Molecular Phylogenetics and Evolution 82 (2015) 245–257

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Mitogenomic phylogenetics of the bank vole *Clethrionomys glareolus*, a model system for studying end-glacial colonization of Europe

Karolína Filipi [a,b], Silvia Marková [a], Jeremy B. Searle [c], Petr Kotlík [a,*]

[a] Laboratory of Molecular Ecology, Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, Rumburská 89, 27721 Liběchov, Czech Republic
[b] Department of Genetics and Microbiology, Faculty of Science, Charles University in Prague, Viničná 5, 12844 Prague 2, Czech Republic
[c] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 29 August 2014
Revised 17 October 2014
Accepted 22 October 2014
Available online 30 October 2014

Keywords:
Adaptation
Glacial refugia
Mitogenome
mtDNA
Myodes glareolus
Numt

ABSTRACT

We have revisited the mtDNA phylogeny of the bank vole *Clethrionomys glareolus* based on Sanger and next-generation Illumina sequencing of 32 complete mitochondrial genomes. The bank vole is a key study species for understanding the response of European fauna to the climate change following the Last Glacial Maximum (LGM) and one of the most convincing examples of a woodland mammal surviving in cryptic northern glacial refugia in Europe. The genomes sequenced included multiple representatives of each of the eight bank vole clades previously described based on cytochrome *b* (*cob*) sequences. All clades with the exception of the Basque – likely a misidentified pseudogene clade – were highly supported in all phylogenetic analyses and the relationships between the clades were resolved with high confidence. Our data extend the distribution of the Carpathian clade, the marker of a northern glacial refugium in the Carpathian Mountains, to include Britain and Fennoscandia (but not adjacent areas of continental Europe). The Carpathian sub-clade that colonized Britain and Fennoscandia had a somewhat different history from the sub-clade currently found in or close to the Carpathians and may have derived from a more northwesterly refugial area. The two bank vole populations that colonized Britain at the end of the last glaciation are for the first time linked with particular continental clades, the first colonists with the Carpathian clade and the second colonists with the western clade originating in a more southerly refugium in the vicinity of the Alps. We however found no evidence that a functional divergence of proteins encoded in the mitochondrial genome promoted the partial genetic replacement of the first colonists by the second colonists detected previously in southern Britain. We did identify one codon site that changed more often and more radically in the tree than expected and where the observed amino acid change may affect the reductase activity of the cytochrome bc_1 complex, but the change was not specific to a particular clade. We also found an excess of radical changes to the primary protein structure for geographically restricted clades from southern Italy and Norway, respectively, possibly related to stronger selective pressure at the latitudinal extremes of the bank vole distribution. However, overall, we find little evidence of pervasive effects of deviation from neutrality on bank vole mtDNA phylogeography.

1. Introduction

Since its introduction as a phylogenetic marker, mitochondrial DNA (mtDNA) has made an enormous contribution to our understanding of the processes of evolutionary divergence (Brown and Wright, 1975; Avise et al., 1979, 1987). The 'golden age' of mtDNA in phylogenetics has clearly passed, but the molecule does retain a number of features for which it remains appealing as a marker (Galtier et al., 2009), even in the 'genomics era' (see special issue on Mitogenomics and Metazoan Evolution in Mol. Phylogenet. Evol. 69, 2013). After decades of relying on restriction endonuclease cleavage maps, or profiles (Brown and Vinograd, 1974), and short, single gene, or single region sequences, the focus is now increasingly on complete mitochondrial genome sequences, a transition enabled by technological advancements and the steadily growing affordability of sequence data. Therefore, mitogenome sequences now are often the marker of choice in studies of species where there is little or no previous mtDNA data available. However, they also offer the possibility to critically revisit and refine earlier results for species with mtDNA phylogenies characterized based on short sequences (Carr and Marshall, 2008; Jacobsen et al., 2012; Keis et al., 2013; Pabijan et al., 2013).

* Corresponding author. Fax: +420 315 639 510.
E-mail address: kotlik@iapg.cas.cz (P. Kotlík).

http://dx.doi.org/10.1016/j.ympev.2014.10.016
1055-7903/

With our increasing knowledge of mtDNA evolution it has become clear that some of the key features earlier claimed for mtDNA do not always strictly hold, with a deviation from neutrality perhaps being the most worrisome (Nachman et al., 1996; Foote et al., 2011; McClellan et al., 2005; Soares et al., 2013). Although the consequences may be neither as detrimental nor pervasive as once thought (Bazin et al., 2006), a deviation from neutrality certainly can seriously mislead the inference about evolutionary history, for example due to a selection-driven mitotype fixation effectively mimicking a rapid range expansion (Rato et al., 2011), or when fixation of allospecific mtDNA causes more distantly related species to appear as sister clades (Shaw, 2002; Marková et al., 2013).

Vertebrate mtDNA contains 13 genes encoding subunits of the protein complexes of the oxidative phosphorylation system (OXPHOS), all potential targets of adaptive evolution. For instance, geographical differences in the rate of protein evolution (measured by the ratio of non-synonymous to synonymous nucleotide substitutions) were found by comparing 104 complete human mtDNA sequences, with the highest rates in the Arctic as compared to the temperate and tropical zones, possibly as a results of functional adaptation of the OXPHOS proteins to climate (Mishmar et al., 2003). Interestingly, a genome-wide study found a significant increase in the rate of protein evolution for genes expressed in the mitochondria (i.e. not only mtDNA-encoded) in the dolphin lineage as compared to other mammalian lineages, suggestive of an OXPHOS adaptation specific to the dolphin lineage (McGowen et al., 2012).

