# Phylogeography of the Striped Field Mouse, Apodemus agrarius (Rodentia: Muridae), throughout its distribution range in the Palearctic region

SCHOLARONE™
Manuscripts

Article type: Original Article

**Phylogeography of the Striped Field Mouse, *Apodemus agrarius* (Rodentia: Muridae), throughout its distribution range in the Palearctic region**

Alice Latinne[1,2], Miguel Navascués[3,4], Marina Pavlenko[5], Irina Kartavtseva[5], Rainer G. Ulrich[6], Marie-Laure Tiouchichine[1], Gilles Catteau[1], Hela Sakka[1], Jean-Pierre Quéré[3], Galina Chelomina[5], Aleksey Bogdanov[7], Michal Stanko[8], Lee Hang[9], Karsten Neumann[10], Heikki Henttonen[11] & Johan Michaux[1]


[1]Laboratoire de génétique de la conservation, Institut de Botanique, Boulevard du rectorat, 27, 4000 Liège, Belgium.
[2]EcoHealth Alliance, New York 10001, NY, USA.
[3]CBGP, Univ Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, Montpellier, France.
[4]Institut de Biologie Computationnelle, Montpellier, France.
[5]Federal Scientific Center of the East Asia Terrestrial Biodiversity, Far East Branch of Russian Academy of Sciences, Vladivostok, 690022, Russia.
[6]Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Institute of Novel and Emerging Infectious Diseases, Südufer 10, 17493 Greifswald - Insel Riems, Germany.
[7]Institute of developmental biology RAS, Vavilov str., 26, 119334 Moscow, Russia.
[8]Institute of Parasitology, Slovak Academy of Sciences, Hlinkova 3, 04001 Košice, Slovakia.
[9]Program for Veterinary Science, Seoul National University, South Korea.
[10]Institute of Pathology, City Hospital Dessau, Auenweg 38, D-06847 Dessau-Roßlau, Germany.
[11]Natural Resources Institute Finland, POB 9, FI-00790 Helsinki, Finland.



Corresponding author : Alice Latinne, Laboratoire de génétique de la conservation, Institut de Botanique, Boulevard du Rectorat, 27, 4000 Liège, Belgium. Email: alice.latinne@gmail.com

Running title: Phylogeography of *Apodemus agrarius*

**ABSTRACT**

In order to better understand the evolutionary history of oriental wildlife newcomers in Europe, we studied the phylogeography and demographic history of the striped field mouse, *Apodemus agrarius*, throughout its Paleartic distribution area. Genetic datasets including cytochrome *b* gene sequences and microsatellite markers *were analyzed using a large range of population genetics methodologies.*

   *A. agrarius* likely appeared in China during the Middle Pleistocene, after which it colonised the Russian Far East around 150 kya. The western colonisation also likely occurred simultaneously from a low number of founders from the Russian Far East. The species progressively increased in these regions and spread throughout Central Asia to colonise Siberia and Central Russia around 76-55 kya, followed by a last expansion in Europe and Turkey around 40-50 kya. Gene flow still happened between populations in the two main distribution ranges, mainly from the Eastern to Western populations.

   Our study, for the first time, provides an overview of the evolutionary and demographic history of the striped field mouse throughout the Palearctic region. *A. agrarius* appears to be an Asiatic immigrant and a relatively new member of the European fauna community. This study further confirms the important role of Far East Asian regions as the origin of European biodiversity.

**Keywords**

Glacial refugia, Palearctic region, *Apodemus agrarius*, mitochondrial DNA, phylogeography, continental colonisation

**INTRODUCTION**

Quaternary climatic oscillations have played a major role in shaping the present geographical distribution of both species and their genetic diversity. In the Northern hemisphere, this resulted in the extinction of northern populations during ice ages, followed by northward expansion from refugia during interglacial periods (Hewitt, 1996, Taberlet *et al.*, 1998, Hewitt, 1999, Hewitt, 2000). Refugial areas for European small mammals were mainly located in the Mediterranean, Ural and Caucasus/Carpathian regions (Bilton *et al.*, 1998, Michaux *et al.*, 2003, Deffontaine *et al.*, 2005, Michaux *et al.*, 2005, Deffontaine *et al.*, 2009). However, some other authors have also proposed that much more northern regions such as Western Scandinavia, Southern Great Britain or the Baltic area would have also provided additional refuges for some boreal as well as temperate mammal species (*e.g.* lemming, *Lemmus lemmus*; red squirrel, *Sciurus vulgaris*) (Fedorov & Stenseth, 2001, Stewart & Lister, 2001, Stewart & Dalén, 2008). However, this hypothesis has been strongly debated and now appears unlikely (Hughes *et al.*, 2016). Europe was also recolonised by some species from Central Asian refugia after the last glacial maximum, *i.e.* voles *Myodes glareolus* (Deffontaine *et al.*, 2005); *Microtus agrestis* (Jaarola & Searle, 2002); *Microtus oeconomus* (Brunhoff *et al.*, 2003); *Microtus arvalis* (Haynes *et al.*, 2003); the arctic fox, *Vulpes lagopus* (Dalén *et al.*, 2007) or the wood lemming, *Myopus schisticolor* (Fedorov *et al.*, 2008). These species are characterized by a weak phylogeographic structure and there are signs of a recent population expansion, suggesting their survival in a small number of refugial areas followed by continental expansion.

Finally, a few mammal species probably colonised Europe from much farther regions, like the Russian Far East or China, *e.g.* the common hamster, *Cricetus cricetus* (Neumann *et al.*, 2005) and the harvest mouse, *Micromys minutus* (Yasuda *et al.*, 2005). However, these studies

were generally based on a small sample size, particularly with regard to Far Eastern populations. Many questions regarding the relationships between populations from Eastern and Western Palearctic areas also remain unanswered: Where did they survive during the Quaternary glaciations? How and when did the oriental populations colonise the western regions? Were they characterized by particular ecological features allowing them to colonise large areas?

In order to gain further insight into the genetic structure of oriental wildlife newcomers in Europe, we studied the phylogeography and demographic history of the striped field mouse, *Apodemus agrarius* (Pallas, 1771), throughout its distribution area. This species is widely distributed over the entire Palearctic region, from Central Europe to the Korean Peninsula and Russian Far East. However, its distribution range is divided into two separate fragments (Europe–Western Siberia and Russian Far East–China), which are about 600–700 km apart, with the disjunction zone running along Transbaikalia and Mongolia (Fig. 1). Although some parts of this disjunction zone seem to be presently colonised by the species due to recent human activities and introductions (Bazhenov *et al.*, 2015), there is currently no permanent contact between *A. agrarius* populations from both fragments. In the future, *A. agrarius* could spread further into agricultural areas of the South-East Transbaikalia region because of its synanthropic and agrophilic habits (Khlyap *et al.*, 2011), but colonisation of the western part of the Transbaikalia region by this species appears highly unlikely as these areas are under a continental climate, with long cold and dry winters, with mountainous landscapes covered by taiga or dry steppe vegetation (Reshchikov, 1961, Perekrest, 2017).

From a taxonomic viewpoint, more than 25 subspecies of striped field mouse have been described, mainly based on body size and coat color (Musser & Carleton, 2005). However, many authors consider that most of these striped field mouse subspecies are non-valid, while accepting

4

only two to ten subspecies, because variations in body size and coat color are primarily determined by landscape and/or microclimatic environmental conditions and do not reflect taxonomic differences (Koh *et al.*, 2014). An in-depth taxonomic revision using a set of diagnostic traits, including genetic ones, therefore appears essential.

Finally, *A. agrarius* is a host of the Hantaan virus in the Far East and China, and of Kurkino and Saaremaa strains of the Dobrava hantavirus in the Western Palearctic, which causes hemorrhagic fever with renal syndrome (HFRS) in humans (Lee *et al.*, 1978, Nemirov *et al.*, 1999, Jonsson *et al.*, 2010, Klempa *et al.*, 2013, Xiao *et al.*, 2018). This species also serves as the reservoir of pathogenic *Leptospira*, *Rickettsia*, *Orientia* and *Bartonella* bacteria (Kraljik *et al.*, 2016, Gajda *et al.*, 2017, Latif *et al.*, 2017, Fischer *et al.*, 2018). Greater knowledge of its evolutionary history and genetic diversity is therefore also important in order to gain insight into its colonisation dynamics and therefore the risks of disease transmissions in new areas.

Most phylogeographic studies of *A. agrarius* have focused on a small part of its distribution range (Gortat *et al.*, 2013, Koh *et al.*, 2014, Andersen *et al.*, 2017, Jo *et al.*, 2017, Pereverzeva *et al.*, 2017, Sheremetyeva *et al.*, 2017), but phylogeographic information from datasets including samples from the two main population fragments and based on allozymes (Mezhzherin & Zykov, 1991, Atopkin *et al.*, 2007), karyotype variations (Kartavtseva & Pavlenko, 2000) or random amplified polymorphic DNA (RAPD) markers (Atopkin *et al.*, 2007) has shown very weak genetic differences among animals from the two fragments, but the Eastern group nevertheless seems to be more heterogeneous than the Western group. These findings were confirmed by the findings of another study based on mitochondrial cytochrome *b* gene (cyt*b*) sequences (Sakka *et al.*, 2010). The low karyotype and allozyme differentiation in the striped field mouse suggests a recent and rapid spread of the species from the Eastern to the Western

Palearctic (Kartavtseva & Pavlenko, 2000, Atopkin *et al.*, 2007). According to Kartavtseva & Pavlenko (2000), this spread occurred after the last glaciation, during the humid and warm ecological Holocene optimum (7 - 4.5 thousand years ago, kya), which was accompanied by high development and growth of mixed forests. However, this assumption is not supported by recent paleontological data, which suggested that *A. agrarius* was already present 50 kya in Central Europe (Popov, 2017) and in Southwestern France during the Late Pleistocene (19 kya), athough it is currently not present in this region (Aguilar *et al.*, 2008).

According to Atopkin *et al.* (2007), the disjunction of the *A. agrarius* range in Transbaikalia occurred after the warm ecological Holocene optimum and was associated with a heavy dry period in this region, which caused the decline of woody vegetation, and in some cases, the disappearance of trees and shrubs. However, this hypothesis would need to be confirmed by better sampling throughout the species' distribution range as well as by the use of more sensitive genetic methods based on rapidly evolving genes. The present study was carried out to better understand the phylogeographic structure and demographic history of the striped field mouse from the species' two main distribution areas using mitochondrial cyt*b* gene sequencing and microsatellite marker genotyping. Further investigation of the demographic history of the striped field mouse using recent statistical methods based on the coalescent theory will also be useful to gain greater insight into the demographic and expansion history of this species.

**MATERIAL AND METHODS**

**Samples and DNA extraction**

A total of 158 *Apodemus agrarius* individuals was sequenced for the mitochondrial cyt*b* gene. Twenty-four cyt*b* sequences from *Apodemus agrarius* available in GenBank were also added to

6

this dataset to cover the entire *A. agrarius* distribution range (68 localities in 20 countries) (Table 1 & Fig. 1; Supporting Information, Table S1). Moreover, we also genotyped 340 *A. agrarius* specimens originating from 88 localities in 17 countries using 9 microsatellite markers (Table 1 & Fig. 1; Supporting Information, Table S1). These specimens were obtained from collaborators, museums and field trips performed by our laboratories. All samples used in the present study were tissue samples stored in ethanol. Genomic DNA was extracted using the DNeasy™ Tissue kit (Qiagen Inc., Valencia, CA) according to the manufacturer's instructions.

**cyt*b* gene amplification**

The cyt*b* gene was amplified using the universal PCR primers L7 (5'-ACCAATGACATGAAAAATCATCGTT-3') and H16 (5'-ACATGAATYGGAGGY-CAACCWG-3') (Kocher *et al.*, 1989). Amplifications were carried out according to the protocol of Michaux *et al.* (2003) and performed in a Labover PTC100 thermal cycler through 39 cycles (30 s/94°C, 1 min/52°C, 2 min/68°C) with a final 10 min extension cycle at 68°C. All sequencing procedures were performed by Macrogen Inc. (Seoul, South Korea). The sequences were aligned using the ClustalW algorithm in BIOEDIT 7.0.5.2 (Hall, 1999).

**Microsatellite genotyping**

We genotyped 340 *A. agrarius* specimens for nine microsatellite markers selected from Makova *et al.* (1998) according to their amplification protocol. One primer of each of the nine primer pairs was labelled with one of the fluorochromes used in the ABIGeneScanTM system (ABI). Reaction mixtures contained approximately 100 ng of genomic DNA, 2.5 units of Taq DNA polymerase (Promega), 10 units of Promega buffer, 1.5 mM of $MgCl_2$, 0.6 mM of each primer (labelled and unlabelled), 250 mM of dNTPs (Perkin Elmer), and water to achieve a final volume of 25 µL. The thermal conditions included an initial 3 min denaturation step at 94°C, followed

7

by 35 cycles (1 min/94 °C, 30–45 s/annealing temperature, 30-60 s/72 °C), and a final 3 min extension at 72°C. After amplification, the nine microsatellite loci were combined in two multiplexes for each animal and analysed on an ABI 3100 automatic sequencer. The results were compiled and analysed with the GeneScanTM and GenotyperTM software packages (ABI).

**Mitochondrial data analysis**

The final cyt*b* dataset included 182 sequences from *A. agrarius*. Phylogenetic reconstructions were performed using the maximum-likelihood criterion (ML) algorithm implemented in the PHYML program (Guindon *et al.*, 2010). We used jMODELTEST (Posada, 2008) to determine the most suitable DNA substitution model for the cyt*b* dataset studied. The robustness of the tree was assessed by 1,000 bootstrap resamplings. Haplotype networks were also used as they may more effectively portray relationships among sequences for populations with low sequence diversity (Crandall & Templeton, 1993). A median-joining network was constructed using NETWORK 4.5 software (Bandelt *et al.*, 1999).

