Kirkland, Corey (2021) *Genetic diversity of British water voles (Arvicola amphibius) and phylogenetics of the rodent subfamily Arvicolinae.* Master of Research (MRes) thesis, University of Kent,.

Genetic diversity of British water voles (*Arvicola amphibius*) and phylogenetics of the rodent subfamily Arvicolinae.

A thesis to the University of Kent for the degree of:

MSc by Research in Genetics

2020

Corey Kirkland

School of Biosciences

Declaration:

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institutions of learning.

Acknowledgements:

I would firstly like to say a huge thanks to my supervisor, Dr Marta Farré-Belmonte, for accepting me onto this MSc-R project and for her support throughout this project. She has provided me with technical support both in the laboratory and in bioinformatics and provided me with an amazing opportunity to develop as a research scientist. These skills will be incredibly useful as I hope to continue my career in research. It has been a pleasure to be a part of her research group at the University of Kent. I would also like to acknowledge all the members of the group – Cristina, Frances, and visiting researcher Lucía – who have been wonderfully supportive and who I have really enjoyed working alongside, as well as members of the ERGO group.

Also, a big thanks to Wildwood Trust and Hazel who have allowed me to sample water voles at their wildlife park. Plus, Kent Mammal Group for providing a research grant for consumables.

Lastly, I would like to thank Matthew, my friends and my family for their love and support.

Contents:

List of Figures:

List of Tables:

Abbreviations:

AIC	Akaike Information Criterion
AMOVA	Analysis of Molecular Variance
BP	Before Present
Cytb	Cytochrome b
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
GHR	Growth Hormone Receptor
GTR	Generalised Time Reversible
HMW	High Molecular Weight
IRBP	Interphotoreceptor Retinoid-Binding Protein
LGM	Last Glacial Maximum
MCMC	Markov Chain Monte Carlo
ML	Maximum-Likelihood
MSA	Multiple Sequence Alignment
mtDNA	Mitochondrial DNA
NCBI	National Centre for Biotechnology Information
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulfate
STR	Short Tandem Repeat
YD	Younger Dryas

Abstract:

The European water vole (*Arvicola amphibius*) has experienced rapid decline in Britain, resulting in numbers declining by approximately 90% in the last century, making them a conservation priority. Its phylogeny within the subfamily Arvicolinae, and its relationship to other taxa, remains debated. Additionally, the impact of captive breeding programs on the genetic diversity of water voles is unknown.

We firstly optimise DNA extraction protocols for tail tissue, hair, faeces, buccal swabs, and cell culture to achieve high DNA yields and purity. We then sequence the mitochondrial control region for 17 captive water voles at Wildwood Trust using tissue, hair, and faecal samples collected at the park to assess their genetic diversity and population structure. Lastly, we use mitochondrial genomes and individual phylogenetic markers to construct the phylogeny of the subfamily Arvicolinae.

Our study provides protocols for the extraction of DNA achieving high yields and improved purity, with cell culture resulting in the highest median yield, followed by buccal swabs and tail tissue. Our results reveal considerable genetic diversity in the mitochondrial control region of captive water voles sampled at Wildwood Trust, with relatively high haplotype diversity. Similar haplotype diversity was also seen in natural populations in Britain. Captive water voles clustered closely with South East of England haplotypes, and were found within the English/Welsh clade. Additionally, we provided support for the phylogeny of several genera of Arvicolinae. Using mitogenomes provided the most resolved phylogenetic tree when compared with other genetic markers and approaches but lacked sequences for all genera to fully resolve phylogeny.

The mitochondrial genome provides a useful marker to study both the conservation and population genetics of water voles, as well as the phylogenetics of the subfamily Arvicolinae. Our study provides support for the breeding program at Wildwood Trust and provides a framework for future conservation studies.

Word Count: 299

1. Introduction:

1.1. Population Genetics of British Water Voles:

Population genetics is a field which examines the genetic variation within populations and is becoming increasingly useful to study declining populations and halt the decline of species. IUCN recognises the importance of conserving genetic diversity as a key conservation priority (McNeely *et al.*, 1990). Genetic diversity can be defined as the variation in the amount of genetic information within and among individuals of a population, a species, an assemblage, or a community (defined by the United Nations in 1992). It enables populations to evolve and adapt to environmental change. It can be measured by polymorphisms such as single-nucleotide polymorphisms (SNPs), average heterozygosity, and allelic diversity.

BOX 1 – Population Genetics Terms:

Heterozygosity – Two different alleles at a given locus.

Homozygosity – Two identical alleles at a given locus.

Inbreeding Depression – A reduction in biological fitness in a population caused by inbreeding.

Outbreeding Depression – A reduction in biological fitness in a population caused by the breeding of genetically distant groups or populations.

Haplotype – A set of DNA variations or polymorphisms that are inherited together.

Loss of genetic diversity is associated with small or declining populations and can lead to inbreeding, reduction in reproductive fitness, and inbreeding depression. Inbreeding depression reduces heterozygosity and increases homozygosity, increasing the number of alleles with deleterious effects in the population. Threatened species often have lower heterozygosity and it has been proven that genetic factors affect species before extinction does (Spielman, Brook and Frankham, 2004).

This process is known as the extinction vortex. Small populations undergo inbreeding and random genetic drift, which leads to a loss of genetic diversity. This reduces an individual's fitness and a populations ability to adapt to changing environmental conditions, leading to lower rates of reproduction and higher mortality. The process continues with populations becoming smaller until they then become extinct.

Captive breeding programs and reintroductions (ex-situ conservation) can be a useful management tool to halt the decline of individuals in small populations, but only when the maintenance of genetic diversity is considered. It is vital to reduce inbreeding during these programs, such as using pedigrees that minimise mean kinship. Outbreeding depression and genetic swamping can also occur in captive breeding programs and reintroductions, as local distinct genetic features and genetic variability can be lost due to the integration of alien gene pools, decreasing evolutionary potential (Laikre *et al.*, 2010; Canu *et al.*, 2013).

Genetic markers, such as nuclear and mitochondrial DNA (mtDNA) are extensively used to determine the genetic diversity of natural and captive animal populations. Mitochondria are found in abundance throughout cells, resulting in thousands of copies of mitochondrial DNA in each cell. In mammals, the mitochondrial genome contains 37 genes. It is inherited maternally, thus offspring inherit a single mtDNA genotype from their mother. The mitogenome is haploid, unlike the nuclear genome, and there is generally no recombination of DNA.

The European Water Vole:

The European or northern water vole (*Arvicola amphibius*) is a species of rodent in the subfamily Arvicolinae, a group which contains voles, lemmings, and muskrats. It has a wide distribution throughout Europe and Asia, with the overall population trend being stable and the species listed as 'Least Concern' (Batsaikhan *et al.*, 2016). The mammal is mostly found in riparian zones, in habitat alongside rivers or streams.

