MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 5309–5323 doi: 10.1111/mec.12018

Deep mitochondrial introgression and hybridization among ecologically divergent vole species

CRISTIANE BASTOS-SILVEIRA,*† SARA M. SANTOS,‡ RITA MONARCA,‡ MARIA DA LUZ MATHIAS†‡ and GERALD HECKEL§¶

*Museu Nacional de História Natural e da Ciencia, Universidade de Lisboa, Rua da Escola Politécnica, 58, 1269-102, Lisboa, Portugal, †Centre for Environmental and Marine Studies (CESAM), Universidade de Aveiro, 3810-193, Aveiro, Portugal, ‡Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016, Lisbon, Portugal, §Computational and Molecular Population Genetics (CMPG), Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012, Bern, Switzerland, ¶Swiss Institute of Bioinformatics, Genopode, 1015, Lausanne, Switzerland

Abstract

The completion of speciation is typically difficult to ascertain in rapidly diverging taxa but the amount of hybridization and gene flow in sympatry or parapatry contains important information about the level of reproductive isolation achieved. Here, we examined the progress in speciation between the Mediterranean (*Microtus duodecimcostatus*) and the Lusitanian pine vole (*M. lusitanicus*), which are part of the most rapid radiation of species known in mammals. These two Iberian pine voles are classified as separate species because of differences in morphology and ecology, but relatively many ambiguous individuals can be found in sympatric conditions. Our phylogenetic analyses of rangewide data from the mitochondrial cytochrome *b* gene (mtDNA) demonstrated high levels of diversity and a basal separation in two parapatric lineages. However, mtDNA affiliation was at odds with morphological classification or geographical distribution of the taxa. In contrast, statistical analyses of microsatellites (nucDNA) showed two clear genetic clusters in allopatry and sympatry generally matching morphological classification. This cytonuclear discordance over a large geographic area suggests historical introgression of mtDNA from *M. duodecimcostatus* to *M. lusitanicus*. There was statistical evidence for at least two recent hybrids in the sympatry zone but gene flow is apparently low given clear-cut differences in nucDNA. Our results indicate a relatively advanced speciation process in these Iberian pine voles without fully established reproductive isolation. This situation enables use of combined population genomic and experimental approaches for the separation of patterns and mechanisms in the ongoing explosive diversification of these and other Arvicoline rodents in the future.

Keywords: Iberian Peninsula, introgression, microsatellites, *Microtus*, mitochondrial DNA, speciation

Received 10 April 2012; revision received 27 July 2012; accepted 30 July 2012

Introduction

Speciation requires the termination of gene flow between taxa. This is less likely to occur in sympatric or parapatric conditions with ample opportunity for interbreeding than in situations where populations are separated in allopatric distribution ranges and reproductive isolation may evolve over time (Mallet *et al.* 2009; Petit & Excoffier 2009; Smadja & Butlin 2011). The secondary contact of such populations or taxa is a decisive step in which the progress of the speciation process is tested, and where (experimental) hybridization may provide means to determine the relative importance of differences in ecological, genetic, or phenotypic traits between taxa for reproductive isolation (Coyne & Orr

Correspondence: Cristiane Bastos-Silveira, Fax:+351213969784; E-mail: cbsilveira@fc.ul.pt

2004; Nosil *et al.* 2009; Maan & Seehausen 2011). However, divergence among nascent species is rarely consistent across different traits or parts of their genomes, and thus the position of recently diverged taxa in the speciation process is often difficult to assess (Petit & Excoffier 2009; Smadja & Butlin 2011; Beysard *et al.* 2012).

The genus *Microtus* represents probably the most rapid mammalian radiation resulting in more than 65 extant species over the course of only 1.2–2 Myr (Jaarola *et al.* 2004; Fink *et al.* 2010). Taxonomic classifications are particularly difficult in the genus owing to the speed of the *Microtus* radiation and often gradual variation in morphological and molecular traits between extant taxa (Mitchell-Jones *et al.* 1999). A recent molecular analysis of the radiation demonstrated the importance of geographic isolation in the genus with subradiations in North America, Asia and Europe and secondary colonization (Fink *et al.* 2010). Phylogeographical analyses have provided evidence of relatively deep divergence between parapatric evolutionary lineages within recognized species (Jaarola & Searle 2002; Brunhoff *et al.* 2003; Fink *et al.* 2004; Heckel *et al.* 2005). However, the taxonomic status of these lineages is often unclear and some may represent unrecognized species (e.g. *Microtus agrestis*: Hellborg *et al.* 2005; Beysard *et al.* 2012; *M. arvalis*: Heckel *et al.* 2005; Braaker & Heckel 2009).

The Mediterranean pine vole (*M. duodecimcostatus*, De Selys-Longchamps 1839) and the Lusitanian pine vole (*M. lusitanicus,* Gerbe 1879) are typically classified as separate species (Mitchell-Jones *et al.* 1999). Morphological discrimination is mainly based on the larger size and weight of *M. duodecimcostatus* (e.g. Cabrera 1914; Ellerman & Morrison-Scott 1951) but craniometrical analyses allow the identification of the taxa based on skulls and mandibles alone (Madureira 1981). These pine voles differ also in a range of ecological characters, such as the type of inhabited soils and their commitment to a life underground. The Lusitanian pine vole occurs in regions dominated by fertile, well-drained and easily mobilized soils (cambisols), and colder and moister climate compared with the Mediterranean pine vole, which occupies mainly grassland landscapes with higher abundance of stony soils (leptosols) and clayey soils with lower acidity (luvisols; Santos *et al.* 2009, 2010a). Both voles occupy preferentially open habitats, which may be agriculturally influenced (Cotilla & Palomo 2002; Mira & Mathias 2002), typically in close association with verges (Santos *et al.* 2011). They feed on roots, tubers, fruits and other parts of plants (Cotilla & Palomo 2002; Mira & Mathias 2002), but there are no data regarding differences in diet between these two species. The mating systems of the two taxa have not been studied in detail but limited data indicated social monogamy for *M. duodecimcostatus* (Paradis & Guédon 1993; Santos *et al.* 2010b) while *Microtus* species like most mammals are generally rather promiscuous or polygynous (Fink *et al.* 2006, 2007).

The distribution ranges are largely allopatric with *M. duodecimcostatus* in the southern-central Iberian Peninsula and southern France and *M. lusitanicus* in the north of Iberia including the French Pyrenees, but there is an extensive area of sympatry in parts of Portugal and Spain (Madureira 1984; Mitchell-Jones *et al.* 1999; Cotilla & Palomo 2002; Mira & Mathias 2002; Fig. 1). There is no information regarding natural hybridization

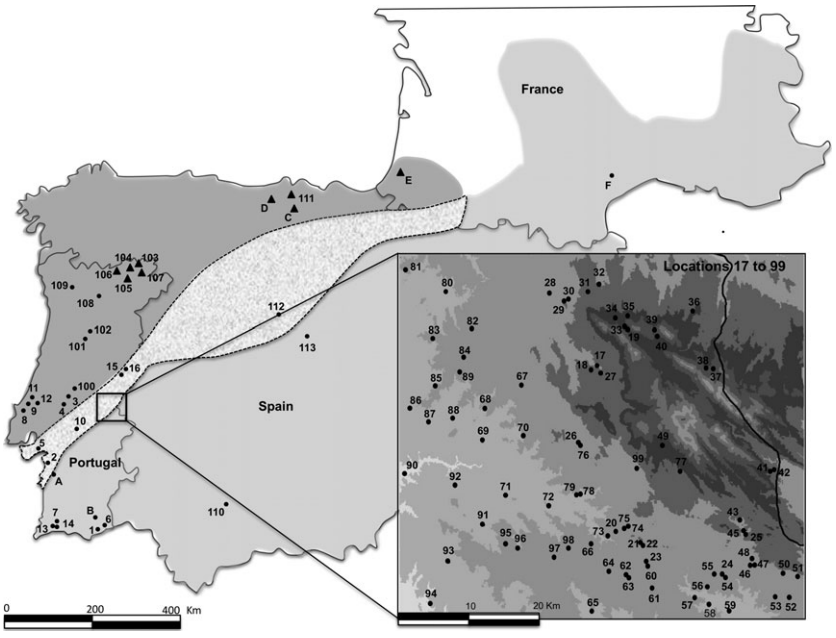

Fig. 1 Map of the ranges of the southwestern European endemic rodents *Microtus lusitanicus* (dark grey shading) and *M. duodecimcostatus* (light grey shading). The area of sympatric occurrence according to morphological data is delimited by the dashed lines. Dots and triangles indicate the sampling sites for mitochondrial and nuclear analyses of the present study, which are cross-referenced with Table 1. Capital letters indicate the geographical origin of *cytb* sequences from GenBank: *M. lusitanicus* from France (E) and Spain (C, D), and *M. duodecimcostatus* from France (F) and Portugal (A, B). Black triangles indicate sampling sites with *M. lusitanicus* mtDNA haplotypes; dots those with *M. duodecimcostatus* mtDNA.

between these two taxa, but it is possible to cross them under laboratory conditions (Wiking 1976).

