

The Speciation History of North American Hares (*Lepus* spp.): Divergence with gene flow?

Fernando Seixas

Mestrado em Biodiversidade, Genética e Evolução,
Departamento de Biologia
Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO)
2012

Orientador
José Melo Ferreira

Coorientador
Paulo Célio Alves

Agradecimentos

A realização deste trabalho deveu-se ao apoio de várias pessoas que para ele contribuíram diretamente ou indiretamente e às quais quero expressar os meus agradecimentos.

Em primeiro lugar gostaria de agradecer aos meus orientadores por todo o apoio, pela ajuda prestada ao longo de todo o meu mestrado e por me terem permitido integrar este projeto.

Ao meu orientador, Dr. José Melo Ferreira, agradeço pela sua paciência em me ensinar, à sua dedicação e empenho na minha formação, e ao facto de se mostrar sempre disponível em me ajudar e guiar ao longo deste trabalho. Um muito obrigado por seres um orientador tão espetacular.

Ao meu co-orientador, Dr. Paulo Célio Alves, pela palavra amiga e pelo entusiasmo demonstrado desde o momento que integrei a equipa das lebres e ao longo de todo este trabalho.

To Scott Mills and Ellen Cheng for providing the samples and for their previous work on the system as well as their scientific input.

Às minhas companheiras da fantástica equipa das lebres, Liliana Farelo e Joana Vilela. Obrigado pela vossa ajuda dentro e fora do laboratório, pela vossa atenção e preocupação constantes.

A todos os colegas do CIBIO e CTM, de laboratório, da biblioteca, das brincadeiras e das longas viagens de metro, pelo seu espírito animado e de grande companheirismo. Obrigado por me fazerem sentir tão integrado.

A todos os meus amigos pela sua amizade e companheirismo. À Susana Neto, pela sua infindável amizade, pelas longas conversas e por ser um apoio constante desde que nos conhecemos. À Catarina Vila-Pouca por me continuar a aturar e por me ter feito crescer. À Filipa Sampaio, pelo seu apoio e constante simpatia. À Teresa Tavares pela sua duradoura amizade. Ao Bruno Rabaçal, o meu parceiro do “crime” e asa para todas as ocasiões. À Vanessa Mata pelas longas discussões, constante disponibilidade em ajudar, debater ou simplesmente conversar. À Joana Nogueira, a minha tranquilizadora.

Por último, gostaria de agradecer à minha família, pela vossa infindável paciência para com todas as minhas manias e maluqueiras, por estarem presentes em todos os momentos e pelo vosso carinho. Em suma, obrigado por serem o meu apoio de todos os dias.

Este trabalho foi financiado pela Fundação para a Ciência e a Tecnologia, através do projecto de investigação com referência PTDC/BIA-EVF/115069/2009 (co-financiado pelo programa COMPETE e pelo Fundo Europeu de Desenvolvimento Regional) e da bolsa de pós-doutoramento atribuída ao meu orientador José Melo Ferreira com referência SFRH/BPD/43264/2008 (co-financiado pelo Fundo Social Europeu).

Resumo

A perspetiva tradicional que perdurou durante grande parte do século XX, em resultado da influência do Conceito Biológico de Espécie, de que o fluxo génico entre espécies animais é um fenómeno raro tem sido contestada por estudos recentes que mostram que frequentemente as espécies partilham uma história de hibridação introgressiva. Apesar de recorrentemente ser considerada como uma menos valia, dificultando a inferência da história das espécies, a introgressão pode ser a única testemunha de processos relevantes do passado das espécies como adaptação, alterações de distribuição das espécies ou de competição entre espécies com substituição de uma delas por outra mais apta. No entanto, as inferências de introgressão são frequentemente assumidas sem que se considerem fatores alternativos que podem levar aos mesmos padrões gerados por introgressão. Particularmente complicada é a distinção entre introgressão secundária e a coalescência incompleta das linhagens entre espécies próximas, precisamente a escala mais relevante para o estudo da especiação.

As lebres (género *Lepus*) são um modelo promissor para a realização de tais estudos. Este género sofreu uma radiação recente e explosiva á cerca de 5-7 milhões de anos gerando mais de 30 espécies atualmente aceites, com distribuições por todo o globo. Além disso, inúmeros casos de fluxo génico entre as espécies foram descritos. No entanto, estes estudos encontram-se enviesados para as espécies Euroasiáticas sendo que pouco se conhece sobre as espécies do Norte da América, a região da qual se pensa que o género radiou. Relativamente às lebres desta região, existem descrições de hibridação na natureza entre a lebre-de-cauda-negra (*L. californicus*) e a lebre-de-cauda-branca (*L. townsendii*) no entanto sem nunca se ter avaliado as repercussões deste fenómeno na história evolutiva e no genoma destas espécies. Um estudo recente descreveu uma terceira espécie, a lebre americana (*L. americanus*), como contendo haplótipos mitocondriais que são filogeneticamente mais relacionados com *L. californicus* do que com outros haplótipos de *L. americanus*, tendo sido sugerido que tal padrão poderia ter resultado de introgressão mitocondrial. No entanto, é também possível que a segregação incompleta das linhagens tenha gerado este padrão

Neste estudo investigamos a história evolutiva destas três lebres do Norte da América – *L. americanus*, *L. californicus* e *L. townsendii* – através da análise de múltiplos marcadores. Nove marcadores (cinco autossómicos, um mitocondrial, um do cromossoma Y e dois do X) foram sequenciados perfazendo um total de 5633 pares

de bases analisadas. A reconstrução filogenética dos marcadores nucleares resultante do método coalescente para múltiplas espécies e marcadores sugere que a *L. americanus* foi a primeira espécie a divergir há cerca de 3 milhões de anos, seguida da divergência de *L. californicus* e *L. townsendii* há cerca de 2 milhões de anos. Aplicando um modelo de isolamento com migração aos nossos dados, encontramos evidências de fluxo génico nuclear entre *L. californicus* e *L. americanus*, no entanto em níveis muito reduzidos, sendo pouco provável que influenciem a reconstrução filogenética. Finalmente, a hipótese de introgressão massiva do ADN mitocondrial de *L. californicus* para *L. americanus* da zona do Noroeste Pacifico dos Estados Unidos da América foi testada através da modelação da divergência esperada na ausência de fluxo génico. Os resultados demonstram que a distância genética observada entre as *L. americanus* do Noroeste Pacifico e *L. californicus* é menor do que a esperada de acordo com um modelo de estrita coalescência, o que suporta a hipótese de introgressão massiva do DNA mitocondrial. Estimou-se que esta introgressão terá ocorrido há 680 milhares de anos e que terá levado a uma substituição massiva do ADN mitocondrial do grupo de *L. americanus* do Noroeste do Pacifico. Como a hibridação e introgressão devem ter ocorrido durante períodos de alterações da distribuição das espécies devido aos avanços e recuos dos glaciares durante o Pleistoceno, é possível que a introgressão tenha resultado de competição com substituição das espécies, tal como já foi proposto noutros casos de introgressão massiva do ADN mitocondrial em lebres. No entanto, se a introgressão massiva do ADN mitocondrial resultou de fenómenos demográficos neutrais ou seleção natural é uma questão ainda em aberto.

Este trabalho sugere que a reticulação do genoma é um fenómeno ubíquo entre as lebres, não estando restrito a espécies ou regiões, afetando particularmente o ADN mitocondrial. Estes resultados reforçam a noção de que as barreiras entre as espécies são porosas e que o fluxo génico é comum no processo da especiação. Na situação particular deste estudo, a população de *L. americanus* potencialmente mais afetada pela introgressão de ADN mitocondrial de *L. californicus* apresenta um fenótipo variável de muda sazonal da cor da pelagem, o que contrasta com a norma da espécie. Se tal é uma simples coincidência ou foi originado por novas combinações resultantes de evolução reticulada é uma questão interessante para se abordar em estudos futuros.

Abstract

The traditional perspective that prevailed for most of the XX century, following Mayr's Biological Species Concept, that gene flow between animal species is rare has been challenged by modern studies that have shown that species often share a history of introgressive hybridization. Although it can often be seen as a confounding parameter, hindering the reconstruction of species relationships, gene introgression can witness relevant events on the history of species, such as adaptation, past range expansions/retractions or competitive replacement of species. However, the inference of introgressive hybridization is often over-simplified, misled by other confounding factors. It is particularly challenging to disentangle the relative contribution of secondary introgression and incomplete lineage sorting in the study of speciation between closely related taxa, the scale that is the most relevant to understand speciation.

Hares (genus *Lepus*) are promising models to undertake such studies. This genus has recently experienced a rapid and explosive radiation *ca*. 5-7 million years ago (Mya) with over 30 species currently being accepted, which are spread all over the world. Moreover, numerous instances of interspecific gene flow have been reported. However, studies within the genus are biased towards Eurasian species little, being known about the North American hares, the region from where the genus is thought to have radiated. Among hares from this region, the black-tailed jackrabbit (*L. californicus*) and the white-tailed jackrabbit (*L. townsendii*) have been reported to hybridize in the wild but the repercussions of this phenomenon on the evolutionary history and genome of these species have never been assessed. Also, a recent study has showed that a third American species, the snowshoe hare (*L. americanus*), harbors mitochondrial DNA (mtDNA) haplotypes that are phylogenetically more closely related to *L. californicus*, and suggested this could be due to mtDNA introgression. However, that incomplete lineage sorting originated this pattern could not be discarded.

In this study we have investigated the evolutionary history of these three North American hares - *L. americanus*, *L. californicus* and *L. townsendii* – using a multilocus approach. Sequence variation at nine DNA markers from all inheritance compartments (five from autosomes, one mitochondrial, one from the Y and two from the X-chromosomes) was sampled, resulting in a total of 5633 bp analysed. The phylogenetic reconstruction from a multispecies-multilocus coalescent-based approach of the nuclear DNA markers suggests that *L. americanus* split first at about 3 million years

ago, and then *L californicus* and *L. townsendii* diverged at about 2 million years ago. Using an isolation-with-migration model to our data, evidences of nuclear gene flow were suggested to have occurred possibly between *L. californicus* and *L. americanus*, though at very limited levels, unlikely to have affected the phylogenetic reconstruction. Finally, the hypothesis of massive introgressive hybridization of the mtDNA of *L. californicus* into *L. americanus* of the Pacific Northwest (PacNW) region of the United States of America was tested by explicitly modeling the expectations of divergence with no gene flow. The results show that the empirical divergence between the PacNW *L. americanus* and *L. californicus* is lower than expected under the strict lineage sorting model, thus supporting the massive mtDNA introgression hypothesis. This introgression was estimated to be old (around 680 thousand years ago; kya), and to have led to the massive replacement of the mtDNA of *L. americanus* in the PacNW population cluster. Since hybridization and introgression must have occurred during periods of important range shifts of species due to the glacial oscillations of the Pleistocene, introgression may have resulted from the competitive replacement of species ranges with hybridization, as has been proposed in other cases of massive mtDNA introgression among hares. However, whether the massive nature of mtDNA introgression among hares results from neutral demography or from natural selection remains an open question.

This work suggests that genome reticulation among hares is a ubiquitous phenomenon, not restricted to some species or regions, and particularly affects the mtDNA. These results thus reinforce the notion of porous species boundaries and the common nature of gene flow in the process of speciation. In the particular situation studied here, the *L. americanus* population that was likely most affected by introgression of *L. californicus* origin mtDNA presents a variable seasonal coat color change phenotype, which contrasts with the norm in the species. Whether this is a coincidence or originated from novel combinations resulting from reticulate evolution is an interesting question that should be addressed in future studies.

Index

List of Figures, Tables and Appendixes

INTRODUCTION

Speciation and Gene Flow

Understanding the basis of speciation has always been a major challenge for biologists. The divergence of species being a continuous process, the first endeavour is establishing the definition of species itself as a non-discrete constantly evolving entity. However, several concepts of species have been proposed from different and within several areas of biology (De Queiroz 1998; Coyne and Orr 2004) differing in the biological properties that are of interest to the different biologists (see De Queiroz 2005). As a result, depending on which concept is adopted, different species may be recognized, particularly at the early stages of divergence, where the species concepts tend to disagree the most.

Perhaps the most widely adopted definition of species, the biological species concept (Dobzhansky 1937; Mayr 1942), defines species as “groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups” (Mayr 1942). According to this concept, reproductive isolation is considered fundamental for speciation to occur because gene flow is expected to rapidly breakdown linkage disequilibrium and to homogenize gene pools, preventing the formation of genetically distinct groups. The BSC assumes that the genomes of species behave as cohesive units due to co-adapted genes involved in speciation. Moreover, during the process of speciation, isolating mechanisms are developed that maintain this cohesion of the genome.

The reproductive barriers underlying speciation under the BSC are traditionally considered intrinsic biological factors rather than purely extrinsic, as the latter, *per se*, do not assure species status – two populations geographically separated do not necessarily constitute distinct species. Nevertheless, Mayr (1942) considered that most speciation tends to occur only after effective geographical isolation, even if only temporary. Following Mayr, allopatric speciation was considered for many years the most common and plausible mode of speciation, and taken as the null hypothesis (Coyne and Orr, 2004). During this process, the establishment of a physical barrier preventing gene flow between populations eventually allows the development of reproductive isolation as a result of divergence through genetic drift or selection. The

modes of allopatric speciation can be divided, depending on whether the ancestral population is divided equally in two populations - vicariant - or a small peripheral population derives from it - peripatric - and whether reproductive isolation was attained in full allopatry or it is completed after secondary contact - alloparapatric (Coyne and Orr, 2004). Sympatric speciation represents another mode of geographic speciation according to which species emerge from an initial population where mating is random and without geographic segregation of the daughter populations. These two contexts of geographic speciation can be considered the extremes of a continuum of initial levels of gene flow among diverging populations, gene flow [m] being 0 in allopatry and 0.5 in sympatry (Gavrilets 2003). Parapatric speciation encompasses the intermediate cases in which gene flow decreases in neighbouring populations causing them to diverge [$0 < m < 0.5$].

Traditionally, because geographical structure of populations is easily observed and given the growing realization that spatial structure of populations is very important, speciation is classified regarding its geographical context – thus allopatric, sympatric or parapatric (e.g. Gavrilets, 2003). However, speciation is a process that generally extends over several generations (Coyne and Orr, 2004). Considering the duration of the process, it is more plausible to assume that speciation has different stages, each one at a different spatial context rather than occurring in strict allopatry or sympatry (Butlin et al, 2008). For example, the Quaternary climatic oscillations are known to have forced different taxa to retract, expand, displace and/or fragment their ranges, inducing considerable shifts on their distributions and this would have occurred repeatedly (e.g. Hewitt, 2000, 2004). The alternate glacial and interglacial periods would thus provide the chance for periods of differentiation in allopatry and periods of secondary contact (Hewitt 2004). During the latter, if complete reproductive isolation was not yet achieved, hybridization between units would be possible to occur, which could result in the exchange of genetic variants between the hybridizing populations.

Understanding the conditions under which new species can arise remains a central question in Biology. One of the most debated topics is the geographical context of speciation, more precisely, the frequency of speciation that occurs in the face of homogenizing gene flow. During most of the XXth century, following the traditional view of the Biological Species Concept, hybridization was thought to play a minor role in speciation, and was generally considered as a breakdown mechanism of reproductive isolation (Dobzhansky 1937; Mayr 1963). When occurring, hybridization was explained via species range changes and environmental disturbance, but mostly as a result of

human habitat disturbance (Mayr 1963). Moreover, since F1 hybrids are generally less viable, less fertile or sterile, and even backcrossed animals tend to be less fit due to the breakdown of “internally balanced chromosome sections”, they would be easily eliminated and thus introgression was thought to rarely occur (Mayr 1963), a vision that was broadly shared among zoologists. On the contrary, being more common in plant species botanists tended to consider hybridization and introgression to be common during speciation and important in adaptive evolution (see Arnold, 1997). While recognizing the importance of allopolyploidy in plants, Dobzhansky thought that hybridization should tend to occur less in animals: sexual reproduction and the complexity of tissues and organ systems (each depending in a larger number of genes) when compared to plants, would increase the importance of reproductive isolation in animals, thus being more constrained to hybridize (see Dowling and Secor, 1997; reference therein).

However, the advances in molecular genetics and the consequent increasing amount of genetic data on natural populations have been challenging the notion that gene flow between distinct animal species is rare, and showing that closely related animal species often share a history of introgressive hybridization (e.g. Good *et al.* 2008; Sequeira *et al.* 2011; Melo-Ferreira *et al.* 2012). Instead, hybridization may potentially be an important, transient phase before the completion of reproductive isolation in animals. Indeed, Mallet (2005) has estimated that on average 25% of plant and 10% of animal species hybridize with at least one other closely related species, which may thus be potentially affected by introgression (although Schwenk et al (2008) presented a more conservative hybridization rate of approximately 1% in animals). A recent work by Pinho and Hey (2010) reviewed studies that assessed gene flow between recently diverged taxa, and suggest that nonzero gene flow between sister species is often reported, although no gene flow is generally the rule. The increasing number of studies reporting gene flow between sister taxa has led to the view of speciation as a process of divergence with gene-flow, in which the genetic factors that underlie speciation may continue to accumulate despite incomplete reproductive isolation. Interestingly, numerous studies have been showing that when sister species hybridize, patterns of introgression are not homogeneous along the genome (see Avise, 2004; Baack and Rieseberg, 2007). These studies consistently found that different portions of the genome vary in their permeability to foreign alleles, some showing extensive gene flow, whereas others remain differentiated, which lead to the recognition that species borders are often semipermeable or porous (Wu 2001).

It has been hypothesized that genomic regions associated to differential adaptation and/or reproductive isolation are expected to show little or no introgression while unlinked regions will be more easily exchanged between hybridizing species (Rieseberg et al, 1999; Wu, 2001). Accordingly, the ease of introgression across the genome will depend on a selection/recombination/dispersal relationship (Barton 2001; Wu 2001). Furthermore, it is also likely that alleles which can either confer advantage in the alternative habitat or interact positively in the foreign genetic background will be more prone to introgress. If disadvantageous in the foreign species, they will contribute to further differentiation (see Wu, 2001). In fact, recent studies confirmed lower levels of introgression in regions of low levels of recombination such as centromeric regions compared to the higher recombinant telomeric regions (e.g. Carneiro et al., 2009) and regions adjacent to chromosomal breakpoints (Machado et al., 2007; Yatabe et al., 2007). However, fewer evidences have been gathered on linking patterns of introgression to the fitness effects, although some studies show reduced introgression in X-linked loci in mice compared to autosomal loci which is consistent with the greater abundance of species incompatibilities on sex chromosomes (Macholán et al., 2007; Payseur et al, 2004).

Different patterns of introgression can occur not only among genomic compartments but also among the involved species, which can e.g. result from sexual asymmetries in hybridization. Haldane's rule, sex-biased dispersal, behavioral asymmetry and density-dependant assortative mating can lead to differential introgression of male and female specific markers (see e.g. Avise, 2004 and references therein), especially when prezygotic isolation models like female-preference or male-male competition are implicated (Chan and Levin, 2005).

Extensive introgression and capture of mtDNA versus nuclear loci appears to be common among hybridizing animal species (Avise, 2004). This led to the assumption that mitochondrial DNA might be more likely to introgress than nuclear DNA (e.g. Ballard and Whitlock, 2004; Funk and Omland, 2003). Traditionally, the ease of mitochondrial introgression was explained by the mtDNA neutrality and its independent segregation resulting in less subjection to indirect selection acting over speciation genes relative to linked nuclear genes (Funk and Omland, 2003 and references therein). However, loosely linked nuclear loci should also be relatively free to introgress, and if positively selected introgression may be enhanced (see Chan and Levin, 2005). Moreover, mitochondrial and nuclear genomes interact in important physiological functions (e.g. oxidative phosphorylation, OXPHOS) and likely co-evolve

(Blier et al, 2001; Rand et al, 2004). Also, despite being for long considered a neutral marker, mtDNA may indeed be under various sorts of selective pressures, which could even drive to massive mtDNA introgression (Ballard and Whitlock, 2004; Galtier et al, 2009). Some studies have hypothesized that adaptive introgression of mtDNA might have occurred in response to environmental factors, such as altitude (Ropiquet and Hassanin, 2006) and temperature (Doiron et al, 2002). Such adaptive value of mtDNA haplotypes in response to climatic factors has been also suggested in humans (Ruiz-Pesini et al, 2004).

However, massive introgression of mtDNA may also result from purely neutral phenomena. It has been suggested that the replacement of a resident species by an invading one with hybridization at the front of invasion might also lead to massive introgression, through a purely demographic and drift process. At the edge of the wave, introgressed alleles may rapidly be fixed (or lost) by drift in the invading species, in a process coined “allele surfing on an expansion wave” (Currat et al, 2008; Excoffier and Ray, 2008). This is particularly relevant for mtDNA as the surfing hypothesis predicts that the prevalence of introgression should be higher for markers transmitted by the least dispersing sex. In case of female-biased philopatry, as often in mammals, introgression of the maternally transmitted mtDNA tends to be more massive. In addition, the reduced effective size of mtDNA when compared to nuclear DNA may also explain mtDNA biased introgression as it can lead to faster fixation (or loss) of the introgressed alleles due to stronger genetic drift (Takahata and Slatkin, 1984).

Nonetheless, the significance of numerous studies that report interspecific mtDNA capture should be regarded considering two factors: 1) until recently mtDNA has been the marker of choice for resolving phylogenies, particularly between closely related species (Avise, 2009); 2) the absence of recombination on mtDNA increases the chances of detecting mtDNA introgression (Funk and Omland 2003; Avise 2004). These factors may be contributing to the higher number of reports of mtDNA relative to nuclear introgression.

Inferring introgression is not straightforward and several factors may confound interpretations of introgressive hybridization. Interspecific gene flow has been often inferred from polyphyly at a given locus (see Funk and Omland, 2003) and incongruence among gene trees built from independent loci (e.g. Sang and Zhong, 2000). However, other biological phenomena such as phylogenetic error, recombination, gene duplication or loss, horizontal gene transfer and incomplete

lineage sorting may lead to these same patterns (see Funk and Omland, 2003; Linder and Rieseberg, 2004). Phylogenetic error is an artifact of phylogenetic reconstruction that results from insufficient variation in the data either because of the evolution rate of a gene being too slow or the fragment analyzed from the gene being too small. On the other hand, fast evolution rate may lead to homoplasy, which also results in inaccurate gene trees. Both these causes of gene tree incongruence may be identified by assessing statistical confidence of a given clade/tree topology and using more complex models of sequence evolution that account for homoplasy (Funk and Omland, 2003). Recombination is expected to affect gene trees inferences, particularly when analyzing large DNA fragments or at genome level studies. This, however, may be corrected by using approaches with a sliding window along the alignment (Hobolth et al, 2011) or reducing the fragments to the largest non-recombining blocks when leading with individual loci. Gene duplication at a given loci may also lead to a polyphyletic gene tree and gene trees incongruence as it reflects the history of duplication at that locus and generally can be detected from genetic analysis with adequate sampling (Small et al, 2004). For horizontal gene transfer completion, genes to be transferred require a vector or other means of transfer but also to be incorporated in the receiving genome to become functional but animals seem to be largely unaffected (Andersson 2005). Finally, the coalescence of sampled alleles into an ancestral allele predating the speciation event is referred to as incomplete lineage sorting or deep coalescence and can potentially affect any single locus in any taxa. It results from the random sorting of alleles between taxa during the process of speciation. At early stages of speciation, incipient species are expected to have alleles inherited from the ancestral population, and some of them may be more closely related to those of its own daughter species (exclusive) while others are more closely related to those in the other species (shared). This results in polyphyletic gene trees that may be different among loci. As speciation moves forward, some alleles are expected to be lost by drift and novel alleles to arise from mutations to the point that eventually intraspecific variation results from post-speciation mutation from only one of the ancestral lineages. At this stage, sorting is complete and species are reciprocally monophyletic at any of the analyzed loci. Because of its lower N_e, mtDNA is expected to achieve complete sorting more rapidly than nuclear loci (Moore 1995) and thus is often assumed to better reflect species phylogeny, one reason for mtDNA being widely used to resolve relationships among closely related taxa. However, for taxa resulting from rapid radiations, incomplete lineage sorting may have major effects on sorting of mtDNA alleles (see Funk and Omland, 2003).

Because all these factors may reflect similar topological patterns to those resulting from introgression, it is necessary to previously eliminate these potential causes before accepting that gene tree discordance results from gene-specific introgression. While the first four factors may easily accounted for, disentangling the relative contribution of lineage sorting and secondary introgression is a major challenge. This is particularly difficult for closely related taxa, the level that is the most important for the study of speciation. Nevertheless, recent methodological advances and creative approaches provide us the power to assess the influence of these confounding factors (e.g. Meng and Kubatko 2009; Joly *et al.* 2009; Gerard *et al.* 2011).

Reticulate evolution in genus Lepus

Hares and jackrabbits (genus *Lepus*) provide an excellent opportunity to study speciation. These belong to the most speciose and widespread leporid genus. The genus has diverged from rabbits *ca* 11.8 Mya and radiated likely from North America *ca* 5-7 Mya (Matthee et al, 2004) forming over 30 species currently distributed all over the world (Alves and Hackländer, 2008). This radiation was suggested to have resulted from the development of grasslands at that period and the formation of the west Antarctic ice sheet (Matthee et al, 2004; Yamada et al, 2002). These species currently occur in a great variety of habitats with some species having restricted distribution ranges - e.g. *L. castroviejoi* occurs in the Cantabrian Mountains in Northeast Spain - while others occur over vast areas - e.g. *L timidus* occurs in all northern Europe and Asia - (Alves and Hackländer, 2008).

Numerous cases of interspecific gene flow have been described in the genus (reviewed in Alves et al, 2008). Most often, instances of introgressive hybridization relate to areas of present contact of species. In Russia, *L. timidus* and *L. europaeus* distributions meet to form a natural hybrid zone with introgression being detected in both directions (Thulin et al, 2006). Current hybridization between these two species has also been reported in Sweden and Denmark where the latter has been introduced by man (Thulin *et al.* 1997; Thulin and Tegelström 2002; Fredsted *et al.* 2006). In Corsica, the frequent introductions of *L. granatensis* and *L. europaeus* are also suggested to have led to the introgression into the native *L. corsicanus* (Pietri et al. 2011). However, instances of introgression were also reported between currently allopatric species. Although *L. timidus* went locally extinct in the Iberian Peninsula, a large portion of haplotypes found

in *L. granatensis, L. castroviejoi* and *L. europaeus* are of *L. timidus* origin (Alves et al 2003; Melo-Ferreira et al, 2005). Also, introgression from *L. timidus* into *L. corsicanus* has been reported, the latter inhabiting the Italic Peninsula where *L. timidus* is currently absent (Alves et al., 2008; Melo-Ferreira et al., 2012).

Remarkably, mtDNA introgression can be massive in some regions. For example, in *L. corsicanus* and its sister species *L. castroviejoi*, introgression from *L. timidus* was suggested to have led to the complete replacement of the original mtDNA of these species (Alves et al., 2008; Melo-Ferreira et al., 2012) and in Iberia the frequency of the introgressed haplotypes reaches quasi-fixation in the *L. europaeus* and in the northern populations of *L. granatensis* (Melo-Ferreira et al, 2005). In contrast, nuclear gene flow tends to be sporadic and geographically spread (Melo-Ferreira *et al.* 2009, 2012). However, we must take into consideration that the mtDNA has been the most studied locus while a minor portion of the nuclear genome was analyzed. Extending the range of nuclear loci analyzed could reveal other patterns of introgression within this genomic compartment. Indeed, two recent studies (Farelo, 2011; Melo-Ferreira et al, 2011) have found extensive gene flow to occur at an X chromosome gene. Several explanations have been hypothesized to explain such asymmetries of introgression among inheritance compartments (e.g. Melo-Ferreira et al., 2009). Sex-biased dispersal can lead to differential introgression of the sex-linked alleles. In hares, it is ambiguous whether, similarly to what happens in other mammals, males are more mobile than females with some studies reporting this trend while not found in others (e.g. Bray et al., 2007; Hamill et al., 2007), but such pattern would favour initial asymmetries in hybridization that could explain the increased mtDNA introgression compared to male transmitted parts of genome in which the continuous flow of native alleles would erase the traces of introgression. According to Chan and Levin (2005) when females of the rarer species often fail to encounter a conspecific male they tend to mate with a male from the other, more abundant, species. Being recurrent, this frequency-dependent female-biased assortative mating, could favour introgression of the maternally transmitted inheritance compartments. The prevalence of mtDNA introgression could also result from male competition. In Sweden, mating between females of the smaller species (*L. timidus*) and males from the larger (*L. europaeus*) was suggested to be favoured in accordance with the expectations suggested by Grant and Grant (1997) concerning mating preferences in crosses between different sized species (Thulin and Tegelström 2002; Thulin *et al.* 2006). However, the introgression of *L. timidus* mtDNA into the smaller *L. granatensis*, in Iberia, is not conformant with this hypothesis.

Interestingly, introgression also tends to be asymmetric in its direction, generally occurring from *L. timidus* into other hares (see Alves et al., 2008). By facilitating the introgression of advantageous lineages, natural selection could promote the asymmetry in the direction of introgression. Particularly, given that the mitochondrial metabolism is involved in thermoregulation and and studies suggest mtDNA evolution may be shaped by positive selection (Ballard and Whitlock, 2004), mtDNA introgression of the arctic/boreal *L. timidus* may confer a selective advantage related to cold. This could explain the northward increase in the frequency of introgression of *L. timidus* mtDNA was reported in the Iberian *L. granatensis* populations (Melo-Ferreira *et al.* 2005). However, a purely demographic neutral process could also be invoked to explain this pattern. During the climate warming, *L. timidus* has likely been replaced by *L. granatensis* has the north expansion of the latter would be favoured by the changing climate (Melo-Ferreira *et al.* 2007). If this competitive replacement has been accompanied by hybridization this could have led to the gradient of introgression observed in *L. granatensis*. Indeed, it has been shown that introgression could be a likely incidental outcome of the process of replacement of a resident species by an invading one, even if hybridization is rare at the front of invasion (Currat et al, 2008; Excoffier and Ray, 2008).

