



HAL
open science

Out of Africa: demographic and colonization history of the Algerian mouse (*Mus spretus* Lataste)

Aude Lalis, Stefano Mona, Emmanuelle Stoetzel, François Bonhomme, Karim Souttou, Ali Ouarour, Stéphane Aulagnier, Christiane Denys, Violaine Nicolas

► To cite this version:

Aude Lalis, Stefano Mona, Emmanuelle Stoetzel, François Bonhomme, Karim Souttou, et al.. Out of Africa: demographic and colonization history of the Algerian mouse (*Mus spretus* Lataste). *Heredity*, 2019, 122 (2), pp.150-171. 10.1038/s41437-018-0089-7 . hal-02173272

HAL Id: hal-02173272

<https://hal.science/hal-02173272v1>

Submitted on 19 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Out of Africa: demographic and colonization history of the Algerian mouse (*Mus spretus* Lataste)

Aude Lalis¹, Stefano Mona^{1,2}, Emmanuelle Stoetzel³, François Bonhomme⁴, Karim Souttou⁵, Ali Ouarour⁶, Stéphane Aulagnier⁷, Christiane Denys¹ & Violaine Nicolas¹

¹ *Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 - CNRS, MNHN, UPMC, EPHE, Sorbonne Universités,*

Paris, France

² *EPHE, PSL Research University, Paris, France*

³ *Histoire Naturelle de l'Homme Préhistorique, HNHP—UMR 7194 —CNRS, MNHN, UPVD, Sorbonne Universités, Paris, France*

⁴ *Institut des Sciences de l'Evolution, ISEM—UMR 4554, CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France*

⁵ *Department of Agronomy, Faculty of Natural Science and Life, University Ziane Achour, Djelfa, Algeria*

⁶ *Laboratoire de Biologie et Santé, Faculté des Sciences, Université Abdelmalek Essâadi, Tétouan, Morocco*

⁷ *Institut National de la Recherche Agronomique, UR35 Comportement et Ecologie de la Faune Sauvage, Caytanet-Tolosan, France*

* Corresponding author : vnicolas@mnhn.fr

Abstract

North Africa is now recognized as a major area for the emergence and dispersal of anatomically modern humans from at least 315 kya. The Mediterranean Basin is thus particularly suited to study the role of climate versus human-mediated changes on the evolutionary history of species. The Algerian mouse (*Mus spretus* Lataste) is an endemic species from this basin, with its distribution restricted to North Africa (from Libya to Morocco), Iberian Peninsula and South of France. A rich paleontological record of *M. spretus* exists in North Africa, suggesting hypotheses concerning colonization pathways, and the demographic and morphologic history of this species. Here we combined genetic (3 mitochondrial DNA loci and 18 microsatellites) and climatic niche modeling data to infer the evolutionary history of the Algerian mouse. We collected 646 new individuals in 51 localities. Our results are consistent with an anthropogenic translocation of the Algerian mouse from North Africa to the Iberian Peninsula via Neolithic navigators, probably from the Tingitane Peninsula. Once arrived in Spain, suitable climatic conditions would then have favored the dispersion of the Algerian mice to France. The morphological differentiation observed between Spanish, French and North African populations could be explained by a founder effect and possibly local adaptation. This article helps to better understand the role of climate versus human-mediated changes on the evolutionary history of mammal species in the Mediterranean Basin.

Introduction

The Mediterranean Basin, between southern Europe and North Africa, is particularly suited to study the role of climate versus human-mediated changes on the evolutionary history of species for several

reasons. First, this area has yielded a rich paleontological record allowing a better understanding of the morphological evolution of small mammal species and of community assemblages over time in relation to Quaternary climatic changes (Stoetzel 2013). Second, North Africa is now recognized as a major area for the emergence and dispersal of anatomically modern humans from at least 315 kya (Garcea 2012; Hublin et al. 2017; Richter et al. 2017). It thus constitutes a region of interest for investigating how the arrival of early representatives of *Homo sapiens* and their recent impact on the environment influenced faunal genetic diversity and distribution. During the Neolithic and after, numerous contacts between the two shores of the Mediterranean Sea occurred, due to the development of navigation (Mulazzani et al. 2010; Souville 1998; Zilhão 2014). This led to the translocation (Dobson 1998) and an almost complete turnover of mammal species in the Mediterranean islands (Vigne 1992).

The Algerian mouse, *Mus spretus* Lataste, is an emblematic species endemic to the Mediterranean Basin. It occurs in North Africa, from Morocco through northern Algeria and Tunisia to northern Libya, and in south-west Europe (southern France, most of the Iberian Peninsula and the Balearic Islands; Palomo et al. 2009). The oldest fossil of *M. spretus* is dated from the Middle Pleistocene, about 600 kyr BP, in Tunisia (Jebel Ressay 2, 3; Stoetzel 2013). This species is also known from several sites in Morocco in the Middle-Late Pleistocene, early Holocene and middlelate Holocene (Stoetzel 2013), while it is reported from Europe only in the Neolithic (Palomo et al. 2009). Fossil evidence thus suggests a North African origin of the species and recent spreading to southwestern Europe via Neolithic navigators (Dobson 1998). Based on restriction fragment length polymorphism (RFLP) data, Boursot et al. (1985) showed a strong genetic isolation between European and North African populations, but they could not characterize the tempo and mode of such divergence. Based on few Dloop mitochondrial DNA sequences (mtDNA), Desmarais (1989) showed the absence of actual gene flow between European and African populations, and a probable recent colonization event of Europe from North Africa with a strong founder effect. He also proposed that during an unfavorable climatic period, allopatric diversification occurred in two refugia in North Africa (Tunisia and Morocco).

The Late Pleistocene to middle Holocene succession of El Harhoura 2 cave (Morocco, region of Témara), yields an exceptionally rich small vertebrate assemblage and showed a continuous fossil record for *M. spretus* from 107 ± 7 kyr BP to 5800 years BP, with nonsignificant change in abundance through time, despite pronounced environmental changes (glacial–interglacial cycles, Stoetzel et al. 2011b). The morphology of the molars of Algerian mouse through this fossil record was also surprisingly stable (Stoetzel et al. 2013). The authors proposed that such evolutionary stability despite environmental changes could be explained by the fact that preferred habitats (grasslands, dry shrubland and open woodland) persisted in the surroundings of El Harhoura 2 throughout the record as a “mosaic” of habitats despite variations in their extension. In addition, the coastal location of this cave probably reduced the impact of global climatic variation compared with further inland. This longterm morphological stability contrasts with the molar size decrease observed between the last El Harhoura 2 Holocene level (5800 years BP) and modern Moroccan mouse populations (Stoetzel et al. 2013). Geographical morphological variation have been documented across the modern distribution range of this mouse (Darviche and Orsini 1982; Darviche et al. 2006; Khammes-El Homsni and Aulagnier 2010; Palomo et al. 1983; Palomo et al. 1985; Stoetzel et al. 2013; Valenzuela-Lamas et al. 2011). The modern geographical variation was hypothesized to be due to relative isolation between populations leading to divergence and possibly local adaptation. The differentiation of Algerian mice from a pool corresponding to Morocco and Southern Europe (Boursot et al. 1985; Khammes-El Homsni and Aulagnier 2010) may have been triggered by North African mice populations being fragmented by

places where the desert reaches the sea or in altitude, isolating favorable places (“continental islands”) for the mice (Stoetzel et al. 2013). However, even at a local scale, great morphological variation is observed between present day Moroccan populations that cannot be attributed to such large-scale biogeographical structure (Stoetzel et al. 2013). The three last millennia are known as a period of important environmental perturbations, with an increase of aridity (Hassan 1997; Le Hou  rou 1997) enhanced by the progressive increase in anthropogenic activities (Brun 1989; Brun 1991; Stambouli-Essassi et al. 2007) and concomitant introduction of commensal rodents such as the house mouse, and the black and brown rats (Audoin-Rouzeau and Vigne 1994; Bonhomme et al. 2011; Cucchi and Vigne 2006). The competition and possible hybridization with the house mouse *Mus musculus domesticus*, together with anthropogenic changes in landscape, may have particularly impacted the local evolution of *M. spretus* by restricting its habitat, triggering divergence of isolated populations, which were previously interconnected. The anthropogenic action may also have modified the environment in a more global way. By the fact, among other effects, it was observed a significant size decrease of several wild animals during historical time (Sheridan and Bickford 2011). The size decrease of the Western Mediterranean mouse during historical times may be the discrete signature of a perturbed ecosystem, pushing the species away from its previous stasis as it was evidenced in islands for other species (Sondaar 2000).

In the present study, we combined mitochondrial and nuclear data to infer the evolutionary history of the Algerian mouse. Specifically, we wanted to answer the following questions: (1) When and how did the Algerian mouse arrive in Europe? (2) Can we detect a recent population bottleneck in North African populations linked to the development of anthropogenic activities and/or the arrival of the house mouse? (3) Is the high morphological variability observed in present day populations due to isolation? Moreover, we evaluated these genetic data with climatic niche modeling to reveal the evolutionary history of *M. spretus* in response to climate change during the Quaternary.

The combination of molecular, fossil, palaeoenvironmental and human context data should allow us to reconstruct the historic fluctuations of populations within this taxon in North Africa and Europe.

Materials and methods

Ethics statement

Animals were live-trapped and handled under the guidelines of the American Society of Mammalogists (Sikes and Mammalogists atACaUCotASo 2016). The protocol was approved by “Comit   Cuvier” (permission no. 68.009). All manipulations of animals were made in Morocco in agreement with the global law 11-03 relative to the protection and the development of the environment. Alive animals were euthanized by the injection of a lethal dose of isoflurane, followed by cervical dislocation. Capture permits were obtained through the “Haut Commissariat aux Eaux et for  ts et la Lutte contre la desertification” (Authorisation no. 15 HCEFLCD/DLCDPN/DPRN/CFF) in Morocco, and through the Ministry of Forestry in Algeria. No capture permit is necessary in Europe, because this species is under no specific status.

Sampling and DNA extraction

A total of 635 newly collected individuals were sampled in 51 localities (81 specimens from 4 Algerian localities, 448 specimens from 28 Moroccan localities, 7 specimens from 4 Tunisian localities,

61 specimens from 7 French localities, 1 specimen from one locality in Portugal and 37 specimens from 7 Spanish localities; Table 1, Table S1 and Fig. 1). Genomic DNA was extracted from tissues preserved in 95% ethanol using the NucleoSpin Tissue Core kit (MACHEREY-NAGEL, Hoerd, France).

DNA amplification and sequencing

Three mitochondrial genes (Cytochrome b—Cytb-, Cytochrome oxidase I—CO1- and control region—CR-) were amplified following (Ducroz et al. 2001; Rajabi-Maham et al. 2008; Robins et al. 2007). All newly determined sequences were submitted to GenBank (accession numbers MH241060 to MH242499 and MH253776 to MH253817).

We genotyped 536 individuals from 19 localities (Table 1) for 18 microsatellite loci using primers and genotyping protocols described in Lalis and Orth (2014).

Sequence analyses

Median-joining networks

In our phylogenetic analyses, we included 471 specimens for Cytb (coming from Algeria, Morocco, Spain, France and Portugal; Table S1), 499 specimens for CO1 (Algeria, Morocco, Tunisia, Spain, France and Portugal) and 518 specimens for CR (Algeria, Morocco, Tunisia, Spain, France and Portugal). Phylogenetic relationships between haplotypes were inferred by constructing a network using the median-joining (MJ) method available in NETWORK v4.500 (Bandelt et al. 1999).

Sequences of 1085 bp for Cytb, 655 bp for CO1 and 840 bp for CR were retained for the network analysis to minimize the number of incomplete sequence as adding ambiguous data in MJ trees is problematic. Analyses were performed for each gene separately and for the three mtDNA genes combined. For the concatenate analysis, only the 425 specimens for which the three mtDNA regions were sequenced were included in the analysis.

Genetic diversity and population differentiation

Only populations with at least 12 individuals were retained in our analyses. Nucleotide diversity, haplotype diversity (Nei 1987) and the mean number of pairwise differences were calculated using DnaSP 5.10 (Librado and Rozas 2009). Population differentiation was inferred from the mtDNA dataset using F_{st} estimates, computed with ARLEQUIN version 3.5.1 (Excoffier and Lischer 2010).

Demographic history and divergence time estimates

We tested for signatures of population expansion through the F_u 's F_s (Fu 1997), Tajima's D (Tajima 1989) and Ramos-Onsins and Rozas R2 (Ramos-Onsins and Rozas 2002) tests of neutrality. These analyses were performed with Arlequin 3.5 (Excoffier and Lischer 2010) and DnaSP 5.10 (Librado and Rozas 2009), and the significance of the neutrality statistics was assessed using 1000 coalescent simulations under a model of constant population size using empirical estimates of theta.