Phylogenetic analyses of the *Microtus* genus based on the mitochondrial *cytochrome b* gene revealed 4–5% sequence divergence between *M. duodecimcostatus* and *M. lusitanicus* individuals sampled in the allopatric parts of the distribution ranges (Jaarola *et al.* 2004; see Fig. 1). This divergence is at the lower end of the values between recognized *Microtus* species and at the upper end of the divergence between allopatric lineages within some other species (e.g. Fink *et al.* 2004; Jaarola *et al.* 2004). Phylogeographic analyses of *M. duodecimcostatus* and *M. lusitanicus* are lacking, and without data from the area of sympatry it is impossible to distinguish whether genetic divergence between the two taxa is only owing to the potential effect of isolation by distance between allopatric genetic lineages within a species (see e.g. Heckel *et al.* 2005) or owing to evolutionary divergence between reproductively isolated species.

'Molecular taxonomy' based on a single genetic marker like mitochondrial DNA (mtDNA) may be misleading or remain inconclusive because of a marker-specific evolutionary history (Ballard & Whitlock 2004) or because taxa continue to exchange genes or share polymorphisms predating their separation (Petit & Excoffier 2009). Thus, in a scenario of low divergence and potential hybridization such as in Iberian *Microtus*, only the combined analysis of molecular markers with different speed of evolution and different modes of inheritance may be powerful enough to characterize the extent of overall divergence between taxa, detect historical and recent hybridization, and reveal the direction of introgression (e.g. Berthier *et al.* 2006; Good *et al.* 2008; Renoult *et al.* 2009). The aims of the present study were thus (i) to assess the extent of genetic divergence between the two pine vole taxa in both the allopatric and sympatric areas of the distribution ranges; (ii) to examine the concordance between species delimitations based on morphology, mitochondrial and nuclear DNA; and (iii) to test for evidence of hybridization between the morphologically defined and ecologically divergent taxa in the area of sympatry.

Material and methods

Samples and DNA extraction

The present study is based on 296 samples from 113 locations spanning almost the entire distribution ranges of *M. duodecimcostatus* and *M. lusitanicus* in south-western Europe with a particular focus on the area of sympatry (Portalegre district, Portugal; Fig. 1; Table 1). Most of the samples ($N = 283$) came from animals that were captured using subterranean live traps of multiple captures ('mole type'). Global positioning system (GPS) was used to register capture locations. Tissue samples (tail; muscle from dead animals) were stored in absolute ethanol at −20 °C.

Species identification of the live-trapped animals in the field was based on morphological characteristics: hind feet length (*M. lusitanicus* < 16 mm; *M. duodecimcostatus* > 16 mm) and body length (*M. lusitanicus* < 105 mm and *M. duodecimcostatus* > 89 mm). Coat colour patterns were additionally used for classification of 16 of 197 voles from the sympatric area presenting feet and body length in the overlapping range (Madureira 1984). Individuals were further classified as adult or subadult/juvenile based on external signs of reproductive activity (testes visible, pregnancy or lactation signs). Based on these morphological traits, our data set included fresh samples from 127 *M. duodecimcostatus* (including 49 adults), 104 *M. lusitanicus* (30 adults) and 52 (33 adults) further individuals for which assignment to either species based on morphology was difficult (hereafter termed *Microtus* sp). Thirteen additional samples came from the Tissue and DNA Collection of the Museu Nacional de História Natural e da Ciência (MB), Portugal, the Museo Nacional de Ciencias Naturales (MNCN), Spain, and one dry skin from the vertebrate collection of the Estación Biológica Doñana (EBD), Spain. Genomic DNA was extracted using standard protocols requiring incubation with SDS and digestion with proteinase K, followed by a phenol–chloroform DNA extraction (Sambrook *et al.* 1989).

DNA sequencing

The complete mitochondrial cytochrome *b* (*cytb)* gene (1140 bp) was amplified and sequenced using *Microtus* specific primers L14727-SP and H-15195-SP (Jaarola & Searle 2002). All reactions contained 100 ng of template DNA, 0.3 mM of each primer, 1U of *Taq* DNA Polymerase, 1× buffer, 2.5 mM $MgCl_2$ and 0.2 mM of each dNTP to a final volume of 25 μL. DNA amplifications were performed using the following conditions: denaturation at 95 °C for 5 min, followed by 40 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 1 min and extension at 72 °C for 1 min. An extension step at 72 °C for 10 min was added at the end. We purified PCR products and directly sequenced on an ABI 3100 Genetic Analyzer (Applied Biosystems, USA).

Sequence analysis

The *cytb* sequences were aligned and translated into amino acids using Sequencher™ 4.8. To ensure the mitochondrial origin of the sequences, they were checked for the presence of stop codons and compared against GenBank entries of *M. duodecimcostatus* and *M. lusitanicus*.

Table 1 Sampling sites (country, location and coordinates) of *Microtus* sp (M), *M. lusitanicus* (ML) and *M. duodecimcostatus* (MD) used for this study. Map reference numbers are listed as displayed in Fig. 1 and samples locations within the sympatric area are marked with x. Given are numbers of samples analysed for microsatellites and *cytb*, and *cytb* haplotypes

Map ref.	Location	UTMx	UTMy	MorphoID	Microsat	Cyt*b*	Cyt*b* haplotypes	Accession numbers
Portugal								
1	Monte Gordo*	608312	4407400	MD	2	–	–	
2	Setubal*	475998	4316844	MD	1	–	–	
3	Almerin*	531953	4340281	ML	1	–	–	
4	Boquilobo*	540084	4359595	ML	2	–	–	
5	Pancas*	475998	4316844	ML	1	–	–	
6	Castro Marin*	637945	4120515	MD	1	–	–	
7	Silves	549722	4116257	MD	1	–	–	
8	Obidos	483192	4355308	ML	2	1	R0QP	JX424119
9	Caldas	490897	4364192	ML	2	1	R0SJ	JX424120
10	Cabeção	578654	4313479	M	3	1	ID006	JX424121
11	Alcobaça	503321	4377793	ML	1	1	ID010	JX424122
12	Candeeiros	506672	4364861	ML	1	1	ID011	JX424123
13	Porto Lagos	541831	4116913	MD	7	3	ID015.1, ID015.2	JX424124-125
14	Pomar jovem	540453	4120826	MD	6	2	ID016.1, ID016.2	JX424126-127
15	Medelim	656188	4432895	MD	5	4	ID018.1, ID018.2, ID018.3	JX424128-129 JX424130
16	Medelim	653227	4429596	M	1	2	ID019.1, ID019.2	JX424131
17(x)	Rib Nisa	632170	4358562	ML	2	1	ID025	JX424133
18(x)	Matinha	631490	4358186	ML	5	2	ID026	JX424134
19(x)	Barretos	635141	4362647	ML	3	1	ID027	JX424135
20(x)	Microcolonia	634379	4340922	MD	2	1	ID030	JX424136
21(x)	Vacas	635287	4341222	MD	3	1	ID031	JX424137
22(x)	Ilha	637362	4339325	MD	1	–	–	
23(x)	Plantação sobreiro	637817	4337196	MD	7	1	ID034	JX424138
24(x)	Mosteiros	646355	4336018	MD	2	1	ID035	JX424139
25(x)	Barulho/Besteiros	648442	4340706	ML	1	1	ID036	JX424140
26(x)	Subestação EDP	630222	4350230	MD	4	3	ID038.1, ID038.2	JX424141-142
27(x)	Castanheira	632629	4357811	ML	2	1	R032	JX424143
28(x)	Pegos Dobrados	626910	4366208	ML	2	1	R041	JX424144
29(x)	Lavradores I	628547	4365368	M	2	1	R042	JX424145
30(x)	Lavradores II	628984	4365546	M	1	1	R041	JX424144
31(x)	Tinhosa	631137	4366362	ML	3	2	ID027, R044	JX424135-146
32(x)	Canto das Nogueiras	632242	4367129	ML	1	1	R045	JX424147
33(x)	Est Vale Rodao I	634160	4363570	ML	3	1	ID027	JX424135
34(x)	Est Vale Rodao II	634134	4363581	ML	2	2	ID036, R049	JX424140-148
35(x)	Amieira	635370	4363765	ML	2	1	R052	JX424149
36(x)	St. Ant. Areias	642471	4364401	ML	4	1	R045, R057	JX424147-150
37(x)	Penha Esparoeira	644781	4358273	ML	5	1	R059	JX424151
38(x)	Braçais	644077	4358333	ML	1	1	R062	JX424152
39(x)	Marvão	638378	4362395	ML	2	1	R049	JX424148
40(x)	Marvão/Jardim	638717	4361656	ML	6	1	R065	JX424153
41(x)	Rabaça I	651434	4347527	ML	1	1	R070	JX424154
42(x)	Rabaça II	651510	4347570	ML	2	2	R071.1, R071.2	JX424155-156
43(x)	Vale das Púdicas I	647818	4342220	M	3	2	R075	JX424157
44(x)	Vale das Púdicas Ia	647825	4342196	ML	2	2	R044, R075	JX424146-157
45(x)	Vale das Púdicas II	647794	4342125	M	4	–	–	
46(x)	Monte dos Moços I	649345	4337370	MD	1	1	R082	JX424158
47(x)	Monte dos Moços II	649375	4337423	MD	1	1	R082	JX424158
48(x)	Azenha nova	648395	4340704	ML	3	3	ID036, R082, R085	JX424140-158-159
49(x)	Reguengo	639362	4350070	ML	4	3	R045, R071.1	JX424147-155
50(x)	Serrinha	654187	4336171	MD	1	1	ID034	JX424138
51(x)	Nave fria	652562	4336594	ML	5	1	R086	JX424160
52(x)	Figueira	653245	4333969	MD	5	5	R092	JX424161
53(x)	Martin Tavares	651812	4334034	MD	2	1	ID035	JX424139