Hares in North America

The origin of genus *Lepus* has been suggested to have occurred in North America *ca* 11.8 Mya from where an ancestral lineage would have dispersed to other continents crossing the Bering Strait to Asia *ca* 7-5 Myr ago (Matthee et al, 2004). This geographical origin for the genus is also supported by recent fossil records (e.g. Lopez-Martinez, 2008). Despite the importance of North America to the understanding of the evolutionary history of hares, there is a substantial lack of knowledge about the evolution of species from this region. Indeed, few phylogenetic works have been done regarding hares from this continent using molecular markers.

Of the more than 30 hare species described in the genus, nine occur in the North American continent – *L. arcticus*, *L. othus*, *L. americanus*, *L. townsendii*, *L. californicus*, *L. insularis*, *L. alleni*, *L. callotis* and *L. flavigularis* (see Flux and Angermann, 1990). These species have very distinct distributions which greatly vary in their range. *L. articus* and *L. othus* are confined to the extreme north of the continent being replaced

at the south by *L. americanus,* as open tundra is replaced by boreal forest. At the south of the continent, the most widely distributed species is *L. californicus* which occurs at the south of the USA north border extending into Mexico while *L. insularis*, *L. alleni*, *L. callotis* and *L. flavigularis* occur in Mexico, *L. insularis* and *L. flavigularis* having the most restrict range of North American hares. *L. townsendii* is almost as widespread as *L. californicus* and occurs between these two geographical groups at the centre of North America.

Halanych et al (1999) tried to resolve the phylogenetic relationships between hare species from this region using a mitochondrial marker (cytochrome *b*). They analyzed 11 recognized species of *Lepus* predominantly from North America and showed that the North America *Lepus* group is not monophyletic. The resulting phylogeny showed the existence of an Arctic clade (high northern latitudes) including *L. timidus*, *L. othus*, *L. arcticus* and *L. townsendii* and a Western clade (middle and western Northern America) including *L. americanus*, *L. californicus*, *L. alleni* and *L. callotis*. The presence of Old World taxa at the base and within the arctic clade suggests that some hares invaded North America secondarily, perhaps via an Asian-American land connection (Bering Strait). However, relying in a single marker to reconstruct the relationship of species may often lead to estimate the phylogeny of that marker rather than the species evolutionary history. Moreover, Alves et al (2006) have suggested that the mitochondrial DNA is not a reliable marker for the reconstruction of hares phylogenies since introgression of mtDNA has been reported between several species. As recognized by the authors, their study was an initial estimate of the evolutionary relationships within North American *Lepus* species.

Indeed, studies focused mostly in Eurasia, show that gene flow is common between species of hares, which on the one hand complicates phylogenetic inferences, but on the other hand makes these species privileged models to understand the process of population divergence, speciation and the establishment of reproductive isolation. Also, given the large distribution areas of many species, which has likely shifted dramatically during the last glacial periods, hares can be studied to understand the impact of past glaciations on biodiversity and predict future population trends, relating it to global warming. The snowshoe (*Lepus americanus*) hare is taxonomically very distinct from other *Lepus* and has the largest area of distribution among the New World hares (Flux and Angermann, 1990). It occurs in boreal and mixed deciduous forests of North America, requiring fairly dense vegetation, which it uses as cover (Murray and Smith, 2008). Species distribution extends from Alaska to Newfoundland, and

penetrates far south into America down the Coastal Range, the Rockies, and the Appalachians, to mid California, northern New Mexico and Tennessee, respectively (Flux and Angermann, 1990) – Figure 1. Given the strong association with the boreal forest, which has greatly shifted its range throughout the glacial periods, and the large distribution area, the Snowshoe hare can be used to understand the major biogeographic impacts of Pleistocene climatic oscillations in North American biota. A recent work (Cheng 2010) used microsatellites and mitochondrial DNA sequences (cytochrome b and control region) to infer the population history of these species. The results show that the species is structured in three major evolutionary groups with well-defined geographic distributions: Boreal (Canada and northwestern border of United States), Rockies (central west United States) and PacNW (Pacific coast of United States). Burton et al (2002) had already reported high genetic distance between snowshoe hares from the Yukon region (Canada) and one snowshoe hare sample from Montana (USA). Interestingly, this pattern of population structure is similar to that inferred from other North American species (Arbogast and Kenagy, 2001), suggesting that common phenomena, such as glaciations, must have forced species to retreat and fragment their distributions into common pockets of suitable habitat, i.e. glacial refugia. Another key finding of the work of (Cheng 2010) is that the PacNW populations of the snowshoe hare possess a mitochondrial DNA lineage more closely related to the one currently present in another North American hare, *Lepus californicus,* the black-tailed jackrabbit, that to the remainder of lineages present in the Boreal and Rockies populations of the species. Although this can result from incomplete lineage sorting along the speciation history of these species, it may also result from secondary introgression, as often described among other species of hares.

Part of the distribution of the black-tailed jackrabbit overlaps with that of the snowshoe hare, which may have set the conditions for hybridization to occur. *Lepus californicus* is the most common jackrabbit in western United States. Generally, it occurs in arid regions and areas of short grass rangeland (Flux and Angermann, 1990). It has a wide distribution across Mexico and USA, extending from Washington and Idaho to Baja California in the West and from South Dakota to central Mexico in the East, and from the Pacific coast almost to the Mississippi (Flux and Angermann 1990; Flinders and Chapman 2003) – Figure 1. This species is currently expanding, out-competing the white-tailed jackrabbit (*L. townsendii*) in the northeast and the antelope jackrabbit (*L. alleni)* and white-sided jackrabbit (*L. callotis*) in the south (Flux and Angermann, 1990). Indeed, where the black-tailed jackrabbit and white-tailed jackrabbit co-exist, the former displaces the white-tailed jackrabbit onto higher, sparser vegetation and can out-

compete it in adaptability to a wider range of habitats (see Flux and Angermann, 1990; references therein). *Lepus californicus* has been reported to hybridize with *Lepus townsendii* (Flux 1983), however, no assessment of the genetic consequences of this hybridization is known to have been reported.

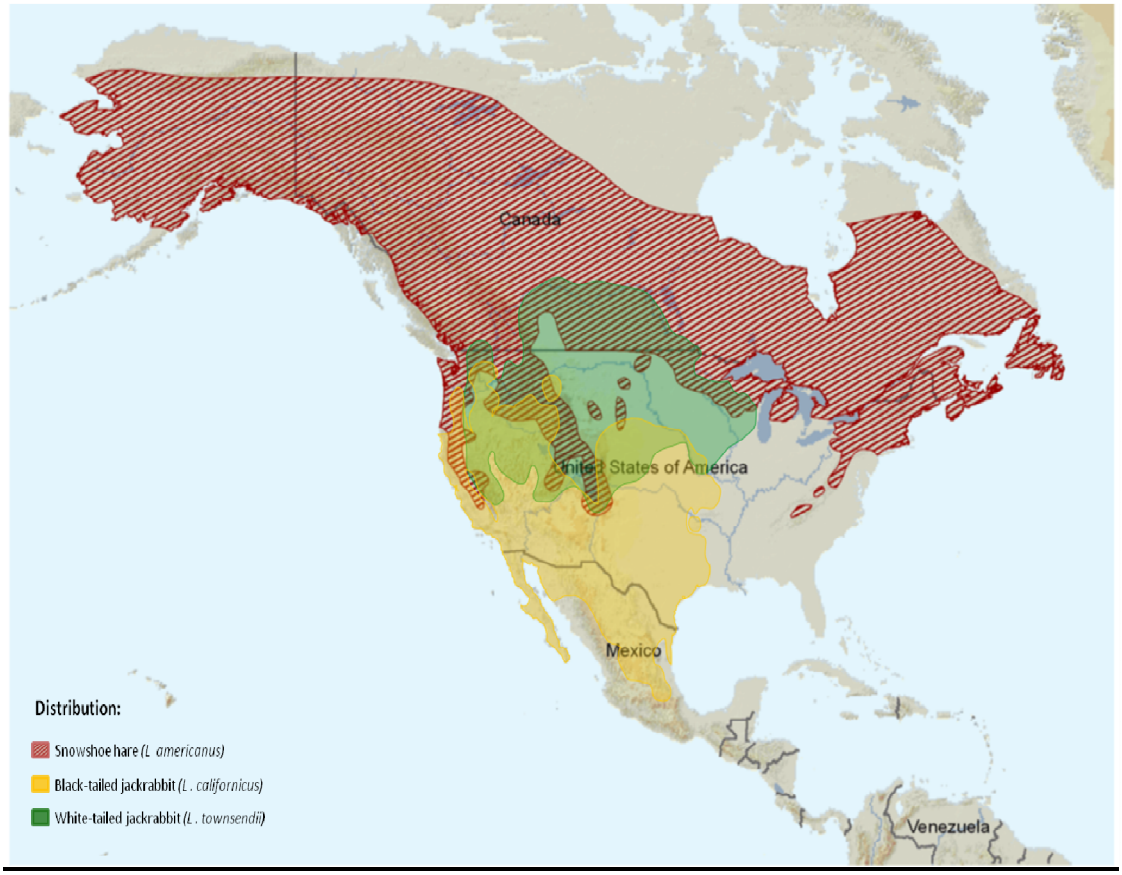


Figure 1 – Distribution of the three North American hare species analyzed in this study**.**

The white-tailed jackrabbit (*Lepus townsendii*) can be readily distinguished from *L. californicus* as the latter is characterized by its long ears and black upper surface on the tail that extends as a line through the back, while the former presents a white tail with a stripe on top (Flux and Angermann, 1990). This species typically inhabits open terrains such as plains and open prairie – (Flux and Angermann, 1990). Its present range extends from Midwestern states of the USA and southern Canada westward to the high mountain slopes of the Rockies, Cascades and Sierra Nevada (Flinders and Chapman 2003) - Figure 1. Yet its distribution has changed in historical times, extending to the northeast in the beginnings of the XXth century (in result human induced landscape transformations), while receding in the southwest in the half-past century - see e.g. Lim, 1987 ;Flinders and Chapman, 2003; references therein). In the Yellowstone National Park (YNP) and Grand Teton National Park (GTNP) the white-

tailed jackrabbits were considered extinct (Berger et al, 2006; 2008a) while recent sightings have been reported in the former (Gunther et al. 2009). The phylogenetic history of *L. townsendii* remains somewhat unclear. Mitochondrial DNA phylogenies suggest it is closely related with the Eurasian *L. timidus* and the American *L. arcticus* and *L. othus* and thus to have resulted from a re-colonization of North America from the Old World (Halanych *et al.* 1999). The analyses of nuclear markers are somewhat unclear, placing this species in a clade with *L. timidus* (Matthee *et al.* 2004), in accordance with mtDNA phylogenies, or not (Melo-Ferreira et al. 2012). This uncertainty questions whether the reflux from the Old World reflects the history of the species or of mtDNA alone, due to mtDNA introgression (Alves et al. 2008; Melo-Ferreira et al. 2012).

These three North American hares – *L. americanus*, *L. californicus* and *L. townsendii* – provide an invaluable model to study the eventual occurrence of gene flow in the divergence of species and the impact of incomplete lineage sorting on the genetic architecture of closely related taxa. This is the major goal of this work.

Objectives

Previous studies have reported multiple instances of reticulate evolution in hares, suggesting that introgressive hybridization may have played a major role in the evolution of some species within the genus *Lepus*. However, most of these studies focus on Eurasian species while few studies have considered their North American counterparts. Nevertheless, as hybridization in the wild has been reported between two of the North American species, *L. californicus and L. townsendii*, and that massive mtDNA introgression from the former to a cluster of another North American species, *L. americanus* it is likely that these species, similarly to other cases in hares, have undergone reticulated evolution, although this has never been investigated.

In this study we have propose to investigate the speciation history of these three North American species - *L. americanus*, *L. californicus* and *L. townsendii* - from the sequence variation at nine DNA markers from all inheritance compartments (autosomal, mitochondrial, X and Y).

With this aim, the following objectives were established:

- Reconstruct the branching pattern of the evolutionary tree of these three species;
- Estimate relevant population history parameters (e.g. effective population sizes, divergence times and gene flow);
- Test the hypothesis of a massive introgression of mitochondrial DNA of *L. californicus* into the *L. americanus* PacNW cluster against the alternative hypothesis of incomplete lineage sorting.

MATERIAL AND METHODS

Sampling and Data Collection

Sampling

Tissue sampling collection for genomic DNA extraction was done in the context of previous and ongoing projects, in the framework of the collaboration with the University of Montana (USA). Samples were available either at CIBIO or at the University of Montana.

Our sampling was based on the available samples of the three analysed American hares, and particularly focused in the regions near to or in the overlap of the ranges of distribution. It is nevertheless expected that it generally represent the extant variation in each species. For *L. americanus*, it was additionally taken into account the existence of three major population clusters described by Cheng (2010). Therefore, populations from of each of these clusters – Boreal, Rockies and Pacific Northwest – were included in this study. Sampled populations are depicted in Figure 2.

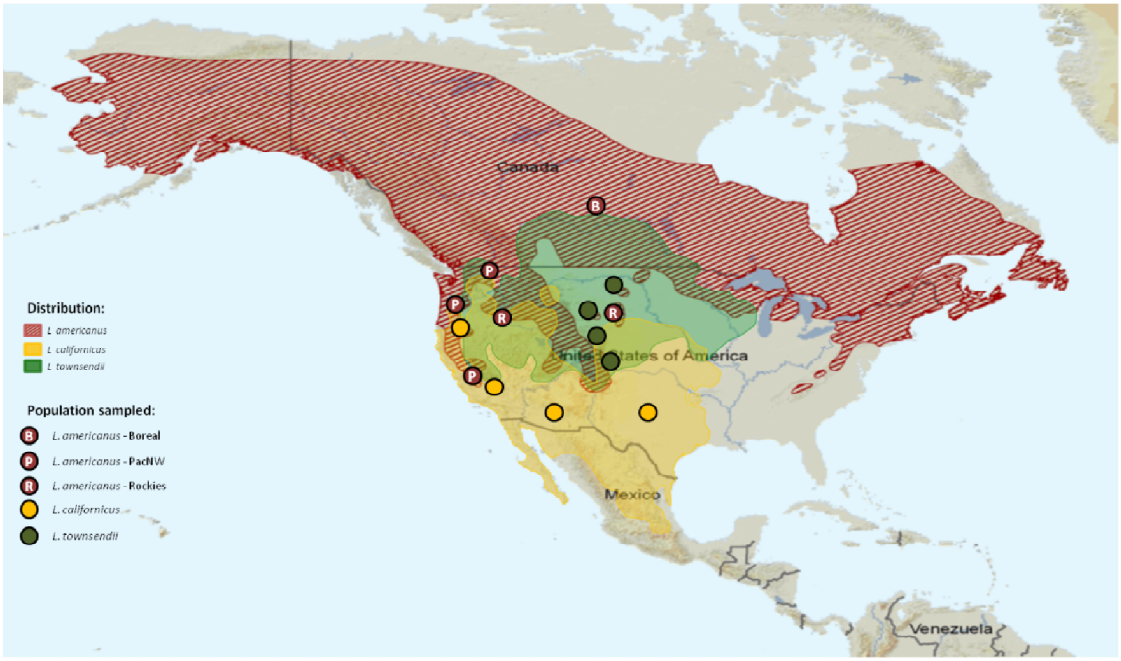


Figure 2 – Sampling Scheme. Clusters are those defined Cheng (2010) from microsatellite data.

The final dataset was composed of a total of 94 individuals - 48 *L. americanus*, 30 *L. californicus* and 16 *L. townsendii* - from 14 populations (Table 1; Appendix 1). Two *Sylvilagus audubonii* and one *Oryctolagus cuniculus* (downloaded from GenBank) were included as outgroups for some of the analyses.

Table 1 **-** Latin and common names of the species included in this study, cluster and population of the sampled specimens and number of specimens per population.

Species	Common Name	Cluster	Pop	# Samples
Lepus americanus	Snowshoe hare	Boreal	SK1	6
		Rockies	OR2	8
			WY1	8
		PacNW	WA4	10
			WA1	8
			CA1	8
Lepus californicus	Black-tailed Jackrabbit		Oregon	10
			California	6
			Arizona	6
			Texas	8
Lepus townsendii	White-tailed Jackrabbit		Montana	2
			Yellowstone	5
			GTP	1
			Wyoming	8
Sylvilagus audubonii	Desert cottontail			2
Oryctolagus cuniculus	European rabbit			1

Notes: *L. americanus* clusters based on Cheng (2010)**.**

DNA Extraction, PCR amplification and Sequencing

Lepus americanus and *L. californicus* DNA extractions were available at University of Montana. For *L. townsendii*, total genomic DNA was extracted from muscle or ear tissues using the JETQUICK Tissue DNA Kit (Genomed) according with manufacturer's instructions.

Nine molecular markers from the different genomic inheritance components (five autosomal, one mitochondrial, one from the Y and two from X chromosomes) were amplified by polymerase chain reaction (PCR) – (Table 2;Appendix 2**)**.

Purified PCR products were automatically sequenced (Macrogen Inc, Netherlands) with forward primers. Due to its larger length the SRY fragment was sequenced with forward and reverse primers and in some cases also with an internal primer (Appendix 2). GRIA 3 fragment was in some cases also sequenced with the reverse primer. Individuals presenting insertion-deletions in the nuclear markers were sequenced both with primers forward and reverse for that marker.

Table 2 - List of the loci analyzed in this study, chromosome location and position in the gene.

	Marker		Chromosome	Gene position
1	CYTB	Cytochrome *b*	mtDNA	exon
2	POLA1	Polymerase, alpha 1, catalytic subunit	X	intron
3	GRIA3	Glutamate receptor, ionotrophic, AMPA 3	X	intron
4	SRY	Sex determining region Y	Y	intron
5	SPTBN	Spectrin, beta, non-erythrocytic 1	autosome	intron
6	PRKCI	Protein kinase C, iota	autosome	intron
7	DARC	Duffy blood group, chemokine receptor	autosome	exon
8	KITLG	KIT ligand	autosome	intron
9	TF	Transferrin	autosome	intron/exon

Data treatment

All sequences were visually inspected in order to detect any base miscalling resulting from the automated reading. Sequences alignment was performed using ClustalW (Thompson et al. 1994) as implemented in BioEdit v7.0.5.3 (Hall, 1999). Polymorphic regions resulting from tandem repeats were removed from alignments and as immediately adjacent regions are known to be prone to sequencing reading errors this buffer was also extended 5bp on both sides of those regions.

For each nuclear dataset, allelic phase was determined using PHASEv2.1.1 (Stephens et al. 2001; Stephens and Scheet, 2005), which implements a Bayesian statistical method for reconstructing haplotypes from population genotype data. Input files were produced on the online software SeqPHASE (Flot, 2010). Two separate files were created: one with haplotype sequences of heterozygous individuals for which phases are known; the other with sequences of homozygous individuals and heterozygotes to be phased. Known haplotypes were defined from individuals with heterozygous insertion-deletions, following Flot et al. (2006) and incorporated in the analysis, since including them can considerably improve phase determination. As insertions-deletions cause discrepancy in the sequencer reading of the two DNA chains from the insertions/deletion onwards due to their different lengths (a pattern of double peaks appears from that point on in the chromatogram) this allows manually determining the phases of each polymorphic site relatively to other polymorphic sites and to the insertion-deletion. In such cases, individuals were sequenced with forward and reverse primers to deduce the complete sequences (see previous section). Five replicate runs

of 1000 iterations after an initial burn-in of 1000 generations were performed, with a thinning interval of 1. The run with the best average goodness of fit was retained.

PHASE has been shown to generate a very low number of false-positives, the majority of unresolved genotypes including an allele occurring only once in the dataset (Garrick et al. 2010). The complete dataset, including some low-probability calls, was thus kept since estimated levels of diversity, hence effective population sizes, would be considerably reduced if unresolved genotypes were excluded (Garrick et al, 2010).

Given that the downstream analysis rely on the assumption of no recombination within each locus, sequence datasets were reduced to the largest non-recombining blocks, using IMgc software (Woerner et al. 2007). IMgc determines the most data-rich recombination-filtered block, inferring recombination from violations of the four-gamete rule as defined by Hudson and Kaplan (1985). Species were analyzed separately and alignments then reconstructed.

Also, we tested each population for Hardy-Weinberg equilibrium in GENEPOP v1.4.1 (Rousset 2008). According to the Hardy-Weinberg model the expected genotype frequencies of a population can be calculated from the frequencies of the alleles observed from that same population if certain assumptions are met, among which random mating (Hardy 1908; Weinberg 1908). We have also tested the alternative hypotheses to Hardy-Weinberg equilibrium, heterozygote excess or deficiency, from the multilocus nuclear dataset.

Finally the multilocus test of selection, HKA (Hudson, Kreitman and Aguadé, 1987), was used as implemented in the software HKA (http://genfaculty.rutgers.edu/hey/software#HKA). This test relies on the prediction that under the neutral model of molecular evolution (Kimura 1968, 1969) the ratio of intraspecific polymorphism and interspecific divergence should be proportional to the effective population size (Ne). If all loci evolve under the neutral model, this ratio should be the same for all loci. By testing for homogeneity across two or more loci in the balance of fixed interspecific differences to intraspecific polymorphisms, the HKA test infers selection when there are significant deviations of expected variation between loci. For each hare species this test was both performed using the rabbit or each of the other hare species as outgroup.

Population Demography

Diversity and Demography

In order to assess the population diversity levels, nucleotide diversity (π) and ThetaS (θ_S) were calculated in Arlequin v.3.11 (Excoffier et al. 2005). Also, haplotype diversity was calculated in DnaSP v5.10 (Librado and Rozas 2009).

To infer the demographic history of the populations and species, mismatch distributions (MMD) analysis and neutrality tests (Tajima's D and Fu's Fs; Tajima 1989a; Fu 1997) were performed in Arlequin v.3.11. This was performed for all populations and also for each *L. americanus* cluster and each species. The MMD were only performed for the Cyt b fragment. The MMD were calculated and fitted to the Sudden Expansion Model (Rogers and Harpending, 1992). This model assumes that an initial population at equilibrium with $\theta = \theta_0$ grows rapidly to a new size with $\theta=\theta_1$, τ units of mutational time ago, where $\theta=N_e u$ and $\tau=2\mu t$ (N_e=effective population size, μ=mutation rate and t=time since the expansion in generations). Goodness-of-fit tests (Schneider and Excoffier, 1999) of the observed to the expected distribution were computed. The confidence intervals for τ were obtained from 5000 bootstrap replicates. MMD depicts the distribution of the number of pairwise differences between haplotypes in the sampled population. This distribution is usually multimodal for populations at demographic equilibrium reflecting the stochastic shape of gene trees, but generally unimodal when populations experienced recent demographic expansion (Slatkin and Richard, 1991; Rogers and Harpending, 1992) or range expansions with high levels of migration between neighboring demes (Ray *et al.* 2003; Excoffier 2004).

The frequency spectrum of mutations was examined using two statistics, Tajima's D (Tajima, 1989a) and Fu's Fs (Fu, 1997), in Arlequin v.3.11. Both tests are based on the infinite site model without recombination. The significance of these statistics was examined by generating 5000 random samples under the hypothesis of selective neutrality and population equilibrium. Significant negative values of these statistics evidence demographic expansion (Tajima 1989b; Aris-Brosou and Excoffier 1996; Fu 1997). Particularly, Fu's Fs has been shown to be especially sensitive to departure from population equilibrium as in case of a population expansion (Fu 1997). However, natural selection can also lead to significant non-zero values in both tests.

The demographic history of the species was also investigated by means of an Extended Bayesian Skyline Plot (EBSP) analysis (Heled and Drummond 2008) using software BEAST v1.7.1. Being able to incorporate multiple unlinked loci, and taking into account their specific mode of inheritance, the EBSP analysis builds up over previous skyline methods which rely in single locus. While from single locus analysis there is considerable coalescent error associated with estimates (an individual genealogy only represents a single realization of a stochastic process), increasing the number of independent loci allows the uncertainty in the coalescent to be assessed. Therefore, this multiple loci approach has been shown to provide better estimates of population size and substantially reduce estimation error (Heled and Drummond, 2008; Ho and Shapiro, 2011). Moreover, it enhances the possibility of detecting past population bottlenecks and also recovering the population history prior to it.

The EBSP analysis was performed for each species separately and for each *L. americanus* cluster (Cheng, 2010) separately, including both nuclear and mitochondrial data (except the KITLG and Cyt *b* markers regarding the analysis of the Boreal and PacNW, respectively, due to absence of variation). Given the possibility that the inclusion of *L. americanus* Cyt *b* sequences more closely related to *L. californicus* results from introgression, which would bias estimates, these were excluded from this locus dataset. Analyses were run using substitution models, unlinked clock models and trees across markers, using the substitution rate estimated for the Cyt *b* (PRK for the PacNW cluster), estimated from the *Lepus-Oryctolagus* uncorrected genetic distance, Dxy (Nei 1987) and considering a split time of 11.8 Myr (Matthee et al, 2004), as a stable reference for all other markers (Heled 2010). Since it is generally a good approximation for analysis at the intrapopulation level (Yang, 2006) and because, by simplifying the coalescent model, it helps analyses to converge (Heled 2010) the strict molecular clock was applied to all loci. The best-fitted model determined by jModelTest or, if not implemented on BEAST, the next most parameterized model available was assigned for each locus. The ploidy was set according to the nature of each locus to account for differences in the mode of inheritance. A piecewise-linear model was defined to describe the demographic history. Prior and operators settings were adjusted as suggested by the authors (Heled 2010). Three independent runs of 200’000’000 generations, sampling every 20’000 generations, were performed to assess the consistency of the estimates. Convergence, stationarity and ESS for each parameter of interest in the analysis were evaluated using the software TRACER 1.5 (Rambaut and Drummond 2007).

Reconstruction of the Speciation Process

Determination of the adequate models of molecular evolution

jModelTest v0.1.1 (Guindon and Gascuel 2003; Posada 2008) was employed to determine the model of sequence evolution best suited for each marker, among 88 possible models, based on likelihood scores and under the Akaike Information Criterion with correction (AICc).

Isolation with Migration

One of the major aims of this work is to reconstruct the history of divergence of three North American hares and to infer whether gene flow has played an important role in the speciation process – and to specifically test whether massive mtDNA introgression occurred from *L. californicus* into PacNW *L. americanus* (as suggested by Cheng, 2010). This aim demands reconstructing the branching pattern of the evolutionary tree (topology) and the relevant population history parameters – e.g. effective population sizes, divergence times and migration (gene flow) rates. However, since phylogeny reconstruction methods rely on the assumption of no between-species gene flow, we first tested if nuclear gene flow occurred between any of the three studied *Lepus* species after the divergence using IMa2 (Hey 2010) considering all three possible branching patterns. This method co-estimates multi-locus effective population sizes (past and present), divergence times and migration rates, under a Isolation with Migration (IM) model which may be applied for multiple populations (taxa) (Nielsen and Wakeley, 2001; Hey and Nielsen, 2004; Hey, 2010) - Figure 3. The dataset reduced to the largest non-recombining blocks as retrieved by IMgc was used as described in the *Data treatment* section. To assess the consistency of the estimates, three independent runs were performed varying the parameters upper bound priors and the starting seeds. The HKY model (Hasegawa et al, 1985) was applied for the eight nuclear loci. The likelihood ratio test described by Nielsen and Wakeley (2001) was applied to assess whether migration was significantly different from zero. The locus specific mutation rates were estimated from the *Lepus-Oryctolagus* uncorrected genetic distance, Dxy (Nei 1987), considering a split time of 11.8 Myr (Matthee et al, 2004) and a generation time of two years (Marboutin and Peroux, 1995). The effective population sizes and the divergence times from IMa2 were calculated from the highest posterior density of each parameter using the geometric mean of the locus-specific mutation rates, following the instructions of the IMa manual.

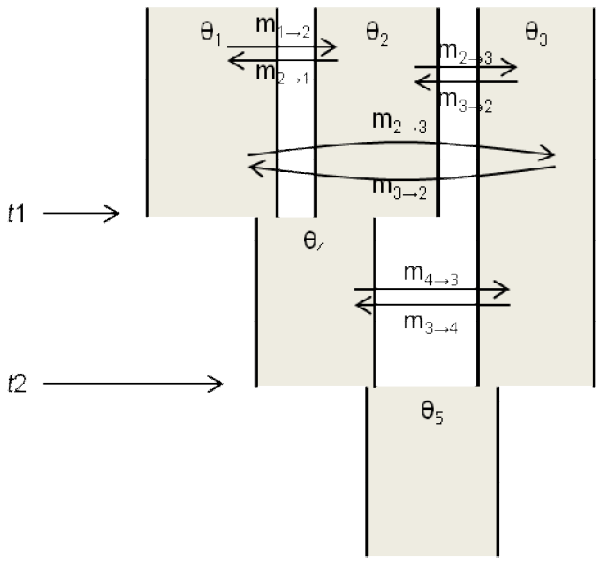


Figure 3 - Schematic representation of the IM model for three taxa. θ_i - effective population size of population *i*; $m_{i \rightarrow j}$ - migration from population *i* to population *j*; t_s - split between two populations at a time of divergence *s*.

Gene tree phylogenies

Maximum-Likelihood (ML) phylogenies were estimated for each nuclear locus using the software Garli v2.0 (Zwickl 2006). The best suited model as determined with jModelTest was fit to the data, although not fixing the model parameters. No starting topology was defined. The program was set to run until no significantly better scoring topology (defined by an lnL increase of 0.01) was encountered after 1'000'000 generations but not longer than 5'000'000 generations. For each locus, five independent search replicate runs were performed to appraise the consistency of the estimates. For the cytb phylogeny, 500 bootstrap replicated were run in Garli.

