leishmaniasis in humans and canine reservoir hosts in the region (Rioux and Lanotte, 1990). We report here the use of polymorphic microsatellite DNA loci, isolated by Aransay *et al* (2001), to investigate the population structure of *P. perniciosus* in Spain. Such studies of insect transmitters of disease are potentially important for planning intervention strategies, as they can help predict the dispersal not only of vectorial traits associated with specific populations but also of genetically modified organisms and the genes they carry. The interest of this approach has been demonstrated for the *Anopheles gambiae* complex, which contains most of the major vectors of human malaria in sub-Saharan Africa (Lehmann *et al*, 1996, 2000; Donnelly and Townson, 2000).

Important changes in the distributions of 'tropical diseases' like leishmaniasis (Cross and Hyams, 1996) can be expected to result from climate warming, which in Europe could shift northwards the boundary of the Mediterranean subregion. *P. perniciosus* is restricted to the western Palaearctic region, where it is found in Portugal, Spain, France (north to Paris), southern Switzerland, the northwest Balkans, Italy, Malta, Libya, Tunisia, Algeria and Morocco (Seccombe *et al*, 1993). It is more abundant at lower altitudes in the Mediterranean subregion, and in northwest Africa the subhumid and semiarid bioclimate zones are preferred (Rioux *et al*, 1984). This subtropical distribution makes it a novel insect model for investigating the effects of climate warming in Europe, studies of which have relied on temperate insects (Hewitt, 1996, 1999).

The current population structure of *P. perniciosus* is consistent with it having suffered a range contraction during the last Ice Age, in the Pleistocene epoch, when its European populations could have survived only in southern Spain and Italy. This conclusion comes from molecular evolution studies of mitochondrial cytochrome *b* gene sequences (Cyt *b*) (Esseghir *et al*, 1997, 2000) and from population genetics based mostly on three polymorphic isoenzyme loci (Benabdennbi, 1998). Two main lineages were discovered: the typical lineage, extending from Morocco, through Tunisia and Malta (the type locality) to Italy; and an Iberian lineage, within which two northeastern populations were slightly differentiated at one isoenzyme locus (Benabdennbi, 1998; Mahamdallie *et al*, 2003). Lack of polymorphism in these markers, however, meant that population structure within the two lineages was poorly resolved.

Microsatellites are often highly polymorphic (Loxdale and Lushai, 1998) and, therefore, should be better able to resolve the current population structure of the Iberian

Correspondence: PD Ready, Molecular Systematics Laboratory, Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. E-mail: P.Ready@nhm.ac.uk

lineage of *P. perniciosus* and its postglacial dispersal during the last 8000–12 000 years. Commonly used classes of microsatellites can occur in low copy numbers in sandflies (Day and Ready, 1999), and so an enrichment protocol had been modified to help isolate two classes of trinucleotide repeats (AGC and AGG) from *P. perniciosus* (Aransay *et al*, 2001). Five polymorphic trinucleotide microsatellite loci were used in the present study, one of which was not described by Aransay *et al* (2001).

Materials and methods

Sample collection
A total of 21 population samples were collected from 16 localities (Table 1, Figure 1) within four Mediterranean bioclimate zones in Spain (Anon, 1965; Morillas-Marquez *et al*, 1983). Sandflies were captured overnight in CDC (Centers for Disease Control) miniature light traps, placed peridomestically inside or near stables and houses. They were stored either dry at −80°C or in 96% ethanol at −20°C.

DNA extraction
The heads and genitalia of individual sandflies were dissected and slide-mounted in Berlese fluid, prior to morphological identification (Martinez-Ortega and Conesa-Gallego, 1987). Total genomic DNA was extracted from the thorax and anterior abdomen of each specimen (Ready *et al*, 1991), before being resuspended in $1 \times$ TE buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0) and stored at −20°C.

PCR amplification of microsatellite loci and genotype scoring
Microsatellites were previously isolated from sandflies collected in Turre (Almería province), a putative Pleistocene Ice Age refuge in southern Spain. Locus-specific primers were designed based on the sequences flanking the microsatellites, and PCR conditions were established for four of the five loci used in the present study – AAm13, AAm20 and AAm82 with AGG-class repeats and AAm24 with AGC-class repeats (Aransay *et al*, 2001). A new AGC-class locus is reported here for the first time. AAm10 contained seven CAG repeats in the original clone (GenBank accession no. AJ516955 and its alleles were amplified with the primers (5′–3′) CACTCCCGG-CACTGCTCAC and **6FAM-**GCAACTGAAGCTGGAG-CAGC (with fluorescent marker dye in bold) at annealing temperatures of 57°C (first stage) and 55°C (second stage).

Microsatellite alleles were amplified by PCR from single-sandfly DNA extracts and aliquots of the products were quantified on 2% agarose gels to estimate dilution factors for the fluorescent dye labels (Aransay *et al*, 2001). The amplified allelic DNA fragments were sized using an automated ABI377 sequencer in the PE-GeneScan mode, by multiplexing all five loci of each sandfly in a single lane of a polyacrylamide gel. Genotyper software (PE Applied Biosystems) was used to score the alleles.

PCR amplification and sequencing of a fragment of mitochondrial cytochrome *b* (Cyt *b*)
A 488-base pair (bp) fragment of Cyt *b* was amplified with primers CB1-SE (5′-TATGTACTACCCTGAGGA-CAAATATC-3′) and CB3-R3A (5′-GCTATTACTCC(T/C)CCTAACTT(A/G)TT-3′). Cycle sequencing was carried out on both strands with CB3-R3A and a modified sense-strand primer, CB1 (5′-TATGTACTACCATGAG-GACAAATATC-3′), using an ABI PRISM Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit v. 2.0 and ABI373/377 sequencing systems (ABI, PE Applied Biosystems). The primers and conditions were described by Esseghir (1998) and Esseghir *et al* (2000). It should be noted that the latter failed to mention that CB1-SE, not CB1, was used for PCR amplification.