We further analyzed the demography of the main populations using the Extended Bayesian Skyline Plot (EBSP) implemented in BEAST v.1.8 (Drummond and Rambaut 2007). EBSP is a non-parametric model that does not specify any prior hypothesis on the tempo and mode of changes of the effective

population size through time (Heled and Drummond 2008). EBSM considers populations as fully isolated, meaning that they have evolved separately from any others up to the time to the most recent common ancestor of the locus(i) under investigation. This hypothesis is clearly violated in our case. However, EBSM can provide important insights if correctly interpreted in the light of the presence of population structure (Maisano Delser et al. 2016; Mazet et al. 2016). We ran 100 million Markov chain Monte Carlo (MCMC) iterations with a thinning interval of 10,000. We checked convergence by visually inspecting the trace and computing the effective sample size (which was always higher than 100) for each parameter in two independent runs using the program TRACER 1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>). EBSM were plotted using R version 3.3.2 (R Development Core Team 2016) setting, the burn-in to 10%. We set a strict clock with rate of 3×10^{-7} substitutions per site per year and the Hasegawa–Kishino–Yano plus Gamma model of molecular evolution. It is now widely recognized that intraspecific and pedigree-based estimates of substitution rates are generally higher than interspecific phylogenetically calibrated rates (Henn et al. 2009; Ho et al. 2007b; Loogväli et al. 2009). Estimating intraspecific substitution rate is not trivial (Bonhomme et al. 2011). A rate of 4×10^{-7} substitutions per site per year was proposed for the CR of the congeneric rodent species *Mus musculus domesticus* (Rajabi-Maham et al. 2008). A Cytb clock rate of 3×10^{-7} substitutions per site per year was estimated using radiocarbon dated postglacial remains in another rodent, *Microtus arvalis* (Martinkova et al. 2013). Similar clock rates have been obtained with dated ancient DNA from a wide range of mammals and birds (Ho et al. 2007a).

Microsatellite analyses

Genetic diversity and differentiation

Genetic variability of the microsatellite markers was measured for each population by the number of alleles (N_a), gene diversity (H_e , expected heterozygosity) and observed heterozygosity (H_o) using R package ADEGENET V3.3.1 (Jombart 2008). The departures from Hardy–Weinberg equilibrium (HWE) and the linkage disequilibrium between pairs of loci, with significance set at $P < 0.05$, were assessed with GENEPOP v. 4.1 (Raymond and Rousset 1995) and were later adjusted using a sequential Bonferroni correction (Rice 1989).

The program BOTTLENECK was used to test for past decrease in effective population size (Piry et al. 1999). Most microsatellites fit a two-phase mutation model (TPM) better than a strict stepwise mutation model (SMM) or an infinite alleles model (IAM). However, microsatellites with imperfect repeat patterns are more likely to evolve under the IAM model (Cornuet and Luikart 1996). As we found no strong reason to select or reject one model over another, we ran Bottleneck under all three mutational models. For the TPM model, the proportion of SMM was set to 70% and the variance to 30 (default values). The significance was assessed by performing 10,000 replicates.

Population structure, assignment and admixture analyses

STRUCTURE v.2.3.353 (Pritchard et al. 2000) was used to infer geographical genetic structure. We assumed an admixture model with correlated allele frequencies (Falush et al. 2003). For each value of the number (K) of clusters set between 1 and 19 (number of sampled sites), we performed 10 independent runs of 106 MCMC iterations with a burn-in of 105. The optimal K value was inferred using the ad-hoc

Evanno statistic Delta K (Evanno et al. 2005). For the selected values of K, the software CLUMPP v1.2 (Jakobsson and Rosenberg 2007) was used to align cluster membership coefficients from the 10 replicate cluster analyses using the Greedy algorithm with 10,000 random input orders. The results were graphically displayed with DISTRUCT v 1.1 (Rosenberg 2004).

We also performed a discriminant analysis of principal components (DAPC), using R package ADEGENET version 1.2-7 (Jombart 2008) for the analysis of genetically structured populations. Dataset was also depicted in an unrooted neighbor-joining (NJ) tree (Takezaki and Nei 1996) based on the pairwise Reynold's distances computed with POPULATION 1.2.32 (Langella 2010). The robustness of nodes was evaluated by carrying out 1000 bootstrap replicates. Population differentiation was further analyzed by computing estimates of FST (Weir and Cockerham 1984) between all population pairs using GENEPOP V4.1.3, whose significance was assessed through 10,000 permutations. P-values were adjusted using sequential Bonferroni correction.

Genetic inference of colonization pathways and demographic history of the three main populations (Africa, Spain, France)

We devised four models to decipher the colonization pathways and the demographic history of *M. spretus* by means of approximate Bayesian computation (ABC) (Bertorelle et al. 2010). These models were designed to address two main issues: (i) the timing and colonization dynamics of Europe; (ii) the response of *M. spretus* to recent environmental changes in North Africa. The four models and the associated parameters are shown in Figure S1 with the prior distributions presented in Table S2. Briefly, models 1 and 2 represent an independent colonization of France and Spain from North Africa, whereas models 3 and 4 represent a sequential colonization of Europe, with *M. spretus* invading first Spain and then France. Models 1 and 3 assumed an instantaneous change in effective population size in North Africa followed by an exponential change in effective population size until present, whereas models 2 and 4 assumed a constant effective population size through time. In all the four models, we assumed an instantaneous bottleneck during the colonization process of both France and Spain followed by an exponential change in effective population size until present. The coefficient of the exponential change was computed as follows: first, we extract a time in generations when the change occurred, then a new effective population size (i.e., an instantaneous change), and last, a new modern effective population size. Taking as example MOD1 (or MOD3), this would correspond to extract, respectively, "T change Africa", "N change Africa" and "N Africa". The variation of the Ne from "N change Africa" to "N Africa" follows an exponential distribution whose coefficient is determined by these three parameters. Depending on the prior extraction, the modern population would be constant (if "N change Africa" = "N Africa"), expanding (if "N change Africa" < "N Africa") or contracting (if "N change Africa" > "N Africa"). The model is therefore free to vary, in order to detect if modern population is stable or is it changing through time. We note that MOD2 and MOD4 are simply nested in MOD1 and MOD3 (i.e., "N change Africa" = "N Africa"). Exponential size change in France and Spain are modeled following the same scheme. Each of the three populations is a pool of the sampled populations coming from that geographic area. This was done to lower the number of parameters. We generated 500,000 simulations for each demographic model using fastsimcoal2 version 2.5.1 (Excoffier et al. 2013). Random values from the prior distribution of each parameter were extracted using an R script. Model posterior probabilities were calculated by a weighted multinomial logistic regression (Beaumont 2008) for which we retained the best 60,000 simulations. The demographic parameters within each model were estimated from the 5000 simulations closest to the observed dataset using a local linear

regression (Beaumont et al. 2002). Analyses were performed in the R environment with the library abc (Csilléry et al. 2012).

We run the four models with two independent set of markers: (i) 18 microsatellites ("MSAT"); (ii) mtDNA, combining d-loop, Cytb and CO1 ("MTDNA"). For each set, we considered a generation time of 0.5 year (Palomo et al. 2009; Vargas et al. 1991). In MSAT, we used per generation per locus mutation rate of 10^{-4} (Dib et al. 1996) and we computed the following summary statistics: mean and variance over loci for each population of the number of alleles, Garza–Williamson statistic, allelic range and the pairwise R_{st} . In MTDNA, we used per generation per site mutation rate of 3×10^{-7} (Martinkova et al. 2013) and we compute the following summary statistics: number of haplotypes, number of segregating sites, Tajima's D , mean pairwise difference in each population and the pairwise molecular F_{st} . All summary statistics were computed with arlsumstat (Excoffier and Lischer 2010).

To test the validity of model assumptions we performed a principal component analysis (PCA) on the reference table of summary statistics simulated from prior distributions adding the vector computed from real data. We visually inspect the resulting two first dimensions of the PCA plot to check if the observed vector fell within the clouds of simulated points. The four models passed this test for both datasets (Figure S2).

A cross-validation procedure was implemented to evaluate the model selection performance using the same threshold as in the real data. A similar cross-validation procedure was implemented to evaluate the precision and accuracy of ABC parameter estimates: pseudo-observed datasets (pods) were simulated from prior distributions and re-analyzed using the reference table and the same parameters of the ABC algorithm. We computed the 95% coverage, the scaled mean error (SME) and the scaled root mean square error (SRMSE) calculated as in Walther and Moore (2005) following Maisano Delser et al. (2016). Finally, we performed a Bayesian posterior predictive test (Gelman et al. 2004) using the global R_{st} in MSAT or F_{st} in MTDNA as predictive statistics (Figure S3). We chose these two statistics as they were not used in the estimation process as suggested by Bertorelle et al. (2010).

Finally, preliminary results (see below) suggested the presence of significant population structure in the metapopulation from North Africa. Pooling the populations without taking structure into account may bias the demographic estimates, but the extent of this bias is unknown. To remove the influence of population structure, we followed the sampling theory proposed by previous works (Wakeley 1999; Stadler et al. 2009; Mona et al. 2014) by sampling few lineages per deme and pooling them. This scattered sample (sensu Wakeley 1999) eliminates the influence of local coalescence and describes the behavior of the whole metapopulation. To this end, we created 1000 resampled dataset by randomly sampling 25 and 20 African individuals from the MSAT (RES-MSAT) and MTDNA (RESMTDNA) dataset, respectively. We performed 500,000 simulations for each of the four models as before and applied the same ABC framework for each resampled dataset (both model selection and parameter estimation). Parameters estimations are based on the average of the 1000 posterior distributions.

Species distribution modeling

We modeled the climatic niche of *M. spretus* to approximate its current, last glacial maximum (LGM) and mid-Holocene distributions. We applied ecological niche modeling methods, where environmental data are extracted from current occurrence records, and habitat suitability is evaluated across the landscape using program-specific algorithms (Elith et al. 2006). The present day models were then projected on the climatic reconstructions of the LGM (a cold and dry period, about

22,000 years ago) and mid-Holocene (a wetter period, about 6000 years ago) under the assumption that the climatic niche of the species remained conserved through time (Elith et al. 2010).

Distribution records of Algerian mouse were assembled from literature (Aulagnier et al. 2017; Brahmi et al. 2010; Brahmi et al. 2012; Khammes-El Homsy and Aulagnier 2010), museum collections (Global Biodiversity Information Facility, GBIF.org (7th September 2016) GBIF Occurrence Download <http://doi.org/10.15468/dl.pcgarno>) and our own unpublished databases. Datasets of this type often exhibit sampling bias, which can lead to unreliable ecological niche modeling results (Fourcade et al. 2014). To correct sampling bias, a subsample of records regularly distributed in the geographical space was selected using ENMtools ver. 1.3 (Warren et al. 2010). The resolution of the reference grid was set to 2.5 arc min (nearly 5 × 5 km), which is the same as that used for the environmental layers. This coarse resolution was chosen to better match the coordinate uncertainties associated with georeferenced, textual localities of museum specimens. After filtering, the final dataset used for modeling consisted of 702 distribution records. However, this final dataset still exhibited sampling bias: most of these records came from Europe and few from Africa (Figure S4). To overcome this bias it is possible to use a bias file: MAXENT can be fed with a bias grid that is a sampling probability surface. Creating this bias file is difficult because we need to know the actual sampling intensity across the study area. We used a different approach: we considered that sampling intensity was different in Europe and North Africa, and performed two distinct analyses. One analysis with only the distribution records from Europe, which was used to estimate present and past habitat suitability for *M. spretus* in Europe, and one analysis with only the distribution records from Africa, which was used to estimate habitat suitability in Africa.

As environmental layers, we used available climatic data from the WorldClim database (Hijmans et al. 2005). We removed from our analyses variables BIO2, BIO3, BIO5, BIO7 and BIO15 because those variables show a high level of discrepancy between general circulation models (GCMs) for North Africa (Varela et al. 2015). Colinearity of the initial variables was measured by Pearson's correlation coefficient in ENMtools v1.3 (Warren et al. 2010). A total of five variables, all of which had a Pearson correlation coefficient lower than 0.75, were retained. The final set of environmental predictor variables used for the species distribution models (SDM) consisted of: annual temperature (BIO1), precipitation of wettest quarter (BIO16), precipitation of driest quarter (BIO17), precipitation of warmest quarter (BIO18) and precipitation of coldest quarter (BIO19).

Climatic variables were used for present conditions and for the LGM and mid-Holocene. Paleoclimate data were drawn from the GCM simulations based on two climate models: the Community Climate System Model (CCSM, version 3) (Collins et al. 2006) and the Model for Interdisciplinary Research on Climate (MIROC, version 3.2) (Hasumi and Emori 2004). For a discussion of the uncertainties associated with the climatic data, see Schorr et al. (2012) and Varela et al. (2015). The use of these two different climate models enabled us to assess and account for modeling uncertainty due to LGM climate data. To predict the potential distribution of the species in current conditions and in the LGM, we used MAXENT ver. 3.3.3 (Phillips et al. 2006), which returns a model with relative occurrence probability of a species within the grid cells of the study area. To ensure consistency of model predictions among repeated runs, we performed a 10-fold cross-validation with random seed. We set the regularization multiplier to 2 (Radosavljevic and Anderson 2014). To determine whether the predictions for current conditions generated by Maxent were better than random predictions, we used the area under the receiver operating characteristic curve (AUC), a commonly used measurement for comparison of model performance (Elith et al. 2006). The AUC

ranges from 0 to 1, with greater scores indicating better discrimination ability. An AUC >0.5 indicates that the model discriminates better than random.

