Molecular Phylogenetics and Evolution xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene

Sylvain Dubey [a,*], Johan Michaux [b], Harald Brünner [c], Rainer Hutterer [d], Peter Vogel [e]

[a] School of Biological Sciences, University of Sydney, Sydney, NSW 2006, Australia
[b] Unité de Recherches Zoogéographiques, Institut de Zoologie, Quai Van Beneden, 22, 4020 Liège, Belgium
[c] Hohenwettersbacher Strasse 10, D-76228 Karlsruhe, Germany
[d] Section of Mammals, Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, D-53113 Bonn, Germany
[e] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 8 October 2008
Revised 27 November 2008
Accepted 8 December 2008
Available online xxxx

Keywords:
Cytochrome *b* gene
Pseudogene
Apodemus
Phylogeography
Phylogeny

ABSTRACT

The phylogeny and phylogeography of the Old World wood mice (subgenus *Sylvaemus*, genus *Apodemus*, Muridae) are well-documented. Nevertheless, the distributions of species, such as *A. fulvipectus* and *A. ponticus* remain dubious, as well as their phylogenetic relationships with *A. sylvaticus*. We analysed samples of *Apodemus* spp. across Europe using the mitochondrial cytochrome-*b* gene (*cyt-b*) and compared the DNA and amino-acid compositions of previously published sequences. The main result stemming from this study is the presence of a well-differentiated lineage of *Sylvaemus* including samples of various species (*A. sylvaticus*, *A. fulvipectus*, *A. ponticus*) from distant locations, which were revealed to be nuclear copies of the mitochondrial *cyt-b*. The presence of this cryptic pseudogene in published sequences is supported by different pathways. This has led to important errors in previous molecular trees and hence to partial misinterpretations in the phylogeny of *Apodemus*.

1. Introduction

The phylogeny and phylogeography of the Old World wood mice (subgenus *Sylvaemus*, genus *Apodemus,* Muridae) are well-documented in numerous publications involving the mitochondrial cytochrome-*b* gene (*cyt-b*; e.g. Martin et al., 2000; Hille et al., 2002; Michaux et al., 2002, 2003, 2004, 2005; Reutter et al., 2003; Balakirev et al., 2007; Hoofer et al., 2007; Suzuki et al., 2008) or both mitochondrial and nuclear genes (e.g. Michaux et al., 2002, 2005; Suzuki et al., 2008). The western Eurasian species of the subgenus *Sylvaemus*, i.e. the alpine mouse *A. alpicola*, the wood mouse *A. sylvaticus*, the yellow-necked mouse *A. flavicollis*, the pygmy woodmouse (*A. uralensis*), the yellow-breasted mouse *A. witherbyi,* the Mt Hermon mouse *A. hermonensis* (a synonym of *A. witherbyi;* see Musser and Carleton, 2005) and the Caucasus mouse *A. ponticus* have been revealed to be genetically closely related (Michaux et al., 2002; Suzuki et al., 2008).

Within the species, the phylogeography of *A. flavicollis*, based on the *cyt-b* (Michaux et al., 2004) showed two well differentiated clades, a first in Turkey, Syria, Israel, and Iran (Southern of the Caucasus), and a second including western, central, and eastern Europe, and Russia (Fig. 1). For *A. alpicola*, a species restricted to the alpine region, Michaux et al. (2002) and Reutter et al. (2003) documented its monophyly. A thorough phylogeography of *A. sylvaticus*, based on the *cyt-b* (Michaux et al., 2003), highlighted a clear pattern with slightly divergent sub-clades involving different geographic areas such as (i) western, northern and central Europe, (ii) Italy and Balkans, or (iii) Sicily. Interestingly, Reutter et al. (2003), based on restricted sampling, found two additional well differentiated sequences of *A. sylvaticus* (G3.4 and G3.1), both from Karlsruhe in the southern part of Germany, from animals clearly identified as *A. sylvaticus* in the field. The first one (G3.4) was considered to belong to a lineage that separated very early, before the split separating the Italian animals (Michaux et al., 2003) and animals from the Pyreneans to the Ukraine. The authors were not able to explain the syntropic occurrence of these two lineages. The second one (G3.1) revealed to be closely related to some sequences ascribed to *A. fulvipectus* (but the identity is still uncertain) and *A. ponticus* from Georgia (Hille et al., 2002) and to a unique sequence of *A. sylvaticus* from Konstanz, Germany (Martin et al., 2000), a locality 150 km away from Karlsruhe. Both tissue samples were from biopsies of released animals without preserved voucher specimens. Reutter et al. (2003) suggested that an unrecognized population of *A. fulvipectus* might occur in southern Germany. As these sequences were fully coding the authors excluded the presence of a nuclear pseudogene and refuted the possibility of a DNA contamination in their laboratory, as they had never handled samples of *A. fulvipectus* from the Caucasus before. Moreover, the situation in Georgia, as studied by Hille et al. (2002), was all but clear. The assignment of some samples by morphology was not concordant

* Corresponding author. Fax: +61 2 9351 5609.
E-mail address: sylvain.dubey@bio.usyd.edu.au (S. Dubey).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.12.008

