New molecular data favour an anthropogenic introduction of the wood mouse (Apodemus sylvaticus) in North Africa

Aude Lalis, Raphaël Leblois, Sohaib Liefried, Ali Ouarour, Champak Beeravolu Reddy, Johan Michaux, Adel Hamani, Christiane Denys, Violaine Nicolas

JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

New molecular data favor an anthropogenic introduction of the wood mouse (Apodemus sylvaticus) in North Africa

Journal:	*Journal of Zoological Systematics and Evolutionary Research*
Manuscript ID:	JZS.201500002.R2
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Lalis, Aude; Muséum National d'Histoire Naturelle, UMR ISYEB 7205 Leblois, Raphael; INRA, UMR 1062 Liefrid, Sohaib; Université Abdelmalek Essaâdi, Faculté des Sciences OUAROUR, Ali; Université Abdelmalek Essaâdi, Faculté des Sciences REDDY BEERAVOLU, Champak; INRA, UMR 1062 MICHAUX, Johan; Université de Liège, Unité de génétique de la conservation Hamani, Adel; ENSA, Département de Zoologie Denys, Christiane; Muséum National d'Histoire Naturelle, UMR ISYEB 7205 Nicolas, Violaine; Muséum National d'Histoire Naturelle, Paris, UMR 7205
Keywords:	Anthropogenic introduction, Demographic history, Mediterranean basin, Population expansion, Rodent
Abstract:	According to fossil data the wood mouse arrived in North Africa 7,500 ya, while it was present in Europe since early Pleistocene. Previous molecular studies suggested that its introduction in North Africa probably occurred via the Strait of Gibraltar more than 0.4 Mya ago. In this study, we widely sampled wood mice in order to get a better understanding of the geographic and demographic history of this species in North Africa, and possibly to help resolving the discrepancy between genetic and paleontological data. Specifically we wanted to answer the following questions: (1) when and how did the wood mouse arrive in North Africa? and (2) What is its demographic and geographic history in North Africa since its colonization? We collected in the field 438 new individuals and used both mtDNA and six microsatellite markers to answer these questions. Our results confirm that North African wood mice have a southwestern European origin and colonized the Maghreb through the Gibraltar strait probably during the Mesolithic or slightly after. They first colonized the Tingitane peninsula and then expanded throughout North Africa. Our genetic data suggest that the ancestral population size comprised numerous individuals reinforcing the idea that wood mice did not colonize Morocco accidentally through rafting of a few individuals, but via recurrent/multiple anthropogenic translocations. No spatial structuring of the genetic variability was recorded in North Africa, from Morocco to Tunisia.

New molecular data favor an anthropogenic introduction of the wood mouse (*Apodemus sylvaticus*) in North Africa

Aude LALIS[1], Raphael LEBLOIS[2], Sohaib LIEFRIED[3], Ali OUAROUR[3], Champak REDDY BEERAVOLU[2], Johan MICHAUX[4], Adel HAMANI[5], Christiane DENYS[1], Violaine NICOLAS[1]

[1] Institut de Systématique, Evolution, Biodiversité ISYEB UMR 7205 - CNRS MNHN UMPC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, 57 rue Cuvier, CP 51, 75005 Paris, France

[2] INRA, UMR1062 CBGP, F-34988 Montferrier-sur-Lez, France

[3] Laboratoire de Biologie et Santé, Université Abdelmalek Essaâdi, Faculté des Sciences, BP 2121, Tétouan, Maroc

[4] Unité de génétique de la conservation. Institut de Botanique, Université de Liège (Sart Tilman), 4000 Liège, Belgique.

[5] Laboratoire d'Ornithologie, Département de Zoologie, ENSA El Harrach, Alger, Algeria

Corresponding author:
Violaine NICOLAS, Institut de Systématique, Evolution, Biodiversité ISYEB UMR 7205 - CNRS MNHN UMPC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, 57 rue Cuvier, CP 51, 75005 Paris, France. Tel : (33) 1 40 79 48 52, vnicolas@mnhn.fr

Keywords
Anthropogenic introduction; Demographic history; Mediterranean basin; Population expansion; Rodent

Abstract

According to fossil data the wood mouse arrived in North Africa 7,500 ya, while it was present in Europe since early Pleistocene. Previous molecular studies suggested that its introduction in North Africa probably occurred via the Strait of Gibraltar more than 0.4 Mya ago. In this study, we widely sampled wood mice in order to get a better understanding of the geographic and demographic history of this species in North Africa, and possibly to help resolving the discrepancy between genetic and paleontological data. Specifically we wanted to answer the following questions: (1) when and how did the wood mouse arrive in North Africa? and (2) What is its demographic and geographic history in North Africa since its colonization? We collected in the field 438 new individuals and used both mtDNA and six microsatellite markers to answer these questions. Our results confirm that North African wood mice have a southwestern European origin and colonized the Maghreb through the Gibraltar strait probably during the Mesolithic or slightly after. They first colonized the Tingitane peninsula and then expanded throughout North Africa. Our genetic data suggest that the ancestral population size comprised numerous individuals reinforcing the idea that wood mice did not colonize Morocco accidentally through rafting of a few individuals, but via recurrent/multiple anthropogenic translocations. No spatial structuring of the genetic variability was recorded in North Africa, from Morocco to Tunisia.

Introduction

North-Africa belongs to the Mediterranean region which is considered today as a biodiversity hotspot (Myers et al. 2000; 2002). Its remarkable diversity is derived from three sources: the northern Palearctic, sub-Saharan Africa and the arid Palearctic (Dobson and Wright 2000). For several species the timing and path of colonization of North Africa are under debate. Genetic data provide powerful tools to infer the geographical origin and colonization time of a given population. The wood mouse, *Apodemus sylvaticus* (Linnaeus 1758), is one example of these widely distributed species in both the Maghreb (Northwest Africa) and Western Europe. According to fossil data it only arrived recently, i.e. during the Holocene, in Morocco and Algeria (Stoetzel 2013), while it was present in Europe since early Pleistocene (Cuenca-Bescós et al. 2010; Michaux and Pasquier 1974). Several hypotheses have been formulated to explain the arrival of the wood mouse in the Maghreb region (Libois et al. 2001): the species would have colonized North Africa either through populations established in the Near East, where it is

actually absent, *via* the Libyco-Egyptian route, or from Western European regions such as Sicily, *via* the Siculo-Tunisian strait, or Spain, *via* the strait of Gibraltar (Fig. 1). First, as a result of glacial and inter-glacial cycles during the Pleistocene, the current desert regions of northern Egypt, Libya and the Near East were not always arid (Groves and Di Castri 1991; Langgut et al. 2011; Migowski et al. 2006) allowing the migration of several species. According to fossil data, *A. sylvaticus* was present in Libya during the upper Paleolithic from 35,000 to 10,000 ya (Jaeger 1975), in Israel from Mindel/Riss (i.e. 450- 300 ka) to recent times (Cheylan 1991; Tchernov 1979), and in Turkey from 350 to 60 ka (Demirel et al. 2011). All these arguments favor the existence of a potential Libyco-Egyptian route of migration. However, it is well known that *A. sylvaticus* and *A. flavicollis* (Melchior 1834) are phenotypically highly similar species, unidentifiable in South and Eastern Europe at the individual level by morphological methods used by paleontological approach (Jojić et al. 2014). Moreover, the species identification of specimens from Turkey and Israel should be confirmed since several taxa from the Middle East previously assigned to *A. sylvaticus* were recently identified as a different species (Filippucci et al. 2002). Second, the two migration routes through southern Europe are often proposed to explain the close affinity between European and North African mammal fauna at different periods of the Pleistocene (Arambourg 1962; Jaeger 1975; Stoetzel 2013). Moreover, several genetic studies recently showed that both the strait of Gibraltar and the strait of Sicily allowed the crossing in both senses of numerous animals during the Pleistocene period, either *via* natural colonization or *via* incidental human introduction (Cosson et al. 2005; Guillaumet et al. 2006; Habel et al. 2009; Recuero et al. 2007; Stöck et al. 2008).

Based on mitochondrial DNA (mtDNA) restriction patterns and cytochrome b (cytb) sequences Libois *et al*. (2001) and Michaux *et al*. (2005; 2003) suggested that the introduction of the wood mouse in North Africa probably occurred *via* the Strait of Gibraltar 0.4 Ma. This dating does not fit the paleontological data; the oldest fossil of wood mouse in North West Africa being dated of 7,500 ya (Stoetzel 2009, 2013). Discrepancies between molecular and paleontological data may be partly due to the limited number of North African specimens (8 to 28) and localities (2 to 7) previously considered in molecular studies. These authors stressed the necessity to get a better sampling in North Africa and South Western Europe to better identify the centre of origin of African populations and to confirm the lack of genetic variability throughout the African range of the species. Moreover previous divergence time analyses were based on substitution rates

inferred from a phylogeny. It is now well known that the mutation rate is much higher than the substitution rate, and that the relationship between mutation and substitution rates can be described by an exponential curve (Ho et al. 2005). This means that molecular rates should be interpreted in the context of calibration point age and that short-term mutation rates can only be extrapolated to older times after accounting for the relationship between short-term and long-term rates of change. Ho *et al.* stressed that taking rate variation into account is particularly important for analyses of sequences on timescales of less than about 1–2 Myr before the present, such as studies on populations, which often incorrectly apply phylogenetic substitution rates to population-level analyses. Because Michaux *et al*. (2005; 2003) did not take this into account, their results need to be reevaluated.

In this study, we widely sampled wood mice in North Africa (Fig. 1) in order to get a better understanding of the geographic and demographic history of this species in this region, and possibly to help resolve the discrepancy between genetic and paleontological data. We also considered new specimens from Europe in order to get a better idea of the geographic origin of North African populations. Finally, we used a new approach to estimate the mutation rate for our intraspecific phylogeny. Specifically we wanted to answer the following questions: (1) when and how did the wood mouse arrive in North Africa? and (2) What is its demographic and geographic history in North Africa since its colonization? Due to its numerous advantages such as high rate of evolution, lack of recombination and haploidy, mtDNA has been widely used as a classical phylogeographic marker (Brito and Edwards 2009). However, because of its maternal inheritance, the risks of introgression and the absence of independent information coming from unlinked locus, mtDNA also presents some inconveniences and could yield to biased historical inferences. Therefore, we used both mtDNA cytb sequences and six microsatellite markers to study the genetic structure and demographic history of the wood mouse in North Africa.

Material and methods

Ethics Statement

Animals were live-trapped and handled under the guidelines of the American Society of Mammalogists (Sikes et al. 2011). The protocol was approved by Comité 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 isofluorane, followed by cervical dislocation. Capture permits were obtained through the "Haut Commissariat aux Eaux et forêts et à la Lutte contre la désertification" (autorisation n°15 HCEFLCD/DLCDPN/DPRN/CFF) in Morocco, and through the Ministry of Forestry in Algeria.