Data analysis
Genotype distributions were tested for departure from Hardy–Weinberg equilibrium at each locus in each population, using Arlequin version 2.000 (Schneider *et al*, 2000) that incorporated a modified version of the Markov-chain random walk algorithm described by Guo and Thomson (1992).

FSTAT version 2.9.3.1 (Goudet, 2001) was used to calculate allele frequencies at each locus, to estimate the pairwise F_{ST} statistic with a weighted analysis of variance according to Weir and Cockerham (1984), and to test the statistical significance of the pairwise F_{ST} values after standard Bonferroni corrections.

Geographical distances between sample sites were calculated using the measuring tool in Atlas Mundial Microsoft® Encarta® 99 (Anon, 1998), so that isolation by distance could be examined by a Mantel test (with 1000 permutations) using Genepop (version 3.3) (Raymond and Rousset, 1995). Genepop (version 3.3) was also used to test for linkage disequilibrium and to calculate the *Rho* statistic (Rho_{ST}) as described by Michalakis and Excoffier (1996).

Cyt *b* sequences were aligned and analysed in Sequencher™ 3.0 (ABI, PE Applied Biosystems).

## Results

Allelic variation at five microsatellite loci
Alleles at the five microsatellite loci were successfully amplified from all 384 *P. perniciosus* (187 males and 197 females) obtained from 21 samples (16 localities) in Spain (Table 1, Figure 1). No locus failed to amplify in any single sandfly DNA extract. The number of sandflies available and characterized per sample ranged from 1 to 38, and so some samples had to be pooled for population genetics analysis. This was based on 13 populations, four of which were pools. Pooled populations were only made with samples collected on different dates (1–20 days apart) but from the same location (MAD-Boa, RIO-Tud, BAR-San), or on different dates (1–3 days apart) from nearby locations (RIO-NW). This was justified by the homogeneity of samples within each of the two geographical regions (Figure 1), explained in the next section.

The total number of alleles per locus ranged from 4 to 9. Allele frequencies for all loci in each population are shown in Table 2. Sex-specific alleles were not observed for any of the loci. Locus AAm24 was neither informative for analysing population structure nor significantly polymorphic: the frequency of its most common allele (208 bp) was greater than 0.95 except in the MUR-Sis population (0.943). However, locus AAm13 was infor-

Table 1 Descriptions of the populations analysed, the samples and their collection details

Population code	HUE-Rio	GRA-Alf	GRA-Tor-01	GRA-Tor-02	GRA-Tre	ALM-Tur	MUR-Sis	MAD-Boa
Sample	Rio Tinto	Alfacar	Torvizcon	Torvizcon	Trevelez	Turre	Siscar	Boadilla del Monte
Location	Huelva	Granada	Granada	Granada	Granada	Almeria	Murcia	Madrid
Sample date	30.08.99	00.06.99	30.09.99	14.06.00	29.06.00	31.07.99	21.09.95	24.08.00 07.09.00
Altitude[a] (m)	321	1019	655	655	1375	25	9	675
Bioclimate[b]	Meso-Mediterranean continental	Meso-Mediterranean continental	Meso-Mediterranean	Meso-Mediterranean	Supra-Mediterranean	Thermo-Mediterranean	Meso-Mediterranean	Meso-Mediterranean continental
CDC position[c]	Out	Out	In+out	Out	In	In	In	Out
Available hosts	Mules, horses	Goats, dogs	Pigeons, pigs, chickens, dogs, rabbits, mule	Rabbits, chickens, horse, mule	Horses	Horses	Pigs	Dogs, humans
Sample sizes[d]	19 (9 M, 10 F)	38 (35 M, 3 F)	34 (8 M, 26 F)	33 (18 M, 15 F)	32 (28 M, 4 F)	31 (2 M, 29 F)	35 (15 M, 20 F)	28 (12 M, 16 F)

Population code	RIO-NW				RIO-Rob	RIO-Tud	TAR-Tor	BAR-San
Sample	Cuzcurrita de Rio Tiron	Rodezno	Ventosa	Medrano	Robres del Castillo	Tudelilla	Torroja del Priorat	Sant Just Desvern
Location	Rioja	Rioja	Rioja	Rioja	Rioja	Rioja	Tarragona	Barcelona
Sample date	28.07.00 29.07.00	28.07.00	29.07.00	30.07.00 31.07.00	31.07.00	01.08.00 02.08.00	26.07.00	18.08.00 07.09.00
Altitude[a] (m)	525	540	650	600	800	495	332	122
Bioclimate[b]	Meso-Mediterranean continental	Meso-Mediterranean continental	Meso-Mediterranean continental	Meso-Mediterranean continental	Supra-Mediterranean	Meso-Mediterranean continental	Meso-Mediterranean	Meso-Mediterranean
CDC position[c]	In	In	In	Out	Out	In	In	In
Available hosts	Chickens, dogs, cats, sheep	Sheep, dogs, horse	Sheep, dogs, chickens	Sheep	Cows, dogs	Sheep	Rabbits	Humans, chickens, horses
Sample sizes[d]	3+2 (0 M, 5 F)	1 (0 M, 1 F)	1 (0 M, 1 F)	2+4 (4 M, 2 F)	33 (17 M, 16 F)	1+28 (15 M, 14 F)	36 (6 M, 30 F)	8+15 (18 M, 5 F)