Bayesian inferences (BI) were performed using BEAST v1.7.1 (Drummond and Rambaut, 2007), applying the best-fitted model as defined by jModelTest. An uncorrelated lognormal relaxed clock (Drummond et al, 2006) and the Yule tree prior were applied. Posterior probabilities were determined after 25'000'000 generations sampling every 2'500 generations. The stability of the runs and convergence of the Markov chain Monte Carlo (MCMC) were assessed using Tracer v1.5 (Rambaut and Drummond, 2007). The first 10% trees of each run were discarded as burn-in.

Species tree reconstruction based on nuclear loci

Given the stochasticity of the coalescent process, there is the potential for gene trees to differ among loci and to not match the underlying species tree. This is particularly true for closely related species or species with large population sizes. We thus used the species-tree Bayesian inference method *BEAST (Heled and Drummond, 2010), as implemented in software BEAST v1.7.1 (Drummond and Rambaut, 2007), to infer the phylogeny of the three North American *Lepus* species based on the sequenced nuclear loci. The *BEAST is a multispecies/multilocus method that explicitly takes into account the possibility of differential lineage sorting across individual loci. This kind of methods are expected to perform better in multi-locus datasets (see e.g. Edwards et al., 2007; Kubatko and Degnan, 2007). In order to avoid violating the assumption of no-recombination within each locus, we used the dataset reduced to the largest non-recombining blocks as retrieved by IMgc (see Data Treatment). Moreover, *BEAST assumes no gene-flow between taxa. According to the isolation with migration analyses described previously, limited gene flow was determined among two of the analyzed species. However, this is unlikely to affect the inference of our phylogeny as multilocus coalescent based methods have been shown to be robust considering low levels of gene flow (Eckert and Carstens, 2008).

Input files for BEAST were generated in BEAUti, part of the package. Specimens were a priori assigned to species, a requirement of *BEAST. Two datasets, one with outgroups and another without outgroups were used. Nucleotide substitution models were set for each locus as assigned by jModelTest. In cases for which the best-fit model as determined by jModelTest was not implemented on BEAST, the next most parameterized model available on BEAST was used. Posterior phylogenies were determined in *BEAST using an uncorrelated lognormal relaxed clock (Drummond et al, 2006) and the Yule tree prior. Prior settings were set as default except for the relaxed clock standard deviation prior which was set to an exponential distribution with a mean of 0.5 as recommended by the authors. Three independent runs of 250’000’000 generations sampling every 25’000 generations were performed. Convergence of the Markov chain Monte Carlo (MCMC), stationarity of the runs and effective sample size (ESS) for each parameter of interest in the analysis were evaluated using the software TRACER 1.5 (Rambaut and Drummond, 2007). The initial 10% of the runs were discarded as burn-in. Summary trees were generated in TreeAnnotator v1.7.1, part of BEAST package. The resulting trees were then analyzed in FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). The same analysis was

also performed but considering the three *L. americanus* clusters separately in order to assess the occurrence and phylogenetic history of these clusters.

Taxa Delimitation Analysis

In order to assess whether the three *L. americanus* clusters described by Cheng (2010) were also recovered in our nuclear data, we performed a Bayesian species delimitation analysis as implemented in the software Bayesian Phylogenetics and Biogeography (BPP) v2.0 (Rannala and Yang, 2003; Yang and Rannala, 2010). This is a multilocus coalescent-based method that accommodates the species phylogeny as well as lineage sorting due to ancestral polymorphism and uses reversible-jump Markov chain Monte Carlo (rjMCMC) to estimate the posterior probability of the different possible taxa delimitation models (resulting from collapsing nodes of the *a priori* given guide tree).

We used the species tree inferred from *BEAST analysis (considering the three *L. americanus* clusters) as the guide tree. Because ancestral effective population size (θ) and root age (τ_0) priors can affect the posterior probabilities of the models (Yang and Rannala, 2010) different combinations of prior distributions, of large and small θ and shallow and deep τ_0, were used (see Appendix 9). Other divergence time parameters were assigned the Dirichlet prior (Yang and Rannala, 2010: equation 2)

Coalescent simulations

To disentangle the relative contribution of incomplete lineage sorting and introgression to the conflicting patterns observed in *L. americanus* and *L. californicus* between nuclear and mitochondrial phylogenies, we used a similar methodology to that used by Melo-Ferreira et al (2012) which is adapted from that developed by Joly et al. (2009) was employed. This method relies on the reasoning that under a given reconstructed speciation history, the expected divergence under a strict lineage sorting model for any given locus can be simulated and compared with the observed divergence. Smaller than expected divergences suggest introgression.

Divergence times and population size estimates obtained between *L. americanus* and *L. californicus* under the IM model were used as input on SimCoal2 v2.1.2 (Laval and Excoffier, 2004) to simulate Cytb datasets under the assumption of no introgression. 10’000 datasets with the same characteristics (sequence length and number of

sequences per species) of the empirical Cyt b dataset were simulated. 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 and without gene flow occurring between the two descendant populations was applied. The mtDNA mutation rate was estimated from the *Lepus-Oryctolagus* average uncorrected distance, Dxy and again considering a split at 11.8 Mya (Matthee et al, 2004) and a generation time of two years (Marboutin and Peroux, 1995). The minimum pairwise uncorrected *p*-distance between the descendant populations was retained for each replicate, producing a distribution of the expected minimum distances under the assumption of no gene flow. The empirical *p*-distance was considered to reject the null hypothesis (phylogeny explained strictly by lineage sorting) if lower than the 5^{th} percentile of the simulated minimum distances distribution, and thus introgression inferred.

RESULTS

Sequence Data

Eight nuclear and one mitochondrial DNA genes were sequenced in this study, in a total of 6184 bp for nuclear genes and 580 bp for mtDNA (Table 3). After removing gaps, which are ignored in most of the performed analysis, the nuclear dataset was composed by 5987bp which was further reduced to 5476 bp after narrowing the analyzed fragments to their largest non-recombining blocks (Table 3; see Data Treatment). The nuclear and mitochondrial DNA final datasets were composed by a total of 5127 bp and 506 bp respectively, when including the two outgroups, *S. audubonii* and *O. cuniculus* in the case of the nuclear dataset (97 specimens; Table 3) and only *O. cuniculus* as outgroup in the case of the mtDNA dataset (Table 3).

Table 3 - Length and appropriate mutation models of the sequenced markers used in this study

	Marker	Inheritance Compartment	Number of sites						Mutation Model[d]
			Total	NG[a]	LNRB[b]	OUT[c]	Inheritance Compartment	Genomic Compartment	
1	CYTB	mtDNA	580	580	580	506	506 bp	506 bp	TPM3uf+G
2	POLA1	X	812	806	566	562	1217 bp	5127 bp	F81+G
3	GRIA3	X	969[f]	947	947	655			TrN
4	SRY	Y	1608	1557	1557	1554	1554 bp		TIM2
5	SPTBN	autosome	636[f]	579	509	495	2356 bp		K80
6	PRKCI	autosome	436	411	411	400			HKY
7	DARC	autosome	784	774	732	732			TPM2uf+G
8	KITLG	autosome	552	529	439	420			JC
9	TF	autosome	387	384	315	309			JC

[a]NG: No gaps (alignment with gaps removed and); [b]LNRB: Largest nonrecombining blocks; [c]OUT: alignment including outgroups; [d]See (Posada 2008) for a description of the mutation models and references; [f]Microsatellites and buffer regions, one in SPTBN (19 bp) and two in GRIA3 (34 bp; 26 bp) not considered (see Material and Methods).

Diversity and Demography

To assess whether the populations were panmitic we tested conformance to Hardy-Weinberg equilibrium. In general, whenever informative, there were no significant deviations to expectations under the Hardy Weinberg equilibrium (see Appendix 3). The alternative hypotheses to Hardy-Weinberg equilibrium, heterozygote excess or deficiency, were also tested considering all markers. In these, none of the populations presented significant deviations (excess or deficiency) except for the Wyoming population which presented heterozygote deficiency (Appendix 3).

The neutral equilibrium of the loci analyzed was tested by performing an eight-loci HKA test. No significant deviations from neutral expectations in ratios of polymorphism to divergence between species were detected ($P > 0.05$), either comparing *Lepus* species to another *Lepus* or to *Oryctolagus* (data not shown).

Mean values of diversity (h, π and θ_S) across analyzed inheritance compartments are depicted in Table 4. Levels of diversity among species were generally similar between *L. californicus* and *L. americanus* while *L. townsendii* presented the lowest diversity overall. Taking into account the *L. americanus* population clusters identified using microsatellite data (Cheng 2010), overall there was no tendency for one of the clusters being more diverse over the others. However, Rockies cluster presented the highest diversity values at the Y-linked locus while the Boreal tended to be more diverse at the X-linked markers. When sampled populations were separately taken into account, *L. californicus* populations tended to be the most diverse except at the Y-linked loci were diversity values were generally the lowest. Among *L. californicus*, the more peripheral Oregon (OR) population was generally the least diverse. In *L. townsendii*, levels of diversity are generally similar across populations. In the set of *L. americanus* populations, the only Boreal population, Saskatchewan (SK1), was often the most diverse. For the Rockies cluster, the Oregon (OR2) was consistently the most diverse while regarding the PacNW cluster no population tended to be more diverse than the others.

Table 4 - Summary of the analysis of nuclear polymorphism in species analyzed in this study

		autossomal[a]					X[b]					Y					mitochondrial			
Groups		**Mean N**	**Mean N_h**	**Mean *h***	**Mean π**	**Mean θs**	**Mean N**	**Mean N_h**	**Mean *h***	**Mean π**	**Mean θs**	**N**	**N_h**	***h***	**π**	**θs**	**N**	**N_h**	***h***	**π**	**θs**
Species																					
Lam		81	9	0.670	0.72	2.34	49	9	0.701	0.52	3.49	26	8	0.637	0.630	3.440	26	11	0.883	1.45	8.12
Lcf		55	13	0.786	1.06	3.54	36	9	0.670	0.47	3.13	19	4	0.731	0.650	6.290	30	16	0.936	0.86	7.07
Ltw		31	5	0.489	0.87	1.16	22	8	0.762	0.22	2.06	10	3	0.378	0.210	1.770	44	14	0.834	0.77	5.06
Clade																					
Lam Boreal		9	4	0.515	0.28	1.40	5	4	0.900	0.43	2.75	4	2	0.500	0.060	1.090	11	5	0.709	0.62	4.10
Lam Rockies		29	3	0.582	0.29	1.02	15	4	0.494	0.38	2.02	10	5	0.822	1.200	1.770	15	6	0.790	0.72	4.31
Lam PacNW		43	4	0.490	0.76	1.00	30	3	0.448	0.13	0.63	12	2	0.167	0.020	0.330	16	7	0.775	0.73	3.32
Population																					
Boreal	SK1	9	4	0.515	0.28	1.40	5	4	0.900	0.43	2.75	4	2	0.500	0.060	1.090	5	4	0.900	0.62	4.32
Rockies	OR2	16	2	0.332	0.23	0.73	7	3	0.432	0.13	1.11	6	2	0.533	0.070	0.880	8	4	0.750	0.54	3.47
	WY1	14	2	0.117	0.05	0.51	9	2	0.232	0.03	0.39	4	3	1.000	0.080	1.330	7	2	0.286	0.10	0.82
PacNW	WA4	19	3	0.447	0.71	1.06	12	2	0.214	0.08	0.39	5	2	0.400	0.050	0.480	6	1	0.000	0.00	0.00
	WA1	13	2	0.401	0.18	0.78	11	3	0.630	0.17	0.87	5	1	0.000	0.000	0.000	8	5	0.786	0.26	1.54
	CA1	11	2	0.462	0.42	1.29	8	2	0.194	0.07	0.37	2	1	0.000	0.000	0.000	8	2	0.250	0.04	0.39
Lcf	OR	19	5	0.621	0.54	2.04	16	3	0.371	0.05	0.46	3	1	0.000	0.000	0.000	10	5	0.822	0.71	4.24
	CA	11	6	0.859	1.23	2.79	7	3	0.652	0.11	0.65	3	2	0.667	0.040	0.670	6	4	0.867	0.76	3.94
	TX	13	7	0.864	1.39	3.43	9	5	0.804	0.26	2.07	7	1	0.000	0.000	0.000	8	4	0.750	0.25	1.93
	AZ	12	5	0.764	0.94	2.76	5	4	0.817	0.77	3.06	6	1	0.000	0.000	0.000	6	4	0.800	0.51	3.50
Ltw	MT	3	2	0.533	1.41	1.26	4	3	0.667	0.14	1.10	-	-	-	-	-	2	1	0.000	0.34	0.00
	YLL	9	2	0.466	0.94	0.74	5	3	0.750	0.23	1.44	5	1	0.000	0.000	0.000	15	4	0.371	0.00	2.15
	GTP	2	1	0.400	0.96	0.60	2	2	1.000	0.46	2.50	-	-	-	-	-	1	1	-	-	-
	WY	16	4	0.494	0.89	1.02	11	5	0.773	0.19	1.54	5	3	0.700	0.420	2.400	19	11	0.906	0.54	4.29

Notes: [a]Unweighted average values across the five autosomal markers; [b]Unweighted average values across the two X chromosome markers; (Polymorphism values of each individual marker are given in Appendix 4). Lam – *Lepus americanus*; Lcf – *Lepus californicus*; Ltw – *Lepus townsendii*; N – number of sequences; N_h – number of observed haplotypes; *h* – haplotype diversity; π – nucleotide diversity; θ_S – computed from the number of segregating sites (Tajima 1983).

Neutrality tests could not be performed for some loci in some populations due to lack of polymorphism (Tables 5 and 6). Among populations, in general there were no significant deviations from zero of Tajima's D or Fu's Fs, indicating no deviations relative to the mutation-drift expectations. When values were found to differ significantly from zero, they were always negative, although the rejection of the null model was not consistent in both tests for most of the cases.

Table 5 - Tajima's D neutrality test

	Tajima's D								
	SPTBN	PRKCI	DARC	KITLG	TF	POLA1	GRIA3	SRY	CYTB
Species									
LAM	-0.56	0.91	-0.04	-0.23	*-1.48	-0.07	-1.39	-0.45	0.13
LCF	-0.87	-1.33	-1.28	-0.55	0.36	*-1.69	-0.37	2.54	-1.04
LTW	-0.05	-0.3	0.26	*-1.89	-1.37	-0.53	-1.15	*-1.74	-0.37
Clade									
Boreal	*-1.56	0.2	-0.5	0	-0.57	0.27	0	-0.71	-0.53
Rockies	2.07	-1.45	1.81	-0.1	1.62	1.57	*-1.83	0.12	-0.12
PacNW	-0.54	2.76	2.2	1.07	-0.49	2.11	-0.13	-1.14	1.02
Population									
SK1	*-1.56	0.2	-0.5	-	-0.57	0.27	0	-0.71	-1.18
OR2	0	-1.27	1.93	0.65	-	-1.05	-1.31	1.03	-0.46
WY1	-1.48	-1.14	-	*-1.89	-	-	-1.31	0	-1.24
WA4	-	1.65	1.64	0.67	-0.78	0.41	-	-0.82	-
WA1	-	-0.47	-1.45	0.63	1.03	1.64	1.19	-	-0.12
CA1	1.5	-1.54	1.82	0.62	-	0.2	-	-	-1.05
OR	-1.03	-1.42	-0.44	0.35	-1.07	-1.49	1.24	-	-0.16
CA	-0.7	0.29	-0.63	-0.13	1.38	0.31	1.34	0	0.69
TX	0.93	-0.41	0.35	-0.12	1.33	-0.92	0.47	-	-1.18
AZ	-1.13	-1.16	-1.18	-0.22	1.3	-0.68	-0.45	-	-0.96
MT	-0.71	0	0	-	-0.61	-0.61	1.09	-	-
YLL	1.23	1.6	0.85	-	0.01	-0.97	1.64	0	-0.29
GTP	0	0	-	-	-	0	0	-	-
WY	0.41	-0.07	0.83	*-1.85	-1.5	-0.25	-1.14	-1.12	-0.99

Notes: Significant values ($P < 0.05$) are given by an *.

When all data of species was taken together, both Tajima's D and Fu's Fs presented significant negative values for *L. americanus* and *L. californicus* but only at one locus while 4 of the 9 loci suggested signs of expansion in *L. townsendii* - see Tables 3 and 4. Also when separating *L. americanus* clusters no significant negative values were found except in sporadic loci in the Rockies and Boreal clusters - see Tables 3 and 4. At the population level, only four populations presented significant deviations from zero of Tajima's D and/or Fu's Fs values, but generally only for one locus the exception being *L. townsendii* WY population which presented signals of expansion at four of the loci - see Tables 3 and 4.

The mismatch analysis of sequences of each of the three species and of *L. americanus* clusters showed a multimodal distribution of the number of pairwise differences rejecting the Sudden Expansion Model. Regarding at separate sampled populations, only the distributions of the number of pairwise differences for *L. californicus'* CA and *L.*

townsendii's WY populations fitted the expectation of the Sudden Expansion Model (Appendix 5).

Table 6 - Fu's Fs neutrality test

	Fu's Fs								
	SPTBN	PRKCI	DARC	KITLG	TF	POLA1	GRIA3	SRY	CYTB
Species									
LAM	1.84	9.58	0.07	0.27	*-5.50	3.5	-2.6	4.88	1.21
LCF	-2.06	1.9	-2.3	-3.89	0.55	*-5.19	4.92	10.38	-4.33
LTW	11.72	7.28	0.47	-0.98	*-2.31	-0.18	*-4.68	3.49	-1.54
Clade									
Boreal	1.22	-0.11	-2.02	-	*-3.41	0.6	-0.69	1.1	1.01
Rockies	3.42	3.12	3.58	0.91	1.7	5.14	*-3.31	6.19	1.19
PacNW	1.03	12.79	2.01	1.64	-0.47	3.37	0.82	0.43	0.44
Population									
SK1	1.22	-0.11	-2.02	-	*-3.41	0.6	-0.69	1.1	-0.04
OR2	0	3.64	2.80	0.87	-	1.69	-1	1.72	1.1
WY1	0.3	-0.48	-	0.18	-	-	-1	-1.22	0.86
WA4	-	6.42	2.83	1.36	-0.73	1.65	-	1.04	-
WA1	-	0.62	0.43	0.6	1.1	2.34	1.26	-	-1.8
CA1	2.34	5.47	0.84	0.91	-	1.59	-	-	-0.18
OR	-1.02	3.52	-0.83	-1.08	2.3	*-1.55	1.23	-	1.06
CA	-0.98	4.32	-1.28	-1.4	1.64	-0.3	0.86	0.2	0.94
TX	0.15	0.31	0.58	-1.47	2.88	-1.75	0.83	-	-0.52
AZ	-1.85	1.99	-1.59	0.58	1.60	-0.99	2.96	-	0.19
MT	5.65	3.04	1.06	-	0.17	0.17	0.01	-	-
YLL	10.69	6.87	1.52	-	0.42	-0.19	0.95	0.00	1.28
GTP	0.69	2.94	-	-	-	1.39	0.69	-	-
WY	9.99	6.94	0.28	-0.12	*-1.61	0.06	*-3.01	2.94	*-3.91

Notes – significant values ($P < 0.02$) are given by an *.

The Extended Bayesian Skyline Plot analyses, which infers the historical demographic trend of populations is presented in Figure 4. Given that large 95% HPD confidence intervals were inferred these inferences must be regarded with caution. It may nevertheless be useful to better understand the population demographic trends of the analyzed species. The demographic history of *L. americanus* suggests no changes over time of the effective population size (Fig. 4a). Also when considering the population clusters of this species separately, no obvious changes in population size through time were found for the Rockies and PacNW clusters (Figs. 4e and 4f) while a recent increase, starting *ca.* 330 kya, is found for the Boreal cluster (Fig. 4d). Regarding *L. californicus* and *L. townsendii* and according to the EBSP, after a period of stability of the effective population sizes, these two species have experienced a tenuous expansion since *ca.* 1 Mya, (Figs. 4b and 4c).

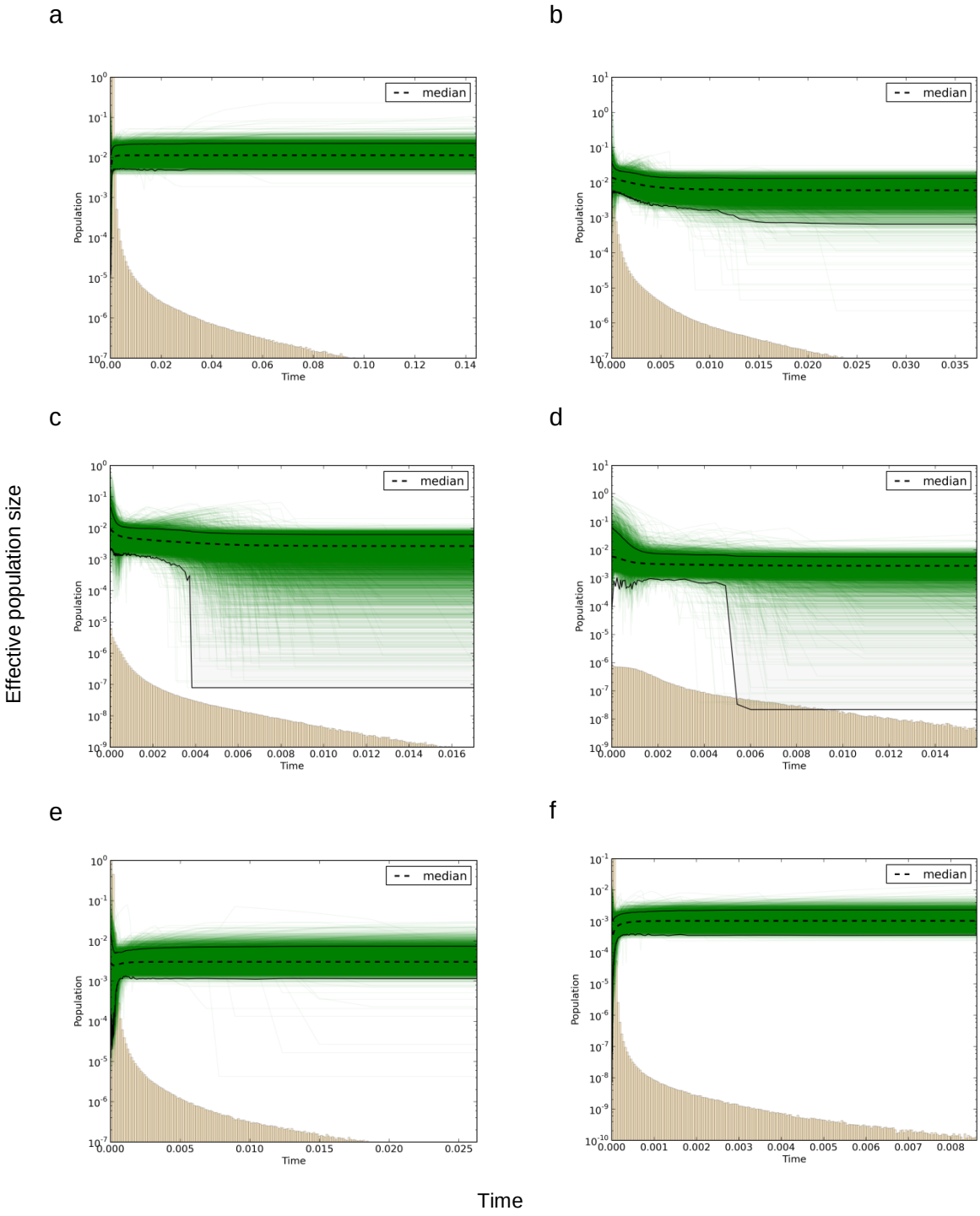


Figure 4 – Extended Bayesian Skyline Plot (EBSP) showing the demographic history of: a - *L. americanus*; b - *L. californicus*; c - *L. townsendii*; d - *L. americanus* Boreal cluster; e - *L. americanus* Rockies cluster; f - *L. americanus* PacNW cluster. The median population size is given by the dashed line and the 95% HPD intervals by the solid black lines. The height of each bar is proportional to the number of demographic functions which had an X point at that interval of time and green lines depict the demographic reconstruction of each of the iterations. The X-axis represents time backwards (in substitutions) and the effective population size is represented in the Y axis (on a log scale for clarity).

Isolation with Migration model – inferences of gene flow

Given that *BEAST does not incorporate gene flow after the initial split in the reconstruction of species phylogenies, its prevalence in the nuclear genes was assessed fitting our data to the IM model (Nielsen and Wakeley 2001; Hey and Nielsen 2004). Since relations among the three species in analysis were not known *a priori* this analysis was performed considering all three possible branching patterns. Gene flow, either ancient or between extant species was always not significantly different from zero regardless of the considered topology. However, ancient migration parameters were generally poorly estimated with the right tail of the posterior density curves failing to reach zero, which is likely due to the complexity of the model used with three populations, which results in difficulties to estimate some of the model parameters. Overall there were no consistent evidences of nuclear gene flow among the studied species.

Phylogenetic Inferences

The models of sequence evolution best suited for each marker as determined by jModeltest are depicted in Table 3. The Maximum Likelihood (ML) phylogenetic reconstructions for each individual nuclear locus yielded consistent best likelihood scores among replicates runs and only slightly different topologies (data not shown) probably due to low variation in each single gene. Accordingly, also the Bayesian Inference analyses recovered trees with overall low support values. Moreover, phylogenies reconstructed from each of the loci presented different topologies, either in BI and ML analysis (Appendixes 6 and 7, respectively).

Given the inconsistency of single-gene phylogenetic reconstructions among nuclear loci, phylogenetic reconstruction from the concatenation of these fragments could mislead inference of species relationships (Kubatko and Degnan, 2007). The species tree was thus inferred using the multispecies/multilocus coalescent method implemented in *BEAST (Heled and Drummond, 2010). This method makes use of multi-locus data, embedding individual gene trees in a multi-locus species tree, to co-estimate the species tree topology and effective population sizes of tip and ancestral taxa using a Bayesian MCMC approach. Because *BEAST assumes no recombination,

nuclear loci were reduced to the largest non-recombining blocks in IMgc software (Woerner et al, 2007). The resulting phylogeny (both including and excluding the outgroups – since this method estimates the root of each single gene tree and uses the multispecies coalescent of the species tree, outgroups need not be included; Heled and Drummond (2010)) places *L. californicus* and *L. townsendii* as sister taxa and an earlier split originating *L. americanus*. This topology was consistent across the three replicate runs and in all of them clades were supported with high posterior probability values - above 91% (Figure 5a; Appendix 8).

The *BEAST analysis was also performed considering the three putative *L. americanus* clusters separately, in order to infer the phylogenetic history of the three *L. americanus* clusters. This allowed investigating whether the *L. americanus* PacNW cluster could also be retrieved more closely related to *L. californicus* in the nuclear genome, as is found in the mtDNA. The resulting phylogeny retrieved the same topology as that of the species tree. Moreover, the monophyly of *L. americanus* was highly supported (Figure 5b).

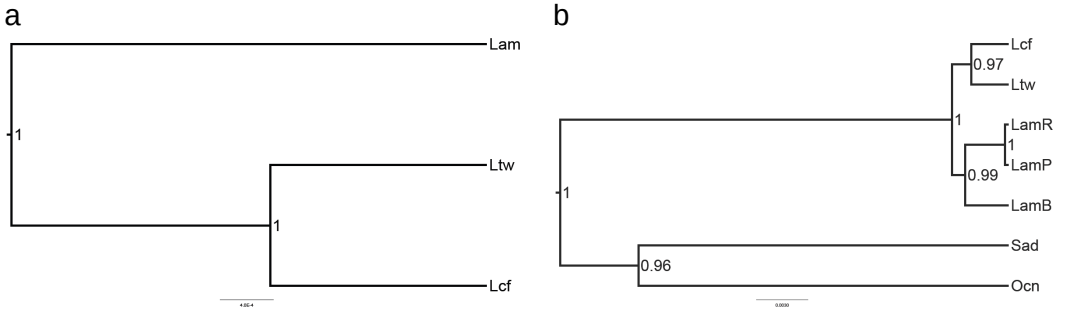


Figure 5 - a) Nuclear DNA *BEAST species tree for the eight nuclear loci (the posterior probability of each clade is depicted in front of each node); b) Nuclear DNA *BEAST species tree for the eight nuclear loci considering the three *L. americanus* clusters separately (the posterior probability of each clade is depicted in front of each node); Lam - *L. americanus,* LamB - *L. americanus* Boreal cluster, LamR - *L. americanus* Rockies cluster, LamP - *L. americanus* PacNW cluster, Lcf - *L. californicus, Ltw - L. townsendii, Ocn – O. cuniculus, Sad – S. audubonii*.

Given that taxa composition must be defined a priori in, which is a strong assumption, we assessed whether the three *L. americanus* clusters were recovered from the nuclear data with no a priori imposition by performing a Bayesian taxa delimitation using BPP. The validity of these three clusters was highly supported (p>0.99) in all analysis (Appendix 9).