Sampling and DNA extraction

438 newly collected individuals were included in this study (23 specimens from 5 French localities, 2 specimens from one locality in Portugal, 40 specimens from 4 Spanish localities, 19 specimens from 2 Algerian localities, 334 specimens from 12 Moroccan localities, 1 specimen from Denmark, 19 specimens from Sweden; Supplementary Table S1). Genomic DNA was extracted from tissues preserved in 95% ethanol using NucleoSpin Tissue Core kit (MACHEREY-NAGEL).

Mitochondrial DNA amplification and sequencing

The cytb gene was amplified for 375 individuals using polymerase chain reaction (PCR) primers L14723, H15915 (Ducroz et al. 2001), ApoIntL2 (CTGGATCWAAYAACCCAACA) or ApoIntH1 (GTGGGGTRTTWAGTGGGT; this study). The internal primers designed in this study were used to amplify the DNA of the Iberian specimens, for which DNA was degraded. PCR conditions and sequencing followed Nicolas *et al*. (2014). The presence of cytb pseudogenes is well documented in *Apodemus* (Dubey et al. 2009). Pseudogenes are usually characterized by the presence of indels, stop codons, frame-shift mutations and amplification of heterozygotes (Frezal and Leblois 2008; Trian and DeWoody 2002). We have not observed any of these indications in our dataset. Moreover the base composition per codon was not significantly different between individuals. So we believe that pseudogenes were not present in our dataset. Sequences were submitted to the Genbank database (KM581675 to KM582049).

Microsatellite genotyping

Six loci were genotyped for 295 individuals. These six loci, As-20, As-34 (Harr et al. 2000) and GTTC4A, GTTD8S, GTTD9A, TNF(CA) (Makova et al. 1998) were selected based on length, annealing temperature, and quality of allele amplification. PCR conditions followed Harr *et al*. (2000) and Makova *et al*. (1998). PCR products were run and genotypes were scored according

to Lalis *et al*. (2012). Those loci were genotyped on the seven Moroccan populations with sample sizes varying from 23 individuals (ElKhizana) to 92 individuals (Taza) (Supplementary Table S1)

Mitochondrial DNA analyses

Phylogenetic analysis

In our phylogenetic analyses we included all the newly sequenced specimens and all specimens available in the Genbank database for which the cytb gene was sequenced, except those considered as pseudogenes by Dubey *et al*. (2009). This represents 545 individuals (Supplementary Table S1: 375 newly sequenced specimens and 170 specimens from Genbank). 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). This method accounts for the coexistence of ancestral and descendent haplotypes, multifurcations, and reticulate relationships (Posada and Crandall 2001) and it is therefore suitable for studying intraspecific relationships. We used the MP post-processing option, which removes all superfluous median vectors and links that are not contained in the shortest trees of the network. Sequences of 701 bp were retained for the network analysis in order to minimize the number of incomplete sequence as adding ambiguous data in median joining trees is problematic.

Genetic diversity and population differentiation

Nucleotide diversity and haplotype diversity (Nei 1987) were calculated using DnaSP 5.10 (Librado and Rozas 2009). This paper focuses on the colonization and subsequent demographic and geographic history of *A. sylvaticus* in North Africa. Thus, estimates of demographic history and spatial structure are only provided for this geographical region. 864 bp were available for all sequences from Maghreb, thus our demographic analyses are based on this sequence length. We analyzed population structure with an analysis of molecular variance (AMOVA). A population was defined as all individuals coming from one geographical locality. F_{ST} values were also calculated between all pairs of populations. Moreover, the plausibility of an isolation-by-distance scenario was explicitly tested by performing Mantel's tests (Mantel 1967) following the procedure described in Nicolas *et al*. (2014). All these analyses were performed using ARLEQUIN 3.11 (Excoffier et al. 2005)

Demographic history

The demographic history of populations was inferred using Fu's F_s test of population growth (Fu 1997). This statistics was estimated using ARLEQUIN 3.11, and its significance was assessed using 1000 coalescent simulations. As suggested in the ARLEQUIN manual, the F_s statistics was considered significant when the *p*-value was below 0.02.

We also used a test based on mismatch distributions in each population to determine if a population expansion occurred in the past, and to characterize it (Rogers and Harpending 1992). Excoffier *et al*. (2005) proposed to use these mismatch distributions to select between two models: a 'pure demographic expansion' and a 'spatial expansion'. Both assume that a stationary haploid population of size N_0 suddenly grew T generations ago to reach a population size of N_1 haploid individuals. However, while the 'pure demographic expansion' model assumes that the growing population is panmictic, the 'spatial expansion' model involves a spatial range expansion and spatially structured populations. To test the fit of these two models to our data, as well as to estimate the scaled expansion time $\tau = T*2\mu/G$ (μ is the mutation rate per sequence per generation; G is the generation time) and migration rate parameter $M=Nm$ in the second model, we used the least square fitting algorithm implemented in ARLEQUIN 3.11. Model choice and CI for parameter estimates are based on a parametric bootstrap approach. The generation time can vary within this species according to ecological conditions (Fons and Saint Girons 1993), but it is 0.5 year in Morocco and Algeria (Harich and Benazzou 1990).

Demographic history was also explored using the MIGRAINE software and the newly developed model of a single population with past variations in population size (Leblois et al. 2014). The model of past change in population size implemented in MIGRAINE is similar to that used in the mismatch analysis except that past variation in population size is exponential and not discrete/sudden. MIGRAINE was used to estimate ancestral theta ($\theta_{anc}=2N_{anc}\mu$, where N_{anc} is the ancestral haploid population size and μ the mutation rate of the whole sequence), current theta ($\theta=2N\mu$, where N is the current population size) and D, the time of occurrence of the demographic change scaled by population size (i.e. $D=T/2NG$, where T is expressed in years and G is the generation time). Because MIGRAINE is based on the infinite sites model (ISM) for analysis of sequence data, two datasets were produced for the mtDNA cytb region to fit this model. For one data set, we chose to systematically remove incompatible sites for all individuals, for the second, we chose to remove haplotypes with incompatible sites. For all analyses, we pooled all individuals from Morocco because of the clear lack of genetic structure observed in

our sample. All runs with MIGRAINE were done using 1,000,000 trees, 2400 points and two iterations.

To get an inference of the time of occurrence of the past expansion, we need to compute the unscaled parameter *T* from the scaled time parameters inferred by the different methods using a given mutation rate and a generation time. However an accurate estimation of the mutation rate is usually difficult to obtain. Both intraspecific and pedigree-based estimates of substitution rates are generally higher than interspecific phylogenetically calibrated rates (Ho et al. 2005). This difference is due to purifying selection. To accurately estimate mutation rate for intraspecific phylogenies it is thus recommended to focus on synonymous mutations because under the assumption of neutral evolution, the substitution rate for synonymous mutations is equal to the mutation rate (Kimura 1968). Nabholz *et al*. (2008) recently re-evaluated the evolutionary substitution rate at the third codon position of the cytb using a multi-point calibration procedure of lineage-specific mutation rates across 1696 mammalian species. They found that Rodentia is the fastest evolving order, with an average of 1.76 10^{-07} substitution per site per year, and that the mutation rate can vary greatly among rodents taxa. Thus we decided to infer a specific mutation rate using cytb data, from the third codon position only, for the genera *Apodemus* and two calibration points derived from paleontological data. The divergence time between *A. mystacinus* (Danford and Alston 1877) and *A. flavicollis*/*A. sylvaticus* was estimated to be approximately 7 My old, and the divergence between *A. sylvaticus* and *A. flavicollis* to be approximately 4 My old (Michaux et al. 2004; Michaux et al. 2003). All cytb sequences of 864 bp of *A. mystacinus* and *A. flavicollis* available in the Genbank database were included in our analyses (i.e. 1 and 34 sequences respectively), and considering the third codon positions only results in a data set containing 288 bp per individual sequence. The mean number of substitutions between *A. mystacinus* and *A. flavicollis*/*A. sylvaticus* was 116, and 75 between *A. sylvaticus* and *A. flavicollis*. Assessing the variation of synonymous substitution rates between lineages is technically problematic because of saturation. To minimize this effect of multiple mutations at one site, we based our computations on the number of sites that are similar between the two species rather than on the segregating ones. According to Felsenstein (2004), the probability of observing a site with a similar state is $\frac{1+e^{-2\mu t}}{2}$, where μ is the mutation rate and t is the divergence time, both expressed in the same unit (i.e. generations or years) between the two

species. Given this formula, we obtained a mutation rate of 1.2 10^{-07} substitution per site per year for the calibration *A. mystacinus* and *A. flavicollis*/*A. sylvaticus*, and a mutation rate of 0.9 10^{-07} substitution per site per year for the calibration *A. sylvaticus/A. flavicollis*.

Microsatellites analyses

Genetic diversity

Genetic variability of the microsatellite markers was measured for each locus by the number of alleles (N_a), gene diversity (H_e, expected heterozygosity), and observed heterozygosity (H_o) using R package adegenet v1.2–7 (Jombart 2008). Allelic richness, tests for Hardy-Weinberg equilibrium (*HWE*) and linkage disequilibrium were conducted according Lalis *et al*. (2012). We then used the software FREENA (Chapuis and Estoup 2007) to estimate null allele frequencies (*a*) for each population and locus following Dempster *et al.* (1977).

Population structure

We applied STRUCTURE v2.3.3 (Pritchard et al. 2000) to the data with K varying from 1 to 6, with 5 runs for each K value. The number of contributing populations was statistically tested using the ad-hoc Evanno statistic DK (Evanno et al. 2005). This procedure is sensitive to pronounced changes in mean log likelihood values between successive K values and the degree of variance of any given mean.

We also analyzed the spatial genetic structure with the software GENELAND v2.5.0 (Guillot et al. 2005) which uses geographic information to identify spatial discontinuities in the genetic structure of the sample. We first performed a preliminary analysis with 10 runs of 1 000 000 iterations with a thinning of 500 and a burn-in of 50%, considering values for K from 1 to 6 with a starting value of 2, to infer the number of populations K maximizing the posterior probability of the data. Then longer runs (ten replicates, each) of 20 000 000 iterations with a thinning of 500 and burn-in of 50% were analyzed to precisely set the spatial limits for K=2 (first split). For all analyses, the uncertainty attached to spatial coordinates was set to 0.2 km and the maximum number of nuclei in the Poisson- Voronoi tessellation fixed at 1800 (roughly three times the number of analyzed individuals).