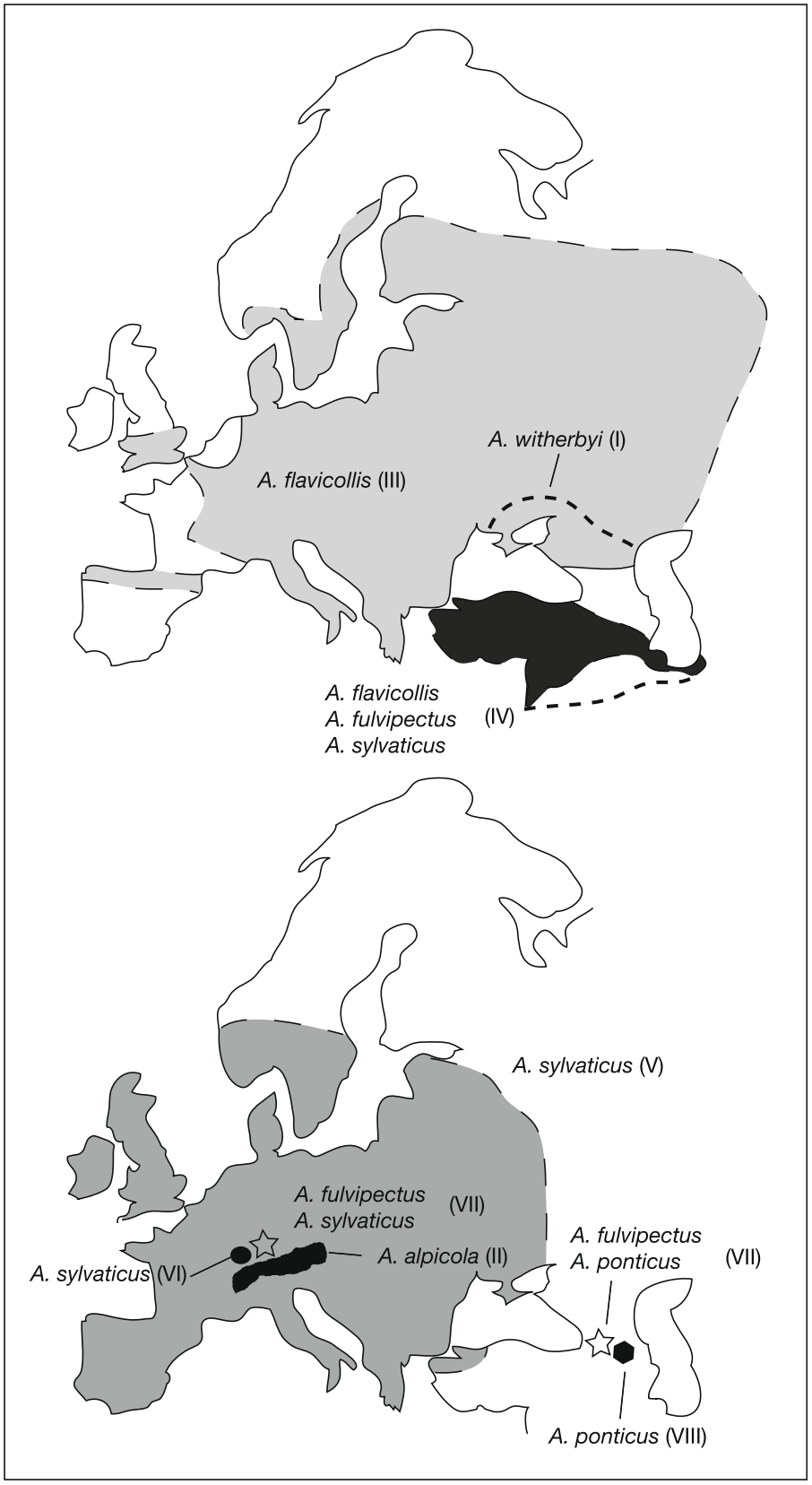

Fig. 1. Distribution of *Apodemus* spp. in Europe and the Near East, according to previous publications based on molecular phylogenetic studies and corresponding lineage (I–VIII) of the present study. Areas of lineages I–V are shaded. The locations of the samples of lineage VI and VII shown by dot, stars or polygon.

with the genetic assignment (*cyt-b*), especially the assignment of *A. ponticus* and *A. flavicollis* from the same locality appeared often problematic. Finally, a recent phylogenetic study (Suzuki et al., 2008) including some of these sequences yielded similar results, which led the authors to the conclusion that the systematics of these species needs to be revised.

Please cite this article in press as: Dubey, S. et al., False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene, Mol. Phylogenet. Evol. (2009), doi:10.1016/j.ympev.2008.12.008

In the present study, we therefore analysed 102 *cyt-b* sequences of different species and lineages of *Apodemus*, and compared their DNA and amino-acid compositions in order to explain the presence of three divergent *cyt-b* lineages of *A. sylvaticus* in Germany, i.e. (i) the lineage widely distributed in central and northern Europe (Michaux et al. 2003), (ii) the unexpected lineage (G3.4) found by Reutter et al. (2003), and (iii) the lineage including samples of *A. fulvipectus* from Germany (Martin et al., 2000; Reutter et al., 2003: G3.1), and *A. fulvipectus* and *A. ponticus* from Georgia (Hille et al., 2002).

2. Materials and methods

2.1. Samples

We analyzed 102 samples of the *Apodemus* spp. collected in Europe (Table 1). We used *Mus musculus* (AB205312) as an outgroup. The set of samples included material from the collections of JR Michaux (JM) located at the University of Liège, Belgium, and from the Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany (ZFMK). Specimens labelled with "HB" were biopsies without voucher specimens. Additional sequences (Table 1) were taken from GenBank (Martin et al., 2000; Hille et al., 2002; Fillipucci et al., 2002; Reutter et al., 2003; Michaux et al., 2003, 2004).

2.2. DNA analysis

The DNA extraction was carried out using the QIA Amp DNA Mini Kit (Qiagen). Double-stranded DNA amplifications of the cytochrome *b* gene (*cyt-b*) were performed using the primer pair CB-AF/CB-AR2 (Reutter et al., 2003), L7/H16 (Michaux et al., 2003). In some samples, the resulting electropherogram of *cyt-b* sequences were ambiguous, with the presence of numerous double peaks, in samples of *A. sylvaticus* from Germany, Sweden or Ireland (see Fig. 2 for example), suggesting the co-amplification of e.g. mitochondrial nuclear copies (pseudogenes). Based on these results, we designed a primer specific to the additional co-amplified sequences, Lpseudo (5′-TTTGGTTCTCTACTAGGAATT-3′), which allows amplification of the pseudogene, coupled with the primer H16.

Amplification conditions consisted of 35 thermal cycles of denaturation at 94 °C for 45 s, annealing at 50 °C for 45 s, and extension at 72 °C for 90 s.

The PCR products were checked on a 1% agarose gel and then purified using the QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instructions. DNA sequencing was performed in a total volume of 10 μl containing 1–3 μl of amplified PCR product, 1 μl of 10 μM primer, and 4 μl of ABI PRISM™ Dye Terminator 1 (Perkin-Elmer). Sequence reactions were visualized on an ABI 3100 genetic analyzer (Applied Biosystems, USA).

2.3. Phylogenetic methods

Nucleotide sequences of *cyt-b* were edited using Sequence Navigator (Parker, 1997) and manually aligned. Parsimony analyses (MP) were performed using the following options: heuristic search, stepwise addition of sequences, 200 replicates of random additions of taxa, and TBR branch swapping, using PAUP* version 4.0b10 PPC (Swofford, 2001). MP bootstrap support values were obtained with 1000 pseudo-replicates. Tests were conducted on the complete fragment, all codon positions were used, and trees were rooted using a sequence from *Mus musculus* (AB205312; Terashima et al., 2006). Fast maximum likelihood (ML) heuristic searches and bootstrap analyses (1000 replicates) were performed using PHYML (Guindon and Gascuel, 2003) with a GTR + I + G model (Rodriguez et al. 1990), which had been selected previously using Modeltest 3.06 according to the protocol of Posada and Crandall (1998), with base frequencies estimated from the data (A = 0.31697, C = 0.28192, G = 0.11727, T = 0.28384), an unequal distribution of substitution rates at variable sites (a = 0.596), a proportion of invariant sites (0.381), and six different substitution types (rate [A – C] = 2.20094, rate [A–G] = 16.38917, [C – T] = 26.73843, rate [A – T] = 3.58426, rate [C – G] = 1.72395, rate [G – T] = 1.0).