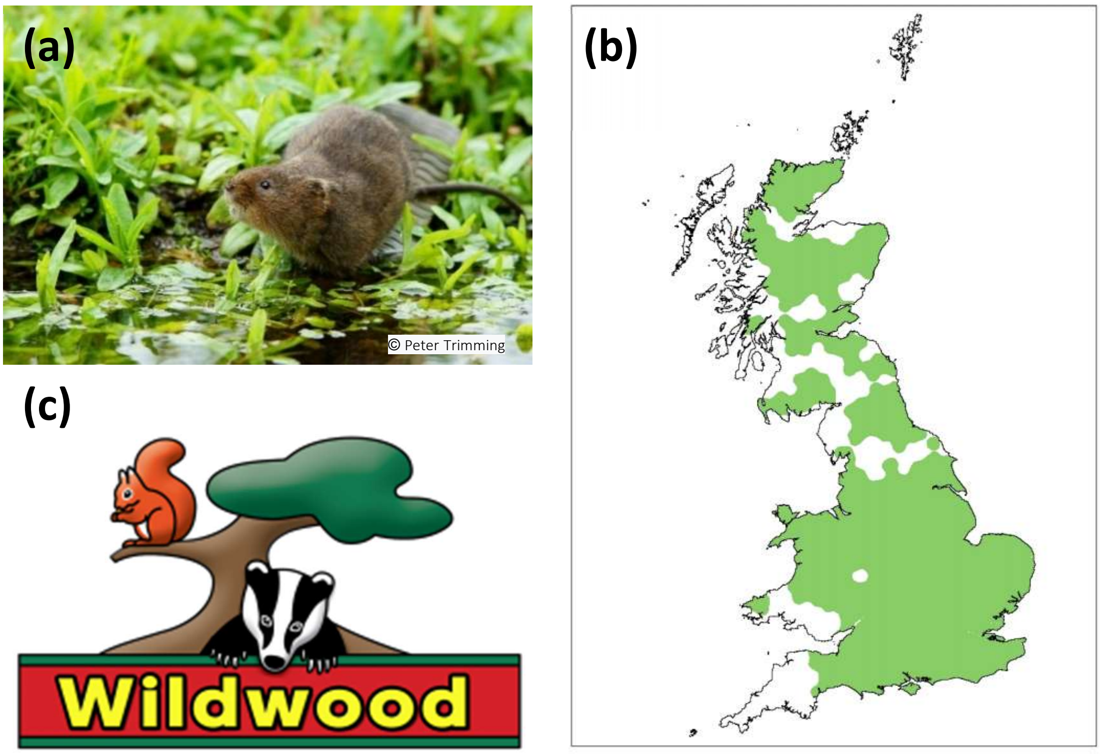


Figure 1: (a) European water vole (*Arvicola amphibius*) in riparian habitat, (b) distribution in Britain (Mathews *et al.*, 2018) and (c) Wildwood Trust logo, Canterbury.

In Britain, populations of water voles are rapidly declining, faster than any other British mammal in the last century (e.g. Strachan and Jefferies, 1993; Strachan, 2004). Regionally, water voles are classified by the 'IUCN Red List' as 'Endangered' in England, 'Near Threatened' in Scotland, and 'Critical' in Wales (Mathews *et al.*, 2018). The main cause of their decline has been increased predation by the invasive American mink (*Neovison vison*), with habitat loss and pollution of watercourses also having a negative effect on populations (Jefferies, Morris and Mulleneux, 1989; Barreto *et al.*, 1998). Therefore, water voles are a conservation priority in Britain, with several projects seeking to increase numbers and manage suitable habitat. Wildwood Trust near Canterbury UK have a conservation program offering mitigation services, health screening, captive breeding, and reintroductions of British water voles.

Population Genetics of British Water Voles:

The phylogeographic structure of European water voles, as with the majority of northern hemisphere extant species, has been determined by glaciation events (Hewitt, 1996, 1999; Bernatchez and Wilson, 1998). The Younger Dryas (YD), between 12,800 to 11,500 before present (BP), was a period of retreat and re-advance of glaciers (Walker *et al.*, 2012). Northern Europe was fully glaciated and central Europe was mostly permafrost during this period [4]. Regions in southern Europe (Iberia, Italy, and the Balkans) were ice and permafrost free, creating pocket refugia for populations, known as a refugial peninsula (Hewitt, 1999). Populations then recolonised north with the retreat of the glaciers.

The first study using molecular markers to study European water vole phylogeography used the mitochondrial cytochrome b (*Cytb*) gene (summarised in Taberlet *et al.*, 1998). The study found that samples were from three lineages: (1) eastern and central Europe, the Balkans, and Fennoscandia; (2) Italy; and (3) France and Spain. The small number of English samples used were clustered in the first lineage.

Later studies used the mitochondrial control region to study the evolutionary history of British water voles. Piertney *et al.*, 2005 sampled 62 individuals from 57 locations, both from extant individuals and museum samples less than 75 years old. The study found a major division between English/Welsh and Scottish water voles, with individuals from the two regions found in separate haplogroups and phylogenetic clades **(Figure 2)**. A minimum of 16 mutational steps was found between the two haplogroups, indicating considerable mitochondrial divergence. Previous studies have debated this distinction, with some stating that Scottish water voles morphologically have darker hair than English and Welsh water voles (Miller, 1912) and differences in the size of their tail and hind feet. Whilst more recent studies have found no considerable differences morphologically between the two water voles (Corbet *et al.*, 1970; Telfer *et al.*, 2003).

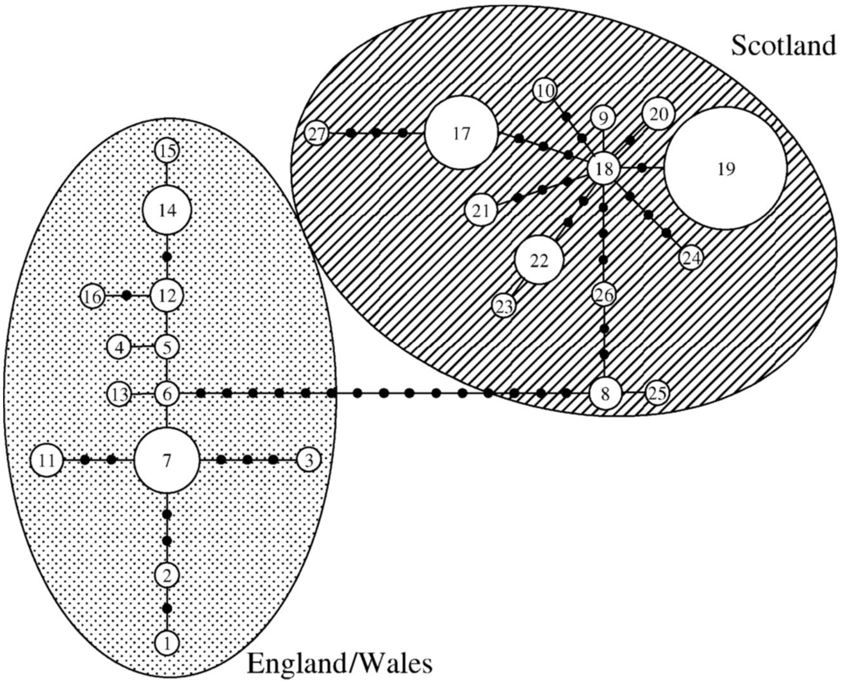
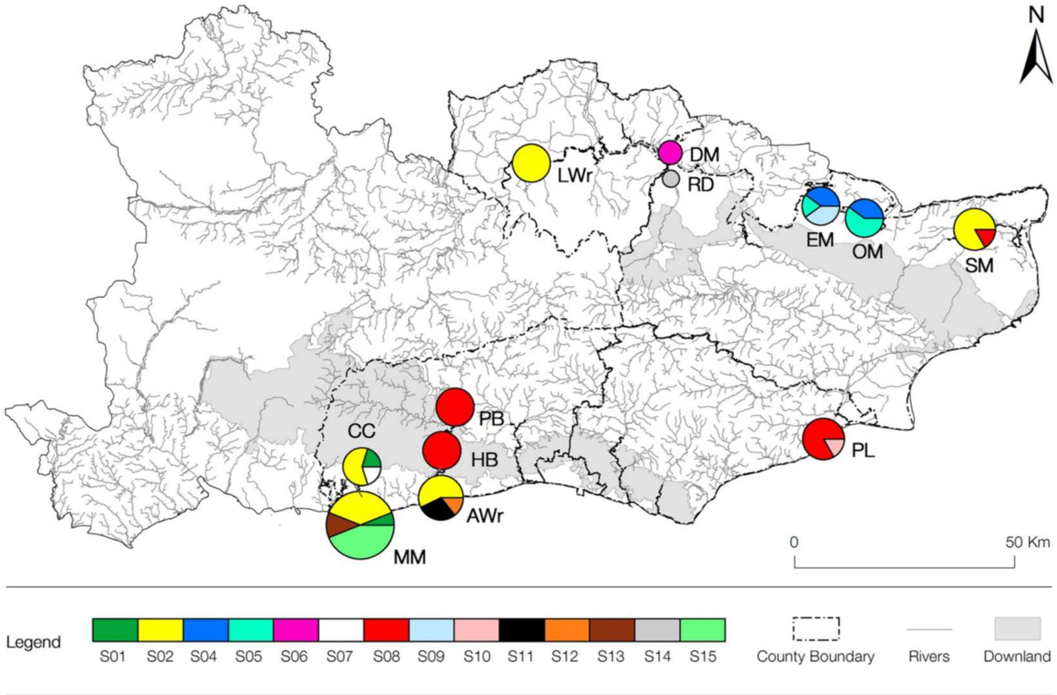