Population differentiation was further analyzed by computing estimates of F_{ST} (Weir and Cockerham 1984) between all population pairs using GENEPOP v4.1.3 and significance was

tested by permutation using FSTAT (Goudet 1995; Goudet et al. 1996). Using GENEPOP, we also looked for isolation by distance patterns by regressing $F_{ST}/(1-F_{ST})$ between populations over the logarithm of geographical distances as recommended by Rousset (1997), and significance of the correlations between genetic and geographic distances was tested using Mantel tests with 30,000 permutations.

Demographic history

The MIGRAINE software was also used on the microsatellite data set to infer past changes in population sizes. All MIGRAINE runs for microsatellites used 20,000 to 200,000 trees, 2,400 points and 3 iterations. To convert our estimates of scaled parameters into unscaled demographic parameters we considered a fixed value of 5 10^{-4} mutation per locus per generation for all microsatellite loci (Dib et al. 1996; Ellegren 2000; Sun et al. 2012).

Results

Genetic diversity and structure of the wood mouse across its geographic range

Our MJ network analysis shows that the 545 *A. sylvaticus* cytb sequences fell into two major lineages (Fig. 2, Supplementary Figure S1); the first one comprising the Italian, Balkan and Sicilian animals (lineage 1), and the second corresponding to all specimens from North Africa and western, northern and central Europe (lineage 2). The first lineage is divided into two sublineages: a Sicilian one (lineage 1b) and an Italo-Balkan one (lineage 1a). The second lineage is also divided into two sublineages: a North African group (lineage 2a), and a Western, Northern and Central Europe group – lineage 2b). In North Africa the network has a starlike pattern with one very common ancestral haplotype widely distributed in Morocco, Algeria and Tunisia. The pattern obtained in lineage 2b is much more complex, with a high number of haplotypes represented by few individuals and high genetic distances between haplotypes. The number of Spanish haplotypes is especially high (Fig. 2). Several starlike patterns are observed within lineage 2b: the central and most common haplotypes are often found in Spain, except one case where the central haplotype is found in Northern Europe (Sweden, Netherlands, Denmark, Czech Republic and Belgium). The European haplotype closest to Maghrebian ones is found in Sweden. According to our MJ network analysis, we have at least 7 mutations (all in third codon position) between Maghrebian and European haplotypes. This corresponds to a time of divergence of

85,000-165,000 ya, depending on mutation rate (Calibration 1 (mutation rate of 1.2 10^{-07} substitution per site per year obtained from comparison between *A. mystacinus* and *A. flavicollis*/*A. sylvaticus* haplotypes) gave a divergence time of 125,000 ya. Calibration 2 (mutation rate of 0.9 10^{-07} substitution per site per year obtained from comparison between*A. sylvaticus* and*A. flavicollis* haplotypes) gave a divergence time of 165,000 ya. Calibration 3 (mutation rate of 1.76 10^{-07} substitution per site per year according to Nabholtz *et al*. (2008)) gave a divergence time of 85,000 ya).

Haplotype diversity is similar between lineages 1a, 2a and 2b and tends to be a little higher than in lineage 2a (Table 1). Nucleotide diversity is 2.1 to 2.9 times lower in the Maghrebian lineage than in the three other lineages. Within the Maghrebian lineage, haplotype diversity is lower in Merja Zerga than in other populations. Within lineage 2b haplotype diversity is 1.2 to 1.5 times higher in the two Spanish populations than in France or Sweden, while nucleotide diversity is 1.4 to 3.7 times higher.

Genetic diversity and structure of the wood mouse in North Africa

All microsatellite loci show a high genetic diversity: the total number of alleles per locus ranged from 1 to 19 with a mean number of 9 alleles per locus (Supplementary Table S3). Among the 6 loci, four (As-34, GTTC4A, TNF(CA) and As-20) show significant heterozygote deficiencies and have deviations (<P=0.05) from mutation-drift equilibrium for an excess of heterozygosity (Supplementary Table S3). Using FREENA, we show that the most probable hypothesis to explain heterozygote deficiencies in these loci is the existence of null alleles. Mean estimated null allele frequencies are moderate (mean frequency $\bar{a}$ (As-34) = 0.067, $\bar{a}$ (GTTC4A) = 0.079, $\bar{a}$ (TNF(CA) = 0.037 and $\bar{a}$ (As-20) = 0.012). Overall the loci were judged statistically independent. The number of alleles per population ranges from 8 to 13 with a mean number of alleles per population of 9 (Table 4). The expected heterozygosity is relatively high. According to locus and population, it varies between 0.600 and 0.775 (Supplementary Table S3).

We applied two complementary clustering algorithms to infer rodent population structure and to probabilistically assign individuals to populations or clusters based on individual multilocus genotypes. STRUCTURE 2.3.3 provided consistent results over 5 replicated runs and the probability of the data (LnP(K)) increased from K=1 to K=6 although with a clear tendency to reach a plateau at K=4 and higher values (Figure 3). According to the Evanno test, K=2 and K=3

are the most likely scenario: all populations are grouped except population MerjaZerga. STRUCTURE results for K = 2 are fully congruent with the GENELAND bipartition (Fig3). The plot is based on the highest-probability run for K = 2 (the same split and similar posterior probabilities were obtained for all 20 replicates).

For the mtDNA data 9% of the genetic variation is partitioned among populations and 91% within populations (F_{ST} = 0.091, P < 0.001). F_{ST} values between most pairs of populations are low (range from 0.007 to 0.199) but significantly different from 0 (P < 0.05), except between the population of El Khizana and the populations of Chrouda, Parc Talassemtane, Taza and Ifrane which are not significant (Supplementary Table S2). For microsatellite data, multilocus estimates of F_{ST} for pairs of populations range from -0.004 to 0.095 (Supplementary Table S4).

No significant correlation between geographic and genetic distances is recorded in all Maghreb samples based on both mt DNA (P value of Mantel test = 0.682) and microsatellite data (P = 0.211).

Demographic history of the wood mouse in North Africa

A clear signal of population expansion is observed in the North African clade based on F_s (Table 1), mismatch analyses (Table 2, Supplementary Figure S2) and MIGRAINE analyses (Table 3, Supplementary Figure S3). These results are also corroborated by the starlike pattern observed in the MJ network. Based on mismatch analyses, a signal of demographic and/or spatial expansion is recorded in the populations of Cap Djinet, Beni Hadifa, Chrouda, El Khizana, Ifrane, Parc Talassemtane and Taza. Estimates of the migration parameter M (M=2Nm) are very large (99,999) for these populations, meaning that a very low level of population structure is inferred under the spatial model for these samples. For every values of the migration parameter, both models are equivalent. On the contrary, for the population of Merja Zerga, the test based on F_s is not significant, and the mismatch distribution fits the spatial expansion model with an M value of 3 (CI: 0-34) but not the demographic expansion model.

Based on mismatch analyses, the timing of the expansion was calculated for three distinct mutation rates (Table 2) and a generation time of 0.5 year. Values vary of a factor 2 according to the mutation rate used, and confidence intervals are large. However all analyses show that the expansions probably occurred in early Holocene or late Pleistocene (mean values vary between

7,319-15,674 ya to 14,287-30,596 ya according to the mutation rate; CI vary from 819-12,503 to 14,181-45,259 ya).

MIGRAINE analyses of both modified data sets fitting the ISM give very similar results suggesting that modifications done to fit the ISM were not too drastic. In both cases, a highly significant signal of past expansion is found, with (1) very high and precise estimates of current scaled population sizes around θ=58; (2) very low but imprecise estimates of ancestral scaled population sizes around θ_{anc}<0.002; and (3) estimations of the time from present to the start of the expansion around D=0.053 with an intermediate precision (Table 3, Supplementary Figure S3). Considering a mutation rate of 10^{-7} per site and per year, and correcting this mutation rate for the modifications made on the data sets to fit the ISM (i.e. using a correction ratio of CR=66/45 and 67/75 kept sites), we can convert scaled population sizes into diploid effective population sizes with the following computations $N= \theta/(CR$ * sequence length * mutation rate site per site per year * generation time). For a generation time of 0.5 year, these calculations give point estimates of current effective population size around 4.5 millions individuals [CI range: 2,900,000 – 8,300,000], ancestral population sizes of few hundred individuals [1 – 165,000] and a time of occurrence around 125,000 years [22,000 – 340,000]. However, taking into account uncertainty contains in the 95% confidence intervals of the scaled parameters makes converted IC very wide, showing very limited precision for higher estimates. For example, 95% CI for the time in years are [14,000–620,000].

For the microsatellite data, MIGRAINE was initially run on all population samples independently. However, probably because of low sample sizes, results by population do not show any significant signal of past changes in population size, except for the Taza population, which had by far the largest sample size, and for which a significant signal of population expansion is found (data not shown). For this reason, and, because almost no population structure is observed on the whole Moroccan samples, MIGRAINE was thus finally run on the pooled Moroccan sample (i.e. all Moroccan populations analyzed as a single population; Table 3). MIGRAINE results show a highly significant signal of past expansion with parameter estimations that are concordant with those obtained with the mtDNA data, with large current scaled population sizes (i.e > 8.6), intermediate ancestral population sizes and recent timing in terms of $T/2N$. However, precision of the estimations are almost opposite to that obtained on the mitochondrial data as the best precision is obtained for the scaled ancestral population size estimate with a very narrow CI of 5.4-20.4 for a

generation time of 0.5 year, whereas current population sizes and timing show much wider CIs. We obtained point estimates of current population size of 160,000 [11,600–1,400,000] individuals, ancestral population sizes of 7,200 [2,800-10,200] individuals and a time of 135 [1–344,000] years. The very high uncertainty level attached to the inference of the time in years is due to the incertitude of the scaled time inferences multiplied by the incertitude of the diploïde population size used for conversion (i.e. $D=T/2N$).

Discussion

Wood mouse phylogeography and origin of North African populations

Our results confirm the phylogeographical structure previously obtained by Libois *et al*. (2001) and Michaux *et al*. (2005; 2003) with four main lineages: a Sicilian lineage, an Italo-Balkan lineage, a North African lineage and a western, northern and central Europe lineage. Compared to these three previous studies, new specimens from Bosnia and Herzegovina, Macedonia and Montenegro fall, as expected, in the Italo-Balkan clade. New specimens from Ireland, Swizerland, Denmark, Sweden, France, Spain and Portugal fall, as expected, in the western, northern and central Europe clade. New specimens from Morocco and Algeria fall in the North African lineage.