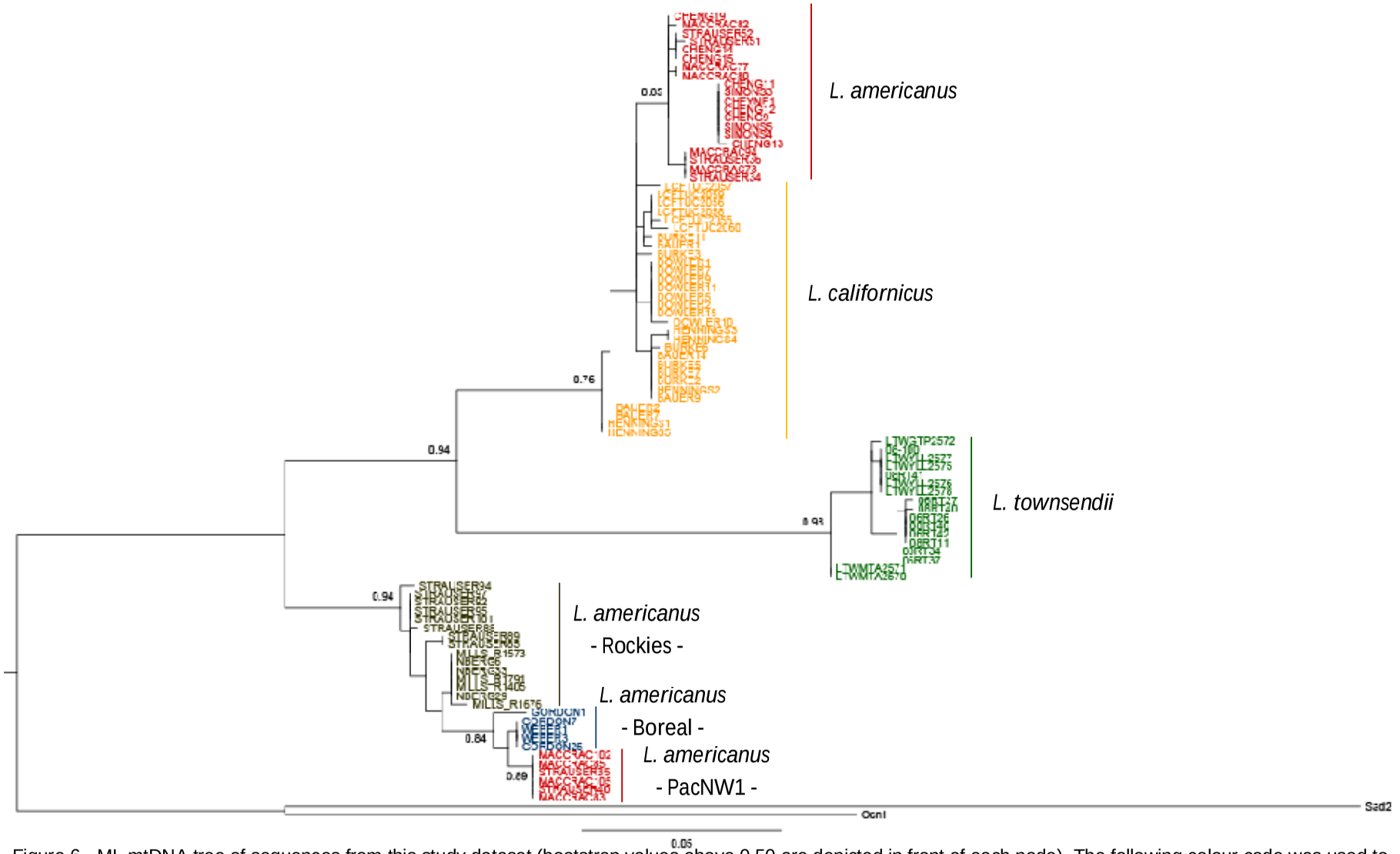


Figure 6- ML mtDNA tree of sequences from this study dataset (bootstrap values above 0.50 are depicted in front of each node). The following colour code was used to identify species and cluster to which individuals belong (as defined by Cheng (2010) based on microsatellite analysis): *L. americanus* Boreal cluster (blue), *L. americanus* Rockies cluster (brown), *L. americanus* PacNW cluster (red), *L. californicus* (yellow) and *L. townsendii* (green).

Using the dataset generated in this study, the inferred mtDNA phylogeny does not conform, as expected and found by Cheng (2010), to that inferred for the nuclear DNA (Figure 6). The ML analysis, *L. americanus* does not form a monophyletic group. Instead, while a group of *L. americanus* PacNW individuals *(PacNW1)* clusters together with the other *L. americanus* clusters another group of individuals from the PacNW cluster *(PacNW2)* is more closely related to *L. californicus* rather than to the other two *L. americanus* clusters. This cluster however is not shared with *L. californicus* and is retrieved as a subgroup of *L. californicus* (Figure 6).

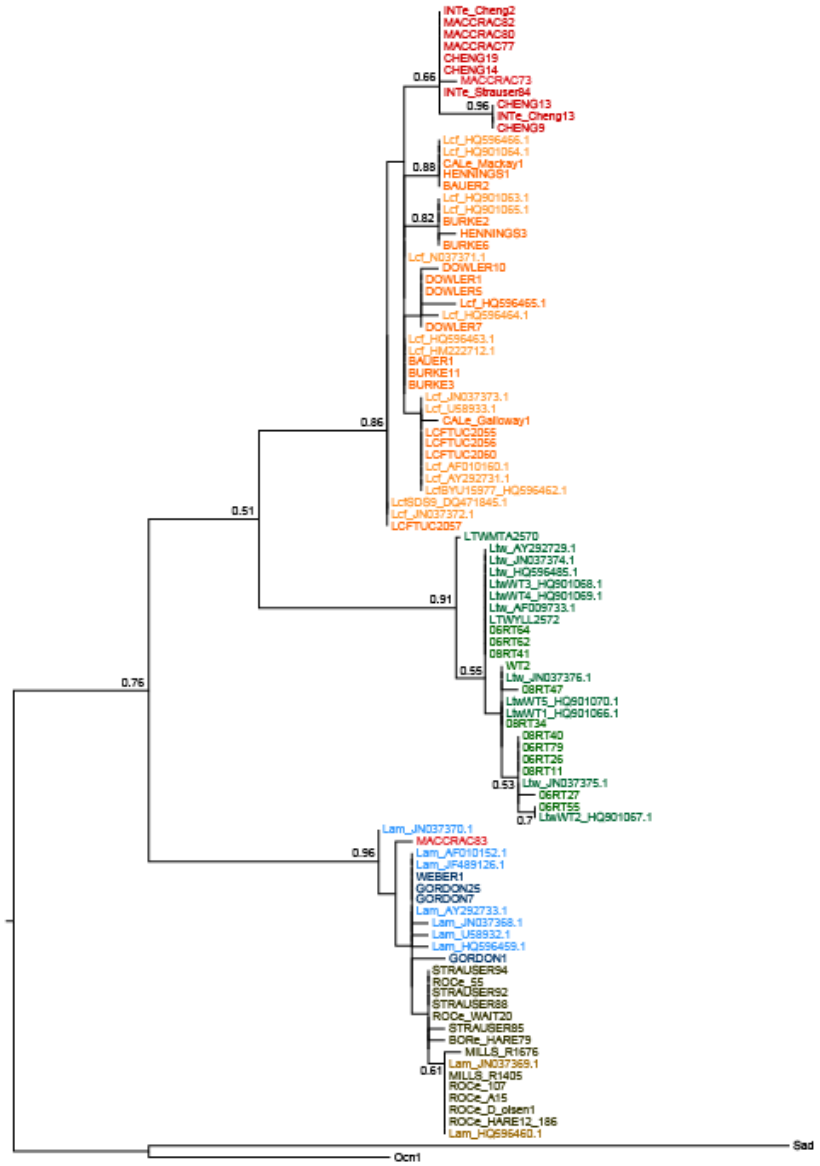


Figure 7 - ML tree of the mtDNA alignment with sequences of *L. americanus*, *L. californicus* and *L. townsendii* Cyt *b* sequences from Genbank (bootstrap values above 0.50 are depicted in front of each node). The following colour code was used to identify species and cluster to which individuals belong and to distinguish between GenBank (GB) and our dataset (OD) sequences: *L. americanus* PacNW cluster (OD – red; GB – light red), *L. americanus* Boreal cluster (OD – blue; GB – light blue), *L. americanus* Rockies cluster (OD – brown; GB light brown), *L. californicus* (OD – yellow; GB – light yellow) and *L. townsendii* (OD – green; GB – light green).

To investigate whether any other North American mtDNA lineage exists that may not have been sampled in this work, all *L. americanus*, *L. californicus* and *L. townsendii* Cyt b sequences deposited on Genbank were aligned to sequences representative of lineages found in our dataset. As depicted in Figure 7, all GenBank sequences are within lineages found in our dataset, reassuring that for the moment no other lineage was overlooked by our dataset. Moreover, we confirmed that the GenBank sequences appear correctly assigned to the species according to our data.

Furthermore, GenBank sequences of other hare species either from North American and other continents were also aligned to our dataset, the PacNW cluster continuing to be more closely related to *L. californicus* (Figure 8).

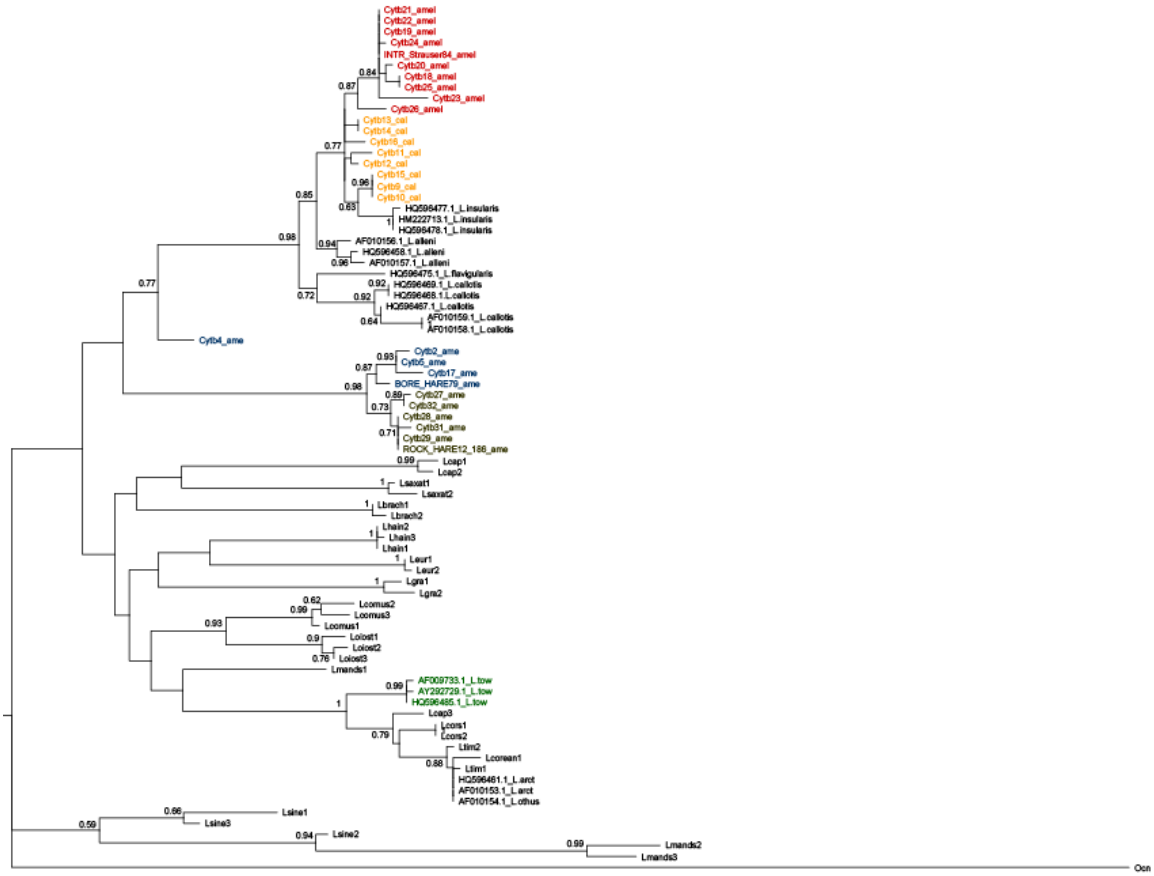


Figure 8 - ML tree of the mtDNA alignment including sequences from other North American species deposited on GenBank (posterior values above 0.50 are depicted in front of each node). The following colour code was used to identify species and cluster to which individuals belong: *L. americanus* PacNW cluster (red), *L. americanus* Boreal cluster (blue), *L. americanus* Rockies cluster (brown), *L. californicus* (yellow) and *L. townsendii* (green).

The cause of the discrepancy between mitochondrial and nuclear DNA was further investigated using coalescent approaches to disentangle the relative contribution of incomplete lineage sorting and introgression in the history of speciation of the studied American hares (see Coalescent Simulations section).

Isolation with Migration model – speciation history

Considering the IMa2 results only for the *BEAST inferred topology (Figure 5a), the posterior density curves of the parameters' estimates for the extant populations were generally consistent across the three replicate runs. However, for estimates of the most ancient divergence time, the two ancient population sizes and ancient migration, the right tail of the posterior density curves systematically failed to reach zero in all replicate runs.

The mutation rate of nuclear DNA (geometric mean of the eight loci) was estimated, from the average distance between *Lepus* and *Oryctolagus*, to be of 3.45 x 10^{-9} substitutions/site/ generation. This estimate was used to calculate the estimated parameters' values (Table 7). According to the model, *L. californicus* and *L. townsendii* were estimated to have diverged about 2.3 Mya and no significant gene flow is inferred neither between these two species nor between these and *L. americanus*. Among the extant species *L. townsendii* had the lowest estimates of effective population size while *L. californicus* presented the highest.

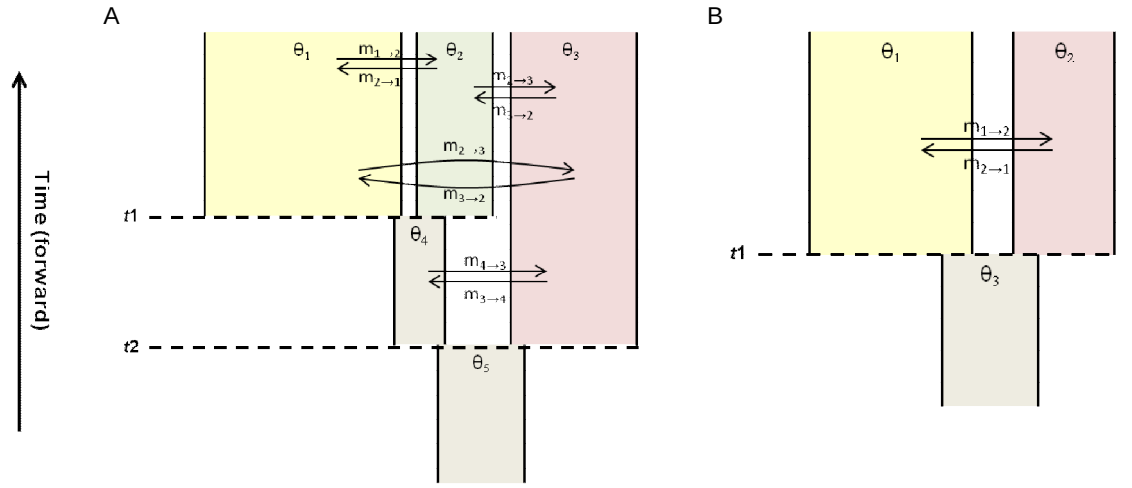


Figure 9 - Schematic representation of the IM model applied to the data: a) IM model for three populations – *L. californicus* (yellow), *L. townsendii* (green) and *L. americanus* (red); b) IM model for two populations - *L. californicus* (yellow) and *L. americanus* (red). θ_i - effective population size of population *i*; $m_{i \rightarrow j}$ - migration from population *i* to population *j*; t_s - split between two populations at a time of divergence *s*. Estimated values of each of these parameters are presented in Table 5.

One of the main objectives of this work was to test the hypothesis of massive mtDNA introgression between *L. californicus* and *L. americanus*. Given that ancestral population size estimates and time estimates of the first split between *L. americanus* and the ancestral population of *L. californicus* and *L. townsendii* were poorly estimated,

greatly varying among replicate runs, the IMa2 analysis was also performed considering only *L. americanus* and *L. californicus* (Figure 9). These two species were estimated to diverge about 3.2 Mya. Estimates of effective population sizes were similar to those inferred from the three species IMa2 analysis. However, although very little in magnitude, significant gene flow was inferred from *L. californicus* to *L. americanus* being detected in all three replicate runs. The inferred values of migration were low, similar to those found among other *Lepus* species, e.g. in Europe (Melo-Ferreira *et al.* 2012). Parameter estimates are depicted in Table 7.

Table 7 - ML estimates (95% posterior density intervals in parentheses) of demographic parameters obtained with IMa2 between pairs of species.

Sp.1 - Sp.2	N_{e1} [a]	N_{e2} [a]	N_{eA} [a]	t [b]	$2Nm_{1\rightarrow2}$ [c]	$2Nm_{2\rightarrow1}$ [c]
Three-population analysis						
Lcf - Ltw	591'232 (441'016-773'343)	221'110 (148'169-328'355)	NC	2'274'412 (1'455'932-3'429'914)	NC	NC
Lcf - Lam	-	341'475 (263'478-441'618)	-	-	0.0177 (0.0177-0.1596)	0.0307 (0.0028-0.0140)
Ltw - Lam	-	-	-	-	0.0177 (0.0177-0.1241)	0.0115 (0.0115-0.0804)
Lam - Anc	-	-	NC	NC	NC	NC
Two-population analysis						
Lcf - Lam	588'343 (464'849-759'742)	374'575 (293'449-488'922)	191'019 (22'568-753'844)	3'213'257 (2'082'791-4'561'342)	*0.0584 (0.0195-0.1751)	0.0306 (0.0306-0.1528)

Notes: [a]Effective population size of population 1 (N_{e1}), 2 (N_{e2}) and the ancestral population (N_{eA}); [b]Time in years since species 1 and 2 split; [c]Population migration rate into population 1 ($2N_{m2\rightarrow1}$) and population 2 ($2N_{m1\rightarrow2}$) (significant values indicated *P < 0.05; Nielsen and Wakeley, 2001). NC – non-confidently estimated.

Coalescent Simulations

The effective population size and divergence time estimates obtained between the *L. americanus-L. californicus* under the IM model were used to simulate *cyt*b datasets which could result from the inferred speciation history from the nuclear loci. A mutation rate of 1.3 x 10^{-8} substitutions/site/generation estimated from the Lepus-*Oryctolagus* average distance was applied in the simulations. The expected distributions of the minimum pairwise uncorrected *p*-distances under a model with no gene flow were produced from the simulations. Under this conservative approach, the observed average pairwise distance between *L. americanus* PacNW2 cluster and *L. californicus* was lower than the 5th percentile of simulated minimum pairwise distances, which indicates that introgression, and not incomplete lineage sorting, is the most likely

responsible for the close similarity between the PacNW mtDNA clade and *L. californicus* (Figure 10).

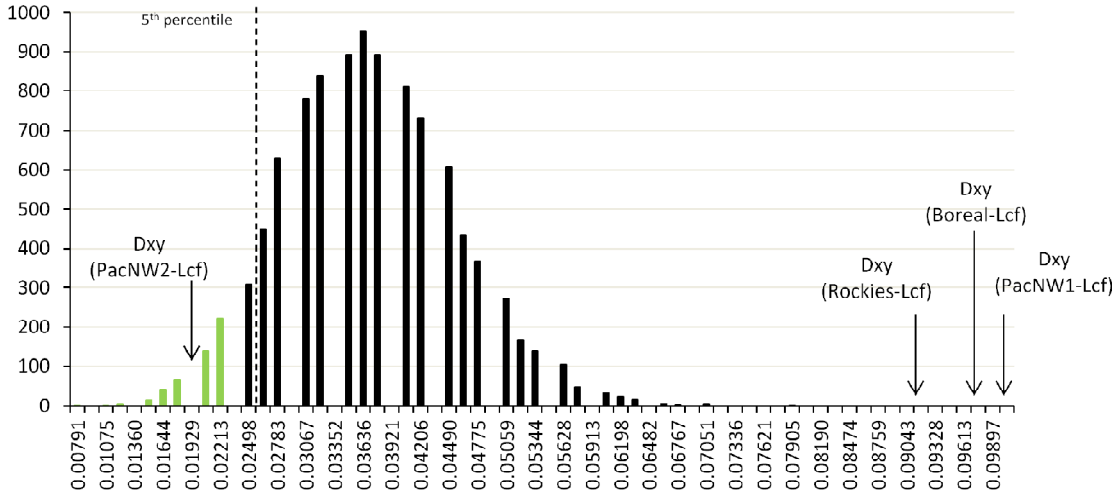


Figure 10 - Results of the coalescent simulations of mtDNA sequences from population parameters estimated with multiple nuclear loci. The distribution of the simulated minimum pairwise uncorrected *p*-distances between *L. americanus* and *L. californicus* is given by the bars - lowest 5th percent (green bars) and highest 95% (black bars). The observed Dxy values between *L. americanus* PacNW1 cluster and *L. californicus, L. americanus* PacNW2 cluster and *L. californicus, L. americanus* Rockies cluster and *L. californicus, L. americanus* Boreal cluster and *L. californicus* are given by the arrows

.

DISCUSSION

The reticulate evolution of hare species has been the subject of several studies with numerous cases of interspecific gene flow being described (reviewed in Alves et al, 2008). Along with their reticulate nature, this genus has undergone an extremely rapid and explosive radiation, rendering this genus as an ideal model to understand speciation under such conditions.

Although nuclear gene flow among hare species is not negligible, with gene flow being inferred in some species and in some cases leading to massive introgression (Farelo, 2011; Melo-Ferreira et al, 2011), the instances of mitochondrial DNA introgression are remarkable within this genus, both because of their taxonomic extent and geographic pervasiveness and prevalence in some of the introgressed species.

However, all cases of secondary introgression among hares have been described in Eurasian species, most of the times with the mountain hare, *L. timidus*, being involved as the donor species of mtDNA introgression (see Alves et al, 2008). Less attention has been given to the North American hares, which is reflected in the scarce number of studies centered on hares from this region. Most such studies focus on restricted areas (e.g. Burton et al, 2002; Tapia, 2010) or are limited in the number of markers and specimens used (Halanych et al, 1999), possibly causing the researcher to disregard many interesting evolutionary processes that have passed unnoticed. Indeed, a recent study on the snowshoe hare, *L. americanus,* covering its whole range of distribution, noticed that one mtDNA lineage of this species is more closely related to that of another North American hare, the black-tailed jackrabbit, *L. californicus*, that the remainder of the mtDNA variation within *L. americanus* (Cheng 2010).

This inference led to the suggestion that extensive mtDNA introgression might have occurred from *L. californicus* into *L. americanus*. However, the proper understanding of such possible reticulation events, separating it from alternative hypotheses of retention of incomplete lineage sorting, demands a strict knowledge of the speciation history of the involved species and of the biogeography of these organisms in the region. In this work we use a multilocus population-level approach using coalescent methods to investigate these issues.

Population Demography of North American hares

The climatic oscillations that characterized the Pleistocene, particularly through the last 700 ka, induced considerable shifts on species distributions with alternate periods of expansion and retraction (e.g. Hewitt, 2000, 2004). This led in many instances, to the establishment of secondary contact areas providing the opportunity for hybridization to occur. In hares, these demographic oscillations have often been suggested to have shaped the evolution of species (e.g.Melo-Ferreira et al., 2007), and that is likely the case in North America.

Given the strong association with the boreal forest, which has greatly shifted its range throughout the glacial periods, and the large distribution area, the snowshoe hare is likely to reflect such patterns. From the analysis of the MMD of pairwise differences at the mtDNA control region (D-loop), Cheng (2010) suggested the expansion of the Boreal and PacNW clusters 211 kya and 90 kya respectively. However, we found no consistent signs of expansion from the neutrality tests, and the MMD from the Cyt b fragment rejected the Sudden Expansion Model, either for the three clusters and for the species. However, these are single locus analysis and while MMD tests only one specific model of demography, neutrality tests present summary statistics which can be affected by other factors other than demography of the species. We have thus used a multilocus coalescent based approach, considering all nuclear and mitochondrial locus of our dataset, to reconstruct the demographic history by means of an EBSP analysis. This method has been shown to provide better estimates of population size under the stochastic coalescence process than single-locus analyses (Heled and Drummond, 2008). Nonetheless, also no strong evidence of demographic changes were found except for the Boreal cluster. The EBSP analysis for this cluster suggests an increase in the effective population size *ca*. 330 kya, which might reflect an expansion during the Pleistocene glacial-interglacial cycles.

Regarding the more warm-adapted *L. californicus*, the demographic reconstruction based on the multilocus EBSP analysis may suggest an expansion around 1 Mya. The Cyt b MMD of the pairwise distances of *L. californicus* California (CA) population also suggests that this population might have expanded. Similarly to *L. californicus*, we found evidence that *L. townsendii* might have started to expand at about 1 Mya. Interestingly, at the population level, we also found evidences from the MMD and the neutrality tests that an expansion could have occurred in the Wyoming (WY) population.

Although it is not the aim of this study to present an exhaustive survey of the population status of these species, we have also investigated the patterns of diversity at the population level. We have estimated *L. americanus* and *L. californicus* to be the most diverse species in accordance with the estimates of effective population sizes. However, at the population level, only *L. californicus* populations maintained this status of being the more diverse. Interestingly, within this species populations at the center of distribution tend to be more diverse than peripheral ones in keeping with a core-periphery hypothesis (Vucetich and Waite 2003; Eckert *et al.* 2008). While some *L. townsendii* populations have been considered extinct (Berger et al, 2006; 2008a; but see Gunther et al. 2009) we found no evidences of these or other *L. townsendii* populations being considerably less diverse than populations of the other two species (generally more diverse than *L. americanus* species). Within *L. americanus,* diversity was similar among clusters while interestingly, at the population level, Saskatchewan population (the only Boreal) generally presented the highest values of diversity overall *L. americanus* populations. Cheng (2010) suggested the existence of multiple Boreal refugia (eastern and western refugia) from where populations would have expanded, Saskatchewan possibly being in that area of contact of haplotypes coming from these refugia.

Speciation history of North American hares

The advances in molecular genetics have allowed obtaining larger amounts of genetic data. With the ease to employ a diverse suite of molecular markers, sequencing both mitochondrial and nuclear markers became commonplace and the number of studies reporting discordant patterns between mtDNA and nuclear markers is increasing. Gene tree polyphyly and incongruence among gene trees are frequently taken as evidence of interspecific gene flow (see Sang and Zhong 2000; Funk and Omland 2003). Particularly instances of mtDNA introgression have often been suggested based on the discordance between mtDNA and nuclear DNA phylogenies (e.g. Buckley *et al.* 2006; Bossu and Near 2009; Spinks and Shaffer 2009). Nevertheless, incongruence among markers could also be the result of other evolutionary phenomena. For example, the incomplete sorting of lineages alone can result in discrepancies between phylogenies. However, this alternative hypothesis is frequently not tested (e.g. Buckley *et al.* 2006; Bossu and Near 2009; Spinks and Shaffer 2009).

In this study we investigated a case of a putative massive mtDNA introgression from one North American hare species, *L. californicus*, to a partially sympatric species, *L. americanus*. A previous study (Cheng 2010) has reported that *L. americanus* PacNW populations possess mitochondrial lineages more closely related to those of *L. californicus* populations rather than to the lineages present in populations of the remainder *L. americanus* clusters (Boreal and Rockies). In keeping with these results, the same pattern is inferred from our analyzed mtDNA fragment (Figure 7). On the contrary, among the collection of nuclear DNA trees resulting from *BEAST MCMC chain, the monophyly of *L. americanus* is sustained in 99% of them. As reported for other hare species, mitochondrial-nuclear discordance and the non-monophyly of *L. americanus* at the mtDNA could have resulted of massive mtDNA introgression from *L. californicus* into PacNW region of *L. americanus* clade following secondary hybridization. However, the alternative hypothesis of incomplete lineage sorting along the speciation history of these species could not be discarded. To disentangle these two confounding factors we have first inferred the history of speciation of North American hares using eight nuclear DNA markers, determining relevant parameters of such history, taking into account the possibility of the occurrence of gene flow.

Along with the evidences of mtDNA introgression in Europe, generally from *L. timidus* into the other four hare species naturally occurring in Europe (*L. granatensis*, *L. europaeus*, *L. castroviejoi* and *L. corsicanus*), instances of limited nuclear gene flow

have been suggested to occur always in the same direction of that of the mtDNA (Melo-Ferreira *et al.* 2009, 2011, 2012). Also, mtDNA of Northern Iberian *L. europaeus* has been shown to be introgressed in some extent from *L. granatensis* which is paralleled by bidirectional nuclear gene flow (Freitas 2006; Melo-Ferreira *et al.* 2009). Moreover, nuclear gene flow from the former species to the latter has been inferred by Melo-Ferreira et al (2012), when considering *L. europaeus* specimens from a broader range over the European species distribution. Other instances of nuclear gene flow have been suggested between *L. granatensis* and *L. castroviejoi* (Melo-Ferreira *et al.* 2012). These cases suggest that mtDNA introgression in European hares is often accompanied by nuclear introgression either in the same or/and in the opposite direction.

The three species analyzed in this study have current partially overlapping distribution ranges, setting the conditions for hybridization to occur, and for the exchange of genetic material between these species. Thus it is possible that introgression, similarly to the cases in Europe, might extend to the nuclear loci accompanying that of the mtDNA. The analyses using the IM model, however, do not provide strong evidences for high nuclear gene flow. Using the three-population model, gene flow is never suggested. However, in the two-population model (*L. californicus* vs. *L. americanus*), nuclear gene flow was found to be significant in the same direction of that suggested for the mtDNA (from *L. californicus* into *L. americanus*) – Table 7. Given that the methods used to infer the phylogeny of species assume the absence of gene flow, this could influence such estimates. However, the levels of gene flow found here are very low (0.0584), unlikely to affect phylogeny reconstruction methods based on the multispecies coalescent (Eckert and Carstens, 2008). Overall, our results thus suggest that nuclear gene flow may have occurred into *L. americanus* during its speciation history, but at very low rates, a result that is similar to what was inferred among European hares (Melo-Ferreira *et al*. 2012).