Bayesian analyses (BA) were run with MrBayes version 3.1.2.1 (Huelsenbeck et al., 2001) using a GTR model. Two independent runs were performed, each consisting of four parallel MCMC chains of 3 million generations, allowing a good convergence of the independent runs. Tree parameters reached stationarity after a burn-in period of 600,000 generations. Optimal trees were then sampled every 100 generations to obtain the final consensus BA tree and associated posterior probabilities.

3. Results

3.1. Molecular phylogeny

We found 80 different haplotypes of 818 bp within the 102 analysed sequences, including 277 variable sites, of which 209 were parsimony informative. We did not find any insertions or deletions. In addition, no stop codons were found. Because the three phylogenetic methods yielded identical arrangements of the main branches, we show the relationship between haplotypes only for the ML analysis (Fig. 3).

Our phylogenetic analyses revealed seven different lineages, which included:

(I) Samples of *A. witherbyi* from Turkey (Michaux et al., 2002, as *hermonensis*).

(II) Samples of *A. alpicola* from Switzerland (Serizawa et al., 2000; Reutter et al., 2003).

(III) Samples of *A. flavicollis* from France, Germany, European Turkey, Greece, Italy, Slovenia, and Russia (Michaux et al., 2004; Reutter et al., 2003).

(IV) Samples of *A. flavicollis* from Turkey (Michaux et al., 2004), *A. sylvaticus* from Russia (Dekonenko et al., 2003), and *A. fulvipectus* from Georgia (new material and Hille et al., 2002). This last sample is not included in the molecular phylogenetic analyses as the sequence was short (246 bp). All these samples are genetically close relatives and should be considered as *A. flavicollis* of the "Near East" lineage (see detailed phylogeographic study of Michaux et al., 2004).

(V) Our new samples of *A. sylvaticus* from various locations, sequenced using classical primers to amplify the mammal mitochondrial *cyt-b*, e.g. from Germany, Sweden, France, Ireland, Belgium, Netherlands, Czech Republic, Yugoslavia, Italy, Slovenia, Greece, as well as published sequences of *A. sylvaticus* from Switzerland and Germany from Reutter et al. (2003; G3.3 and S1.5), corresponding to the widespread clade of *A. sylvaticus* (Michaux et al., 2003).

(VI) One samples of *A. sylvaticus* from Germany (G3.4, Reutter et al., 2003).

(VII) Samples of *A. ponticus* from Georgia (Hille et al., 2002; same locality as lineage V, not shown), *A. fulvipectus* from Georgia (Hille et al., 2002, not shown), *A. sylvaticus* from Germany (Martin et al., 2000; Reutter et al., 2003, G3.1), and our new samples of *A. sylvaticus* individuals coming from Sweden, Denmark, Ireland, Netherland and Germany, but which were amplified this time using specific primers for mitochondrial nuclear copies (primers Lpseudo/H16, see Section 2).

Please cite this article in press as: Dubey, S. et al., False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene, Mol. Phylogenet. Evol. (2009), doi:10.1016/j.ympev.2008.12.008

Table 1
Species (according to original assignment), collection code, location, GenBank accession number with reference and lineage assignment resulting from this study (as shown in Fig. 2).