Figure 2: Haplotype network of British water voles showing England/Wales and Scotland haplogroups (Piertney *et al.*, 2005). Each pie chart represents a different haplotype. The size of the pie chart is proportional to the number of individuals in each haplotype. Dotted lines between haplotypes show the number of mutational steps between two haplotypes. The circle filled with dots shows the England/Wales haplogroup, whilst the circle with lines shows the Scotland haplogroup.

Another study used ancient museum specimens from Britain and mainland Europe, both before the last glacial maximum (Pleistocene period) and following the YD (Holocene period), to understand the colonisation history of British water voles (Brace *et al.*, 2016). The study supported the hypothesis that there were two colonisation events by water voles into Britain. The first colonisation occurred before the last glacial maximum (LGM) throughout England. A second colonisation event then occurred after the Pleistocene, displacing the first colonisers, while the newer colonisers remained in England throughout the Holocene until the present day. The first colonisers were displaced north into Scotland, creating the two modern evolutionary distinct groups.

Studies have also looked at the genetic diversity of natural water vole populations, using both the mitochondrial control region and microsatellite markers. Populations within the two locations have been shown to have significant genetic structure when using mtDNA (Piertney *et al.*, 2005). Analysis of Molecular Variance (AMOVA) revealed that haplotype frequencies between regions and between populations within regions was a significant proportion of the observed haplotype frequency. The remaining haplotype frequency was between the two major divisions. Several studies looking at the phylogeography of water voles in the South East and East of England found significant genetic diversity (Baker, 2015; Baker *et al.*, 2020; **Figure 3**). Using mtDNA alone uncovered substantial genetic structure between watershed populations, whilst finer scale structure between populations within watersheds was found when using both mtDNA and microsatellites (Baker *et al.*, 2020). Mitochondrial DNA also showed high levels of haplotype diversity among natural populations.



Contains National Statistics and Ordnance Survey data © Crown copyright and database right 2018. Contains Natural England data © Natural England copyright.

Figure 3: South East of England haplotypes (Baker *et al.*, 2020). Colours represent the 15 haplotypes. The pie charts represent the proportion of a given haplotype at each population sampled.

Genetic Sampling:

In order to study the genetic diversity of animal populations, we need to obtain DNA from several individuals. Non-invasive genetic sampling is a method of collecting DNA that is left behind by animals, without disturbing them (Taberlet, Waits and Luikart, 1999). Non-destructive sampling is more invasive and can include capturing the animal for a short period of time to obtain the sample, but less invasive than destructive sampling, where the animal is killed for the sample. Non-invasive sampling is becoming more widely used in population genetic studies, with the development of DNA extraction, amplification, and sequencing technologies. It can include sampling hair, faeces, urine, feathers, shed skin, saliva, and eggshells that are left behind by animals (Waits, Lisette and Paetkau, 2005).

One aim of genetic sampling is to obtain sufficient quantities of high-quality DNA. Using non-invasive sampling frequently results in lower quantities of DNA and DNA samples that contain contaminants causing lower polymerase chain reaction (PCR) success rates. When using nuclear loci this can result in genotyping errors such as allelic dropout, where only one of the two alleles present at a heterozygous locus is amplified (Taberlet, Waits and Luikart, 1999). This can arise when concentrations of the DNA template are below 0.05ng/10ml (Gagneux, Boesch and

Woodruff, 1997). Mitochondrial DNA is more reliable when it comes to PCR amplification, as it is haploid, so avoids allelic dropout, and is in larger quantities in non-invasive samples than nuclear DNA.

DNA extraction of non-invasive samples has improved greatly, resulting in higher DNA concentrations and improved DNA quality. There are many different approaches to extracting DNA, but all contain the same main steps of lysis with an extraction buffer, precipitation with alcohol, then resuspension in water or an elution buffer. The extraction buffer contains salts (such as NaCl and Tris-HCl) that stabilises the pH, protecting the negatively charged phosphate groups on the DNA backbone. Detergents, such as sodium dodecyl sulfate (SDS), dissolve the lipid membranes of the cell, releasing the contents of the cell and nucleus. Chelating agents (e.g., ethylenediaminetetraacetic acid (EDTA)) reduce protease or DNAse activity. A proteinase (e.g., proteinase K) can also be added to the extraction buffer which enzymatically breaks down proteins, that would otherwise degrade DNA. The concentrations of each of these components differ between extraction protocols and can be optimised for different sample types.

Traditionally, phenol-chloroform was the preferred method for extracting DNA from samples. However, this method uses extremely hazardous chemicals, and it requires significant bench time compared to DNA extraction kits (Schiebelhut *et al.*, 2017). Phenol-chloroform-isoamyl alcohol separates lipids and cellular debris in the solvent phase and DNA in the aqueous phase. Between the two phases is a layer which contains protein. The aqueous phase containing DNA can then be precipitated to increase DNA concentration and DNA purity, using ethanol or isopropanol. Phenol-chloroform DNA extraction consistently produces the highest DNA yield and purity, when compared to other extraction methods (Schiebelhut *et al.*, 2017).

Silica membrane-based extraction kits require less harmful chemicals and quicker extraction times than the more laborious phenol-chloroform extraction (Schiebelhut *et al.*, 2017). Spin columns contain silica beads which cause DNA to bind to them when high concentrations of chaotropic salt are passed through the column. Contaminants also pass through the column during multiple washes and centrifugations. DNA is then eluted into water or a buffer. This method is widely used in conservation and population genetics studies but produces lower DNA concentrations and lower DNA purity. Studies have found that mitochondrial DNA is lost from silica membrane-based kits at a higher rate than nuclear DNA (Guo *et al.*, 2009).

Other extraction protocols have been developed, such as an extraction method for mouse tails using ethanol instead of phenol-chloroform (Wang and Storm, 2006) or using isopropanol instead of ethanol. These methods provide fewer tube changes, increasing speed of extraction and are more cost effective than buying commercial kits.

1.2. Phylogenetics of Arvicolinae:

Phylogenetics is the study of evolutionary relationships in biology, from individual genes to populations, species, and groups of species. Traditionally, morphological characteristics were used to construct phylogenies of species. More recently molecular markers and whole genomes are being used to further resolve phylogenies and offer a greater understand of evolutionary processes. Phylogenetics is incredibly important in conservation management by resolving taxonomic uncertainties between species and defining evolutionary significant units (ESUs) within species that need to be protected separately. Phylogenetics is also used in various other biological disciplines, such as pathology to identify emerging pathogens, understanding the relationship to other pathogens, and the likely source of transmission.

BOX 2 – Phylogenetic Terms:

Monophyletic – A group of organisms all sharing a common ancestor.

Paraphyletic – An artificial group of organisms sharing a common ancestor but that does not include all descendants.