[a]Altitude according to Anon (2000).
[b]Bioclimates as deduced from Anon (1965) and Morillas Marquez *et al*. (1983).
[c]CDC light-trap position: 'In' means traps were placed overnight inside stables or houses, and 'Out' means they were placed outside but nearby stables or houses.
[d]Sample size for males (M) and females (F).
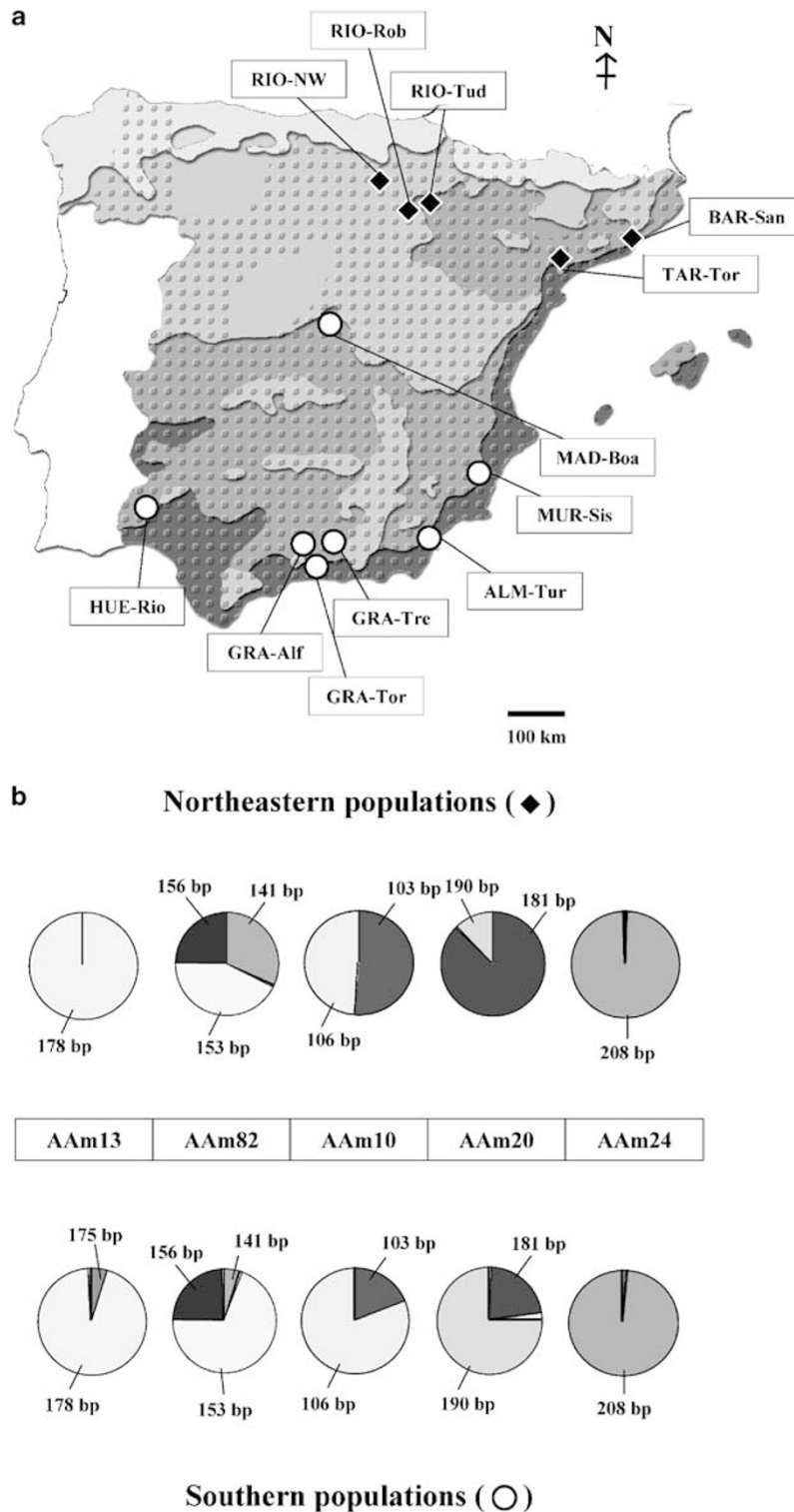
Figure 1 Distributions of *P. perniciosus* populations and their microsatellite alleles in Spain. (a) Map of Spain showing the locations of the populations analysed. Bioclimates (Anon, 1965; Morillas Marquez *et al*, 1983): light grey for temperate, medium grey for supramediterranean, dark grey for mesomediterranean and black for thermomediterranean. The dotted net covers the distribution of *P. perniciosus* (Houin, 1965; Gil Collado *et al*, 1989). (b) Variation in allele frequencies at five microsatellite loci in the northeastern and southern regions. The base pair (bp) numbers refer to the sizes of the more common alleles.

mative, even though the frequency of its most common allele (178 bp) was 0.96 overall. In contrast, the loci AAm10, AAm20 and AAm82 were polymorphic by definition in each and every population – the frequency

of the most common allele at each being <0.95 (Wright, 1978) – as well as being informative for analysing population structure. Southern populations (5/8), not northeastern populations, had private alleles at four

Table 2 The allele frequencies at each microsatellite locus recorded by population. *N*=population size