Table 1 *Continued*

Map ref.	Location	UTMx	UTMy	MorphoID	Microsat	Cyt*b*	Cyt*b* haplotypes	Accession numbers
54(x)	Ronseiras	646028	4336337	MD	2	2	R094.1, R094.2	JX424162-163
55(x)	Algeireirinhas	645053	4336337	MD	3	1	R095	JX424164
56(x)	Freirinha	644428	4334998	MD	2	–	–	
57(x)	Monte Faia	643064	4333813	MD	2	1	R097	JX424165
58(x)	Fragosa	644564	4333044	MD	3	1	R098	JX424166
59(x)	Arronches	646723	4332355	MD	1	1	–	
60(x)	Monte Ferrador	637684	4337575	MD	3	1	R100	JX424167
61(x)	Assmar	638395	4334883	MD	2	2	R101.1, R101.2	JX424168-169
62(x)	Monte Alcaide I	635566	4336187	MD	2	1	R102	JX424170
63(x)	Monte Alcaide II	635514	4336204	MD	4	1	R103	JX424171
64(x)	Escudeiros	633684	4336569	MD	2	1	R104	JX424172
65(x)	Monte Évora	631843	4332291	MD	4	1	R105	JX424173
66(x)	Sobreiral	631756	4339455	MD	4	1	R106	JX424174
67(x)	Cerejeira-Alagoa	623986	4356367	ML	11	1	R114	JX424175
68(x)	Monte Alegre	620029	4353824	MD	1	–	–	
69(x)	Monte Romeiras	619762	4350322	ML	1	–	–	
70(x)	Sra Martires	624167	4351032	MD	1	1	R118	JX424176
71(x)	Abodaneira	622447	4344566	MD	2	–	R120	JX424177
72(x)	Cabeco Malhadio	627075	4343453	MD	2	1	R121	JX424178
73(x)	Inglesinha	633496	4340295	MD	7	2	R082, R122	JX424158-179
74(x)	Monte Bonito I	635633	4341489	M	4	1	R041	JX424144
75(x)	Monte Bonito II	635598	4341478	M	5	1	R045	JX424147
76(x)	Subestacao EDP	630226	4350175	M	4	–	R125	JX424180
77(x)	Penha hortas	641233	4347376	ML	1	1	R071.1	JX424155
78(x)	Entreribeiras I	630014	4344679	MD	1	1	R128	JX424181
79(x)	Entreribeiras II	630001	4344689	MD	1	1	R129	JX424182
80(x)	Pedreira	615705	4366292	ML	2	1	R131	JX424183
81(x)	Miguel Pestana Bastos	611255	4368583	M	3	1	R133	JX424184
82(x)	Alpalhao	618518	4362286	M	1	–	–	
83(x)	Gafete Cortesoes	614282	4361253	M	1	–	–	
84(x)	Vale Peso	617646	4359268	M	2	–	–	
85(x)	Campainhas	614628	4356117	M	2	1	R137	JX424185
86(x)	Atalainha	611880	4353774	M	4	1	R138	JX424186
87(x)	Ponte velha	613934	4352313	M	3	1	R139	JX424187
88(x)	Flor Rosa	616555	4352673	M	1	1	R141	JX424188
89(x)	Vale Vaqueiros	617312	4357663	M	4	1	R142	JX424189
90(x)	Ponte Rib Seda	611417	4346845	M	1	1	R143	JX424190
91(x)	Coutada Pina	619989	4341421	MD	1	1	R147	JX424191
92(x)	Rib Linhais	616839	4345535	MD	5	–	–	
93(x)	Outeiro Ferreiros	616223	4337419	MD	3	1	R151	JX424192
94(x)	Monte Marinheira	614359	4332957	MD	2	1	R153	JX424193
95(x)	Silveira I	622389	4339454	MD	6	1	R160	JX424194
96(x)	Silveira II	623698	4338902	MD	7	1	R161	JX424195
97(x)	Ribeiro Freixo	627679	4338033	MD	1	1	R162	JX424196
98(x)	Cabeca Alta	629341	4338957	MD	1	1	R163	JX424197
99(x)	Portalegre	636696	4347623	ML	2	1	R167	JX424198
100	Tomar	559328	4387052	ML	15	3	R200.1, R200.2	JX424199-200
101	Senhorim	598925	4488440	ML	1	–	–	
102	Germil	606428	4499718	ML	2	1	Ri050	JX424201
103	Nogueira	681967	4625688	ML	1	1	RI052	JX424202
104	Santuario	678155	4624532	ML	1	1	Ri053	JX424203
105	Ervedosa	662054	4617841	ML	3	3	Ri057.1, Ri057.2, Ri057.3	JX424204-205 JX424206
106	Rebordelo	655770	4626801	ML	1	1	Ri058	JX424207
107	Grijo da Parada	690830	4620892	ML	1	1	Ri060	JX424208
108	Alijó	627718	4574842	ML	1	1	Ri064	JX424209
109	Rio Longo	570953	4605326	ML	1	1	Ri075	JX424210

Table 1 *Continued*

Map ref.	Location	UTMx	UTMy	MorphoID	Microsat	Cyt*b*	Cyt*b* haplotypes	Accession numbers
Spain								
110	Cordoba*	343687	4193264	MD	1	–	–	
111	Ventorrillo*	411890	4769087	M	1	–	–	
112	Segovia†	404784	4532876	MD	1	1	DS3	JX424211
113	Madrid†	441665	4470636	MD	1	1	DS3	JX424211
Total					295	122	93	

*Samples from MNHNC-Portugal.
†Samples from Museo Nacional de Ciencias Naturales-Spain.

Six additional *cytb* sequences, from specimens captured in France (AJ717744 and AJ717746), Spain (AY513812 and AY513813) and Portugal (AY513796 and AY513797), were included in our data set to complement the range-wide coverage of our analyses (see Fig. 1). The identity and frequency of haplotypes in the data set was determined with ARLEQUIN 3.1 (Excoffier *et al.* 2005). Net average divergence (Da) (Nei & Li 1979) between mtDNA clusters was estimated using the Tamura–Nei model (Tamura & Nei 1993) in MEGA 5.05 (Tamura *et al.* 2011). Standard errors were estimated by the bootstrap method using 10 000 replicates. Redundant haplotypes were removed for phylogenetic analyses.

Phylogenetic relationships among *cytb* haplotypes were reconstructed analogous to the analyses described in Braaker & Heckel (2009). Briefly, phylogenetic analyses were performed using the Neighbour-Joining method implemented in MEGA 5.05 (Tamura *et al.* 2011) and Bayesian algorithms implemented in MRBAYES 3.1.2 (Ronquist & Huelsenbeck 2003). The closely related species *M. gerbei* (Jaarola *et al.* 2004) was used as outgroup to root the phylogenetic trees (AY513800). JMODELTEST 0.0.1 (Posada 2008) was used to select the best-fitting model of nucleotide substitution. Different selection criteria (Akaike Information Criterion and Bayesian Information Criterion) identified the same model: the general time-reversible substitution model with a proportion of invariable sites and a gamma-distributed rate variation across sites (GTR + I + G; Tavaré *et al.* 1997). The estimated parameters of the model were as follows: substitution rate matrix [AC] = 1.1368; [AG] = 23.9209; [AT] = 0.4109; [CG] = 1.1577; [CT] = 9.4057 and [GT] = 1.0000; gamma shape parameter = 0.6900, with 72.8% of invariable sites; and nucleotide frequencies of A = 0.3079; C = 0.3041; G = 0.1311 and T = 0.2568.

The Bayesian analysis was performed three times using the GTR + I + G model, each analysis starting from two different random trees for two million generations with every 100th generation sampled. Each run used one cold and three heated chains simultaneously. For each analysis, the average standard deviation of split frequencies at completion between the independent runs was compared with check for convergence on a stationary distribution (Ronquist & Huelsenbeck 2003). The first 25% of the trees were discarded as burn-in and the remaining trees were used to reconstruct a consensus and estimate Bayesian posterior probabilities.