Clade – A monophyletic group.

Node – A branch point on a phylogenetic tree.

Polytomy – An internal node of a phylogenetic tree that has more than two immediate descendants.

Basal Group – The earliest diverging group within a clade.

Rodents (order Rodentia) are one of the most speciose orders within the mammalian kingdom, containing 2,552 species (513 genera) out of the 6,495 species (1,314 genera) of mammals (Burgin *et al.*, 2018). Within Rodentia is the family Muroidea (mice, rats, voles, hamsters, etc.) and within this group the subfamily Arvicolinae (containing voles, lemmings, and muskrats). The phylogeny of genera within Arvicolinae remains debated, with the evolutionary relationships of many genera and species unresolved. One reason is the subfamily contains 150 species in 30 genera (Carleton and Musser, 2005), with very few sequenced genomes nor a wide selection of sequenced phylogenetic markers.

Several studies have looked at the phylogenetics of rodents and species within Arvicolinae, using mitochondrial and nuclear molecular markers. An early study used 1.2kb of the interphotoreceptor retinoid-binding protein (*IRBP*) nuclear gene with 22 rodent species (DeBry and Sagel, 2001). This validated the monophyletic group Muroidea, which contains the subfamily Arvicolinae, but did not cover sufficient arvicoline taxa. Later studies, such as Blanga-Kanfi *et al.*, 2009, used more molecular markers to resolve the evolutionary relationships between Rodentia families, confirming the monophyletic group Arvicolinae.

Studies focusing on the phylogenetics of Arvicolinae started by sequencing and analysing the mitochondrial cytochrome b (*Cytb*) gene. Several studies found rapid, near simultaneous radiations when using this marker (summarised in Robovský, Řičánková and Zrzavý, 2008). Another hypothesis is that substitution saturation has occurred in this gene, reducing its value as a phylogenetic marker. Genetic saturation was found at both transitions and transversions of the *Cytb* gene in arvicoline species (Triant and DeWoody, 2008).

Multiple genetic markers have been used in subsequent studies. For example, one study used mitochondrial *Cytb* and nuclear growth hormone receptor (*GHR*) genes, as well as morphological characters (Robovský, Řičánková and Zrzavý, 2008). They found the basal arvicoline in the proposed phylogeny tree **(Figure 4a)** to be *Ellobius*, *Prometheomys*, *Hyperacrius*, *Eolagurus* and *Lagurus*. The next to branch was the clade 'Dicrostonychini' containing genera *Dicrostonyx*, *Phenacomys*, and *Arborimus*. The next group to diverge was not fully resolved in this study, with *Dinaromys*, *Neofiber* and *Ondatra*, and the clade 'Lemmini' (containing *Synaptomys*, *Lemmus*, and *Myopus*) forming a polytomy at this node. The clade 'Clethrionomyini' (*Eothenomys, Myodes, and Alticola*) was well-supported in both this study and other studies. The final clade 'Arvicolini' contains genera *Arvicola*, *Lemmiscus*, *Stenocranius*, *Chionomys*, and *Microtus*. Other genera were originally found within this group, but this led to the paraphyly of *Microtus*. The genera *Neodon*, *Alexandromys*, *Mynomes*, *Lasiopodomys*, *Proedromys*, and *Terricola* were reclassified in this study to *Microtus*. The relationships between taxa in this clade is poorly resolved due to polytomies.

A more recent tree of Arvicolinae can be found in a later study which sampled 900 Muroidea species, with substantial numbers of arvicolines included, using six molecular markers (Steppan and Schenk, 2017; **Figure 4b**). Although not all genera were accounted for, there is better resolution at some nodes. The basal arvicolines in this tree were *Prometheomys*, followed by a clade containing *Ondatra* and *Neofiber*. This was followed by a monophyletic group containing *Dicrostonyx*, *Arborimus*, and *Phenacomys* in one clade and another clade containing *Lemmus*, *Myopus*, and *Synaptomys*. This supports the previously described study with the clades 'Dicrostonychini' and 'Lemmini'. The well supported clade containing *Eothenomys*, *Myodes*, and *Alticola* diverged next. *Dinaromys* and *Lagurus* and *Eolagurus* diverged much later in this tree, followed by *Ellobius*. This is significantly different to their proposed phylogeny in Robovský et al., 2008. *Arvicola* grouped together with *Lemmiscus*, as the basal taxa within 'Arvicolini', and followed by the branching of *Chionomys*. There is then a polytomy between a group containing

Volemys and *Proedromys bedfordi*, a branch containing *Proedromys liangshanensis*, and a clade containing *Microtus*, *Lasiopodomys*, *Neodon*, and *Blanfordimys* species. *Microtus* is also paraphyletic in this study. This demonstrates even with extensive molecular markers the phylogeny of all Arvicolinae genera is still unresolved due to lack of support at some nodes.

