MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 4617–4630 doi: 10.1111/mec.12886

# The hidden history of the snowshoe hare, *Lepus americanus*: extensive mitochondrial DNA introgression inferred from multilocus genetic variation

JOSÉ MELO-FERREIRA,*[1] FERNANDO A. SEIXAS,*†[1] ELLEN CHENG,‡§ L. SCOTT MILLS‡¶ and PAULO C. ALVES*†‡

*CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO - Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal, †Departamento Biologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal, ‡Wildlife Biology, University of Montana, 32 Campus Drive, Missoula, MT 59812, USA, §Ugyen Wangchuck Institute for Conservation and Environment, Lamai Goempa, Bumthang, Bhutan, ¶Fisheries, Wildlife and Conservation Biology Program, North Carolina State University, Raleigh, NC 27695-7617, USA

## Abstract

**Hybridization drives the evolutionary trajectory of many species or local populations, and assessing the geographic extent and genetic impact of interspecific gene flow may provide invaluable clues to understand population divergence or the adaptive relevance of admixture. In North America, hares (*Lepus* spp.) are key species for ecosystem dynamics and their evolutionary history may have been affected by hybridization. Here we reconstructed the speciation history of the three most widespread hares in North America – the snowshoe hare (*Lepus americanus*), the white-tailed jackrabbit (*L. townsendii*) and the black-tailed jackrabbit (*L. californicus*) – by analysing sequence variation at eight nuclear markers and one mitochondrial DNA (mtDNA) locus (6240 bp; 94 specimens). A multilocus–multispecies coalescent-based phylogeny suggests that *L. americanus* diverged ~2.7 Ma and that *L. californicus* and *L. townsendii* split more recently (~1.2 Ma). Within *L. americanus*, a deep history of cryptic divergence (~2.0 Ma) was inferred, which coincides with major speciation events in other North American species. While the isolation-with-migration model suggested that nuclear gene flow was generally rare or absent among species or major genetic groups, coalescent simulations of mtDNA divergence revealed historical mtDNA introgression from *L. californicus* into the Pacific Northwest populations of *L. americanus*. This finding marks a history of past reticulation between these species, which may have affected other parts of the genome and influence the adaptive potential of hares during climate change.**

*Keywords*: coalescent, cryptic divergence, hares and jackrabbits, lagomorphs, reticulate evolution, species tree

*Received 6 May 2014; revision received 7 August 2014; accepted 8 August 2014*

## Introduction

The modern view of interspecific dynamics recognizes that closely related species, even when divergence is irreversible, may exchange genetic material and that introgressive hybridization plays an important role in shaping the genetic diversity of taxa. Mallet (2005), for example, estimated that 10% of animal species hybridize with at least one other closely related species (see also Pinho & Hey 2010). Understanding patterns of introgression is therefore important to unveil the determinants of major processes of species evolution, such as the genetic nature of population divergence or the generation of adaptive genetic innovation (Seehausen 2004, 2013; Feder *et al.* 2012; Abbott *et al.* 2013).

Correspondence: José Melo-Ferreira, Fax: +351 252 661 780; E-mail: jmeloferreira@cibio.up.pt

[1]These authors contributed equally.

Inferences of introgression have often been based on gene tree polyphyly or paraphyly and incongruence among gene trees (Bossu & Near 2009; Spinks & Shaffer 2009). However, discordance among markers may arise from the stochasticity of the evolutionary process itself, due to the incomplete sorting of lineages along the divergence of species. Distinguishing these two causes of gene tree discordance is not straightforward, particularly for closely related taxa (Edwards 2009). Nevertheless, several methodological strategies have been created to assess the relative influence of retention of ancestral polymorphism and gene flow in observed patterns of multilocus genetic variation (Meng & Kubatko 2009; Hey 2010; Gerard *et al.* 2011).

Natural hybridization often occurs among species with a rapid and young radiation, and hares (*Lepus* spp.) have emerged as a particularly suitable model to study reticulate evolution (Thulin *et al.* 2006a,b; Alves *et al.* 2008; Melo-Ferreira *et al.* 2009, 2011, 2012; Liu *et al.* 2011). Even though most instances of introgressive hybridization described among hares relate to areas of present species contact (e.g. between *L. timidus* and *L. europaeus* in Sweden or Russia; Thulin *et al.* 2006a,b), cases of ancestral introgression between currently allopatric species have also been reported (Alves *et al.* 2003; Melo-Ferreira *et al.* 2012). Even though these reticulation events are more pronounced in the mtDNA, they also occur at the nuclear genome, but at different degrees across inheritance pathways and chromosome regions (Melo-Ferreira *et al.* 2009, 2011, 2012).

Given the widespread nature of genome reticulation and extensive introgression in hares (reviewed by Alves *et al.* 2008), introgression is expected to have also impacted the evolution of North American species in the United States and Canada, with potential consequences to their conservation and adaptive potential. In North America, hares are strong interactors in ecosystem dynamics (Tyson *et al.* 2010; Krebs 2011; Lewis *et al.* 2011) and model systems for basic ecological studies ranging from cyclic population dynamics to mechanisms of top-down versus bottom-up population control (Griffin & Mills 2009; Krebs 2011), to the ecology of stress (Boonstra 2013). Also, two of the most widespread hare species in North America (snowshoe hares, *Lepus americanus*, and white-tailed jackrabbits, *L. townsendii*) undergo seasonal coat colour changes, a trait vulnerable to being compromised by climate change, as the number of days of white hares on brown backgrounds increases in the future (Mills *et al.* 2013; Zimova *et al.* 2014). Despite these studies, information on the evolutionary history of North American hares is still very scarce. Recently, a comprehensive study by Cheng *et al.* (2014) based on microsatellites and mtDNA sequences and covering the entire range of the snowshoe hare suggested that this species is structured in three major evolutionary population clusters with well-defined geographic distributions: Boreal (entire northern and eastern range of the species), Rockies and Pacific Northwest. This pattern of population structure is similar to that inferred for other boreal North American mammals, implying that common phenomena such as climatic oscillations may have shaped the phylogeography of this species. Cheng *et al.* (2014) also show that the Pacific Northwest population of *L. americanus* possesses an mtDNA lineage that is more closely related to that of the black-tailed jackrabbit, *L. californicus*. This pattern of mtDNA divergence may result from secondary introgression following interspecific hybridization, as often described among species of hares, or from incomplete lineage sorting. However, distinguishing between these competing hypotheses requires reconstructing the speciation history of these species. In addition, Flux (1983) reported that *L. californicus* hybridizes in the wild with the white-tailed jackrabbit, but no study of the genetic consequences of this hybridization has been conducted.

In this study, we aimed to infer the evolutionary history of the three most widespread North American hare species *L. americanus*, *L. californicus* and *L. townsendii*, by analysing the sequence variation at nine loci from all inheritance pathways. In addition, we determine the extent and timing of gene introgression in these species and discuss the potential adaptive importance of hybridization in their evolution.

## Materials and methods

### *Sampling and data collection*

A total of 94 individuals (48 *L. americanus*, 30 *L. californicus* and 16 *L. townsendii*) from 14 sampling locations were used in this study (Table 1; Tables S1 and S2, Supporting information), including the three *L. americanus* population clusters described by Cheng *et al.* (2014) (Fig. 1). The European rabbit, *Oryctolagus cuniculus*, was used as outgroup for some of the analyses.