Results

MtDNA results

The three mtDNA genes were sequenced for 425 specimens, resulting in 263 haplotypes and 299 polymorphic sites. The average number of nucleotide differences is 17.725.

Results obtained for the three mtDNA genes were congruent (see Figure S5 for the results obtained for each gene), thus only the results obtained for the combined dataset (Cytb + CO1 + CR) are shown on Fig. 2. Several haplogroups can be identified on the network. Nearly all specimens (67/69) from Europe cluster together in the network; the only exceptions being one specimen from Grenada (Spain), which clusters with African haplotypes of several localities and one specimen from Barcelona (Spain) who shares its haplotype with one specimen from Tetouan (Morocco). Within this European haplogroup, two subgroups can be identified: one grouping all specimens from France, and one grouping most specimens from the Iberian Peninsula. In both France and the Iberian Peninsula, a clear starlike pattern is observed, suggesting a recent demographic expansion. Several haplogroups grouping together specimens from Morocco and Algeria can be identified, none of which showed a starlike pattern. In continental Europe, branches lengths are short (always <4 mutations between haplotypes), whereas in North Africa branches lengths tend to be longer (many branch lengths >6 mutations, with a maximum of 13 mutations).

Haplotype diversity is high in all localities (Table 2), with values ranging from 0.874 ± 0.048 to 0.999 ± 0.004 . Nucleotide diversity is low in the two European localities (Madrid in Spain = 0.00146 ± 0.00013 and Montbazin in France = 0.00093 ± 0.00015). In North African localities, nucleotide diversity is higher, varying from 0.00468 ± 0.00071 to 0.00742 ± 0.00064 . The high haplotype and low nucleotide diversities observed in European localities suggest a rapid population growth from an ancestral population with small N_e , provided the time was sufficient for recovery of haplotype variation via mutation yet too short for an accumulation of large sequence differences. The high haplotype and nucleotide diversities observed in North African populations suggest stable populations with large long-term N_e or admixed sample of individuals from historically sundered populations.

We tested for signatures of population size change for each population with neutrality tests (Table 2) and EBSP (Figure S6). In all North African populations, neutrality tests are nonsignificant, except in Berrechid. In Berrechid, a signal of demographic expansion is revealed by the F_s test and the EBSP (started ca 7000 ya). In the two European localities (Madrid and Montbazin), the neutrality tests and EBSP analyses reveal a signal of demographic expansion, which probably started ca 1000 ya.

The results of the differentiation analyses using F_{ST} are shown in Table S3. In total, 87% of pairwise comparisons are significant. MerjaZerga, Oued Nfifikh, BenGuerir, Ifrane and Taza are the localities that showed fewer significantly different comparisons with other populations, whereas BenSlimane, DayetErroumi, Montbazin and Sidi- Boughaba showed significant differentiation in all cases.

μ sat results

All the 18 microsatellites were polymorphic in the genotyped 536 individuals. Tests for genotypic linkage disequilibrium confirmed the null hypothesis of the independence of loci pairs. According to

locus and population, the expected heterozygosity (H_e) values vary between 0.587 and 0.744 (Table S4). After sequential Bonferroni corrections were applied, all populations were inHWE with FIS estimate ranging from 0.034 to 0.167 (Table S4).

Three models were used to test for recent genetic bottlenecks based on allele frequency data. Although tests based on the IAM mutation model suggested potential signatures of past genetic bottlenecks in samples collected in Algeria (Meftah) and Morocco (BenGuerir, BeniHadifa, BenSlimane, Berrechid, Cherouda, Clementine, Ifrane, Oued Nfifikh, SidiBoughaba and Taza), those based on the most realistic SMM mutation model were only significant for five Moroccan localities (Berrechid, Cherouda, Clementine, Oued Nfifikh and Sidi Boughaba; Table S4). In contrast, no significant values were observed using the TPM model ($P > 0.05$, Wilcoxon test).

The results of the Bayesian clustering analysis (using STRUCTURE) and Evanno's method supported $K = 4$ as the most likely number of clusters, which corresponded to a "Moroccan cluster" (yellow) including all populations from Morocco excepted Cherouda, Tetouan and Ifrane, an "Algerian cluster" (green) consisting of Algerian populations and Ifrane, a "Spanish cluster" (pink) with the population from Madrid and the Moroccan Cherouda and Tetouan populations, and a "French cluster" (blue) with the two French populations (Fig. 3a). Assuming $K = 3$, the two French populations, the Spanish population and the Moroccan populations of Cherouda and Tetouan were regrouped in the same cluster (blue, Figure S7). A second cluster includes the two Algerian populations and the Moroccan Ifrane population (green). And finally, the last cluster includes all other Moroccan populations (yellow). At K varying from 5 to 6, the Moroccan populations were gradually separated into distinct clusters. The Moroccan population from Ifrane (orange) is isolated from other Moroccan populations, and hence, the Algerian populations form an isolated cluster (green). In contrast, the Spanish population and the Moroccan Cherouda and Tetouan populations remained regrouped in the same cluster, suggesting that genetic divergence among Moroccan populations was larger than between Spanish and Moroccan Cherouda—Tetouan populations (Figure S7).

On the DAPCs (Fig. 3b), the first two axes explained >97% of the total inertia. The first axis allows to discriminate three groups of populations: the French populations (Montbazin and La Gardiole), the Spanish population (Madrid) and all African populations. On axis 2, the populations from Cherouda is distinct from all other populations. Results of the NJ tree (Fig. 3c) are congruent with those of the DAPC and confirms the close affinity between the two French populations and the distinctiveness of the Spanish population. All African populations are grouped together, and the populations from Cherouda and Tetouan are the closest to the Spanish population.

Despite the geographical distances involved, pairwise F_{ST} estimates based on microsatellite data remained relatively moderated ($F_{ST} \leq 0.152$; Table 3).

When considering all the populations, 128 of the 171 pairwise comparisons tested were significantly different from zero (Table 3). F_{ST} was not significant between: the Moroccan population Ifrane versus the two Algerian populations; the Moroccan populations Tetouan and Cherouda versus the Spanish population Madrid; between the two Algerian populations Cap Djinet versus Meftah; between the two French populations La Gardiole versus Montbazin (Table 3), which appeared homogenous, with no significant differentiation following the use of Bonferroni correction, suggesting the occurrence of extensive gene flow among these populations. Un-significant genetic differentiation was also observed among most remaining Moroccan populations (37/91 comparisons). Significant results were mainly due to the populations of BenSlimane, MerjaZerga and SidiBoughaba. Interestingly, BenSlimane seems to be an isolated population as it was significantly differentiated from the other populations except Aounate and Clementine.

Genetic inference of colonization pathways and demographic history of the three main populations (Africa, Spain, France)

We tested the potential routes of colonization of *M. spretus* from North Africa to Europe using ABC methods. Results of model selection for the two datasets are shown in Table 4, whereas cross-validation experiments are presented in Table S5. First, cross-validation suggests that the four models are not easily distinguishable. If we sum the posterior probability of models supporting independent (model 1 + model 2) versus serial (model 3 + model 4) colonization of Europe, MSAT was not able to distinguish between the two hypotheses (both having approximately $P = 0.5$), whereas MTDNA attained ~ 0.8 probability of a serial colonization (Table 4). If we consider the results from the cross-validation experiment, it is clear that such a high probability value obtained with MTDNA cannot occur if the independent colonization was the true model. Moreover, both RES-MSAT and RES-MTDNA favored the serial colonization ($P = 0.6$ and $P = 0.87$, respectively). Considering only the serial models (models 3 and 4), MSAT and RES-MSAT supported a colonization of Spain and France very close in time: ~ 5000 and ~ 3000 years BP for MSAT, respectively; and $\sim 13,000$ and $\sim 11,000$ years BP for RES-MSAT, respectively (Table 5). For MTDNA and RES-MTDNA, the posterior distribution of this parameter is flat, implying this parameter is not well estimated under such model (Table S6).

We found a higher probability of having a change in the effective population size in Africa under MSAT (model 1 + model 3 have $p \sim 0.63$), whereas a signature of constant population size under MTDNA (model 2 + model 4 have $p \sim 1$). When looking at the posterior distribution of the effective size in Africa under MSAT, we note that we have an ancestral population of $\sim 400,000$ individuals that instantaneously expand to $\sim 2,800,000$ ~ 8000 years BP and go down again to $\sim 550,000$ individuals. However, effective size 95% credible intervals are largely overlapping (model 1 and model 3, Table 5). The SRMSE of the ancestral population size, the time when the change occurred and the instantaneous size change suggest low accuracy in parameter estimation, supporting the idea that if a change occurred it did not leave a strong signature in the pattern of genetic variability (Table S6). RES-MSAT favor a constant population in Africa ($P = 0.7$). RES-MTDNA was not able to distinguish between the two hypotheses (both having approximately $P = 0.5$), but the scenario with the highest probability (MOD3, $P = 0.46$) corresponds to a scenario with population size change in Africa. Values obtained for RES-MTDNA suggest that if a population size change occurred (MOD3), it would have started from an ancestral population of $\sim 29,000$ individuals, followed by an instantaneous population expansion to $\sim 531,000$ individuals ~ 4000 years BP and then a population reduction to $\sim 120,000$ individuals. Finally, we note that the EBSP highlight a complex demography in Africa with a strong population expansion ~ 5000 years BP and a following recent bottleneck (~ 600 years BP; Fig. 4).

Independently of the model and of the markers, modern effective population size in Africa is ~ 3 to ~ 60 times higher than in Europe (Table 5), consistently with the EBSP on mtDNA (Fig. 4). EBSP and ABC (under MSAT, MTDNA, RES-MSAT and RES-MTDNA) both highlighted a recent expansion in Spain and France starting from a low effective population size (few thousands individuals, independently on the model considered, Table 5).

Species distribution modeling

We built a SDM for *M. spretus* based on the known presence localities of this species (Fig. 5). The AUC value in both analyses (Europe and Africa) is higher than 0.985, which is considered to correspond

to a useful predictive model. Under present day climatic conditions, the model reveals high climatic suitability for this species across most the Iberian Peninsula (except the western and northern parts, and the Pyrenean Mountains) and coastal Mediterranean part of France, in lowland northern Morocco (< 1000 m) from the Atlantic and Mediterranean coasts to the Atlas mountains, and in northern Algeria and Tunisia on a small coastal band of approximately 50–100 km large.

For the LGM, the CCSM and MIROC models both inferred patches of climatically suitable habitat for *M. spretus* on the Atlantic coast of Morocco, and in the southern part of the Atlas Mountains in Morocco. According to the CCSM model high habitat suitability could also have occurred, as small patches, on the Mediterranean coast at the Moroccan-Algerian border, and the Algerian-Tunisian border. Climatically suitable habitat for this species also occurred at this period in the Iberian Peninsula (except the western and northern parts, and the Pyrenean Mountains) and the coastal Mediterranean part of France.

For the mid-Holocene, the distribution of suitable habitat for *M. spretus* is predicted to be more or less similar to present day conditions, except lower habitat suitability in the Rif and High Atlas Mountains, and in the coastal Mediterranean part of France.

Discussion

When and how did the Algerian mouse arrive in Europe?

Currently, for a small terrestrial species, the Strait of Gibraltar is an important barrier to dispersal, being 14 km wide at its narrowest point and exceeding 200 m in depth. However, the fossil evidence suggests a North African origin of *M. spretus* (Stoetzel 2013) and recent spreading to southwestern Europe (Dobson 1998; Palomo et al. 2009). Previous molecular data (few allozyme markers, RFLP and CR sequences) suggested strong genetic differentiation between European and North African populations (Boursot et al. 1985; Desmarais 1989), as well as a likely recent colonization of Europe from North Africa with a strong founding effect (Desmarais 1989; Jacquart 1986). However, the timing and colonization pathways were not determined.

Here we confirmed the strong genetic isolation between African and European populations for both mitochondrial and nuclear datasets (Figs. 2 and 3, Table 3). By sampling individuals from both France and Spain, we have further shown a significant genetic divergence between these two areas with mitochondrial and nuclear data.