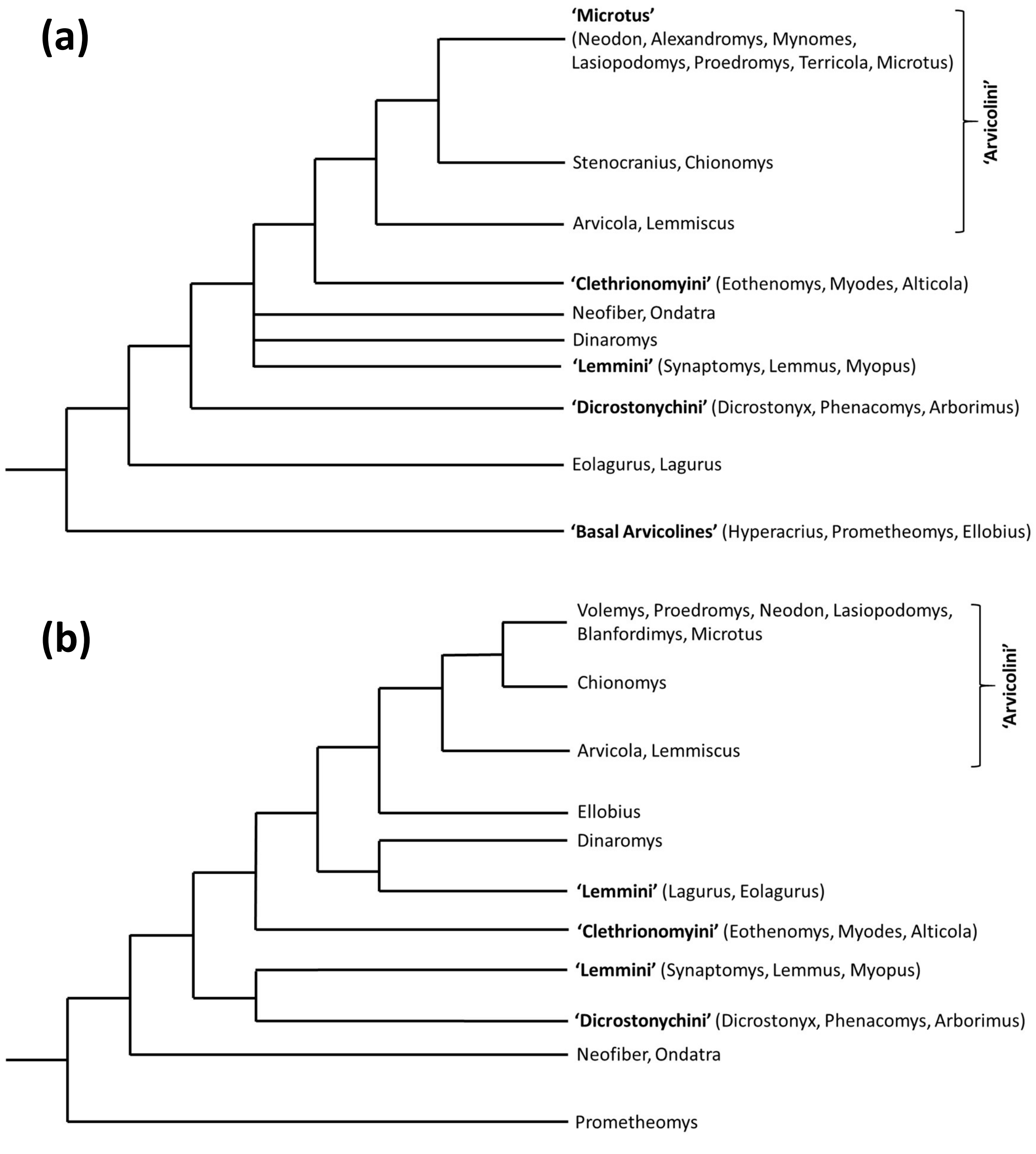

(a)
'Microtus'
(Neodon, Alexandromys, Mynomes, Lasiopodomys, Proedromys, Terricola, Microtus)
Stenocranius, Chionomys
Arvicola, Lemmiscus
'Arvicolini'
'Clethrionomyini' (Eothenomys, Myodes, Alticola)
Neofiber, Ondatra
Dinaromys
'Lemmini' (Synaptomys, Lemmus, Myopus)
'Dicrostonychini' (Dicrostonyx, Phenacomys, Arborimus)
Eolagurus, Lagurus
'Basal Arvicolines' (Hyperacrius, Prometheomys, Ellobius)
(b)
Volemys, Proedromys, Neodon, Lasiopodomys, Blanfordimys, Microtus
Chionomys
Arvicola, Lemmiscus
'Arvicolini'
Ellobius
Dinaromys
'Lemmini' (Lagurus, Eolagurus)
'Clethrionomyini' (Eothenomys, Myodes, Alticola)
'Lemmini' (Synaptomys, Lemmus, Myopus)
'Dicrostonychini' (Dicrostonyx, Phenacomys, Arborimus)
Neofiber, Ondatra
Prometheomys

Figure 4: Simplified cladograms from (a) the proposed phylogeny of Arvicoline by Robovský, Říčánková and Zrzavý, 2008 and (b) the phylogenetic tree of Arvicolinae by Steppan and Schenk, 2017.

Recently, the mitochondrial genome has been sequenced for a number of arvicoline species (e.g. (Folkertsma *et al.*, 2018; Bondareva and Abramson, 2019; Zhu *et al.*, 2019; Alqahtani *et al.*, 2020). These studies focus on genomic sequencing and mapping of mitochondrial genes, with brief phylogenetic analyses using available mitogenomes. However, the analyses are not sufficient to fully resolve the phylogenetic relationships between genera and species within Arvicolinae, due to small sample sizes and lack of extensive phylogenetic analysis.

1.3. Hypothesis:

We hypothesise that firstly sufficient quantities of DNA will be extracted from both tissue and non-invasive samples to sequence the mitochondrial DNA control region. Secondly, the genetic diversity of water vole populations in Britain will be low and considerably lower in the captive population, due to the impact that captive breeding programs can have on genetic diversity. Thirdly, using the mitochondrial genome will improve and further resolve the phylogeny of taxa within the rodent subfamily Arvicolinae.

1.4. Research Aims:

This research project is divided into three sections with the following aims:

1. To optimise DNA extraction protocols from various sample types, prioritising non-invasive genetic sampling.
2. To assess the genetic diversity of British water voles (*Arvicola amphibius*) at Wildwood Trust through the random sequencing of individuals in captivity, comparing the genetic diversity between the captive population and natural populations in the South East of England, Britain, and the rest of Europe.
3. To improve the phylogeny of the rodent subfamily Arvicolinae using available molecular markers and various phylogenetic approaches.

2. Materials and Methods:

2.1. Optimisation of DNA Extraction:

Sample Collection:

Water vole samples were collected from Wildwood Trust on several occasions totalling 20 individuals **(Table 1)**. Four samples were collected from tail tissue of deceased water voles and stored at -20°C. Six samples were collected from hair tufts collected in 2019 and stored in paper envelopes at room temperature. Another 10 samples were collected in 2019 from faecal pellets found in water bowls within enclosures and stored at -20°C. All individuals were randomly chosen, and non-invasive sampling was prioritised. Faecal samples were only collected from enclosures with single voles or those containing mother and offspring.

Table 1: Wildwood Trust samples.

Sample No.	Sample Type	Enclosure No.	Local ID	Sex
1	Tissue	TB31	-	-
2	Tissue	WW46	-	-
3	Tissue	WW0304/34	-	Male
4	Tissue	WW34/39	-	-
5	Hair	Q88	-	Male
6	Hair	Q100	-	Male
7	Hair	R95	-	Male
8	Hair	R12	-	Male
9	Hair	R28	-	Male
10	Hair	Q100	-	Male
11	Faecal	R2	2228	Male
12	Faecal	Q52	2245	Female
13	Faecal	Q42	2218	Female
14	Faecal	Q7	2264	Female
15	Faecal	Q75a	2326	Female
16	Faecal	R50	2232	Male
17	Faecal	R51	2225	Male
18	Faecal	Q58	2314	Male
19	Faecal	Q100	2185	Female
20	Faecal	R27	2445	Female

DNA Extraction of Tissue Samples:

Tail tissue was obtained from frozen deceased water voles due to its accessibility and ease when dissecting. Approximately 1-2 cm of tail tissue was used and fragmented into smaller pieces. Water vole samples 1 and 2 were used to optimise DNA extraction of tissue samples. DNA was then extracted from samples 3 and 4 using the optimised protocol. Based on preliminary studies the ‘Qiagen DNeasy Blood and Tissue Kit’ was omitted due to very low concentrations of DNA and poor purity ratios. All DNA concentrations and ratios were measured using NanoDrop and is applicable to all sample types.

Extraction buffers were assessed using phenol-chloroform DNA extraction. Three buffers were selected: buffer 1 (Jain *et al.*, 2017) containing 10 mM Tris-Cl (pH 8.0), 25 mM EDTA (pH 8.0), 100 mM NaCl, and 0.5% SDS; buffer 2 (Green, M. R., Hughes, H., Sambrook, J. and MacCallum, 2012) containing 20 mM Tris-HCl (pH 8.0), 5 mM EDTA (pH 8.0), 400 mM NaCl, and 1% SDS; and buffer 3 (Wang and Storm, 2006) containing 100 mM Tris-HCl (pH 8.0), 5 mM EDTA (pH 8.0), 200 mM NaCl, and 0.2% SDS. For the lysis step, 500 µl of the selected buffer and 20 µl of proteinase K was added to the sample and incubated overnight at 55°C on a shaking platform. Phenol-chloroform extraction followed protocol by Green and Sambrook, 2012.

For the phenol-chloroform extraction, an equal volume of phenol:chloroform:isomyl alcohol was added to tissue samples and placed on a rocking platform for 30 mins. Samples were then centrifuged at 15,000x g for five minutes (at room temperature) and the upper aqueous phase was transferred to a clean Eppendorf tube. DNA was precipitated by adding an equal volume of isopropanol and centrifuged for 15 mins at 13,000x g (at 4°C). Isopropanol was removed, the pellet was rinsed with 70% ethanol, and dried at room temperature. The pellet was dissolved overnight in 100 µl of ddH_2O at 4°C.

Ethanol DNA extraction followed (Wang and Storm, 2006). A total of 300 µl of extraction buffer and 6 µl of proteinase K were added to the tissue sample, then incubated overnight at 55°C on a rocking platform. Next, 1 mL of 100% ethanol was added and centrifuged at 15,000x g for 30 mins. Ethanol was poured out and DNA pellets were washed with 70% ethanol. Samples were centrifuged at 15,000x g for 20 mins, ethanol poured out, 300 µl of ddH_2O added, and then incubated at 55°C for two hours with lids open. An overnight incubation at 4°C allowed the DNA to fully dissolve.