Total genomic DNA was extracted from muscle and ear tissues using the JETQUICK Tissue DNA Kit (Genomed) following manufacturer's instructions. The sex of the individuals was determined following the PCR approach described by Wallner *et al.* (2001). Nine loci from all inheritance pathways – five autosomal (SPTBN1, PRKCI, DARC, KITLG and TF), one mitochondrial (CYTB), two X-linked (POLA1 and GRIA3) and one Y-linked (SRY) – were amplified by polymerase chain reaction (PCR) (Table 2; see Table S3, Supporting information, for primers and PCR conditions). Purified PCR products were automatically sequenced (Macrogen

**Table 1** Species and geographic location of the samples collected in this study

| Species | Locality number | Locality code* | Locality | Sample size |
|---|---|---|---|---|
| *Lepus americanus* | 1 | CA1 | California, U.S.A. | 8 |
| | 2 | WA1 | Washington, U.S.A. | 8 |
| | 3 | WA4 | Washington, U.S.A. | 10 |
| | 4 | OR2 | Oregon, U.S.A. | 8 |
| | 5 | SK1 | Saskatchewan, Canada | 6 |
| | 6 | WY1 | Wyoming, U.S.A. | 8 |
| | | | Total *L. americanus* | 48 |
| *Lepus californicus* | 7 | LCA_OR | Oregon, U.S.A. | 10 |
| | 8 | LCA_CA | California, U.S.A. | 6 |
| | 9 | LCA_TE | Texas, U.S.A. | 8 |
| | 10 | LCA_AR | Arizona, U.S.A. | 6 |
| | | | Total *L. californicus* | 30 |
| *Lepus townsendii* | 11 | LTO_ID1 | Idaho, U.S.A. | 8 |
| | 12 | LTO_MO1 | Montana, U.S.A. | 2 |
| | 13 | LTO_WY1 | Wyoming, U.S.A. | 1 |
| | 14 | LTO_MO2 | Montana, U.S.A (Yellowstone N.P.) | 5 |
| | | | Total *L. townsendii* | 16 |
| | | | Total | 94 |

*Locality codes in *L. americanus* as in Cheng *et al.* (2014).

Inc, Netherlands) using forward and reverse PCR primers and occasionally internal primers as indicated in Table S3.

### *Analysis of sequence data sets*

Sequences were visually inspected and aligned using CLUSTALW (Thompson *et al.* 1994) as implemented in BIOEDIT v7.0.5.3 (Hall 1999). Polymorphic tandem repeats and the 5-bp adjacent regions were excluded. Allelic phases were determined using PHASEv2.1.1 (Stephens *et al.* 2001; Stephens & Scheet 2005). Input files were produced with the online software SEQPHASE (Flot 2010). Haplotypes defined from individuals with heterozygous insertion–deletions, following Flot *et al.* (2006), were incorporated in the analysis to improve phase determination (Stephens *et al.* 2001). Five replicate runs of 1000 iterations after an initial burn-in of 1000 generations were performed, with a thinning interval of 1, and the run with the best average goodness of fit was retained. As PHASE has been shown to generate a very low number of false positives (Garrick *et al.* 2010), the complete data set including some low-probability calls was kept to avoid biasing levels of diversity and the frequency spectra of mutations. Sequence alignments were reduced to the largest nonrecombining blocks using IMGC (Woerner *et al.* 2007).

Finally, we assessed conformation of the multilocus variation to neutral expectations using the HKA test (Hudson *et al.* 1987) as implemented in the software HKA (https://bio.cst.temple.edu/~hey/software/software.htm#HKA) and using both the rabbit or each of the other hare species as outgroup.

### *Phylogenetic and species delimitation analysis*

To estimate phylogenies of the individual nuclear loci, the European rabbit was used as the outgroup, while for the cytochrome *b* phylogeny, both the European rabbit and the eastern cottontail (*Sylvilagus floridanus*) were used (GenBank accession nos in Table S1, Supporting information). The best-fit model of sequence evolution for each sequenced locus was determined among 88 possible models using JMODELTEST v0.1.1 (Guindon & Gascuel 2003; Posada 2008) under the Akaike information criterion with correction (AICc).

Maximum likelihood (ML) and Bayesian inference (BI) phylogenies were estimated for each nuclear locus using GARLI v2.0 (Zwickl 2006) and BEAST v1.7.4 (Drummond *et al.* 2012), respectively, using European rabbit sequences as outgroup. For Garli, five replicate runs of 1 million generations were performed using the best-fit mutation model and without fixing the model parameters. For BEAST, three independent runs of 50 million generations were performed, applying the best-fit mutation model or the next-most complex model implemented in the software, a Yule tree prior and an uncorrelated lognormal relaxed clock (Drummond *et al.* 2006). Runs were examined in TRACER v1.5 (Rambaut & Drummond 2007) and concatenated using LOGCOMBINER, and post-burn-in trees were summarized using TREEANNOTATOR, part of the BEAST package. For the cytochrome *b*, both the European rabbit and the eastern cottontail were used as outgroups and similar phylogeny reconstruction analyses were conducted, but running 250 million generations for the BI and 5 million generations and 500 bootstrap replicates for the ML estimate.

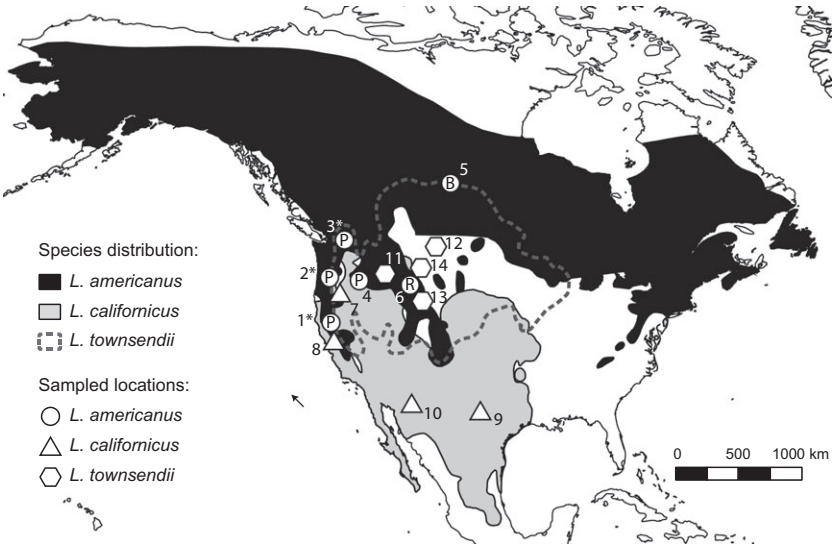


**Fig. 1** Distribution of *L. americanus*, *L. californicus* and *L. townsendii* in North America, and approximate locations of samples used in this study. Letters in *L. americanus* sample locations indicate the microsatellite cluster identified by Cheng *et al.* (2014): B – Boreal; R – Rockies; P – Pacific Northwest (the localities where the *L. californicus*-like mtDNA was found are indicated by '*'). See Table 1 for the detailed location of sampling sites (depicted by numbers).

**Table 2** Loci included in this study, length of obtained sequences and inferred mutation models

| Locus | | | Number of characters | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | LNRB* | Out† | Variable‡ | Noncoding§ | Exon§ | Mutation model¶ |
| 1 | SPTBN1 | Spectrin, beta, nonerythrocytic 1 | 636** | 561 | 561 | 26 | 1–561 | — | K80 |
| 2 | PRKCI | Protein kinase C, iota | 436 | 432 | 426 | 36 | 10–432 | 1–9 | HKY |
| 3 | DARC | Duffy blood group, chemokine receptor | 783 | 741 | 741 | 26 | — | 1–741 | TPM2uf+Γ |
| 4 | KITLG | KIT ligand | 552 | 461 | 461 | 23 | 1–461 | — | JC |
| 5 | TF | Transferrin | 387 | 316 | 320 | 29 | 1–316 | — | JC |
| 6 | POLA1 | Polymerase, alpha 1, catalytic subunit | 813 | 572 | 572 | 34 | 1–572 | — | F81 + Γ |
| 7 | GRIA3 | Glutamate receptor, ionotrophic, AMPA 3 | 969** | 969 | 677 | 41 | 1–969 | — | TrN |
| 8 | SRY | Sex determining region of the Y chromosome | 1608 | 1608 | 1608 | 40 | 1–220; 836–1608 | 221–835 | TIM2 |
| *Total nuclear DNA* | | | 6184 | 5660 | 5366 | 255 | | | — |
| 9 | CYTB | Cytochrome *b* | 580 | 580 | 580 | 127 | — | 580 | TPM3uf+Γ |
| *Total* | | | 6764 | 6240 | 5946 | 382 | | | — |

*Largest nonrecombining blocks.
†Alignment including outgroup.
‡Only ingroup taxa were considered.
§Coordinates of the LNRB alignment.
¶See Posada (2008) for a description of models and references.
**Microsatellites and buffer regions, two in GRIA3 (34 bp; 16 bp) and one in SPTBN1 (19 bp), not considered (see Materials and methods).