Our MJ network analysis strongly suggests that North African wood mice have a western/northern/central European origin. This is supported by the absence of genetic affinities between all North African wood mice with either the Sicilian, Italian, or Balkan populations. Gemmeke *et al*. (1987) also found that the A allele of transferrin is shared by Tunisian and western European (Portugal, Spain, France, Germany) wood mice, whereas the animals of the Tyrrhenian-Drisatic region (Italy, Sardinia, Croatia) are characterized by the presence of the B and C alleles. Both Libois *et al*. (2001) and Michaux *et al*. (2003) suggested that wood mice introduction into North Africa occurred *via* the Strait of Gibraltar, the genetically nearest European haplotype coming either from the central part of the Iberian Peninsula (Libois et al. 2001) or from the central part of Portugal (Michaux et al. 2003). However, based on larger sample sizes, our analyses show that the nearest European haplotype to the Maghrebian ones comes from Sweden. The high haplotype and nucleotide diversities observed in the two Spanish populations (Table 1) suggest that the Iberian Peninsula was a refuge region for *A. sylvaticus* during the last glacial maximum, and that wood mice recolonized and expanded in the main part

of the Western Palearctic region from there at the end of the last ice age (Michaux et al. 2005; Michaux et al. 2003). Thus, the close affinity between haplotypes from Morocco and Sweden may be due to the large geographic genetic variability with limited sampling: we sampled only a restricted number of localities in Spain (7) and Portugal (1) and probably underestimated the genetic diversity within the Iberian Peninsula. Moreover, in a recent review Gomez and Lunt (2007) showed that Iberia was not a single refuge during the Pleistocene glacial maxima, but that at least seven glacial refugia existed for terrestrial taxa ('refugia within refugia hypothesis'). We sampled less than four of them. Extensive sampling of wood mice in Iberia would be necessary to test the 'refugia within refugia' hypothesis on this species and its impact on the phylogeographic history of European and North African wood mice. It is interesting to note that the sequencing of the entire human mtDNA reveals that the Saami of Scandinavia and the Berbers of North Africa share an extremely young branch, aged merely approximately 9,000 years (Achilli et al. 2005). According to these authors this finding not only confirms that the Franco-Cantabrian refuge area of southwestern Europe was the source of late-glacial expansions of hunter-gatherers that repopulated northern Europe after the last glacial maximum, but also indicates that European hunter-gatherers crossed the Strait of Gibraltar.

According to our mismatch analyses, oldests time of expansion were recorded in the populations of Chrouda, Taza and El Khizana (i.e. close to the Tingitane Peninsula and the Gibraltar Strait) and the youngest one in the Algerian population of Cap Djinet (i.e. Far East from the Gibraltar Strait). This result reinforces the idea that the Maghreb was colonized from Iberia through the Gibraltar Strait.

Crossing the Gibraltar Strait: the when and how of African colonization

Today, for a small terrestrial species, the Gibraltar Strait is an important barrier to dispersal, being 14km wide at its narrowest point and (currently) exceeding 200m in depth. It is therefore interesting to evaluate the timing and dynamics of colonization of African wood mouse from the Iberian Peninsula to the Maghreb. European wood mice could have colonized Africa either *via* a land bridge connecting the two continents, or after the opening of the Gibraltar Strait, either by rafting on vegetation, or in recent times *via* anthropogenic means.

Geological evidence indicates that Morocco and the Iberian Peninsula have been connected by a land bridge only twice (Blondel and Aronoson 1999; Duggen et al. 2003; Krijgsman et al. 1999):

during the Betic crisis (16–14 Ma) and during the Messinian salinity crisis (5.59–5.33 Ma). Our mtDNA analysis suggests that Africa was colonized less than 85,000-165,000 ya. These dates are too recent to be consistent with dispersal *via* either of these land bridges.

An alternative possibility is suggested by the bathymetry of the Strait of Gibraltar and climatic events during the Pleistocene period. The floor of the strait has a very complex topography including several ridges, so that depths vary greatly (Brandt et al. 1996). The shallowest sections are on an almost straight line from Cape Malabata in Morocco to Punta Paloma in Spain. On this eminence, known as the Camarinal Sill, the present maximum water depth is 290 m, but in many places it is much shallower, ranging between 40 and 150 m (Brandt et al. 1996). Given that sea levels in the area of the Strait of Gibraltar dropped by approximately 130 m during Pleistocene glaciations (Andersen and Borns Jr. 1997), some of the higher parts of this area of the Camarinal Sill are likely to have been exposed at that time as temporary small islands. Islands probably formed visible land masses covered by vegetation, completely changing the appearance of the strait from either shore. When sea levels were low, the maximum distance between two land masses from Morocco to Spain was only about five kilometers (Straus 2001). This may have enabled some terrestrial vertebrates to 'hop' across the Strait of Gibraltar quite recently, as suggested for snakes (Carranza et al. 2006). *Apodemus sylvaticus* is clearly unable to swim the distance between Africa and Europe, and even between either of the two continents and the Mediterranean islands. However, rafting on a natural support may potentially have occurred, even though biogeographical data concerning the western Mediterranean (Dobson 1998) and other parts of the world (Heaney 1986) suggest that such events are extremely rare.

It is also possible that human activities led to translocation of *A. sylvaticus* from Spain to Morocco, as previously shown for several other mammal species (Dobson 1998). Controversy about possible trans-Gibraltar human movements in the Lower, Middle and even Upper Pleistocene has reigned for over a century and continues to do so. According to the most recent review (Rolland 2013), Europe was peopled independently by converging population movements from both the Western Asian and Ibero-Moroccan staging posts during the Early Pleistocene, between ca. 1.85-1.40 Ma. During brief, though favorable warm to cold transition periods, purposeful dispersal took place by swimming and/or wading from coast to coast, possibly via ephemeral small islands, perpendicular to currents. No migration event would have occurred between Iberia and Morocco during the mid-Upper Pleistocene. According to Straus (2001) it is

only in the terminal Paleolithic (10,500-12,000 ya) that, "with clear evidence of marine fishing and probable navigation, a credible case can be made for trans-Gibraltar human contacts". However more recent studies, taking into account the new Aterian chronology, do not support this result (Derricourt 2005; Garcea 2004). During the Mesolithic (9,000 ya), 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). Human genetic data also indicates that crossing of the strait of Gibraltar occurred for humans about 9,000 ya (Achilli et al. 2005; Semino et al. 2004). Divergence values between European and Moroccan wood mice are too small to be consistent with dispersal during the Early Pleistocene. The time of divergence obtained between European and North African population (85,000-165,000 ya) is greater than the point estimates obtained for the time of expansion in North Africa according to mismatch mtDNA analyses (between 7,319 and 30,596 ya [CI: 819-45,259]) and MIGRAINE microsatellites analyses (135 ya [CI: 1-344,000])). MIGRAINE mtDNA point estimates tended to be higher (125,000 ya [CI: 22,000-340,000]), but with a low precision. It is very difficult to obtain robust age estimates for recent divergent events and for the start of a past expansion for several reasons: 1) the difficulty to robustly estimate mutation rate; 2) large confidence intervals are obtained with MIGRAINE for the dating of the expansion. Those large confidence intervals are partly due to the lack of information in the data but also to the model of continuous exponential increase. Contrarily to sudden expansions, two exponentially growing populations with an expansion starting at different moments but with a similar change in effective population sizes will lead to very similar patterns of polymorphism because of the shape of the exponential. For example, Leblois *et al*. (2014) showed that the precision of the inference of the timing of such past progressive expansion is very limited compared to other parameters and to other demographic scenarios. Confidence intervals obtained from mismatch curves are much narrower than for MIGRAINE, probably because of both the sudden expansion model and the statistical method used. Beside these methodological arguments explaining the lack of precision in the dating, the discrepancy between divergence and expansion time estimates can be explained by two biological hypotheses:

1) A recent expansion from a small area of original 'inoculation', in which the inoculation occurred 85,000-165,000 ya. MIGRAINE estimates of ancestral population size (several hundreds or thousands of individuals) before expansion favor this hypothesis. Wood mice arrival in North

Africa may have been progressive, with plenty individuals arriving during a long period of time, and the demographic expansion would have occurred only thousand years later. However this hypothesis is invalidated by : i) fossil data indicating that the wood mouse arrived during the Holocene in Morocco and Algeria (Stoetzel 2013); ii) trans-Gibraltar human movements are attested only after 9,000 ya (see above); iii) owing to the ability of wood mice to live in numerous habitat types (dense and humid forests, dry pine forests, high mountain cedar forests, meadows, sand dunes near the sea, shrubs; to sea level up to 2,000 m) and to its opportunistic feeding habits (Aulagnier et al. 2008; Kowalski and Rzebik-Kowalska 1991), once it have reached North Africa it would have find suitable conditions to undergone a demographic and spatial expansion throughout the Maghreb. Few local competitors probably existed at that time, other rodent species being present avoiding forests (*Meriones shawi* (Duvernoy 1842), *Lemniscomys barbarus* (Linnaeus 1766), *Mus spretus* Lataste 1883, *Dipodillus campestris* (Loche 1867), *Arvicanthis niloticus* (E. Geoffroy 1803), *Psammomys obesus* Cretzschmar 1828) (Stoetzel 2013). The commensal species *Mus musculus domesticus* Schwarz and Schwarz 1943 and *Rattus* spp. were present in North Africa more recently (Stoetzel 2013).

2) A colonization event later than 85,000-165,000 ya. The over-estimation of the time of divergence between European and North African wood mice would be explained by the recent invasion of some European haplotypes already divergent from the other haplotypes. Indeed, refuge regions are generally characterized by a high diversity of mitochondrial types (Avise 2000) that evolved separately, and high genetic divergence between Spanish haplotypes is observed in the MJ network (with up to 16 mutations). Moreover, as stated above, Gómez and Lunt (2007) showed that Iberia was not a single refuge during the Pleistocene glacial maxima, but that several Iberian refugia existed. Extensive sampling of wood mice in Iberia would be necessary to test if it has an impact on our estimates of divergence time between European and North African wood mice populations. Moreover it would be interesting to sequence longer sequences, since our estimate of the time of divergence between North African and European wood mouse is only based on a short mtDNA sequence length of 701 bp.