For the isopropanol protocol 485 µl of extraction buffer and 15 µl of proteinase K were added to the tissue sample, and then incubated overnight at 55°C on a rocking platform. Samples were then centrifuged for 20 mins at 3000 rpm (at room temperature) and the supernatant transferred to a new Eppendorf tube with 500 µl of isopropanol, followed by incubation for 30 mins at -80°C and further centrifugation for 20 mins at 3000 rpm (at 4°C). The supernatant was discarded, 70% ethanol was added, and tubes were centrifuged for 20 mins at 3000 rpm (4°C). The supernatant was discarded again, and the pellet was dried at 37°C. A total of 100 µl of ddH_2O was added and incubated overnight at 4°C.

The phenol-chloroform extraction protocol was then optimised. 'Optimised Phenol-Chloroform (1)' was identical to the previous protocol described above and used as a control. 'Optimised Phenol-Chloroform (2)' was identical to the control apart from an ethanol precipitation rather than an isopropanol precipitation. 'Optimised Phenol-Chloroform (3) was different in that samples were centrifuged with isopropanol for 30 mins rather than 15 mins. Protocol 3 was chosen for the remaining tissue samples. A 200 µl ddH$_2$O elution volume was used in all three phenol-chloroform optimisation samples.

DNA Extraction of Hair Samples:

Prior to DNA extraction, tufts of hair were collected from water voles and stored in paper envelopes. Each tuft contained multiple hairs and follicles. The 'Qiagen DNeasy Blood and Tissue Kit' (following the manufactures protocol) was compared with the optimised phenol-chloroform extraction for tissue samples (described previously). Multiple hairs were placed into each Eppendorf tube. Hairs from both protocols were incubated overnight at 56°C with the protocols stated volume of extraction buffer and proteinase K. DNA was eluted into 100 µl of AE ('Qiagen') or 100 µl ddH$_2$O (phenol-chloroform).

The buffer was modified to improve the lysis process. The new extraction buffer (Pfeiffer *et al.*, 2004) contained 100 mM Tris HCl (pH 8.0), 100 mM NaCl, 3 mM CaCl$_2$, 2% SDS, and 40 mM dithiothreitol (DTT). A total of 1360 µl of extraction buffer and 80 µl of lysis buffer (proportional to the number of repeats) was added to a falcon tube containing the tuft of hair and incubated overnight at 56°C with agitation. The optimised phenol-chloroform protocol for the tissue samples was used, but with an addition of 5 minutes of centrifugation with isopropanol at maximum speed. For this experiment elution volumes differed (100 µl, 75 µl, and 50 µl).

The remaining hair samples were extracted using the optimised buffer and protocol. Each hair tuft was placed into a falcon tube with 680 µl of extraction buffer and 80 µl of proteinase K, which allowed for a repeat and a change in concentration. DNA was finally eluted in 50 µl of ddH$_2$O.

DNA Extraction of Faecal Samples:

DNA from the faecal samples was extracted using the 'Qiagen QIAamp DNA Stool Mini Kit' and/or the 'Qiagen QIAamp PowerFecal DNA Kit' following the manufactures protocols. The former method eluted 100 µl of Qiagen buffer ATE, whilst the latter eluted 75 µl of ATE. DNA was precipitated using 3M Na-Acetate (pH 5.2) and 100% ethanol, with an incubation at -20°C, followed by two washes with 70% ethanol. DNA was resuspended in 50 µl of ddH$_2$O.

DNA Extraction of Additional Sample Types:

Because we could not establish a cell culture for water vole, due to lack of sample availability, DNA was extracted from waterbuck (*Kobus ellipsiprymnus*) cell culture using a phenol-chloroform extraction protocol adapted from (Green, M. R., Hughes, H., Sambrook, J. and MacCallum, 2012). Cells were washed with phosphate-buffered saline (PBS) before DNA extraction. Proteinase K was added to a final concentration of extraction buffer of 400 µg/ml containing 20 mM Tris-Cl (pH 8.0), 5 mM EDTA (pH 8.0), 400 mM NaCl, and 1% SDS. An appropriate volume of buffer was added to cells and incubated overnight at 56°C with agitation. DNA pellets were resuspended in TE buffer.

Cattle (*Bos taurus*) buccal swabs were collected from three individuals and samples were incubated with agitation overnight at 56°C in an extraction buffer (Ghatak, Muthukumaran and Nachimuthu, 2013) containing 10 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 2% SDS, and 20 mg/ml proteinase K. The previous optimised tissue and hair phenol-chloroform extraction protocol was used. DNA was eluted into 50 µl of ddH$_2$O.

2.2. Amplification and Sequencing of Mitochondrial Control Region:

Amplification of Marker:

Forward and reverse primers for the mitochondrial DNA control region were selected from a previous publication (5’-TTAATCTACCATCCTCCGTGAAACC-3’ and 5’-TKGACACTGGTCTAGGGATATTTGC-3’, respectively; Piertney et al., 2005). All 20 samples were amplified using a PCR reaction mix containing 1x PCR buffer, 200 µM of each dNTP, 0.5 µM forward primer, 0.5 µM reverse primer, and 2.5 units/reaction ‘Qiagen HotStarTaq DNA Polymerase’. Template DNA was then added at a separate workstation (between 9-47 ng/µl). A 15-minute denaturation stage was required at 95°C, followed by 35 cycles of annealing and elongation (94°C for 1 min, 50°C for 1 min, and 72°C for 1 min), ending with 10 mins at 72°C. A negative control and a separate PCR workstation was used for the preparation of the PCR reaction mix to prevent contamination. Amplification was viewed with gel electrophoresis (1% agarose gel, 90 V, ~60 mins, and viewed with Syngene Gel Doc. Samples were purified using ‘Qiagen QIAquick PCR Purification Kit’ following the manufactures protocol. DNA concentrations and purity were measured following PCR clean-up using NanoDrop.

Sequencing:

The amplified DNA was sent for sequencing at DBS Genomics, Durham, UK. Both the forward and reverse strands for each individual water vole were sequenced. We used the package ‘Geneious Prime 2020.1 (https://www.geneious.com)’ to create a consensus sequence from both strands of DNA. This included reverse complementing the reverse strand and subsequently aligning both strands using the global alignment tool, with free gaps and 93% similarity. Consensus sequences were exported as FASTA files.

2.3. Population Genetics of the Water Vole:

Multiple Sequence Alignments (MSA):

We aligned DNA sequences from Wildwood Trust and DNA sequences from selected papers (Piertney *et al.*, 2005; Baker, 2015; Brace *et al.*, 2016) using the 'R' package 'Ape v5.3' (Paradis and Schliep, 2018) and the program 'Clustal W v2.0' (Larkin *et al.*, 2007). MSAs were carried out using the 'clustal()' command in 'Ape v5.3', with default parameters. Alignments were trimmed based on gaps on the borders of the alignment.

Haplotype Networks:

Haplotypes were computed from the MSA results using the 'R' package 'Pegas v0.13' (Paradis, 2010) with the 'haplotype()' command and haplotype networks were computed using the 'haploNet()' command, both with default parameters.