Species label	Collection code	Location	GenBank	Lineage
A. fulvipectus		Georgia	AF249765, Hille et al. (2002)	IV
A. fulvipectus	ZFMK 2004.018	Georgia	FJ389657, this study	IV
A. fulvipectus	ZFMK 2004.019	Georgia	FJ389658, this study	IV
A. fulvipectus	ZFMK 2004.020	Georgia	FJ389659, this study	IV
A. fulvipectus		Georgia	AF249762, Hille et al. (2002)	VII
A. fulvipectus		Georgia	AF249763, Hille et al. (2002)	VII
A. fulvipectus		Georgia	AF249764, Hille et al. (2002)	VII
A. fulvipectus	G3.1 Karlsruhe	Germany	AY179491, Reutter et al. (2003)	VII
A. alpicola		Austria	ABO32854, Serizawa et al. (2000)	II
A. alpicola		Switzerland	AF159391, Martin et al. (2000)	II
A. alpicola		Switzerland	AY179494, Reutter et al. (2003)	II
A. flavicollis		France	AJ298602, Michaux et al. (2001)	III
A. flavicollis		Germany	AY179498, Reutter et al. (2003)	III
A. flavicollis	ZFMK2008.312 (671, Karlsruhe)	Germany	FJ389660, this study	III
A. flavicollis	ZFMK2008.313 (674, Karlsruhe)	Germany	FJ389661, this study	III
A. flavicollis		Greece	AJ605630, Michaux et al. (2004)	III
A. flavicollis		Italy	AJ605635, Michaux et al. (2004)	III
A. flavicollis		Russia	AJ605654, Michaux et al. (2004)	III
A. flavicollis		Slovenia	AJ605657, Michaux et al. (2004)	III
A. flavicollis		Turkey	AJ605673, Michaux et al. (2004)	III
A. flavicollis		Turkey	AJ605677, Michaux et al. (2004)	IV
A. hermonensis		Turkey	AJ311156, Michaux et al. (2002)	I
A. hermonensis		Turkey	AJ311157, Michaux et al. (2002)	I
A. ponticus		Georgia	AF249767, Hille et al. (2002)	VII
A. ponticus		Georgia	AF249768, Hille et al. (2002)	VII
A. ponticus		Georgia	AF249766, Hille et al. (2002)	VIII
A. sylvaticus	JM212	"Yugoslavia"	AJ511941, this study	V
A. sylvaticus	JM105	Belgium	AJ511878, this study	V
A. sylvaticus	JM107	Belgium	AJ511879, this study	V
A. sylvaticus	JM373	Czech Rep.	AJ511889, this study	V
A. sylvaticus	JM374	Czech Rep.	AJ511890, this study	V
A. sylvaticus	JM574	France	FJ389652, this study	V
A. sylvaticus	JM575	France	FJ389653, this study	V
A. sylvaticus	HB706, Karlsruhe	Germany	FJ389656, this study	V
A. sylvaticus	HBG3.3, Karlsruhe	Germany	AY180339, Reutter et al. (2003)	V
A. sylvaticus	JM181	Greece	AJ511940, this study	V
A. sylvaticus	JM1235	Ireland	FJ389650, this study	V
A. sylvaticus	JM1239	Ireland	FJ389651, this study	V
A. sylvaticus	JM160	Italy	AJ511923, this study	V
A. sylvaticus	JM162	Italy	AJ511924, this study	V
A. sylvaticus		Netherlands	ABO33695, Suzuki et al. (2000)	V
A. sylvaticus	JM304	Sicily/Italy	AJ511960, this study	V
A. sylvaticus	JM305	Sicily/Italy	AJ511959, this study	V
A. sylvaticus	JM417	Slovenia	AJ511931, this study	V
A. sylvaticus	JM434	Slovenia	AJ511932, this study	V
A. sylvaticus	JM132	Spain	AJ511881, this study	V
A. sylvaticus	JM118	Sweden	FJ389649, this study	V
A. sylvaticus	JM557	Sweden	FJ389648, this study	V
A. sylvaticus	JM1196	Sweden	FJ389664, this study	V
A. sylvaticus	S1.5	Switzerland	AY179493, Reutter et al. (2003)	V
A. sylvaticus	JM1193	Denmark	FJ389639, FJ389601, this study	V/VII
A. sylvaticus	JM1195	Denmark	FJ389622, FJ389603, this study	V/VII
A. sylvaticus	JM1196	Denmark	FJ389664, FJ389662, this study	V/VII
A. sylvaticus	JM1200	Denmark	FJ389621, FJ389589, this study	V/VII
A. sylvaticus	JM1201	Denmark	FJ389640, FJ389594, this study	V/VII
A. sylvaticus	JM1187	Ireland	FJ389642, FJ389602, this study	V/VII
A. sylvaticus	JM1188	Ireland	FJ389645, FJ389584, this study	V/VII
A. sylvaticus	JM1190	Ireland	FJ389646, FJ389604, this study	V/VII
A. sylvaticus	JM1197	Ireland	FJ389647, FJ389585, this study	V/VII
A. sylvaticus	JM1198	Ireland	FJ389666, FJ389587, this study	V/VII
A. sylvaticus	JM1199	Ireland	FJ389643, FJ389586, this study	V/VII
A. sylvaticus	JM1233	Ireland	FJ389644, FJ389605, this study	V/VII
A. sylvaticus	JM1234	Ireland	FJ389665, FJ389663, this study	V/VII
A. sylvaticus	JM446	Netherlands	FJ389641, FJ389588, this study	V/VII
A. sylvaticus	JM4583	Sweden	FJ389632, FJ389607, this study	V/VII
A. sylvaticus	JM4584	Sweden	FJ389633, FJ389608, this study	V/VII
A. sylvaticus	JM4590	Sweden	FJ389634, FJ389606, this study	V/VII
A. sylvaticus	JM4592	Sweden	FJ389635, FJ389618, this study	V/VII
A. sylvaticus	JM4596	Sweden	FJ389638, FJ389619, this study	V/VII
A. sylvaticus	JM4598	Sweden	FJ389636, FJ389600, this study	V/VII
A. sylvaticus	JM4600	Sweden	FJ389620, FJ389615, this study	V/VII
A. sylvaticus	JM4624	Sweden	FJ389631, FJ389616, this study	V/VII
A. sylvaticus	JM4625	Sweden	FJ389637, FJ389617, this study	V/VII
A. sylvaticus	JM4661	Sweden	FJ389626, FJ389592, this study	V/VII
A. sylvaticus	JM4676	Sweden	FJ389624, FJ389595, this study	V/VII

Please cite this article in press as: Dubey, S. et al., False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene, Mol. Phylogenet. Evol. (2009), doi:10.1016/j.ympev.2008.12.008

Table 1 (*continued*)

Species label	Collection code	Location	GenBank	Lineage
A. sylvaticus	JM4707	Sweden	FJ389627, FJ389591, this study	V/VII
A. sylvaticus	JM4775	Sweden	FJ389623, FJ389599, this study	V/VII
A. sylvaticus	JM4796	Sweden	FJ389628, FJ389593, this study	V/VII
A. sylvaticus	JM4814	Sweden	FJ389629, FJ389597, this study	V/VII
A. sylvaticus	JM4833	Sweden	FJ389630, FJ389598, this study	V/VII
A. sylvaticus	JM4854	Sweden	FJ389625, FJ389583, this study	V/VII
A. sylvaticus	G3.4	Germany	AY179492, Reutter et al. (2003)	VI
A. sylvaticus		Germany	AF159395, Martin et al. (2000)	VII
A. sylvaticus	HB702	Germany	FJ389611, this study	VII
A. sylvaticus	HB715	Germany	FJ389613, this study	VII
A. sylvaticus	HB721	Germany	FJ389614, this study	VII
A. sylvaticus	ZFMK2008.317 (711)	Germany	FJ389610, this study	VII
A. sylvaticus	ZFMK2008.318 (713)	Germany	FJ389612, this study	VII
A. sylvaticus	ZFMK2008.319 (728)	Germany	FJ389609, this study	VII
A. sylvaticus	JM4788	Sweden	FJ389582, this study	VII
A. sylvaticus	JM4812	Sweden	FJ389590, this study	VII
A. sylvaticus	JM4786	Sweden	FJ389596, this study	VII
A. sylvaticus	JM4788	Sweden	FJ389582, this study	VII
A. sylvaticus	JM4812	Sweden	FJ389590, this study	VII
Mus musculus		Japan	AB205312, Terashima et al. (2006)	

An additional lineage (VIII) is represented by one sample of *A. ponticus* from Georgia (Hille et al., 2002), which is closely related to the lineages III and IV. This lineage is not shown in Fig. 2, for the same reason as the other samples from Hille et al. (2002).

The mean Kimura two-parameter genetic distance (K2P; Kimura, 1980) between *Apodemus* (*Sylvaemus*) lineages varied from 5.1% (IV–III) to 15.3% (I–IV; Table 2), and between *Mus musculus* and *Apodemus* lineages from 15.0% (II) to 19.4% (VII).