Microsatellite genotyping

Nuclear DNA variability of 295 individuals was assessed by genotyping 13 microsatellite loci: MM1 and MM2 (Ishibashi *et al.* 1999), CRB5 and CRB7 (Ishibashi *et al.* 1995), MAG6 and MAG25 (Jaarola *et al.* 2007), MAR3, 12, 16, 63, 76, 80 and AVP-INb (Walser & Heckel 2008; Hahne *et al.* 2011). PCR amplification was performed using the Qiagen Multiplex Kit according to the protocol described in Braaker & Heckel (2009). Fragment separation was carried out on an ABI 3100 sequencer and the genotypes were scored using GENEMAPPER software 3.7 (Applied Biosystems) against the internal LIZ 500 size standard. Five per cent of the genotypes were independently reamplified and scored to ensure repeatability and quality of genotyping.

Microsatellite analysis

The software CONVERT 1.2 (Glaubitz 2004) and FORMATOMATIC 0.8.1 (Manoukis 2007) were used to prepare the input files for all software used for microsatellite data analyses. ARLEQUIN 3.1 (Excoffier *et al.* 2005) was used to calculate general diversity estimates such as the number of alleles per locus, observed (H_o) and expected heterozygosity (H_e).

Principal component analysis (PCA), performed in PCAGEN (Goudet 1999), was used as an exploratory method to visualize patterns of multilocus genotypic differentiation among individuals without using information regarding their morphology-based classification as *M. duodecimcostatus*, *M. lusitanicus* or *M.* sp. Further, several Bayesian methods were applied to investigate possible natural hybridization at the Alto Alentejo sympatric area. First, we employed an individual-based

approach implemented in GENELAND 2.0.12 (Guillot *et al.* 2005) aiming to determine the number and distribution of genetic clusters present in the sympatric area in consideration of the spatial coordinates where voles were trapped. We inferred the most probable number of genetic clusters (K) in the data set by performing 500 000 Markov chain Monte Carlo (MCMC) iterations with a maximum rate of Poisson processes fixed to 200, which is a value close to the number of individuals in our data set, as suggested by Guillot *et al.* (2005). K was allowed to vary between 1 and 10, and the uncertainty attached to spatial coordinates was specified as 0.01. We used 600 as the maximum number of nuclei in the Poisson–Voronoi tessellation (three times the maximum rate of Poisson processes) as suggested by Guillot *et al.* (2005), and an independent Dirichlet distribution model for allele frequencies. To check the consistency of the results, the MCMC was run 10 times. Once K was determined, we performed a second series of 100 independent runs using the same parameters but K fixed to the value identified in the first step. The average logarithm of the posterior probability was computed for each of the 100 runs, and the 10 runs with the highest mean posterior probability values were kept for inference.

Additionally, we applied two Bayesian methods, which are tailored to identify hybrid individuals but use different approaches. STRUCTURE 2.3.2 (Pritchard *et al.* 2000) assigns probabilities for individuals to have recent ancestry in two or more populations, while NEWHYBRIDS 1.1 (Anderson & Thompson 2002) estimates the probability of individuals belonging to distinct hybrid or pure-bred classes.

STRUCTURE was run with ten repetitions of 100 000 MCMC iterations following a burn-in period of 20 000 steps, the admixture model, correlated allele frequencies and no prior population information. The number of clusters (K = 1 to 10) was tested according to Evanno *et al.* (2005). For the optimal K, we assessed the average proportion of membership (Q_i) of each individual to the inferred clusters and assigned them based on threshold values of the individual membership proportion (q_i). NEWHYBRIDS computes the posterior probability of a multilocus genotype for being either of pure origin for one of two parental classes or for belonging to one of four potential hybrid classes arising from two generations of interbreeding. For our study, the six classes were pure *M. duodecimcostatus*, pure *M. lusitanicus*, F1 hybrid, F2 hybrid, backcross with *M. duodecimcostatus* and backcross with *M. lusitanicus*. For the analyses, individuals were not classified a priori by morphological species identification. NEWHYBRIDS was run ten times using the default parameters with a uniform prior for the six classes, a burn-in phase of 10 000 steps and 100 000 further MCMC sweeps as suggested by the authors. The posterior probability for each individual and the hybrid categories estimated by NEWHYBRIDS were used to generate Q-values according to Berthier *et al.* (2006). In our study, the Q-value represents an overall estimate of the proportion of the genome of an individual being of pure *M. lusitanicus* ancestry.

An intrinsic issue with Bayesian analyses is that validity of the assumed distribution of prior(s) cannot be assessed statistically; thus, simulations were used for our data set to evaluate the statistical power for correctly identifying the individuals as parentals or hybrids (Nielsen *et al.* 2006). We used the software HYBRIDLAB 1.0 (Nielsen *et al.* 2006), which generates multilocus hybrid genotypes based on given parental populations. The parental genotypes consisted of individuals morphologically classified as *M. duodecimcostatus* and *M. lusitanicus* and assigned to their species cluster ($q > 0.95$) using STRUCTURE. From this pool, we randomly selected 50 individuals from each species to generate 100 genotypes of each parental and hybrid class (F_1, F_2, $F_1 \times$ *M. duodecimcostatus*, $F_1 \times$ *M. lusitanicus*). This procedure was replicated 10 times and STRUCTURE was run with $K = 2$ on the generated genotypes. The efficiency and accuracy of the assignment of the individuals was calculated as suggested by Vähä & Primmer (2006), using the threshold values of 0.05, 0.10, 0.20 and 0.30. The same set of simulated genotypes was used in NEWHYBRIDS for analogous analyses.

Results

Genetic variation and taxonomic classification based on mtDNA

Sequencing of the complete *cytb* of 122 individuals yielded 93 different haplotypes defined by 186 polymorphic and 109 parsimony informative sites. We found 42 haplotypes in 54 (77%) sequenced individuals morphologically classified as *M. duodecimcostatus*, and 38 in 52 (73%) *M. lusitanicus*, with nucleotide diversities of 0.017 and 0.023, respectively. Haplotypes were only shared by individuals from the same locality or from geographically close sites (Table 1). GenBank accession numbers are given in Table 1.

Phylogenetic reconstructions showed that mtDNA haplotypes do not segregate overall according to morphological classification of individuals into *M. lusitanicus* and *M. duodecimcostatus* (Fig. 2). The topology of the phylogenetic tree is relatively shallow with a most remarkable and well-supported basal split separating the sequences of the two taxa from a previous phylogenetic analysis (Jaarola *et al.* 2004) as well as the new sequences (net divergence: 3.3%). Only six new haplo-