To gain more insights into the evolutionary history of *M. spretus*, we reconstructed the pathways of colonization of Europe from North Africa by means of ABC. Our mtDNA supported a serial colonization of Europe (i.e., first Spain and then France), whereas nuclear data conferred equal posterior probability to the serial and independent colonization scenario. We argue that the serial colonization scenario is more likely than the independent colonization model because the cross-validation experiment suggest that the high posterior probability value obtained with mtDNA is unlikely to occur if the independent colonization was the true model. Consistently, France has less genetic diversity at both markers compared with Spain. Moreover, coalescent argument can explain why nuclear data were not able to discriminate between the two scenarios under these demographic settings: the colonization of Europe has been recent and it started from a large ancestral population in North Africa with a relative mild bottleneck (see below). Both factors limited the influence of drift particularly at the nuclear level, where the effective population size is four times higher than at the mtDNA level (considering a sex ratio ~ 1). This explains why mtDNA had more power to distinguish

between models in our particular case. Another explanation of the difficulty for ABC analyses to distinguish the two scenarios (serial versus independent colonization scenario) could be due to inadequate sampling in Spain, as almost all samples used are from Madrid. Madrid is located on a high plateau in Central Spain, and does not match the typical climatic niche of *M. spretus* (Fig. 5). Thus, it was probably colonized recently. It could have been easier to discriminate the two scenarios in our ABC analyses if we had included more specimens from the South of Spain, and particularly from coastal regions. Finally, we highlight that modeling the African metapopulation as panmictic, although simplifying the ABC analyses, may come at some risks. Previous works have shown that the shape of the gene genealogy of lineages sampled from a single deme, a pool of demes, or a pool of single chromosome per deme can vary as a function of the extent of differentiation among demes from the metapopulation (Wakeley 1999; Stadler et al. 2009; Mona et al. 2013, 2014). Several biases can arise if metapopulation structure and the exact sampling scheme are not correctly modeled (Maisano Delser et al. 2016; Mazet et al. 2016), but the influence of this choice on the outcome of the model selection procedure is difficult to evaluate. For this reason, we adopted a resampling strategy to mimic a scatter sample sensu Wakeley (1999) hence removing the influence of population structure on our modeling approach. Results of both RES-MSAT and RESMTDNA confirmed that the serial colonization model is largely favored. It would be interesting to test using simulations, which of the two approaches leads to more accurate results (pooling of demes vs scatter sample), but this is beyond the scope of this work.

Based on our ABC analyses, the microsatellites data support a colonization of Spain and France very close in time, between ~5000 and ~3000 years BP for MSAT, respectively, and between ~13,000 and ~11,000 years BP for RES-MSAT. Credible intervals are quite large and more data are needed to refine the estimates. These parameters were poorly estimated with ABC from mtDNA. All our analyses (neutrality tests, EBSP and ABC) highlight a recent demographic expansion in Spain and France. Our ecological niche modeling analyses suggest that suitable climatic conditions occurred throughout most central and eastern Iberian Peninsula during the last 6000 years (Fig. 5). This would have favored the demographic and spatial expansion of the Algerian mice after its arrival in Europe. It would have first spread throughout Spain and then in the South of France. The lower genetic diversity observed in French *M. spretus* populations compared with Spanish ones is consistent with this hypothesis, as well as the branching pattern observed in the NJ tree (Fig. 3c) and the haplotype network obtained for the Cyt b gene (Figure S5A). This hypothesis is further strengthened by the study of Jacquart (1986) who observed, based on allozyme data, a gradual diminution of the enzymatic polymorphism in *M. spretus* from southern Spain to France, which could be explained by founder events.

Our genetic and ecological niche modeling data are consistent with fossil data (Palomo et al. 2009) and favor an anthropogenic translocation from North Africa to the Iberian Peninsula via Neolithic navigators. During the Neolithic and after, numerous contacts between the two shores of the Mediterranean Sea occurred (seafaring evidence attested at ca 7000–9000 ya), due to the development of navigation (Mulazzani et al. 2010; Souville 1998; Zilhão 2014). Moreover, our genetic data show that the ancestral population size (before expansion) comprised a high number of individuals, reinforcing the idea that the Algerian mouse did not colonize the Iberian Peninsula through rafting of a few individuals, but via recurrent temporal anthropogenic translocations. These translocations are likely to have been accidental, mice being transported with agricultural and timber products (Brahmi et al. 2010; Dobson 1998). Several other mammals were also probably translocated by humans from one shore of the Mediterranean Sea to the other during the Neolithic period (Dobson 1998). Moreover from the beginning of the Neolithic to the present day, the turnover of mammal

species is almost complete in the Mediterranean islands: most of the mammal fauna has been exterminated and replaced by a diverse assemblage of introduced species (Vigne 1992).

At first sight, it could be surprising that the Algerian mice colonized Europe via anthropogenic translocation several thousand years ago and that its dispersal between the two continents did not continue until today since maritime trade has increased. A similar phenomenon was previously observed in another rodent species, the wood mouse (Lalis et al. 2016; Stoetzel 2017). One hypothesis could be that at the beginning of the Holocene these two species were more commensal than today. Molecular and zooarchaeological data showed that the commensal species *Mus musculus* only reached Western Europe and North Africa during the first millennium BC and onwards, related to the generalization of maritime trade (Bonhomme et al. 2011). The arrival of the house mouse at this time could have led to a shift in the degree of commensality of *M. spretus*. Our mtDNA network analyses show that nearly all specimens of *M. spretus* from Europe cluster together in the network, with two exceptions (two specimens from Spain cluster with African specimens). This result suggests that one main event of colonization of Europe occurred several thousand years ago, but that several limited exchanges between Africa and Spain could also have occurred more recently. Our microsatellite results show a strong affinity between the Spanish population from Madrid (locality 49 on Fig. 1) and the Moroccan Cherouda (locality 11) and Tetouan (locality 32) populations. These two Moroccan populations are located in the Tingitane Peninsula, close to the Gibraltar Strait, where the sea distance between Africa and Iberia is the lowest. Even if the navigation is not easy in this area due to the presence of strong ocean currents and counter currents, human navigation probably occurred across the Strait of Gibraltar at the Neolithic (Souville 1998; Zilhão 2014). Thus, it could be hypothesized that the Algerian mice colonized Iberia from the Tingitane Peninsula.

Can we detect a recent population bottleneck in North African populations linked to the development of anthropogenic activities and/or the arrival of the house mouse?

The El Harhoura 2 cave succession (Morocco, region of Témara) showed a continuous fossil record for *M. spretus* from 107 ± 7 kyr BP to 5800 years BP, with no significant change in abundance through time (Stoetzel et al. 2011a). Our climatic niche modeling analyses show that, according to the MIROC model, climatically suitable habitats (high or medium probability of occurrence) probably occurred around Témara throughout the Pleistocene (Fig. 4), which could explain the demographic stability observed in El Harhoura 2 cave.

An important molar size decrease was observed between the El Harhoura 2 Pleistocene and Holocene levels, including the most recent ones (5800 years BP) on one hand, and modern Moroccan mouse populations on the other hand, which contrasts with the stability observed through the long-term fossil record of the El Harhoura 2 cave. It was hypothesized that this could be the result of the competition with the house mouse *M. m. domesticus*, which only reached northwestern Africa and the Iberian Peninsula in the last millennium BC (Bonhomme et al. 2011), together with anthropogenic changes in landscape (Stoetzel et al. 2013). Important differences in the physiological, spatial and social behavior of the two species exist. It was shown that when food resources are scarce or variable in time, *M. m. domesticus* is favored and is the only of the two species being able to colonize partly covered habitats (Cassaing and Croset 1985). Moreover, in wild enclosure where both species co-occur only *M. m. domesticus* is able to reproduce normally (Orsini et al. 1982). Thus, we could expect to find a recent bottleneck (after 3000 ya) in Algerian mice African populations, after the arrival of the house mouse. According to our ABC microsatellite analyses, we found a higher probability of having a change in the

effective population size in Africa ($p \sim 0.65$), with an instantaneous expansion followed by a bottleneck. However, when looking at the posterior distribution of the effective sizes, we note that 95% credible intervals are largely overlapping. Results of our tests for recent genetic bottlenecks based on allele frequency data are also inconclusive: tests based on the IAM mutation model suggested potential signatures of past genetic bottlenecks in samples collected in Algeria and Morocco, whereas those based on the most realistic SMM or TPM mutation models were significant for few or no localities. Based on our ABC analyses, a signature of constant population size was observed for mtDNA based on all specimens ($p \sim 1$). However, the EBS highlight a more complex demography in Africa with a strong population expansion ~ 5000 years BP (in agreement with our climatic niche modeling analyses) and a following recent bottleneck (~ 600 years BP; Fig. 4 and Figure S6). RES-MSAT favor a constant population size ($p \sim 0.7$), and RES-MTDNA tend to favor a population size change. Results on the recent demography of *M. spretus* in North Africa are therefore not conclusive. Our ABC framework could have failed to retrieve the recent bottleneck because we modeled only a change of effective size in North Africa: we detected the oldest expansion as it had more time to shape the genetic diversity. Another explanation is that likelihood approaches have more power to detect more recent events than approximate methods (Girod et al. 2011): bottlenecks are particularly difficult to identify when the N_e before the decrease is large and the intensity of the bottleneck is mild, as in our case (Table 5). If we assume that this bottleneck occurred at the mtDNA level, it remains to be understood why nuclear data do not support the same conclusion. One possibility is that nuclear loci having a larger N_e than mtDNA, it renders harder to detect recent changes of effective size. The power to detect subtle and/or recent changes in the effective size increases with the number of loci studied: genome-wide approaches (i.e., whole-genome sequencing, Rad-seq, transcriptomics) will help us to shed light on this issue in the very next future.

Is the high morphological variability observed in present day populations due to isolation?

Geographical differences in body size, molar or skull morphology have been documented across the modern distribution range of *M. spretus* (Darviche and Orsini 1982; Darviche et al. 2006; Khammes-El Homsni and Aulagnier 2010; Palomo et al. 1985; Valenzuela-Lamas et al. 2011). The observed morphological variability could be hypothesized to be due to relative isolation between populations, leading to divergence (founder effect, drift) and possibly local adaptation. Our genetic results suggest that these phenomena could easily explain the morphological differentiation observed between the Spanish, French and North African populations.

Our climatic niche modeling analyses reveals that the LGM cold and dry period was unfavorable for *M. spretus*: climatically suitable habitat occurred mostly on the Atlantic coast of Morocco, and possibly also as small patches on the Mediterranean coast at the Moroccan-Algerian border, and at the Algerian-Tunisian border. Based on these results, one could expect to find genetically distinct populations in Algeria and Atlantic Morocco. However, our genetic data do not indicate a strong phylogeographic structure in North Africa. To further test this hypothesis, we run an additional model under the ABC framework (data not shown): we simulated an ancestral population in North Africa that around LGM splits into two. After $\sim 15,000$ years of separation, the two populations merged ~ 3000 years BP. For each model (1–4), we run the corresponding split and merge alternative. All pairwise comparison for both MSAT and MTDNA returned a posterior probability ~ 0.50 , suggesting that either we do not have enough discriminatory power with the present data, or the two hypotheses do not lead to large differences in the predicted diversity. Further test and more genomic data will help clarify

this issue, but our present data do not allow us to conclude that the geographical variability in morphology observed in North Africa (Khammes-El Homsni and Aulagnier 2010; Stoetzel et al. 2013) can be explained by long-term isolation. Numerous environmental factors such as diet, habitat, elevation, temperature, precipitation, urbanization and pollution can affect morphometric variation in rodents (Caumul and Polly 2005; Khammes-El Homsni and Aulagnier 2010; Nunes et al. 2001). More detailed analyses incorporating these factors are necessary to fully understand factors triggering phenotypic variability in the Algerian mice.

To conclude, our genetic and ecological niche modeling data are consistent with fossil and human context data, and favor an anthropogenic translocation of the Algerian mouse from the Tingitane Peninsula to the Iberian Peninsula via Neolithic navigators. More genetic data should be added to strengthen this finding. Once arrived in Spain, suitable climatic conditions would then have favored the dispersion of the Algerian mice to France. Finally, the morphological differentiation observed between the Spanish, French and North African populations could be due to founder effect and possibly local adaptation. This article highlights the importance of both climate and human-mediated changes on the evolutionary history of mammal species in the Mediterranean Basin.

Acknowledgements

The present study was supported by ANR 6eme extinction ANR-09-PEXT-004 MOHMIE “Modern Human installation in Morocco, Influence on the small terrestrial vertebrate biodiversity and Evolution” and the project CMEP TASSILI MDU09MDU755. Fieldwork was made possible through collaboration with the “Institut Scientifique de Rabat” (A. El Hassani, M. Fekhaoui), the University of Tetouan (Department of Biology), and the “Haut Commissariat aux Eaux et Forêts et Lutte contre la Désertification”. We are grateful to all collectors, particularly Abderrahmane Mataame, Hicham El Brini, Arnaud Delapré, Sohaib Liefried Loubna Tifarouine, Adel Hamani, Lylia Amrouche Larabi and Yolanda Fernandez-Jalvo. Molecular analyses were financially supported by the “ATM MNHN: Taxonomie moléculaire, DNA Barcode & Gestion Durable des Collections”, the “Service de Systématique Moléculaire” of the MNHN (UMS 2700, Paris, France), particularly Josie Lambourdière for her help in the microsatellites analysis, and the network “Bibliothèque du Vivant” funded by the CNRS, MNHN, INRA and CEA (Genoscope). We thank the Genotoul bioinformatics platform Toulouse Midi- Pyrenees (Bioinfo Genotoul) for providing computing resources. We are grateful to three anonymous reviewers for their helpful comments on an earlier version of this paper.

Data archiving

Sequence data available from GenBank: accession numbers MH241060 to MH241525 for Cytb, MH241526 to MH242042 for CR, MH242043 to MH242499 and MH253776 to MH253817 for CO1 sequences.