Locus	Allele size (bp)	HUE-Rio	GRA-Alf	GRA-Tor-01	GRA-Tor-02	GRA-Tre	ALM-Tur	MUR-Sis	MAD-Boa	RIO-NW	RIO-Rob	RIO-Tud	TAR-Tor	BAR-San	All
AAm10	100	0	0	0	0	0	0	0.029	0	0	0	0	0	0	0.003
	103	0.184	0.197	0.191	0.197	0.141	0.177	0.200	0.179	0.385	0.424	0.621	0.472	0.652	0.299
	106	0.816	0.776	0.809	0.803	0.859	0.823	0.771	0.821	0.615	0.576	0.379	0.528	0.348	0.695
	109	0	0.026	0	0	0	0	0	0	0	0	0	0	0	0.003
AAm13	175	0.053	0.066	0.044	0.061	0.031	0.048	0.043	0.054	0	0	0	0	0	0.033
	178	0.947	0.921	0.956	0.924	0.922	0.952	0.957	0.929	1	1	1	1	1	0.960
	181	0	0	0	0.015	0.047	0	0	0.018	0	0	0	0	0	0.007
	184	0	0.013	0	0	0	0	0	0	0	0	0	0	0	0.001
AAm20	175	0	0.013	0	0	0	0.016	0	0.018	0	0	0	0	0	0.004
	178	0.053	0	0	0	0	0	0	0	0	0	0	0	0	0.003
	181	0.237	0.276	0.221	0.212	0.250	0.145	0.243	0.143	0.923	0.833	0.845	0.889	0.913	0.447
	184	0	0	0.029	0	0.031	0.032	0.029	0	0	0	0	0.014	0	0.012
	187	0	0	0	0.015	0.016	0	0.014	0	0	0	0	0	0	0.004
	190	0.711	0.711	0.735	0.773	0.703	0.806	0.686	0.839	0.077	0.167	0.155	0.097	0.087	0.527
	193	0	0	0	0	0	0	0.014	0	0	0	0	0	0	0.001
	196	0	0	0.015	0	0	0	0	0	0	0	0	0	0	0.001
	202	0	0	0	0	0	0	0.014	0	0	0	0	0	0	0.001
AAm24	202	0	0	0.015	0	0.016	0.016	0.057	0	0	0	0	0.014	0	0.010
	205	0	0	0	0.015	0	0.016	0	0	0	0.015	0	0	0	0.004
	208	1	1	0.985	0.970	0.984	0.952	0.943	0.982	1	0.970	1	0.986	0.978	0.979
	211	0	0	0	0	0	0.016	0	0	0	0.015	0	0	0	0.003
	214	0	0	0	0.015	0	0	0	0	0	0	0	0	0.022	0.003
	217	0	0	0	0	0	0	0	0.018	0	0	0	0	0	0.001
AAm82	141	0.026	0.066	0.029	0.045	0.016	0.113	0.086	0	0.385	0.333	0.293	0.278	0.348	0.143
	147	0	0	0.015	0	0	0	0	0	0	0.015	0	0	0	0.003
	150	0	0.013	0.015	0	0	0.016	0	0.018	0	0	0.017	0	0	0.007
	153	0.684	0.671	0.706	0.606	0.781	0.661	0.743	0.679	0.538	0.288	0.483	0.458	0.457	0.600
	156	0.289	0.224	0.221	0.348	0.203	0.210	0.171	0.304	0.077	0.364	0.207	0.264	0.196	0.243
	159	0	0.026	0.015	0	0	0	0	0	0	0	0	0	0	0.004
	N	19	38	34	33	32	31	35	28	13	33	29	36	23	

Table 3 Observed (H_O) and expected (H_E) heterozygosities at each microsatellite locus. *P*=significance of the deviation from Hardy–Weinberg expectations; N = population size

Locus		HUE-Rio	GRA-Alf	GRA-Tor-01	GRA-Tor-02	GRA-Tre	ALM-Tur	MUR-Sis	MAD-Boa	RIO-NW	RIO-Rob	RIO-Tud	TAR-Tor	BAR-San
AAm10	H_O	0.368	0.342	0.323	0.333	0.281	0.354	0.314	0.285	0.615	0.606	0.413	0.555	0.347
	H_E	0.351	0.362	0.337	0.391	0.298	0.296	0.396	0.355	0.538	0.496	0520	0.505	0.463
	P	1.000	0.77	1.000	1.000	1.000	0.552	0.306	1.000	0.569	0.284	0.691	0.743	0.351
AAm13	H_O	0.105	0.157	0.088	0.151	0.156	0.096	0.085	0.142	N/A	N/A	N/A	N/A	N/A
	H_E	0.152	0.173	0.113	0.225	0.149	0.124	0.110	0.201	N/A	N/A	N/A	N/A	N/A
	P	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	N/A	N/A	N/A	N/A	N/A
AAm20	H_O	0.473	0.421	0.470	0.333	0.406	0.387	0.514	0.321	0.153	0.333	0.310	0.222	0.173
	H_E	0.590	0.424	0.437	0.436	0.499	0.358	0.495	0.338	0.147	0.282	0.266	0.251	0.341
	P	1.000	1.000	0.847	0.695	0.323	1.000	0.456	1.000	1.000	0.558	1.000	1.000	1.000
AAm24	H_O	N/A	0.026	0.029	0.060	0.031	0.096	0.057	0.035	N/A	0.060	N/A	0.027	0.043
	H_E	N/A	0.052	0.058	0.145	0.091	0.125	0.109	0.104	N/A	0.117	N/A	0.081	0.236
	P	N/A	1.000	1.000	1.000	1.000	1.000	0.085	1.000	N/A	1.000	N/A	1.000	1.000
AAm82	H_O	0.526	0.500	0.382	0.696	0.437	0.419	0.428	0.428	0.307	0.666	0.724	0.750	0.695
	H_E	0.495	0.488	0.464	0.568	0.401	0.541	0.421	0.599	0.578	0.683	0.681	0.652	0.646
	P	1.000	0.373	0.580	0.108	0.428	0.120	1.000	1.000	0.027	0.632	0.099	0.310	0.005
	N	19	38	34	33	32	31	35	28	13	33	29	36	23