3.2. Comparison of nucleotidic and amino acid composition of previously published sequences

The sequences G3.1 (lineage VII, *A. fulvipectus*), G3.3 (lineage V, *A. sylvaticus*), and G3.4 (lineage VI, *A. sylvaticus*) of Reutter et al. (2003) are characterized by numerous ambiguous positions e.g., D (G or A or T), K (G or T), M (A or C), N (G or A or T or C), R (G or A), W (A or T) and Y (T or C). The number of uncertain mutations between sequences varied from 14% to 34% (Table 3). In addition, 92.3% of ambiguous positions in the sequence G3.4 are positions showing a mutation between G3.3 and G3.1. The sample G3.4 of *A. sylvaticus* was considered by Reutter et al. (2003) to belong to an unexpected lineage of *Apodemus* in Germany sympatric to lineage VII (including the sample of *A. fulvipectus* from Reutter et al., 2003) and to lineage V of *A. sylvaticus*.

However, considering these previous results, it strongly suggests that the unique sequence of lineage VI of *A. sylvaticus* (G3.4; Reutter et al., 2003) is due to a co-amplification by PCR of lineages V and VII, resulting in ambiguous determination of the correct base for numerous positions, leading to an unexpected phylogenetic position of this sequence in Reutter et al. (2003). This co-amplification could result from a DNA contamination, or from the presence of a nuclear copy of the mitochondrial *cyt-b* (pseudogene).

The mean number of amino acid differences between lineages varied from 0.667 (II–VI) to 4.028 (IV–VII; Table 2). In addition, the overall mean number of amino acid differences between lineages of *Apodemus* spp. varied from 1.70 (II) to 3.40 (VII).

The relation between the mean K2P distance between lineages and the corresponding mean number of amino acid differences is shown in Fig. 4. It reveals that lineage VII shows a higher level of amino acid differences with other lineages compared to its respective mean K2P distances (Fig. 4).

4. Discussion

Nuclear copies of mitochondrial genes have been described in a variety of animals and plants, and revealed to be common in mammals, particularly in Rodentia (e.g. Cooper et al., 2003; DeWoody et al., 1999; Mirol et al., 2000; Jaarola and Searle 2004; Rat Genome Sequencing Project Consortium, 2004; Richly and Leister, 2004; Triant and DeWoody 2007, 2008). These nuclear sequences can cause major problems in systematic analyses, including DNA barcoding and phylogeography, because the mitochondrial DNA genome is haploid, and is under different selection pressures than the nuclear genome in addition to exhibiting a faster rate of nucle-

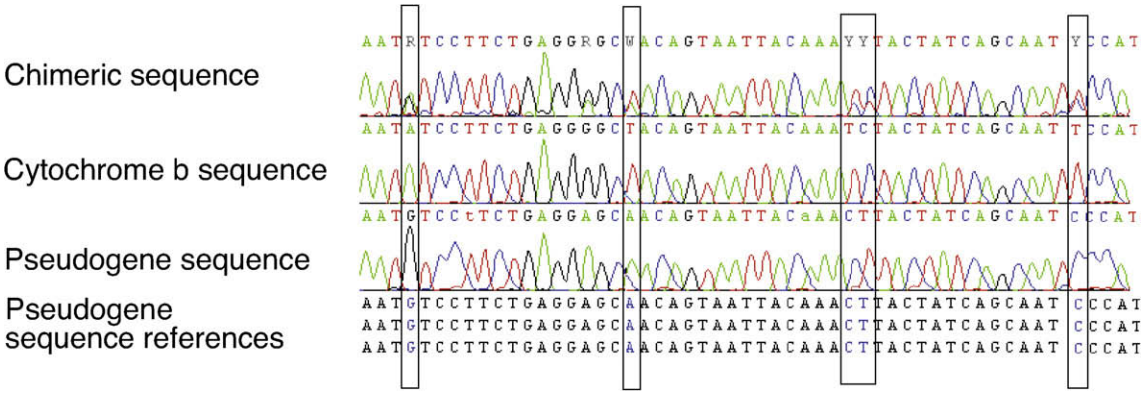

Fig. 2. Examples of electropherograms of a chimeric *cyt-b* sequence of lineage V/VII (VI) and of pure lineages V and VII.