Haplotype (h) and nucleotide (π) diversities and their standard deviations (Tajima, 1989), Fu's *Fs* and genetic differentiation (using population pairwise $F_{ST}$) among populations were estimated using ARLEQUIN 3.1 (Excoffier *et al.*, 2005). Fu's *Fs* is a powerful test that is used to detect population expansion under assumptions of neutrality (Fu, 1997, Ramos-Onsins & Rozas, 2002). These indices were calculated for the Eastern and Western Palearctic groups. Moreover, to assess whether genetic diversity was higher within the potential refugial regions, these two main groups were divided into five regional subgroups: the first one corresponding to Russian Far Eastern populations; the second one to animals from Siberia (Western side of the Baikal lake, Novosibirsk region, and Altai region) and Kazakhstan; the third one to populations from Central Russia (from the Ural mountains to the Moscow region) and Ukraine; the fourth

8

one to all European and Turkish populations; and the fifth one to South Korean, Taiwanese and Chinese striped field mice. The genetic structure of populations was also examined using an analysis of molecular variance (AMOVA) performed in ARLEQUIN. AMOVA was conducted at three hierarchical levels of the range subdivision: among the two main geographic ranges (Eastern and Western Palearctic groups), among regional subgroups (Europe + Turkey, Western Siberia, Central Russia, Russian Far East and China + South Korea + Taiwan) and within each regional subgroup and its significance of these parameters estimated by 10,000 permutations.

Demographic histories of the two main striped field mouse groups (Eastern and Western Palearctic groups) were inferred using our cyt*b* gene dataset and an isolation-with-migration (IM) model implemented in the IM program (Hey & Nielsen, 2004). The model uses coalescent simulation within a Bayesian inference framework to estimate posterior probability distributions for five parameters, including: contemporary and ancestral effective population sizes ($\theta = N_e\mu$), divergence times ($T = t\mu$) and rates of gene flow between the Eastern and Western fragments. We assumed an HKY model of sequence evolution (Hasegawa *et al.*, 1985) and equal migration rates in both directions (i.e. just one migration parameter, *m*). However the effective number of migrants ($2N_em$) from each population can differ as $\theta$ estimates differed between Eastern and Western Palearctic groups. We used a burn-in of 200,000 steps followed by a run of 1 million steps. Prior boundaries were empirically determined to ensure that the posterior distributions fell completely within the prior distributions. The posterior mode was used as point estimate. We estimated the credibility intervals as the 90% highest probability density (HPD) intervals (i.e. the shortest span that includes 90% of the probability density of a parameter). To ensure reliable convergence toward the stationary distribution, we monitored multiple independent runs, each with 70 to 100 independent chains under Metropolis coupling, in order to improve mixing.

Mixing properties of the Markov-chain-Monte-Carlo method (MCMC) were assessed by examining the level of autocorrelation between the final and initial parameter values and by visual inspection of the parameter trend plots. The analyses were considered to have converged upon the stationary distribution if independent runs generated similar posterior distributions, with each having at least an effective sample size of 50 for each estimated parameter.

To convert the parameter estimates scaled by $\mu$ (i.e. $T$ and $\theta$) to demographic units, we used a per generation mutation rate of 2.7%/Myr as generally used for rodents of the *Apodemus* genus (*e.g.* Michaux *et al*. (2003), Michaux *et al*. (2005)). Assuming a generation time ($G$) for *A. agrarius* of 0.5 year (Pereverzeva *et al.*, 2017), the population divergence time ($T$) can be converted to calendar years ($t$ in years) and estimates of population mutation rates ($\theta_1$, $\theta_2$, and $\theta_A$) can be converted to estimates of effective population size parameters ($N_1$, $N_2$, $N_A$, respectively, in number of individuals). The migration parameters in the model can be used to obtain population migration rates (i.e. the effective number of migrants per generation), using an estimate of $\theta$ (i.e. $2Nm = \theta m/2$) (Fontaine *et al.*, 2010). $2N_1m$ and $2N_2m$ are the effective number of migrants per generation in populations 1 and 2, respectively. We also roughly estimated the timing of demographic expansion of each group and subgroup using the mismatch distribution $\tau$ mode calculated in DNASP (Librado & Rozas, 2009) and expressed as $\tau = 2\mu t$, where t is the expansion time in number of generations and $\mu$ is the mutation rate for the whole sequence.

**Microsatellite data analysis**

The proportion of null alleles (NA) at each locus and for each population was estimated with FREENA (Chapuis & Estoup, 2007). Genetic diversity was assessed by calculating expected (He) and observed (Ho) heterozygosities with ARLEQUIN over all loci for each group and confirmation of Hardy-Weinberg equilibrium (HWE) was tested using GENEPOP (Rousset,

10

2008). Multi-locus $F_{IS}$ was calculated for each group with FSTAT 2.9.3.2 (Goudet, 2001). The allelic richness (AR) was calculated using the rarefaction procedure implemented in FSTAT.

STRUCTURE 2.3.1 (Pritchard *et al.*, 2000) was used to infer the number of populations (K) and assign individuals to genetic clusters independently of spatial sampling. Ten iterations were run for each K value from 1 to 15 using an admixture model with a burn-in of $1\times10^5$ and MCMC values of $1\times10^6$. We used CLUMPAK (Kopelman *et al.*, 2015) to average the results of multiple iterations for a given K and to generate a visual output of the STRUCTURE results.

The demographic history of *A. agrarius* was also inferred from microsatellite data using an approximate Bayesian computation (ABC) approach via random forests (Marin *et al.*, 2016, Pudlo *et al.*, 2016). In this approach, data is simulated from the demographic model with parameter values taken from prior probability distributions and data is transformed into summary statistics. Random forests are used to learn about parameters from the simulated summary statistics. The resulting random forests (i.e. sets of decision trees) can then be used to estimate the posterior probability distributions of concurrent models and their parameters from the observed summary statistics. Forty thousand simulations were generated to create the reference table and random forests of 1,000 trees were used for the parameter and posterior probability estimation. Prior errors from the ABC-random forests were calculated with an out-of-bag approach (see Marin *et al.* (2016) and Pudlo *et al.* (2016) for details).

In our analysis, a model of two populations (Eastern and Western clusters) was evaluated. Each population was characterized by a parameter $\theta$ ($\theta_W = 4N_W\mu$ and $\theta_E = 4N_E\mu$, where $N_W$ is the effective population size of the Western population, $N_E$ is the effective population size of the Eastern population, and $\mu$ is the mutation rate). The Western population was founded by individuals from the Eastern population at time $T=t/4N_W$ (time t measured in number of

11

generations). Two concurrent models were evaluated regarding the presence or absence of gene

flow between the two populations. If gene flow was detected, an additional parameter, i.e. the

scaled migration rate $M=4N_W m$, was included. Microsatellites were assumed to mutate according

to a generalized stepwise mutation model (GSM) in which the number of repeat units gained or

lost in each mutation is taken from a geometric distribution with parameter $P_{GSM}$. Data under this

model were generated by simulation using coalescent simulator ms (Hudson, 2002) with a

custom script (see below) to transform its output into microsatellite data. Each simulated dataset

was summarized on the basis of population genetics statistics to characterize microsatellite

genetic diversity and population differentiation, which are known to be informative about

demographic patterns (Supporting Information, Table S2). Parameter values at each simulation

were sampled from the prior probability distribution specified in Table 2. Point estimates of

parameters at the natural scale were not estimated as estimates of microsatellite mutation rate are

not available for *A. agrarius* or other *Apodemus* species.

Approximate Bayesian computation was performed with microsatABC-IM (Navascués,

2017), which uses R (R Core Team, 2013) with functions from pegas (Paradis, 2010), mmod

(Winter, 2012) and adegenet (Jombart, 2008) for calculating summary statistics and abcrf (Pudlo

*et al.*, 2016) for performing random forest analyses.

**RESULTS**

**Mitochondrial DNA analysis**

Phylogenetic and phylogeographic analysis

A total of 121 haplotypes were identified within our cyt*b* dataset (Supporting Information, Table

S3). All new sequences have been deposited in GenBank (accession numbers MH257777 -

MH257893). The complete data matrix included these 121 haplotypes as well as two *Apodemus*

12

*chevrieri* haplotypes as outgroups. This matrix provided 923 base pairs (bps), of which 208 sites were variable. The nucleotide frequencies were 29.2%, 26.7%, 30.4% and 13.6% for T, C, A and G, respectively. ML analyses were performed using the HKY85 + I + Gamma model suggested for the dataset using the Akaike information criterion estimated using jModelTest, with a proportion of invariable sites of 0.616 and gamma distribution shape parameter of 0.8323.

Three main lineages were recovered in the ML phylogenetic tree (Supporting Information, Fig. S1). The first two lineages to diverge were well supported and included all haplotypes from Jeju Island in South Korea (BS = 83%) and Taiwan (BS = 94%), respectively. The third lineage was weakly supported (BS = 20%) and did not show any clear phylogeographic structure. Haplotypes corresponding to animals coming from the whole distribution area were mixed in this lineage and not associated within the supported clades (Supporting Information, Fig. S1). However, two genetic groups corresponding to *A. agrarius* specimens from Europe, Turkey, Central Russia and Western Siberia (Western fragment), on the one hand, and all striped field mice from the Eastern fragment, on the other, were documented in our median joining network (Fig. 2). Interestingly, some cyt*b* haplotypes are shared among individuals from Russian Far East and Europe, from Europe and Central Russia and from Western Siberia and Central Russia.

Analysis of genetic diversity and differentiation

We calculated nucleotide and haplotype diversities for the two main striped field mice groups (Eastern and Western Palearctic) and the five regional subgroups. The results of these analyses are summarized in Table 3 and indicated that populations from within the Western Palearctic range were characterised by weaker nucleotide diversity values (from 0.0062) as compared to those within the Eastern Palearctic range (0.0159). Haplotype diversity was high and similar in

all groups, except in Central Russia where it was lower. Findings of Fu's *Fs* test of neutrality were significant for all groups (Table 3), which indicated population expansion. $F_{ST}$ estimates among the five regional subgroups (Table 4) confirmed stronger genetic differentiation ($F_{ST} >$ 0.20) between the Eastern and Western subgroups.

The AMOVA results showed that the largest part of the total mtDNA variation (71.7%) occurred within the regional subgroups, whereas a low percentage of this variation (17.4%) was observed between the two defined groups (Eastern and Western Palearctic) and among subgroups within the two groups (10.9%).

Demographic history (IM model)

The estimated current population size of the Eastern lineage was threefold larger than that of the Western lineage (Fig. 3). The divergence time between these two lineages was estimated at 154 kya (95% CI: 97 – 228 kya) under the IM model (Table 5). Gene flow was estimated at around 4.5 female migrants per generation from East to West and 1.4 migrant per generation from West to East.

**Microsatellite data analysis**

Genetic diversity

The NA frequency values determined in FreeNA were very low for each locus in each group, except for the AGRA11 locus in Korea and Russian Far East (>10%). Observed heterozygosity and allelic richness were higher in the Eastern group and subgroups (Korea and Russian Far East) (Table 6). Tests for HWE showed deviation from the expected frequencies in all groups. All inbreeding coefficients (*F*is) were significant (in bold in Table 6).

Population structure

We used the ΔK method described by Evanno *et al.* (2005) to interpret the STRUCTURE output. The highest ΔK was found at K = 2 (Supporting Information, Fig. S2). For K = 2, the Korean populations clustered with populations from the Russian Far East (Eastern group) (Fig. 4). The second cluster corresponded to the Western group (European, Turkish, Russian, Ukrainian and Kazakh populations). The Eastern cluster (Korea + Russian Far East) was recovered until K = 5 (Supporting Information, Fig. S3).

<u>Demographic history</u>

Distinguishing between models was difficult with the ABC-random forest approach, with a prior error rate of 0.42. A model with migration between Western and Eastern populations was slightly favored over a model of pure divergence with a posterior probability estimated at only 0.54. Because of this low posterior probability, parameters common to both models were estimated from the reference table for both models. The migration rate was also estimated for the isolation-with-migration model. Point estimates (median of posterior probability distribution) and 95% HPD intervals are reported in Table 2 and a more detailed description of the posterior distribution is presented in the Supporting Information (Fig. S4). The estimated effective population size of the Eastern group ($\theta_E$) was 2.25-fold higher than that of the Western group ($\theta_W$).

**DISCUSSION**

**Origin and colonisation history of *Apodemus agrarius***

Our results showed high mitochondrial genetic homogeneity among *A. agrarius* populations throughout the Palearctic region, but microsatellite markers—which mutate more rapidly—detected a finer population structure with the genetic differentiation of populations from the Eastern and Western distribution ranges. These findings suggest a recent separation between the *A. agrarius* groups, with relatively frequent gene flow among them.

15

Striped field mouse populations in the Eastern range were characterized by higher genetic diversity than those in the Western range, which confirmed that the species originated in Eastern Asia, likely around 800 kya (Suzuki *et al.*, 2008). The origin of *A. agrarius* in Eastern Asia was also corroborated by our estimates of effective population size for both cyt*b* sequences and microsatellites, which highlighted a population size around threefold higher in the Eastern part of the range as compared to the Western part. The center of *A. agrarius* origin might be located in the China/Korea area as these populations had the highest genetic diversity and the oldest expansion time estimate (206 kya). Later, *A. agrarius* expanded into other East Asian areas, such as the Primorye region (Russian Pacific Far East), around 145 kya, and colonised the entire Far East region. *A. agrarius* might also have colonised the Western part of its distribution range at that time, at the end of the Middle Pleistocene. The lower genetic diversity of Western populations could be the result of founder events associated with a low number of colonisers coming from the East, followed by a recent population expansion in the West. This western colonisation likely originated from Far East Russian populations as the median-joining network showed closer relationships and shared haplotypes between these two populations. Western colonisation via China appears unlikely as substantial biogeographic barriers, such as the Himalayan Mountains or the Gobi Desert, have hemmed in this region for several million years. The Western colonisation from the Russian Far East could have happened during one of the last interglacials, before 150 kya, when the warmer climatic conditions would have given rise to a mosaic of forests, meadows, bushes and forest-steppe grasslands in the Transbaikalia region (Batuyev *et al.*, 2000, Velichko, 2009). Indeed, these habitats correspond to the ecological preference of *A. agrarius*, and this species cannot survive in taiga or tundra habitats (Karaseva *et al.*, 1992, Okulova *et al.*, 2012).