Given the stochasticity of the coalescent process, methods that explicitly take into account the possibility of differential lineage sorting across individual loci are expected to perform better in multilocus data sets (Edwards *et al.* 2007; Kubatko & Degnan 2007). We therefore used the multilocus/multispecies Bayesian inference method *BEAST (Heled & Drummond 2010), as implemented in software BEAST v1.7.4 (Drummond *et al.* 2012), to infer the phylogeny of the three focal North American *Lepus* species based on the eight nuclear loci. Two strategies of species assignation were used as follows: (i) specimens were assigned to the three sampled species and (ii) *L. americanus* specimens were split into three units that correspond to the three population clusters described by Cheng *et al.* (2014). Given that this method estimates the root of each single-gene tree and uses the multispecies coalescent of the species tree (Heled & Drummond 2010), outgroup sequences were not included. Model choice

and postrun examination followed the previously described BEAST analyses, but, in this case, three independent *BEAST runs of 500 million generations were performed. The substitution rates of the multiple loci were estimated relative to PRKCI, and the rate for this locus was calibrated using the *Lepus–Oryctolagus* uncorrected genetic distance, considering a split time of 11.8 Myr (Matthee *et al.* 2004).

To assess whether the three *L. americanus* population clusters described by Cheng *et al.* (2014) based on microsatellite data reflect long-term sequence evolution, we performed a Bayesian species delimitation analysis using the nuclear data as implemented in the software BP&P v2.0 (Rannala & Yang 2003; Yang & Rannala 2010). The posterior probability of different possible taxa delimitation models was estimated by collapsing nodes of the species tree considering the three separate population clusters of *L. americanus* (assignation strategy (ii) described above). Different combinations of ancestral effective population size ($\theta$) and root age ($\tau_0$) priors were used (Yang & Rannala 2010) (see Table S4, Supporting information). Two runs of 2 500 000 generations were performed. These analyses were also performed randomizing the assignment of the sequences to groups to assess the robustness of inferences.

### *Isolation-with-migration analysis*

Given that the multispecies–multilocus phylogeny reconstruction method used here relies on the assumption that no introgression occurred between species (Heled & Drummond 2010), we attempted to quantify gene flow levels regardless of the inferred phylogeny by applying the isolation-with-migration (IM) model implemented in IMa2 (Hey 2010) to pairs of species and/or populations. Three independent runs were performed, varying the parameters' upper bound priors and the starting seeds and using the HKY mutation model (Hasegawa *et al.* 1985). Significance of gene flow estimates was assessed using Nielsen & Wakeley (2001) approach and also the likelihood ratio tests of different models implemented in IMa2's L mode. Substitution rates (per generation) were estimated from the *Lepus–Oryctolagus* uncorrected genetic distance, considering a split time of 11.8 Myr (Matthee *et al.* 2004) and a generation time of 2 years (Marboutin & Peroux 1995).

### *Demographic analyses*

The demographic history of the species was also investigated using the Extended Bayesian Skyline Plot (EBSP) analysis (Heled & Drummond 2008) using software BEAST v1.7.4. The EBSP analysis was performed for each species and for each *L. americanus* cluster (Cheng *et al.* 2014) separately. As the *californicus*-like cytochrome *b* sequences of *L. americanus* from the Pacific Northwest may be the result of introgression, they were not included in this analysis. Three independent runs of 200 million generations were performed using the best-fit mutation model selected with JMODELTEST or the next-most complex model implemented in the program. TRACER v1.5 was used to evaluate the combined runs, and EBSPs were plotted using the GraphfromCSV python script provided with BEAST package v1.6.4. The mtDNA substitution rate, estimated from the *Lepus–Oryctolagus* average corrected distance and considering a divergence time of 11.8 Myr (Matthee *et al.* 2004), was used to calibrate the demographic plots.

### *Coalescent simulations*

We followed a methodology similar to that used by Melo-Ferreira *et al.* (2012) to understand the contribution of incomplete lineage sorting and introgression to the mtDNA phylogeny. Divergence time and population size estimates obtained between the Pacific Northwest population of *L. americanus* (the one possessing the discordant mtDNA lineage) and *L. californicus* under the IM model were used as input for SIMCOAL2 v2.1.2 (Laval & Excoffier 2004) to simulate 10 000 cytochrome *b* data sets mimicking the empirical data set. Alternatively, the IM parameter values inferred considering *L. americanus* as a single population were also tested. A model where an ancestral haploid population of size $N_{eA}/2$ splits into two descendant populations of sizes $N_{e1}/2$ and $N_{e2}/2$, $t$ generations ago, with no gene flow occurring between the two descendant populations, was applied. An unequal transition–transversion rate was considered (estimated in JMODELTEST), and the mtDNA substitution rate per generation was again estimated from the *Lepus–Oryctolagus* average corrected distance. The minimum pairwise corrected p-distance between the descendent populations was retained for each replicate. The empirical p-distance was considered to reject the incomplete lineage sorting hypothesis if found to be lower than the 5th percentile of the simulated distribution of minimum distances. This analysis was also performed using the 95% HPD bounds of the IM estimates that maximize incomplete lineage sorting (lower bound of divergence time and upper bounds of effective population sizes).

## Results

### *Sequence data and phylogenetic inferences*

Eight nuclear markers and one mtDNA locus were sequenced in this study, for a total of 6184 bp of nuclear DNA and 580 bp of mtDNA (Table 2). Limiting

the analyses to the largest nonrecombining blocks, the nuclear data set was reduced to 5660 bp (5366 bp with the inclusion of *O. cuniculus* as outgroup) (Table 2). The HKA test did not detect deviations from neutral expectations ($P > 0.05$).

The maximum likelihood (ML) and Bayesian inference (BI) phylogenetic reconstructions showed extensive sequence sharing among species (Figs S1 and S2, Supporting information). The multilocus nuclear phylogeny resulting from *BEAST suggests that *L. californicus* and *L. townsendii* are more closely related than either is to *L. americanus*, which is consistent across the replicate runs (Fig. S3, Supporting information). Additionally, the BP&P species delimitation analyses demonstrated high support for the topology considering the three *L. americanus* clusters separately (posterior probability >0.99; Table S4, Supporting information) but not when sequences were randomly assigned to clusters (not shown). The *BEAST analysis taking these three *L. americanus* clusters into account recovered a monophyletic *L. americanus* (posterior probability >0.99) (Fig. 2).

Our divergence estimates suggest that *L. americanus* split from the other two species around 2.7 Ma, while *L. townsendii* and *L. californicus* diverged at 1.2 Ma (Fig. 2). Within *L. americanus*, the Boreal group was estimated to have diverged 2.0 Ma, while the Rockies and Pacific Northwest groups split much more recently, at about 0.36 Ma (Fig. 2).