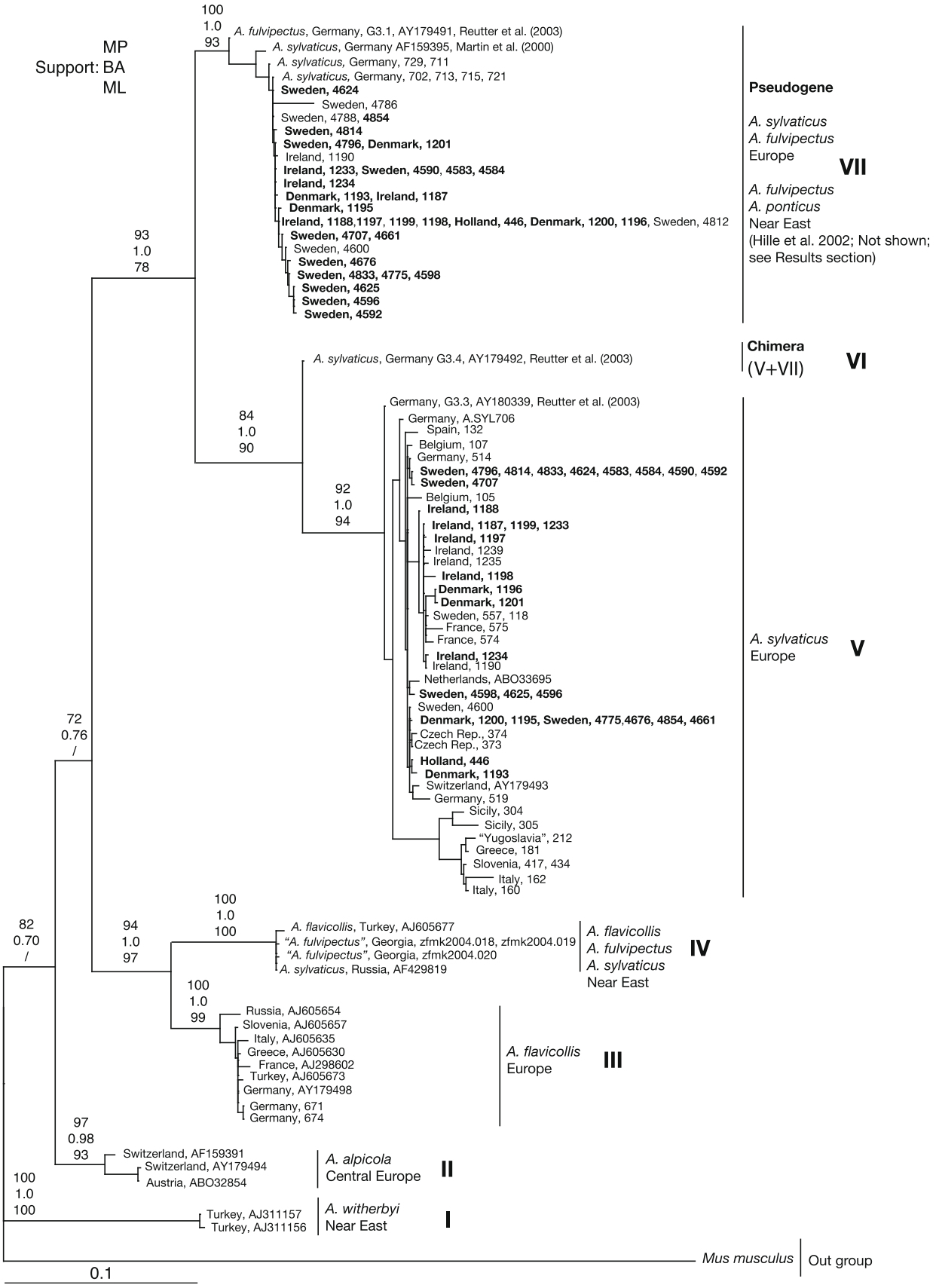

Fig. 3. Phylogeny of mitochondrial *cyt-b* analysed with maximum likelihood. Values in branches show indices of support for the major branches for maximum parsimony analyses (MP), maximum likelihood analyses (ML), and Bayesian posterior probabilities (BA). *A. sylvaticus* specimens analysed for both *cyt-b* and pseudogene appearing in two lineages (V and VII) are marked in bold. Codes are as in Table 1.

Please cite this article in press as: Dubey, S. et al., False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene, Mol. Phylogenet. Evol. (2009), doi:10.1016/j.ympev.2008.12.008

Table 2
Mean number of amino acids differences between lineages of *Apodemus* spp. and between *Mus musculus* and *Apodemus* spp. (lower matrix), overall mean number of amino acids differences between lineages of *Apodemus* spp. (without lineage VI) without and with lineage VII, respectively (diagonal), and mean K2P distance between lineages (upper matrix).

	I	II	III	IV	V	VI	VII	Mus
I	1.922/1.915	0.112	0.143	0.153	0.146	0.149	0.134	0.179
II	1.667	1.700/2.048	0.078	0.062	0.107	0.096	0.079	0.150
III	1.889	1.667	1.755/2.093	0.051	0.096	0.109	0.108	0.161
IV	2.250	1.917	1.861	1.977/2.316	0.105	0.109	0.108	0.161
V	1.880	1.547	1.880	2.130	1.859/2.293	0.043	0.103	0.159
VI	1.000	0.667	1.000	1.250	0.880	–/–	0.064	0.176
VII	**1.889**	**3.444**	**3.673**	**4.028**	**3.658**	**2.778**	**–/3.338**	**0.194**
Mus	7.000	5.333	7.000	6.250	6.160	6.000	8.778	6.349/6.754

Table 3
Mutations observed between sequences of Reutter et al. (2003) of lineage V–VII (in bold, ambiguous position). The following letters are ambiguous positions: D (G or A or T), K (G or T), M (A or C), N (G or A or T or C), R (G or A), W (A or T) and Y (T or C).

Samples	Mutations observed between sequences of Reutter et al. (2003)																							
VII, AY179491, G3.1	T	C	C	A	**Y**	A	C	C	C	C	C	A	C	T	T	A	C	C	Y	A	T	T	C	C
VI, AY179492, G3.4	**Y**	.	**R**	.	**Y**	G	.	.	Y	D	T	.	.	.	**K**	.	T	.	T	G	T	.	**W**	**N**
V, AY180339, G3.3	C	T	T	T	C	G	T	A	T	T	T	G	T	C	G	T	T	Y	T	G	C	C	C	T
VII, AY179491, G3.1	A	G	**Y**	**R**	T	G	A	**Y**	T	A	**W**	T	C	Y	G	C	C	A	C	**Y**	**Y**	T	A	C
VI, AY179492, G3.4	G	A	Y	G	C	.	T	C	C	T	T	C	.	T	.	A	**Y**	**M**	.	T	C	**Y**	G	T
V, AY180339, G3.3	G	A	T	G	C	A	T	C	C	T	T	C	T	T	A	A	**Y**	C	T	C	C	C	G	T
VII, AY179491, G3.1	G	A	T	T	G	A	C	C	**Y**	C	T	T	G	T	T	T	C	C	C	T	T	C	C	T
VI, AY179492, G3.4	.	T	C	C	.	C	T	T	G	T	C	C	A	**Y**	A	.	T	.	.	C	.	T	T	C
V, AY180339, G3.3	A	T	C	C	A	C	T	T	T	T	C	C	A	C	A	C	T	T	T	C	**Y**	T	T	C

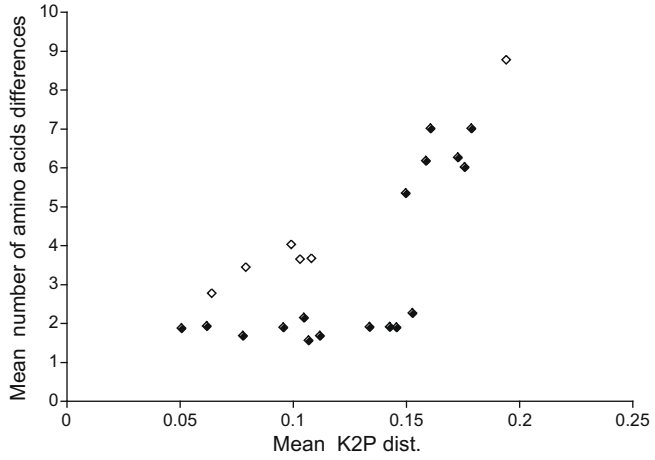

Fig. 4. Relation between the mean K2P distance between lineages and the corresponding mean number of amino acids differences (excluding lineage VI and including the out group *Mus musculus*); closed lozenge, relation between lineages I–V, and the out group; open lozenge, relation between lineage VII and the other lineages, and the out group.

otide substitutions. The inclusion of nuclear DNA in mtDNA data sets can therefore lead to inaccurate species identifications, divergence estimates, or phylogenetic groupings (Zhang and Hewitt, 1996, 2003; Triant and DeWoody, 2007).