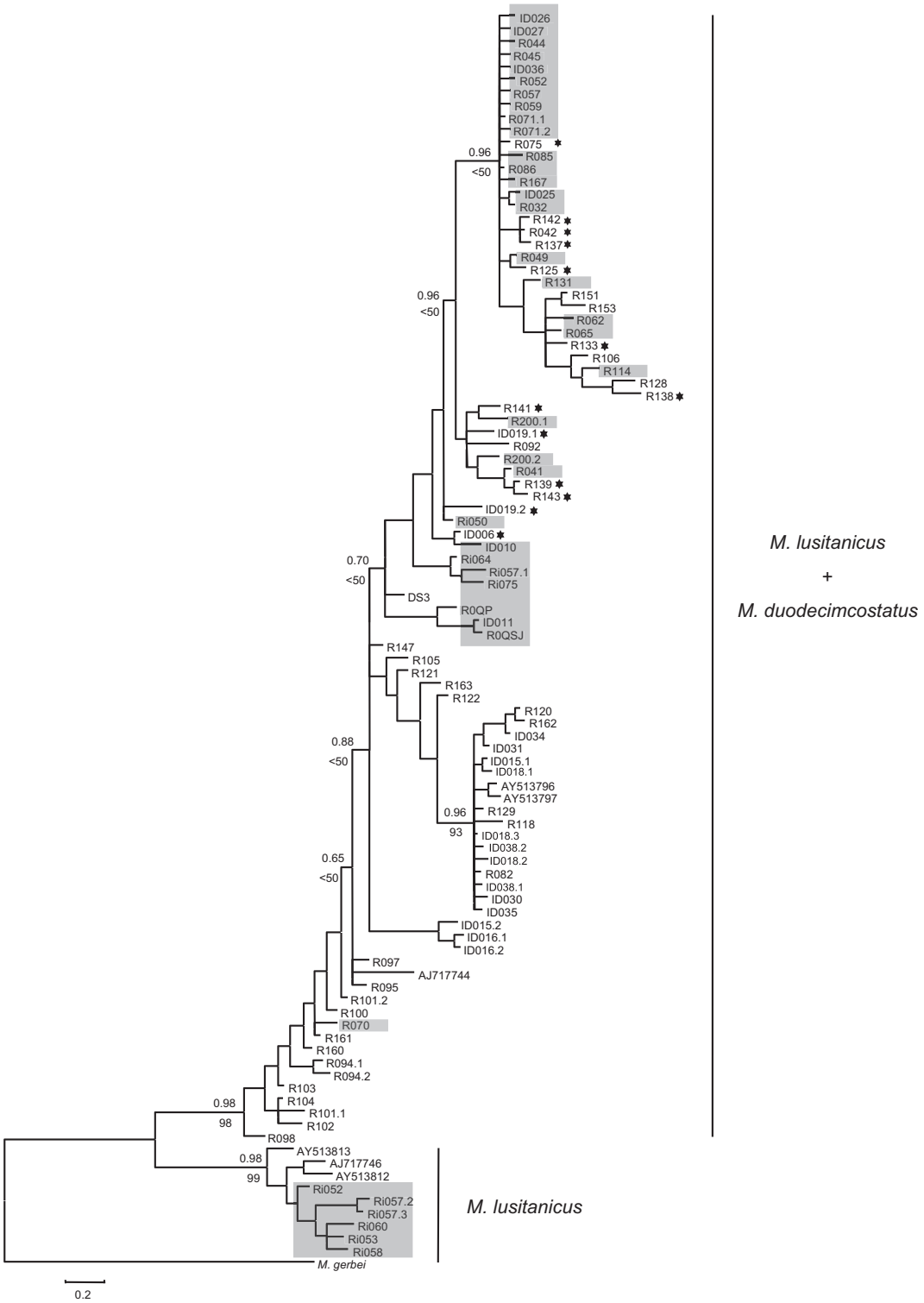

Fig. 2 Reconstruction of phylogenetic relationships of *cytb* haplotypes from individuals morphologically classified as *Microtus lusitanicus* or *M. duodecimcostatus* with *M. gerbei* as outgroup. Haplotype labels are according to Table 1. Bootstrap values from Neighbour-Joining algorithms and Bayesian posterior probabilities are listed above and below the main nodes, respectively. Morphological identification of individuals from this study is indicated as follows: *Microtus* sp (*), *M. lusitanicus* (grey shadow) and *M. duodecimcostatus* (blank).

types (Ri052, Ri053, Ri057.2, Ri057.3, Ri058, Ri060) clustered with the *M. lusitanicus* sequences from Spain and France (locations C, D, E in Fig. 1). The according six individuals were captured at the northernmost sampling sites in Portugal (Table 1). All remaining haplotypes clustered with *M. duodecimcostatus* sequences from France and Portugal (locations A, B, F in Fig. 1), regardless of their morphological classification or of their origin in the allopatric or sympatric parts of the distribution ranges.

Genetic structure and geographical patterns based on nuclear DNA

Genetic variation at the 13 microsatellite loci was high with a mean number of alleles per locus of 15.7 across all samples, and individual loci showing between three (MAR016) and 34 (AVP-INb) alleles. Morphologically identified *M. duodecimcostatus* (N = 132) had a mean of 12.7 alleles (range 3–28) and *M. lusitanicus* (N = 108) had on average 13.2 alleles per locus (range 3–28). Mean observed and expected heterozygosity were 0.5 and 0.61, respectively, for *M. duodecimcostatus* and 0.63 and 0.74 for *M. lusitanicus*. Considering the distribution of our samples across most of the distribution ranges of the two taxa, the deficit in observed heterozygosity can be explained by the presence of several populations.

Principal component analysis provided overall evidence of two distinct genetic clusters in our nuclear data set largely consistent with the morphological classification of individuals in *M. duodecimcostatus* and *M. lusitanicus*. In the analysis of the complete data set including individuals from the allopatric and sympatric parts of the ranges, PCA axis 1 explained 17.8% of the variability and axis 2 3.9% (Fig. 3A). Axis 1 separated genotypes from most morphologically identified *M. duodecimcostatus* and *M. lusitanicus* irrespective of their sympatric or allopatric origin. Individuals that were morphologically not assignable to either taxon in the field (N = 51; including 18 juveniles or subadults) clustered all with *M. lusitanicus* genotypes. Furthermore, 14 individuals identified as *M. duodecimcostatus* based on morphology clustered also within the *M. lusitanicus* group. Given that discrimination between the two taxa in the field is mostly based on size-related traits with partially overlapping distribution, we hypothesized that large individuals of the smaller taxon (*M. lusitanicus*) might have been wrongly classified as the larger taxon (*M. duodecimcostatus*) owing to age or geographical variation. A new PCA

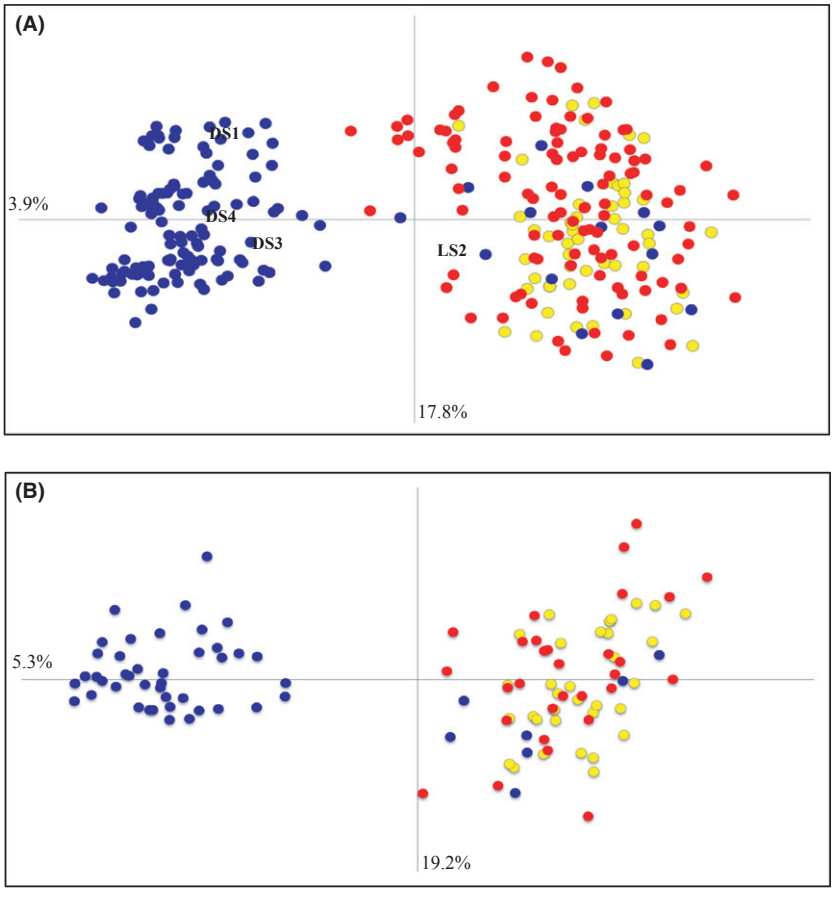

Fig. 3 Principal component analysis (PCA) performed on microsatellite genotypes of morphologically determined *Microtus duodecimcostatus* (blue), *M. lusitanicus* (red) and individuals of undetermined species (yellow). Shown are the first two principal component axes (PC1 and PC2) and the variance explained. (A) PCA including all individuals from the allopatric and sympatric parts of the ranges. Individuals from Spain are indicated by their ID number and the corresponding geographical localities in Fig. 1 are as follows: DS1 (110), LS2 (111), DS3 (112) and DS4 (113). (B) PCA including only adult individuals from the sympatric area.

based only on adult individuals with clear signs of reproductive activity from the sympatric area (N = 112) showed that there were still seven individuals that had *M. duodecimcostatus* morphology but genotypes similar to *M. lusitanicus* (Fig. 3B).

The integration of genetic and spatial information in GENELAND supported the presence of two distinct clusters in nuclear DNA in the sympatric area with a relatively complex geographical distribution (Fig. 4). All posterior distributions peaked at K = 2 with very little support for other K values. The 10 best runs with K = 2 converged to a geographical pattern where the two genetic clusters occupy ranges with a complex omega-shaped line of genetic discontinuity, which matches only loosely the distribution of soil types in the region (Fig. 4). All individuals, except one (R087-1) at the borderline, were assigned to one of the two clusters with posterior probabilities exceeding 0.9. This was generally consistent with morphology, but in agreement with PCA results, the 14 individuals classified morphologically as *M. duodecimcostatus* were assigned to the *M. lusitanicus* population. Morphologically defined *M. lusitanicus* were found north of the genetic discontinuity line whereas most morphologically defined *M. duodecimcostatus* were present south of it.

The Evanno *et al.* (2005) approach for the STRUCTURE analyses strongly suggested two genetic clusters in the sympatric area. All individuals identified as *Microtus* sp. and *M. lusitanicus* were assigned to cluster I, resulting in an average proportion of membership of Q = 0.994 and Q = 0.987, respectively (Fig. S1A, Supporting information). *M. duodecimcostatus* were overall assigned to cluster II with Q_i = 0.857. At the individual level, the 14 voles classified morphologically as *M. duodecimcostatus* were assigned to cluster I with q > 0.99, in agreement with PCA and GENELAND results. After removal of these individuals, the Q-value of *M. duodecimcostatus* rose to 0.98. NEWHYBRIDS showed for 202 of 227 voles posterior probabilities larger than 0.95 for having a pure genotype from one of the two taxa. In agreement with the analyses above, 45 of 47 *Microtus* sp. were likely *M. lusitanicus*, and 12 of 14 *M. duodecimcostatus* had posterior probabilities exceeding 0.95 for being pure *M. lusitanicus* (Fig. S1B, Supporting information).