The mtDNA phylogeny does not conform to that of the nuclear DNA, as *L. americanus* is not recovered as monophyletic (Fig. 3) given that one group from the Pacific Northwest population cluster (here named PacNW2) is more closely related to *L. californicus*. The remaining *L. americanus* form a distinct clade, classified as Boreal, Rockies and Pacific Northwest (PacNW1) based on the microsatellite groups of Cheng *et al.* (2014). The discordant clade (PacNW2) does not include *L. californicus* haplotypes. This result agrees with the observations of Cheng *et al.* (2014), suggesting either the occurrence of mtDNA introgression from *L. californicus* into *L. americanus* or incomplete mtDNA lineage sorting in the evolution of these species.

### *Isolation-with-migration and demographic analyses*

IMa2 was used to quantify gene flow among species/clusters (see parameter estimates in Table 3; Tables S5–S7, Supporting information). Among species, gene flow was only significant from *L. americanus* into *L. townsendii* and *L. californicus*, but at very low levels unlikely to affect phylogenetic reconstruction (Eckert & Carstens 2008). In general, parameter estimates did not differ much when considering *L. americanus* as a single evolutionary unit (Table S5, Supporting information). Within *L. americanus*, the three population clusters were found to be remarkably isolated. Nuclear gene flow was only suggested as significant from the Boreal into the Pacific Northwest population and from this to the Rockies population but, again, at very low levels. Estimates of divergence among populations according to the IM model (Table 3) were consistent with those inferred with *BEAST (Fig. 2). *L. californicus* was the species suggested to have the highest effective population size ($N_e$) and *L. townsendii* the lowest. However, the Pacific Northwest and Rockies populations of *L. americanus* have the lowest $N_e$ across all analysed populations (Table 3).

The EBSP analysis does not suggest any drastic shifts in population size for the species and intraspecific clusters analysed here, particularly if the large confidence intervals of the inference are taken into account (Fig. 4).

### *Coalescent simulations*

The effective population sizes and divergence times obtained from IM analysis of the two population

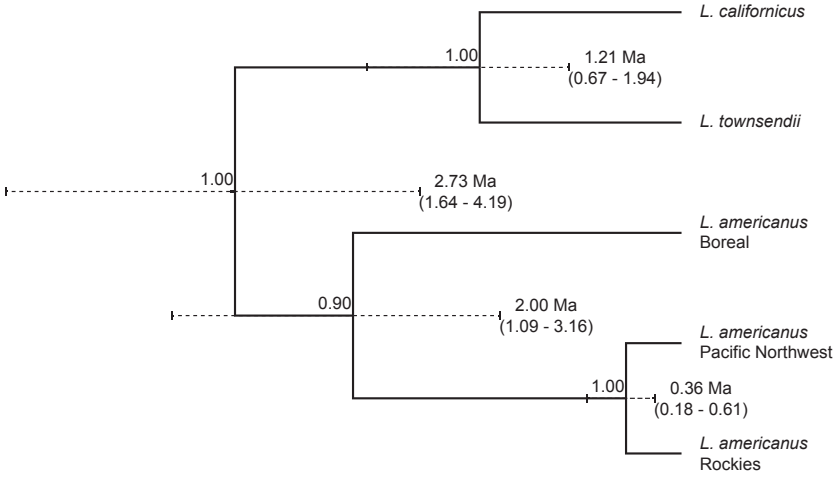


**Fig. 2** Species tree of *L. californicus*, *L. townsendii* and *L. americanus* inferred with *BEAST, considering the partition of the latter into three discrete populations (Boreal, Rockies and Pacific Northwest). Numbers above branches indicate the posterior probabilities and dashed line the 95% confidence intervals of node ages (mean value and 95% CI are indicated next to the line). The tree was calibrated using a substitution rate of $1.65 \times 10^{-9}$ substitutions/site/year for the PRKCI fragment.

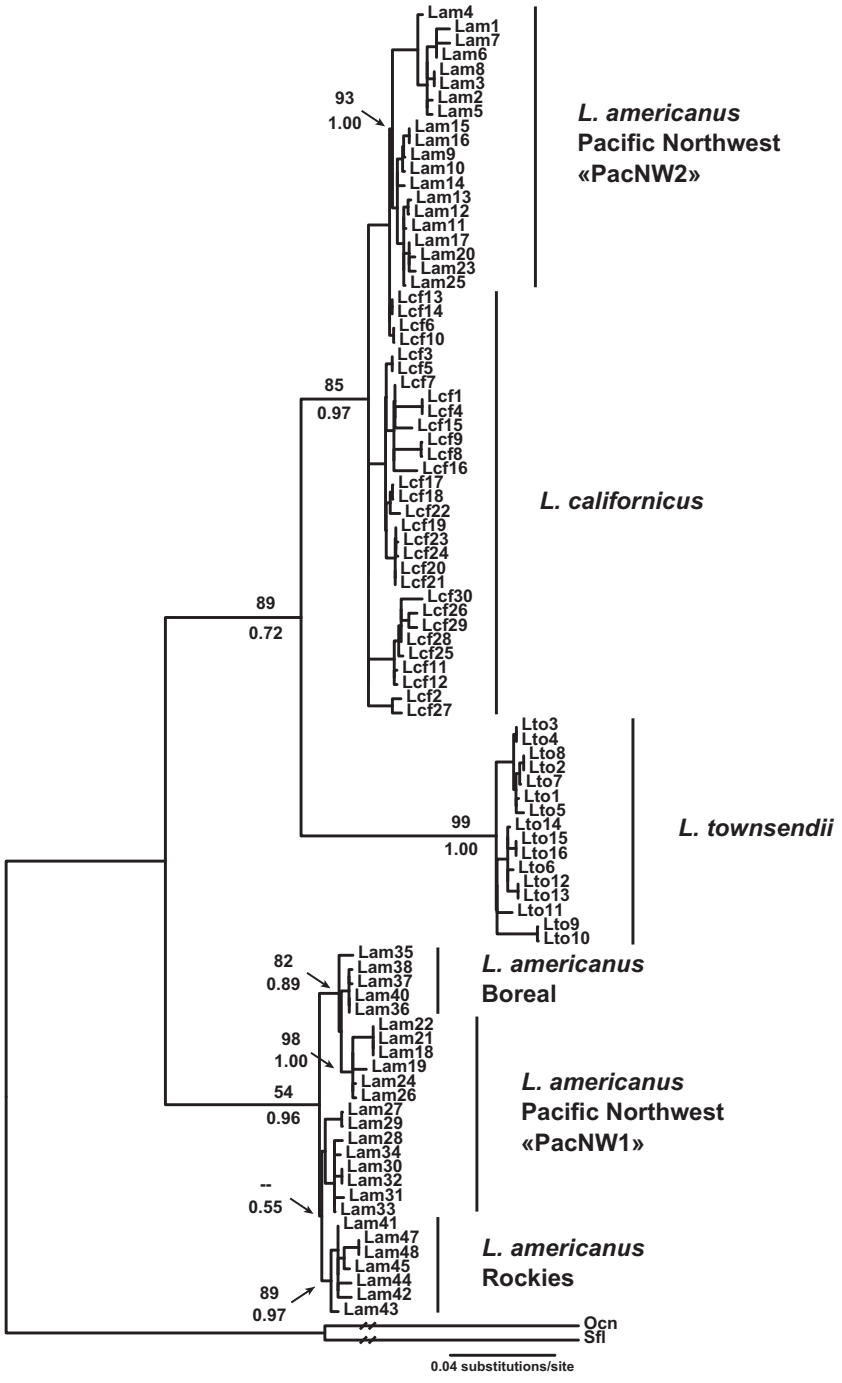


**Fig. 3** Cytochrome *b* Bayesian inference phylogeny of *L. californicus*, *L. townsendii* and *L. americanus*. Sequences from *Oryctolagus cuniculus* (Ocn) and *Sylvilagus floridanus* (Sfl) were used as outgroups. Maximum-likelihood bootstrap supports and BI posterior probabilities of the most relevant clades are shown above and below branches, respectively (if bootstrap support was higher than 50% or posterior probability higher than 0.5). See specimen codes and GenBank accession numbers in Table S1 (Supporting information).

clusters possibly involved in the mtDNA introgression events – *L. americanus* Pacific Northwest and *L. californicus* – were used to simulate cytochrome *b* data sets under a model with no gene flow. The observed pairwise distances between *L. americanus* PacNW1 cluster and *L. californicus* were found to lie within the range of distances expected under a strict lineage sorting scenario (the same was found for the Boreal and Rockies groups). On the contrary, all pairwise distances between *L. americanus* PacNW2 and *L. californicus* fell below the 5th percentile of the minimum distances simulated assuming no gene flow, suggesting introgression (Fig. 5). The same results were obtained maximizing the probability of retention of ancestral polymorphism (Fig. S6, Supporting information). The geographic distribution of mtDNA introgression is shown in Fig. 1.