When divergence between lineages is recent, it is difficult to obtain robust age estimates, and therefore to test the alternative hypotheses of anthropogenic translocation and natural colonization, as already stressed by Husemann et al. (2013). However, for *A. sylvaticus*, we have the chance to have a good fossil record in Maghreb indicating that the wood mouse arrived

552 during the Holocene in North Africa (Stoetzel 2009, 2013). While it is recorded from 2,500-4,000
553 ya from the Capelleti cave in Algeria, it is recorded since 7,500-6,000 ya in the Tingitana
554 Peninsula (Kahf-That-El-Ghar, Bou Saria). This result fits our time of expansion, with an older
555 expansion near the Tingitana Peninsula than in Algeria. Taken together, genetic and fossil data
556 are consistent (given the difficulty to accurately date recent divergence events) and favor an
557 anthropogenic translocation from the Iberian Peninsula to Morocco. Moreover our genetic data
558 show that the ancestral population size (before expansion) comprised a high number of
559 individuals, reinforcing the idea that wood mice did not colonize Morocco accidentally through
560 rafting of a few individuals, but via recurrent temporal anthropogenic translocations. At first sight
561 it could be surprising that woodmice colonize North Africa via anthropogenic translocation
562 several thousand years ago, and that its dispersal between the two continents did not continues
563 until today since maritime trade has increased. One hypothesis could be that at the beginning of
564 the Holocene *A. sylvaticus* was more commensal than today. Molecular and zooarcheological
565 data showed that the commensal species *Mus musculus* only reached Western Europe during the
566 first millennium BC and onwards, related to the generalization of maritime trade (Bonhomme et
567 al. 2010). The arrival of the house mouse in Western Europe at this time could have led to a shift
568 in the degree of commensality of *Apodemus*.
569
570 *Geographical structure in the Maghreb: taxonomical implications*
571 Three subspecies were described in North Africa: *A. s. rufescens* inhabits the High Atlas, the arid
572 forest of the Rif and high plateau of Algeria; *A. s. ifranensis* is present in the medium Atlas and
573 the region of Oulmès, and *A. s. hayi* inhabits the Mediterranean regions of the Maghreb (Saint
574 Girons 1974; Saint Girons and Van Bree 1962). Our mitochondrial and microsatellite analyses
575 reveal low variability in the North African lineage from Morocco to Tunisia, and nearly no
576 spatial structuring: 1) no significant pattern of isolation by distance was detected with both
577 genetic markers, 2) AMOVA on mtDNA indicates that most of the genetic variation is
578 partitioned within populations, 3) STRUCTURE and GENELAND clustering analyses based on
579 microsatellite data suggest that North African wood mice form a single population, except
580 perhaps for the Merja Zerga population (strong drift as indicated also by low haplotype
581 diversity), and 4) estimates of F_{ST} among populations were not significantly different from 0 for
582 microsatellite data, and were low for mtDNA data. The number of microsatellite loci used in this

583 analysis is relatively low and may not confer sufficient power to discern fine-scale structure. The
584 use of additional loci could help to elucidate patterns of genetic structure not identified in this
585 study. However, this great similarity throughout North Africa was already highlighted, on a few
586 number of specimens, by the allozymic study of Filippucci (1992) (Nei's distance, D = 0.008),
587 the mtDNA restriction patterns of Libois (2001) and the cytb sequencing of Michaux (2003). Our
588 data confirm a lack of differentiation, even between animals that were caught either at long
589 distances from each other or in the loci typici of the North African subspecies, i.e., where some
590 genetic differences could a priori be expected. Thus, from a taxonomic point of view, our
591 molecular data reinforce the opinion of Kowalski and Rzebik-Kowalska (1991), who, based on
592 morphological characters, invalidated the taxa *A. s. ifranensis* and *A. s. rufescens* and considered
593 that the wood mouse is monotypic throughout the region.
594 Most molecular biogeographical studies performed in North Africa yielded high estimates of
595 genetic diversity, and the majority of taxa exhibited multiple endemic lineages dating back to the
596 Plio-Pleistocene or even longer (reviewed by Husemann et al. 2013; Nicolas et al. 2014). Our
597 results on wood mice are strikingly different, but can easily be explained by its recent
598 colonization of the Maghreb.
599 To conclude, wood mice colonized the Maghreb through the Gibraltar strait, probably during the
600 Mesolithic or slightly after, by recurrent/multiple anthropogenic translocations, and then
601 expanded rapidly throughout North Africa without any geographical structuring. Extensive
602 sampling in Iberia and more genetic markers would be necessary to test the 'refugia within
603 refugia' hypothesis and to obtain more accurate dating of the African time of colonization.
604

605 **Acknowledgments**

606 This study was supported by the ANR call "6ème extinction" project "ANR-09-PEXT-004 and the
607 project CMEP TASSILI MDU 09MDU755. RL and CRB have been partially funded by ANR
608 projects EMILE 09-blan-0145-01 and IM-Model@CORAL.FISH 2010-BLAN-1726-01, and by
609 the Project INRA Starting Group "IGGiPop". Molecular analyses were supported by the ATM
610 "Taxonomie moléculaire, DNA Barcode et Gestion Durable des Collections", the Service de
611 Systématique Moléculaire of the MNHN (UMS 2700, Paris, France) and the network
612 "Bibliothèque du Vivant" funded by CNRS, MNHN, INRA and CEA (Genoscope). Part of this
613 work was carried out by using the resources from the MNHN UMS 2700, INRA GENOTOUL

and MIGALE bioinformatics platforms and the computing grid of the CBGP lab. We are grateful to all collectors, particularly E. Stoetzel, A. Delapre, L. Tifarouine, A. Rihane. The fieldwork was possible through the collaboration of the "Institut Scientifique de Rabat" (Dr A. El Hassani, A. Qninba, M. Fekhaoui, M. Abderhamane, H. El Brini) and of the "Haut Commissariat aux Eaux et forêts et lutte contre la desertification". Prof. S. Fahd (University of Tetouan), Prof. H. Aouraghe (University of Oujda), Prof. T. Benazzou (University of Rabat), Prof. S. Doumandji (IRA, El Harrach Alger), Dr Y. Fernandez-Jalvo and Dr C. Sésé (MNCN, Madrid, Spain) must be thanked for all their support and assistance.

References

Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, Bandelt HJ, Semino O, Torroni A (2005) Saami and Berbers--an unexpected mitochondrial DNA link. *Am J Hum Genet* **76**:883-886.

Andersen BG, Borns Jr. HW (1997) *The Ice Age world: An Introduction to Quaternary History and Research with Emphasis on North America and Northern Europe During the Last 2.5 Million Years*. Scandinavian University Press, Oslo.

Arambourg C (1962) Les faunes mammalogiques du Pléistocène circumméditerranéen. *Quaternaria* **6**:67–109.

Aulagnier S, Haffner P, Mitchell-Jones AJ, Moutou F, Zima J (2008) *Guide des mammifères d'Europe, d'Afrique du Nord et du Moyen-Orient*. Delachaux et Niestlé, Paris.

Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard University press, Cambridge.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**:37-48.

Blondel J, Aronoson J (1999) *Biology and Wildlife of the Mediterranean Region*. Oxford University Press, Oxford.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffray JC, Britton-Davidian J (2010) Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. *Proc Biol Sci*.

643 Brandt P, Alpers W, Backhaus JO (1996) Study of the generation and propagation of internal
644 waves in the Strait of Gibraltar using a numerical model and synthetic aperture radar
645 images of the European ERS 1 satellite. *J Geophys Res* **101**:14237–14252.
646 Brito PH, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-
647 based markers. *Genetica* **135**:439-455.
648 Carranza S, Arnold EN, Pleguezuelos JM (2006) Phylogeny, biogeography, and evolution of two
649 Mediterranean snakes, *Malpolon monspessulanus* and *Hemorrhois hippocrepis*
650 (Squamata, Colubridae), using mtDNA sequences. *Mol Phylogenet Evol* **40**:532-546.
651 Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population
652 differentiation. *Mol Biol Evol* **24**:621-631.
653 Cheylan G (1991) Patterns of Pleistocene turnover, current distribution and speciation among
654 Mediterranean mammals. In: Groves RH, Di Castri F eds., *Biogeography of*
655 *Mediterranean Invasions*. Cambridge University Press, Cambridge, pp. 227-262.
656 Cosson JF, Hutterer R, Libois R, Sara M, Taberlet P, Vogel P (2005) Phylogeographical
657 footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western
658 Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula*
659 (Mammalia: Soricidae). *Mol Ecol* **14**:1151-1162.
660 Cuenca-Bescós G, Rofes J, López-García JM, Blain H-A, De Marfá RJ, Galindo-Pellicena MA,
661 Bennásar-Serra ML, Melero-Rubio M, Arsuaga JL, de Castro JMB, Carbonell E (2010)
662 Biochronology of Spanish Quaternary small vertebrate faunas. *Quatern Int* **212**:109-119.
663 Demirel A, Andrews P, Yalçınkaya I, Ersoy A (2011) The taphonomy and the
664 palaeoenvironmental implications of the small mammals from Karain Cave, Turkey. *J*
665 *Archaeol Sci* **38**:3048-3059.
666 Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the
667 EM algorithm. *J Roy Stat Soc B* **39**:1-38.
668 Derricourt R (2005) Getting "Out of Africa": Sea crossings, land crossings and culture in the
669 hominin migrations. *J World Prehist* **19**:119-132.
670 Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J,
671 Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive
672 genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**:152-154.

673 Dobson M (1998) Mammal distributions in the western Mediterranean: the role of human
674 intervention. *Mammal Rev* **28**:77–88.
675 Dobson M, Wright A (2000) Faunal relationships and zoogeographical affinities of mammals in
676 north-west Africa. *J Biogeogr* **27**:417-424.
677 Dubey S, Michaux J, Brunner H, Hutterer R, Vogel P (2009) False phylogenies on wood mice
678 due to cryptic cytochrome-b pseudogene. *Mol Phylogenet Evol* **50**:633-641.
679 Ducroz JF, Volobouev V, Granjon L (2001) An assessment of the systematics of arvicanthine
680 rodents using mitochondrial DNA sequences: evolutionary and biogeographical
681 implications. *J Mammal Evol* **8**:173-206.
682 Duggen S, Hoernle K, van den Bogaard P, Rüpke L, Morgan JP (2003) Deep roots of the
683 Messinian salinity crisis. *Nature* **422**:602-606.
684 Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences.
685 *Nat Genet* **24**:400-402.
686 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the
687 software STRUCTURE: a simulation study. *Mol Ecol* **14**:2611-2620.
688 Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for
689 population genetics data analysis. *Evol Bioinform Online* **1**:47-50.
690 Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Inc, Sunderland, Massachusetts.
691 Filippucci MG (1992) Allozyme variation and divergence among European, Middle Eastern, and
692 North African species of the genus *Apodemus* (Rodentia, Muridae). *Isr J Zool* **38**:193–
693 218.
694 Filippucci MG, Macholan M, Michaux JR (2002) Genetic variation and evolution in the genus
695 *Apodemus* (Muridae: Rodentia). *Biol J Linn Soc* **75**:395-419.
696 Fons R, Saint Girons MC (1993) Le cycle sexuel chez le mulot sylvestre, *Apodemus sylvaticus*
697 (L., 1758), (Muridae) en région méditerranéenne. *Z Säugetierk* **58**:38-47.
698 Frezal L, Leblois R (2008) Four years of DNA barcoding: current advances and prospects. *Infect*
699 *Genet Evol* **8**: 727–736.
700 Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking
701 and background selection. *Genetics* **147**:915-925.
702 Garcea E (2004) Crossing deserts and avoiding seas: Aterian North African-European relations. *J*
703 *Anthropol Res* **60**:27-53.