Table 4 Matrix of genetic differentiation and geographical distances between pairs of populations. Below diagonal: pairwise F_{ST} statistics; above diagonal: geographical distances (km)

	HUE-Rio	GRA-Alf	GRA-Tor-01	GRA-Tor-02	GRA-Tre	ALM-Tur	MUR-Sis	MAD-Boa	RIO-NW	RIO-Rob	RIO-Tud	TAR-Tor	BAR-San
HUE-Rio	—	272	299	299	304	420	491	382	623	627	639	747	850
GRA-Alf	−0.0138	—	43	43	38	149	243	354	588	571	578	583	675
GRA-Tor-01	−0.0156	−0.0097	—	0	16	135	246	395	623	607	613	603	687
GRA-Tor-02	−0.0110	−0.0018	−0.0020	—	16	135	246	395	623	607	613	603	687
GRA-Tre	−0.0077	−0.0016	−0.0078	0.0122	—	123	231	383	610	592	598	587	671
ALM-Tur	−0.0051	−0.0006	−0.0067	−0.0009	0.0059	—	128	401	590	570	573	509	582
MUR-Sis	−0.0075	−0.0062	−0.0087	0.0100	−0.0047	−0.0010	—	360	518	480	480	384	455
MAD-Boa	−0.0085	0.0034	−0.0047	−0.0073	0.0075	−0.0062	0.0112	—	248	246	257	411	513
RIO-NW	0.2963*	0.2579*	0.3027*	0.3068*	0.3139*	0.3276*	0.2574*	0.3717*	—	42	55	326	418
RIO-Rob	0.2532*	0.2357*	0.2735*	0.2572*	0.2915*	0.2925*	0.2478*	0.3182*	0.0320	—	16	284	377
RIO-Tud	0.2935*	0.2657*	0.3047*	0.3016*	0.3240*	0.3292*	0.2697*	0.3578*	0.0285	0.0331	—	271	363
TAR-Tor	0.2769*	0.2532*	0.2911*	0.2873*	0.3042*	0.3183*	0.2592*	0.3451*	0.0017	0.0063	0.0055	—	108
BAR-San	0.3466*	0.3142*	0.3560*	0.3505*	0.3759*	0.3783*	0.3171*	0.4104*	0.0316	0.0414	−0.0133	0.0111	—

*Values significantly greater than could be expected by chance: $P < 0.01$, after standard Bonferroni corrections executed in FSTAT (Goudet, 2001).

loci (not AAm82): mean frequency 0.033 (range: 0.013–0.053).

The ranges of observed and expected heterozygosity values were 0.026–0.750 and 0.052–0.683, respectively (Table 3). Deviations from Hardy–Weinberg (HW) expectations were only significant ($P < 0.05$) for locus AAm82, in the two populations RIO-NW ($P = 0.027$) and BAR-San ($P = 0.005$), undoubtedly because these were the most heterogeneous of the pooled populations in terms of locations and dates, respectively.

No significant linkage disequilibrium was detected between pairs of loci in any population, except for AAm10/AAm13 in population GRA-Tre ($P = 0.021$) and for AAm20/AAm82 in RIO-NW ($P = 0.027$).

Population structure indicated by the allelic variation at the microsatellite loci

The genetic differences between pairs of populations, quantified with the F_{ST} statistic, ranged from 0.0006 to 0.4104, and clearly indicated the existence of two regional groups of populations (Table 4, Figure 2). Thus, F_{ST} estimates were very low (0.0006–0.0156) within a group containing all eight populations from seven southern locations (HUE-Rio, GRA-Alf, GRA-Tor-01 (1999), GRA-Tor-02 (2000), GRA-Tre, ALM-Tur, MUR-Sis and MAD-Boa); and, they were also very low (0.0017–0.0414) within a second group consisting of all five populations from eight northeastern locations (RIO-NW, RIO-Rob, RIO-Tud, TAR-Tor, BAR-San). In comparison, the F_{ST} values for inter-regional pairs of populations were very high (0.2357–0.4104). Within each region, F_{ST} values were not significantly greater than could be expected by chance ($P > 0.05$, after standard Bonferroni corrections), but all F_{ST} values between regions were significant at the 1% level (Table 4). Most pairwise F_{ST} values were greater when locus AAm13 was removed from the analysis (in consideration of the absence of polymorphism in all northeastern populations), but the statistical significance of the results remained the same.
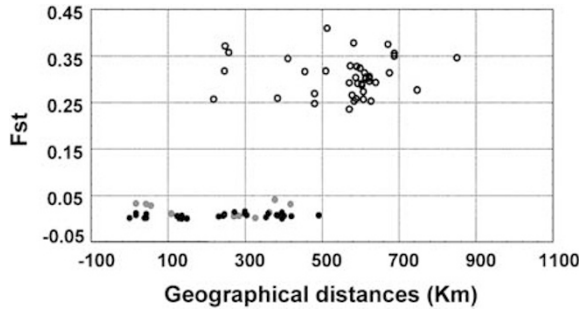
Figure 2 Variation of genetic differentiation between populations with geographical distance. Pairwise F_{ST} estimates are plotted against geographical distances (km) between pairs of populations within the southern region (Black circles. Spearman Rank correlation coefficient to fit F_{ST} to distance: $a = -0.002525$, $b = -0.000004$; $P = 0.732$ for greater correlation and $P = 0.268$ for lesser correlation, using Mantel test with 1000 permutations), within the northeastern region (Grey circles. $a = 0.022355$, $b = -0.000025$; $P = 0.763$ for greater correlation and $P = 0.270$ for lesser correlation), and between regions (open circles).

Rho_{ST} estimates (data not shown) were larger than the F_{ST} ones for most of the comparisons within the southern group, but smaller for most of the comparisons within the northeastern group. For inter-regional comparisons, Rho_{ST} values were higher than F_{ST} estimates.