**Table 3** ML estimates (95% posterior density intervals in parentheses) of demographic parameters obtained with IMa2 between pairs of populations

| Pop. 1 | Pop. 2 | $N_{e1}$† | $N_{e2}$† | $N_{eA}$† | $t$‡ | $2Nm_1$§ | $2Nm_2$§ | ABCDD¶ | ABC0D¶ | ABCD0¶ | ABC00¶ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Lam-Bor | Lam-Roc | 226 647 | 38 216 | — | 2 007 684 | — | — | n.s. | n.s. | n.s. | n.s. |
| | | (127 346; 404 666) | (15 347; 83 353) | (0; –) | (986 991; –) | (0.0000; 0.3085) | (0.0000; –) | | | | |
| Lam-Bor | Lam-PacNW | 225 443 | 77 936 | — | 2 421 739 | — | 0.0176* | n.s. | n.s. | * | * |
| | | (128 429; 394 435) | (47 845; 120 605) | (0; –) | (948 474; –) | (0.0000; 0.2761) | (0.0003; 0.1591) | | | | |
| Lam-Roc | Lam-PacNW | 22 111 | 69 968 | 196 315 | — | 0.0904* | — | n.s. | * | n.s. | * |
| | | (7673; 50 108) | (39 263; 114 190) | (0; –) | — | (0; –) | (0.0000; 0.3504) | | | | |
| Lam-Bor | Lca | 265 404 | 574 742 | 207 629 | 2 972 528 | — | — | n.s. | n.s. | n.s. | * |
| | | (152 261; 455 581) | (442 340; 739 641) | — | (1 240 238; –) | (0.0000; 0.6103) | (0.0000; 0.2688) | | | | |
| Lam-Roc | Lca | 60 038 | 582 686 | 363 381 | 3 429 914 | — | — | n.s. | n.s. | n.s. | n.s. |
| | | (26 637; 117 813) | (453 655; 744 336) | — | (1 452 080; –) | (0.0000; 0.1759) | (0.0000; 0.1036) | | | | |
| Lam-PacNW | Lca | 110 579 | 597 129 | 309 217 | 2 520 919 | — | — | n.s. | n.s. | n.s. | n.s. |
| | | (69 968; 164 779) | (465 330; 763 233) | (4062; 747 946) | (1 537 780; 4 311 946) | (0.0000; 0.0934) | (0.0000; 0.1094) | | | | |
| Lam-Bor | Lto | 320 050 | 175 612 | — | — | — | 0.0301* | n.s. | n.s. | * | * |
| | | (179 223; 568 242) | (115 105; 262 275) | — | — | (0.0000; 0.5460) | (0.0000; 0.2029) | | | | |
| Lam-Roc | Lto | 37 470 | 174 649 | — | — | 0.0068 | 0.0263* | n.s. | n.s. | * | * |
| | | (148 89; 78 093) | (114 190; 260 469) | — | — | (0.0000; 0.0629) | (0.0018; 0.1421) | | | | |
| Lam-PacNW | Lto | 78 983 | 181 871 | — | — | — | 0.0312* | n.s. | n.s. | * | * |
| | | (47 388; 123 253) | (120 485; 268 534) | — | — | (0.0000; 0.0877) | (0.0000; 0.1687) | | | | |
| Lca | Lto | 641 424 | 228 813 | 264 923 | 1 357 714 | 0.0033 | 0.0012 | n.s. | n.s. | n.s. | n.s. |
| | | (491 570; 830 998) | (152 984; 334 494) | (91 622; 550 187) | (856 997; 2 166 565) | (0.0000; 0.4184) | (0.0000; 0.2241) | | | | |

Lam-Bor: *L. americanus*, Boreal; Lam-Roc: *L. americanus*, Rockies; Lam-PacNW: *L. americanus*, Pacific Northwest; Lca: *L. californicus*; Lto: *L. townsendii*; Missing values correspond to cases where parameters could not be reliably estimated; a substitution rate of $3.45 \times 10^{-9}$ substitutions/site/generation was estimated.

†Effective population size of population 1 ($N_{e1}$), 2 ($N_{e2}$), and ancestral population ($N_{eA}$).

‡Time in years as populations 1 and 2 split.

§Population migration rate into population 1 ($2Nm_1$) and population 2 ($2Nm_2$) (*significant values, $P < 0.05$; Nielsen & Wakeley 2001).

¶Likelihood ratio test of nested models with equal gene flow between populations (ABCDD), no gene flow into population 1 (ABC0D), no gene flow into population 2 (ABCD0) and with no gene flow (ABC00).

The test statistic was calculated as follows: ABCDD (2LLR against ABCDE) follows a chi-square distribution with 1 degree of freedom with critical value *$P < 0.05$ at 2LLR > 3.84; ABC0D and ABCD0 (2LLR against ABCDE) and ABC00 (2LLR against ABCDD) follow a chi-square distribution that is 1/2 × chi-square(1) + 1/2 × chi-square(0) with critical value *$P < 0.05$ at 2LLR > 2.70.