Microsatellite data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.3sv4634>.

Compliance with ethical standards

Conflict of interest : The authors declare that they have no conflict of interest.

References

- Audoin-Rouzeau F, Vigne JD (1994) La colonisation de l'Europe par le rat noir (*Rattus rattus*). *Rev Paléontol* 13:125–145
- Aulagnier S, Cuzin F, Thévenot M (2017) Mammifères sauvages du Maroc: peuplement, répartition, écologie. Société française pour l'étude et la protection des mammifères: Paris.
- Bandelt HJ, Forster P, Rohlf A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1):37–48
- Beaumont MA (2008) Joint determination of topology, divergence time, and immigration in population trees. In: Matsumura S, Forster P, Renfrew C (eds) *Simulation, genetics, and human prehistory*. McDonald Institute for Archaeological Research, Cambridge, p 135–154
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol Ecol* 19(13):2609–2625
- Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P et al. (2011) Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. *Proc Biol Sci* 278(1708):1034–1043
- Boursot P, Jacquart T, Bonhomme F, Britton-Davidian J, Thaler L (1985) Différenciation géographique du génome mitochondrial chez *Mus spretus* Lataste. *C R Acad Sci Paris III* 301(5):157–161.
- Brahmi K, Khechekhouche EA, Mostefaoui O, Bebbi K, Hadjoudj M, Doumandji S et al. (2010) Extralimital presence of small mammals in north-eastern Algerian Sahara. *Mammalia* 74:105–108
- Brahmi K, Khechekhouche EA, Mostefaoui O, Doumandji S, Baziz B, Aulagnier S (2012) First quantitative data on the diet of the fennec fox, *Vulpes zerda* (Canidae, Carnivora), in Algeria. *Folia Zool* 61(1):61–70
- Brun A (1989) Micoflores et paléovégétations en Afrique du Nord depuis 30 000 ans. *Bull Soc Géol Fr* 1:25–33
- Brun A (1991) Réflexions sur les Pluviaux et Arides au Pléistocène supérieur et à l'Holocène en Tunisie. *Palaeoecol Afr* 22:157–183
- Cassaing J, Croset H (1985) Organisation spatiale, compétition et dynamique des populations de Souris (*Mus spretus* Lataste et *Mus musculus domesticus* Ruddy) du Midi de la France. *Z Säugetierk* 50(5):271–284
- Caumul R, Polly PD (2005) Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (*Marmota*, Rodentia). *Evolution* 59 (11):2460–2472
- Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, et al. (2006) The Community Climate System Model version 3 (CCSM3) *J. Climate* 19:2122–2143
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144(4):2001–2014
- Csillery K, Francois O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol* 3 (3):475–479
- Cucchi T, Vigne JD (2006) Origin and diffusion of the house mouse in the Mediterranean. *Human Evol* 21:95–106
- Darviche D, Orsini P (1982) Critères de différenciation morphologique et biométrique de deux espèces de souris sympatriques: *Mus spretus* et *Mus musculus domesticus*. *Mammalia* 46:205–217
- Darviche D, Orth A, Michaux J (2006) *Mus spretus* and *M. musculus* (Rodentia, Mammalia) in the Mediterranean zone: biometric and morphological differentiation: application to Pleistocene Moroccan fossils. *Mammalia* 70(1-2):90–97

- Desmarais E (1989) Phylogénies intraspécifiques et histoire évolutive des populations de souris *Mus spretus* Lataste: analyse des lignées matriarcales de l'ADN mitochondrial. PhD Thesis, Université de Montpellier, Montpellier.
- Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A et al. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380(6570):152–154
- Dobson M (1998) Mammal distributions in the western Mediterranean: the role of human intervention. *Mammal Rev* 28(2):77–88
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
- Ducroz JF, Volobouev V, Granjon L (2001) An assessment of the systematics of arvicanthine rodents using mitochondrial DNA sequences: evolutionary and biogeographical implications. *J Mammal Evol* 8:173–206
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- Elith J, Kearney M, Phillips S (2010) The art of modelling range shifting species. *Meth Ecol Evol* 1(4):330–342
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14(8):2611–2620
- Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9(10):e1003905
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587
- Fourcade Y, Engler JO, Rodder D, Secondi J (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS ONE* 9(5):e97122
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147(2):915–925
- Garcea E (2012) Successes and failures of human dispersals from North Africa. *Quatern Int* 270:119–128
- Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis. Chapman & Hall/CRC, London, UK
- Girod C, Vitalis R, Leblois R, Freville H (2011) Inferring population decline and expansion from microsatellite data: a simulation based evaluation of the MSVAR method. *Genetics* 188:165–179
- Hassan FA (1997) Holocene palaeoclimates of Africa. *Afr Archaeol J* 14:213–230
- Hasumi H, Emori S (eds) (2004) K-1 coupled GCM (MIROC) description. K-1 model developers tech Rep 1. Center for Climate System Research, University of Tokyo, Tokyo, Japan, p 34
- Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evol Biol* 8:289
- Henn BM, Gignoux CR, Feldman MW, Mountain JL (2009) Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol Biol Evol* 26(1):217–230
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978

- Ho SY, Kolokotronis SO, Allaby RG (2007a) Elevated substitution rates estimated from ancient DNA sequences. *Biol Lett* 3 (6):702–705
- Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007b) Evidence for time dependency of molecular rate estimates. *Syst Biol* 56(3):515–522
- Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MW, Bergmann I, Le Cabec A, Benazzi S, Harvati K, Gunz P(2017) New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature* 546:289–292
- Jacquart T (1986) Structure génétique et phylogénie intraspécifique chez la souris sauvage *Mus spretus* Lataste: distribution spatiale du polymorphisme des gènes nucléaires de structure et de l'ADN mitochondrial. PhD Thesis thesis, Université des Sciences et Techniques du Languedoc, Montpellier.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23 (14):1801–1806
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405
- Khammes-El Homsy N, Aulagnier S (2010) Unexpected morphometric differentiation of the Algerian mouse, *Mus spretus* (Rodentia: Muridae) from Kabylie of Djurdjura (Algeria). *Mammalia* 74 (2):199–207
- Lalis A, Leblois R, Liefried S, Ouarour A, Beeravolu C, Michaux J et al. (2016) New molecular data favour an anthropogenic introduction of the wood mouse (*Apodemus sylvaticus*) in North Africa. *J Zool Syst Evol Res* 54(1):1–15
- Lalis A, Orth A (2014) A set of 18 polymorphic microsatellite loci in the Algerian mouse, *Mus spretus* Lataste 1883 (Rodentia: Muridae). *Conserv Genet Res* 6(2):255–258
- Langella O (2010) CNRS, UPR9034, France Populations.1.2.3.1 Available at <http://bioinformatics.org/populations/>.
- Le Houérou HN (1997) Climate, flora and fauna changes in the Sahara over the past 500 million years. *J Arid Environ* 37:619–647
- Librado P, Rozas J (2009) DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25 (11):1451–1452
- Loogväli EL, Kivisild T, Margus T, Villems R (2009) Explaining the imperfection of the molecular clock of hominid mitochondria. *PLoS ONE* 4(12):e8260
- Maisano Delsler P, Corrigan S, Hale M, Li C, Veuille M, Planes S et al. (2016) Population genomics of *C. melanopterus* using target gene capture data: demographic inferences and conservation perspectives. *Sci Rep* 6:33753
- Martinkova N, Barnett R, Cucchi T, Struchen R, Pascal M, Pascal M et al. (2013) Divergent evolutionary processes associated with colonization of offshore islands. *Mol Ecol* 22(20):5205–5220
- Mazet O, Rodriguez W, Grusea S, Boitard S, Chikhi L (2016) On the importance of being structured: instantaneous coalescence rates and human evolution-lessons for ancestral population size inference? *Heredity* 116(4):362–371
- Mona S, Mordret E, Veuille M, Tommaseo-Ponzetta M (2013) Investigating sex-specific dynamics using uniparental markers: West New Guinea as a case study. *Ecol Evol* 3(8):2647–2660
- Mona S, Ray N, Arenas M, Excoffier L (2014) Genetic consequences of habitat fragmentation during a range expansion. *Heredity* 112:291–299

- Mulazzani S, Le Bourdonnec F-X, Belhouchet L, Poupeau G, Zoughlami J, Dubernet S et al. (2010) Obsidian from the Epipalaeolithic and Neolithic eastern Maghreb. A view from the Hergla context (Tunisia). *J Archaeol Sc* 37(10):2529–2537
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York, NY
- Nunes AC, Auffray JC, Mathias ML (2001) Developmental instability in a riparian population of the Algerian mouse (*Mus spretus*) associated with a heavy metal-polluted area in central Portugal. *Arch Environ Contam Toxicol* 41:515–521
- Orsini P, Cassaing J, Duplantier J-M, Croset H (1982) Premières données sur l'écologie des populations naturelles de souris, *Mus spretus* Lataste et *Mus musculus domesticus* Ruddy dans le midi de la France. *Rev Ecol (Terre Et Vie)* 36(3):321–336
- Palomo LJ, Antúñez A, Ibanez C, Vargas JM, Sans-Coma V (1985) Estudio taxonomico de *Mus spretus* Lataste, 1883 del norte de Marruecos. *Misc Zool* 9:367–374
- Palomo LJ, España M, López-Fuster MJ, Gosalbez J, Sans-Coma V (1983) Sobre la variabilidad genetica y morfometrica de *Mus spretus* Lataste, 1883 en la Peninsulalberica. *Misc Zool* 7:171–192
- Palomo LJ, Justo ER, Vargas JM (2009) *Mus spretus* (Rodentia: Muridae). *Mamm Species* 840:1–10
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modelling of species geographic distributions. *Ecol Model* 190:231–239
- Piry S, Luikart G, Cornuet JM (1999) Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Hered* 90:502–503
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- R Development Core Team (2016) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna
- Radosavljevic A, Anderson RP (2014) Making better MAXENT models of species distributions: complexity, overfitting and evaluation. *J Biogeogr* 41:629–643
- Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe. *Mol Ecol* 17 (2):627–641
- Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19 (12):2092–2100
- Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution* 49:1280–1283
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43:223–225
- Richter D, Grün R, Joannes-Boyau R, Steele TE, Amani F, Rué M, Fernandes P, Raynal JP, Geraads D, Ben-Ncer A, Hublin JJ, McPherron SP (2017) The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature* 546(7657):293–296
- Robins JH, Hingston M, Matisoo-Smith E, Ross HA (2007) Identifying *Rattus* species using mitochondrial DNA. *Mol Ecol Notes* 7 (5):717–729
- Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138
- Schorr G, Holstein N, Pearman PB, Guisan A, Kadereit JW (2012) Integrating species distribution models (SDMs) and phylogeography for two species of Alpine *Primula*. *Ecol Evol* 2 (6):1260–1277
- Sheridan JA, Bickford D (2011) Shrinking body size as an ecological response to climate change. *Nat Clim Change* 1:401–406
- Sikes RS, Mammalogists atACaUCotASo (2016) 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *J Mammal* 97(3):663–688

- Sondaar PY (2000) Early human exploration and exploitation of islands. *Tropics* 10:203–230
- Souville G (1998) Contacts et échanges entre la péninsule Ibérique et le Nord-Ouest de l’Afrique durant les temps préhistoriques et protohistoriques. *CR Acad Inscr Belle* 142(1):163–177
- Stadler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P (2009) The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* 182:205–216
- Stambouli-Essassi S, Roche E, Bouzid S (2007) Evolution de la végétation et du climat dans le nord-ouest de la Tunisie au cours des 40 derniers millénaires. *Geo-Éco-Trop* 31:171–214
- Stoetzel E (2013) Late Cenozoic micromammal biochronology of northwestern Africa. *Palaeogeogr Palaeoclimatol Palaeoecol* 392:359–381
- Stoetzel E (2017) Adaptations and dispersals of anatomically modern humans in the changing environments of North Africa: the contribution of microvertebrates. *Afr Archaeol Rev* 34:453–468
- Stoetzel E, Denys C, Bailon S, El Hajraoui MA, Nespoulet R (2011a) Taphonomic analysis of amphibian and squamate remains from El Harhoura 2 (Rabat-Témara, Morocco): contributions to palaeoecological and archaeological interpretations. *Int J Osteoarchaeol*. <https://doi.org/10.1002/oa.1275>
- Stoetzel E, Denys C, Michaux J, Renaud S (2013) Mus in Morocco: a quaternary sequence of intraspecific evolution *Biol J Linn Soc* 109:599–621
- Stoetzel E, Marion L, Nespoulet R, El Hajraoui MA, Denys C (2011b) Taphonomy and palaeoecology of the late Pleistocene to middle Holocene small mammal succession of El Harhoura 2 cave (Rabat-Temara, Morocco). *J Hum Evol* 60(1):1–33
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144 (1):389–399
- Valenzuela-Lamas S, Baylac M, Cucchi T, Vigne JD (2011) House mouse dispersal in IronAge Spain: a geometric morphometrics appraisal. *Biol J Linn Soc* 102:483–497
- Varela S, Lima-Ribeiro MS, Terribile LC (2015) A short guide to the climatic variables of the last glacial maximum for biogeographers. *PLoS ONE* 10(6):e0129037
- Vargas JM, Palomo LJ, Palmqvist P (1991) Reproduction of the Algerian mouse (*Mus spretus* Lataste, 1883) in the south of the Iberian Peninsula. *Bonn Zool Beitr* 42(1):1–10
- Vigne JD (1992) Zooarchaeology and the biogeographical history of the mammals of Corsica and Sardinia, since the last IceAge. *Mamm Rev* 22:87–96
- Wakeley J (1999) Nonequilibrium migration in human history. *Genetics* 153:1863–1871
- Walther BA, Moore JL (2005) The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28:815–829
- Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33(3):607–611
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Zilhão J (2014) Early prehistoric navigation in the western Mediterranean: implications for the Neolithic transition in Iberia and the Maghreb. *Eurasia Prehistory* 11(1-2):185–200