Overall, there was profound unidirectional cytonuclear discordance in the data set, which extends north beyond the sympatric area far into the *M. lusitanicus* range (see Fig. 1). Seventy-six individuals carried *M. duodecimcostatus* mtDNA and *M. lusitanicus* nuclear DNA, but the reverse combination was not detected. Forty individuals were classified as *M. duodecimcostatus* for both genomes and six voles were *M. lusitanicus* for both marker types.

Evidence of recent hybridization in the sympatric area

Genotype simulations with HYBRIDLAB indicated highest robustness of hybrid identification at q = 0.20 (overall performance was 96% for parental individuals and 94% for hybrids; Fig. S2, Supporting information). At this threshold, the efficiency of assigning simulated individuals with

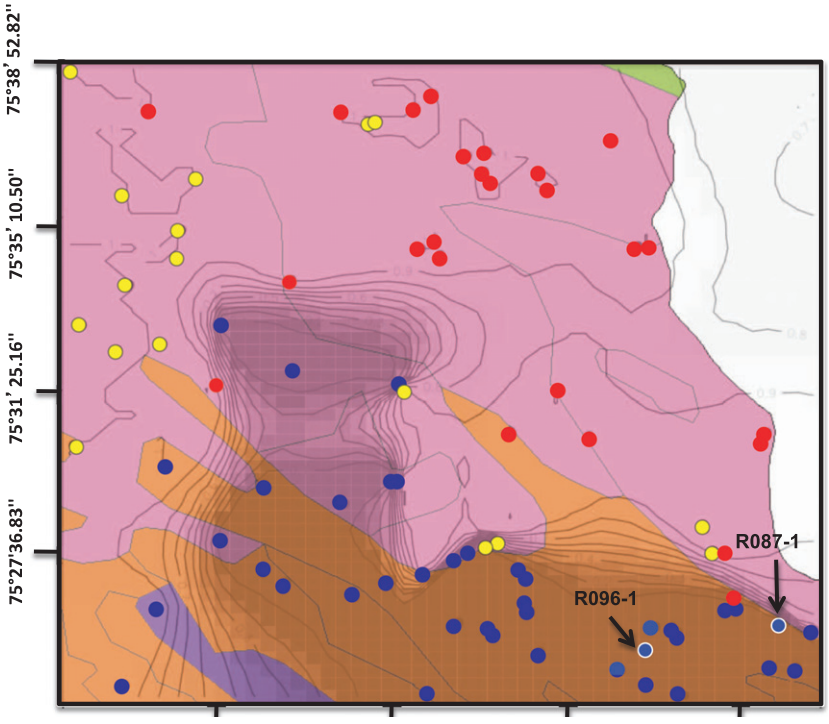

Fig. 4 Map of genetic clustering in the sympatric area of *Microtus lusitanicus* and *M. duodecimcostatus* generated by Geneland overlaid on a map showing the types of soil in the study area (pink = cambisol; orange = luvisol; green = leptosol; purple = vertisol). Dots represent sampling sites (see Fig. 1), and colours represent morphological identification as in Fig. 3: *Microtus* sp. (yellow); *M. lusitanicus* (red) and *M. duodecimcostatus* (blue). The probability of nucDNA class membership is indicated with brightness: morphologically identified *M. duodecimcostatus* were mostly found in the area covered by the dark-shaded (southern) cluster. The arrows indicate the capture location of the two voles identified as likely hybrids (see text). Note that all individuals shown on this figure harboured *M. duodecimcostatus*-like mtDNA.

STRUCTURE was 100% for pure *M. duodecimcostatus* ($Q = 0.904$) and 96% for pure *M. lusitanicus* ($Q = 0.928$) genotypes, and 99% and 94% for F_1 and F_2, respectively. The proportion of backcross individuals identified as parentals was 47% in the case of $F_1 \times$ *M. duodecimcostatus* and 35% in $F_1 \times$ *M. lusitanicus*. For NEWHYBRIDS, none of the simulated *M. duodecimcostatus* individuals were assigned to the *M. lusitanicus*, F_1 or F_2 classes, but 21% appeared as $F_1 \times$ *M. duodecimcostatus* backcross. Similarly, none of the *M. lusitanicus* were assigned to the *M. duodecimcostatus*, F_1 and F_2 classes, but 41% were assigned to the class $F_1 \times$ *M. lusitanicus*. Simulated F_1 and F_2 individuals were all correctly assigned except for two F_2, which appeared as a backcross $F_1 \times$ *M. duodecimcostatus*.

Individual assignments with STRUCTURE and NEWHYBRIDS provided evidence of very limited recent hybridization between *M. duodecimcostatus* and *M. lusitanicus*. STRUCTURE indicated two individuals as admixed (R087-1: $q = 0.607$; R096-1: $q = 0.585$), which were both morphologically determined as *M. duodecimcostatus* and trapped within the range of the southern cluster identified by GENELAND (Fig. 4). NEWHYBRIDS suggested the same two individuals as admixed with overall $Q = 0.75$ and $Q = 0.5$, respectively, and posterior probabilities lower than 0.5 for being purebred. R087-1 shared its *cytb* haplotype with a *M. duodecimcostatus* (ID034-1) captured within the southern part of the sympatric area. R096-1 had a unique haplotype. NEWHYBRIDS suggested a third individual from the southern part as potential hybrid (R149-4: $Q = 0.28$), which was assigned a likely *M. duodecimcostatus* by STRUCTURE ($q = 0.18$) and morphology.

Discussion

The results of this study highlight the importance and necessity of multiple sources of information for resolving evolutionary processes in systems with comparatively low evolutionary divergence and potential interbreeding at different time scales. In the Iberian pine voles analysed here, morphological, ecological, nuclear and mitochondrial data taken each separately provide only incomplete and partially misleading portraits of their geographical distribution, evolutionary divergence and the extent of gene flow among them. The combined analysis resolved evidence of deep and ancient mitochondrial introgression between these pine vole taxa that are otherwise apparently genetically mostly isolated but relatively difficult to classify based on morphology alone.

Cytonuclear discordance

The comparison of geographic patterns in mtDNA, nucDNA and morphotypes demonstrated extensive cytonuclear discordance in these ecologically divergent Iberian voles. The basal dichotomy in our mtDNA data is consistent with phylogenetic surveys of the *Microtus* genus in terms of the level of sequence divergence between these lineages and their allopatric distribution in Iberia (Jaarola *et al.* 2004; Fink *et al.* 2010). Within these basal clusters, there is relatively high haplotypic variation with strongly localized and nonoverlapping geographic distributions—a feature which is similar to phylogeographic patterns in other European *Microtus* species (e.g. Jaarola & Searle 2002; Heckel *et al.* 2005). However, the phylogeographic patterns in our mtDNA data reflect neither the well-established morphology-based distribution ranges of the two species (e.g. Madureira 1984 studied 1800 specimens across Iberia) nor the profound genetic subdivision detected in nucDNA. In contrast to mtDNA, the nuclear subdivision corresponds largely to the *M. duodecimcostatus* and *M. lusitanicus* morphotypes and PCA results for the individuals from Spain suggest that this extends over large distances. It is thus most likely that the distribution of mtDNA polymorphisms in these taxa was affected by specific evolutionary processes leading to discordance with the morphological species designation.

Morphologically established taxa may be difficult to distinguish with genetic markers, for example, owing to the incomplete sorting of ancestral polymorphisms or the effects of historical or ongoing gene flow (e.g. Petit & Excoffier 2009). mtDNA markers are less likely to be affected by incomplete lineage sorting compared with nuclear loci owing to the smaller effective size of mitochondrial loci but the stochasticity of these processes may nevertheless produce inconsistencies with the rest of the genome (Hudson & Turelli 2003). However, the geographically coherent distribution of the haplotypes belonging to the basal mtDNA clusters and the mixing of morphotypes in only one of them render pure stochasticity unlikely and rather suggest restricted gene flow between the two taxa in the form of introgressive hybridization as the likely cause of incongruence with morphology and nucDNA.

Asymmetric and geographically widespread introgression

Hybridization beyond the F1 generation may—in principle—lead to the presence of both parental types of mtDNA in hybrids but asymmetric gene flow between closely related species and parapatric evolutionary lineages is indeed relatively common at least in the immediate zone of contact (e.g. Jaarola *et al.* 1997; Buggs 2007; Plötner *et al.* 2008; Petit & Excoffier 2009; Beysard *et al.* 2012). Our data show that introgression of *M. duodecimcostatus* mtDNA into the range occupied by *M. lusitanicus* extends relatively far beyond the region

of sympatry. The direction of introgression is consistent with results from limited breeding experiments, which indicated that hybridization is possible under laboratory conditions (Wiking 1976). A detailed characterization of the depth of mtDNA introgression in this system will require a wider coverage of the northern parts of the range of *M. lusitanicus* but it is clear that it extends several dozens or hundreds of kilometres beyond the immediate sympatry or contact zone (see Fig. 1).