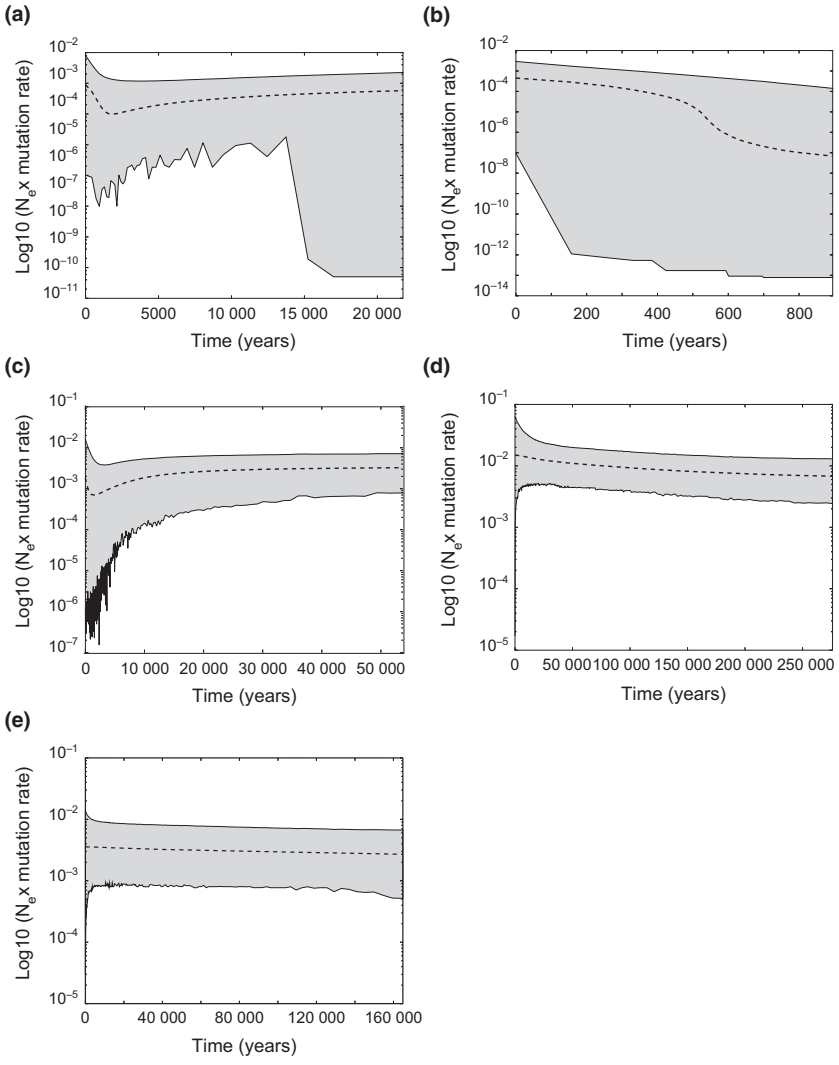


**Fig. 4** Demographic profiles of *L. americanus* Boreal (a), *L. americanus* Rockies (b), *L. americanus* Pacific Northwest (c), *L. californicus* (d) and *L. townsendii* (e), based on Extended Bayesian Skyline Plot analyses. The last 10% of the time points are not shown, except for plot b (see full plots in Fig. S5, Supporting information). Time is in units of years before the present (calibrated using a mtDNA substitution rate of $1.8 \times 10^{-8}$ substitutions/site/year).

## Discussion

### *Speciation history of North American hares suggests cryptic divergence*

Understanding the relative importance of introgression in the evolution of North American hares requires estimating the most relevant parameters of their history of speciation. The *BEAST phylogeny suggest that *L. americanus* diverged from the common ancestor of the three focal species at around 2.7 Ma and that the jackrabbits, *L. californicus* and *L. townsendii*, diverged 1.2 Ma (Fig. 2). These estimates are generally consistent with those obtained from the IM analyses (2.4–3.1 and 1.4 Ma, respectively; Table 3) and are more recent than previous estimates based on a molecular supermatrix (4.8 Ma for the stem divergence of *L. americanus*; Matthee *et al*. 2004) or mtDNA (5.6 Ma for the stem divergence of *L. americanus*; Wu *et al*. 2005). Interestingly, our analysis suggests that the Pacific Northwest and Rockies populations of *L. americanus* may have diverged ~360 kya, which is consistent with the fragmentation of the Western forest of the Pacific coast and Rocky Mountains (see Weir & Schluter 2004); however, the Boreal population diverged from the other two at a deeper evolutionary timescale (2.0 Ma; Fig. 2). This estimate roughly places the event of fragmentation and divergence in the same period of the split between *L. townsendii* and *L. californicus*, which conforms to the presumed time frame of speciation events in North American mammals (Arbogast & Kenagy 2001; Demboski & Cook 2001) and birds (Weir & Schluter 2004), and may thus have resulted from common environment-driven fragmentation pressures (Weir & Schluter 2004). The unexpected depth of the snowshoe hare's intraspecific divergence suggests that genetic isolation among groups arose from historic processes and not from recent geographic fragmentation. In addition, the extremely limited levels of gene flow inferred

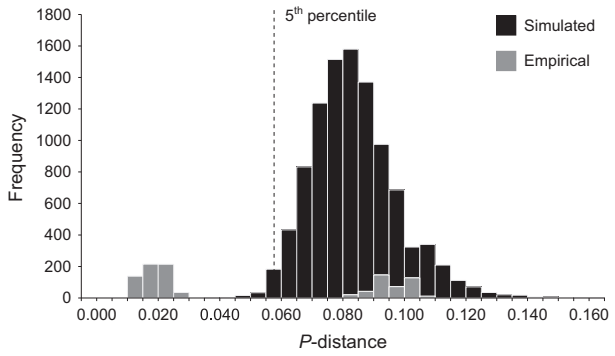


**Fig. 5** Empirical (grey bars) and simulated (black bars) mtDNA distances between *L. californicus* and the Pacific Northwest cluster of *L. americanus*. Simulations were performed under the assumption of no gene flow and a cytochrome *b* substitution rate of $3.6 \times 10^{-8}$ substitutions/site/generation. Vertical line indicates the 5th percentile of the distribution of simulated distances.

between the three *L. americanus* genetic groups using the IM framework (Table 3) suggest that some degree of reproductive isolation may exist. Interestingly, Nagorsen (1985) found no indication of morphological distinctiveness of the Boreal snowshoe hares or conformation to subspecific classifications, and it thus appears that we detected cryptic divergence within *L. americanus*. Although it would be useful to perform coalescent-based analyses with extended sampling of the Boreal group to confirm levels of divergence, we note that Cheng *et al.* (2014) showed that genetic variation within the Boreal group is homogeneous and thus our sampling may adequately describe genetic variation in that group. Whether or not the inferred divergence and low levels of gene flow justify a taxonomic revision of *L. americanus* must be addressed with an integrative analysis including data from multiple genetic and non-genetic sources.

### *Demographic history*

Although *L. americanus* has the largest distribution among the North American hares, *L. californicus* has the largest effective population size among the three species (Table 3), which could reflect different evolutionary histories. Northern *L. americanus* is likely to have been more susceptible to demographic and geographical oscillations due to the repeated advance and retreat of glaciers throughout the Pleistocene (Hewitt 2004). The lower estimated effective population sizes of the Pacific Northwest and Rockies clusters may reflect an increased susceptibility of peripheral populations to demographic fluctuations (Cheng *et al.* 2014; see Eckert *et al.* 2008 for a review on the central–marginal hypothesis) (Table 3; Fig. 4). The large distribution of *L. californicus* may have been less affected by climatic oscillations, allowing the species to maintain larger population sizes (Fig. 4). Our EBSP does not suggest changes in population sizes through time for any species or cluster. This may be due to the relatively small sample size in this work for such inferences, particularly in the Boreal snowshoe hare group. Indeed, using only mtDNA but a larger sample size, Cheng *et al.* (2014) inferred a Late Pleistocene demographic expansion of the Boreal group.

Little is known about the population history of *L. townsendii*. Our results suggest that this species has the lowest long-term effective population size among the three studied species (Table 3), but no dramatic shift of population size through time was inferred (Fig. 4). However, fossils suggest that over the past few thousand years, this species may have been excluded from some southern regions due to competitive exclusion by *L. californicus* (Lim 1987 and references therein). In addition, *L. townsendii* may have disappeared from some areas due to land use and habitat fragmentation (Berger 2008; but see Gunther *et al.* 2009). Our analysis suggests that gene flow from *L. americanus* into *L. townsendii* has occurred since the divergence of these species (Table 3), although it is difficult to assess whether this corresponds to recent introgression in populations of *L. townsendii*. No evidence of gene flow was found from this species to/from *L. californicus* contrary to the suggestion that these species hybridize in nature (Flux 1983).

### *Extensive mtDNA introgression from L. californicus into L. americanus*

Even though nuclear gene flow among the three North American species seems rare or absent (Table 3), previous results of Cheng *et al.* (2014) suggested that mtDNA introgression might have occurred between *L. californicus* and *L. americanus*, considering the sharing of mtDNA lineages visible in the mtDNA phylogeny (seen also in this work; Fig. 3). This contrasts with the monophyly of *L. americanus* that we estimate for nuclear DNA (Fig. 2). Our coalescent simulations show that the genetic similarity between the PacNW2 mtDNA haplotypes of *L. americanus* and *L. californicus* is incompatible with simple incomplete lineage sorting (contrary to the remaining divergences to *L. californicus*, which are within expectations: PacNW1 $D_{XY} = 0.096$; Boreal $D_{XY} = 0.101$; Rockies $D_{XY} = 0.098$; see Fig. 5). However, *L. californicus* and *L. americanus* PacNW2 do not share mtDNA haplotypes, which could result from (i) ancient introgression, (ii) introgression of an extant but unsampled *L. californicus* haplogroup or (iii) introgression from another species not included in this study.