FIGURES

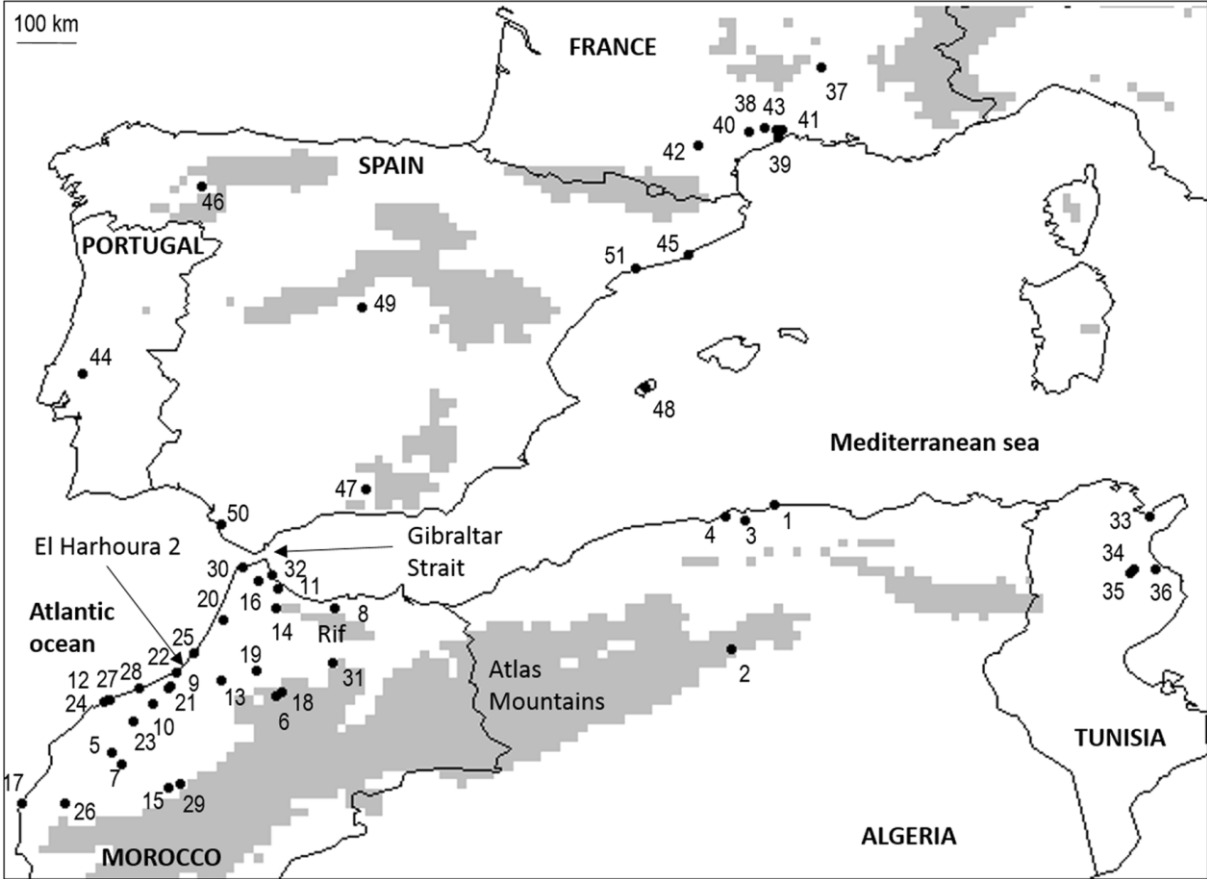


Fig. 1. Map of sampling localities and the names of the main geographical features discussed in the text. For locality names, see Table 1. Gray areas indicate mountains higher than 1000 m

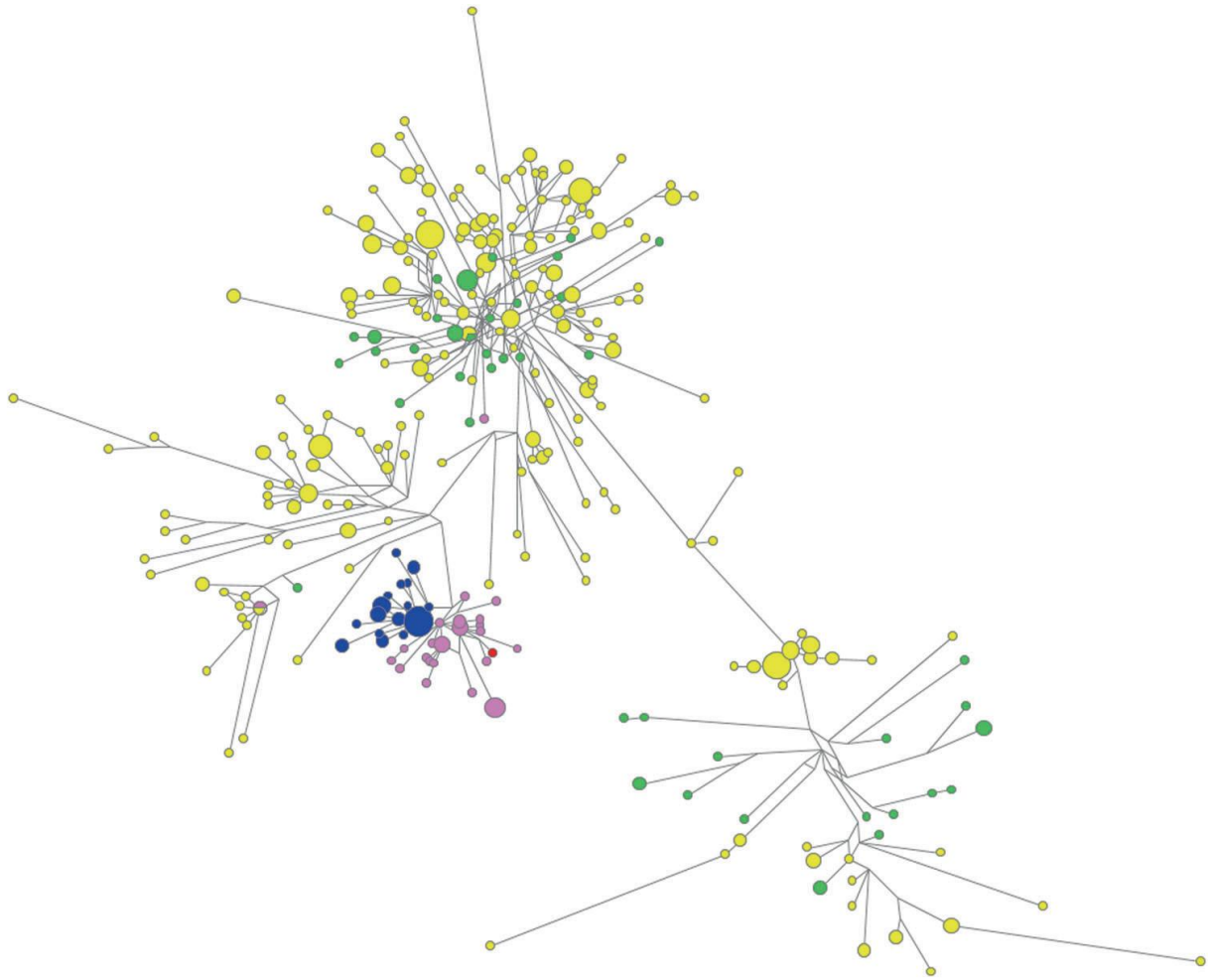


Fig. 2. Minimum spanning networks of *M. spretus* obtained for the three mtDNA regions concatenated. Circle sizes are proportional to the number of similar haplotypes observed in the dataset. Branch lengths are proportional to the number of mutations between haplotypes. Colors represent the country of sampling (red = Portugal; blue = France; pink = Spain; yellow = Morocco; green = Algeria).

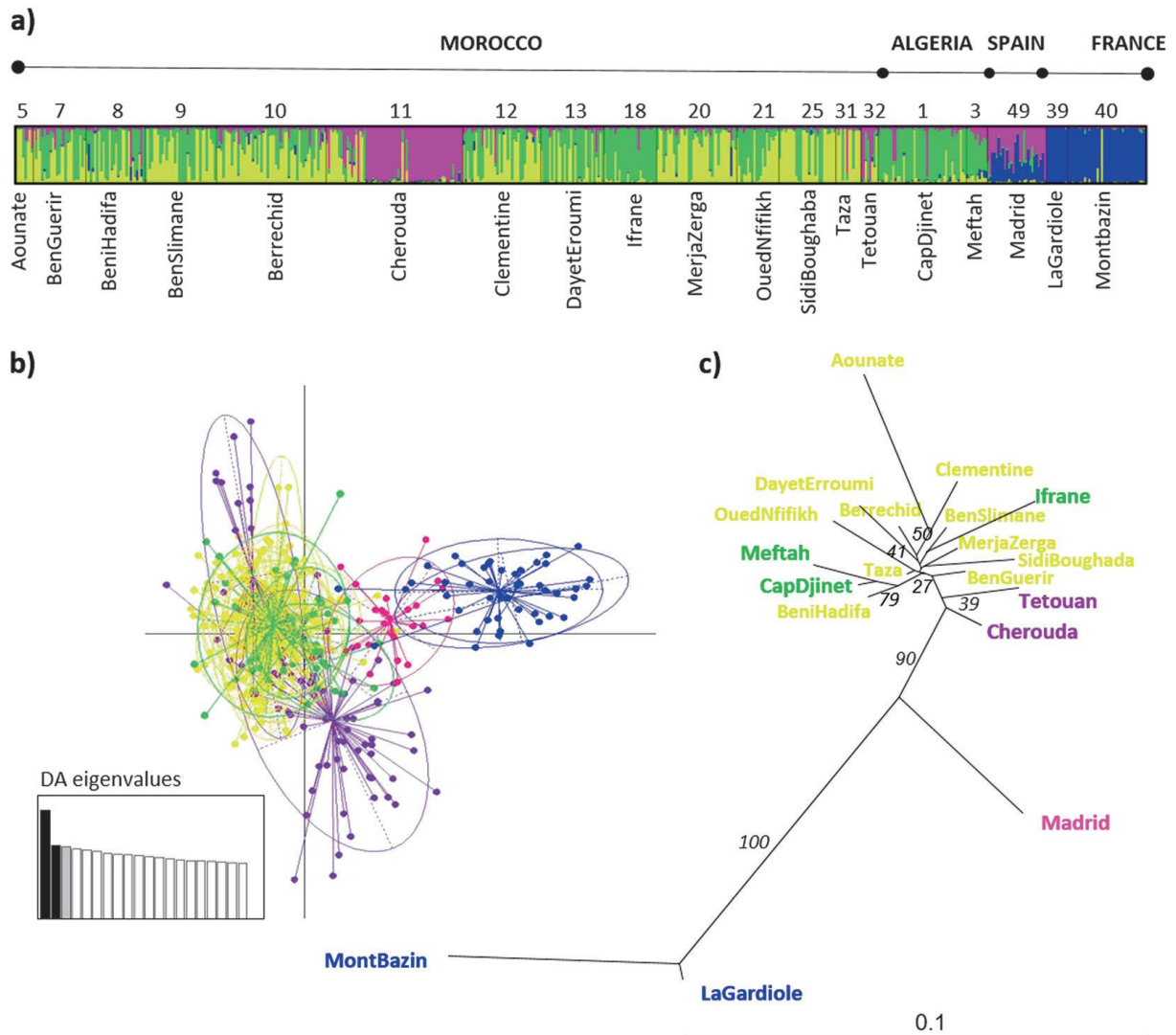


Fig. 3. a: Bayesian clustering analyses of *M. spretus* populations genotyped with 18 microsatellite loci. Clustering was performed in STRUCTURE. Each vertical line represents an individual, and each color represents a cluster (in blue: France; in pink: Madrid—Cherouda —Tetouan; in green: Algerian populations and Ifrane, and in yellow: the others Moroccan populations). Individuals are grouped by sampling location (locality number as in Table 1). **b:** Discriminant analysis of principal components (axis 1 and 2). The axes represent the first two discriminant functions (DA) respectively. Individuals (dots) and populations (colored ellipses) are plotted within the orthogonal space defined by the first two 2 DA eigenvalues (inserts). **c:** Neighbor-joining tree based on Reynold's distance. Bootstrap values are calculated over 1000 replicates. For b, c, each color represents a cluster: in blue: France; in pink: Madrid; in purple: Moroccan populations Cherouda and Tetouan; in green: Algeria and in yellow: others Moroccan populations.