16

Soon after their expansion in the Central Asian region, the Central Palearctic populations became isolated from their Eastern relatives. The IM analyses suggested a separation between these two distribution ranges at around 150 kya. This dating estimation for the disjunction of the *A. agrarius* range appears much older as compared to those proposed by Atopkin *et al.* (2007), who estimated that this separation occurred later, during the Holocene (<12 kya), and that it was associated with a heavy dry period in this region, which caused the decline and sometimes even the disappearance of trees and shrubs. However, similar climatic events might have occurred around 150 kya, at the end of the Riss Ice Age, which was characterized by particularly cold and dry climates, and would have led to similar isolation (Velichko, 2009). Fossil records in the Transbaikalia region indicated that this period (Middle to Late Pleistocene) was characterized by a faunal transition and the expansion of dry cold steppes and small mammal species associated with this environment (Erbajeva *et al.*, 2013).

From this period, the Central-Western Palearctic populations (Western group) started to differentiate from the Russian Far East and Chinese populations, even though some gene flow still took place between the isolated ranges, as suggested by cyt*b* and microsatellite data. This gene flow seemed to be higher from East to the West than vice-versa. Populations of the Western group progressively increased and expanded throughout Central Asia to colonise the European and Turkish regions. According to the low levels of nucleotide diversity as well as the sign of recent expansion revealed by Fu's Fs index, this expansion throughout Central Asia and Europe would have occurred relatively quickly. Our expansion time estimates suggest that the population expansion started in Central Russia and Siberia around 76-55 kya, followed by a last expansion in Europe 40-50 kya. This scenario is corroborated by the findings of several paleontological studies, which suggested the presence of *A. agrarius* in Central Europe around

50 kya (Popov, 2017) and in Southwestern France around 19 kya, although the species is now extinct in this latter region (Aguilar *et al.*, 2008). This confirmed the assumption that *A. agrarius* is a relatively new member of the European fauna (Martín Suárez & Mein, 1998, Kowalski, 2001, Knitlová & Horáček, 2017, Popov, 2017).

The Central Asian expansion could have been favored during an interglacial period when some habitats preferred by the striped field mouse (e.g. mixed vegetation with grasslands, bushes, shrubs, mosaic of forests and meadows) were distributed at the interface between taiga and tundra or steppe habitats in a large area of Central Asia (Velichko, 2009). Herbivorous megafauna could also have favored the persistence of such habitats in the region (zoogenic landscape) (Bakker *et al.*, 2016). In contrast, the last Ice Age probably enabled *A. agrarius* to expand into Western Europe, as during this period the European deciduous forest was replaced by a mosaic of open habitats and coniferous forests (Fletcher *et al.*, 2010). At the beginning of the Holocene, the striped field mouse distribution range probably regressed from most of Western Europe when deciduous forests started their postglacial recolonisation. This resulted in the species only surviving in Central European open habitats. It is only during recent periods, when human activities have created open habitats via agricultural development, that the striped field mouse has been able to recolonise some Western European regions (Germany, Denmark, Italy, and more recently Austria, Hungary, Slovakia and Czech Republic) (Spitzenberger & Engelberger, 2014).

This type of colonisation pattern from Far East Asia is quite rare for mammals. It has probably been facilitated by the high ecological plasticity and synanthropic habits of the striped field mouse. To our knowledge, this pattern has only been observed in the harvest mouse (*Micromys minutus*) (Yasuda *et al.*, 2005). These latter authors described a close genetic

relationship between Western and Eastern Palearctic *M. minutus* populations and a process of recolonisation of Europe from refugia located in Central and Eastern Asia around 80 kya. A similar pattern has also been reported in roe deer (*Capreolus pygargus*) as populations from Lithuania and Poland appear to be genetically closely related to those from Central and Far East Asia, suggesting recent colonisation of Europe from these Eastern regions (Lorenzini *et al.*, 2014).

**Eastern refuge areas**

*Apodemus agrarius* showed a complex genetic structure within the Eastern range. China, the Russian Far East and Korea correspond to important centers of diversification for this species as the genetic diversity levels (particularly the nucleotide diversity of the cyt*b* gene) are significantly higher within these regions as compared to others. This diversification could be the result of repeated population isolation during the Quaternary Ice Ages, which led to allopatric differentiation, as suggested by the high levels of nucleotide diversity characterising striped field mice within these regions. During these periods, the cooler climate allowed the extension of the Gobi Desert towards Pacific areas. According to Zhou *et al.* (2004), the extension of arid zones in China during the Ice Ages was probably linked to the high irregularity in the mode of the monsoon seasons. This in turn probably led to isolation of the Russian Far East (Primorye and Khabarovsk regions) from several Chinese regions (Zhang *et al.*, 2008) as well as from populations in the Korean Peninsula (Harrison *et al.*, 2001, Zhou *et al.*, 2004, Zhang *et al.*, 2008, Koh *et al.*, 2014, Kim & Park, 2015). The Korean Peninsula, characterized by a temperate mountain climate in its southern part, was less deeply affected by the global Quaternary climate changes (Liu & Li, 1996, Kim & Park, 2015) and therefore could have acted as glacial refugia for many organisms during the Quaternary coldest phases. The Russian Far East as well as China

could also have been a potential Quaternary refugium for *A. agrarius* (Atopkin *et al.*, 2007, Dokuchaev *et al.*, 2008, Sakka *et al.*, 2010, Pereverzeva & Pavlenko, 2014, Pereverzeva *et al.*, 2017). Our findings corroborate results obtained in other studies carried out on the Asian *Apodemus* species (Suzuki *et al.*, 2003, Sakka *et al.*, 2010) but also on several other vertebrate species, e.g. fish, amphibians, birds and mammals (Kryukov, 2010).

**Taxonomic implication**

Corbet (1978) classified *A. agrarius* populations from Europe and Western and Central Asia as the subspecies *A. a. agrarius*, while populations from Eastern Asia were considered as the subspecies *A. a. ningpoensis*. Later, in their Russian taxonomical review, Gromov & Erbajeva (1995) proposed three subspecies on the Russian territory: *A. a. karelicus*, *A. a. agrarius* s.str. and *A. a. mantchuricus*. This latter subspecies was also described in continental China in addition to two other subspecies: *A. a. pallidior* and *A. a. ningpoensis* (Zhang, 1997), but Liu *et al.* (1991) considered that *A. a. pallidior* should be invalidated and therefore, only *A. a. mantschuricus* and *A. a. ningpoensis* should be recognized in China. Some morphological (Jones & Johnson, 1965) and genetic studies (Koh *et al.*, 1998) also suggested that striped field mice from the Korean peninsula could be differentiated from the other Far East populations and considered them as two different subspecies: *A. a. pallescens* in South Korea and *A. a. coreae* in Central Korea. However, more recent studies based on morphometric features (Koh *et al.*, 1998), karyotype analyses (Kartavtseva & Pavlenko, 2000) and genetic markers (Atopkin *et al.*, 2007, Suzuki *et al.*, 2008, Sakka *et al.*, 2010, Koh *et al.*, 2014), as well as the findings of the present study, did not indicate any distinction between these subspecies. Our cyt*b* gene sequence analyses showed high genetic homogeneity between mainland populations from the Eastern and Western distribution ranges. According to these findings, no subspecies should be recognized

within all *A. agrarius* Palearctic continental populations, including those from China and the Russian Far East.

However, Koh *et al.* (2014) suggested that some differentiation would exist in insular populations from Taiwan and Jeju Island and therefore proposed to consider these insular populations as two different subspecies, *A. a. insulaemus* and *A. a. chejuensis*, respectively. The results of the present study also confirmed the genetic distinctness of these two populations (phylogenetic tree and median-joining network). Further taxonomic investigations are needed to confirm these results.

**CONCLUSION**

Our study, for the first time, provides an overview of the evolutionary and demographic history of the striped field mouse throughout the Palearctic region based on our findings using sensitive genetic markers. Our results suggest that this species appeared around 800 kya in the Asian Far East, likely in China. Around 150 kya, it colonised the Central Palearctic region, probably from a low number of founders from the Russian Far East. From these regions, the species progressively increased its range and expanded relatively quickly throughout Central Asia during the last 80 kyrs, to finally colonise the Western European region, until it reached Southwestern France, around 19 kya, where it is presently extinct. *A. agrarius* is therefore an Asiatic immigrant and a relatively new member of the European fauna. This peculiar phylogeographic pattern highlights the importance of Far East Asian regions as a center of origin and diversification for several Palearctic species and as a source for the European biodiversity. This highlights the complexity of the origin of the existing European fauna, where many species have survived in European refugia during the Quaternary glaciations, whereas several others came from much more distant origins like Central Asia (*e.g. Microtus arvalis* (Haynes *et al.*, 2003), *M. oeconomus* (Brunhoff *et*

21

*al.*, 2003) and *Cricetus cricetus* (Neumann *et al.*, 2005)) or even Far East Asia (*e.g. Micromys minutus* (Yasuda *et al.*, 2005) and *A. agrarius*, present study).

## ACKNOWLEDGEMENTS

**REFERENCES**

**Aguilar J-P, Pélissié T, Sigé B, Michaux J. 2008.** Occurrence of the Stripe Field Mouse lineage (*Apodemus agrarius* Pallas 1771; Rodentia; Mammalia) in the Late Pleistocene of southwestern France. *Comptes Rendus Palevol* **7:** 217-225.

**Andersen LW, Jacobsen M, Vedel-Smith C, Jensen TS. 2017.** Mice as stowaways? Colonization history of Danish striped field mice. *Biology Letters* **13:** 20170064.

**Atopkin DM, Bogdanov AS, Chelomina GN. 2007.** Genetic variation and differentiation in striped field mouse *Apodemus agrarius* inferred from RAPD-PCR analysis. *Russian Journal of Genetics* **43:** 665-676.

**Bakker ES, Gill JL, Johnson CN, Vera FWM, Sandom CJ, Asner GP, Svenning J-C. 2016.** Combining paleo-data and modern exclosure experiments to assess the impact of megafauna extinctions on woody vegetation. *Proceedings of the National Academy of Sciences* **113:** 847-855.

**Bandelt HJ, Forster P, Rohl A. 1999.** Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37-48.

**Batuyev AR, Buyantuyev AB, Snytko VA. 2000.** *Geosistemy i kartografirovaniye ekologo-geograficheskikh situatsiy Priselenginskikh kotlovin Baykalskogo regiona [The geosystems*

*and mapping of the ecogeographical situations of the Priselenginsk depressions of the Baikal region]*. Novosibirsk [In Russian].

**Bazhenov YA, Pavlenko MV, Korablev VP, Kardash AI. 2015.** Current distribution of the striped field mouse (*Apodemus agrarius* Pallas, 1771) in Eastern Transbaikalia: New findings in the disjunction area. *Russian Journal of Biological Invasions* **6:** 1-5.

**Bilton DT, Mirol PM, Mascheretti S, Fredga K, Zima J, Searle JB. 1998.** Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265:** 1219-1226.

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

**Chapuis MP, Estoup A. 2007.** Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24:** 621-631.

**Corbet GB. 1978.** *The Mammals of the Palaearctic Region : a taxonomic review*. London and Ithaca, NY: British Museum (Natural History) and Cornell University Press.

**Crandall KA, Templeton AR. 1993.** Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134:** 959-969.

**Dalén L, Nyström V, Valdiosera C, Germonpré M, Sablin M, Turner E, Angerbjörn A, Arsuaga JL, Götherström A. 2007.** Ancient DNA reveals lack of postglacial habitat

tracking in the arctic fox. *Proceedings of the National Academy of Sciences* **104:** 6726-6729.

**Deffontaine V, Ledevin R, Fontaine MC, Quere JP, Renaud S, Libois R, Michaux JR. 2009.** A relict bank vole lineage highlights the biogeographic history of the Pyrenean region in Europe. *Molecular Ecology* **18:** 2489-2502.

**Deffontaine V, Libois R, KotlÍK P, Sommer R, Nieberding C, Paradis E, Searle JB, Michaux JR. 2005.** Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology* **14:** 1727-1739.

**Dokuchaev NE, Lapinskii AG, Solovenchuk LL. 2008.** Genetic diversity of the striped field mouse (*Apodemus agrarius* Pallas, 1771) in the Russian Far East as assessed by RAPD-PCR. *Biology Bulletin* **35:** 368-373.

**Erbajeva MA, Khenzykhenova FI, Alexeeva NV. 2013.** Aridization of the Transbaikalia in the context of global events during the Pleistocene and its effect on the evolution of small mammals. *Quaternary International* **284:** 45-52.

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

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

**Fedorov VB, Stenseth NC. 2001.** Glacial survival of the Norwegian lemming (*Lemmus lemmus*) in Scandinavia: inference from mitochondrial DNA variation. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268:** 809-814.

**Fedorov VB, Goropashnaya AV, Boeskorov GG, Cook JA. 2008.** Comparative phylogeography and demographic history of the wood lemming (*Myopus schisticolor*): implications for late Quaternary history of the taiga species in Eurasia. *Molecular Ecology* **17:** 598-610.

**Fischer S, Mayer-Scholl A, Imholt C, Spierling NG, Heuser E, Schmidt S, Reil D, Rosenfeld UM, Jacob J, Nöckler K, Ulrich RG. 2018.** *Leptospira* Genomospecies and Sequence Type Prevalence in Small Mammal Populations in Germany. *Vector-Borne and Zoonotic Diseases* **18:** 188-199.

**Fletcher WJ, Sánchez Goñi MF, Allen JRM*, et al.* 2010.** Millennial-scale variability during the last glacial in vegetation records from Europe. *Quaternary Science Reviews* **29:** 2839-2864.

**Fontaine MC, Tolley KA, Michaux JR, Birkun A, Ferreira M, Jauniaux T, Llavona Á, Öztürk B, Öztürk AA, Ridoux V, Rogan E, Sequeira M, Bouquegneau J-M, Baird SJE. 2010.** Genetic and historic evidence for climate-driven population fragmentation in a top cetacean predator: the harbour porpoises in European water. *Proceedings of the Royal Society B: Biological Sciences* **277:** 2829-2837.

**Fu YX. 1997.** Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147:** 915-925.

**Gajda E, Hildebrand J, Sprong H, Buńkowska-Gawlik K, Perec-Matysiak A, Coipan EC. 2017.** Spotted fever rickettsiae in wild-living rodents from south-western Poland. *Parasites & Vectors* **10:** 413.