Phylogenetic Tree Construction:

Neighbour Joining (NJ) and Maximum-Likelihood (ML) trees were constructed using the MSA data in 'R' packages 'Phangorn v2.5.5' (Schliep *et al.*, 2017), 'Ape v5.3' (Paradis and Schliep, 2018) and 'ggtree v2.0.2' (Yu, 2020). Both tree types were rooted on the outgroup. For ML trees, the best nucleotide evolution model was chosen using 'modelTest()' and then 'bootstrap.pml()' was used to perform the bootstrap analysis (1000 replicates with optimised topology). Bayesian phylogenetic trees were constructed in the program 'MrBayes v3.2.7' (Huelsenbeck and Ronquist, 2001) and visualised in 'FigTree v1.4.4'. Standard parameters were used except for changing the nucleotide evolution model (computed in 'R') and the number of Markov Chain Monte Carlo (MCMC) generations, which depended on the average standard deviation of split frequencies and whether additional generations were needed if the standard deviation was above 0.01. A consensus tree was produced using the 'sumt' command and rooted on the outgroup taxon.

Population Genetics Calculations:

Nucleotide diversity, haplotype diversity, and Tajima's D were calculated in the 'R' package 'Pegas 0.13' (Paradis, 2010) using the MSA. Mitochondrial control region sequences of *Myodes glareolus* were used as a comparison to *Arvicola amphibius*. Sequences were obtained from Filipi *et al.*, 2015 and Marková *et al.*, 2020 and aligned, before calculating population genetics statistics.

Population Structure:

All 144 sequences and 642 loci were analysed in the program ‘STRUCTURE 2.3.4’ (Pritchard, Stephens and Donnelly, 2000). Parameters for length of burn-in period was set at 10,000 and number of MCMC generations after burn-in was set at 100,000. The admixture model was chosen, and the number of populations assumed (K) was set from K=1 to K=6, with 5 iterations each. ‘Structure Harvester’ (Earl and vonHoldt, 2012) was then used to identify the most appropriate K value using graphs of L(K) and delta K.

2.4. Phylogenetics of Arvicolinae:

Mitochondrial and Nuclear Markers:

All available mitochondrial genomes, mitochondrial cytochrome b (*Cytb*) genes, nuclear growth hormone receptor (*GHR*) genes, and nuclear interphotoreceptor retinoid-binding protein (*IRBP*) genes were found utilizing the ‘National Centre for Biotechnology Information (NCBI) Nucleotide’ and ‘NCBI BLAST’ databases for available Arvicolinae species. Outgroup taxa were chosen from three subfamilies of the Cricetidae family: Cricetinae (*Cricetulus griseus*), Neotominae (*Peromyscus polionotus*), and Sigmodontinae (*Sigmodon hispidus*). One additional outgroup taxon was chosen from a Muroidea family Muridae (*Mus musculus*) and was used to root the phylogenetic trees.

Phylogenetic Tree Construction:

ML and Bayesian inference phylogenetic trees were constructed using the previously described methods. Trees were constructed for each of the three markers, as well as for the available mitochondrial genomes. A supermatrix approach was used for the three markers. Individual markers were firstly aligned in ‘R’ using ‘Clustal W’ and trimmed. FASTA files of the alignments were concatenated in ‘MEGA X’. Concatenated sequences were then read into ‘R’ as a FASTA file and ML (in ‘R’) or Bayesian inference phylogenetic trees (in ‘MrBayes’) were constructed using the concatenated sequence.

3. Results:

3.1. Optimisation of DNA Extraction:

DNA Extraction of Tissue Samples:

The DNA extraction for tissue samples was optimised using fragmented water vole tail tissue. Firstly, the lysis buffer was optimised by comparing three buffers, containing different concentrations of NaCl, EDTA, and Tris-Cl, using the published phenol-chloroform extraction protocol. Buffer 3 had the highest DNA concentration (172.27 ng/µl) and highest DNA yield (17227 ng), compared to buffers 1 and 2 **(Table 2 'Optimisation of Buffer')**. The two absorbance ratios, 260/280 and 260/230, were 1.51 and 1.56, respectively. The 260/280 absorbance ratios test for the presence of protein and other contaminants in the DNA sample and are expected to be ~1.8 for 260/280 and ~2.0 for 260/230. This indicates a "pure" DNA sample. Buffer 3 was chosen for the following DNA extractions of tissue samples.

Next, three extraction protocols were tested to determine the highest DNA yield. Phenol-chloroform extraction had the highest DNA yield at 17,227 ng, with ethanol extraction and isopropanol extraction having much lower DNA yields **(Table 2: 'Optimisation of Protocol')**. Phenol-chloroform extraction also had a purer DNA sample after DNA extraction. The phenol-chloroform extraction protocol was optimised further by considering the precipitation of DNA, either isopropanol as the control (Phenol-Chloroform 1) or ethanol in 'Phenol-Chloroform 2' **(Table 2: 'Further Optimisation')**. Or the centrifugation of isopropanol to pellet the DNA which was 15 mins in the control (Phenol-Chloroform 1) or 30 mins in 'Phenol-Chloroform 3'. The added 15 mins to the isopropanol centrifugation increased DNA yield (20,666 ng) compared with the control (3,088 ng).

Table 2: Optimisation of DNA extraction lysis buffers and protocols for water vole tissue samples.

Optimisation of Buffer:	**Elution Volume (µl)**	**ng/µl**	**ng**	**260/280**	**260/230**
Buffer 1	100	89.11	8911	1.43	1.47
Buffer 2	100	102.18	10218	1.39	1.54
Buffer 3	100	172.27	17227	1.51	1.56
Optimisation of Protocol:					
Phenol-Chloroform	100	172.27	17227	1.51	1.56
Ethanol	100	37.01	3701	1.42	0.41
Isopropanol	100	32.45	3245	1.51	0.36
Further Optimisation:					
Phenol-Chloroform (1)	200	15.44	3088	1.71	1.42
Phenol-Chloroform (2)	200	45.10	9020	1.6	1.37
Phenol-Chloroform (3)	200	103.33	20666	1.57	1.41

An optimised protocol consisting of the selected buffer, phenol-chloroform extraction, and the added 15 mins of centrifugation with isopropanol was used for the DNA extraction of all tissue samples **(Table 3)**. Gel electrophoresis was used to determine size of the DNA fragments **(Figure 5)**. In all four tissue samples high molecular weight (HMW) bands were seen at around 10 kb. Some samples contained more fragmented DNA than others, as shown by the DNA streak from high to low molecular weight.

Table 3: DNA extraction of all water vole tissue samples. Results for samples 1 and 2 from the previous table. Samples 3 and 4 using the final optimised protocol ('Phenol-Chloroform 3').

Sample	**Elution Volume (µl)**	**ng/µl**	**ng**	**260/280**	**260/230**
1	100	172.27	17227	1.51	1.56
2	200	103.33	20666	1.57	1.41
3	100	30.68	3068	1.59	1.49
4	100	43.00	4300	1.53	1.61