Such nuclear copies can be detected by typical features as (i) the presence of stop codon and frame-shift mutations (if the sequence is long enough), (ii) an accumulation of non-synonymous mutations (if the pseudogene is old enough to exhibit such pattern), or (iii) the DNA amplification of a mixture of a pseudogene and the mitochondrial gene, resulting in chimeric sequences (Jaarola and Searle, 2004; Triant and DeWoody, 2008). In the last situation, the development of primers specific to the pseudogene permits the study of its evolution within and between species (Triant and DeWoody 2008).

In our study, the presence of a cryptic pseudogene in published sequences of *Apodemus* spp. from various authors, i.e. Martin et al. (2000), Hille et al. (2002) and Reutter et al. (2003), was revealed by five independent pathways:

(a) The presence of a chimeric sequence (G3.4, lineage VI; Fig. 3, Table 3) in Reutter et al. (2003), between the mitochondrial *cyt-b* of *A. sylvaticus* (G3.3; lineage V) and its pseudogene (G3.1; lineage VII), resulting in the coamplification by PCR of the nuclear pseudogene with its mitochondrial counterpart. This result is revealed by the high number of ambiguous positions in the sequence of lineage VI (92.3%), which are positions showing a mutation between *A. sylvaticus* G3.3 of lineage V and *A. fulvipectus*, G3.1 of lineage VII.

(b) The amplification of the mitochondrial *cyt-b* (lineage V) or its pseudogene (lineage VII); as well as their co-amplification in samples of *A. sylvaticus* from various locations in Europe, using standard primers (e.g. Michaux et al., 2003, 2004, 2005; Reutter et al., 2003; this study); or strictly the pseudogene with our specific primer (H16), (Fig. 3). Consequently, it revealed that both lineages are present within the same samples of *A. sylvaticus*.

(c) The higher number of amino acid differences between lineage VII and the other lineages (mean number of differences: 3.34), compared to the differences observed between the other lineages (mean number of differences: 1.7–1.98; Table 2). This accumulation of non-synonymous mutations in a nuclear copy of mitochondrial genes compared to its mitochondrial counterpart is a typical characteristic. In addition, the relation between the mean K2P distance between lineages and the corresponding mean number of amino acids differences revealed that lineage VII showed a higher level of amino acid differences with other lineages compared to its respective mean K2P distances (Fig. 4).

(d) The absence of a conclusive biogeographic distribution of lineage VII, which was only found in Germany and Georgia, at localities separated by c. 2700 km, while numerous samples of *Apodemus* spp. analysed from between these localities (e.g. Michaux et al., 2003, 2004, 2005) yielded different lineages. Therefore, from a biogeographic point of view, it was unlikely that such a distance separates conspecific populations of *Apodemus*, and that

Please cite this article in press as: Dubey, S. et al., False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene, Mol. Phylogenet. Evol. (2009), doi:10.1016/j.ympev.2008.12.008

no additional populations of lineage VII were found in the range in between.

(e) The presence of samples of the same species (determined morphologically) within the same localities belonging to two different lineages, e.g. *A. ponticus* and *A. fulvipectus* from Georgia (Hille et al., 2002) and *A. sylvaticus* from Germany (Reutter et al., 2003). In each study the second lineage belonged to lineage VII. In addition, it is unlikely that a cryptic species (lineage VII) showed morphological similarities with *A. fulvipectus*, *A. ponticus*, and *A. sylvaticus,* preventing its morphological determination, whereas morphological traits easily distinguish *A. fulvipectus*, *A. ponticus*, and *A. sylvaticus*.

Considering the strong evidence (a–e) for the presence of a cryptic pseudogene in *Apodemus* spp., the mitochondrial phylogenetic relationships between all the species of the genus *Apodemus* and their biogeographic history should be carefully revisited without the inclusion of the pseudogene (lineage VII) and the erroneous sequences of Reutter et al. (2003). This exclusion should clarify the evolution of *Apodemus*, particularly within the subgenus *Sylvaemus*. In addition, it may solve the taxonomic ambiguities linked to discrepancies between molecular markers and morphological characters of species from the Near East, such as *A. ponticus* and *A. witherbyi*, and therefore clarify their geographic distributions.

In summary, the main results of this study are the presence of (i) a well-differentiated lineage of *Apodemus* (VII) including samples of various species (*A. sylvaticus*, *A. fulvipectus*, *A. ponticus*; Martin et al., 2000; Hille et al., 2002; Reutter et al., 2003, our study; Table 1, Fig. 3) and locations (from Germany to Georgia), which revealed to be a nuclear copy of the mitochondrial *cyt-b* leading to erroneous trees (Martin et al., 2000) including misunderstanding in previous molecular phylogeny of *Apodemus* (Hille et al., 2002; Reutter et al., 2003; Suzuki et al., 2008); (ii) a chimeric sequence in Reutter et al. (2003; G3.4; lineage VI) between the mitochondrial *cyt-b* and its nuclear copy (Table 3).

In a more general context, with the creation of identification systems reliant on the analysis of sequence diversity in small segments of DNA (DNA barcoding; Tautz et al., 2003), the presence of cryptic pseudogenes should be carefully considered. In fact, most of the studies based on DNA barcoding use mitochondrial genes such as *cyt-b* or *COI* (cytochrome *c* oxidase I), for which numerous pseudogenes have be found in a large panel of taxa (Triant and DeWoody, 2007). Consequently, only DNA barcoding studies focussing on a precise group of taxa with extensive sampling will allow detecting such cryptic pseudogenes as shown in Hebert et al. (2004). These conditions should be the norm in future DNA barcoding studies. However, an alternative option will be to use methods allowing the purification of mtDNA, but excluding the nuclear one. Nevertheless, such techniques are often complex or unsuitable, when they required fresh samples, which were never stored in ethanol (see e.g. Ibarguchi et al., 2006).

Acknowledgments

For providing tissue samples, we thank our colleagues A. Bukhnikashvili (Georgia), W. Booth and I. Montgomery (Ireland), J. Searle and X. Lambin (Great Britain), G. Olsson (Sweden) and T.S. Hansen (Denmark). We also thank N. Di Marco for laboratory work. This study was supported by the Swiss National Science Foundation (SNSF) to S.D, and by a Belgian FNRS fellowship (mandat "chercheur qualifié") and a financial grant of the Belgian FNRS (crédits aux chercheurs) to J.M.

References

Balakirev, A.E., Baskevich, M.I., Gmyl, A.P., Okulova, N.M., Andreeva, T.A., Sokolenko, O.V., Malygin, V.M., Khlyap, L.A., Oparin, M.L., Orlov, V.N., 2007. On the taxonomic rank of *ciscaucasicus* and its relationships with the pygmy wood mouse *Sylvaemus uralensis* inferred from the mtDNA cytochrome *b* gene sequence. Russ. J. Genet. 43, 1386–1399.