We aligned all cytochrome *b* haplotypes of the three species included in this study and other species available at GenBank to our data set (Fig. S4, Supporting information). The position of the PacNW2 clade is maintained closer to *L. californicus* in this extended phylogeny, suggesting that introgression was likely ancient and of *L. californicus* origin. We estimated that the split between *L. californicus* and PacNW2 mtDNA occurred 470 000 years ago (200 000–906 000 95% HPD), which can thus indicate the time of introgression.

Historical and ongoing gene introgression has been found among other North American mammals (Good *et al.* 2008; Chavez *et al.* 2011), sometimes with massive mtDNA introgression or 'capture' (Good *et al.* 2008) and little nuclear DNA introgression as found in this work. This may have resulted from the competitive replacement of resident *L. californicus* by invading *L. americanus* during the Pleistocene glaciations, a situation that is expected to lead to introgression into the genome of the invading species (Currat *et al.* 2008; Excoffier & Ray 2008). These two species have different habitat requirements, *L. americanus* inhabiting for example dense boreal forest and *L. californicus* being distributed in southern open arid regions, and glacial cycles would have differentially shifted these distinct habitats. This competitive replacement model predicts that introgression should prevail for markers transmitted by the least-dispersing sex, which are often females in mammals. However, whether this explains massive introgression of mtDNA into *L. americanus* is at this point uncertain. In addition, there is no evidence of sex-biased dispersal in this species (Burton *et al.* 2002). The asymmetric direction of introgression would also be favoured by mechanisms that induce sex-biased matings, such as female choice and frequency-dependent assortative matings, among others (Wirtz 1999; Chan & Levin 2005). Alternatively, mtDNA introgression into *L. americanus* may have been favoured by natural selection. Adaptive introgression of mtDNA has been hypothesized in several species (Ruiz-Pesini *et al.* 2004; Ropiquet & Hassanin 2006), including in hares (Melo-Ferreira *et al.* 2011, 2014), and could explain the pattern observed here if the *L. californicus* mtDNA type is advantageous in the *L. americanus* nuclear background. However, separating the relative contributions of selective and demographic processes to interspecific gene flow is a major challenge and should be the object of future research.

### *Conclusions and future prospects*

Our results uncover hidden evolutionary processes in the North American hares as follows: (i) deep cryptic divergence exists within *L. americanus*, (ii) nuclear gene flow occurred from *L. americanus* into *L. townsendii* and *L. californicus* and (iii) extensive mtDNA introgression occurred from *L. californicus* into the Pacific Northwest populations of *L. americanus*. Introgression is a source of genetic novelty and may set the conditions for adaptation if the introgressed variants underlie favoured phenotypes (reviewed by Arnold & Martin 2009). For example, introgression has been shown to enhance abiotic tolerance in sunflowers (Whitney *et al.* 2010), to induce poison resistance in mice (Song *et al.* 2011) and to generate adaptive wing colour variation in butterflies (Pardo-Diaz *et al.* 2012). It is striking that contrary to the general trend in *L. americanus*, which undergoes seasonal coat colour changes from a brown coat in the summer to a white winter coat, part of the Pacific Northwest group retains their summer coat year-round, mimicking the phenotype of *L. californicus*. The dramatic snow pack decrease caused by global warming and the increased tendency of seasonally changing hares to become more mismatched against a snow-free background (Mills *et al.* 2013) may confer a significant adaptive advantage to the trait present in the Pacific Northwest populations. Although other evolutionary mechanisms can underlie this phenotype, hybridization may have contributed to the retention of the summer coat year-round if introgression affected genomic regions involved in seasonal coat colour change. Although speculative at present, this hypothesis opens new perspectives in the study of the impact of global warming to the survival of boreal species undergoing seasonal coat colour change and deserves further investigation.

## Acknowledgements

This work was supported by Portuguese national funds through the FCT (Fundação para a Ciência e a Tecnologia) and by FEDER funds (Fundo Europeu de Desenvolvimento Regional) through the COMPETE program (PTDC/BIA-EVF/115069/2009 research project), by Programa Operacional Potencial Humano-Quadro de Referência Estratégico Nacional (POPH-QREN) funds from the European Social Fund and Portuguese Ministério da Educação e Ciência (FCT, SFRH/BPD/43264/2008 postdoc grant to J.M.F., PhD grant SFRH/BD/87126/2012 to F.A.S. and SFRH/BSAB/1278/2012 sabbatical grant to PCA), by FLAD (Luso-American Foundation), and by project 'Genomics and Evolutionary Biology' co-financed by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF). North American funding support came from grants to LSM from the National Science Foundation (USA) (DEB 0817078) and the U.S. National Park Service (USA). We thank Kerry A. Gunther and Janet Rachlow for the collection of *L. townsendii* samples; Armando Geraldes, Kris Hennings, Buck Bauer and Robert Dowler for providing *L. californicus* samples; and Christy Cheyne, Lee Simons, James

MacCracken, Don Gordon, Mark Weber and Nate Berg for providing *L. americanus* samples. We also thank the University of Washington Burke Museum for providing *L. californicus* samples (Museum accession numbers Lca1 - UWBM 79850, Lca2 - UWBM 79852, Lca3 - UWBM 79853, Lca4 - UWBM 78698, Lca5 - UWBM 81660, Lca11 - UWBM 78721; see Table S1, Supporting Information, for sample correspondence). We thank Pierre Boursot, Lisette Waits, Link Olson and two anonymous reviewers for their valuable comments on this work.

## References

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

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

Alves PC, Melo-Ferreira J, Freitas H, Boursot P (2008) The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. *Philosophical Transactions of the Royal Society B*, **363**, 2831–2839.

Arbogast BS, Kenagy GJ (2001) Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, **28**, 819–825.

Arnold ML, Martin NH (2009) Adaptation by introgression. *Journal of Biology*, **8**, 82.

Berger J (2008) Undetected species losses, food webs, and ecological baselines: a cautionary tale from the Greater Yellowstone Ecosystem, USA. *Oryx*, **42**, 139–142.

Boonstra R (2013) Reality as the leading cause of stress: rethinking the impact of chronic stress in nature. *Functional Ecology*, **27**, 11–23.

Bossu CM, Near TJ (2009) Gene trees reveal repeated instances of mitochondrial DNA introgression in orangethroat darters (Percidae: Etheostoma). *Systematic Biology*, **58**, 114–129.

Burton C, Krebs CJ, Taylor EB (2002) Population genetic structure of the cyclic snowshoe hare (*Lepus americanus*) in southwestern Yukon, Canada. *Molecular Ecology*, **11**, 1689–1701.

Chan KM, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720–729.

Chavez AS, Saltzberg CJ, Kenagy GJ (2011) Genetic and phenotypic variation across a hybrid zone between ecologically divergent tree squirrels (*Tamiasciurus*). *Molecular Ecology*, **20**, 3350–3366.

Cheng E, Hodges K, Melo-Ferreira J, Alves P, Mills LS (2014) Conservation implications of the evolutionary history and genetic diversity hotspots of the snowshoe hare. *Molecular Ecology*, **23**, 2929–2942.

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

Demboski JR, Cook JA (2001) Phylogeography of the dusky shrew, *Sorex monticolus* (Insectivora, Soricidae): insight into deep and shallow history in northwestern North America. *Molecular Ecology*, **10**, 1227–1240.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PloS Biology*, **4**, e88.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUTI and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Eckert AJ, Carstens BC (2008) Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Molecular Phylogenetics and Evolution*, **49**, 832–842.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Edwards SV (2009) Is a new and general theory of molecular systematics emerging? *Evolution*, **63**, 1–19.