The existence of these two regional groups of populations was also evident from an inspection of allele frequency differences at loci AAm13, AAm10, AAm20 and AAm82 (Table 2, Figure 1). At locus AAm13, alleles 175, 181 and 184 did not occur in any of the northeastern populations, which were homozygous for allele 178. For locus AAm10, the frequencies of the two common alleles showed nonoverlapping ranges in the southern and northeastern groups of populations, respectively: 0.141–0.200 and 0.385–0.652 for allele 103; and, 0.771–0.859 and 0.348–0.615 for allele 106. Similar, nonoverlapping regional variation was observed for locus AAm20 (0.143–0.276 and 0.833–0.923 for allele 181; and, 0.686–0.839 and 0.077–0.167 for allele 190) and locus AAm82 (0.000–0.113 and 0.278–0.385 for allele 141; and, 0.661–0.781 and 0.288–0.538 for allele 153).

There was a total of 29 alleles at all five loci. All occurred in the southern group of populations, but only 16 were found in the northeastern group (Table 2). This reduced allele diversity in the northeastern group was more notable for loci AAm10 (two alleles found out of a total of four), AAm13 (one out of four) and AAm20 (three out of nine). Less differentiated regionally were loci AAm82 and AAm24 (each with five alleles found in the northeastern group, out of a total of six). Furthermore, locus AAm24 was strictly monomorphic for its most common allele (208) in some populations, both in the south (HUE-Rio, GRA-Alf) and in the northeast (RIO-NW and RIO-Tud).

Pairwise F_{ST} estimates were plotted against geographical distances (Figure 2, Table 4) to illustrate the absence of any support for isolation by distance. When assessed by Mantel tests, there was no significant positive or negative correlation between pairwise F_{ST} estimates and geographical distance, neither for the total data nor within each of the southern and northeastern regions, where there was very low genetic differentiation

($F_{ST} < 0.05$) compared with the very high inter-regional genetic differentiation ($F_{ST} > 0.23$). The same result was obtained when $F_{ST}/(1 - F_{ST})$ was substituted for F_{ST} and either of these measures of genetic differentiation (Rousset, 1997) was correlated with $\log_e$ distance.

Identity of the mitochondrial Cyt *b* lineage in northeastern populations

New Cyt *b* sequences were obtained for a total of 24 specimens from the northeastern region: six (three males and three females) from RIO-NW, three (all females) from RIO-Rob, six (three males and three females) from RIO-Tud, three (all females) from TAR-Tor and six (three males and three females) from BAR-San. Additionally, sequences were obtained for the first time from specimens (three males and three females) of the most northerly population (MAD-Boa) of the southern region.

All Cyt *b* sequences belonged to the Iberian lineage according to the fixed nucleotide differences between it and the typical lineage found in northwest Africa, Malta and Italy (Esseghir *et al*, 2000). In the alignment of the nucleotide (nt) sequences of the Iberian and typical lineages (GenBank accession numbers AF161205 and AF161204, respectively), the bases at the diagnostic positions were: nt 9-G (not A), nt 54-A (not G), nt 81-G (not A), nt 138-T (not C), nt 375-C (not T) and nt 420-T (not C).

Discussion

Allelic variation and evolution at the five microsatellite loci

Allelic variation at four out of the five trinucleotide microsatellite loci proved to be informative for investigating the population structure of *P. perniciosus* in Spain. Only locus AAm24 was found to be neither informative nor polymorphic by definition (ie the overall frequency of its common allele was > 0.95). In this sense, locus AAm13 was also not polymorphic, but it did help to define the two regional groups of populations. The small amount of actual polymorphism at these two loci could have resulted from the presence of null alleles, which failed to amplify because of at least one difference in a flanking region sequence. Unlike AAm13, the other three loci (AAm10, AAm20, AAm82) were not only informative about population structure but also showed no evidence of a null allele effect. In all unpooled and large populations ($N = 19+$), the genotype frequencies at these loci conformed to HW expectations, and no significant association between loci was detected by linkage disequilibrium analysis. This indicated that their alleles were selectively neutral and being inherited in an independent and Mendelian fashion.

The number of alleles at each trinucleotide locus (4–9) in *P. perniciosus* was intermediate between the number (3–15) reported for *An. gambiae* (Lehmann *et al*, 1996) and that (2–6) reported for *Drosophila melanogaster* (Schug *et al*, 1998). This variation could reflect the different histories of the populations of these species, as well as differences in the evolutionary rates of different microsatellite classes and loci. Dinucleotide microsatellite loci may be more polymorphic, eg *An. gambiae* had 12–14 alleles per locus in sub-Saharan Africa (Lehmann *et al*, 1996) and its sibling species, *An. arabiensis* had 11–19 alleles per locus in East Africa (Donnelly and Townson, 2000).

F_{ST} values were lower than Rho_{ST} estimates for most of the comparisons between populations of *P. perniciosus*, which is expected from the stepwise mutation model (SMM) of microsatellite evolution, rather than the infinite alleles model (Jarne and Lagoda, 1996). However, as other parameters influence the behaviour of both statistics, this result cannot be taken as proof of the SMM.

Evidence for two distinct regional populations of *P. perniciosus* in Spain

The results provide strong evidence for population subdivision of *P. perniciosus* in Spain. Four out of the five microsatellite loci showed significant variability between, but not within, two regional groups of populations. One regional group included all eight populations from southern Spain, and the other comprised all five populations from the northeast. Pairwise F_{ST} estimates within each regional population were less than 0.05, which according to Wright (1978) is considered as indicating *little* genetic differentiation. In contrast, most inter-regional pairwise F_{ST} values were greater than 0.25 (two were 0.23–0.24), which indicates *very great* genetic differentiation. Some of the intra-regional F_{ST} (and Rho_{ST}) values were negative, indicating that the variation was sometimes higher within populations than between them. This suggests that the populations within regions were panmictic, which is supported by the absence of evidence for isolation by distance (Figure 2).