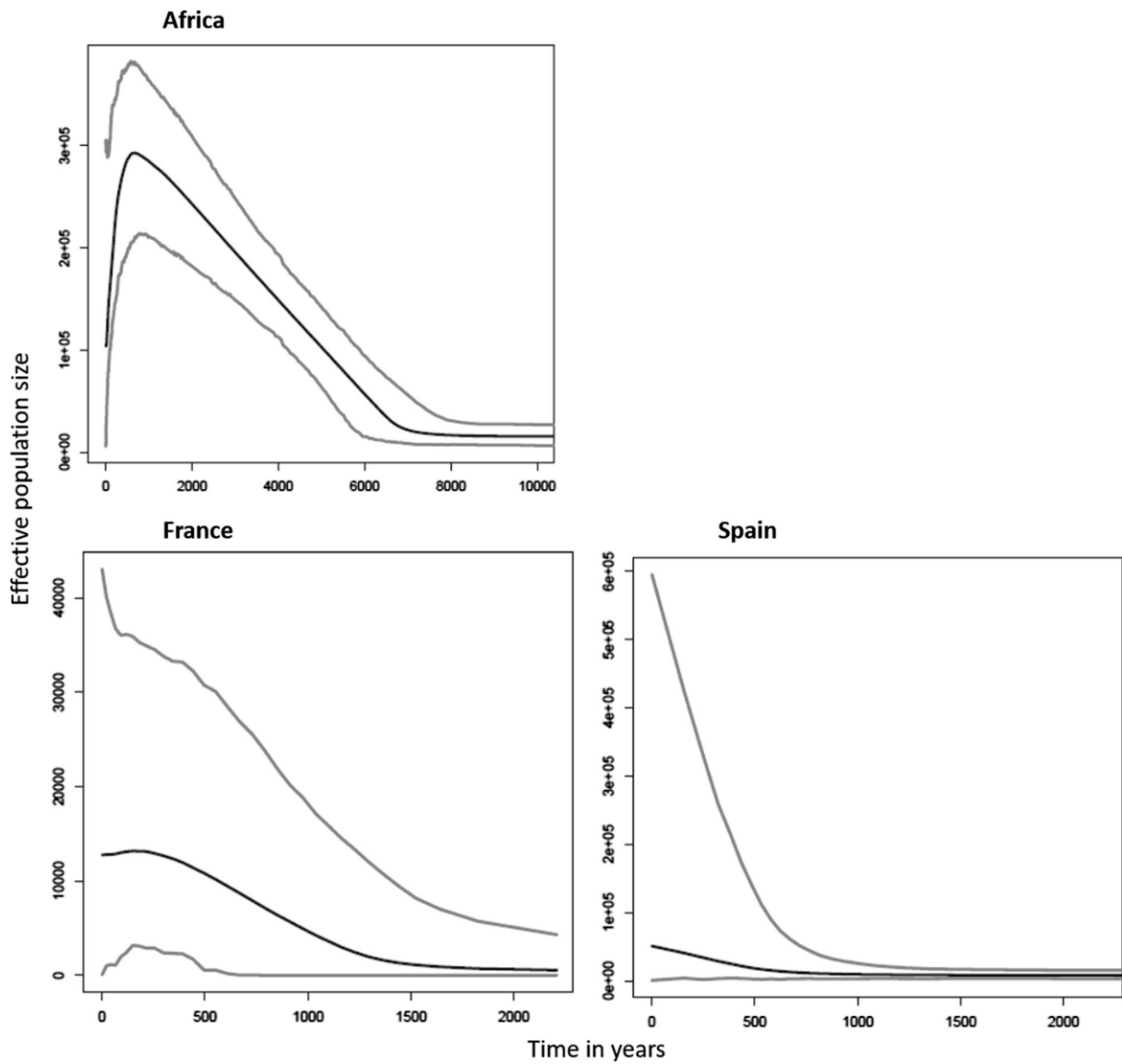


Fig. 4. Extended Bayesian skyline plot (mtDNA) computed in Africa, France and Spain. X axis: times in years; y axis: effective population size. Median values are shown in black, 95% high posterior density in red

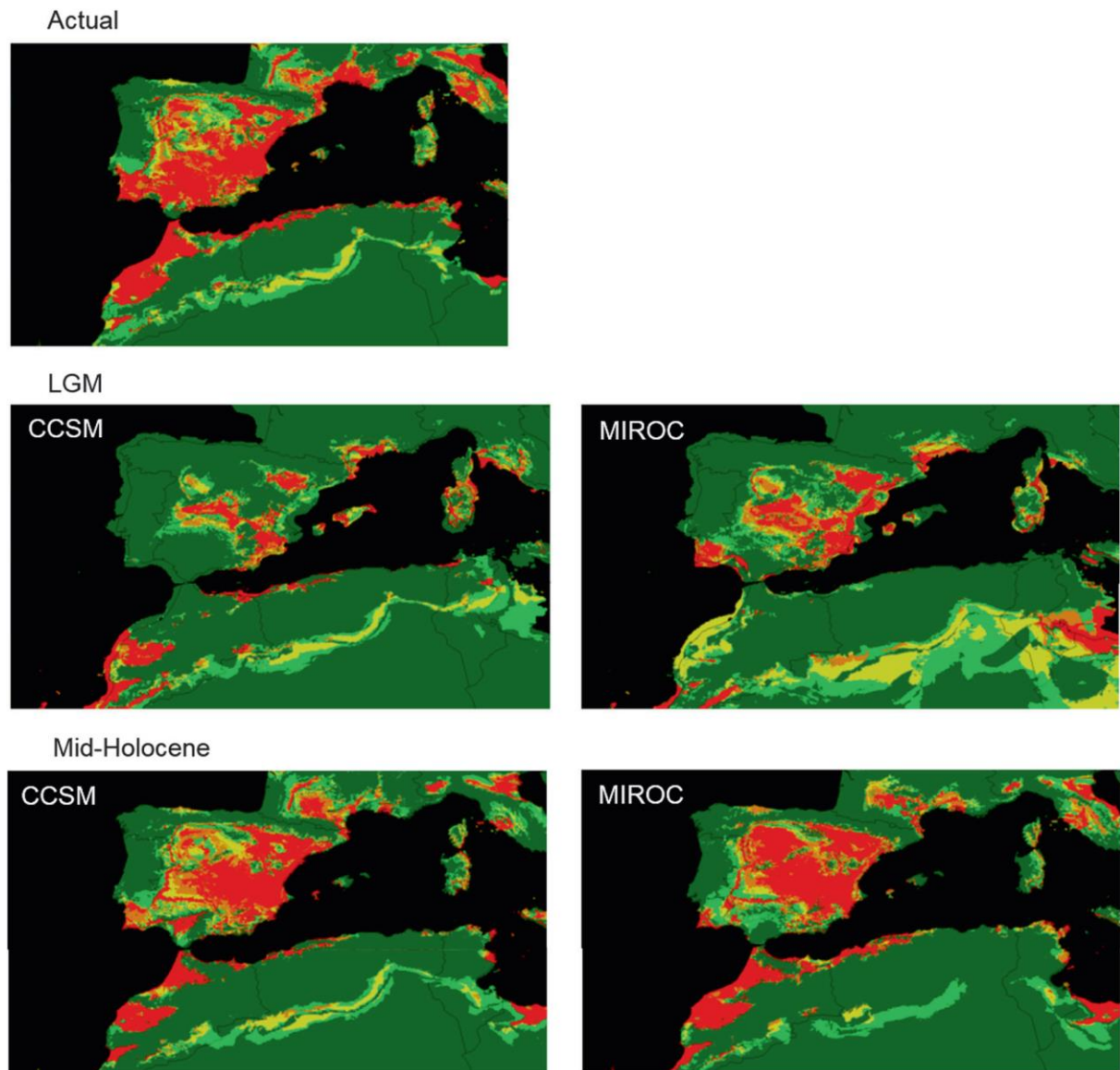


Fig. 5. Species distribution modeling of *M. spretus* as estimated by Maxent for present day conditions (actual), for the last glacial maximum (LGM) and for the mid-Holocene. For paleoclimatic data, estimates are based on the Model for Interdisciplinary Research on Climate (MIROC) and on the Community Climate System Model (CCSM) paleoclimatic models. Warmer colors show areas with higher probability of presence.

TABLES

Continent	Country	Locality	Locality number in Fig. 1	Latitude	Longitude	cytb	COI	CR	Three mtDNA regions	Microsatellites
Africa	Algeria	Cap Djinet	1	36.878	3.721	48	44	52	34	42
Africa	Algeria	Hadjeb El Djemel	2	34.300	2.967	8	6	8	6	-
Africa	Algeria	Meftah	3	36.616	3.216	9	10	8	8	10
Africa	Algeria	Zeralda	4	36.691	2.836	5	5	5	5	-
Africa	Morocco	Aouinate (OuledTounsi)	5	32.450	-8.150	8	7	8	7	8
Africa	Morocco	Azrou	6	33.440	-5.220	-	1	1	-	-
Africa	Morocco	BenGuerir	7	32.220	-7.970	28	28	28	27	25
Africa	Morocco	BeniHadifa	8	35.024	-4.162	29	29	28	28	28
Africa	Morocco	BenSlimane	9	33.635	-7.106	35	34	34	34	34
Africa	Morocco	Berrechid	10	33.296	-7.410	52	52	48	48	52
Africa	Morocco	Cherouda	11	35.371	-5.173	7	8	8	7	65
Africa	Morocco	Clementine	12	33.383	-8.233	9	18	19	5	37
Africa	Morocco	DayetErroumi	13	33.742	-6.189	29	29	31	29	30
Africa	Morocco	EIKhizana	14	35.030	-5.218	3	3	3	3	-
Africa	Morocco	ElSahrij	15	31.819	-7.130	-	1	1	-	-
Africa	Morocco	Esperada	16	35.535	-5.537	1	1	1	1	-
Africa	Morocco	Essaouira	17	31.510	-9.760	-	1	1	-	-
Africa	Morocco	Ifrane	18	33.514	-5.126	23	25	25	23	25
Africa	Morocco	Meknes	19	33.900	-5.550	-	-	-	-	-
Africa	Morocco	MerjaZerga	20	34.821	-6.180	35	32	37	28	38
Africa	Morocco	Oued Nfifikh	21	33.600	-7.130	17	21	18	14	20
Africa	Morocco	OuedYquem	22	33.886	-7.000	1	1	1	1	-
Africa	Morocco	OuladSaïd	23	33.003	-7.791	6	6	6	6	-
Africa	Morocco	Sasma	24	33.350	-8.317	-	1	-	-	-
Africa	Morocco	SidiBoughaba	25	34.238	-6.673	25	24	22	22	27
Africa	Morocco	Sidi' L' Moctar	26	31.533	-8.999	1	1	1	1	-
Africa	Morocco	Sodea	27	33.390	-8.190	-	1	1	-	-
Africa	Morocco	Tamaris	28	33.580	-7.660	-	-	-	-	-
Africa	Morocco	Tanant	29	31.860	-6.950	-	-	-	-	-
Africa	Morocco	Tanger	30	35.780	-5.810	1	6	3	1	-
Africa	Morocco	Taza	31	34.047	-4.186	13	13	14	13	12
Africa	Morocco	Tetouan	32	35.640	-5.280	7	8	8	7	8
Africa	Tunisia	Fondouk Djedid	33	36.671	10.448	-	-	1	-	-
Africa	Tunisia	Hanomed	34	35.722	10.199	-	-	1	-	-
Africa	Tunisia	Kairouan	35	35.673	10.099	-	3	2	-	-
Africa	Tunisia	M'Saken	36	35.731	10.582	-	-	-	-	-
Europe	France	Ardèche	37	44.760	4.560	2	1	2	1	-
Europe	France	Gignac	38	43.650	3.550	5	5	5	5	-
Europe	France	La Gardiole	39	43.490	3.800	1	5	10	1	10
Europe	France	Montbazin	40	43.590	3.260	26	27	27	26	37
Europe	France	Montpellier	41	43.612	3.877	-	1	1	-	-
Europe	France	Salsigne	42	43.330	2.360	-	2	10	-	-
Europe	France	St Georges d'Orques	43	43.610	3.780	1	3	3	1	-
Europe	Portugal	Santarem	44	39.240	-8.680	1	1	1	1	-
Europe	Spain	unknown locality				1	-	-	-	-
Europe	Spain	Barcelone	45	41.380	2.170	1	1	1	1	-
Europe	Spain	Cubillos	46	42.620	-6.560	1	1	1	1	-
Europe	Spain	Grenade	47	37.180	-3.600	1	2	2	1	-
Europe	Spain	Ibiza	48	38.990	1.420	1	1	1	1	-
Europe	Spain	Madrid	49	40.433	-3.683	29	29	29	28	28
Europe	Spain	Puerto Real (near Cadiz)	50	36.529	-6.192	-	-	1	-	-
Europe	Spain	Taragone	51	41.140	1.230	1	1	-	-	-
Total						471	499	518	425	536

Table 1. Collection localities of *Mus spretus* specimens included in this study, with the number of specimens sequenced per DNA region and per locality.