704 Gemmeke H, Radtke M, Niethammer J (1987) Zur innerartlichen Proteinvariation bei der
705 Waldmaus (*Apodemus sylvaticus*). *Z Säugetierkunde* **52**:242–247.
706 Gómez A, Lunt DH (2007) Refugia within Refugia: patterns of Phylogeographic Concordance in
707 the Iberian Peninsula. In: Weiss N, Ferrand N eds., *Phylogeography of Southern*
708 *European Refugia*. Springer, Dordrecht, pp. 155–188.
709 Goudet J (1995) FSTAT: a computer program to calculate F-statistics. *J Hered* **86**:485–486.
710 Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differenciation in diploid
711 populations. *Genetics* **144**:1933–1940.
712 Groves RH, Di Castri F (1991) *Biogeography of Mediterranean invasions*. Cambridge University
713 Press, Cambridge.
714 Guillaumet A, Pons JM, Godelle B, Crochet PA (2006) History of the Crested Lark in the
715 Mediterranean region as revealed by mtDNA sequences and morphology. *Mol Phylogenet*
716 *Evol* **39**:645-656.
717 Guillot G, Mortier F, Estoup A (2005) GENELAND: a program for landscape genetics. *Mol Ecol*
718 *Notes* **5**:712–7115.
719 Habel JC, Dieker P, Schmitt T (2009) Biogeographical connections between the Maghreb and the
720 Mediterranean peninsulas of southern Europe. *Biol J Linn Soc* **98**:693-703.
721 Harich N, Benazzou T (1990) Contribution à l'étude de la biologie du mulot (*Apodemus*
722 *sylvaticus*, Rongeurs, Muridés) de la plaine côtière du Maroc. *Mammalia* **54:** 47-59
723 Harr B, Musolf K, Gerlach G (2000) Characterization and isolation of DNA microsatellite
724 primers in wood mice (*Apodemus sylvaticus*, Rodentia). *Mol Ecol* **9**:1664-1665.
725 Heaney LR (1986) Biogeography of mammals in SE Asia: estimates of rates of colonization,
726 extinction and speciation. *Biol J Linn Soc* **28**:127–165.
727 Ho SYw, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate
728 estimates and systematic overestimation of recent divergence times. *Mol Biol Evol*
729 **22**:1561-1568.
730 Husemann M, Schmitt T, Zachos FE, Ulrich W, Habel JC (2013) Palaearctic biogeography
731 revisited: evidence for the existence of a North African refugium for Western Palaearctic
732 biota. *J Biogeogr* **41**:81-94.
733 Jaeger JJ (1975) *Les Rongeurs, du Miocène à l'actuel, en Afrique nord-occidentale.* Université
734 des Sciences et Techniques du Languedoc, Montpellier.

735 Jojić V., Bugarski-Stanojević V., Blagojević J. Vujošević M. (2014) Discrimination of the sibling
736 species *Apodemus flavicollis* and *A. sylvaticus* (Rodentia, Muridae). *Zoologischer*
737 *Anzeiger* **253**: 261-269.
738 Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers.
739 *Bioinformatics* **24**:1403–1405.
740 Kimura M (1968) Evolutionary rate at molecular level. *Nature* **217**:624-626.
741 Kowalski K, Rzebik-Kowalska B (1991) *Mammals of Algeria*. Polish Academy of Sciences,
742 Institute of Systematics of Evolution of Animals, Wrodow.
743 Krijgsman W, Hilgen FJ, RaY I, Sierro FJ, Wilson DS (1999) Chronology, causes and
744 progression of the Messinian salinity crisis. *Nature* **400**:652–655.
745 Lalis A, Leblois R, Lecompte E, Denys C, Ter Meulen J, Wirth T (2012) The impact of human
746 conflict on the genetics of *Mastomys natalensis* and Lassa virus in West Africa. *PLoS*
747 *ONE* **7**:e37068.
748 Langgut D, Almogi-Labin A, Bar-Matthews M, Weinstein-Evron M (2011) Vegetation and
749 climate changes in the South Eastern Mediterranean during the Last Glacial-Interglacial
750 cycle (86 ka): new marine pollen record. *Quat Sci Rev* **30**:3960–3972.
751 Leblois R, Pudlo P, Néron J, Bertaux F, Beeravolu CR, Vitalis R, Rousset F (2014) Maximum
752 likelihood inference of population size contractions from microsatellite data. *Mol Biol*
753 *Evol* **31**:2805-2823.
754 Libois RM, Michaux JR, Ramalhinho MG, Maurois C, Sara M (2001) On the origin and
755 systematics of the northern African wood mouse (*Apodemus sylvaticus*) populations: a
756 comparative study of mtDNA restriction patterns. *Can J Zool* **79**:1503-1511.
757 Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA
758 polymorphism data. *Bioinformatics* **25**:1451-1452.
759 Makova KD, Patton JC, Chesser RK, Krysanov EY, Baker RJ (1998) Microsatellite markers in
760 wood mouse and striped field mouse (genus *Apodemus*). *Mol Ecol* **7**:247-248.
761 Mantel N (1967) The detection of disease clustering and generalized regression approach. *Cancer*
762 *Res* **27**:209-220.
763 Michaux J, Pasquier L (1974) Dynamique des populations de mulots (Rodentia, *Apodemus*) en
764 Europe durant le Quaternaire. Premières données. *Bull Soc Geol France* **164**:431-439.

765 Michaux JR, Libois R, Filippucci MG (2005) So close and so different: comparative
766 phylogeography of two small mammal species, the yellow-necked fieldmouse (*Apodemus*
767 *flavicollis*) and the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region.
768 *Heredity* **94**:52-63.
769 Michaux JR, Libois R, Paradis E, Filippucci MG (2004) Phylogeographic history of the yellow-
770 necked fieldmouse (*Apodemus flavicollis*) in Europe and in the Near and Middle East.
771 *Mol Phylogenet Evol* **32**:788-798.
772 Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R (2003) Mitochondrial
773 phylogeography of the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic
774 region. *Mol Ecol* **12**:685-697.
775 Migowski C, Stein M, Prasad S, Negendank JFW, Agnon A (2006) Holocene climate variability
776 and cultural evolution in the Near East from the Dead Sea sedimentary record.
777 *Quaternary Res* **66**:421–431.
778 Mulazzani S, Le Bourdonnec F-X, Belhouchet L, Poupeau G, Zoughlami J, Dubernet S, Tufano
779 E, Lefrais Y, Khedhaier R (2010) Obsidian from the Epipalaeolithic and Neolithic eastern
780 Maghreb. A view from the Hergla context (Tunisia). *J Archaeol Sc* **37**:2529-2537.
781 Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity
782 hotspots for conservation priorities. *Nature* **403**:853–858.
783 Nabholz B, Glemin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across
784 mammals--the longevity hypothesis. *Mol Biol Evol* **25**:120-130.
785 Nicolas V, Ndiaye A, Benazzou T, Souttou K, Delapre A, Denys C (2014) Phylogeography of the
786 North African dipodil (Rodentia: Muridae) based on cytochrome-b sequences. *J Mammal*
787 **95**:241-253.
788 Olson DM, Dinerstein E (2002) The global 200: priority ecoregions for global conservation. *Ann*
789 *Missouri Bot Gard* **89**:199-224.
790 Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks.
791 *Trends Ecol Evol* **16**:37-45.
792 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus
793 genotype data. *Genetics* **155**:945–959.

794 Recuero E, Iraola A, Rubio X, Machordom A, Garcia-Paris M (2007) Mitochondrial
795 differentiation and biogeography of *Hyla meridionalis* (Anura : Hylidae): an unusual
796 phylogeographical pattern. *J Biogeogr* **34**:1207-1219.
797 Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise
798 genetic differences. *Mol Biol Evol* **9**:552-569.
799 Rolland N (2013) Europe was peopled independently by converging population movements from
800 both the Western Asian and Ibero-Moroccan staging posts during the Early Pleistocene.
801 *Quatern Int* **316**:59-72.
802 Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under
803 isolation by distance. *Genetics* **145**:1219-1228.
804 Saint Girons MC (1974) Rongeurs, lagomorphes et insectivores du Massif du Toubkal (Haut
805 Atlas marocain). *Bull Soc Sci Nat Ph Maroc* **54**:55–59.
806 Saint Girons MC, Van Bree PJH (1962) Recherches sur la répartition et la systématique de
807 *Apodemus sylvaticus* (Linnaeus 1758) en Afrique du Nord. *Mammalia* **26**:478–488.
808 Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis
809 C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill
810 PA, Santachiara-Benerecetti AS (2004) Origin, diffusion, and differentiation of Y-
811 chromosome haplogroups E and J: inferences on the neolithization of Europe and later
812 migratory events in the Mediterranean area. *Am J Hum Genet* **74**:1023-1034.
813 Sikes RS, Gannon WL, the Animal Care and Use Committee of the American Society of
814 Mammalogists (2011) Guidelines of the American Society of Mammalogists for the use
815 of wild mammals in research. *J Mammal* **92**:235-253.
816 Souville G (1998) Contacts et échanges entre la péninsule Ibérique et le Nord-Ouest de l'Afrique
817 durant les temps préhistoriques et protohistoriques. *CR Acad Inscr Belle* **142**:163-177.
818 Stöck M, Sicilia A, Belfiore NM, Buckley D, Lo Brutto S, Lo Valvo M, Arculeo M (2008) Post-
819 Messinian evolutionary relationships across the Sicilian channel: Mitochondrial and
820 nuclear markers link a new green toad from Sicily to African relatives. *BMC Evol Biol*
821 **8**:56-56.
822 Stoetzel E (2009) *Les microvertébrés du site d'occupation humaine d'El Harhoura 2*
823 *(Pléistocène supérieur-Holocène, Maroc) : systématique, évolution, taphonomie et*
824 *paléoécologie.* Muséum national d'Histoire naturelle, Paris. PhD Thesis.