**Gortat T, Rutkowski R, Gryczynska-Siemiatkowska A, Kozakiewicz A, Kozakiewicz M. 2013.** Genetic structure in urban and rural populations of *Apodemus agrarius* in Poland. *Mammalian Biology - Zeitschrift für Säugetierkunde* **78:** 171-177.

**Goudet J. 2001.** FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.  Updated from Goudet (1995).

**Gromov I, Erbajeva M. 1995.** The Mammals of Russia and Adjacent Territories; Lagomorphs and Rodents, vol. 167. St Petersburg: Russian Academy of Science, Zoological Institute

**Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010.** New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59:** 307-321.

**Hall TA. 1999.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95-98.

**Harrison SP, Yu G, Takahara H, Prentice IC. 2001.** Diversity of temperate plants in east Asia. *Nature* **413:** 129.

**Hasegawa M, Kishino H, Yano T-A. 1985.** Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22:** 160-174.

27

**Haynes S, Jaarola M, Searle JB. 2003.** Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. *Molecular Ecology* **12:** 951-956.

**Hewitt GM. 1996.** Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58:** 247-276.

**Hewitt GM. 1999.** Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68:** 87-112.

**Hewitt GM. 2000.** The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907-913.

**Hey J, Nielsen R. 2004.** Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167:** 747-760.

**Hudson RR. 2002.** Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18:** 337-338.

**Hughes ALC, Gyllencreutz R, Lohne ØS, Mangerud J, Svendsen JI. 2016.** The last Eurasian ice sheets – a chronological database and time-slice reconstruction, DATED-1. *Boreas* **45:** 1-45.

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

**Jo Y-S, Kim H-N, Baccus John T, Jung J. 2017.** Genetic differentiation of the Korean striped field mouse, *Apodemus agrarius* (Muridae, Rodentia), based on microsatellite polymorphism. *Mammalia* **81:** 297.

**Jombart T. 2008.** adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403-1405.

**Jones JKJ, Johnson DN. 1965.** *Synopsis of lagomorphs and rodents of Korea*. Kansas City: University of Kansas Publications Museum Natural History.

**Jonsson CB, Figueiredo LTM, Vapalahti O. 2010.** A Global Perspective on Hantavirus Ecology, Epidemiology, and Disease. *Clinical Microbiology Reviews* **23:** 412-441.

**Karaseva E, Tichonova G, Bogomilov P. 1992.** Distribution of the Striped field mouse (*Apodemus agrarius*) and peculiarities of its ecology in different parts of its range. *Zoologicheskii Zhurnal* **71:** 106-115.

**Kartavtseva IV, Pavlenko MV. 2000.** Chromosomal variability of the field mouse *Apodemus agarius* (Rodentia, Muridae). *Genetika* **36:** 223-236.

**Khlyap LA, Warshavsky AA, Bobrov VV. 2011.** Diversity of alien mammal species in different regions of Russia. *Russian Journal of Biological Invasions* **2:** 293-299.

**Kim HR, Park YC. 2015.** Genetic diversity and genetic structure of the striped field mouse *Apodemus agrarius coreae* (Muridae, Rodentia) in Korea. *Gene* **572:** 292-297.

**Klempa B, Avsic-Zupanc T, Clement J, Dzagurova TK, Henttonen H, Heyman P, Jakab F, Kruger DH, Maes P, Papa A, Tkachenko EA, Ulrich RG, Vapalahti O, Vaheri A. 2013.** Complex evolution and epidemiology of Dobrava-Belgrade hantavirus: definition of genotypes and their characteristics. *Archives of Virology* **158:** 521-529.

**Knitlová M, Horáček I. 2017.** Late Pleistocene-Holocene paleobiogeography of the genus *Apodemus* in Central Europe. *PLoS ONE* **12:** e0173668.

**Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC. 1989.** Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences* **86:** 6196-6200.

**Koh HS, Csorba G, Tiunov MP, Tikhonova G. 1998.** Morphometric analyses of the three subspecies of striped field mouse, *Apodemus agrarius* Pallas (Mammalia: Rodentia) from Far Eastern Asia: taxonomix status of North Korean striped field mice. *The Korean Journal of Systematic Zoology* **14:** 327-334.

**Koh HS, Shaner PJ, Csorba G, Wang Y, Jang KH, Lee JH. 2014.** Comparative genetics of *Apodemus agrarius* (Rodentia: Mammalia) from insular and continental Eurasian populations: cytochrome b sequence analyses. *Acta Zoologica Academiae Scientiarum Hungaricae* **60:** 73-84.

**Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015.** Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15:** 1179-1191.

**Kowalski K. 2001.** Pleistocene rodents of Europe. *Folia Quaternaria* **72:** 3-389.

**Kraljik J, Paziewska-Harris A, Miklisová D, BlaŇArová L, MoŠAnskÝ L, Bona M, Stanko M. 2016.** Genetic diversity of *Bartonella* genotypes found in the striped field mouse (*Apodemus agrarius*) in Central Europe. *Parasitology* **143:** 1437-1442.

**Kryukov AP. 2010.** Comparative phylogeographic patterns of several vertebrates in the east Palearctic. *Moscow University Biological Sciences Bulletin* **65:** 184-186.

**Latif A, Liu B-Y, Chen Z, Sun Y, Shi Y-L, Zong J, Li J-J, Ren C-P, Zhang X-C, Liu X-N, Yu X-J, Liu Y. 2017.** *Orientia tsutsugamushi* infection in rodents in Anhui Province of China. *Infection, Genetics and Evolution* **56:** 14-18.

**Lee HW, Lee PW, Johnson KM. 1978.** Isolation of the Etiologic Agent of Korean Hemorrhagic Fever. *The Journal of Infectious Diseases* **137:** 298-308.

**Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

**Liu C, Wu W, Guo S, Meng J. 1991.** A study of the subspecies classification of *Apodemus agrarius* in eastern continental China. *Acta Theriologica Sinica* **11:** 294-299.

**Liu DS, Li ZG. 1996.** *Geography of Asia*. Beijing: Commercial Press.

**Lorenzini R, Garofalo L, Qin X, Voloshina I, Lovari S. 2014.** Global phylogeography of the genus *Capreolus* (Artiodactyla: Cervidae), a Palaearctic meso-mammal. *Zoological Journal of the Linnean Society* **170:** 209-221.

31

**Makova KD, Patton JC, Krysanov EY, Chesser RK, Baker RJ. 1998.** Microsatellite markers in wood mouse and striped field mouse (genus *Apodemus*). *Molecular Ecology* **7:** 247-255.

**Marin J-M, Raynal L, Pudlo P, Ribatet M, Robert CP. 2016.** ABC random forests for Bayesian parameter inference. *arXiv* 1605.05537.

**Martín Suárez E, Mein P. 1998.** Revision of the genera *Parapodemus*, *Apodemus*, *Rhagamys* and *Rhagapodemus* (Rodentia, Mammalia). *Geobios* **31:** 87-97.

**Mezhzherin SV, Zykov AE. 1991.** Genetic divergence and allozymic variability in mice of the genus *Apodemus* s. lato (Muridae, Rodentia). *TSitologiia i genetika* **25:** 51-59.

**Michaux JR, Libois R, Filippucci MG. 2005.** So close and so different: comparative phylogeography of two small mammal species, the Yellow-necked fieldmouse (*Apodemus flavicollis*) and the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Heredity* **94:** 52-63.

**Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R. 2003.** Mitochondrial phylogeography of the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Molecular Ecology* **12:** 685-697.

**Musser GG, Carleton M. 2005.** Superfamily Muroidea. In: Wilson DE, Reeder DM, eds. *Mammal species of the World: a taxonomic and geographic reference*. Baltimore: Johns Hopkins University Press, 894–1531

**Navascués M. 2017.** microsatABC-IM: Approximate Bayesian computation inferences under the isolation with migration model from microsatellite data (Version v1.0). Zenodo.

32

**Nemirov K, Vapalahti O, Lundkvist A, Vasilenko V, Golovljova I, Plyusnina A, Niemimaa J, Laakkonen J, Henttonen H, Vaheri A, Plyusnin A. 1999.** Isolation and characterization of Dobrava hantavirus carried by the striped field mouse (*Apodemus agrarius*) in Estonia. *Journal of General Virology* **80:** 371-379.

**Neumann K, Michaux JR, Maak S, Jansman HAH, Kayser A, Mundt G, Gattermann R. 2005.** Genetic spatial structure of European common hamsters (*Cricetus cricetus*) — a result of repeated range expansion and demographic bottlenecks. *Molecular Ecology* **14:** 1473-1483.

**Okulova NM, Duvanova IA, Kalinkina EV, Mironova TA, Nedosekin VY, Drozdova VF. 2012.** On the ecology of striped field mouse (*Apodemus agrarius* Pall., 1771) in forest-steppes of the chernozem zone: 1. Abundance. *Biology Bulletin* **39:** 809-815.

**Paradis E.** 2010. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26:** 419-420.

**Perekrest VV. 2017.** *The Great Illustrated Atlas of the World*. Moskow.

**Pereverzeva VV, Pavlenko MV. 2014.** Diversity of the mitochondrial DNA cytochrome b gene of the field mouse *Apodemus agrarius* Pallas, 1771 in the south of the Russian Far East. *Biology Bulletin* **41:** 1-11.

**Pereverzeva VV, Primak AA, Pavlenko MV, Dokuchaev NE, Evdokimova AA. 2017.** Genetic features and the putative sources of formation of isolated populations of the striped

field mouse *Apodemus agrarius* Pallas, 1771 in Magadan oblast. *Russian Journal of Biological Invasions* **8:** 87-100.

**Popov V. 2017.** A Pleistocene record of *Apodemus agrarius* (Pallas, 1771) (Mammalia: Rodentia) in the Magura Cave, Bulgaria. *Acta Zoologica Bulgarica* **69:** 121-124.

**Posada D. 2008.** jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25:** 1253-1256.

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

**Pudlo P, Marin J-M, Estoup A, Cornuet J-M, Gautier M, Robert CP. 2016.** Reliable ABC model choice via random forests. *Bioinformatics* **32:** 859-866.

**R Core Team. 2013.** *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

**Ramos-Onsins SE, Rozas J. 2002.** Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19:** 2092-2100.

**Reshchikov MA. 1961.** *Steppe of Western Transbaikalia* [in Russian].

**Rousset F. 2008.** GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8:** 103-106.

**Sakka H, Quere JP, Kartavtseva I, Pavlenko M, Chelomina G, Atopkin D, Bogdanov A, Michaux J. 2010.** Comparative phylogeography of four *Apodemus* species (Mammalia:

34

Rodentia) in the Asian Far East: evidence of Quaternary climatic changes in their genetic structure. *Biological Journal of the Linnean Society* **100:** 797-821.

**Sheremetyeva IN, Kartavtseva IV, Pavlenko MV, Kostenko VA, Sheremetyev IS, Katin IO, Kosoy ME. 2017.** Morphological and genetic variability in small island populations of the striped field mouse *Apodemus agrarius* Pallas, 1771. *Biology Bulletin* **44:** 159-171.

**Spitzenberger F, Engelberger S. 2014.** A new look at the dynamic western distribution border of *Apodemus agrarius* in Central Europe (Rodentia: Muridae). *Lynx, series nova* **45:** 69-79.

**Stewart JR, Lister AM. 2001.** Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* **16:** 608-613.

**Stewart JR, Dalén L. 2008.** Is the glacial refugium concept relevant for northern species? A comment on Pruett and Winker 2005. *Climatic Change* **86:** 19-22.

**Suzuki H, Filippucci MG, Chelomina GN, Sato JJ, Serizawa K, Nevo E. 2008.** A biogeographic view of *Apodemus* in Asia and Europe inferred from nuclear and mitochondrial gene sequences. *Biochemical Genetics* **46:** 329-346.

**Suzuki H, Sato JJ, Tsuchiya K, Luo J, Zhang YP, Wang YX, Jiang XL. 2003.** Molecular phylogeny of wood mice (*Apodemus*, Muridae) in East Asia. *Biological Journal of the Linnean Society* **80:** 469-481.

**Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998.** Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453-464.

35

**Tajima F. 1989.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585-595.

**Velichko AA. 2009.** *Paleoclimates and Paleoenvironments of Extra-tropical Area of the Northern Hemisphere. Late Pleistocene - Holocene*. Moscow: GEOS Publishers.

**Winter DJ. 2012.** mmod: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources* **12:** 1158-1160.

**Xiao H, Tong X, Huang R, Gao L, Hu S, Li Y, Gao H, Zheng P, Yang H, Huang ZYX, Tan H, Tian H. 2018.** Landscape and rodent community composition are associated with risk of hemorrhagic fever with renal syndrome in two cities in China, 2006–2013. *BMC Infectious Diseases* **18:** 37.

**Yasuda SP, Vogel P, Tsuchiya K, Han S-H, Lin L-K, Suzuki H. 2005.** Phylogeographic patterning of mtDNA in the widely distributed harvest mouse (*Micromys minutus*) suggests dramatic cycles of range contraction and expansion during the mid- to late Pleistocene. *Canadian Journal of Zoology* **83:** 1411-1420.

**Zhang H, Yan J, Zhang G, Zhou K. 2008.** Phylogeography and Demographic History of Chinese Black-Spotted Frog Populations (*Pelophylax nigromaculata*): Evidence for Independent Refugia Expansion and Secondary Contact. *BMC Evolutionary Biology* **8:** 21.

**Zhang Y. 1997.** *Distribution of mammalian species in China*. China Forestry Publishing.

**Zhou SZ, Jijun L, Zhang SQ, Zhao JD, Cui JX. 2004.** Quaternary glaciations in China. In: Ehlers J, Gibbard PL, eds. *Developments in Quaternary Sciences,* Vol. 2. Elsevier, 105-113.