Another caveat, more technical in its nature, is the risk of mistaking nuclear pseudogenes or numts (nuclear copies of mitochondrial DNA) for authentic mtDNA sequences (Bernt et al., 2013a), which has been recognized since shortly after the introduction of PCR amplification into mtDNA analysis (Zhang and Hewitt, 1996). Although frame-shift mutations and premature termination codons may often reveal pseudogenes, not all pseudogene sequences need to be out of the frame or contain premature stop codons (Mirol et al., 2000), and these criteria are not applicable in studies targeting other mtDNA regions than protein coding genes (PCGs). Other means of verifying the mtDNA origin of the data therefore need to be employed. For example, a new, highly divergent and early-branching clade was found in an mtDNA study of an island population of house mouse and only when a whole mitochondrial genome of one individual was sequenced was it recognized that the sequence defining the clade was a pseudogene (Macholán et al., 2012). Clearly, if unrecognized as a pseudogene in origin, such a clade would have seriously misled the inferences about the phylogeny and evolutionary history of the *Mus musculus* complex. The basis for mtDNA pseudogenes is a matter of discussion (Doolittle, 1998; Berg and Kurland, 2000; Blanchard and Lynch, 2000), but they most likely are a result of transposition of mtDNA sequence to the nuclear genome (Zhang and Hewitt, 1996). Nuclear copies of mtDNA vary between taxa in their copy number, length and position on the chromosomes, but as Benasson et al. (2001) reviewed, their presence is widespread across species. Because of the generally lower mutation rate of nuclear DNA compared to mtDNA, nuclear copies may retain sequence characteristics of extinct mitotypes that existed at the time of the transfer to the nucleus. Nuclear integrations predating the origin of current mtDNA diversity of a species would appear as early-branching 'ancient' clades (e.g. Mirol et al., 2000; Macholán et al., 2012). Lineage-specific pseudogenes should in turn emerge as sister-clades to mtDNA clades of the lineage they are specific to (e.g. Triant and DeWoody, 2007a). Several well-studied examples also show that repeated translocations of mtDNA into nuclear DNA are often followed by tandem duplication (Lopez et al., 1994; Triant and DeWoody, 2007b), which may further complicate isolation of authentic mtDNA fragments in cases where mtDNA-specific primers are not available.

The present study revisits the intraspecific phylogeny of the bank vole, *Clethrionomys glareolus*, a small forest dwelling rodent widespread in temperate and boreal forests throughout Europe, using 33 complete mitochondrial genomes and a combination of phylogenomic approaches and statistical tests of the signatures of adaptive protein evolution. The bank vole is one of the key mammals used in genetic studies of the response of European fauna to climate change following the Last Glacial Maximum (LGM; Deffontaine et al., 2005; Searle et al., 2009; Wójcik et al., 2010; Colangelo et al., 2012; Kotlík et al., 2014). The bank vole is also one of the most convincing examples of a woodland mammal surviving in cryptic northern refugia in Europe, that is, refugia located further north of the traditionally recognized southern refugia in the Mediterranean (Bilton et al., 1998; Kotlík et al., 2006; Bhagwat and Willis, 2008).

A number of divergent mtDNA clades have been described from various parts of the bank vole distribution (Fig. 1) based of sequence data from the gene encoding cytochrome *b* (*cob*; the gene nomenclature follows that proposed by Boore, 2006, and adopted by e.g. Bernt et al., 2013b), which has long served as a prime phylogenetic and phylogeographic marker for small mammals (e.g. Conroy and Cook, 2000; Iwasa and Suzuki, 2002; Jaarola and Searle, 2002; Michaux et al., 2004; Dubey et al., 2006). In the first comprehensive study by Deffontaine et al. (2005), five bank vole clades were described – three in southern Europe on the Mediterranean peninsulas of Iberia (Spain), Italy and the Balkans, and two widespread continental clades, a western European clade and an eastern clade with a distribution reaching as far as western Siberia (Fig. 1).

Some bank voles in the study of Deffontaine et al. (2005) collected at various localities, in northern Fennoscandia in particular (Fig. 1), possessed *cob* haplotypes closely related to those of a boreal and tundra species, the northern red-backed vole *Clethrionomys rutilus*. Tegelström (1987) showed that the introgression of mtDNA from *C. rutilus* to *C. glareolus* is a consequence of past hybridization between the two species (see also Tegelström et al., 1988; Abramson et al., 2009). Because of the geographic distribution of the bank voles with *C. rutilus* mtDNA, which are found from Fennoscandia as far east as the Ural Mountains and beyond (Fig. 1; Deffontaine et al., 2005; Abramson et al., 2009), the clade formed by the introgressed haplotypes was termed Ural clade by Deffontaine et al. (2005).

Later on, a detailed study of the eastern European bank vole populations revealed a distinct clade in the vicinity of the Carpathians (Fig. 1; Kotlík et al., 2006), most likely a survivor in cryptic glacial refugia that existed in this mountain chain (Kotlík et al., 2006; Wójcik et al., 2010). Another three years later, a Basque clade was described by Deffontaine et al. (2009), found to be endemic to the French Basque country (Fig. 1; see also Malé et al., 2012). Surprisingly, the Basque clade was highly divergent from the other six bank vole clades and it branched off at the base of the tree, appearing as a sister clade to the remaining bank vole diversity (Deffontaine et al., 2009).

Finally, in a recent study, Colangelo et al. (2012) described a deeply divergent bank vole clade from Calabria in the extreme south of the peninsular Italy (Fig. 1), a region known to harbor endemic genetic diversity for a number of vertebrate species (Vega et al., 2010). The Calabrian clade appeared as a sister clade to all other bank vole clades (including Basque), which caused Colangelo et al. (2012) to consider a separate taxonomic (subspecific) status of the Calabrian bank vole. Interestingly, while the majority of bank voles from Italy north of Calabria carried mtDNA