Country	Locality	N	S	h	Hd	Pi	k	Fs	D	R ²
France	Montbazin	26	16	12	0.874 ± 0.048	0.00093 ± 0.00015	2.397	-6.738	-1.498	0.069
Spain	Madrid	28	27	19	0.955 ± 0.024	0.00146 ± 0.00013	3.749	-10.814	-1.668	0.054
Morocco	BenGuerir	27	80	18	0.963 ± 0.020	0.00661 ± 0.00051	17.011	-1.980	-0.712	0.084
Morocco	BeniHadifa	28	64	17	0.947 ± 0.025	0.00586 ± 0.00073	15.085	-0.094	-0.315	0.096
Morocco	BenSlimane	34	70	20	0.966 ± 0.014	0.00646 ± 0.00050	16.617	-0.814	-0.108	0.110
Morocco	Berrechid	48	117	47	0.999 ± 0.004	0.00713 ± 0.00035	18.34	-24.186	-1.092	0.112
Morocco	Cap Djinet	34	87	27	0.975 ± 0.017	0.00648 ± 0.00048	16.67	-6.180	-0.808	0.071
Morocco	DayetErroumi	29	64	15	0.889 ± 0.046	0.00468 ± 0.00071	12.047	-0.108	-0.988	0.085
Morocco	Ifrane	23	78	16	0.960 ± 0.024	0.00635 ± 0.00067	16.356	-0.512	-0.898	0.099
Morocco	MerjaZerga	28	92	20	0.974 ± 0.016	0.00648 ± 0.00062	16.675	-1.845	-1.136	0.076
Morocco	Oued Nifikh	14	50	12	0.978 ± 0.035	0.00556 ± 0.00033	14.319	-1.310	-0.391	0.112
Morocco	SidiBoughaba	22	67	11	0.905 ± 0.043	0.00742 ± 0.00064	19.104	4.141	0.157	0.135
Morocco	Taza	13	61	8	0.897 ± 0.067	0.00553 ± 0.00109	14.244	2.246	-1.240	0.106

Table 2. Genetic diversity estimates and neutrality tests for the Algerian mouse based on the mitochondrial DNA data. Estimates for the 13 populations for which sample size is >12. Number of specimens (N), number of polymorphic sites (S), number of haplotypes (h), haplotype diversity (Hd), nucleotide diversity (Pi), average number of nucleotide differences (k), Fu's Fs test (Fs), Tajima's D test (D), Ramos- Onsins and Rozas R2 test. Significant values are highlighted in bold (P-value < 0.05 for D and R2 tests, P < 0.02 for Fs test).

		5	7	8	9	10	11	12	13	18	20	21	25	31	32	1	3	49	39	
Morocco	5	Aounate																		
Morocco	7	BenGuerir	0.0400																	
Morocco	8	BeniHadifa	0.0473	0.0178																
Morocco	9	BenSlimane	0.0361	0.0128	0.0172															
Morocco	10	Berrechid	0.0239	0.0077	0.0091	0.0074														
Morocco	11	Cherouda	0.0472	0.0220	0.0289	0.0296	0.0189													
Morocco	12	Clementine	0.0257	0.0180	0.0170	0.0092	0.0068	0.0259												
Morocco	13	DayetErroumi	0.0443	0.0224	0.0268	0.0278	0.0191	0.0283	0.0224											
Morocco	18	Ifrane	0.0576	0.0326	0.0357	0.0283	0.0299	0.0443	0.0282	0.0363										
Morocco	20	MerjaZerga	0.0404	0.0127	0.0197	0.0146	0.0099	0.0224	0.0183	0.0241	0.0356									
Morocco	21	Oued Nfifikh	0.0254	0.0148	0.0198	0.0139	0.0046	0.0268	0.0131	0.0198	0.0321	0.0143								
Morocco	25	SidiBoughaba	0.0407	0.0222	0.0211	0.0184	0.0100	0.0321	0.0178	0.0265	0.0421	0.0186	0.0181							
Morocco	31	Taza	0.0527	0.0236	0.0220	0.0259	0.0134	0.0288	0.025	0.0309	0.0400	0.0265	0.0168	0.0303						
Morocco	32	Tetouan	0.0534	0.0112	0.0174	0.0203	0.0137	0.0081	0.0220	0.0295	0.0355	0.0199	0.0237	0.0270	0.0241					
Algeria	1	Cap Djinet	0.0481	0.0232	0.0125	0.0191	0.0150	0.0300	0.0191	0.0277	0.0417	0.0212	0.0154	0.0236	0.0299	0.0211				
Algeria	3	Meflah	0.0556	0.0202	0.0082	0.0256	0.0137	0.0283	0.0186	0.0311	0.0405	0.0221	0.0207	0.0222	0.0276	0.0333	0.0035			
Spain	49	Madrid	0.0848	0.0679	0.0697	0.0777	0.0574	0.0612	0.0648	0.0784	0.0829	0.0776	0.0702	0.0756	0.0791	0.0575	0.0728	0.0765		
France	39	La Gardiole	0.1136	0.0898	0.0927	0.1052	0.0806	0.0854	0.0987	0.0927	0.1066	0.0915	0.0996	0.0925	0.1051	0.1124	0.0944	0.0922	0.0915	
France	40	Montbazin	0.1520	0.1092	0.1074	0.1254	0.0982	0.0997	0.1160	0.1125	0.1360	0.1045	0.1167	0.1149	0.1413	0.1269	0.1102	0.1260	0.1051	0.0297

Table 3. Pairwise Fst values between *M. spretus* populations samples (microsatellites data). Significant values after Bonferroni correction (P = 0.05) are in bold. Locality number as in Table 1 and Fig. 1.

ABC Model	Posterior probabilities		Details of the model
	MSAT	MTDNA	
Pooled populations			
MOD1	0.3140	0.0033	Independent colonization; size change in Africa
MOD2	0.1806	0.2002	Independent colonization; no size change in Africa
MOD3	0.3280	0.0018	Serial colonization; size change in Africa
MOD4	0.1773	0.7945	Serial colonization; no size change in Africa
Resampling	RES-MSAT	RES-MTDNA	
MOD1	0.0953	0.0704	Independent colonization; size change in Africa
MOD2	0.3024	0.0577	Independent colonization; no size change in Africa
MOD3	0.1914	0.4593	Serial colonization; size change in Africa
MOD4	0.4108	0.4126	Serial colonization; no size change in Africa

Table 4. Posterior probabilities of the four ABC models under microsatellite data and mtDNA data both for the pooled African populations (MSAT and MTDNA, respectively) and the resampling strategy (RESMSAT and RES-MTDNA, respectively).

		MSAT				MTDNA				RES-MSAT				RES-MTDNA			
		MOD1	MOD2	MOD3	MOD4	MOD1	MOD2	MOD3	MOD4	MOD1	MOD2	MOD3	MOD4	MOD1	MOD2	MOD3	MOD4
N africa	MODE	556478	465749	579668	453894	146386	175320	313120	179786	574224	364479	524179	200676	62495	32938	120572	40979
	MEDIAN	685409	465087	967071	469286	10220-50	180715	12523-64	185090	10524-99	402380	682730	203122	136301	38219	299609	46804
	95% low	368355	402992	263468	374430	13133	149436	89801	150237	135052	215645	196861	175839	13208	16800	20079	20540
	95% up	19953-44	539792	28879-30	599568	29997-56	225038	29967-57	236344	29159-96	819485	24120-35	243698	23184-57	78610	26025-59	95062
N spain	MODE	154960	176366	62604	67087	41270	30270	28593	23937	467333	482619	487410	472316	36224	37768	26074	26903
	MEDIAN	195901	200817	112758	109478	68648	51764	53826	37316	456926	405951	456780	437507	60773	59613	39808	45058
	95% low	55427	77588	27927	35006	18322	11224	10805	12248	385741	27715	311798	296089	15445	19200	13652	12413
	95% up	447019	429380	453437	412103	396855	378932	387843	235330	493540	499618	498126	495765	376103	303660	209891	284057
N France	MODE	107883	84854	103594	107555	8770	5924	5070	5919	98547	171270	116253	114568	7853	7832	8358	7236
	MEDIAN	170061	135496	123572	166330	16380	8600	9532	8642	89741	166523	110670	106603	10771	10238	12110	10329
	95% low	44455	35281	54799	51472	3017	2833	2280	2841	27179	54874	14947	36237	3662	4228	3812	3457
	95% up	467834	447266	349777	461447	374921	57199	231415	61224	169309	261987	243435	178171	115152	89905	115486	80359
N ancestral	MODE	548107	NA	338457	NA	7026	NA	22137	NA	18699-06	NA	11856-66	NA	55445	NA	28987	NA
	MEDIAN	737773	NA	420869	NA	21156	NA	58767	NA	16481-32	NA	13978-72	NA	164542	NA	64492	NA
	95% low	141144	NA	103498	NA	1950	NA	3255	NA	193744	NA	531154	NA	4710	NA	6120	NA
	95% up	24766-42	NA	15098-19	NA	527435	NA	15420-57	NA	29678-42	NA	25009-04	NA	28334-72	NA	15870-32	NA
T change Africa	MODE	10771	NA	6677	NA	141	NA	483	NA	25437	NA	44656	NA	5029	NA	4271	NA
	MEDIAN	24019	NA	18446	NA	352	NA	1594	NA	25498	NA	29321	NA	11300	NA	10129	NA
	95% low	1638	NA	1228	NA	51	NA	55	NA	1854	NA	103	NA	663	NA	621	NA
	95% up	48463	NA	47980	NA	41543	NA	44782	NA	48084	NA	49983	NA	44901	NA	44017	NA
T bott Spain	MODE	2920	4344	5416	4583	851	2832	4013	14387	10918	10920	12705	13103	2030	2165	13594	9005
	MEDIAN	6532	5086	7383	6462	1474	5632	4731	11382	9271	10816	11801	12542	4055	3827	10331	9046
	95% low	440	1209	2567	1866	273	628	1771	2766	228	7724	7202	8965	637	664	3114	3087
	95% up	14409	11535	14293	14375	6971	14437	12751	14959	14946	12940	14582	14742	13280	12913	14820	14499
T bott France	MODE	6242	8740	4687	2568	710	1896	840	2551	8747	12745	6925	10154	941	989	1222	1620
	MEDIAN	6444	7997	5172	4315	1424	5074	1375	6197	8864	11718	7069	10175	3351	4120	2437	2334
	95% low	787	657	1885	941	500	690	561	863	3743	6414	4506	6769	504	516	620	770
	95% up	14004	14258	10611	12448	14941	14675	7760	14692	14444	14676	10352	13237	14347	13805	11988	8451
N change Africa	MODE	28491-08	NA	28235-92	NA	91130	NA	53313	NA	27095-17	NA	22564-12	NA	17095-70	NA	530806	NA
	MEDIAN	18137-02	NA	15581-58	NA	493157	NA	165096	NA	20551-33	NA	21619-14	NA	16289-96	NA	11802-61	NA
	95% low	75493	NA	50026	NA	599	NA	110	NA	273603	NA	12256-41	NA	193730	NA	78944	NA
	95% up	29896-93	NA	29884-71	NA	29813-88	NA	29947-47	NA	29733-70	NA	27815-80	NA	29220-06	NA	28464-45	NA
N bott Spain	MODE	7922	6376	3183	7441	1316	932	1197	1524	18289	19035	14814	11302	1791	1867	1219	1146
	MEDIAN	9229	9331	6026	9617	6240	3010	6917	4075	16297	17078	13545	11658	6632	6408	5457	4256
	95% low	4277	2976	842	1667	227	168	178	282	6970	6453	5652	7043	410	481	210	201
	95% up	17210	18556	18306	19164	19838	19032	19834	17857	19766	19909	19358	17059	19421	19093	19617	19241
N bott France	MODE	4118	4483	3365	9723	2026	768	1646	919	19626	19799	19704	19880	1172	1276	1113	1203
	MEDIAN	6162	8648	9624	10646	8983	2241	8173	2824	19336	19617	19557	19811	6424	7908	4616	6185
	95% low	1880	1538	1282	4762	448	137	315	159	13688	14663	17282	18967	162	228	183	190
	95% up	17257	19664	19914	18843	19732	17678	19840	18496	19905	19969	19919	19960	19635	19546	19088	19677

Table 5. Parameters estimation for the four models under microsatellite data and mtDNA data both for the pooled African populations (MSAT and MTDNA, respectively) and the resampling strategy (RES-MSAT and RES-MTDNA, respectively). For each parameter we reported the mode, median and 95% credible interval. Values in bold corresponds to parameters estimated for the most supported models (see Table 4).