36

**TABLES**

Table 1: Number and geographic origin of *A. agrarius* samples used in this study

| Country | cyt*b* dataset | | Microsatellite dataset | |
|---|---|---|---|---|
| | Number of sequences | Number of localities | Number of genotyped individuals | Number of localities |
| Austria | | | 21 | 7 (AU1-7) |
| Bulgaria | | | 8 | 1 (BU) |
| China | 29 | 11 (CH1-11) | | |
| Croatia | 5 | 1 (CRO) | | |
| Czech Republic | 4 | 1 (CZ1) | 10 | 3 (CZ2-4) |
| Denmark | 5 | 1 (DA1) | 19 | 2 (DA1-2) |
| Estonia | 1 | 1 (EST) | 1 | 1 (EST) |
| Finland | 3 | 3 (FIN1-3) | | |
| Germany | 10 | 9 (GE1-9) | 61 | 24 (GE3-26) |
| Hungary | 2 | 1 (HU) | 2 | 1 (HU) |
| Italy | 1 | 1 (IT1) | 1 | 1 (IT2) |
| Kazakhstan | 3 | 2 (KAZ1-2) | | |
| Lithuania | | | 1 | 1 (LIT) |
| Poland | 5 | 1 (PO1) | 31 | 4 (PO1-4) |
| Romania | 4 | 1 (RO1) | 11 | 5 (RO1-5) |
| Russia (Far East) | 52 | 10 (FE1-10) | 68 | 16 (FE1, FE3-5, FE7-18) |
| Russia (European part+Siberia) | 19 | 10 (RU1-10) | 13 | 3 (RU1, RU9-10) |

| Slovakia | 1 | 1 (SLV1) | 79 | 13 (SLV1-13) |
| Slovenia | 6 | 3 (SL1-3) | 1 | 1 (SL1) |
| South Korea | 18 | 5 (KO1-5) | 12 | 4 (KO1-4) |
| Taiwan | 9 | 3 (TAI1-3) | | |
| Turkey | 3 | 1 (TUR) | 1 | 1 (TUR) |
| Ukraine | 2 | 2 (UK1-2) | | |

Table 2. Parameters (coalescent scale) estimated for the isolation using the migration model

|  | prior | prior MSE | median | 95%HPD |
|---|---|---|---|---|
| $\theta_W$ | Log-uniform(0.1,1000) | 1.30 | 5.25 | 1.86–81.69 |
| $\theta_E$ | Log-uniform(0.1,1000) | 1.05 | 11.86 | 6.70–164.53 |
| $T$ | Log-uniform($10^{-5}$,10) | 1.67 | 0.17 | 0.03–8.45 |
| $M$ | Log-uniform($10^{-5}$,100) | 973.82 | $4.96\times10^{-3}$ | $1.24\times10^{-5}$–3.64 |
| $P_{GSM}$ | Uniform(0,1) | $6.39\times10^{-3}$ | 0.50 | 0.09–0.64 |

MSE: mean squared error

Table 3: Diversity estimates and expansion times for *A. agrarius* groups and subgroups

| | Corresponding localities | N | h ± SD | π ± SD | Fu's Fs | Tau | Expansion time (in years) |
|---|---|---|---|---|---|---|---|
| Overall | All | 182 | 0.9933 ± 0.0016 | 0.0135 ± 0.0068 | **-24.05267** | 4.587 | 92 031 |
| | | | | | | | |
| Eastern group | CH1-11, TAI1-3, KO1-5, FE1-10 | 108 | 0.9894 ± 0.0033 | 0.0159 ± 0.0079 | **-24.10376** | 7.244 | 145 339 |
| Western group | CRO, CZ1, DA1, EST, FIN1-3, GE1-9, HU, IT1, PO1, RO1, SLV1, SL1-3, TUR, RU1-10, UK1-2, KAZ1-2 | 74 | 0.9874 ± 0.0056 | 0.0062 ± 0.0034 | **-25.48372** | 2.468 | 49 516 |
| | | | | | | | |
| China, Taiwan, Korea | CH1-11, TAI1-3, KO1-5 | 56 | 0.9890 ± 0.0060 | 0.0180 ± 0.0090 | **-16.00178** | 10.299 | 206 633 |
| Russian Far East | FE1-10 | 52 | 0.9668 ± 0.0114 | 0.0114 ± 0.0058 | **-8.51826** | 5.264 | 105 614 |
| Europe, Turkey | CRO, CZ1, DA1, EST, FIN1-3, GE1-9, HU, IT1, PO1, RO1, SLV1, SL1-3, TUR | 50 | 0.9771 ± 0.0113 | 0.0058 ± 0.0032 | **-25.59017** | 2.014 | 40 408 |
| Central Russia, Ukraine | RU1-5, RU7, RU10, UK1-2 | 14 | 0.8901 ± 0.0807 | 0.0051 ± 0.0029 | **-2.84393** | 3.803 | 76 301 |
| Siberia, | RU6, RU8-9, | 10 | 1.0 ± 0.0447 | 0.0051 ± | **-6.12762** | 2.772 | 55 616 |

| Kazakhstan | KAZ1-2 | | | 0.0029 | | | |

Table 4: $F_{ST}$ among the five subgroups

| | China, Taiwan, Korea | Russian Far East | Europe, Turkey | Central Russia, Ukraine | Western Siberia, Kazakhstan |
|---|---|---|---|---|---|
| China, Taiwan, Korea | | | | | |
| Russian Far East | 0.13 | | | | |
| Europe, Turkey | 0.29 | 0.23 | | | |
| Central Russia, Ukraine | 0.29 | 0.26 | 0.15 | | |
| Siberia, Kazakhstan | 0.27 | 0.23 | 0.19 | 0.04 | |

Table 5: Parameters converted on a demographic scale assuming a mutation rate of 2.7 %/Myr, and a generation time of 0.5 year. The length of the usable sequence was 766 bps.

| | $N_1$ Eastern lineage (inds) | $N_2$ Western lineage (inds) | N Ancestral population (inds) | T (Yrs) | 2N1m1 (inds) | 2N2m2 (inds) |
|---|---|---|---|---|---|---|
| HiSmth | 6,648,293.2 | 2,121,409.9 | 664,829.3 | 154,240.4 | 4.5 | 1.4 |
| HPD90Lo | 4,569,190.6 | 1,468,668.4 | 181,317.1 | 97,186.0 | 1.2 | 0.4 |
| HPD90Hi | 9,162,556.8 | 3,158,543.7 | 1,607,678.2 | 227,734.3 | 19.4 | 6.7 |

For Peer Review

Table 6: Microsatellite genetic diversity within *A. agrarius* groups and subgroups

| | Corresponding localities | n | Ho | He | *F*is | AR |
|---|---|---|---|---|---|---|
| Overall | All | 340 | 0.67633 ± 0.12353 | 0.80644 ± 0.10300 | **0.162** | 21.07 |
| | | | | | | |
| Eastern group | KO1-4, FE1, FE3-5, FE7-18 | 80 | 0.73267 ± 0.10842 | 0.86524 ± 0.05688 | **0.109** | 16.66 |
| Western group | AU1-7, BU, CZ2-4, DA1-2, EST, GE3-26, HU, IT2, LIT, PO1-4, RO1-5, SLV1-13, SL1, TUR, RU1, RU9-10 | 260 | 0.65890 ± 0.15377 | 0.73896 ± 0.16587 | **0.154** | 12.61 |
| | | | | | | |
| Korea | KO1-4 | 12 | 0.73232 ± 0.18672 | 0.84326 ± 0.05644 | **0.137** | 7.84 |
| Russian Far East | FE1, FE3-5, FE7-18 | 68 | 0.73275 ± 0.10193 | 0.85588 ± 0.06176 | **0.145** | 9.01 |
| Europe, Turkey | AU1-7, BU, CZ2-4, DA1-2, EST, GE3-26, HU, IT2, LIT, PO1-4, RO1-5, SLV1-13, SL1, TUR | 247 | 0.65732 ± 0.15898 | 0.73558 ± 0.17095 | **0.107** | 6.94 |
| Central Russia, Siberia | RU1, RU9-10 | 13 | 0.68803 ± 0.13927 | 0.75492 ± 0.11485 | **0.092** | 7.06 |

**Supporting Information**

**Table S1**. Sampling localities and number of samples included in the cyt*b* and microsatellite datasets

**Table S2**. Summary statistics used in ABC

**Table S3**. Haplotype definitions

**Figure S1**. ML phylogenetic tree, including all *A. agrarius* haplotypes

**Figure S2**. Bayesian clustering analysis results obtained with STRUCTURE. (A) Plot of the likelihood of the mean Ln Pr(X|K). (B) ΔK values calculated according to Evanno *et al.* (2005)

**Figure S3**. Population structure estimated using STRUCTURE (K=1 to 10)

**Figure S4**. Prior and posterior probability densities for the demographic and mutational parameters of the model
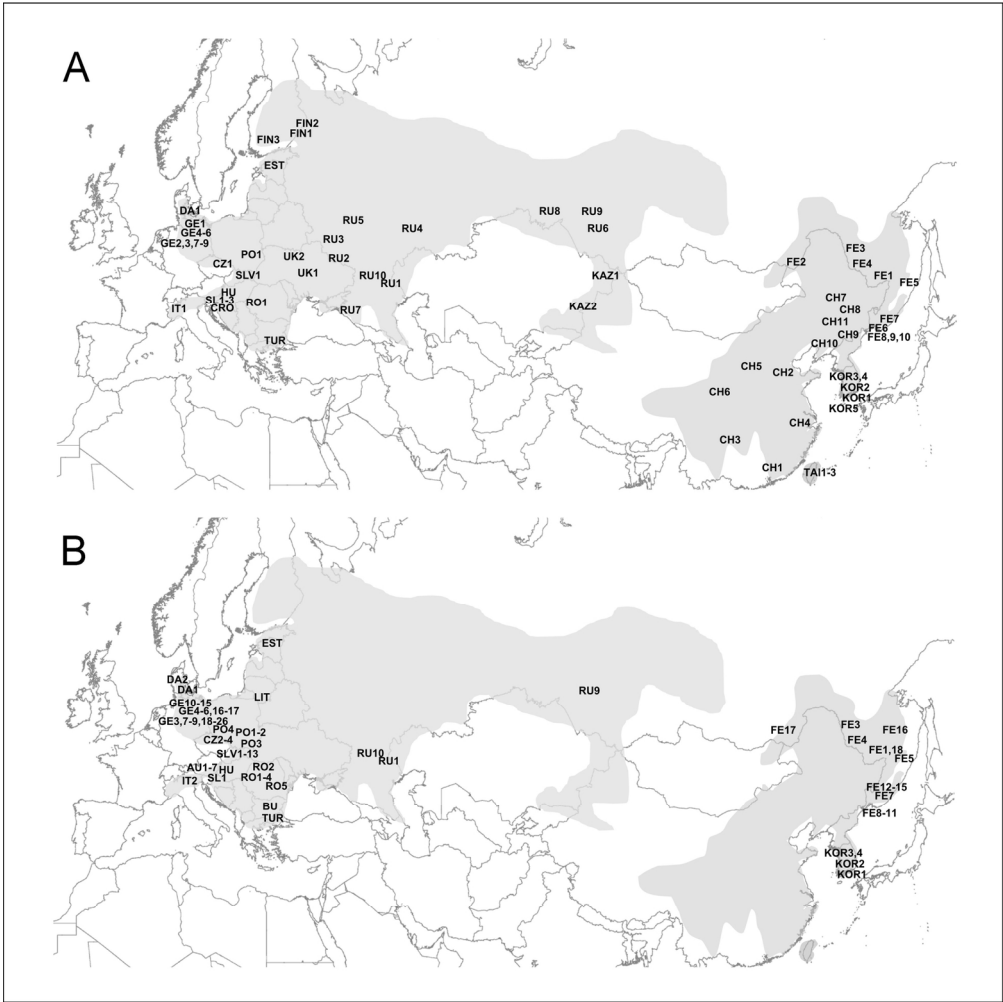
Figure 1: Distribution range of Apodemus agrarius (shaded area) and sampling localities of mitochondrial (A) and microsatellite (B) datasets. The locality codes are given in Table 1.

165x166mm (300 x 300 DPI)
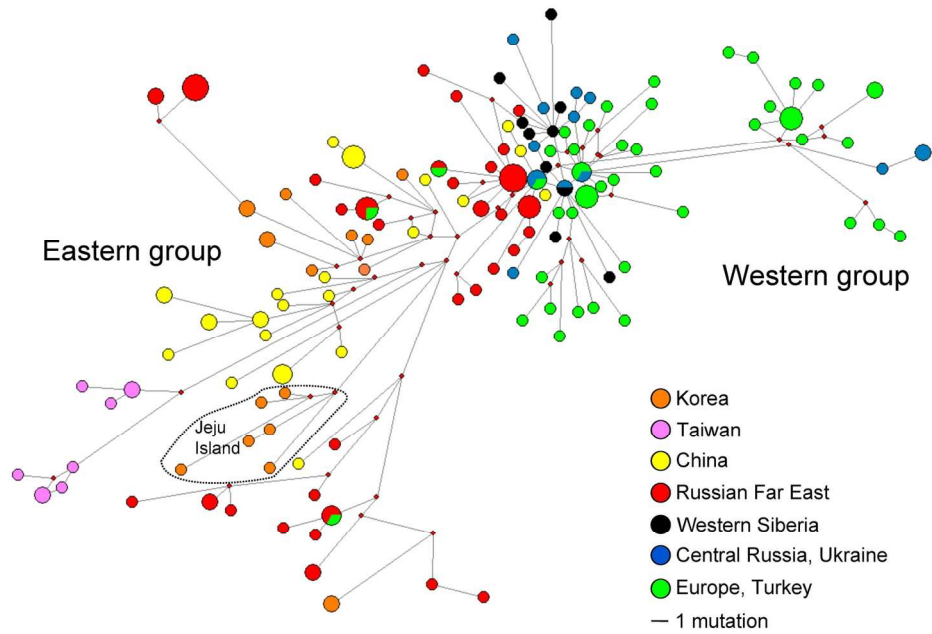
Figure 2: Median-joining network based on the cytb dataset. Circles correspond to distinct haplotypes and circle sizes are proportional to the number of animals sharing this haplotype. Branch lengths are proportional to the number of mutations between haplotypes.