The large introgressed territory together with the low dispersal abilities of voles (Hamilton *et al.* 2005; Schweizer *et al.* 2007; Santos *et al.* 2010b; Le Galliard *et al.* 2012) suggests a relatively old hybridization event between the two taxa with consistent spread of *M. duodecimcostatus* mtDNA afterwards. This could be explained by either neutral processes or a selective advantage of *M. duodecimcostatus* mtDNA in a *M. lusitanicus* nuclear background. Selection has also been considered as a force for mtDNA introgression in other organisms (e.g. Melo-Ferreira *et al.* 2005; Renoult *et al.* 2009), but there is no direct support of non-neutrality of mtDNA and only very localized evidence from another *Microtus* species (Fink *et al.* 2004). In our study, the large diversity of introgressed mtDNA haplotypes in *M. lusitanicus* argues against a recent selective sweep but very old adaptive processes cannot be excluded. In general, the importance of adaptive mutations in mtDNA in natural populations is still unclear (Ballard & Melvin 2010).

A selectively neutral alternative scenario for the deep introgression detected here would consist in an invasion of one taxon into the range of the other with considerable interbreeding. Such range expansion processes with gene flow between the invading and the local species may lead to massive introgression of the invaders at neutral genes (Currat *et al.* 2008). Male-biased dispersal, which is common in *Microtus* (Hahne *et al.* 2011; Le Galliard *et al.* 2012), and limited intra-specific gene flow may cause mtDNA to introgress more than nuclear markers in such invasions (Petit & Excoffier 2009). To be consistent with our data, this would require *M. lusitanicus* to have invaded the range of *M. duodecimcostatus*. At present, there is no evidence of such a range expansion in this particular system although very limited fossil data suggest their presence in the same locality (Caldeirão Cave near Tomar) slightly north of the current sympatry zone approximately 20–30 000 years ago (Brunet-Lecomte & Povoas 1993). It is worth noting, however, that the Iberian Peninsula contains a number of independent cases with relatively deep mitochondrial introgression (e.g. Alves *et al.* 2003; Renoult *et al.* 2009). A wide range of climates and topographical diversity potentially leading to microrefugia within the Iberian refugium (Gomez & Lunt 2007) and repeated vegetation changes associated with the glacial cycles certainly provided the necessary conditions for both neutral and adaptive processes conducive to lead to introgression of mtDNA between hybridizing taxa in this region.

Past and present hybridization

The sharing of a mtDNA lineage between the two pine vole taxa demonstrates ancient hybridization but overall levels of hybridization have been apparently low. Several analytical approaches detected consistently two major genetic clusters in nucDNA in the area of contact between the two morphological taxa where hybridization might actually occur (Figs 3 and 4). The genetic distinctness at this regional scale suggests very limited historical gene flow, which is generally supported by morphological determination, even though the size overlap between the two taxa makes morphology error-prone particularly for large *M. lusitanicus*. Under relatively stringent conditions, we have evidence of two hybrids among 227 individuals; less stringent criteria might suggest a few additional backcrosses more but reliable statistical distinction from parentals is very difficult for these (Vähä & Primmer 2006). However, the regional scale of our analyses prevents us from estimating hybridization rates directly. Dedicated sampling of local populations in the direct contact zone of the two taxa will be necessary to determine if current hybridization is actually very rare or whether partial reproductive barriers only prevent gene flow beyond the F1 generation (see e.g. Schilthuizen *et al.* 2011; Smadja & Butlin 2011).

The factors maintaining the genetic integrity of the two taxa in the sympatric region are currently unknown, but attempts of experimental crossing indicate the contribution of behavioural mechanisms (i.e. differences in aggression levels; Wiking 1976). Very low sample sizes in these experiments prevent further conclusions, but limited evidence from other *Microtus* taxa at different stages of speciation suggests that premating and/or postmating mechanisms leading to partial reproductive isolation may evolve relatively quickly in this rapidly radiating rodent genus (Jaarola *et al.* 1997; Bulatova *et al.* 2010; Beysard *et al.* 2012).

It remains unclear how important the role of ecology or habitat preferences is in the maintenance of the general genetic integrity of these pine voles or whether such differences are rather a consequence of competitive exclusion at the local scale. Competitive exclusion in this system would require an advantage of the smaller species *M. lusitanicus* over *M. duodecimcostatus* to be consistent with the invasion scenario for the asymmetrical introgression of mtDNA from above. In our overlay of a soil map with GENELAND results, *M. lusitanicus* occurs mainly in the cambisol class and *M. duodecimcostatus* in leptosols and luvisols but there are several localities with

exceptions (Fig. 4). Although these mismatches might be explained by a resolution of the soil map, which is too coarse for representing local habitat conditions or the particular interpolation of likelihoods in GENELAND, it is still consistent with ecological data suggesting that *M. duodecimcostatus* is able to occupy a wider range of soil types, while *M. lusitanicus* may not (Santos *et al.* 2010a,b). However, habitat features like soil thickness and the percentage of plant cover differ also in other regions of the sympatry zone between these pine voles (Borghi *et al.* 1994), which indicates the importance of ecological factors for the separation of the two taxa.

Conclusion

The completion of the speciation process is notoriously difficult to determine in nascent species, and the Iberian pine voles studied here exemplify this. Our results show that the current classification as two species is appropriate or not—depending on the species concept favoured (see Mitchell-Jones *et al*. 1999; Baker & Bradley 2006). More importantly, however, these two taxa are apparently at a point in the speciation process where genetic and phenotypic differences have established but interbreeding may still occur. We are thus in a position to combine in the future the comparative characterization of their evolutionary history at a large geographical scale with population-based analyses of genome-wide gene flow patterns and experimental approaches to phenotypic divergence and reproductive isolation. Such comprehensive analyses hold very much potential for a deeper understanding of this particular system and the mechanisms behind the ongoing explosive radiation of Arvicoline rodents in general.

Acknowledgements

We are grateful to Mathias Beysard for helpful suggestions on the manuscript and Isabel Rey from the Museo Nacional de Ciencias Naturales (MNCN), Spain, for providing biological samples. CBS is supported by postdoctoral grant FCT-SFRH/BPD/26612/2005, Synthesis Grant ES-TAF-5106 and a Research Project FCT-PTDC/BIA-BEC/103729/2008. SMS is supported by PhD grant PRAXIS/SFRH/BD/21403/2005. GH acknowledges support from the Swiss National Science Foundation grant 31003A_127377/1.

References

Alves PC, Ferrand N, Suchentrunk F, Harris DJ (2003) Ancient introgression of *Lepus timidus* mtDNA into *L. granatensis* and *L. europaeus* in the Iberian Peninsula. *Molecular Phylogenetics and Evolution*, **27**, 70–80.

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

Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, **87**, 643–662.

Ballard JWO, Melvin RG (2010) Linking the mitochondrial genotype to the organismal phenotype. *Molecular Ecology*, **19**, 1523–1539.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 3101–3109.

Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P (2012) Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). *Journal of Evolutionary Biology*, **25**, 400–408.

Borghi CE, Giannoni SM, Martinez-Rica JP (1994) Habitat segregation of three sympatric fossorial rodents in the Spanish Pyrenees. *Zeitschrift für Säugetierkunde*, **59**, 52–57.

Braaker S, Heckel G (2009) Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole *Microtus arvalis*. *Molecular Ecology*, **18**, 2518–2531.

Brunet-Lecomte P, Povoas L (1993) Voles (Arvicolidae, Rodentia) from Caldeirão Cave (Tomar, Portugal). *Arquivos do Museu Bocage*, **24**, 409–414.

Brunhoff C, Galbreath KE, Fedorov VB, Cook JA, Jaarola M (2003) Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology*, **12**, 957–968.

Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity*, **99**, 301–312.

Bulatova NS, Potapov SG, Lavrenchenko LA (2010) Genomic versus chromosomal polytypy in studies of mitochondrial and nuclear DNA markers in the *Microtus arvalis* group. *Russian Journal of Genetics*, **46**, 586–594.

Cabrera A (1914) *Fauna Ibérica*. Mamíferos. Imprenta de Fontanet. Madrid.

Cotilla I, Palomo LJ (2002) *Microtus duodecimcostatus* (de Selys-Longchamps, 1839). In: Atlas de los mamíferos terrestres de España (eds Palomo LJ, Gisbert J), pp. 378–381. Dirección General de Conservación de la Naturaleza, SECEM-SECEMU, Madrid.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–1920.

De Selys-Longchamps M (1839) Campagnols inédits. Revue Zoologique par La Société Cuvierienne, 8–9.