The first split, as inferred from the *BEAST analysis, occurred between *L. americanus* and the ancestral of *L. californicus* and *L. townsendii*. Assuming that it might be represented by the time inferred for the split between *L. americanus* and *L. californicus*, from the IM two population analysis, the split of *L. americanus* would have occurred about 3.2 Mya (2 Mya - 4.6 Mya; HPD 95% confidence interval - Table 7). This result is consistent with the *BEAST divergence time estimates both when considering only the species and considering the three *L. americanus* clusters separately (1.6 Mya – 3.8 Mya and 1.6 Mya – 3.4 Mya HPD 95% confidence intervals, respectively; data not

shown). The IM estimate for the divergence between *L. townsendii* and *L. californicus* suggests that these species have diverged about 2.3 Mya (1.5 Mya – 3.4 Mya; HPD 95% confidence interval - Table 7). Again, this is consistent with the *BEAST estimates (7 kya – 1.8 Mya and 9 kya – 2.4 Mya HPD 95% confidence intervals, only considering the species or distinguishing the three *L. americanus* clusters respectively; data not shown). Within *L. americanus*, and according to the *BEAST estimate, the Boreal cluster was the first to diverge about 1.8 Mya (1 Mya – 2.6 Mya; HPD 95% confidence interval) from the ancestral of the remaining two clusters. Although this estimate could reflect an older divergence between the Boreal and the PacNW/Rockies clusters relative to that between *L. californicus* and *L. townsendii* we must note that the HPD 95% confidence intervals, not only of these estimates but also of the divergence *L. americanus* – ancestral of *L. californicus* and *L. townsendii*, greatly overlap which reflects a possible rapid radiation of these species. Considering such a rapid radiation, incomplete lineage sorting is likely to affect the different loci, as reflected by the different phylogenies estimated among nuclear loci.

Although *L. americanus* has the largest distribution of North American hares, we estimated *L. californicus* to have the largest effective size among the three species being almost two fold higher than that of the former. This could reflect the different responses of these two species to the glacial cycles. *L. americanus*, having its distribution more to the north, is likely to have been more susceptible to the repeated advance and retreat of the glaciers, which might have forced this species to retreat to refugia, causing bottlenecks. On the contrary, the more southern *L. californicus* would be less exposed to the climate oscillations, possibly suffering less fluctuation in its distribution and allowing it to maintain more stable effective population sizes. *L. townsendii* had the lowest N_e of the three species. Given that it has a similar distribution area to that of *L. californicus*, it is more likely that it was its evolutionary history and not its distribution that conditioned *L. townsendii* to present the lowest effective sizes.

Extensive mtDNA introgression from L. californicus into L. americanus

By simulating the evolution of the mitochondrial fragment under a model that assumes no gene flow, using the population demography estimates from the isolation with migration analyses using nuclear DNA, we were able to test whether the observed pattern found in the mitochondrial phylogeny could be explained by the coalescent process alone or by gene flow. The results of the coalescent simulations of the minimum pairwise distances between *L. americanus* and *L. californicus* suggest that the empirical average pairwise uncorrected *p*-distance between the PacNW2 *L. americanus* clade and *L. californicus* is smaller than that expected under a pure lineage sorting scenario (Figure 10), thus supporting the introgression hypothesis. A minor portion of the empirical individual distances (18%), however, were above the 5th percentile of the simulated minimum distances. Since the PacNW2 individuals are grouped into an mtDNA phylogenetic clade (Figure 6), it is unlikely that these exceptions may correspond to non-introgressed individuals. In fact, we must note that when calculating the Cyt *b* mutation rate calibrating with the *Lepus-Oryctolagus* uncorrected p-distance, we tended to underestimate this rate due to the occurrence of homoplasy. Indeed, Halanych and Robinson (1999) suggested the cytochrome *b* to be likely affected by saturation at the inter-generic level and thus the simulated distribution of *L. americanus-L. californicus* distances using this uncorrected rate are also likely underestimated. This property of the methodology is reflected by the fact that the Boreal, Rockies and PacNW1 clades presented empirical average pairwise distances well above the range of the simulated distribution (Figure 10).

We have thus additionally simulated the expected minimum pairwise distances between *L. americanus* and *L. californicus* using a mutation rate estimated from *Lepus-Oryctolagus* distances calculated using the adequate mutation model suggested by jModeltest (TPM3uf+G, with gamma shape = 0.172; distances estimated in PAUP, Swofford, 2001). This is expected to yield a more realistic mutation rate. Using this approach, the simulated distribution of minimum distances between *L. americanus* and *L. californicus* shifts towards higher divergence values. Indeed, the range of this distribution now coincides with the empirical average distances between both the Rockies, Boreal and PacNW1 mtDNA clades and *L. californicus*, which may indicate that a more realistic model was now used. In this approach, which is indeed the exact same used by Melo-Ferreira *et al.* (2012), not only the average empirical pairwise distance between the PacNW2 *L. americanus* clade and *L. californicus* is lower than

expected, but all individual pairwise distances are so, reinforcing the inference of mtDNA introgression.

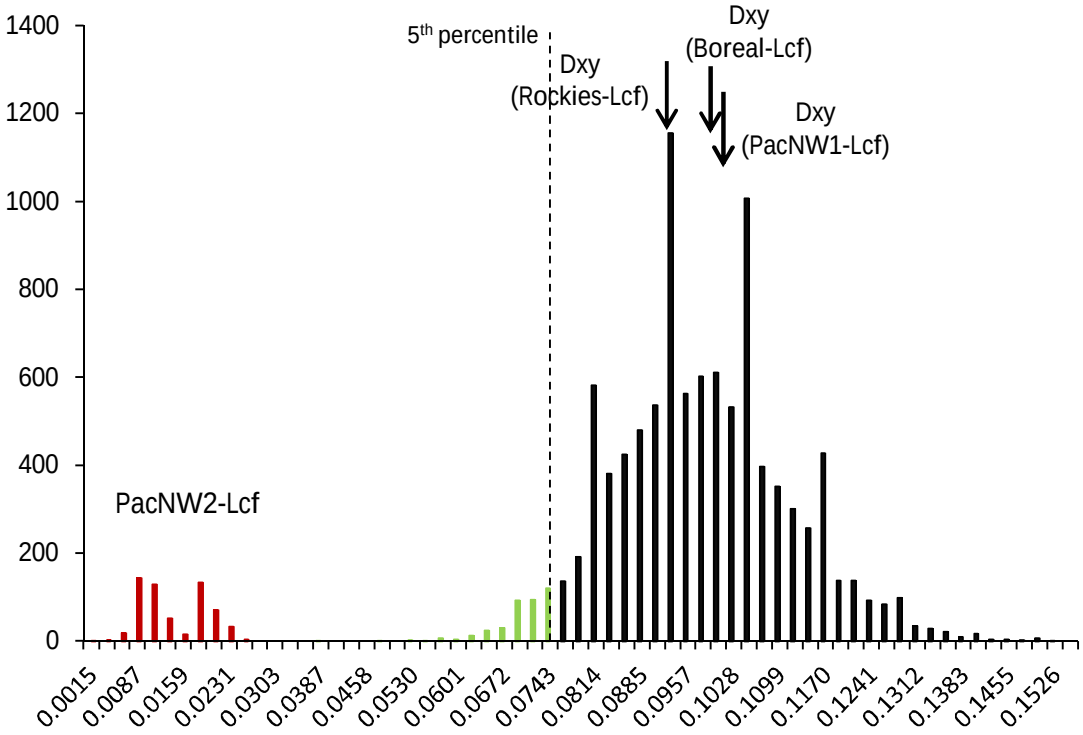


Figure 11 - Results of the coalescent simulations of mtDNA sequences from population parameters estimated with multiple nuclear loci. The distribution of the simulated minimum pairwise corrected *p*-distances between *L. americanus* and *L. californicus* is given by the blue (lowest 5th percent) and black bars (highest 95%) and the distribution of the observed minimum pairwise corrected distances between *L. americanus* PacNW2 cluster and *L. californicus* (Lcf) is given by the red bars. The observed Dxy values between *L. americanus* PacNW1 cluster and *L. californicus, L. americanus* Rockies cluster and *L. californicus and L. americanus* Boreal cluster and *L. californicus* are given by the arrows;

In addition, under a scenario of incomplete lineage sorting the discordance between nuclear and mtDNA is not expected to leave any predictable geographic pattern. In contrast, introgression resulting from secondary contact or gene flow along the process of divergence of species may result in a strong biogeographical discordance between these two marker types (Toews and Brelsford, 2012; see references therein). Our results show that, even though the mtDNA and microsatellite clusters described by Cheng (2010) are generally coincident, an almost complete replacement of the mtDNA of the PacNW cluster occurred via introgression. Some sequenced haplotypes, which we here depict as PacNW1, may correspond to the native PacNW mtDNA type. Analyzing the geographical distribution and frequencies of the mtDNA lineages present in the PacNW cluster, the putative native PacNW lineage (PacNW1) is only found in

the northernmost of the analyzed PacNW populations, Washington 4 (WA4), while the PacNW2 lineage occurs in all populations, its frequency increasing towards the south where *L. americanus* PacNW cluster and *L. californicus* ranges come into contact. Such a cline is more easily explained by a scenario of introgression rather than of incomplete sorting of the lineages, although more populations should be sampled for a better resolution of the geographical distribution of the lineages' frequencies.

Interestingly, *L. californicus* and *L. americanus* PacNW2 do not share mtDNA haplotypes, and the introgressed PacNW2 forms a compact clade. This suggests that introgression occurred at a specific point in the past and no gene flow has likely occurred since then. We attempted to date the introgression applying a simple mutation-drift model do the divergence between the PacNW and *L. californicus* haplotypes ($k = 2\mu t + \pi$, where k is the average number of pairwise differences among alleles from the two groups, μ is the mutation rate, π is the nucleotide diversity of the ancestral population averaged from the descendants, and t is the time of divergence between groups). Using this rationale, introgression was estimated to have occurred about 680'000 years ago which coincides with the beginning of a glacial cycle during the Pre-Illinoian Stage (note that molecular calibrations must be dealt with care). Is thus possible that introgression has occurred in result of past climatic alterations impact on species distribution. Such an influence of the climate is well reflected in the evolution of some hare species. Indeed, although introgression among hares is more often found in areas of present contact and to be an ongoing phenomenon (Thulin and Tegelström 2002; Thulin *et al.* 2006; Pietri *et al.* 2011), instances of ancient introgression have also been recognized in currently allopatric species (Alves *et al*. 2003; Melo-Ferreira *et al*. 2005). Although *L. timidus* went locally extinct in the Iberian Peninsula, a large portion of mtDNA haplotypes found in *L. granatensis*, *L. europaeus* and *L. castroviejoi* are of *L. timidus* origin, being found in high frequencies at some parts of *L. granatensis* distribution and reaching quasi-fixation in *L. europaeus* (Melo-Ferreira *et al*. 2005). In *L. castroviejoi*, two introgression events from *L. timidus* are thought to have occurred. The first is suspected to have led to the complete mitochondrion replacement in the ancestral of *L. castroviejoi* and its sister species *L. corsicanus*, which occurs in the Italic Peninsula from were *L. timidus* is also currently absent. A second introgression event would have only affected *L. castroviejoi* (Alves *et al*. 2008; Melo-Ferreira *et al*. 2012). Similarly to the cases observed in the Iberian and Italic Peninsulas, the observed pattern in *L. americanus* PacnNW2 cluster could result of ancient introgression from *L. californicus*, although this may also result from gene flow from a non-sampled lineage. We have investigated this possibility by aligning all

Cyt *b* sequences of the three species included in this study available at GenBank to our dataset. However, none of these sequences provided any new lineage that could be more closely related to *L. americanus* PacNW2 cluster than those included in our study. As so, if gene flow occurred from a non-sampled *L. californicus* lineage, this either has never been sampled or has gone extinct. We have also considered the possibility that mtDNA introgression into *L. americanus* PacNW2 cluster occurred from a non-sampled donor species but, again, none of lineages from other North American species among the sequences available in GenBank proved to be more closely related to the PacNW2 mtDNA group.

Given the deep divergence estimated between *L. americanus* PacNW2 lineages and current known lineages, either present in *L. californicus* or in the other North American species, the most likely scenario of introgression into *L. americanus* PacNW2 cluster is that of an ancient introgression from a *L. californicus*-type.

Causes and Consequences of Introgression

As an invasion of the genome, introgression is sometimes seen as deleterious. For example, cases of introductions in which introduced and native species hybridize can generate conservation issues (Rhymer and Simberloff 1996; Allendorf *et al.* 2001). However, as a natural process, hybridization and introgression have been suggested to be an import mean of evolution of species (e.g. Arnold, 1997; Mallet, 2005). Leading to diversification, introgression can result in new and advantageous genomic combinations and a mean for species to acquire novel or sometimes lost traits (e.g. Rieseberg, 2009)

Several hypotheses could be considered that might explain the massive mtDNA introgression inferred at the PacNW cluster. According to our estimate of the time of mtDNA introgression this probably occurred during a glacial cycle. Given the retreat of the boreal forest following the advance of the glaciers, *L. americanus* distribution would be likely shifting south possibly out-competing *L. californicus*. The competitive replacement of a species by another that is expanding its range, when accompanied by hybridization at the front of the expansion wave, where the latter is rarer, has been shown to likely lead to introgression into the genome of the invading species (Currat et al, 2008; Excoffier and Ray, 2008). Moreover, it predicts introgression prevalence to be higher for markers transmitted by the least dispersing sex. Assuming *L. americanus* females to disperse less than males (which is a generality in mammals), this would be compatible with the observed pattern. However, a study on *L. americanus* dispersal found no evidence of sex-biased dispersal (Gillis and Krebs, 1999). Another possible explanation is that *L. californicus* mtDNA might confer an advantage relatively to the mtDNA of the native species. Selection favoring *L. californicus* mtDNA would facilitate introgression into *L. americanus*. Other possible factors that might explain introgression, e.g. asymmetric reproductive behavior in situations of hybridization, have already been explored in other works concerning species of the *Lepus* genus (e.g. Melo-Ferreira *et al*. 2009).

Interestingly, in the populations of the *L. americanus* PacNW cluster, while some specimens undergo seasonal coat color changes from a brown summer to a white winter morph (the rule in this species), others present an invariable phenotype all year round. Although other factors can lead to this pattern, given the history of reticulation that has affected this *L. americanus* evolutionary group (massive in the mtDNA, but likely also occurring at the nuclear genome, according to our results), it is tempting to

hypothesize that introgression might have contributed to the peculiar phenotypes. In hares for example, in areas where *L. timidus* and *L. europaeus* hybridize, specimens exhibiting intermediate morphological characters have been tentatively identified as hybrids (Notini 1941), although phenotypic plasticity exists within both species (Thulin *et al*. 2006). Introgression may have led to this new phenotype, if affected genomic regions involved in the coat-color phenotype. This is an issue that deserves future investigation.

Conclusions and Future Prospects

In this study we provide a better understanding on the evolutionary history of three North American hare species – *L. americanus*, *L. californicus* and *L. townsendii*. Our results suggest that these species have undergone a recent and possibly rapid radiation, and that although generally not very frequent considering the nuclear markers analyzed here, gene flow may have played a role during the divergence of these species. Contrary to that found when analyzing the nuclear markers, we show that massive mtDNA introgression occurred between *L. californicus* and *L. americanus*, the introgressed lineage being almost fixed in the PacNW region of the *L. americanus* range. These results add to the debate of whether neutral demography or natural selection are the most important engine of mtDNA introgression among mammals, and should be the object of future studies.

This work thus reinforces the porous nature of species boundaries and the importance of understanding gene flow to infer the causes of genome divergence. It sets the conditions to better understand the genomics of speciation – are some regions of the genome more prone to introgress than others? Which regions are impermeable to introgression and may have thus played an important part in the establishment of reproductive isolation? What is the relative role of demography and natural selection in the divergence of the genome? These questions are pertinent for the understanding of the process of species formation and can be addressed in the future using *L. americanus* and *L. californicus* as models.

Since the *L. americanus* populations from the region most affected by gene flow, show an interesting coat color phenotype – while *L. americanus* generally undergoes coat color change during the spring and autumn molts, in these populations the phenotype is variable – it is tempting to hypothesize that genome reticulation may be the cause of such phenotype. Although speculative at present, this is a hypothesis that should be addressed in the future, also because the genetic bases of coat color change in *L. americanus* is the focus of ongoing research.

Bibliography

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, **16**, 613–622.

Alves P, 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 P, Hackländer K (2008) Lagomorph species: geographical distribution and conservation status. In: *Lagomorph biology: evolution, ecology, and conservation* (eds Alves P, Hackländer K, Ferrand N), pp. 395–405. Springer. Berlin, Germany.

Alves P, Harris DJ, Melo-Ferreira J *et al.* (2006) Hares on thin ice: introgression of mitochondrial DNA in hares and its implications for recent phylogenetic analyses. *Molecular Phylogenetics and Evolution*, **40**, 640–641.

Alves PC, Melo-Ferreira J, Branco M *et al.* (2008) Evidence for genetic similarity of two allopatric European hares (Lepus corsicanus and L. castroviejoi) inferred from nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, **46**, 1191–7.

Alves P, 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 of London. Series B, Biological sciences*, **363**, 2831–2839.

Andersson JO (2005) Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences*, **62**, 1182–1197.

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

Aris-Brosou S, Excoffier L (1996) The Impact of Population Expansion DNA Sequence Polymorphism and Mutation Rate Heterogeneity on. *Molecular Biology and Evolution*, **13**, 494–504.

Arnold M (1997) *Natural Hybridization and Evolution*. Oxford University Press, Oxford, UK.

Avise J (2004) *Molecular Markers, Natural History, and Evolution*. Sinauer & Associates, Sunderland, Massachusetts.

Avise J (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. , 513–518.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.

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

Berger J, Berger KM, Brussard PF *et al.* (2006) Where have all the rabbits gone? Summary and Recommendations from a Workshop Held in Grand Teton National Park, September 23-24, 2005. *Wildlife Conservation Society, Bronx, New York*.

Blier PU, Dufresne F, Burton RS (2001) Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends in genetics : TIG*, **17**, 400–6.

Bossu CM, Near TJ (2009) Gene trees reveal repeated instances of mitochondrial DNA introgression in orangethroat darters (percidae: etheostoma). *Systematic Biology*, **58**, 114–29.

Bray Y, Devillard S, Marboutin E, Mauvy B, Péroux R (2007) Natal dispersal of European hare in France. *Journal of Zoology*, **273**, 426–434.

Buckley TR, Cordeiro M, Marshall DC, Simon C (2006) Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (Maoricicada Dugdale). *Systematic Biology*, **55**, 411–25.

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.

Butlin RK, Galindo J, Grahame JW (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 2997–3007.

Carneiro M, Ferrand N, Nachman MW (2009) Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (Oryctolagus cuniculus). *Genetics*, **181**, 593–606.

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720–9.

Cheng E (2010) Climate change, gene flow, and the legendary synchrony of snowshoe hares. PhD Thesis, University of Montana, Montana, USA.

Coyne J, Orr H (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–20.

Dobzhansky T (1937) *Genetics and the Origin of Species. 364 pp. 1st ed.* Columbia Univiversity Press, New York.

Doiron S, Bernatchez L, Blier PU (2002) A Comparative Mitogenomic Analysis of the Potential Adaptive Value of Arctic Charr mtDNA Introgression in Brook Charr Populations (Salvelinus fontinalis Mitchill). *Molecular Biology and Evolution*, **19**, 1902–1909.

Dowling TE, Secor CL (1997) The Role of Hybridization and Introgression in the Diversification of Animals. *Annual Review of Ecology and Systematics*, 593–619.

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

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, **7**, 214.

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**, 1170–88.

Edwards S V, 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**, 5936–41.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 853–864.

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

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, **23**, 347–51.

Farelo L (2011) The role of the X chromosome on reticulate evolution of hares (Lepus spp.). Faculdade de Ciências da Universidade do Porto.

Flinders JT, Chapman JA (2003) Black-tailed Jackrabbit (Lepus californicus and Allies). In: *Wild Mammals of North America: Biology, Management, and Conservation* (eds Feldhamer GA, Thompson BC, Chapman JA), pp. 126–146. The Johns Hopkins Univesity Press, Baltimore, Maryland, USA.

Flot J (2010) Seqphase: a Web Tool for Interconverting Phase Input/Output Files and Fasta Sequence Alignments. *Molecular ecology resources*, **10**, 162–6.

Flot J, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes*, **6**, 627–630.

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

Flux J, Angermann R (1990) The Hares and Jackrabbits. In: *Rabbits, Hares and Pikas: Status Survey and Conservation Action Plan* (eds Chapman JA, Flux J), pp. 61–94. Gland, Switzerland: International Union for Conservation of Nature and Natural Resources.

Fredsted T, Wincentz T, Villesen P (2006) Introgression of mountain hare (Lepus timidus) mitochondrial DNA into wild brown hares (Lepus europaeus) in Denmark. *BMC ecology*, **6**, 17.

Freitas H (2006) Natural hybridization between the Iberian hare (Lepus granatensis) and the brown hae (L. europaeus) in northern Iberian Peninsula. MSc thesis. Faculdade de Ciências da Unoversidade do Porto.

Fu Y-X (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchiking and Background Selection. *Genetics*, **147**, 915–925.

Funk DJ, Omland KE (2003) SPECIES-LEVEL PARAPHYLY AND POLYPHYLY: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 397–423.

Galtier N, Nabholz B, Glémin S, Hurst G (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular ecology*, **18**, 4541–50.

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.

Gavrilets S (2003) Models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.

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

Gillis EA, Krebs CJ (1999) Natal dispersal of snowshoe hares during a cyclic population increase. *Journal of Mammalogy*, **80**, 933–939.

Good JM, Hird S, Reid N *et al.* (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular Ecology*, **17**, 1313–27.

Grant P, Grant B (1997) Hybridization, sexual imprinting, and mate choice. *The American Naturalist*, **149**, 1–28.

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.

Halanych KM, Demboski JR, Van Vuuren BJ, Klein DR, Cook J a (1999) Cytochrome b phylogeny of North American hares and jackrabbits (Lepus, lagomorpha) and the effects of saturation in outgroup taxa. *Molecular Phylogenetics and Evolution*, **11**, 213–221.

Halanych KM, Robinson TJ (1999) Multiple substitutions affect the phylogenetic utility of cytochrome b and 12S rDNA data: examining a rapid radiation in leporid (Lagomorpha) evolution. *Journal of molecular evolution*, **48**, 369–379.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*, **41**, 95–98.

Hamill RM, D Doyle, Duke JE (2007) Microsatellite analysis of mountain hares (Lepus timidus hibernicus): Low genetic differentiation and possible sex-bias in dispersal. *Journal of Mammalogy*, **88**.

Hardy GH (1908) Mendelian Proportions in a Mixed Population. , **28**, 49–50.

Hasegawa M, Kishino H, Yano T (1985) Journal of Molecular Evolution ~). *Journal of Molecular Evolution*, **22**, 160–174.

Heled J (2010) Extended Bayesian Skyline Plot tutorial.

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**, 570–80.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt G (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 183–195.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular biology and evolution*, **27**, 905–20.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. *Genetics*, **167**, 747–60.

Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular ecology resources*, **11**, 423–34.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Hudson RR, Kreitman M, Aguadé M (1987) A Test of Neutrar Evolution Based on Nucleotide Data. *Genetics*, **116**, 153–159.

Joly S, McLenachan P a, Lockhart PJ (2009) A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist*, **174**, E54–70.

Kimura M (1968) Evolutionary rate at the molecular level. *Nature*, **217**, 624–626.

Kimura M (1969) THE NUMBER OF HETEROZYGOUS NUCLEOTIDE SITES MAINTAINED IN A FINITE POPULATION DUE TO STEADY FLUX OF MUTATIONS. *Genetics*, **61**, 893–903.

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.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)*, **25**, 1451–2.

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

Linder CR, Rieseberg LH (2004) Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany*, **91**, 1700–1708.

Lopez-Martinez N (2008) The Lagomorph Fossil Record and the Origin of the European Rabbit. In: *Lagomorph Biology: Evolution, Ecology, and Conservation* (eds Alves P, Ferrand N, Hackländer K), pp. 27–46. Springer. Berlin, Germany.

Machado C a, Haselkorn TS, Noor M a F (2007) Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in Drosophila pseudoobscura and D. persimilis. *Genetics*, **175**, 1289–306.

Macholán M, Munclinger P, Sugerková M *et al.* (2007) Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution*, **61**, 746–771.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–37.

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

Matthee C, Van Vuuren B, Bell D, Robinson T (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.

Mayr E (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Columbia University Press, New York.

Mayr E (1963) *Animal species and evolution.* Harvard University Press, Cambridge, MA.

Melo-Ferreira J, Alves P, 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**, 2643–58.

Melo-Ferreira J, Alves P, Rocha J, Ferrand N, Boursot P (2011) Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? *Evolution; international journal of organic evolution*, **65**, 1956–68.

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, Boursot P, Randi E *et al.* (2007) The rise and fall of the mountain hare (Lepus timidus) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Molecular Ecology*, **16**, 605–18.

Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves P (2005) Invasion from the cold past: extensive introgression of mountain hare (Lepus timidus) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, **14**, 2459–64.

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.

Moore W (1995) Inferring phylogenies from mtDNA variation - mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49**, 718–726.

Murray D, Smith A (2008) Lepus americanus. *IUCN 2012. In IUCN Red List of Threatened Species. Version 2012.1.*

Nei M (1987) *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–96.

Notini G (1941) Om harens biologi. *Sven Jägareförbundets Medd*, **4**, 1–192.

Payseur B a, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064–2078.

Pietri C, Alves PC, Melo-Ferreira J (2011) Hares in Corsica: high prevalence of Lepus corsicanus and hybridization with introduced L. europaeus and L. granatensis. *European Journal of Wildlife Research*, **57**, 313–321.

Pinho C, Hey J (2010) Divergence with Gene Flow: Models and Data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215–230.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular biology and evolution*, **25**, 1253–6.

De Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In: *Endless Forms: Species and Speciation* (eds Howard D, Berlocher S), pp. 57–75. Oxford Univiversity Press, Oxford, U.K.

De Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, **102 Suppl** , 6600–7.

Rambaut A, Drummond AJ (2007) Tracer v1.5.

Rand DM, Haney R a, Fry AJ (2004) Cytonuclear coevolution: the genomics of cooperation. *Trends in ecology & evolution*, **19**, 645–53.

Rannala B, Yang Z (2003) Using DNA Sequences From Multiple Loci. , **1656**, 1645–1656.

Ray N, Currat M, Excoffier L (2003) Intra-Deme Molecular Diversity in Spatially Expanding Populations. *Molecular Biology and Evolution*, **20**, 76–86.

Rhymer JM, Simberloff D (1996) Extinction By Hybridization and Introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rieseberg LH (2009) Evolution: replacing genes and traits through hybridization. *Current Biology*, **19**, 119–122.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid Zones and the Genetic Architecture of a Barrier to Gene Flow Between Two Sunflower Species.

Rogers A, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

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

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources*, **8**, 103–6.

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.

Sang T, Zhong Y (2000) Testing hybridization hypotheses based on incongruent gene trees. *Systematic Biology*, **49**, 422–434.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Schwenk K, Brede N, Streit B (2008) Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B*, **363**, 2805–2811.

Sequeira F, Sodré D, Ferrand N *et al.* (2011) Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads Rhinella marina and R. schneideri inferred from mtDNA and nuclear markers. *BMC evolutionary biology*, **11**, 264.

Slatkin M, Richard R (1991) Pairwise Comparisons of Mitochondrial. *Genetics Society of America*, **129**, 555–562.

Small RL, Cronn RC, Wendel JF (2004) Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany*, **17**, 145–170.

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**, 449–62.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American journal of human genetics*, **68**, 978–89.

Swofford LD (2001) PAUP* phylogenetic analysis using parsimony (and other methods).

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1989a) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–95.

Tajima F (1989b) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.

Takahata N, Slatkin M (1984) Mitochondrial gene flow. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 1764–1767.

Tapia I (2010) Genetic diversity and connectivity of white-tailed jackrabbit populations in Iowa with notes on seasonal home ranges. Iowa State University.

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**, 4673–80.

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

Thulin C, Jaarola M, Tegelström H (1997) The occurrence of mountain hare mitochondrial DNA in wild brown hares. *Molecular ecology*, **6**, 463–467.

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

Thulin C, Tegelström H (2002) Biased geographical distribution of mitochondrial DNA that passed the species barrier from mountain hares to brown hares (genus Lepus): an effect of genetic incompatibility and mating behaviour? *Journal of Zoology*, **258**, 299–306.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*.

Vucetich JA, Waite TA (2003) Spatial patterns of demography and genetic processes across the species ' range : Null hypotheses for landscape conservation genetics. *Conservation Genetics*, **4**, 639–645.

Weinberg (1908) On the demonstration of heredity in man. In: *Papers on human genetics* (ed Boyer S). Prentice Hall, Englewood Cliffs, NJ (1963).

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–3.

Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Yamada F, Takaki M, Suzuki H (2002) Molecular phylogeny of Japanese Leporidae, the Amami rabbit Pentalagus furnessi, the Japanese hare Lepus brachyurus, and the mountain hare Lepus timidus, inferred from mitochondrial DNA sequences. *Genes & Genetic Systems*, **77**, 107–116.

Yang Z (2006) *Computational Molecular Evolution*. Oxford University Press, Oxford, UK.

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**, 9264–9.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, Helianthus annuus and H. petiolaris. *Genetics*, **175**, 1883–93.