153x119mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
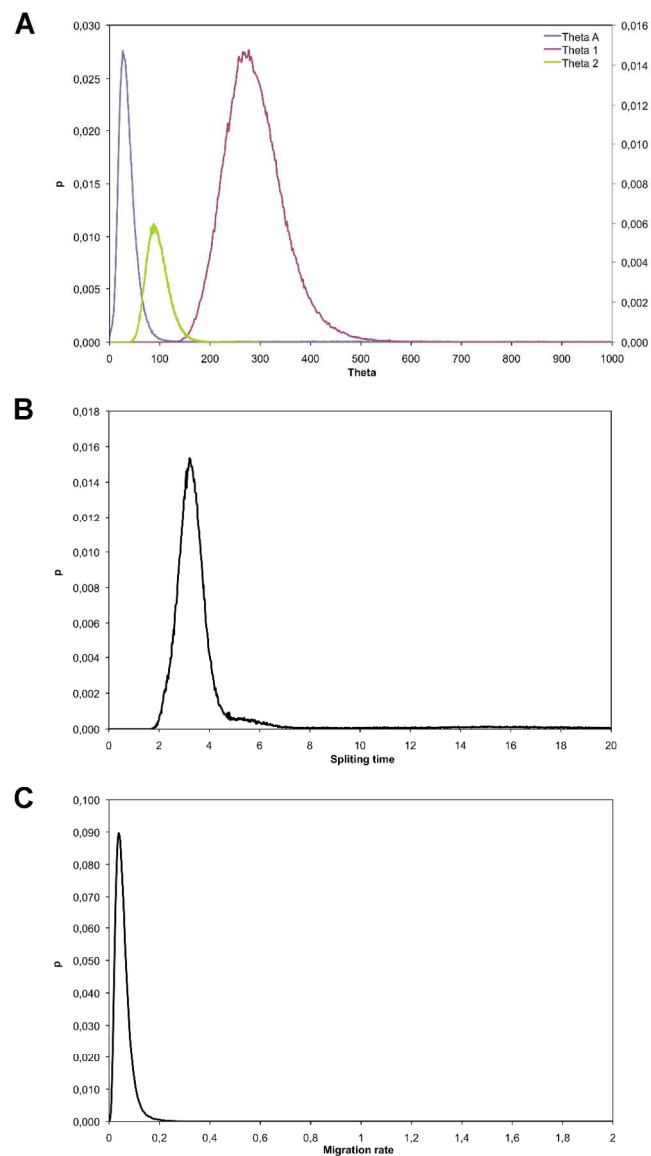


Figure 3: Plots of posterior probability of parameters estimated with the isolation-with-migration model (scaled by the mutation rate μ): (A) Effective population sizes of the Eastern lineage (θ1), Western lineage (θ2) and ancestral population (θA), (B) splitting time between Eastern and Western lineages and (C) migration rate (m) between Eastern and Western lineages

161x268mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
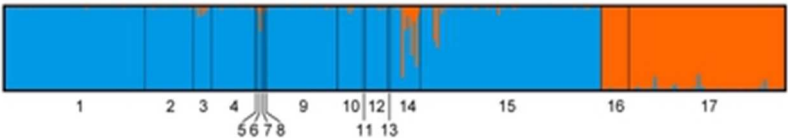


Figure 4: Population structure estimated using STRUCTURE (K=2). Each individual is represented by a vertical line partitioned into K colour segments, with the length of each colour segment being proportional to the estimated membership coefficient. Numbers correspond to the sampling countries: 1 = Germany, 2 = Austria, 3 = Bulgaria, 4 = Denmark, 5 = Estonia, 6 = Hungary, 7 = Italy, 8 = Lithuania, 9 = Poland, 10 = Romania, 11 = Slovenia, 12 = Czech Republic, 13 = Turkey, 14 = Slovakia, 15 = Russia (Central Russia + Siberia), 16 = South Korea, 17 = Russian Far East

34x6mm (300 x 300 DPI)

**Supporting information**

**Table S1**. Sampling localities and number of samples included in cyt*b* and microsatellite datasets

| Country | Locality | Symbol (see Fig 1) | n cyt*b* | n microsatellites | GenBank accession numbers of cyt*b* sequences |
|---|---|---|---|---|---|
| Austria | Apetlonerhof | AU1 | | 6 | |
| | Eselstall south | AU2 | | 1 | |
| | Illmitzer Damm | AU3 | | 6 | |
| | Martentaulacke north | AU4 | | 1 | |
| | Untere Wiesen | AU5 | | 2 | |
| | Xixsee | AU6 | | 1 | |
| | Zicksee | AU7 | | 4 | |
| Bulgaria | | BU | | 8 | |
| China | Taishan, Guangdong | CH1 | 1 | | AY389012 |
| | Shanxii | CH2 | 1 | | AY389011 |
| | Tongzy | CH3 | 1 | | AB096809 |
| | Shanghai, Shanghai | CH4 | 1 | | AB096815 |
| | Xiji, Ningxia | CH5 | 14 | | AM945740-50, AM945754-55, AM945850 |
| | Gansu | CH6 | 3 | | AM945751-53 |
| | Daqing, Heilongjiang | CH7 | 1 | | KJ082011 |
| | Harbin, Heilongjiang | CH8 | 2 | | KJ082012-13 |
| | Longjiang, | CH9 | 2 | | KJ082014-15 |

| | Heilongjiang | | | | |
|---|---|---|---|---|---|
| | Shenyang, Liaoning | CH10 | 2 | | KJ082006, KJ082008 |
| | Changchun, Jilin | CH11 | 1 | | KJ082007 |
| Croatia | | CRO | 5 | | MH257777, MH257782, MH257861, MH257882-83 |
| Czech Republic | Prilepy, Zlin | CZ1 | 4 | | MH257783, MH257863, MH257884-85 |
| | Kněhyně | CZ2 | | 1 | |
| | Koprivnice | CZ3 | | 1 | |
| | Větřkovice | CZ4 | | 8 | |
| Denmark | Lolland | DA1 | 5 | 9 | MH257786, MH257791-92, MH257864-65 |
| | Filskov | DA2 | | 10 | |
| Estonia | Tallinn | EST | 1 | 1 | AJ311145 |
| Finland | Parikala | FIN1 | 1 | | MH257866 |
| | Rautjarvi | FIN2 | 1 | | MH257793 |
| | Jautseno | FIN3 | 1 | | MH257867 |
| Germany | Lübeck | GE1 | 1 | | AB096817 |
| | Halle | GE2 | 1 | | AF159390 |
| | Göttingen | GE3 | 1 | 2 | MH257801 |
| | Elmenhorst | GE4 | 1 | 3 | MH257785 |
| | Biebersdorf | GE5 | 1 | 2 | MH257851 |
| | Westergellersen | GE6 | 2 | 3 | MH257852, MH257856 |
| | Leina | GE7 | 1 | 4 | MH257853 |

| | Beerwalde | GE8 | 1 | 3 | MH257854 |
|---|---|---|---|---|---|
| | Creuzburg | GE9 | 1 | 3 | MH257855 |
| | Bremerhagen | GE10 | | 3 | |
| | Insel Riems | GE11 | | 2 | |
| | Jasnitz | GE12 | | 3 | |
| | Reinberg | GE13 | | 3 | |
| | Horst | GE14 | | 6 | |
| | Wardenburg | GE15 | | 1 | |
| | Eberswalde | GE16 | | 1 | |
| | Groß Schönebeck | GE17 | | 1 | |
| | Altdöbern | GE18 | | 2 | |
| | Dresden | GE19 | | 2 | |
| | Gotha | GE20 | | 4 | |
| | Liebenburg | GE21 | | 1 | |
| | Lohsa | GE22 | | 1 | |
| | Lucka | GE23 | | 4 | |
| | Welzow | GE24 | | 3 | |
| | Wermsdorfer Forst | GE25 | | 2 | |
| | Wolbrechtshausen | GE26 | | 2 | |
| Hungary | Bak | HU | 2 | 2 | MH257794, MH257859 |
| Italy | / | IT1 | 1 | | AB303226 |
| | Cividale | IT2 | | 1 | |
| Kazakstan | Zharminskii | KAZ1 | 2 | | AM945839, AM945845 |
| | Uighentasskii | KAZ2 | 1 | | AM945846 |
| South Korea | Yong Do | KO1 | 2 | 2 | MH257795-96 |
| | Cheongju | KO2 | 2 | 2 | MH257797-98 |
| | Kanghwa Do | KO3 | 4 | 5 | AM945767-69, |

| | | | | | MH257799 |
|---|---|---|---|---|---|
| | Seoul Grand Park | KO4 | 4 | 3 | MH257800, MH257868-70 |
| | Jeju Island | KO5 | 6 | | KJ081981-86 |
| Lithuania | Alytus | LIT | | 1 | |
| Poland | Pulawy | PO1 | 5 | 7 | MH257871-75 |
| | Lublin | PO2 | | 3 | |
| | Bieszczady-Solina | PO3 | | 11 | |
| | Kujawy | PO4 | | 10 | |
| Romania | Gavodjdia | RO1 | 4 | 5 | MH257802, MH257862, MH257886, MH257891 |
| | Baia Sprie | RO2 | | 1 | |
| | Fenes | RO3 | | 1 | |
| | Iles | RO4 | | 2 | |
| | Coman | RO5 | | 2 | |
| Russia | Evreiskaia Oblast, Birobidzhan City | FE1 | 2 | 2 | MH257803-04 |
| | Chita region, Priargunsky raion | FE2 | 2 | | MH257809-10 |
| | Amurskaja Oblast, vil. Arhara | FE3 | 3 | 2 | MH257811, MH257814-15 |
| | Amur region | FE4 | 1 | 1 | MH257816 |
| | Khabarovskii region | FE5 | 13 | 7 | AM94585-86, AF427332-34, MH257787-90, |

| | | | | | MH257807-08, MH257876-77 |
|---|---|---|---|---|---|
| | Vladivostok | FE6 | 1 | | MH257817 |
| | Primorskii region | FE7 | 24 | 11 | AB032851, MH257805-06, MH257812, MH257818-22, MH257827-31, MH257833-36, MH257878-81, MH257887-88 |
| | Askold Island | FE8 | 2 | 2 | MH257837-38 |
| | Reineke island | FE9 | 2 | 5 | MH257823-24 |
| | Russky island | FE10 | 2 | 5 | MH257825-26 |
| | Putyatin Island | FE11 | | 7 | |
| | Khankaysky | FE12 | | 4 | |
| | Khorolsky | FE13 | | 3 | |
| | Pogranichny | FE14 | | 1 | |
| | Spassky | FE15 | | 7 | |
| | Komsomolsky-na-Amyre | FE16 | | 1 | |
| | Onon river | FE17 | | 1 | |
| | Ussuriysky | FE18 | | 9 | |
| | Ural region | RU1 | 5 | 5 | MH257839-41, MH257857, MH257893 |
| | Belgorod | RU2 | 1 | | MH257842 |
| | Kursk | RU3 | 1 | | MH257843 |
| | Samara | RU4 | 1 | | MH257844 |
| | Ryazan region | RU5 | 1 | | MH257845 |

| | Barnaul region | RU6 | 1 | | AM945838 |
|---|---|---|---|---|---|
| | Krasnodar region, North Caucasus | RU7 | 2 | | MH257846-47 |
| | Omsk, West Siberia | RU8 | 1 | | MH257858 |
| | Novosibirsk, West Siberia | RU9 | 5 | 3 | AM945848, MH257778, MH257848, MH257889-90 |
| | Mikailovica (Volgograd district) | RU10 | 1 | 5 | MH257779 |
| Slovakia | Kosice | SLV1 | 1 | 19 | MH257860 |
| | Busince | SLV2 | | 6 | |
| | Eastern Slovakia | SLV3 | | 8 | |
| | Grajciar | SLV4 | | 1 | |
| | Kechnec | SLV5 | | 2 | |
| | Lesnica | SLV6 | | 1 | |
| | Opatka | SLV7 | | 5 | |
| | Plavec | SLV8 | | 3 | |
| | Podvazie | SLV9 | | 12 | |
| | Radvanovce | SLV10 | | 13 | |
| | Slanec | SLV11 | | 3 | |
| | Velka n/lplom | SLV12 | | 4 | |
| | Viničky | SLV13 | | 2 | |
| Slovenia | Koprivicne | SL1 | 1 | 1 | AJ311144 |
| | Gorjanci | SL2 | 1 | | MH257784 |
| | Prekmurje | SL3 | 4 | | MH257780, MH257813, MH257832, |

| | | | | | MH257892 |
|---|---|---|---|---|---|
| Taiwan | / | TAI1 | 1 | | AB096816 |
| | Hualien | TAI2 | 5 | | KJ082030-34 |
| | Dadushan | TAI3 | 3 | | KJ082035-37 |
| Turkey | | TUR | 3 | 1 | MH257781, MH257849-50 |
| Ukraine | / | UK1 | 1 | | AF445756 |
| | Chernobyl | UK2 | 1 | | AF127535 |

**Table S2**. **Summary statistics used in ABC**. An approximate Bayesian computation analysis requires to summarize observed and simulated data by some numerical indexes (or summary statistics) that are, in the ideal case, informative about the underlying processes that we wish to infer. In the present work we chose summary statistics (Table S1) commonly used in population genetics that are known informative about the population size within demes (genetic diversity indexes such as heterozygosity, number of alleles and variance in allele size), genetic differentiation among demes (genetic distances between populations, $F_{ST}$ estimators and analogous statistics), and statistics informative about population size changes (Garza and Williamson statistic, Kimmel's β and the statistic used in the popular software "Bottleneck" (Cornuet et al. 1999)). In addition we introduced an additional summary statistics sensitive to population size changes. Characterization of population size changes through summary statistics is, in most cases, based on the comparisons of two genetic diversity indexes whose relative values follow some theoretically known relationship for population in equilibrium (i.e. constant size populations). Kimmel's β is based on a disequilibrium between heterozygosity and allele size variance (Kimmel et al. 1998) and Bottleneck statistic compares number of alleles and heterozygosity (Cornuet and Luikart 1996). Our new statistic β' is the ratio between an estimator of θ=4$N$μ based on the allele size variance and an estimator θ based on the number of alleles and allele size variance (β'=$\theta_V/\theta_{Na}$), following the same approach as (Kimmel et al. 1998). These estimators are based on the expected values under a stepwise mutation model (SMM) and have been described by (Kimmel and Chakraborty 1996) as $\theta_V$=2$V$ (where $V$ is the variance in allele size) and (Ohta and Kimura 1973) as $\theta_{Na}$=($Na^2$-1)/2 (where $Na$ is the number of alleles). In an equilibrium population values should be around 1 and departures from this values would indicate changes in population size (for loci mutating under the SMM).