825 Stoetzel E (2013) Late Cenozoic micromammal biochronology of northwestern Africa.
826 *Palaeogeogr Palaeoclimatol Palaeoecol* **392**:359–381.
827 Straus LG (2001) Africa and Iberia in the Pleistocene. *Quatern Int* **75**:91-102.
828 Sun JX, Helgason A, Masson G, Ebenesersdottir SS, Li H, Mallick S, Gnerre S, Patterson N,
829 Kong A, Reich D, Stefansson K (2012) A direct characterization of human mutation
830 based on microsatellites. *Nat Genet* **44**:1161-1165.
831 Tchernov E (1979) Polymorphism, size trends and Pleistocene paleoclimatic response of the
832 subgenus *Sylvaemus* (Mammalia: Rodentia) in Israel. *Israel J Zool* **28**:131-159.
833 Trian DA, DeWoody A (2007) The occurrence, detection, and avoidance of mitochondrial DNA
834 translocations in mammalian systematics and phylogeography. *J Mammal* **88**: 908–920
835 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure.
836 *Evolution* **38**:1358–1370.
837
838

Figure legends

Fig. 1: Map showing the actual geographical distribution of the wood mouse (grey shading), the 3 potential colonization routes of North Africa discussed in the text and the four main clades recovered in the Median Joining network analysis (A), and the sampling localities in North Africa (B). Localities codes: 1 = BenSlimane, 2 = SidiBoughaba, 3 = MerjaZerga, 4 = Esperada, 5 = Tétouan, 6 = Chrouda, 7 = BeniHadifa, 8 = Ketama, 9 = Parc Talassemtane, 10 = ElKhizana, 11 = Ifrane, 12 = Moyen Atlas, 13 = Taza, 14 = Zeralda, 15 = Cap Djinet, 16 = Ain Dram. Localities with only mtDNA data are in black, and localities with both mtDNA and microsatellite data are in grey.

Fig. 2: Minimum spanning network of *A. sylvaticus* cytb haplotypes, with geographic provenance of haplotypes. The area of the circle is proportional to the haplotype frequency, and the length of connecting lines to the distance between haplotypes, defined as the number of substitutions estimated by NETWORK v4.500 (Bandelt et al. 1999). Specimens from Spain are in black.

Fig. 3: *Apodemus sylvaticus* populations clustering based:

A. on STRUCTURE Bayesian inference (K= 1 to 6); Burn-in period = 150,000; MCMC repeat length = 350,000). Graph illustrating the log posterior probabilities of the microsatellite data (Ln P(K)) for each number of genetic groups (K) tested for 5 runs each. The likelihood (LnP(K)) and the number of contributing populations was tested using the ad-hoc Evanno statistic (DeltaK) for K=1 to 6. For K2 to K5, each color represents one assumed population cluster K. Multiple colored bars display an individual's estimated membership proportion in more than one population (q), i.e. the admixture level.

B. on GENELAND spatial assignments to clusters for K = 2. The highest membership values are in light yellow and the isolines (grey curves) illustrate the spatial changes in assignment values. The labels correspond to the sampling location indicated in table 7.

866 **Supporting Information**

867 Figure S1: Minimum spanning network of *A. sylvaticus* Cytb haplotypes.

868

869 Figure S2 : Observed and expected mismatch distributions under different models (cytb data).

870

871 Figure S3: bidimensional plots of likelihood ratio profiles for pairs of parameters inferred by MIGRAINE.

872

873 Table S1: List of specimens used in this study, with geographical origins, field numbers, voucher
874 numbers of museum collection, haplotype numbers and GenBank accession numbers.

875

876 Table S2 : FST values between pairs of populations recorded with the cytb data.

877 Table S3. Estimates of diversity for six nuclear microsatellites per population and per locus.

878

879 Table S4: FST statistics calculated between each pair of populations for microsatellite data.

880

Table 1: Diversity estimates and demographic history of the wood mouse based on the cytb gene. Estimates for the four main wood mice lineages recovered in our phylogenetic analyses, and for the four populations of the lineage 2b with greater sample size were based on 701 bp. Estimates for North African populations were based on 864 bp. Number of sequences (*N*), number of polymorphic sites (*S*), number of haplotypes (*h*), haplotype diversity (*Hd*), nucleotide diversity (*Pi*), average number of nucleotide differences (*k*). For North Africa, values of *Hd*, *Pi*, *k* and Fu's *Fs* are only given for populations with more than 10 individuals sampled.

	N	S	h	Hd	Pi	k	Fu's Fs
Estimates based on 701 bp (whole geographical range of the species)							
Lineage 1a	44	79	34	0.984 ± 0.009	0.01067 ± 0.00178	7.477	
Lineage 1b	15	38	14	0.990 ± 0.028	0.01233 ± 0.00133	8.581	
Lineage 2a	298	125	125	0.959 ± 0.007	0.00419 ± 0.00019	2.938	
Lineage 2b	188	127	112	0.978 ± 0.005	0.00887 ± 0.00047	6.215	
Montseny (Spain)	16	31	15	0.992 ± 0.025	0.00907 ± 0.00102	6.358	
Murcia (Spain)	18	21	13	0.948 ± 0.039	0.00524 ± 0.00090	3.667	
Saint-Benoit (France)	16	5	5	0.767 ± 0.080	0.00247 ± 0.00025	1.733	
South Sweden (Sweden)	28	6	3	0.667 ± 0.039	0.00362 ± 0.00034	2.540	
Estimates based on 864 bp (North Africa)							
	298	136	144	0.977 ± 0.005	0.00399 ± 0.00016	3.449	-26.248 (P<0.001)
Cap Djinet	16	21	12	0.942 ± 0.048	0.00456 ± 0.00088	3.942	-6.371 (P<0.001)
Zeralda	3	1	2				
BeniHadifa	20	14	10	0.895 ± 0.043	0.00267 ± 0.00041	2.305	-5.240 (P=0.003)
BenSlimane	7	6	4				
Chrouda	40	34	23	0.962 ± 0.014	0.00462 ± 0.00041	3.994	-14.922 (P<0.001)
ElKhizana	16	24	14	0.983 ± 0.028	0.00422 ± 0.00074	3.65	-10.488 (P<0.001)
Esperada	8	5	3				
Ifrane	51	45	36	0.964 ± 0.017	0.00327 ± 0.00029	2.824	-26.949 (P<0.001)
Ket ama	4	5	4				
MerjaZerga	28	11	6	0.757 ± 0.049	0.00353 ± 0.00035	3.053	2.538 (P=0.856)
Moyen-Atlas ISR	3	4	3				
ParcTalassemtane	25	23	16	0.943 ± 0.030	0.00319 ± 0.00040	2.757	-12.123 (P<0.001)
SidiBoughaba	1						
Taza	71	54	40	0.975 ± 0.007	0.00412 ± 0.00031	3.557	-26.529 (P<0.001)
Tétouan	1						
Ain Dram	4	6	4				

Table 2: Estimated values of Tau, with confidence interval (P = 0.05), obtained in mismatch analyses using ARLEQUIN, and corresponding expansion time in years for three mutation rates and a generation time of 0.5 year. Calibration 1: mutation rate of 1.2 10^{-07} substitution per site per year for the calibration *A. mystacinus* and *A. flavicollis*/*A. sylvaticus*. Calibration 2: mutation rate of 0.9 10^{-07} substitution per site per year for the calibration *A. sylvaticus/A. flavicollis.* Calibration 3: mutation rate of 1.76 10^{-07} substitution per site per year according to Nabholtz *et al*. (2008). The sequence length was 288 bp (only third codon positions were included).

				Expansion time (years)								
	Tau			Calibration 1			Calibration 2			Calibration 3		
	Est val	Low bound	Up bound	Est val	Low bound	Up bound	Est val	Low bound	Up bound	Est val	Low bound	Up bound
Cap djinet	1,484	0,166	2,535	11020	1233	18824	14287	1598	24406	7319	819	12503
Beni Hadifa	1,863	0,355	3,248	13834	2636	24118	17936	3418	31270	9189	1751	16020
Chrouda	2,76	1,955	3,562	20495	14517	26450	26572	18822	34293	13613	9642	17568
ElKhizana	3,178	1,473	4,701	23599	10938	34908	30596	14181	45259	15674	7265	23186
Ifrane	2,336	1,766	3,082	17346	13114	22886	22490	17002	29672	11521	8710	15201
Parc Talassemtane	2,584	1,686	3,637	19188	12520	27007	24877	16232	35015	12749	8316	17938
Taza	2,99	2,172	3,791	22203	16128	28151	28786	20911	36498	14747	10713	18698
Maghreb	2,676	2,318	2,939	19871	17213	21824	25763	22316	28295	13198	11433	14496

899 Table 3 – Inferences on demographic history by the software MIGRAINE on the pooled Moroccan
900 data set. Point estimates and 95% Confidence intervals (brackets) are reported. Inferred parameters are
901 (1) pGSM, the parameter of the geometric distribution of the Generalized stepwise mutation model
902 {Pritchard, 1999 #4576} (2) $\theta = 2N\mu$ and $\theta_{anc}=2N_{anc}\mu$ the scaled current and ancestral population sizes; (3)
903 $D= T_{in\ generation}/2N$ the scaled time of when the past change in population size started. All population sizes
904 are expressed as numbers of genes, i.e. haploïd population sizes. See text and MIGRAINE manual for
905 details about the settings of the analyses and the models and method used.
906

N	pGSM	θ	D	θanc	Pop size eq. θ/θanc
Microsatellites					
290	0.46	160	0.00042	7.1	22.3
	[0.53 - 0.6]	[12-1400]	[2.2E-5 - 0.123]	[2.7 - 10.2]	[1.8 - 212]
mtDNA					
Deleted problematic sites (67 haplotypes and 81 segregating sites left)					
275	NA	58	0.055	0.00040	150,000
	NA	[39 - 95]	[0.01 – 0.15]	[0 – 2.1]	[25 – 900,000]
Deleted problematic individuals (67 haplotypes and 81 segregating sites left)					
229	NA	58	0.053	0.0023	26,000
	NA	[37 – 107]	[0.01 – 0.15]	[0 – 2.0]	[27 – 1,000,000]

907

908 Table 4. Population polymorphism at six microsatellite loci over the seven populations sampled:
909 sample size (N), Allele richness (A_R), observed (H_O) and expected heterozygosity (H_E), within -
910 population coefficient of inbreeding (F_{IS}), and *HWE* probability that the genotype population
911 conformed to the Hardy–Weinberg equilibrium.