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

APPENDIX

Appendix 1 - Final Dataset

Species	Common name	Cluster[a]	Pop	Sample	Sex
Lepus americanus	Snowshoe hare	Boreal	**SK1**	GORDON1	M
				GORDON7	M
				GORDON25	M
				WEBER1	F
				WEBER2	F
				WEBER3	M
		Rockies	**OR2**	STRAUSER85	M
				STRAUSER88	M
				STRAUSER89	F
				STRAUSER92	F
				STRAUSER94	M
				STRAUSER95	M
				STRAUSER97	M
				STRAUSER101	M
			WY1	MILLS R1405	F
				MILLS R1573	M
				MILLS R1676	F
				MILLS R1791	F
				NBERG6	M
				NBERG10	M
				NBERG29	M
				NBERG33	M
		PacNW	**WA4**	MACCRAC73	M
				MACCRAC83	F
				MACCRAC85	F
				MACCRAC94	F
				MACCRAC102	M
				MACCRAC105	F
				STRAUSER34	M
				STRAUSER35	M
				STRAUSER36	F
				STRAUSER40	M
			WA1	CHENG14	F
				CHENG15	M
				CHENG19	M
				MACCRAC77	F
				MACCRAC80	M
				MACCRAC82	M
				STRAUSER51	M
				STRAUSER52	F
			CA1	CHENG9	F
				CHENG11	F
				CHENG12	F
				CHENG13	M
				CHEYNE1	M
				SIMONS3	M
				SIMONS4	F
				SIMONS5	M

(continues)

Lepus californicus	Black-tailed jackrabbit	**Oregon**	BURKE2	F
			BURKE3	F
			BURKE5	F
			BURKE6	F
			BURKE7	F
			HENNINGS1	M
			HENNINGS2	F
			HENNINGS3	M
			HENNINGS4	M
			HENNINGS5	F
		California	BURKE11	M
			BAUER1	M
			BAUER2	F
			BAUER7	F
			BAUER9	M
			BAUER14	M
		Texas	DOWLER1	M
			DOWLER2	M
			DOWLER5	M
			DOWLER7	F
			DOWLER9	M
			DOWLER10	M
			DOWLER11	M
			DOWLER15	M
		Arizona	LCF.TUC.2055	M
			LCF.TUC.2056	M
			LCF.TUC.2057	M
			LCF.TUC.2058	M
			LCF.TUC.2059	M
			LCF.TUC.2060	M
Lepus townsendii	White-tailed jackrabbit	**Montana**	LTW.MTA.2570	F
			LTW.MTA.2571	F
		Yellowstone	LTW.YLL.2575	M
			LTW.YLL.2576	M
			LTW.YLL.2577	M
			LTW.YLL.2578	M
			08-100	M
		GTP	LTW.YLL.2572	F
		Wyoming	08RT11	F
			06RT26	M
			08RT34	F
			06RT37	M
			08RT40	F
			08RT41	M
			06RT42	M
			08RT46	M
Sylvilagus audubonii	Desert cottontail		SAD1	M
			SAD2	F
Oryctolagus cuniculus[b]	European rabbit		OC	M

Notes:
[a]clusters as define by Cheng (2010) from microsatellite data
[b]Sequences obtained from GenBank. The accession numbers are: AJ001588 (CYTB), JN037052 (SPTBN1), JN037024 (PRKCI), JN036940 (DARC), JN036996 (KITLG), JN037078 (TF),

Appendix 2- PCR primers, sequencing primers and PCR annealing temperatures used for the amplification of each X-linked marker.

	Loci	Primer type[a]	PCR Primers	Reference	AT (°C)
			Forward/Reverse (5' – 3')		
1	CYTB	PCR/S	AGCCTGATGAAACTTTGGCTC	Alves et al. 2003	56
		PCR/S	GGATTTTATTCTCGACTAAGC	Alves et al. 2003	
		PCR/S	GTTGGCAGGGGTGTAGTTGT	this work	
2	POLA1	PCR/S	GGTATTTCTGTTTGGCAAGGTTTG	Carneiro et al. 2010	57[b]
		PCR/S	CTTGGACTTGAATTTCATGATTC	Carneiro et al. 2010	
3	GRIA3	PCR/S	CTCAGATCAGCAAATCAGCAATG	Carneiro et al. 2010	57[b]
		PCR/S	CATAGGCTAAGTCTACACAATAG	Carneiro et al. 2010	
4	SRY	PCR/S	CTGTTGCAGCATGCTTTGAG	Melo-Ferreira et al. 2009	56[b]
		PCR/S	GATTTGACGAATGCCAAGTGTTTC	Melo-Ferreira et al. 2009	
		S	ACAGCAAGGTGCAAAACAAGAA	Melo-Ferreira et al. 2009	
5	SPTBN	PCR/S	CTCTGCCCAGAAGTTTGCAAC	Matthee et al. 2004	65
		PCR/S	TGATAGCAGAACTCCATGTGG	Matthee et al. 2004	
6	PRKCI	PCR/S	AAACAGATCGCATTTATGCAAT	Matthee et al. 2004	58
		PCR/S	TGTCTGTACCCAGTCAATATC	Matthee et al. 2004	
7	DARC	PCR/S	CTCTCAGTTGACCCAAATTC	Melo-Ferreira et al. 2009	56
		PCR/S	GCCTTTAATTCAGGTTGACG	Melo-Ferreira et al. 2009	
8	KITLG	PCR/S	AAATATCAGTCTTGAATCTTAC	Matthee et al. 2004	56
		PCR/S	TTTTAGATGAATTACAGTGTCC	Matthee et al. 2004	
9	TF	PCR/S	GCCTTTGTCAAGCAAGAGACC	Matthee et al. 2004	55
		PCR/S	CACAGCAGCTCATACTGATCC	Matthee et al. 2004	

Notes:
[a]PCR - primers used for PCR; S - primers used for sequencing amplified fragments
[b]PCR amplification of this locus required an initial touchdown phase with a decrease of the annealing temperature of 0.5ºC per cicle starting at 64ºC.

Appendix 3 - Tests of Hardy-Weinberg equilibrium and heterozygosity deficiency and excess

	SPTBN	PRKCI	DARC	KITLG	TF	POLA1	GRIA3	Het def	Het exc
WA4n	N.inf .	0.78 (0.00)	1.00 (-)	1.00 (-)	1.00 (-)	N.inf .	N.inf .	0.29 (0.00)	0.82 (0.00)
SK1	N.inf .	0.20 (0.00)	1.00 (-)	N.inf .	0.62 (0.00)	N.inf .	N.inf .	0.16 (0.00)	0.89 (0.00)
WY1	N.inf .	N.inf .	N.inf .	0.08 (-)	N.inf .	N.inf .	N.inf .	0.08 (0.00)	1.00 (0.00)
OR2	N.inf .	*0.01 (-)	1.00 (-)	1.00 (-)	N.inf .	N.inf .	1.00 (-)	0.76 (0.00)	0.26 (0.00)
CA1	0.4286 (-)	1.00 (-)	0.65 (-)	1.00 (-)	N.inf .	N.inf .	N.inf .	0.38 (0.00)	0.62 (0.00)
WA1	N.inf .	N.inf .	N.inf .	0.71 (-)	0.49 (-)	0.20 (-)	1.00 (-)	0.75 (0.00)	0.30 (0.00)
OR	0.14 (0.00)	0.24 (0.00)	0.32 (-)	0.22 (0.00)	1.00 (-)	N.inf .	1.00 (-)	0.68 (0.00)	0.32 (0.00)
CA	*0.03 (0.00)	0.77 (-)	0.65 (0.00)	1.00 (-)	0.08 (0.00)	N.inf .	N.inf .	0.09 (0.00)	0.91 (0.00)
TX	*0.04 (-)	0.19 (0.00)	0.76 (0.00)	1.00 (-)	0.49 (0.00)	N.inf .	N.inf .	0.56 (0.00)	0.45 (0.00)
AZ	0.41 (0.00)	0.83 (0.00)	1.00 (-)	0.3143 (-)	0.47 (0.00)	N.inf .	N.inf .	0.24 (0.00)	0.77 (0.00)
WY1	0.1385 (-)	0.29 (0.00)	0.44 (-)	1.00 (-)	1.00 (-)	1.00 (-)	1.00 (-)	1.00 (0.00)	0.67 (0.00)
MT	N.inf .	N.inf .	N.inf .	N.inf .	N.inf .	N.inf .	1.00 (-)	0.61 (0.00)	0.48 (0.00)
GTP	-	-	-	-	-	-	-	-	-
YLL	0.4286 (-)	1.00 (-)	0.6571 (-)	N.inf .	1.00 (-)	N.inf .	N.inf .	*0.05 (0.00)	0.95 (0.00)

Notes: *deviations from equilibrium

Appendix 4 – Diversity estimates at individual loci

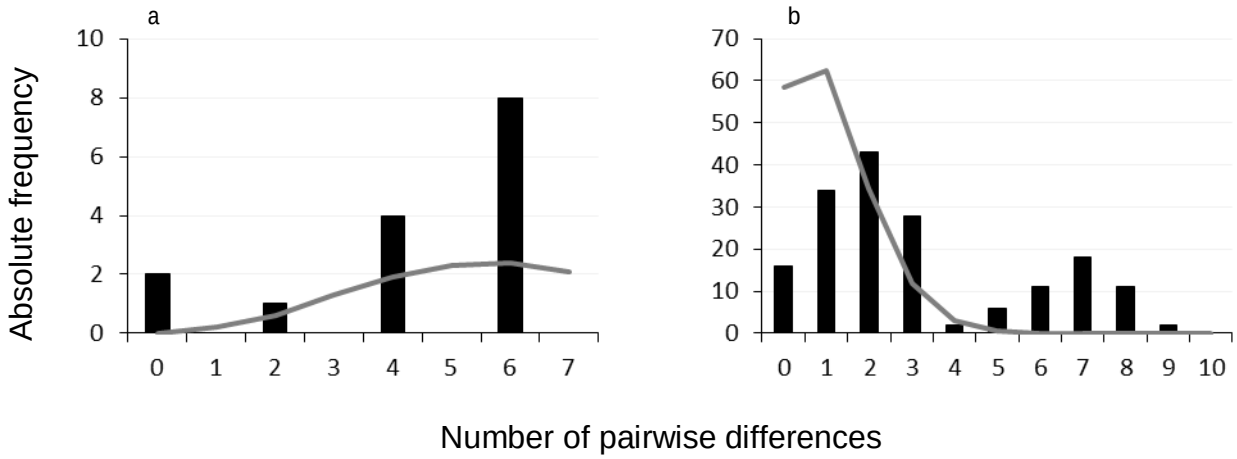
	h		SPTBN	PRKCI	DARC	KITLG	TF	POLA1	GRIA3	SRY
Lam	Boreal	SK1	0.200	0.639	0.844	0.000	0.891	0.800	1.000	0.500
	Rockies	OR2	0.000	0.725	0.533	0.400	0.000	0.400	0.464	0.533
		WY1	0.143	0.167	0.000	0.275	0.000	0.000	0.464	1.000
		WA4	0.000	0.742	0.479	0.647	0.368	0.429	0.000	0.400
	PacNW	WA1	0.000	0.733	0.167	0.648	0.458	0.533	0.727	0.000
		CA1	0.509	0.556	0.621	0.622	0.000	0.389	0.000	0.000
Lcf		OR	0.637	0.616	0.596	0.821	0.437	0.257	0.485	0.000
		CA	0.848	0.867	0.822	0.909	0.848	0.733	0.571	0.667
		TX	0.829	0.945	0.890	0.927	0.731	0.857	0.750	0.000
		AZ	0.758	0.864	0.561	0.822	0.818	0.800	0.833	0.000
Ltw		MT	0.500	1.000	0.667	0.000	0.500	0.500	0.833	-
		YLL	0.500	0.750	0.722	0.000	0.356	0.700	0.800	0.000
		GTP	1.000	1.000	0.000	0.000	0.000	1.000	1.000	-
		WY	0.433	0.767	0.714	0.314	0.242	0.745	0.800	0.700
Lam			0.495	0.778	0.730	0.754	0.594	0.754	0.649	0.637
Lcf			0.782	0.868	0.743	0.847	0.690	0.593	0.746	0.731
Ltw			0.469	0.799	0.719	0.171	0.288	0.701	0.823	0.378
Lam	Boreal		0.200	0.639	0.844	0.000	0.891	0.800	1.000	0.500
	Rockies		0.508	0.577	0.655	0.655	0.515	0.538	0.450	0.822
	PacNW		0.159	0.753	0.527	0.689	0.324	0.519	0.377	0.167

	π %		SPTBN	PRKCI	DARC	KITLG	TF	POLA1	GRIA3	SRY
Lam	Boreal	SK1	0.11 (0.11)	0.18 (0.16)	0.25 (0.18)	0.00 (0.00)	0.84 (0.55)	0.65 (0.44)	0.21 (0.20)	0.06 (0.06)
	Rockies	OR2	0.00 (0.00)	0.91 (0.54)	0.14 (0.11)	0.09 (0.10)	0.00 (0.00)	0.21 (0.19)	0.05 (0.06)	0.07 (0.06)
		WY1	0.05 (0.07)	0.04 (0.06)	0.00 (0.00)	0.16 (0.14)	0.00 (0.00)	0.00 (0.00)	0.05 (0.06)	0.08 (0.09)
		WA4	0.00 (0.00)	3.09 (1.64)	0.13 (0.10)	0.23 (0.18)	0.12 (0.13)	0.15 (0.13)	0.00 (0.00)	0.05 (0.05)
	PacNW	WA1	0.00 (0.00)	0.39 (0.31)	0.05 (0.05)	0.32 (0.23)	0.15 (0.15)	0.19 (0.15)	0.14 (0.11)	0.00 (0.00)
		CA1	0.19 (0.15)	1.51 (0.90)	0.15 (0.12)	0.27 (0.21)	0.00 (0.00)	0.14 (0.12)	0.00 (0.00)	0.00 (0.00)
Lcf		OR	0.19 (0.15)	1.06 (0.61)	0.10 (0.09)	0.61 (0.37)	0.74 (0.47)	0.05 (0.06)	0.05 (0.05)	0.00 (0.00)
		CA	0.41 (0.27)	2.82 (1.58)	0.20 (0.15)	0.72 (0.45)	1.98 (1.14)	0.16 (0.15)	0.06 (0.06)	0.04 (0.05)
		TX	0.35 (0.24)	3.02 (1.62)	0.69 (0.40)	1.08 (0.64)	1.79 (1.03)	0.27 (0.20)	0.25 (0.17)	0.00 (0.00)
		AZ	0.30 (0.21)	1.64 (0.93)	0.09 (0.08)	0.73 (0.46)	1.95 (1.12)	0.27 (0.21)	1.26 (0.86)	0.00 (0.00)
Ltw		MT	1.84 (1.28)	4.88 (5.00)	0.18 (0.18)	0.00 (0.00)	0.16 (0.20)	0.09 (0.11)	0.19 (0.16)	-
		YLL	1.84 (1.05)	2.49 (1.45)	0.24 (0.17)	0.00 (0.00)	0.11 (0.13)	0.21 (0.19)	0.25 (0.19)	0.00 (0.00)
		GTP	0.38 (0.47)	4.42 (4.53)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.70 (0.78)	0.21 (0.25)	-
		WY	1.51 (0.83)	2.45 (1.32)	0.18 (0.13)	0.22 (0.17)	0.08 (0.11)	0.25 (0.19)	0.12 (0.10)	0.42 (0.28)
Lam			0.29 (0.19)	2.21 (1.14)	0.36 (0.21)	0.36 (0.24)	0.37 (0.27)	0.90 (0.49)	0.14 (0.10)	0.63 (0.33)
Lcf			0.32 (0.21)	2.28 (1.17)	0.29 (0.18)	0.75 (0.43)	1.68 (0.91)	0.17 (0.13)	0.76 (0.41)	0.65 (0.35)
Ltw			1.50 (0.80)	2.42 (1.26)	0.21 (0.14)	0.11 (0.11)	0.10 (0.11)	0.24 (0.17)	0.19 (0.13)	0.21 (0.13)
Lam	Boreal		0.11 (0.11)	0.18 (0.16)	0.25 (0.18)	0.00 (0.00)	0.84 (0.55)	0.65 (0.44)	0.21 (0.20)	0.06 (0.06)
	Rockies		0.19 (0.14)	0.57 (0.36)	0.25 (0.16)	0.27 (0.19)	0.16 (0.16)	0.70 (0.42)	0.05 (0.05)	1.20 (0.66)
	PacNW		0.06 (0.07)	3.18 (1.64)	0.14 (0.10)	0.29 (0.20)	0.11 (0.12)	0.18 (0.14)	0.07 (0.06)	0.02 (0.03)

	θ_s %		SPTBN	PRKCI	DARC	KITLG	TF	POLA1	GRIA3	SRY
Lam	Boreal	SK1	1.06 (0.70)	0.74 (0.56)	2.12 (1.16)	0.00 (0.00)	3.07 (1.53)	3.50 (1.97)	2.00 (1.51)	1.09 (0.88)
	Rockies	OR2	0.00 (0.00)	2.71 (1.28)	0.62 (0.46)	0.30 (0.30)	0.00 (0.00)	1.44 (1.02)	0.77 (0.59)	0.88 (0.68)
		WY1	0.63 (0.47)	0.33 (0.33)	0.00 (0.00)	1.57 (0.87)	0.00 (0.00)	0.00 (0.00)	0.77 (0.59)	1.33 (1.10)
		WA4	0.00 (0.00)	3.32 (1.50)	0.56 (0.42)	0.85 (0.54)	0.57 (0.43)	0.77 (0.59)	0.00 (0.00)	0.48 (0.48)
	PacNW	WA1	0.00 (0.00)	1.75 (1.13)	0.66 (0.50)	1.21 (0.71)	0.30 (0.30)	0.71 (0.54)	1.02 (0.68)	0.00 (0.00)
		CA1	0.68 (0.52)	4.05 (2.01)	0.66 (0.50)	1.06 (0.70)	0.00 (0.00)	0.74 (0.56)	0.00 (0.00)	0.00 (0.00)
Lcf		OR	1.43 (0.77)	2.25 (1.07)	0.89 (0.57)	2.54 (1.17)	3.10 (1.36)	0.62 (0.46)	0.30 (0.30)	0.00 (0.00)
		CA	1.99 (1.07)	2.47 (1.31)	1.77 (1.01)	3.41 (1.66)	4.30 (1.98)	0.88 (0.68)	0.41 (0.41)	0.67 (0.67)
		TX	1.54 (0.85)	4.09 (1.83)	2.52 (1.24)	5.12 (2.34)	3.87 (1.78)	1.93 (1.14)	2.21 (1.23)	0.00 (0.00)
		AZ	2.32 (1.20)	2.65 (1.33)	0.99 (0.65)	3.53 (1.75)	4.30 (1.98)	1.75 (1.13)	4.36 (2.68)	0.00 (0.00)
Ltw		MT	1.09 (0.88)	4.00 (3.16)	0.67 (0.67)	0.00 (0.00)	0.55 (0.55)	0.55 (0.55)	1.64 (1.19)	-
		YLL	0.74 (0.56)	1.16 (0.78)	1.47 (0.91)	0.00 (0.00)	0.35 (0.35)	0.96 (0.76)	1.92 (1.27)	0.00 (0.00)
		GTP	1.00 (1.00)	2.00 (1.73)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	3.00 (2.45)	2.00 (1.73)	-
		WY	0.90 (0.58)	1.51 (0.83)	0.92 (0.60)	1.16 (0.68)	0.60 (0.45)	1.37 (0.82)	1.71 (0.97)	2.40 (1.51)
Lam			1.99 (0.78)	2.92 (1.07)	2.40 (0.90)	1.82 (0.74)	2.57 (0.94)	4.53 (1.59)	2.44 (0.97)	3.44 (1.42)
Lcf			2.38 (0.94)	4.14 (1.44)	2.64 (1.03)	4.20 (1.47)	4.34 (1.49)	2.19 (0.95)	4.07 (1.52)	6.29 (2.48)
Ltw			1.00 (0.57)	1.80 (0.85)	1.27 (0.67)	0.98 (0.55)	0.74 (0.47)	1.37 (0.73)	2.74 (1.22)	1.77 (1.01)
Lam	Boreal		1.06 (0.70)	0.74 (0.56)	2.12 (1.16)	0.00 (0.00)	3.07 (1.53)	3.50 (1.97)	2.00 (1.51)	1.09 (0.88)
	Rockies		0.50 (0.37)	2.31 (1.03)	0.76 (0.48)	1.26 (0.66)	0.25 (0.25)	2.83 (1.36)	1.21 (0.71)	1.77 (1.01)
	PacNW		0.45 (0.33)	2.75 (1.15)	0.46 (0.34)	0.91 (0.51)	0.45 (0.33)	0.52 (0.38)	0.74 (0.47)	0.33 (0.33)

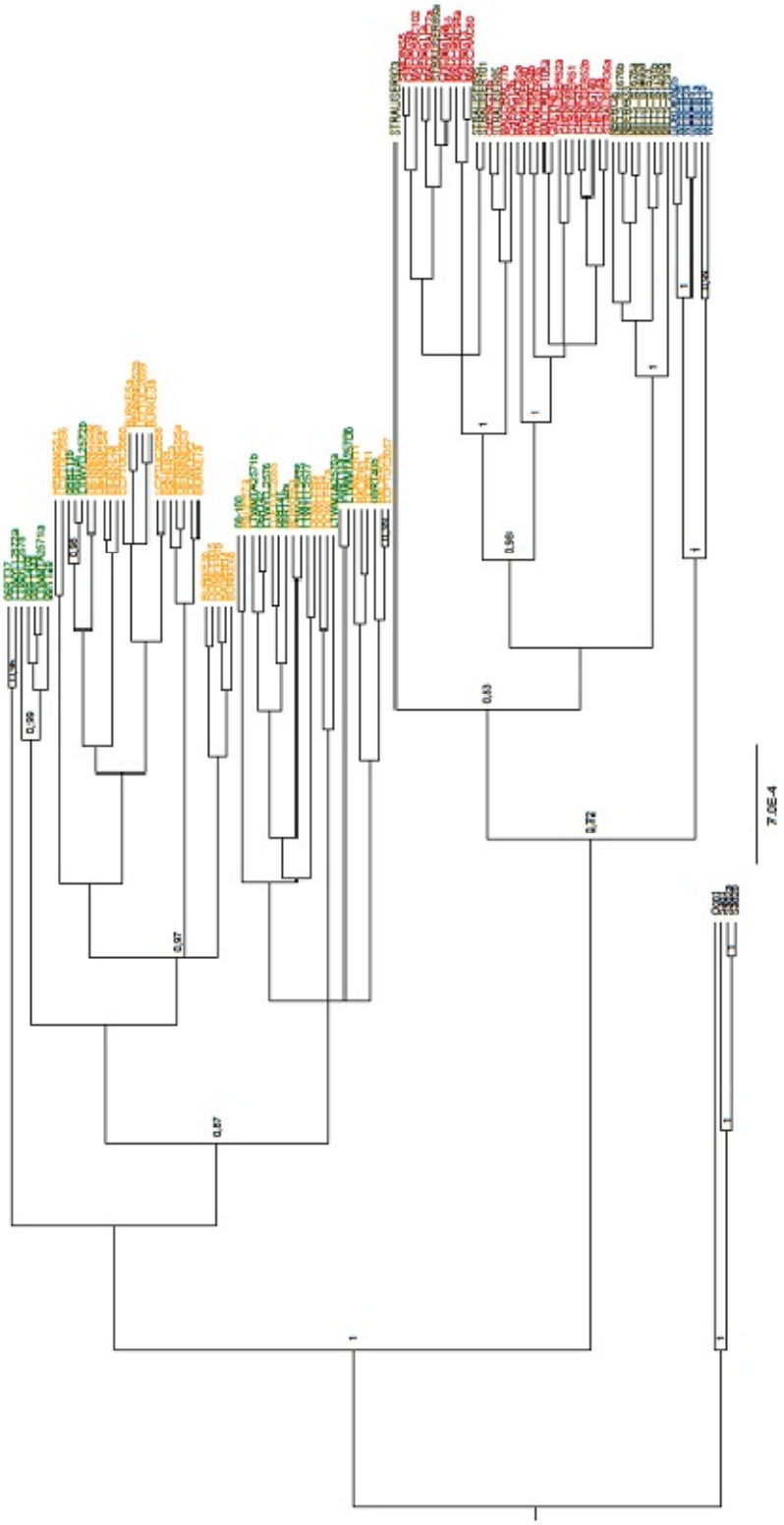
Notes - *h*, haplotype diversity; π, nucleotide diversity and θs, computed from the number of segregating sites Tajima, 1983. Lam - *L. americanus*; Lcf – *L. californicus*; Ltw – *L. townsendii.* Standard deviations (SD) are shown in brackets

Appendix 5 - Observed (bars) and expected (solid lines) mismatch distributions of a) *L. californicus* California (CA) population; b) *L. townsendii Wyoming* (WY) population.

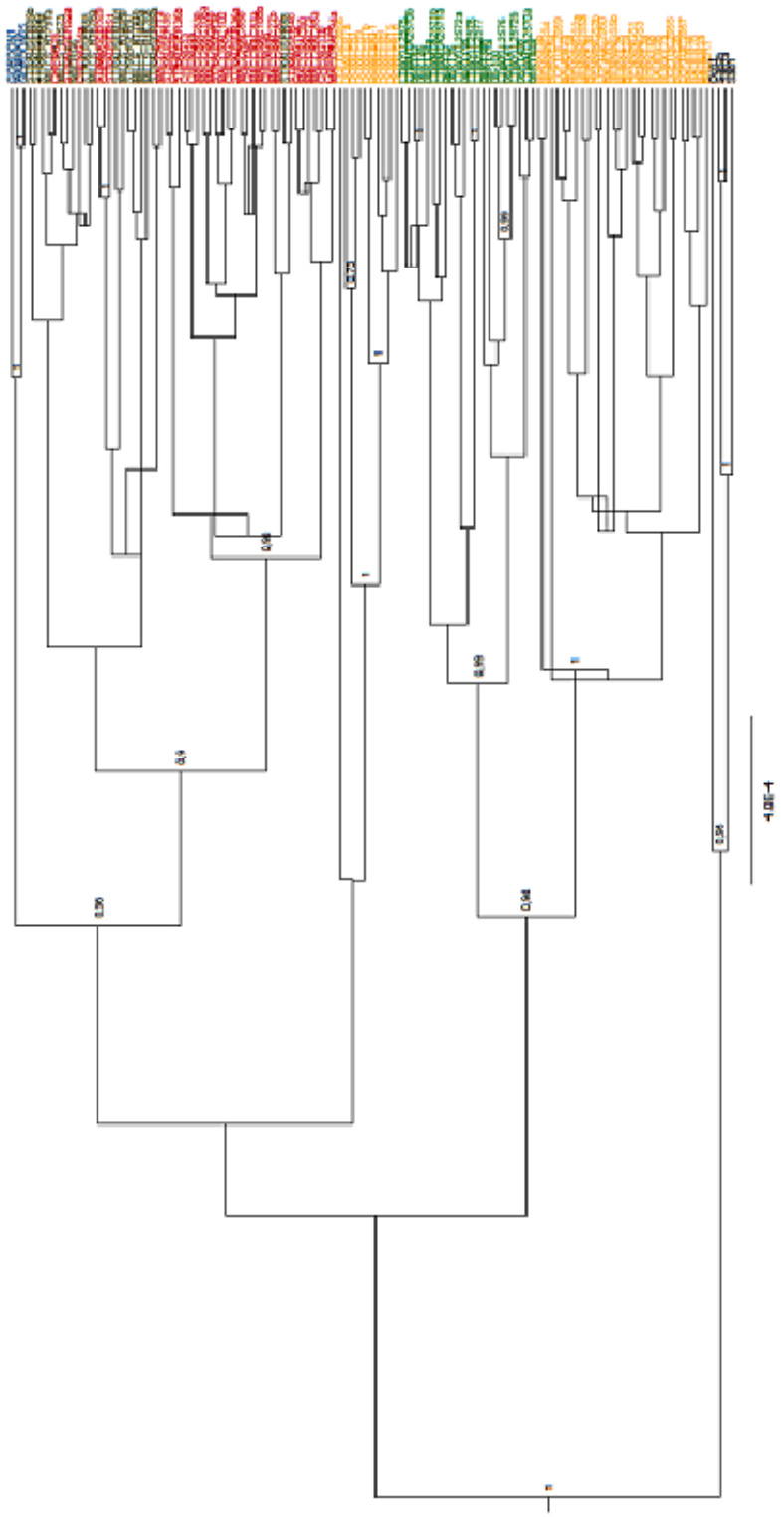


Appendix 6 - Individual nuclear loci trees generated from the outputs of BEAST (Drummond and Rambaut, 2007) (numbers close to nodes indicate the posterior probabilities if higher than 0.50). The following colour code was used to identify species and cluster to which individuals belong: *L. americanus* PacNW cluster (red), *L. americanus* Boreal cluster (blue), *L. americanus* Rockies cluster (brown), *L. californicus* (yellow), *L. townsendii* (green).