| Summary statistic | Population | R package | Reference |
|---|---|---|---|
| Heterozygosity[*] | Eastern, Western, Total | pegas | (Nei 1987) |
| Number of alleles[*] | Eastern, Western, Total | | |
| Number of private alleles[*] | Eastern, Western | | |
| Number of shared alleles[*] | Eastern & Western | | |
| Variance in allele size[*] | Eastern, Western, Total | | (Kimmel and Chakraborty 1996) |
| Range of allele size[*] | Eastern, Western, Total | | |
| Garza and Williamson statistic[*] | Eastern, Western, Total | | (Garza and Williamson 2001) |
| Kimmel's β[*] | Eastern, Western, Total | | (Kimmel et al. 1998) |
| β'[*] | Eastern, Western, Total | | this work (see text) |
| "Bottleneck" statistic[*] | Eastern, Western, Total | | (Cornuet and Luikart 1996) |
| (δμ)[2*] | Eastern & Western | | (Goldstein et al. 1995) |
| $G_{ST}$[*$] | Eastern & Western | mmod | (Nei 1973) |
| $G''_{ST}$[*$] | Eastern & Western | mmod | (Meirmans and Hedrick |

| | | | 2011) |
|---|---|---|---|
| Jost's D$^{*\$\#}$ | Eastern & Western | mmod | (Jost 2008) |
| $\Phi'_{ST}$$^{*\$}$ | Eastern & Western | mmod | (Meirmans 2006) |
| $F_{ST}$$^{*\$}$ | Eastern & Western | pegas | (Weir and Cockerham 1984) |
| Neis's distance$^{\$}$ | Eastern & Western | adegenet | (Nei 1972) |
| Edwards' distance$^{\$}$ | Eastern & Western | adegenet | (Edwards 1971) |
| Reynolds' distance$^{\$}$ | Eastern & Western | adegenet | (Reynolds et al. 1983) |
| Rogers' distance$^{\$}$ | Eastern & Western | adegenet | (Rogers 1972) |
| Prevosti's distance$^{\$}$ | Eastern & Western | adegenet | (Prevosti et al. 1975) |

[*] mean and variance among loci
[\$] multilocus
[#] harmonic mean among loci

**Table S3**. Haplotype definitions

| Haplotype | N | Locality | Genbank accession numbers |
|---|---|---|---|
| HCYTB1 | 1 | CH1 | AY389012 |
| HCYTB2 | 1 | CH2 | AY389011 |
| HCYTB3 | 1 | CH3 | AB096809 |
| HCYTB4 | 1 | CH4 | AB096815 |
| HCYTB5 | 3 | CH5 – CH6 | AM945746 – AM945752 – AM945850 |
| HCYTB6 | 2 | CH5 | AM945748 – AM945749 |
| HCYTB7 | 2 | CH5 | AM945747 – AM945750 |
| HCYTB8 | 4 | CH5 | AM945741 – AM945742 – AM945745 – AM945755 |
| HCYTB9 | 1 | CH5 | AM945743 |
| HCYTB10 | 2 | CH5 | AM945740 – AM945754 |
| HCYTB11 | 1 | CH5 | AM945744 |
| HCYTB12 | 1 | CH6 | AM945751 |
| HCYTB13 | 1 | CH6 | AM945753 |
| HCYTB14 | 3 | CRO – UK1 – UK2 | MH257882 – AF445756 – AF127535 |
| HCYTB15 | 1 | CRO | MH257777 |
| HCYTB16 | 10 | CRO – EST – GE2 – RU10 – SLV1 – SL3 - TUR | MH257883 - AJ311145 - AF159390 – MH257778 – MH257779 – MH257860 – MH257780 – MH257781 |
| HCYTB17 | 1 | CRO | MH257861 |
| HCYTB18 | 1 | CRO | MH257782 |
| HCYTB19 | 6 | CZ1 – GE4 – RO1 – SL2 | MH257783 – MH257884 – MH257885 – MH257862 – MH257784 – MH257785 |
| HCYTB20 | 1 | CZ1 | MH257863 |
| HCYTB21 | 2 | DA1 | MH257786 – MH257864 |
| HCYTB22 | 5 | DA1 – FE5 | MH257865 – MH257787 – MH257788 – MH257789 – MH257790 |
| HCYTB23 | 1 | DA1 | MH257791 |
| HCYTB24 | 1 | DA1 | MH257792 |
| HCYTB25 | 3 | FIN1 – FIN2 – FIN3 | MH257866 – MH257793 – MH257867 |
| HCYTB26 | 1 | GE1 | AB096817 |
| HCYTB27 | 1 | HU | MH257794 |
| HCYTB28 | 1 | KAZ1 | AM945839 |

| HCYTB29 | 1 | KAZ1 | AM945845 |
|---|---|---|---|
| HCYTB30 | 1 | KAZ2 | AM945846 |
| HCYTB31 | 2 | KO1 | MH257795 – MH257796 |
| HCYTB32 | 1 | KO2 | MH257797 |
| HCYTB33 | 1 | KO2 | MH257798 |
| HCYTB34 | 1 | KO3 | MH257799 |
| HCYTB35 | 2 | KO4 | MH257868 – MH257800 |
| HCYTB36 | 1 | KO4 | MH257869 |
| HCYTB37 | 1 | KO4 | MH257870 |
| HCYTB38 | 1 | KO3 | AM945767 |
| HCYTB39 | 2 | KO3 | AM945768 – AM945769 |
| HCYTB40 | 4 | GE3 – PO1 | MH257871 – MH257872 – MH257873 – MH257801 |
| HCYTB41 | 2 | PO1 | MH257874 |
| HCYTB42 | 1 | PO1 | MH257875 |
| HCYTB43 | 1 | RO1 | MH257886 |
| HCYTB44 | 1 | RO1 | MH257891 |
| HCYTB45 | 1 | RO1 | MH257802 |
| HCYTB46 | 7 | FE1 – FE5 – FE7 | MH257803 - MH257804 – MH257805 – MH257806 – AM945856 – AF427333 – AF427332 |
| HCYTB47 | 1 | FE5 | MH257807 |
| HCYTB48 | 1 | FE5 | MH257808 |
| HCYTB49 | 1 | HU | MH257859 |
| HCYTB50 | 2 | FE5 | AM945855 – AF427334 |
| HCYTB51 | 1 | FE5 | MH257877 |
| HCYTB52 | 1 | FE5 | MH257876 |
| HCYTB53 | 1 | FE2 | MH257809 |
| HCYTB54 | 5 | FE2 – FE3 – FE7 – SL3 | MH257810 – MH257811 – MH257878 – MH257812 – MH257813 |
| HCYTB55 | 2 | FE3 – FE7 | MH257814 - MH257879 |
| HCYTB56 | 1 | FE3 | MH257815 |
| HCYTB57 | 1 | FE4 | MH257816 |
| HCYTB58 | 1 | FE6 | MH257817 |
| HCYTB59 | 1 | FE7 | MH257818 |
| HCYTB60 | 2 | FE7 | MH257819 |
| HCYTB61 | 2 | FE7 | MH257887 - MH257880 |
| HCYTB62 | 1 | FE7 | MH257881 |
| HCYTB63 | 1 | FE7 | MH257820 |
| HCYTB64 | 1 | FE7 | MH257821 |
| HCYTB65 | 5 | FE1 – FE7 – FE9 | MH257822 – MH257823 – MH257824 – MH257825 – MH257826 |

| HCYTB66 | 1 | FE7 | MH257827 |
|---|---|---|---|
| HCYTB67 | 1 | FE7 | MH257888 |
| HCYTB68 | 1 | FE7 | MH257828 |
| HCYTB69 | 4 | FE7 – SL3 | MH257829 – MH257830 – MH257831 – MH257832 |
| HCYTB70 | 1 | FE7 | MH257833 |
| HCYTB71 | 1 | FE7 | MH257834 |
| HCYTB72 | 1 | FE7 | MH257835 |
| HCYTB73 | 1 | FE7 | MH257836 |
| HCYTB74 | 2 | FE7 – SL1 | AB032851 – AJ311144 |
| HCYTB75 | 2 | FE8 | MH257837 – MH257838 |
| HCYTB76 | 1 | RU1 | MH257839 |
| HCYTB77 | 3 | RU1 | MH257840 – MH257893 – MH257841 |
| HCYTB78 | 1 | RU2 | MH257842 |
| HCYTB79 | 1 | RU3 | MH257843 |
| HCYTB80 | 1 | RU4 | MH257844 |
| HCYTB81 | 1 | RU5 | MH257845 |
| HCYTB82 | 1 | RU6 | AM945838 |
| HCYTB83 | 1 | RU7 | MH257846 |
| HCYTB84 | 1 | RU7 | MH257847 |
| HCYTB85 | 1 | RU9 | AM945848 |
| HCYTB86 | 1 | RU9 | MH257848 |
| HCYTB87 | 1 | SL3 | MH257892 |
| HCYTB88 | 2 | TAI1 – TAI3 | AB096816 – KJ082035 |
| HCYTB89 | 1 | TUR | MH257849 |
| HCYTB90 | 1 | TUR | MH257850 |
| HCYTB91 | 1 | GE5 | MH257851 |
| HCYTB92 | 1 | GE6 | MH257852 |
| HCYTB93 | 1 | GE7 | MH257853 |
| HCYTB94 | 1 | GE8 | MH257854 |
| HCYTB95 | 1 | GE9 | MH257855 |
| HCYTB96 | 1 | GE6 | MH257856 |
| HCYTB97 | 1 | IT1 | AB303226 |
| HCYTB98 | 1 | TAI2 | KJ082030 |
| HCYTB99 | 2 | TAI2 | KJ082031 – KJ082032 |
| HCYTB100 | 1 | TAI2 | KJ082033 |
| HCYTB101 | 1 | TAI2 | KJ082034 |
| HCYTB102 | 1 | TAI3 | KJ082036 |
| HCYTB103 | 1 | TAI3 | KJ082037 |
| HCYTB104 | 1 | KO5 | KJ081981 |
| HCYTB105 | 1 | KO5 | KJ081982 |
| HCYTB106 | 1 | KO5 | KJ081983 |

| HCYTB107 | 1 | KO5 | KJ081984 |
| HCYTB108 | 1 | KO5 | KJ081985 |
| HCYTB109 | 1 | KO5 | KJ081986 |
| HCYTB110 | 1 | CH10 | KJ082006 |
| HCYTB111 | 1 | CH10 | KJ082008 |
| HCYTB112 | 1 | CH11 | KJ082007 |
| HCYTB113 | 1 | CH7 | KJ082011 |
| HCYTB114 | 1 | CH8 | KJ082012 |
| HCYTB115 | 1 | CH8 | KJ082013 |
| HCYTB116 | 1 | CH9 | KJ082014 |
| HCYTB117 | 1 | CH9 | KJ082015 |
| HCYTB118 | 1 | RU9 | MH257889 |
| HCYTB119 | 1 | RU9 | MH257890 |
| HCYTB120 | 1 | RU1 | MH257857 |
| HCYTB121 | 1 | RU8 | MH257858 |

**Figure S1**. ML phylogenetic tree including all *A. agrarius* haplotypes

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
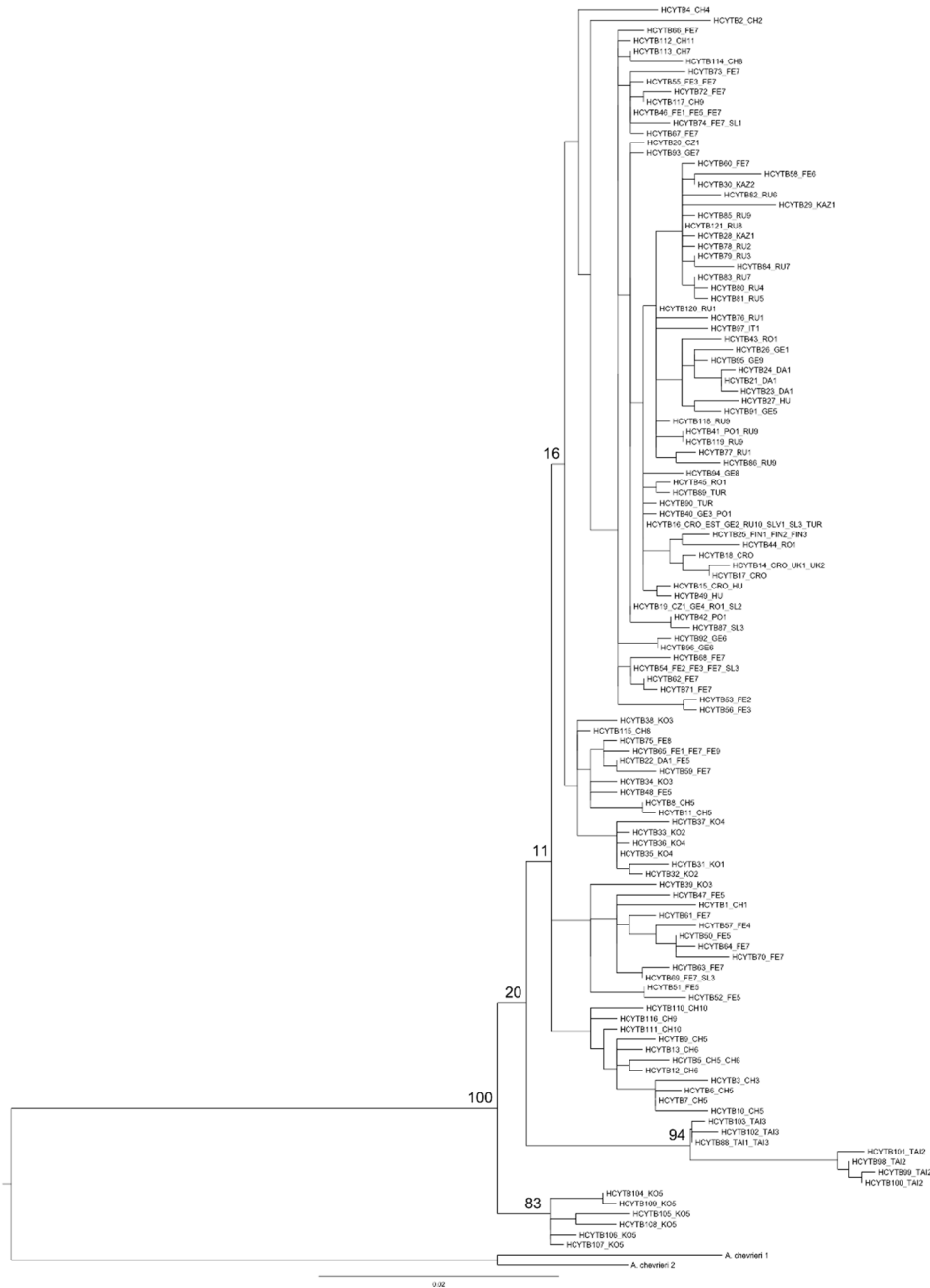42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Figure S2**. Results of the Bayesian clustering analysis with STRUCTURE. (A) Plot of the likelihood of the mean Ln Pr(X|K). (B) Values of ΔK calculated according to Evanno *et al.* (2005)