Population	***N***	***AR***	***Ho***	***HE***	***Fis***	***HWE***
MerjaZerga	35	8	0.749	0.746	0.063	1.000
Chrouda	25	8	0.778	0.753	0.069	0.993
BeniHadifa	27	8	0.564	0.600	0.078	0.975
ParcTalassemtane	31	8	0.666	0.745	0.129	1.000
ElKhizana	23	9	0.738	0.736	0.042	0.907
Ifrane	52	10	0.708	0.766	0.099	1.000
Taza	92	13	0.715	0.775	0.092	1.000

912

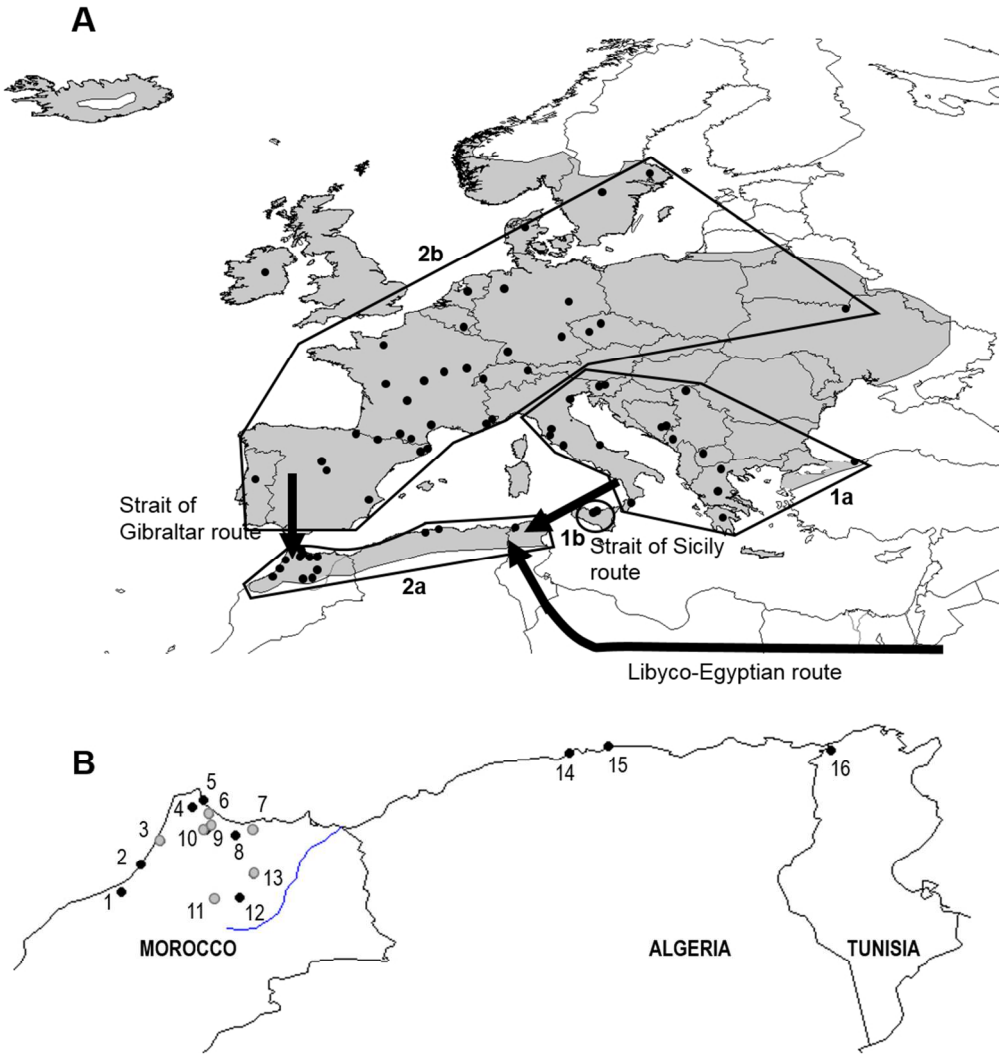

Fig. 1: Map showing the actual geographical distribution of the wood mouse (grey shading), the 3 potential colonization routes of North Africa discussed in the text and the four main clades recovered in the Median Joining network analysis (A), and the localities of collect in North Africa (B). Localities codes: 1 = BenSlimane, 2 = SidiBoughaba, 3 = MerjaZerga, 4 = Esperada, 5 = Tétouan, 6 = Chrouda, 7 = BeniHadifa, 8 = Ketama, 9 = Parc Talassemtane, 10 = ElKhizana, 11 = Ifrane, 12 = Moyen Atlas, 13 = Taza, 14 = Zeralda, 15 = Cap Djinet, 16 = Ain Dram. Localities with only mtDNA data are in black, and localities with both mtDNA and microsatellite data are in grey.
190x210mm (150 x 150 DPI)

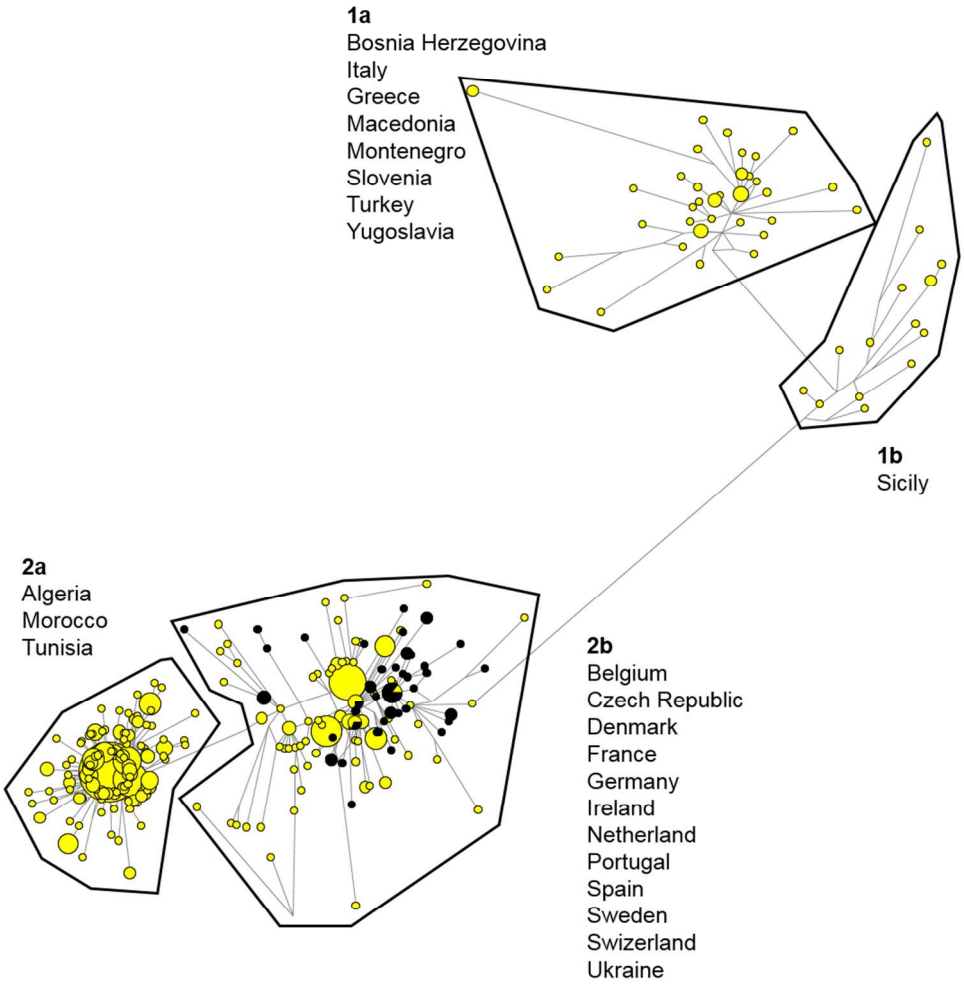

Fig. 2: Minimum spanning network of A. sylvaticus cytb haplotypes, with geographic provenance of haplotypes. The area of the circle is proportional to the haplotype frequency, and the length of connecting lines to the distance between haplotypes, defined as the number of substitutions estimated by NETWORK v4.500 (Bandelt et al. 1999). Specimens from Spain are in black.
188x195mm (150 x 150 DPI)

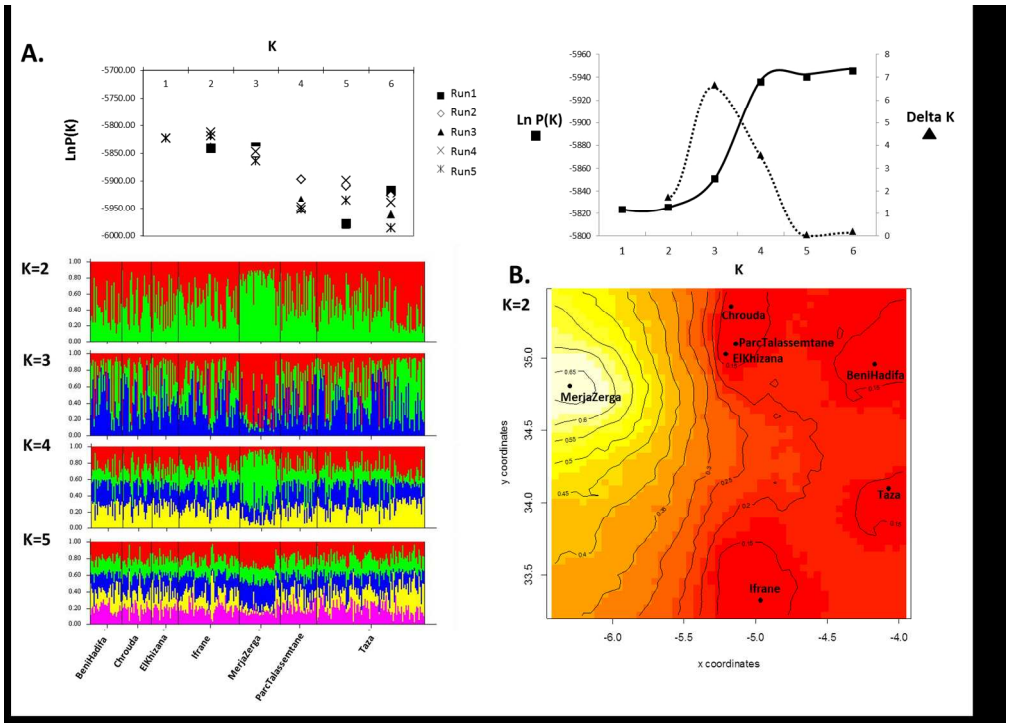

Fig. 3: Apodemus sylvaticus populations clustering based:
A. on STRUCTURE Bayesian inference (K= 1 to 6); Burn-in period = 150,000; MCMC repeat length = 350,000). Graph illustrating the log posterior probabilities of the microsatellite data (Ln P(K)) for each number of genetic groups (K) tested for 5 runs each. The likelihood (LnP(K)) and the number of contributing populations was tested using the ad-hoc Evanno statistic (DeltaK) for K=1 to 6. For K2 to K5, each color represents one assumed population cluster K. Multiple colored bars display an individual's estimated membership proportion in more than one population (q), i.e. the admixture level.
B. on GENELAND spatial assignments to clusters for K = 2. The highest membership values are in light yellow and the isolines (grey curves) illustrate the spatial changes in assignment values. The labels correspond to the sampling location indicated in table 7.

305x219mm (150 x 150 DPI)