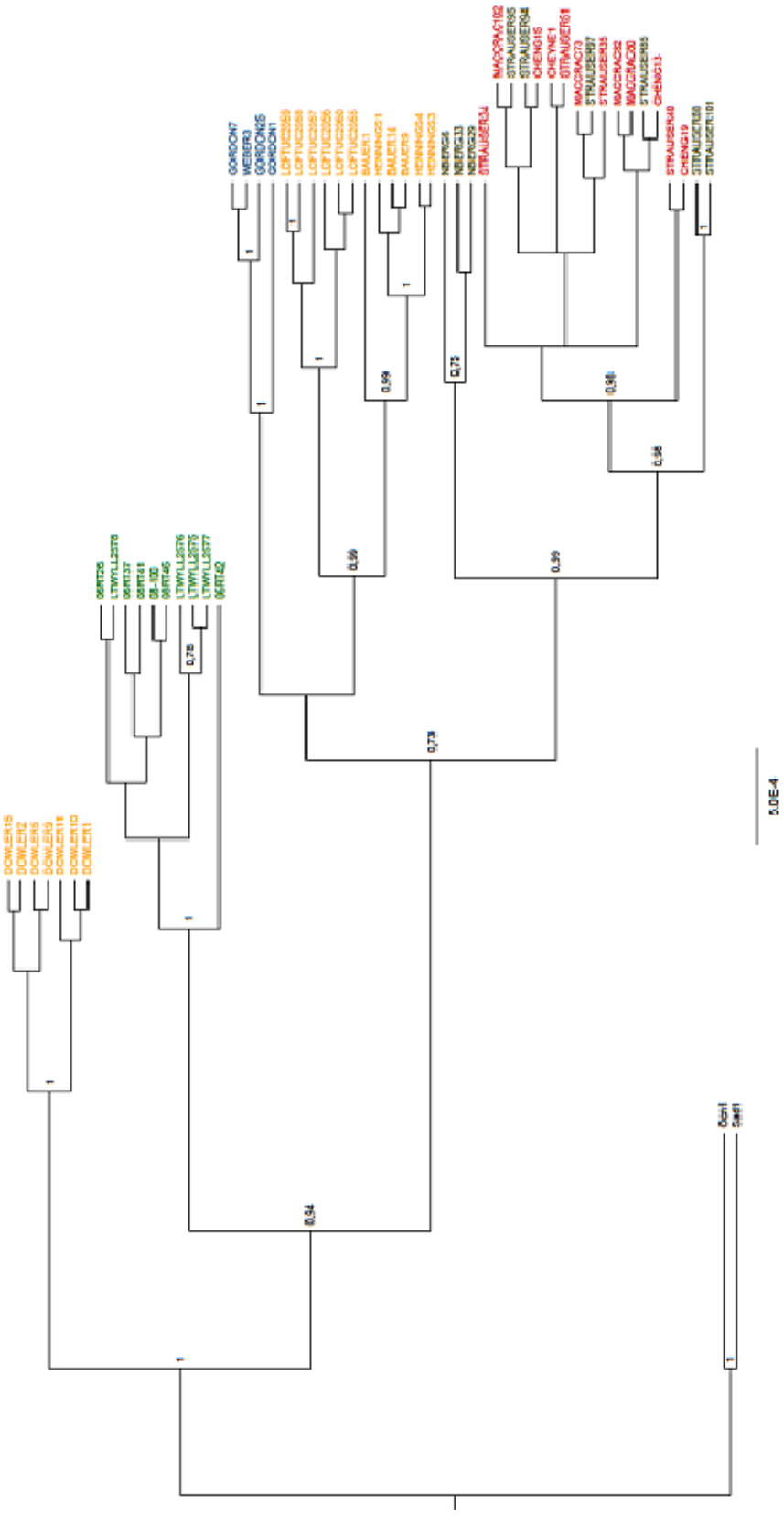
POLA1



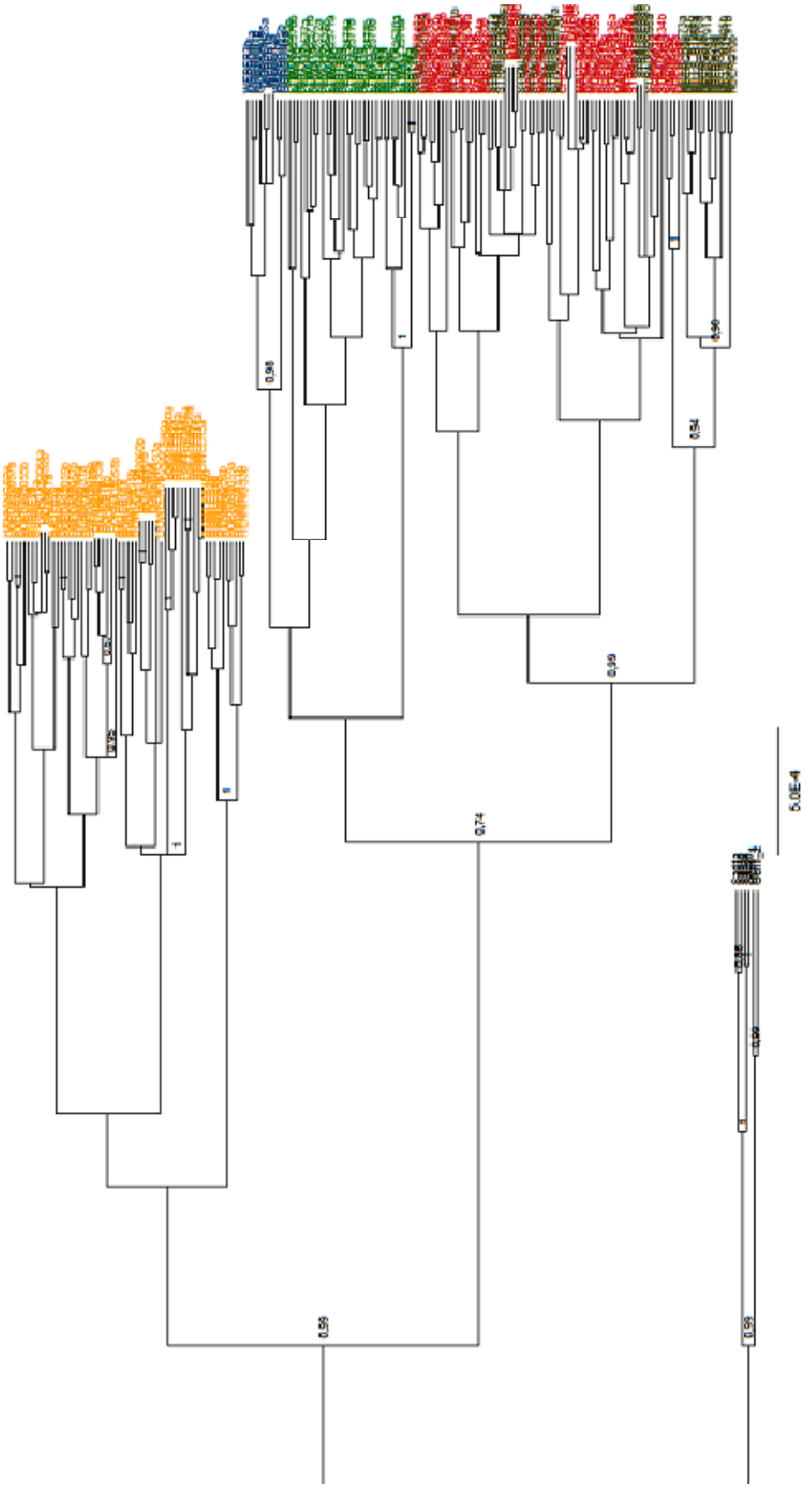
GRIA3



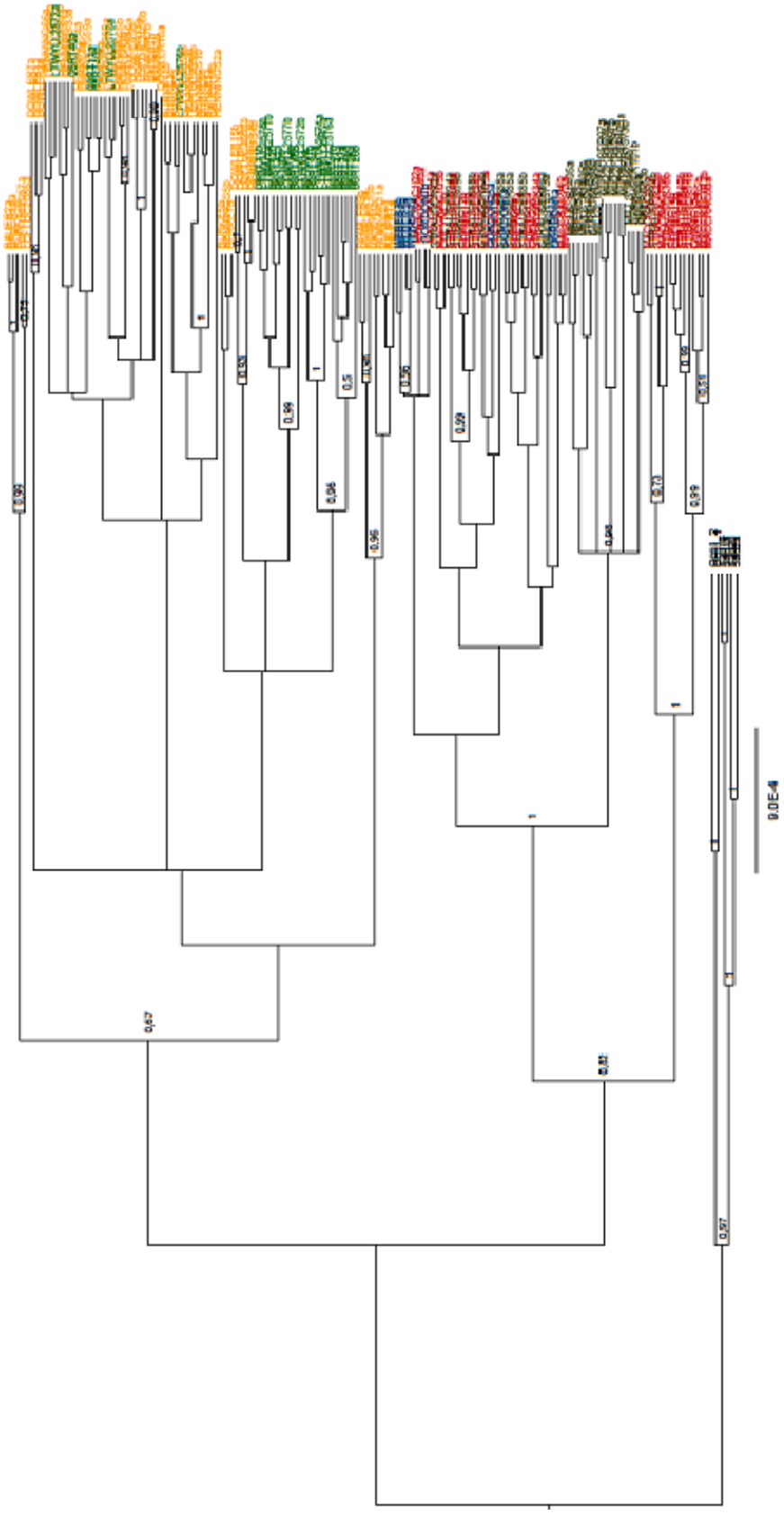
SRY


MAGGRAC102
STRAUSER95
STRAUSER94
CHENG15
CHEYNE1
STRAUSER91
MAGGRAC73
STRAUSER97
STRAUSER35
MAGGRAC82
MAGGRAC80
STRAUSER85
CHENG13
STRAUSER40
CHENG19
STRAUSER30
STRAUSER101
GORDON7
WEBER3
GORDON25
GORDON1
LCFTUC2559
LCFTUC2558
LCFTUC2557
LCFTUC2556
LCFTUC2560
LCFTUC2555
BAUER1
HENNINGS1
BAUER14
BAUER5
HENNINGS4
HENNINGS3
NBERG24
NBERG33
NBERG29
STRAUSER84
GSRT25
LTWYLLS376
GSRT37
GSRT48
GS-100
GSRT45
LTWYLLS375
LTWYLLS373
LTWYLLS377
GSRT42
DOWLER15
DOWLER2
DOWLER5
DOWLER9
DOWLER14
DOWLER10
DOWLER11
Dcin1
Sasi1
0.99
0.99
0.98
0.96
0.79
0.99
0.73
0.78
0.54
5.0E-4

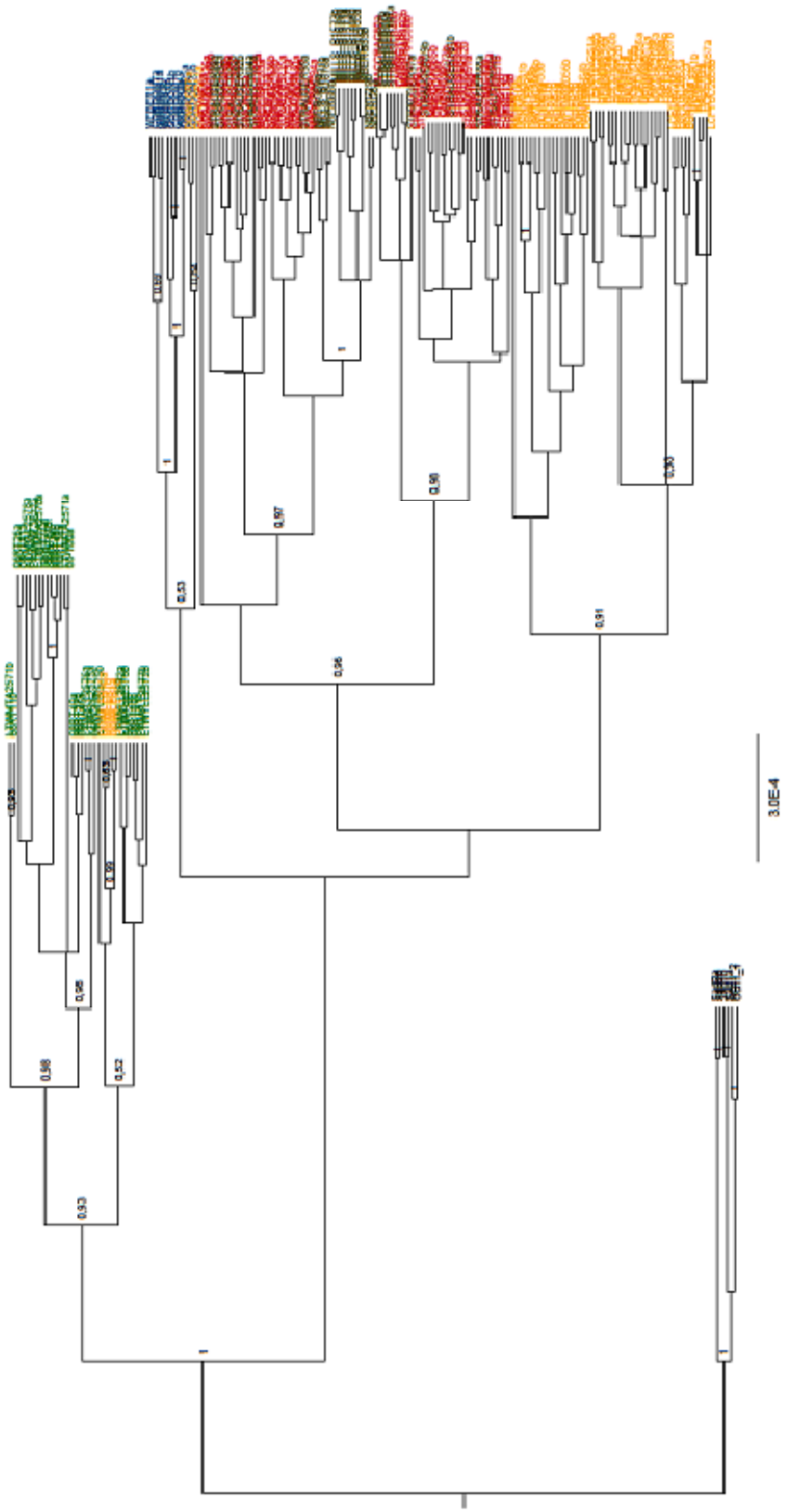
SPTBN1



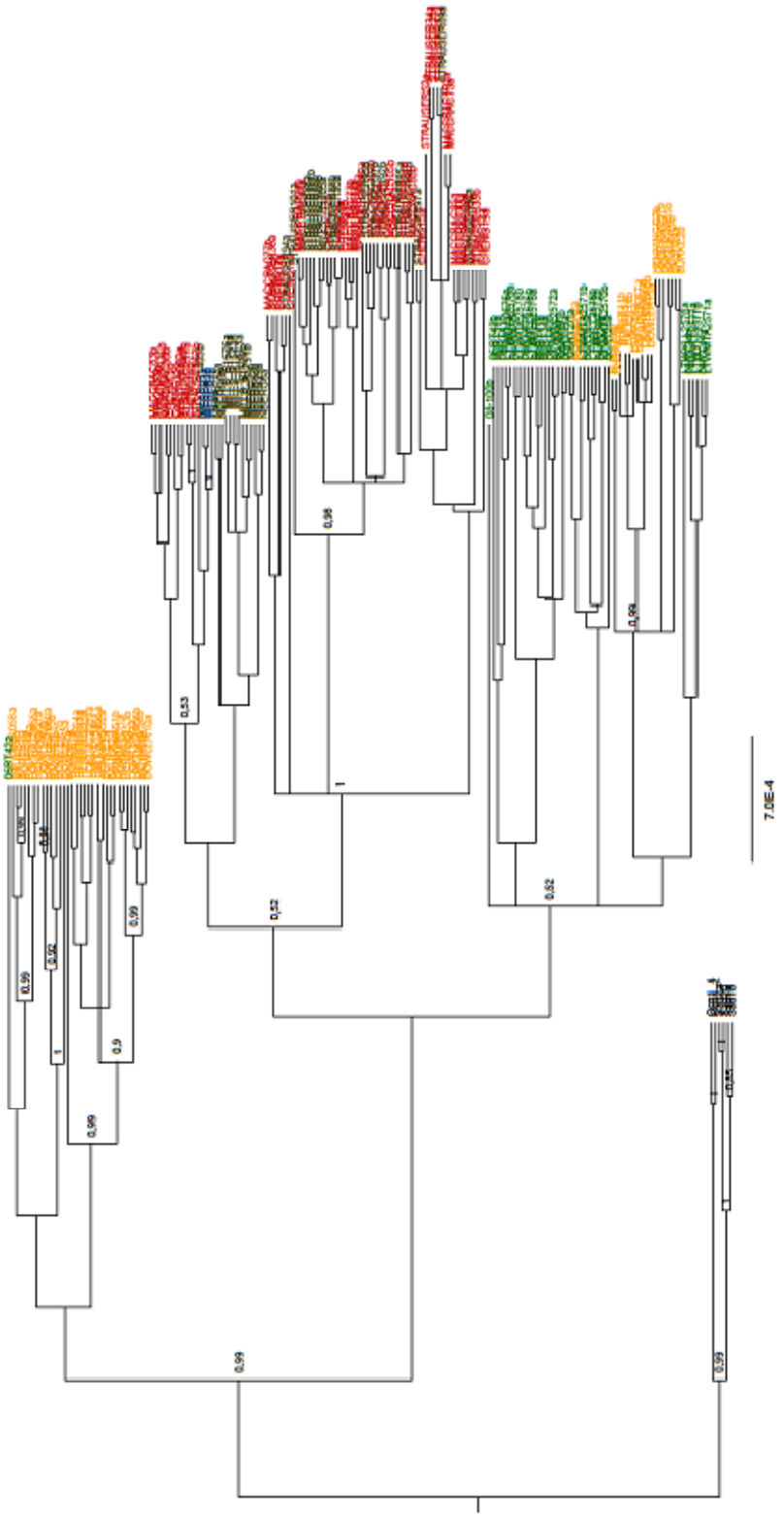
PRKCI



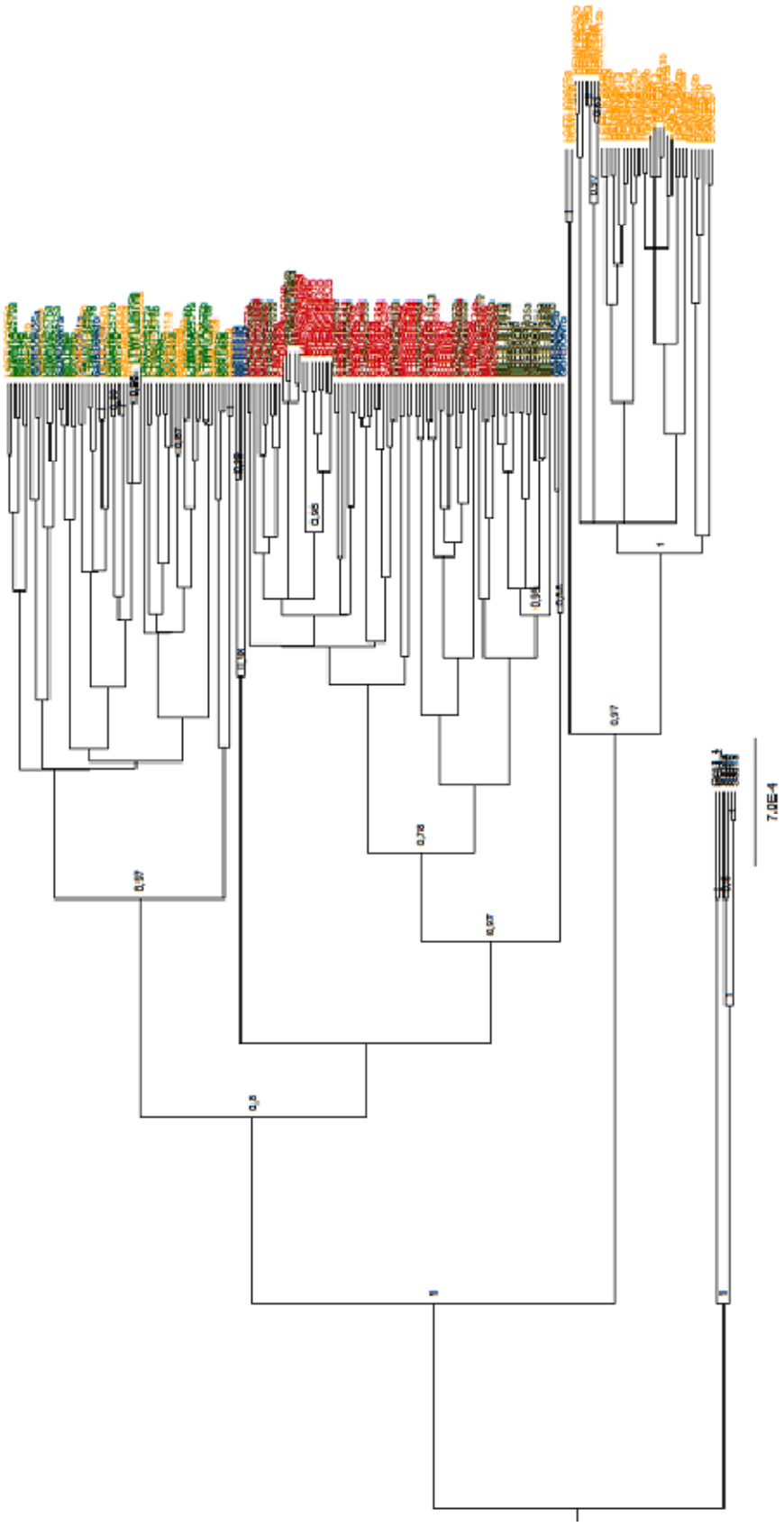
DARC



KITLG

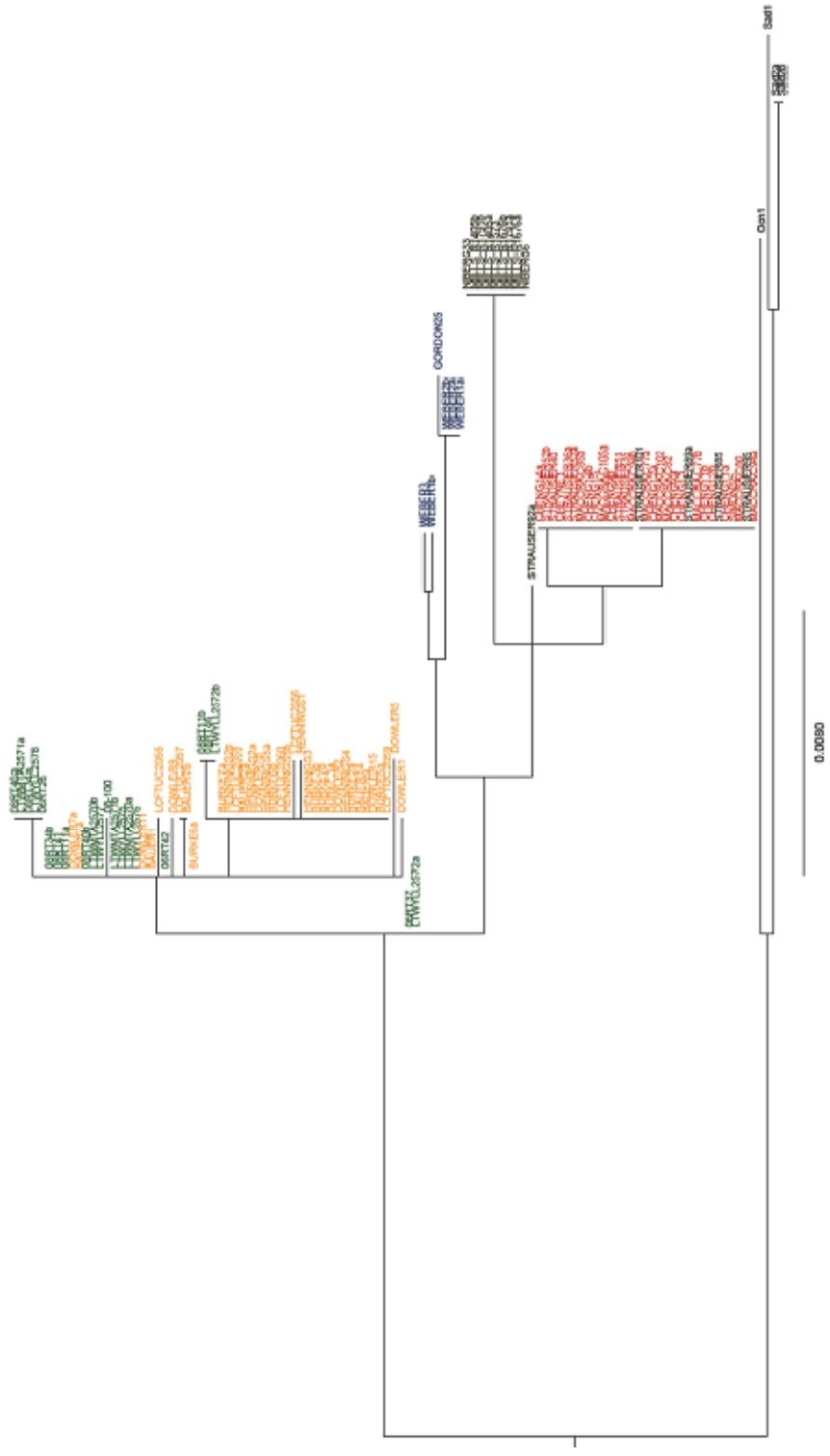


TF

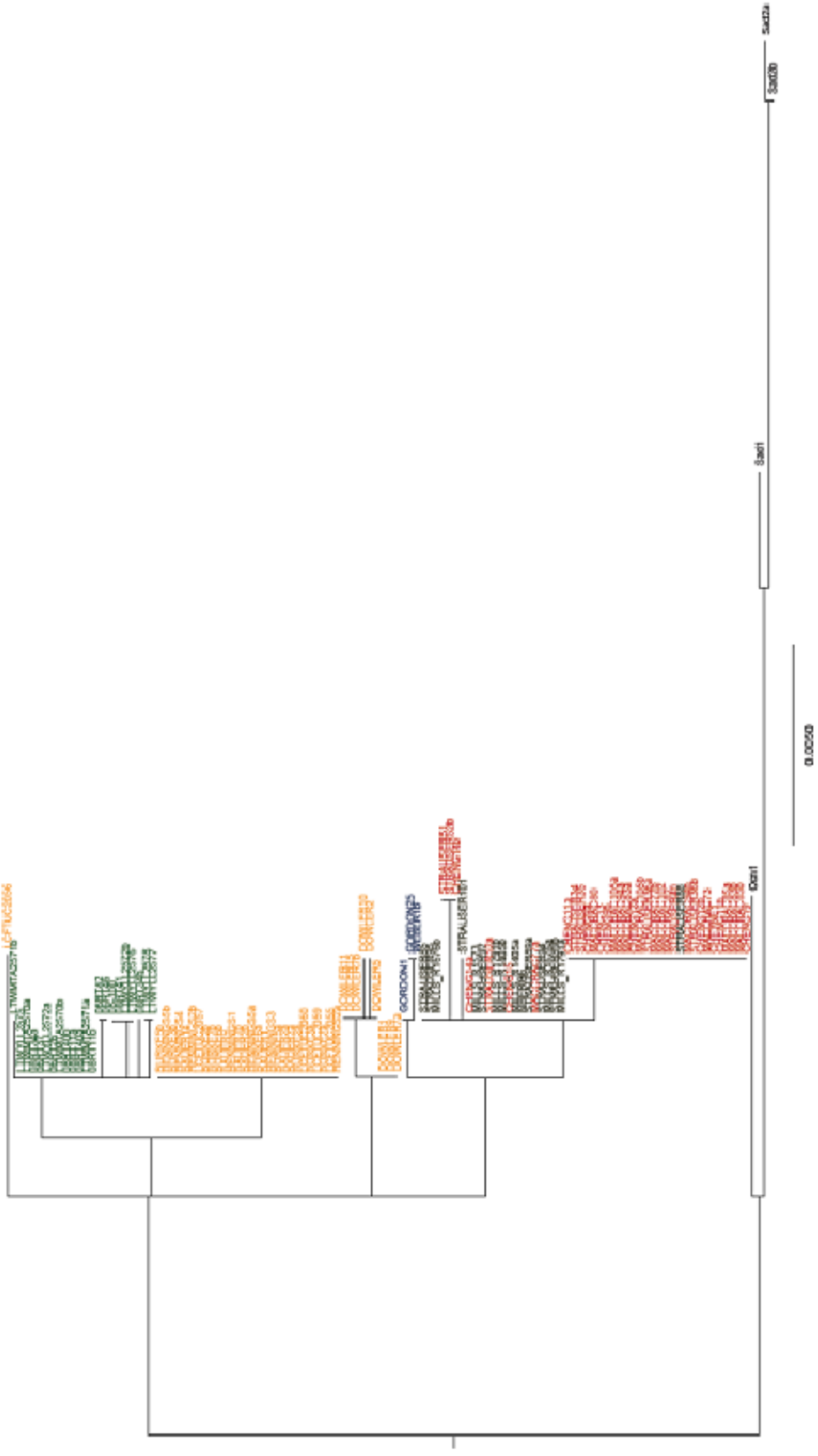


Appendix 7 - Individual nuclear loci best Maximum Likelihood trees inferred from Garli v1.0 (Zwickl, 2006). The following colour code was used to identify species and cluster to which individuals belong: *L. americanus* PacNW cluster (red), *L. americanus* Boreal cluster (blue), *L. americanus* Rockies cluster (brown), *L. californicus* (yellow), *L. townsendii* (green).