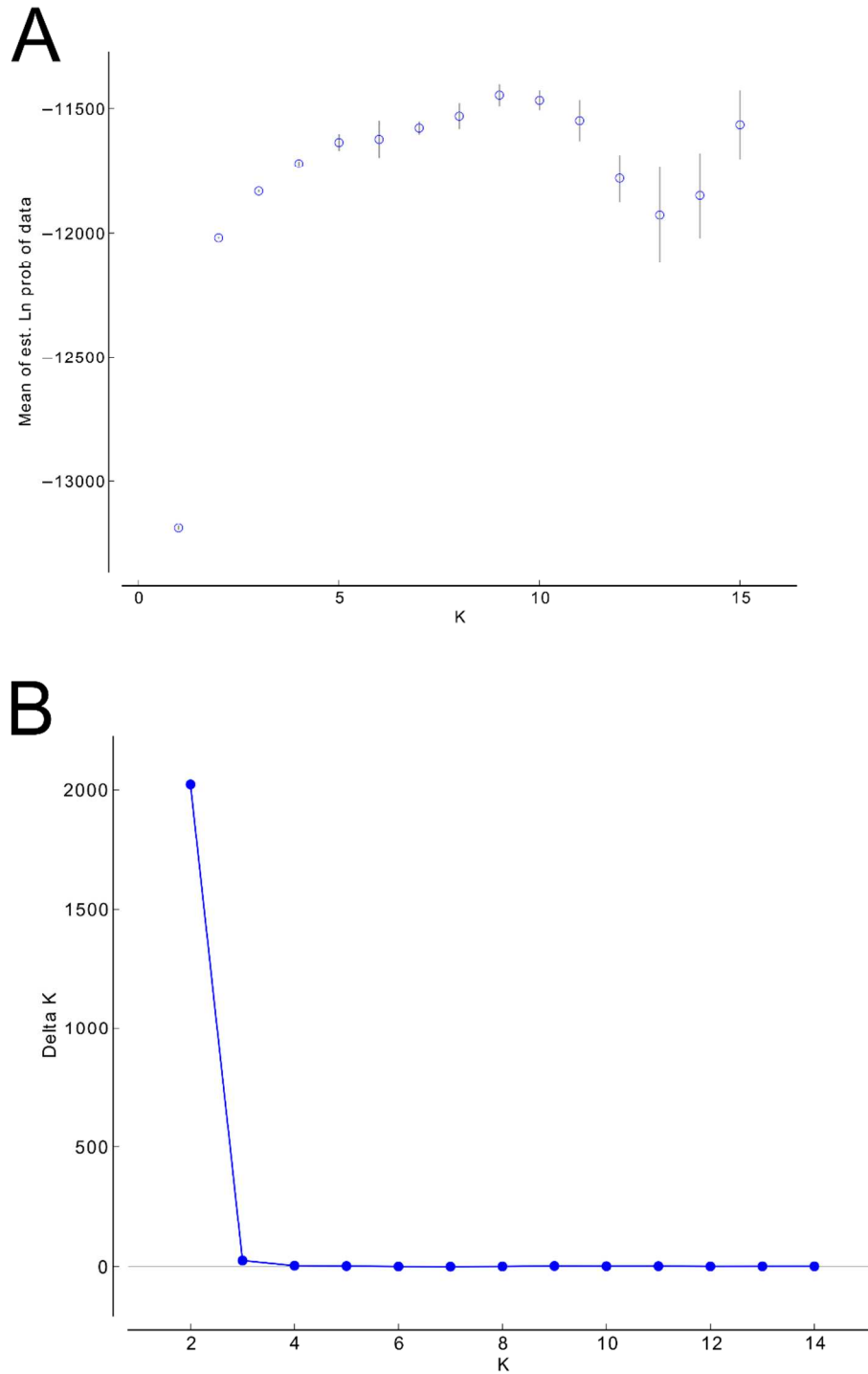**Figure S3**. Population structure estimated using STRUCTURE (K=1 to 10). Each individual is represented by a vertical line partitioned into K colour segments, the length of each colour segment being proportional to the estimated membership coefficient. Numbers correspond to sampling countries: 1 = Germany, 2 = Austria, 3 = Bulgaria, 4 = Denmark, 5 = Estonia, 6 = Hungary, 7 = Italy, 8 = Korea, 9 = Lithuania, 10 = Poland, 11 = Romania, 12 = Russian Far East, 13 = Russia (Central Russia + Western Siberia), 14 = Slovakia, 15 = Slovenia, 16 = Czech Republic, 17 = Turkey
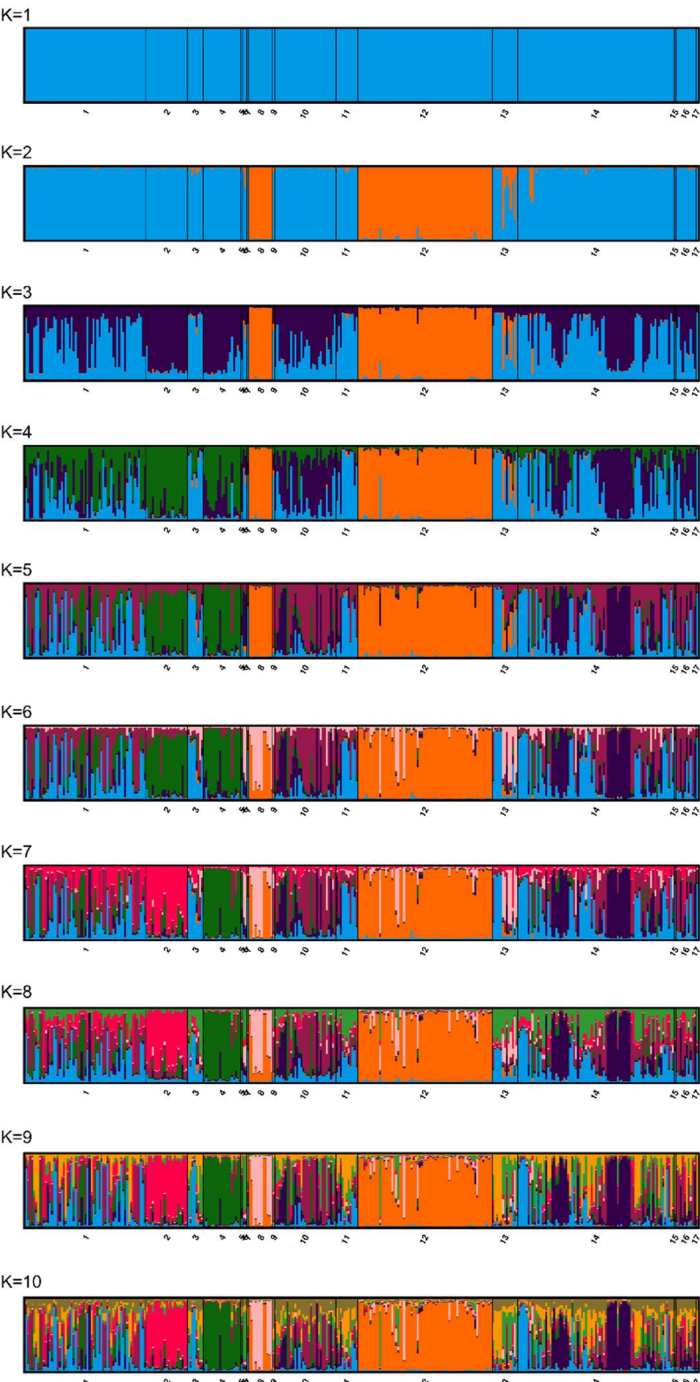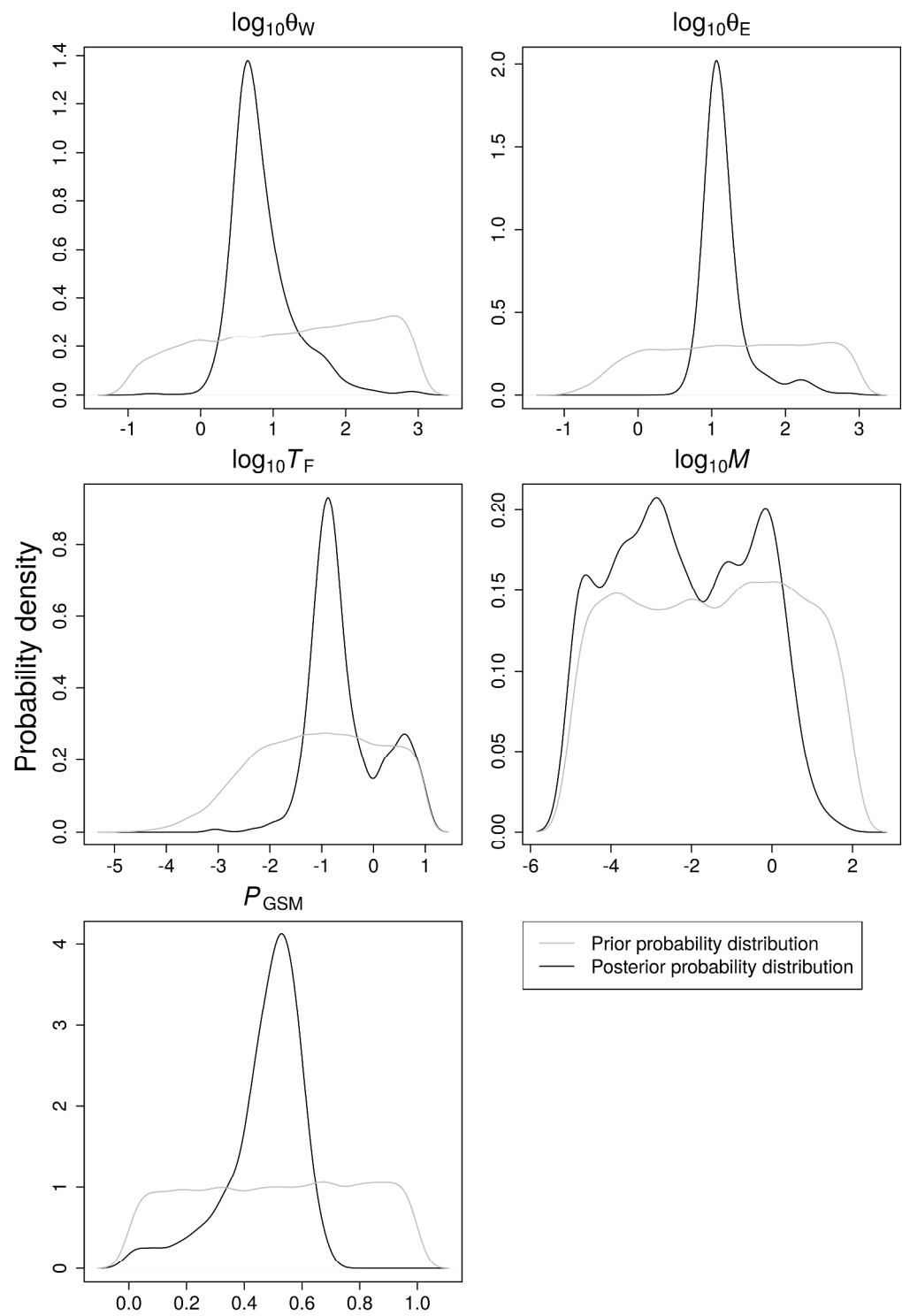
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Figure S4**. Prior and posterior probability densities, for the demographic and mutational parameters of the model.

**References**

Cornuet, J. M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014.

Cornuet, J.-M., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153:1989–2000.

Edwards, A. W. F. 1971. Distances between Populations on the Basis of Gene Frequencies. Biometrics 27:873.

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

Garza, J. C., and E. G. Williamson. 2001. Detection of reduction in population size using data from microsatellite loci. Mol. Ecol. 10:305–318.

Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. 92:6723–6727.

Jost, L. 2008. GST and its relatives do not measure differentiation. Mol. Ecol. 17:4015–4026.

Kimmel, M., and R. Chakraborty. 1996. Measures of variation at DNA repeat loci under a general stepwise mutation model. Theor. Popul. Biol. 50:345–367.

Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins, and L. B. Jorde. 1998. Signatures of population expansion in microsatellite repeat data. Genetics 148:1921–1930.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60:2399–2402.

Meirmans, P. G., and P. W. Hedrick. 2011. Assessing population structure: FST and related measures. Mol. Ecol. Resour. 11:5–18.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. 70:3321–3323.

Nei, M. 1972. Genetic distance between populations. Am. Nat. 106:283–292.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Ohta, T., and M. Kimura. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. 22:201.

Prevosti, A., J. Ocaña, and G. Alonso. 1975. Distances between populations of Drosophila subobscura, based on chromosome arrangement frequencies. Theor. Appl. Genet. 45:231–241.

Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779.

Rogers, J. S. 1972. Measures of genetic similarity and genetic distances. Pp. 145–153 *in* Studies in Genetics. University of Texas Publications.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.