Origins and continued isolation of the two regional populations

There was reduced allele diversity in the northeastern region, where only 16 out of the 29 alleles were recorded, compared with the southern region (all 29 alleles found). By analogy with *An. gambiae* (Lehmann *et al*, 2000), the lack of unique alleles and the higher F_{ST} than Rho_{ST} values among northeastern populations of *P. perniciosus* both suggest that pure drift was the main differentiating process for the northeastern region of Spain. This is consistent with the hypothesis of historical migration from a southern Ice Age refuge, with the northeastern population arising as a northern peripheral isolate in which founder effects and bottleneck events led to the loss of rare alleles and significant changes in allele frequencies, as argued for temperate European insects (Hewitt, 1996, 1999). The two Spanish regional populations were shown to share the mitochondrial Cyt *b* lineage previously reported as characteristic of the Iberian race of *P. perniciosus* (Esseghir *et al*, 1997; Mahamdallie *et al*, 2003), and so they are likely to have become isolated in the postglacial Holocene epoch (starting 12 000 y.a.), long after the divergence of the Iberian and typical races, which was dated well before the end of the Pleistocene by the Cyt *b* molecular clock (Esseghir *et al*, 2000).

The absence of any significant correlation between genetic differentiation and geographical distance suggests that the two regional populations still remain isolated, in spite of the species' continuous distribution (Figure 1). It is unknown whether the barriers are pre-/postmating or environmental. Maternally inherited *Wolbachia* can cause reproductive incompatibility between populations of some Diptera, and a strain has been isolated from laboratory-reared *P. perniciosus* (Ono *et al*,

2001). However, there is no direct evidence for genetic barriers between populations of *P. perniciosus*.

So far, the northeastern group of populations of *P. perniciosus* has been found only in the valley of the River Ebro and the adjacent Mediterranean coastal plain. Much of this region is isolated from the south by a mountain range (Sistema Ibérico), and the Ebro valley has some distinctive climatic, floristic and faunistic characteristics (Montserrat and Montserrat, 1988; Ribera and Blasco-Zumeta, 1998). However, we sampled *P. perniciosus* from a range of bioclimates, altitudes and habitats in both the northeastern and southern regions (Table 1, Figure 1) without detecting any obvious associations between allele frequencies and environment.

Epidemiological significance of the regional populations

The low rates of inter-regional gene flow in *P. perniciosus* suggest that the population structure of this vector might constrain the dispersal of genes and traits of epidemiological importance. Reduced hybrid fitness in a 'tension zone' (Arnold, 1997) between the two regional populations could prevent the dispersal of each through their neighbour's range, as discussed for other European insects by Hewitt (1996, 1999).

There are two likely explanations for the low F_{ST} values between pairs of populations within each region: either some populations have become isolated, but there has been no significant change in allele frequencies in any population since the relatively recent postglacial dispersal (no drift, no natural selection, no evolution of new alleles); or no population has become permanently isolated, and there is continuing high gene flow between contiguous populations over distances up to 500 km. The latter seems more likely, although it could be difficult to reconcile with the supposedly 'weak flight capabilities' of sandflies (Cárdenas *et al*, 2001). Weak flight has been used to help explain the limited gene flow (estimated from alloenzyme data) between Colombian populations of *Lutzomyia longipalpis* (Morrison *et al*, 1993) or *Lutzomyia shannoni* (Cardenas *et al*, 2001; Mukhopadhyay *et al*, 2001). However, the present results are consistent with the knowledge that adults of *Phlebotomus (Larroussius)* species can travel up to 2.2 km in as little as 3 days during the Mediterranean summer (Killick-Kendrick *et al*, 1984). Our results suggest that leading-edge populations of *P. perniciosus* should be able to disperse relatively rapidly into northwest Europe, if climate warming provides suitable environmental conditions.

References

Anon (1965). *Atlas Nacional de España*. Instituto Geográfico y Catastral: Madrid.

Anon (1998). *Atlas Mundial Microsoft® Encarta® 99 (1995–1998)*. Microsoft Corporation, USA.

Anon (2000). *Global Gazetteer*. Presentation Copyright 1998–2000 by Falling Rain Genomics, Inc. (URL: http://www.calle.com/world/spain/index.html.)

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.

Arnold ML (1997). *Natural Hybridization and Evolution*. Oxford University Press: Oxford.

Benabdennbi I (1998). Contribution à l'étude de la biodiversité chez les *Larroussius* (Diptera:Psychodidae) du Bassin méditerranéen. Apports de l'analyse isoenzymatique. Thèse de Doctorat de l'Université Louis Pasteur de Strasbourg.

Cardenas E, Munstermann LE, Martinez O, Corredor D, Ferro C (2001). Genetic variability among populations of *Lutzomyia (Psathyromyia) shannoni* (Dyar 1929) (Diptera: Psychodidae: Phlebotominae) in Colombia. *Mem Inst Oswaldo Cruz* **96**: 189–196.

Cross ER, Hyams KC (1996). The potential effect of global warming on the geographic and seasonal distribution of *Phlebotomus papatasi* in south-west Asia. *Environ Health Perspect* **104**: 724–727.

Day JC, Ready PD (1999). Relative abundance, isolation and structure of phlebotomine microsatellites. *Insect Mol Biol* **8**: 575–580.