Ellerman JR, Morrison-Scott TS (1951) Checklist of Palearctic and Indian Mammals. British Museum (Natural History), London.

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

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

Fink S, Excoffier L, Heckel G (2004) Mitochondrial gene diversity in the common vole *Microtus arvalis* shaped by historical divergence and local adaptations. *Molecular Ecology*, **13**, 3501–3514.

Fink S, Excoffier L, Heckel G (2006) Mammalian monogamy is not controlled by a single gene. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 10956–10960.

Fink S, Excoffier L, Heckel G (2007) High variability and non-neutral evolution in the mammalian avpr1a gene. *BMC Evolutionary Biology*, **7**, e176.

Fink S, Fischer MC, Excoffier L, Heckel G (2010) Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. *Systematic Biology*, **59**, 548–572.

Gerbe MZ (1879) Etude comparative de quelques caractères du campagnol incertain. *Revue et Magasin de Zoologie 3º Série*, **7**, 42–47.

Glaubitz J (2004) CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, **4**, 309–310.

Gomez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Phylogeography of Southern European Refugia. Evolutionary Perspectives on the Origins and Conservation of European Biodiversity (eds Weiss S, Ferrand N), pp. 155–188. Springer, Netherlands.

Good JM, Hird S, Reid N *et al.* (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular Ecology*, **17**, 1313–1327.

Goudet J (1999) PCAGEN, a computer package which performs Principal Component Analysis (PCA) on gene frequency data. Available from: http://www2.unil.ch/popgen/softwares/pcagen.htm.

Guillot G, Mortier F, Estoup A (2005) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Hahne J, Jenkins T, Halle S, Heckel G (2011) Establishment success and resulting fitness consequences for vole dispersers. *Oikos*, **120**, 95–105.

Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L. (2005) Bayesian estimation of recent migration rates after a spatial expansion. *Genetics*, **170**, 409–417.

Heckel G, Burri R, Fink S, Desmet J-F, Excoffier L (2005) Genetic structure and colonization processes in European populations of the common vole *Microtus arvalis*. *Evolution*, **59**, 2231–2242.

Hellborg L, Gunduz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, **14**, 2025–2031.

Hudson RR, Turelli M (2003) Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182–190.

Ishibashi Y, Saitoh T, Abe S, Yoshida MC (1995) Polymorphic microsatellite DNA markers in the grey red-backed vole *Clethrionomys rufocanus bedfordiae*. *Molecular Ecology*, **4**, 127–128.

Ishibashi Y, Yoshinaga Y, Saitoh T *et al.* (1999) Polymorphic microsatellite DNA markers in the field vole *Microtus montebelli*. *Molecular Ecology*, **8**, 163–164.

Jaarola M, Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, **11**, 2613–2621.

Jaarola M, Tegelstrom H, Fredga K (1997) A contact zone with noncoincident clines for sex-specific markers in the fied vole (*Microtus agrestis*). *Evolution*, **51**, 241–249.

Jaarola M, Martínková N, Gunduz I *et al.* (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**, 647–663.

Jaarola M, Ashford RT, Ratkiewicz M, Brunhoff C, Borkowska A (2007) Isolation and characterization of polymorphic microsatellite loci in the field vole, *Microtus agrestis*, and their cross-utility in the common vole, *M. arvalis*. *Molecular Ecology Notes*, **7**, 1029–1031.

Le Galliard J-F, Remy A, Ims RA, Lambin X (2012) Patterns and processes of dispersal behaviour in arvicoline rodents. *Molecular Ecology*, **21**, 505–523.

Maan ME, Seehausen O (2011) Ecology, sexual selection and speciation. *Ecology Letters*, **14**, 591–602.

Madureira ML (1981) Discriminant analysis in Portuguese pine voles: *Pitymys lusitanicus* Gerbe and *Pitymys duodecimcostatus* de Sélys-Longchamps (Mammalia: Rodentia). *Arquivos do Museu Bocage Série A*, **I**, 111–122.

Madureira ML (1984) *A biologia de Microtus (Pitymys) duodecimcostatus de Sélys-Longchamps, 1839 e M. (P.) lusitanicus Gerbe, 1879 em Portugal (Arvicolidae, Rodentia): taxonomia, osteologia, ecologia e adaptações*. Unpublished PhD Thesis, Department of Animal Biology, Faculty of Sciences, University of Lisbon, Lisbon.

Mallet J, Meyer A, Nosil P, Feder J (2009) Space, sympatry and speciation. *Journal of Evolutionary Biology*, **22**, 2332–2341.

Manoukis NC (2007) FORMATOMATIC: a program for converting diploid allelic data between common formats for population genetic analysis. *Molecular Ecology Notes*, **7**, 592–593.

Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, **14**, 2459–2464.

Mira A, Mathias ML (2002) *Microtus lusitanicus* (Gerbe, 1879). In: Atlas de los mamíferos terrestres de España (eds Palomo LJ, Gisbert J), pp. 378–381. Dirección General de Conservación de la Naturaleza, SECEM-SECEMU, Madrid.

Mitchell-Jones AJ, Amori G, Bogdanowicz W, Krystufek B, Reijnders PJH *et al.* (1999) The Atlas of European Mammals. T & AD Poyser, London.

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 5269–5273.

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

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution*, **24**, 145–156.

Paradis E, Guédon G (1993) Demography of a mediterranean microtine: the Mediterranean pine vole, *Microtus duodecimcostatus*. *Oecologia*, **95**, 47–53.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Plötner J, Uzzell T, Beerli P, Spolsky C, Ohst T *et al.* (2008) Widespread unidirectional transfer of mitochondrial DNA: a case in western Palearctic water frogs. *Journal of Evolutionary Biology*, **21**, 668–681.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

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

Renoult JP, Geniez P, Bacquet P, Benoit L, Crochet PA (2009) Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. *Molecular Ecology*, **18**, 4298–4315.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harber Laboratory Press, New York City, New York.

Santos S, Mira A, Mathias ML (2009) Factors influencing large-scale distribution of two sister species of pine voles (*Microtus lusitanicus* and *M. duodecimcostatus*): the importance of spatial autocorrelation. *Canadian Journal of Zoology*, **87**, 1227–1240.

Santos S, Mathias ML, Mira A (2010a) Local coexistence and niche differences between the Lusitanian and Mediterranean pine voles (*Microtus lusitanicus* and *M. duodecimcostatus*). *Ecological Research*, **25**, 1019–1031.

Santos SM, Lourenço RF, Mathias ML, Mira AP (2010b) Spatial and temporal ecology of the Lusitanian pine vole (*Microtus lusitanicus*) in a Mediterranean polyculture. *Animal Biology*, **60**, 209–227.

Santos SM, Mathias ML, Mira AP (2011) The influence of local, landscape and spatial factors on the distribution of the Lusitanian and the Mediterranean pine voles in a Mediterranean landscape. *Mammalian Biology*, **76**, 133–142.

Schilthuizen M, Giesbers MCWG, Beukeboom LW (2011) Haldane's rule in the 21st century. *Heredity*, **107**, 95–102.

Schweizer M, Excoffier L, Heckel G (2007) Fine-scale genetic structure and dispersal patterns in the common vole *Microtus arvalis*. *Molecular Ecology*, **16**, 2463–2473.

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

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505–518.

Vähä JPK, Primmer CR (2006) Detecting hybridization between individuals of closely related populations—a simulation study to assess the efficiency of model-based Bayesian methods to detect hybrid individuals. *Molecular Ecology*, **15**, 63–72.

Walser B, Heckel G (2008) Microsatellite markers for the common vole (*Microtus arvalis*) and their cross-species utility. *Conservation Genetics*, **9**, 479–481.

Wiking VH (1976) Karyologie und Biologie der beiden iberischen Wühlmausarten *Pitymys mariae* und *Pitymys duodecimcostatus*. *Zoologische Systematik und Evolutionsforschung*, **14**, 104–129.

C.B.S. research focuses on the use of natural history collections and molecular data to understand the role of past and present environmental changes in affecting evolution of small mammals in the Iberian Peninsula and Mozambique. S.M.S. is specialised in microtine ecology and currently focuses on road barrier effects and functional landscape connectivity for different mammal and bird species. R.M. is interested in the energy demands and contrains of small mammals and the ecological factors regulating their physiological processes. M.M.L. is currently interested in short and long-term adaptation strategies of small mammal species to cope with ever-changing environments. G.H. has broad interests ranging from speciation and colonisation processes over host-pathogen co-evolution, the molecular ecology of dispersal and mating systems to the evolution of behaviorally-active genes.

Data accessibility

DNA sequences: GenBank accessions JX424119–JX424211.

Microsatellite data: DRYAD entry doi:10.5061/dryad.q3ng0.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Fig. S1 Admixture analyses of observed individuals from the sympatric area of Alto Alentejo, computed by STRUCTURE (A) and NEWHYBRIDS (B).

Fig. S2 The efficiency of detecting parental and F1 individuals simulated by Hybridlab for different STRUCTURE thresholds *q*-values.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.