Cooper, G.M., Brudno, M., NISC Comparative Sequencing Program, Green, E.D., Batzoglou, S., Sidow, A., 2003. Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. Genome Res. 13, 813–820.

Dekonenko, A., Yakimenko, V., Ivanov, A., Vyacheslav, M., Pavel, N., 2003. Genetic similarity of puumala viruses found in Finland and western Siberia and of the mitochondrial DNA of their rodent hosts suggests a common evolutionary origin. Infect. Genet. Evol. 3, 245–247.

DeWoody, J.A., Chesser, R.K., Baker, R.J., 1999. A translocated mitochondrial cytochrome b pseudogene in voles (Rodentia: *Microtus*). J. Mol. Evol. 48, 380–382.

Filippucci, M.G., Macholan, M., Michaux, J., 2002. Genetic variation and evolution in the genus *Apodemus* (Muridae, Rodentia). Biol. J. Linn. Soc. 75, 395–419.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 692–704.

Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc. Natl. Acad. Sci. USA 101, 14812–14817.

Hille, A., Tarkhnishvili, D., Meinig, H., Hutterer, R., 2002. Morphometric, biochemical and molecular traits in Caucasian wood mice (*Apodemus*/*Sylvaemus*), with remarks on species divergence. Acta Theriol. 47, 389–416.

Hoofer, S.R., Gaschak, S., Dunina-Barkovskaya, Y., Makluk, J., Meeks, H.N., Wickliffe, J.K., Baker, R.J., 2007. New information for systematics, taxanomy, and phylogeography of the rodent genus *Apodemus* (*Sylvaemus*) in Ukraine. J. Mammal. 88, 330–342.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Evolution: Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Ibarguchi, G., Friesen, V., Lougheed, S.C., 2006. Defeating numts: semi-pure mitochondrial DNA from eggs and simple purification methods for field-collected wildlife tissues. Genome 49, 1438–1450.

Jaarola, M., Searle, J.B., 2004. A highly divergent mitochondrial DNA lineage of *Microtus agrestis* in southern Europe. Heredity 92, 228–234.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Martin, Y., Gerlach, G., Schlotterer, C., Meyer, A., 2000. Molecular phylogeny of European muroid rodents based on complete cytochrome b sequences. Mol. Phylogenet. Evol. 16, 37–47.

Michaux, J.R., Chevret, P., Filippucci, M.G., Macholan, M., 2002. Phylogeny of the genus *Apodemus* with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers: cytochrome *b* and 12S rRNA. Mol. Phylogenet. Evol. 23, 123–136.

Michaux, J.R., Magnanou, E., Paradis, E., Nieberding, C., Libois, R., 2003. Mitochondrial phylogeography of the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic Region. Mol. Ecol. 12, 685–697.

Michaux, J.R., Libois, R., Paradis, E., Filippucci, M.G., 2004. Phylogeographic history of the yellow-necked fieldmouse (*Apodemus flavicollis*) in Europe and in the Near and Middle East. Mol. Phylogenet. Evol. 32, 788–798.

Michaux, J.R., Libois, R., Filippucci, M.G., 2005. So close and so different: comparative phylogeography of two small mammal species, the Yellownecked fieldmouse (*Apodemus flavicollis*) and the Woodmouse (*Apodemus sylvaticus*), in the Western Palearctic region. Heredity 94, 52–63.

Mirol, P.M., Mascheretti, S., Searle, J.B., 2000. Multiple nuclear pseudogenes of mitochondrial cytochrome b in *Ctenomys* (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. Heredity 84, 538–547.

Musser, G.G., Carleton, M.C., 2005. Superfamily Muroidea, third ed.. In: Wilson, D.E., Reeder, D.M. (Eds.), Mammals Species of the World: A Taxonomic and Geographic Reference, Vol. 2 The Johns Hopkins University Press, Baltimore, pp. 894–1522.

Parker, S.R., 1997. Sequence navigator. multiple sequence alignment software. Methods Mol. Biol. 70, 145–154.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rat Genome Sequencing Project Consortium, 2004. Genome sequences of the Brown Norway rat yields insights into mammalian evolution. Nature 428, 493–521.

Richly, E., Leister, D., 2004. Numts in sequenced eukaryotic genomes. Mol. Biol. Evol. 21, 1081–1084.

Reutter, B.A., Petit, E., Bruenner, H., Vogel, P., 2003. Cytochrome *b* haplotype divergences in West European *Apodemus*. Z. Saugetierkunde 68, 153–164.

Rodriguez, R., Oliver, J.L., Marin, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Serizawa, K., Suzuki, H., Tsuchiya, K., 2000. A phylogenetic view on species radiation in *Apodemus* inferred from variation of nuclear and mitochondrial genes. Biochem. Genet. 38, 27–40.

Suzuki, H., Tsuchiya, K., Takezaki, N., 2000. A molecular phylogenetic framework for the Ryuyku endemic rodents *Tokudaia osimensis* and *Diplothrix legata*. Mol. Phylogenet. Evol. 15, 15–24.

Suzuki, H., Filippucci, M.G., Chelomina, G.N., Sato, J.J., Serizawa, K., Nevo, E., 2008. A Biogeographic View of *Apodemus* in Asia and Europe inferred from nuclear and mitochondrial gene sequences. Biochem. Genet. 46, 329–346.

Swofford, D.L., 2001. PAUP*4.0: phylogenetic analyses using parsimony (*and other methods). Beta version 4.0b10a for PC. Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P., 2003. A plea for DNA taxonomy. Trends Ecol. Evol. 18, 70–74.

Terashima, M., Furusawa, S., Hanzawa, N., Tsuchiya, K., Suyanto, A., Moriwaki, K., Yonekawa, H., Suzuki, H., 2006. Phylogeographic origin of Hokkaido house mice (*Mus musculus*) as indicated by genetic markers with maternal, paternal and biparental inheritance. Heredity 96, 128–138.

Triant, D.A., DeWoody, A.J., 2007. The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. J. Mammal. 88, 908–920.

Triant, D.A., DeWoody, A.J., 2008. Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodents. Genetica 132, 21–33.

Zhang, D.X., Hewitt, G.M., 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol. Evol. 11, 247–251.

Zhang, D.X., Hewitt, G.M., 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Mol. Ecol. 12, 563–584.

Please cite this article in press as: Dubey, S. et al., False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene, Mol. Phylogenet. Evol. (2009), doi:10.1016/j.ympev.2008.12.008