Donnelly MJ, Townson H (2000). Evidence for extensive genetic differentiation among populations of the malaria vector *Anopheles arabiensis* in Eastern Africa. *Insect Mol Biol* **9**: 357–367.

Esseghir S (1998). Phylogénie des phlébotomes méditerranéens vecteurs de *Leishmania major* et de *Leishmania infantum* et marqueurs moléculaires de populations. Thèse de Doctorat en Biologie, Faculté des Sciences, Université de Tunis.

Esseghir S, Ready PD, Ben-Ismail R (2000). Speciation of *Phlebotomus* sandflies of the subgenus *Larroussius* coincided with the late Miocene-Pliocene aridification of the Mediterranean subregion. *Biol J Linn Soc* **70**: 189–219.

Esseghir S, Ready PD, Killick-Kendrick R, Ben-Ismail R (1997). Mitochondrial haplotypes and phylogeography of *Phlebotomus* vectors of *Leishmania major*. *Insect Mol Biol* **6**: 211–225.

Gil Collado J, Morillas Marquez F, Sanchis Marin MC (1989). Los flebotomos en España. *Rev San Pub* **63**: 15–34.

Gou S, Thomson E (1992). Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48**: 361–372.

Goudet J (2001). FSTAT (version 2.9.3.1) *A program to estimate and test gene diversities and fixation indices*. (URL: http://www.unil.ch/izea/softwares/fstat.html.)

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.

Houin R (1965). Nouvelles localisations de phlébotomes en Espagne. *Ann Parasitol (Paris)* **40**: 341–363.

Jarne P, Lagoda PJL (1996). Microsatellites, from molecules to populations and back. *Trends Ecol Evol* **11**: 424–429.

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.

Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F, Collins FH (1996). Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity* **77**: 192–208.

Lehmann T, Blackston CR, Besansky NJ, Escalante AA, Collins FH, Hawley WA (2000). The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya: the mtDNA perspective. *J Heredity* **91**: 165–168.

Loxdale HD, Lushai G (1998). Molecular markers in entomology. *Bull Entomol Res* **88**: 577–600.

Mahamdallie S, Pesson B, Morillas-Marquez F, Perrotey S, Ready PD (2003). Adaptive variation in fringe populations of *Phlebotomus perniciosus*, an important vector of leishmaniasis in Europe. *Trans R Soc Trop Med Hyg* **97**: in press.

Martinez-Ortega E, Conesa-Gallego E (1987). Caracteres morfológicos de interés taxónomico de los flebotomos (Diptera, Psychodidae) de la Península Ibérica. *Anal Biol* **11**(Biol. animal, 3): 43–53.

Michalakis Y, Excoffier L (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**: 1061–1064.

Montserrat J, Montserrat G (1988). Hypothesis on the post-glacial dynamics of thermo-mediterranean plants on the southern slopes of the Pyrenees. *Monogr Inst Pir Ecol* **4**: 649–660.

Morillas Marquez M, Guevara Benitez DC, Ubeda Ontiveros JM, Gonzalez Castro J (1983). Fluctuations annuelles des populations de phlébotomes (Diptera, Phlebotomidae) dans la province de Grenade (Espagne). *Ann Parasitol Hum Comp* **58**: 625–632.

Morrison AC, Ferro C, Morales A, Tesh RB, Wilson ML (1993). Dispersal of the sandfly *Lutzomyia longipalpis* (Diptera: Psychodidae) at an endemic focus of visceral leishmaniasis in Colombia. *J Med Entomol* **30**: 427–435.

Mukhopadhyay J, Ghosh K, Ferro C, Munstermann LE (2001). Distribution of phlebotomine sandfly genotypes (*Lutzomyia shannoni* Diptera: Psychodidae) across a highly heterogeneous landscape. *J Med Entomol* **38**: 260–267.

Ono M, Braig HR, Munstermann LE, Ferro C, O'neill S (2001). *Wolbachia* infections of phlebotomine sandflies (Diptera: Psychodidae). *J Med Entomol* **38**: 237–241.

Raymond M, Rousset F (1995). GENEPOP (version 3.3, March 2001): an update version of GENEPOP (Version 1.2), *Population genetics software for exact test and ecumenism*. *J Heredity* **83**: 248–249.

Ready PD, Lainson R, Shaw JJ, De Souza AA (1991). DNA probes for distinguishing *Psychodopygus wellcomei* from *Psychodopygus complexus* (Diptera: Psychodidae). *Memórias do Instituto Oswaldo Cruz* **86**: 41–49.

Ribera I, Blasco-Zumeta J (1998). Biogeographical links between steppe insects in the Monegros region (Aragón, NE Spain), the eastern Mediterranean, and central Asia. *J Biogeogr* **6**: 969–986.

Rioux JA, Lanotte G (1990). *Leishmania infantum* as a cause of cutaneous leishmaniasis. *Trans R Soc Trop Med Hyg* **84**: 898.

Rioux JA, Rispail 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.

Rousset F (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* **145**: 1219–1228.

Schneider S, Roessli D, Excoffier L (2000). Arlequin ver. 2.000: *A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland (URL: http://anthro.umige.ch/arlequin.)

Schug MD, Wetterstrand A, Gaudette MS, Lim RH, Hutter CM, Aquadro CF (1998). The distribution and frequency of

microsatellite loci in *Drosophila melanogaster*. *Mol Ecol* **7**: 57–70.

Seccombe AK, Ready PD, Huddleston LM (1993). A catalogue of Old World phlebotomine sandflies (Diptera: Psychodidae, Phlebotominae). *Occ Pap Syst Ent* **8**: 1–57.

Weir BS, Cockerham CC (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Wright S (1978). *Evolution and the Genetics of Populations*, Vol. IV. Variability within and among natural populations. University of Chicago Press: Chicago.