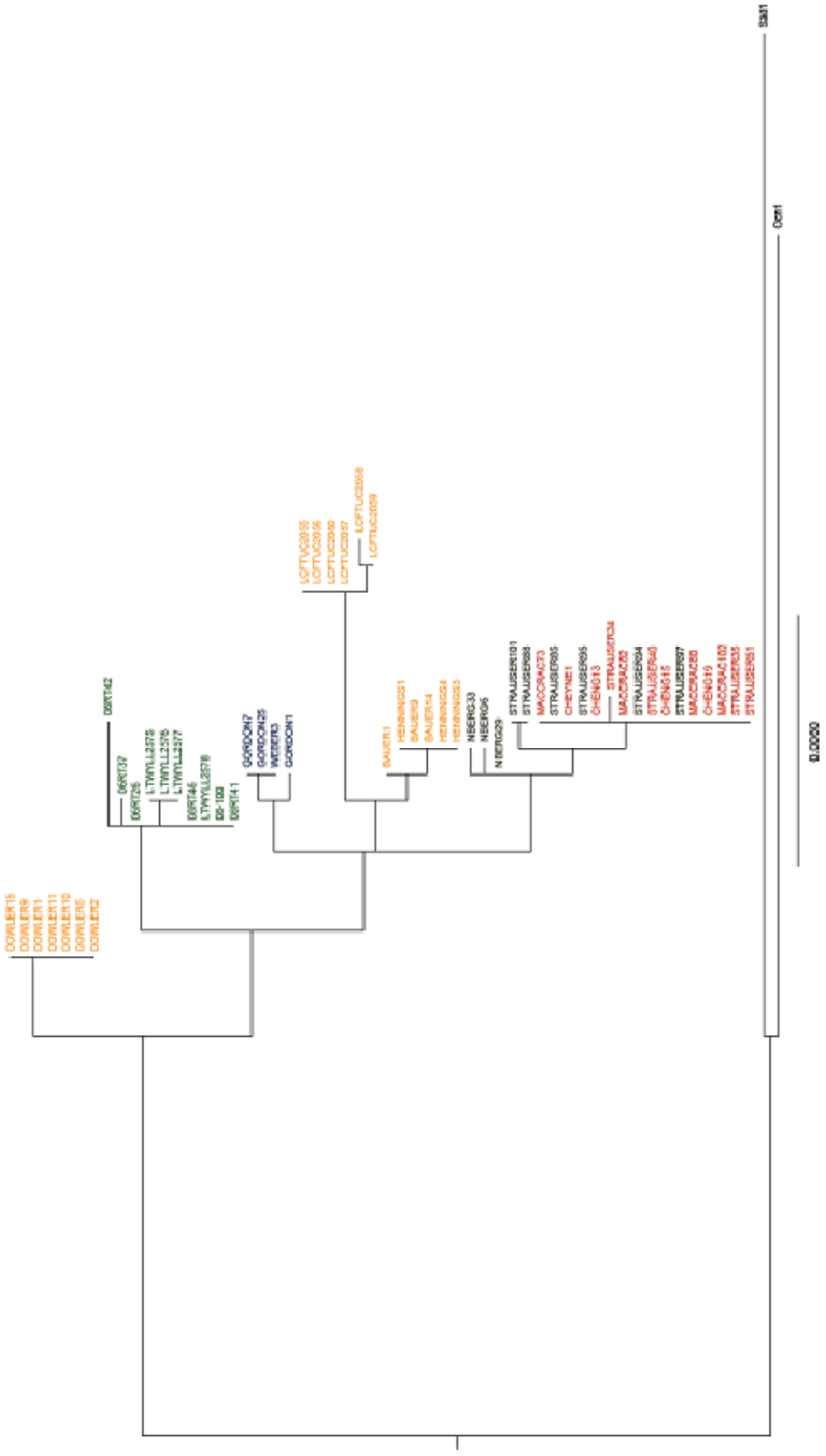
POLA1



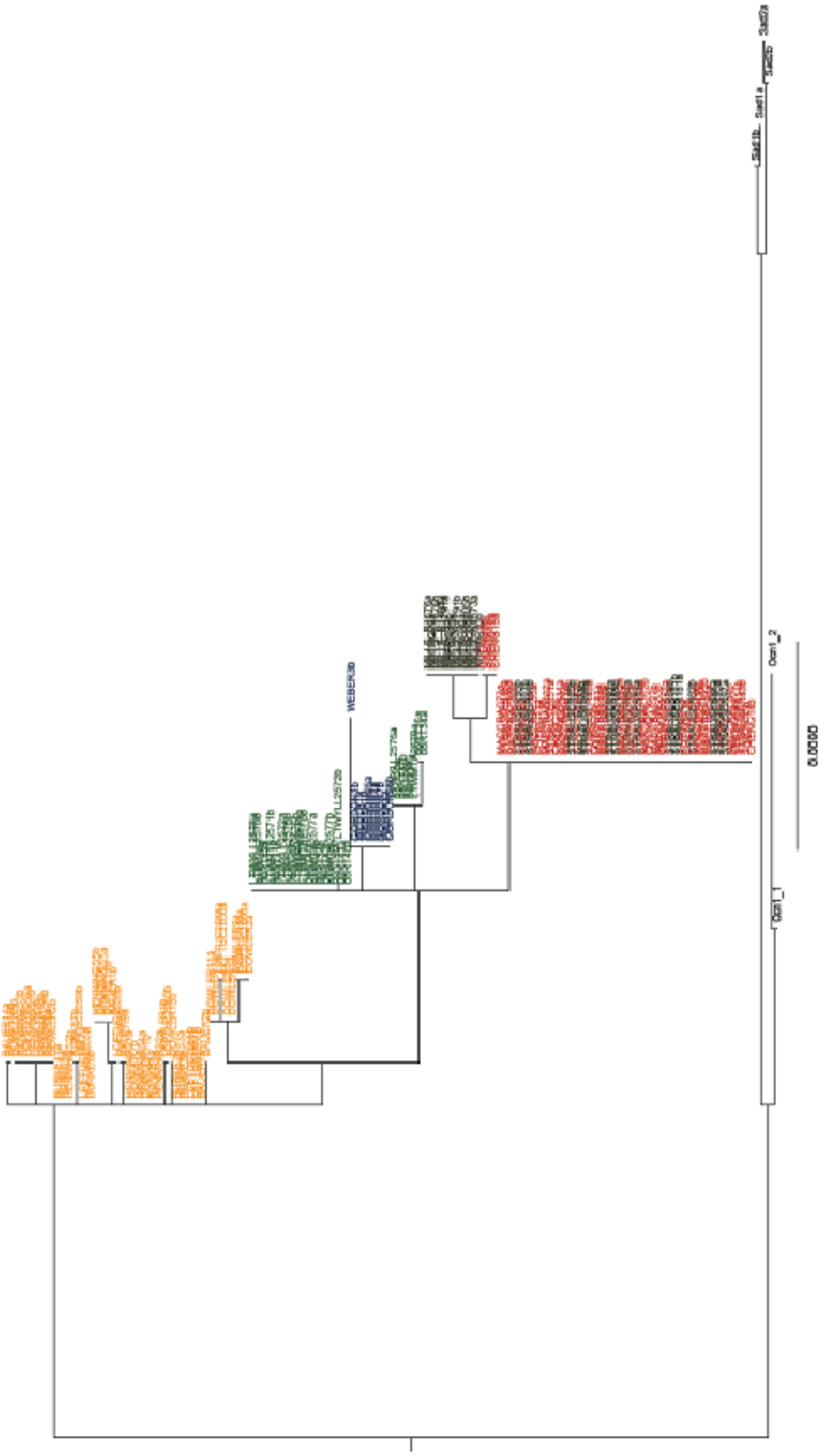
GRIA3



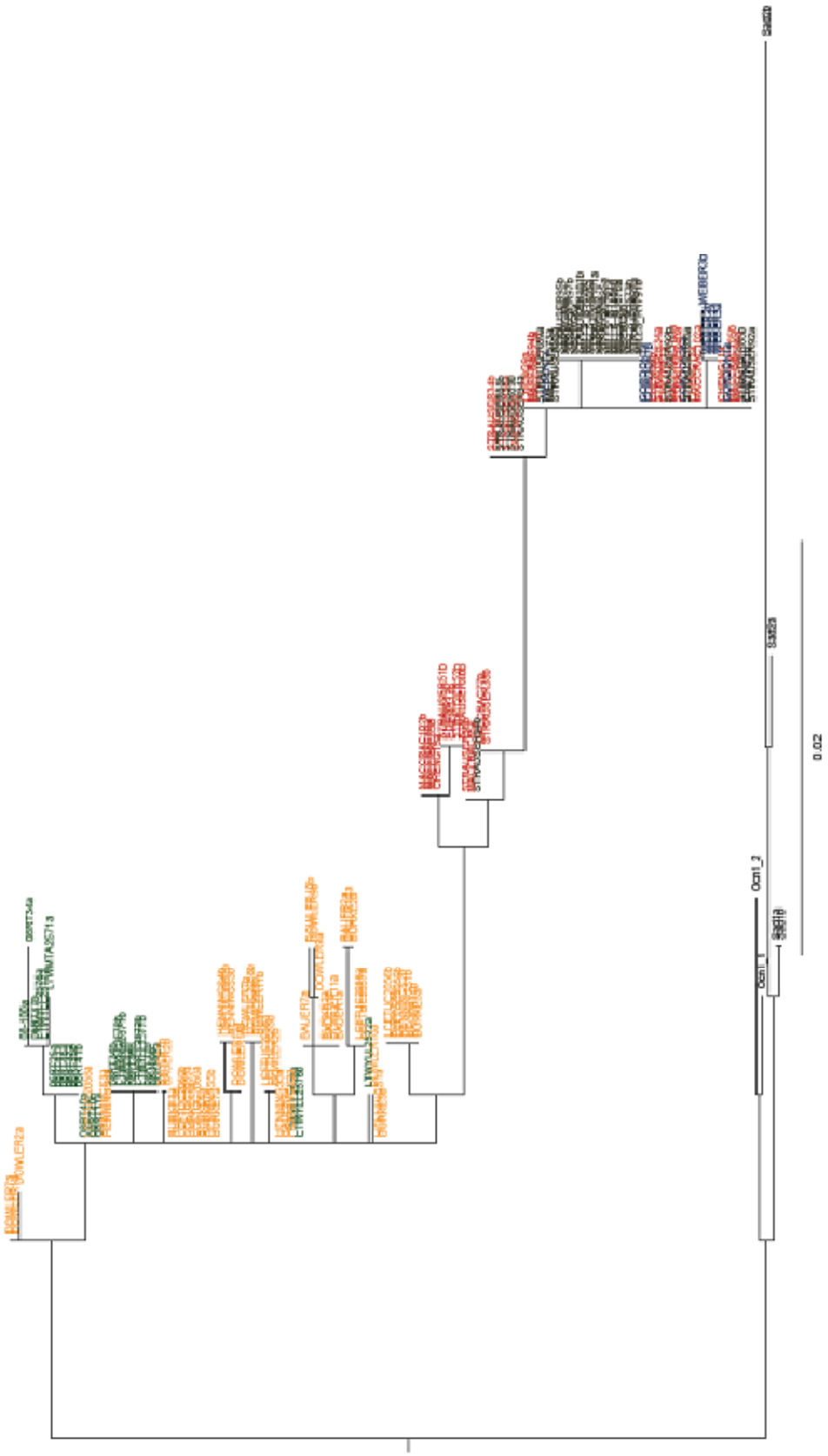
SRY



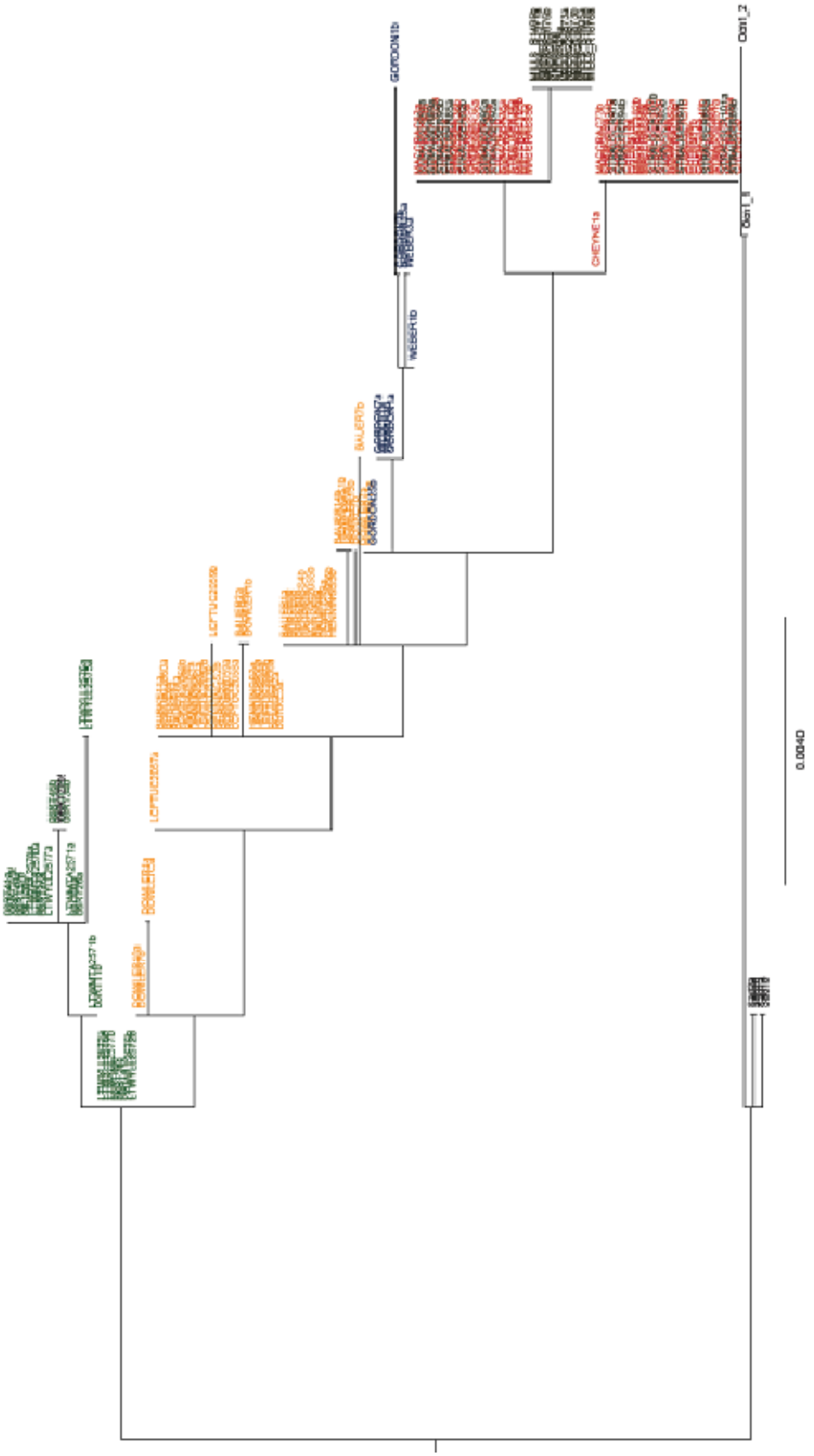
SPTBN1



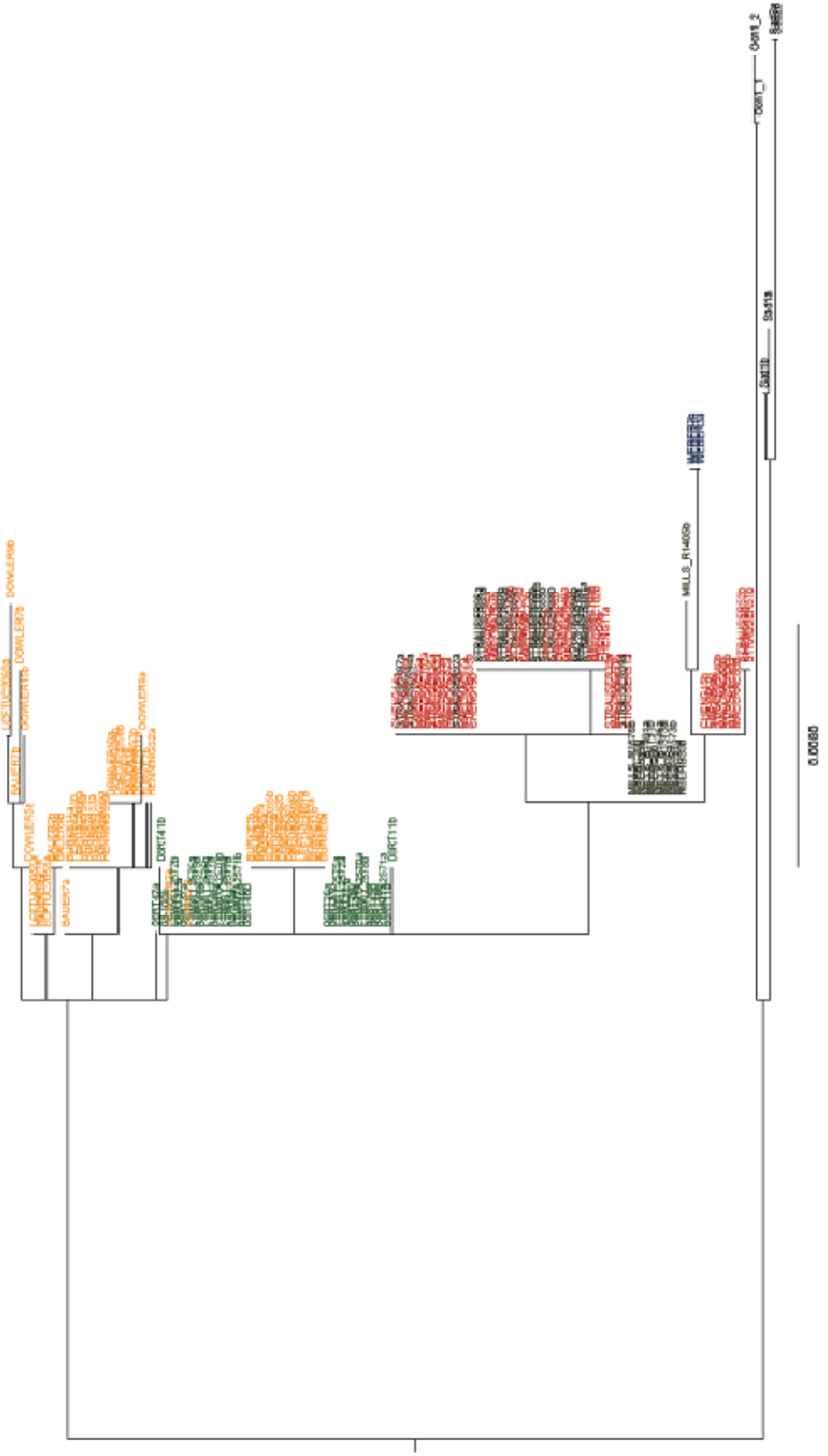
PRKCI



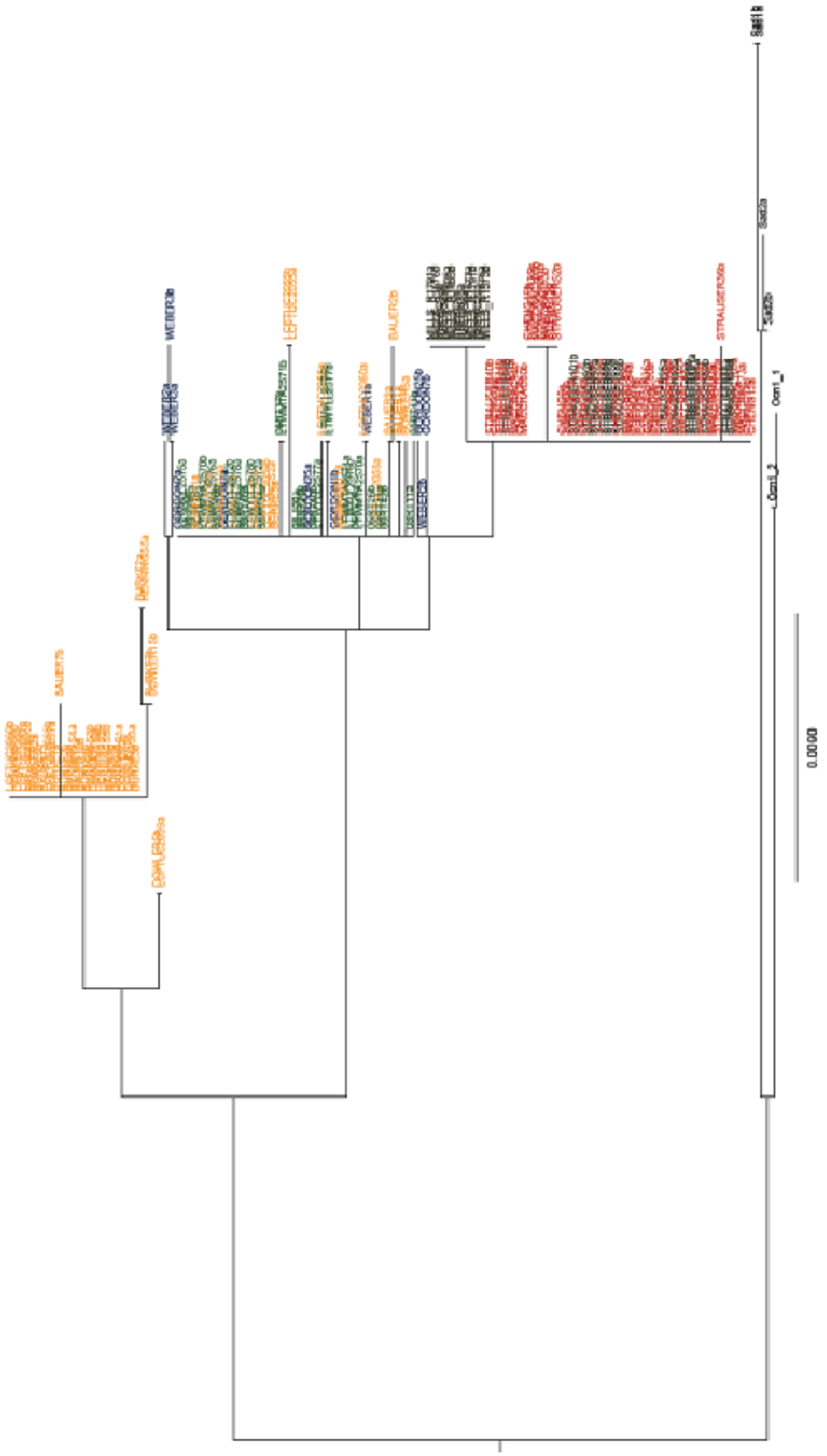
DARC



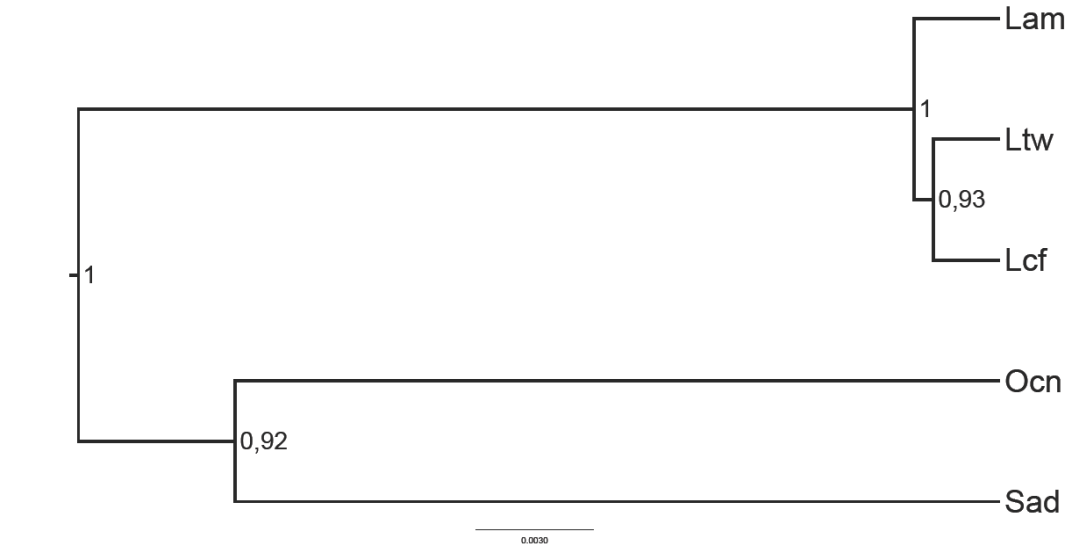
KITLG

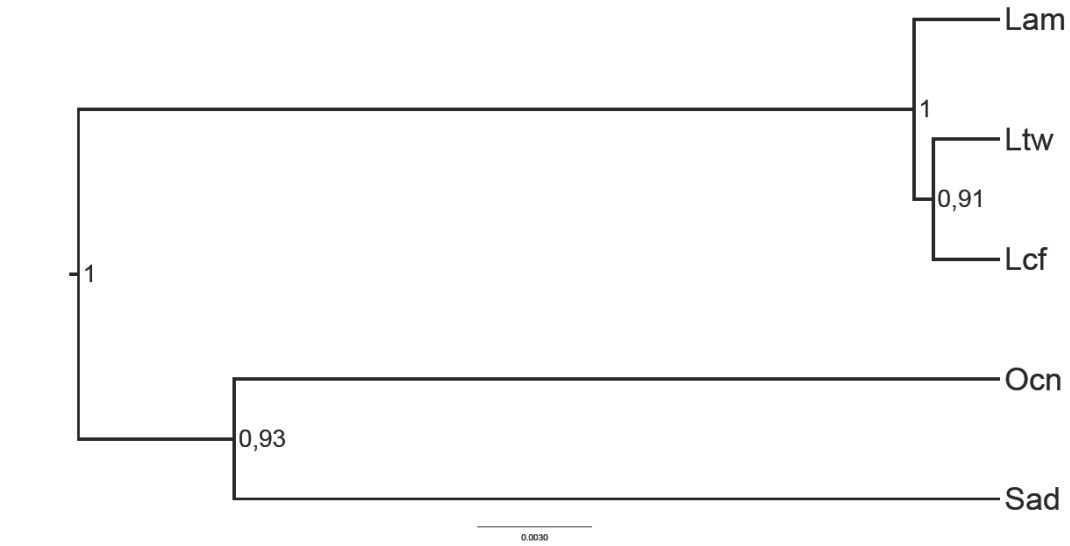


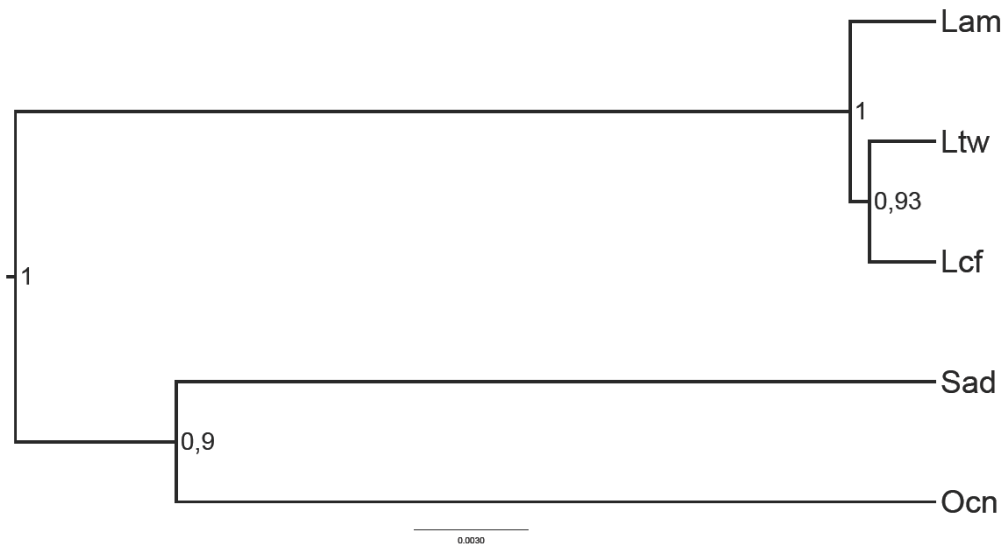
TF



Appendix 8 – Multilocus nuclear species tree inferred from *BEAST (Heled and Drumond, 2010) (numbers close to nodes indicate the posterior probabilities).







Appendix 9 – Bayesian taxa delimitation results using the eight nuclear loci. Posterior probabilities of the different possible delimitation models are only given for those models for which, at least in one of the different combinations of θ and τ priors, the posterior is greater than 0.The guide tree is based on the Bayesian species tree inference based using the eight loci in the *BEAST analysis

θ prior[a]	τ prior[a]	start[b]	Model[c]	
			1101	1111
(2.0, 2000)	(2.0, 2000)	1110	0	1
(2.0, 2000)	(1.0, 10)	1101	0	1
(0.02, 20)	(0.02, 20)	1101	0	1
(0.02, 20)	(0.1, 1.0)	0000	0	1
(1.0, 10)	(2.0, 2000)	1101	0	1
(1.0, 10)	(1.0, 10)	1001	0.0003	0.9997
(0.1, 1.0)	(0.02, 20)	1101	0.0005	0.9995
(0.1, 1.0)	(0.1, 1.0)	1101	0.0002	0.9998

Notes:

[a]Gamma distribution of the priors, considering small ancestral effective population sizes or shallow divergence (2.0, 2000), (0.02, 20), and large effective population sizes or deep divergence (1.0, 10), (0.1, 1.0) (see e.g. Leaché and and Fujita, 2010);

[b]Starting taxa delimitation model of the analysis;

[c]Taxa delimitation models: 1101 – *L. californicus*:*L. townsendii*:Boreal(Rockies,PacNW); 1111 – *L. californicus*:*L. townsendii*:Boreal:Rockies:PacNW. (Colon indicates the delimitation of taxa while parenthesis indicate taxa which are collapsed in the model)