Edwards SV, Liu L, Pearl DK (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5936–5941.

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, **23**, 347–351.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Flot JF (2010) SEQPHASE: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162–166.

Flot JF, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes*, **6**, 627–638.

Flux JEC (1983) Introduction to taxonomic problems in hares. *Acta Zoologica Fennica*, **174**, 7–10.

Garrick RC, Sunnucks P, Dyer RJ (2010) Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology*, **10**, 118.

Gerard D, Gibbs HL, Kubatko L (2011) Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC Evolutionary Biology*, **11**, 291.

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

Griffin PC, Mills LS (2009) Sinks without borders: snowshoe hare dynamics in a complex landscape. *Oikos*, **118**, 1487–1498.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Gunther KA, Renkin RA, Halfpenny JC *et al.* (2009) Presence and distribution of white-tailed jackrabbits in Yellowstone National Park. *Yellowstone Science*, **17**, 24–32.

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.

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

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B-Biological Sciences*, **359**, 183–195.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.

Krebs CJ (2011) Of lemmings and snowshoe hares: the ecology of northern Canada. *Proceedings of the Royal Society B-Biological Sciences*, **278**, 481–489.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, **56**, 17–24.

Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, **20**, 2485–2487.

Lewis CW, Hodges KE, Koehler GM, Mills LS (2011) Influence of stand and landscape features on snowshoe hare abundance in fragmented forests. *Journal of Mammalogy*, **92**, 561–567.

Lim BK (1987) *Lepus townsendii*. *Mammalian Species*, **288**, 1–6.

Liu J, Yu L, Arnold ML *et al.* (2011) Reticulate evolution: frequent introgressive hybridization among Chinese hares (genus *Lepus*) revealed by analyses of multiple mitochondrial and nuclear DNA loci. *BMC Evolutionary Biology*, **11**, 223.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–237.

Marboutin E, Peroux R (1995) Survival pattern of European hare in a decreasing population. *Journal of Applied Ecology*, **32**, 809–816.

Matthee CA, van Vuuren BJ, Bell D, Robinson TJ (2004) A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Systematic Biology*, **53**, 433–447.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P (2009) The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. *Molecular Ecology*, **18**, 2643–2658.

Melo-Ferreira J, Alves PC, Rocha J, Ferrand N, Boursot P (2011) Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? *Evolution*, **65**, 1956–1968.

Melo-Ferreira J, Boursot P, Carneiro M *et al.* (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology*, **61**, 367–381.

Melo-Ferreira J, Vilela J, Fonseca MM *et al.* (2014) The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biology and Evolution*, **6**, 886–896.

Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theoretical Population Biology*, **75**, 35–45.

Mills LS, Zimova M, Oyler J *et al.* (2013) Camouflage mismatch in seasonal coat color due to decreased snow duration. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 7360–7365.

Nagorsen DW (1985) A morphometric study of geographic variation in the snowshoe hare (*Lepus americanus*). *Canadian Journal of Zoology*, **63**, 567–579.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Pardo-Diaz C, Salazar C, Baxter SW *et al.* (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, **8**, e1002752.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review in Ecology, Evolution and Systematics*, **41**, 215–230.

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

Rambaut A, Drummond AJ (2007) TRACER v1.4. Available from http://beast.bio.ed.ac.uk/tracer.

Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645–1656.

Ropiquet A, Hassanin A (2006) Hybrid origin of the Pliocene ancestor of wild goats. *Molecular Phylogenetics and Evolution*, **41**, 395–404.

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, **303**, 223–226.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.

Seehausen O (2013) Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology*, **26**, 279–281.

Song Y, Endepols S, Klemann N *et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, **21**, 1296–1301.

Spinks PQ, Shaffer HB (2009) Conflicting mitochondrial and nuclear phylogenies for the widely disjunct Emys (Testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Systematic Biology*, **58**, 1–20.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics*, **76**, 449–462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Thulin C-G, Stone J, Tegelström H, Walker CW (2006a) Species assignment and hybrid identification among Scandinavian hares *Lepus europaeus* and *L. timidus*. *Wildlife Biology*, **12**, 29–38.

Thulin CG, Fang M, Averianov AO (2006b) Introgression from *Lepus europaeus* to *L. timidus* in Russia revealed by mitochondrial single nucleotide polymorphisms and nuclear microsatellites. *Hereditas*, **143**, 68–76.

Tyson R, Haines S, Hodges KE (2010) Modelling the Canada lynx and snowshoe hare population cycle: the role of specialist predators. *Theoretical Ecology*, **3**, 97–111.

Wallner B, Huber S, Achmann R (2001) Non-invasive PCR sexing of rabbits (*Oryctolagus cuniculus*) and hares (*Lepus europaeus*). *Mammalian Biology*, **66**, 190–192.

Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society of London B-Biological Sciences*, **271**, 1881–1887.

Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist*, **187**, 230–239.

Wirtz P (1999) Mother species-father species: unidirectional hybridization in animals with female choice. *Animal Behaviour*, **58**, 1–12.

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.

Wu C, Wu J, Bunch TD *et al.* (2005) Molecular phylogenetics and biogeography of *Lepus* in Eastern Asia based on mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **37**, 45–61.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9264–9269.

Zimova M, Mills LS, Lukacs PM, Mitchell MS (2014) Snowshoe hares display limited phenotypic plasticity to mismatch in seasonal camouflage. *Proceedings of the Royal Society B-Biological Sciences*, **281**, 20140029.

Zwickl DJ (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD Dissertation, University of Texas at Austin, Austin, Texas.

J.M.-F., P.C.A., L.S.M. and E.C. conceived the work. F.A.S. performed the laboratory work and data analyses under J.M.-F. supervision. J.M.-F. and F.A.S. drafted the manuscript. All authors interpreted the results, revised and approved the final version of the manuscript.

## Data accessibility

Newly obtained DNA sequences were deposited in GenBank with accession numbers KM260760–KM261521. Sequences with the following GenBank accession numbers were also used in the main analysis: AJ001588, AY292724, AY785433, HM028196, HM028509, JN036940, JN036996, JN037024, JN037052, JN037078, KF781358, KF781404, KF781408, KF781413 and KF781423. Sequence alignments were deposited in Dryad: doi:10.5061/dryad.21f62.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Table S1** Detailed information of sequences obtained per individual with GenBank accession numbers.

**Table S2** Geographic coordinates of sampling sites.

**Table S3** Analysed loci, PCR conditions and primers.

**Table S4** Posterior probabilities for models of taxa delimitation estimated using BP&P.

**Table S5** ML estimates of demographic parameters obtained with IMa2 between pairs of species.

**Table S6** ML estimates of demographic parameters obtained with IMa2 among the three species.

**Table S7** ML estimates of demographic parameters obtained with IMa2 among the three *L. americanus* groups.

**Fig. S1** Individual phylogenies of nuclear loci generated from the outputs of BEAST.

**Fig. S2** Individual phylogenies of nuclear loci inferred using Garli.

**Fig. S3** Species tree of *L. californicus*, *L. townsendii*, and *L. americanus* inferred with *BEAST.

**Fig. S4** Cytochrome *b* phylogenies of all North American hare species and one sequence representative of each non-North American hare species available in GenBank.

**Fig. S5** Demographic profiles of *L. americanus* population clusters based on Extended Bayesian Skyline Plot analyses.

**Fig. S6** Empirical and simulated mtDNA distances between *L. californicus* and the PacNW2 group of *L. americanus* (simulations used highest 95% HPD estimate of effective population sizes and the lowest 95% HPD estimate of divergence time inferred using IMa2).