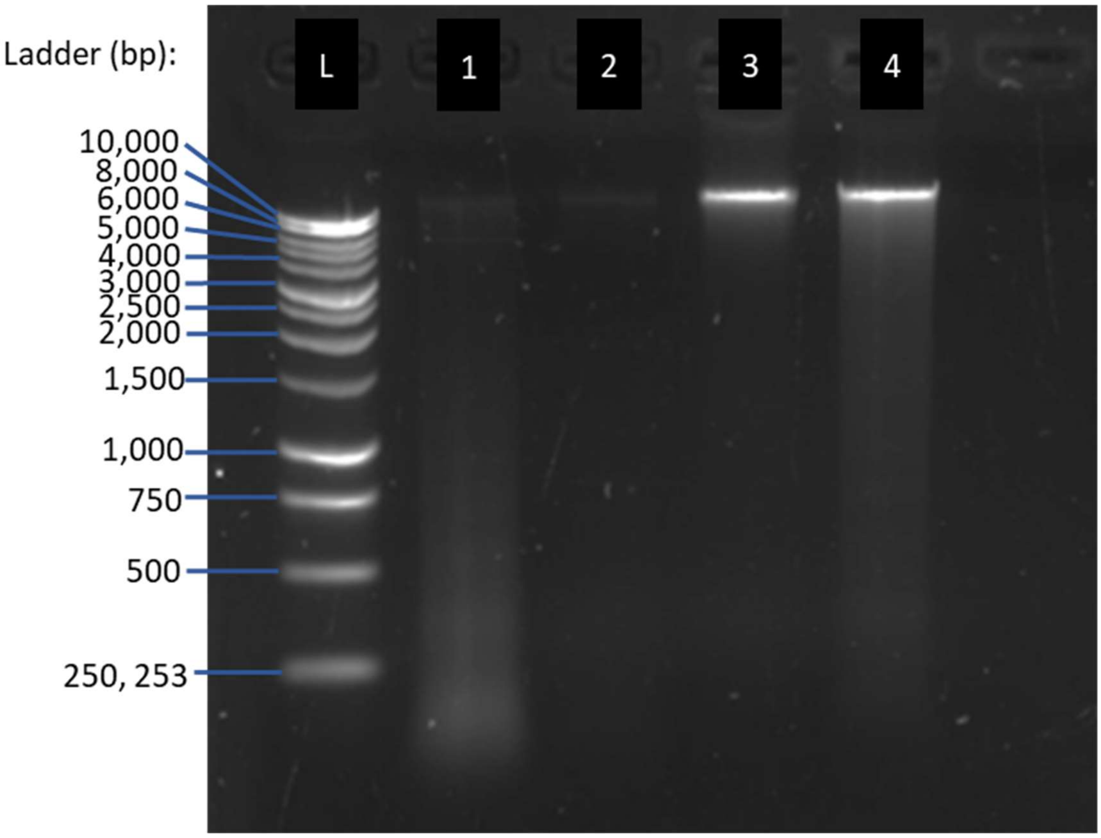


Figure 5: Gel electrophoresis of all water vole tail tissue samples. Lane 'L' shows the 10kb ladder, and lanes 1-4 are the corresponding DNA samples. 1% agarose, 70V, ~60 mins.

DNA Extraction of Hair Samples:

Two hair extraction protocols were compared, the 'Qiagen DNeasy Blood and Tissue DNA Extraction Kit' and the phenol-chloroform extraction, for DNA yield and purity of the DNA sample. Both protocols had low DNA yield and poor absorbance ratios (i.e., not similar to the expected values) **(Table 4)**. A new buffer was used that contained $CaCl_2$ and DTT but without EDTA. This was tested with different elution volumes of ddH_2O. It was found that using this buffer and eluting 50 µl of ddH_2O increased DNA yield (3,794 ng).

Table 4: Optimisation of DNA extraction protocols for hair samples.

	Elution Volume (µl)	ng/µl	ng	260/280	260/230
Qiagen Kit (1)	100	4.38	438	2.41	0.97
Qiagen Kit (2)	100	2.21	221	7.23	5.18
Phenol-Chloroform (1)	100	4.3	430	1.15	0.53
Phenol-Chloroform (2)	100	10.3	1030	1.41	1.23
Optimised Phenol-Chloroform (1)	100	17.88	1788	1.24	1.00
Optimised Phenol-Chloroform (2)	75	17.09	1281.75	1.25	1.19
Optimised Phenol-Chloroform (3)	50	75.88	3794	1.34	1.46

DNA was extracted from the remaining hair samples using the new lysis buffer and optimised hair extraction protocol, using the previously optimised protocol for tissue samples **(Table 5)**.

Table 5: DNA extraction of hair samples.

Sample	Elution Volume (µl)	ng/µl	ng	260/280	260/230
5	50	10.3	515	1.41	1.23
6	50	75.88	3794	1.34	1.46
7	50	87.56	4378	1.44	1.68
8	50	13.83	691.5	1.51	1.49
9	50	9.04	452	1.39	1.30
10	50	47.02	2351	1.36	1.61

DNA Extraction of Faecal Samples:

The ‘Qiagen QIAamp DNA Stool Mini Kit’ and ‘Qiagen QIAamp PowerFecal DNA Kit’ were tested for DNA yield and purity **(Table 6)**. Both extraction kits produced similar DNA yields (a mean of 1,091.25 ng and 1,136.25 ng, respectively). However, values for 260/230 absorbance ratio were considerably different in the two extraction kits. ‘Qiagen QIAamp DNA Stool Mini Kit’ had values greater than 2.0 (the expected value for pure samples), whilst ‘Qiagen QIAamp PowerFecal DNA Kit’ had values considerably lower. Overall, the latter DNA extraction kit performed better for the samples tested.

Table 6: Comparison of different DNA extraction kits for faecal samples.

DNA Extraction Kit	Elution Volume (µl)	ng/µl	ng	260/280	260/230
Qiagen Stool Mini Kit (1)	100	9.89	989	1.53	2.42
Qiagen Stool Mini Kit (2)	100	11.78	1178	1.79	2.17
Qiagen Stool Mini Kit (3)	100	12.24	1224	1.59	2.73
Qiagen Stool Mini Kit (4)	100	9.74	974	1.54	2.02
Qiagen Power Faecal Kit (1)	75	10.81	810.75	2.43	0.44
Qiagen Power Faecal Kit (2)	75	19.49	1461.75	1.86	0.17

DNA was extracted from all 10 faecal samples from Wildwood Trust using one or both extraction kits **(Table 7)**. All samples were precipitated with ethanol to increase DNA yield and improve the purity of the DNA sample. This was the case for the majority of samples, where DNA yield was higher and absorbance values were closer to the expected values following precipitation.

Table 7: DNA concentrations and purity of extracted faecal samples using 'Qiagen QIAamp DNA Stool Mini Kit' (A) and 'Qiagen QIAamp PowerFecal DNA Kit' (B) before and after ethanol precipitation.

	Before Precipitation					**After Precipitation**				
	E.V. (µl)	**ng/µl**	**ng**	**260/ 280**	**260/ 230**	**E.V. (µl)**	**ng/µl**	**ng**	**260/ 280**	**260/ 230**
Q58 (A)	100	6.16	616	2.60	1.17	50	26.78	1339	1.60	1.74
Q75A.2 (A)	100	4.25	425	1.63	0.99	50	11.41	570.5	1.47	1.39
R2 (A)	100	12.03	1203	1.54	1.69	50	38.13	1906.5	1.64	1.87
R51 (A)	100	5.92	592	1.52	1.31	50	16.37	818.5	1.87	1.48
Q42.2 (A)	100	4.05	405	1.65	1.52	50	9.14	457	1.91	1.06
R50 (A)	100	6.05	605	2.05	1.24	50	23.29	1164.5	1.55	1.76
R27 (A)	100	9.72	972	1.78	1.67	50	28.96	1448	1.69	1.65
Q52 (A)	100	5.32	532	1.94	1.23	50	17.44	872	1.39	1.68
Q100 (A)	100	10.68	1068	1.89	1.17	50	27.9	1395	1.53	1.57
Q7 (A)	100	5.21	521	1.10	0.88	50	13.93	696.5	1.89	1.46
R51 (B)	75	10.81	810.75	2.43	0.44	50	14.69	734.5	1.55	1.91
R27 (B)	75	19.49	1461.75	1.86	0.17	50	24.11	1205.5	1.52	2.11

DNA Extraction of Additional Sample Types:

DNA extraction from cell culture was used to demonstrate the benchmark method of extracting large quantities of DNA and pure DNA samples. For this, we used an available waterbuck cell culture and extracted DNA from the flask. This resulted in high DNA yield with a mean of 46,566.67 ng **(Table 8)**. Sample purity was lower than expected for all samples and both absorbance ratios.

Table 8: Extraction of DNA from waterbuck cell culture.

Sample	Elution Volume (µl)	ng/µl	ng	260/280	260/230
1	500	53.20	26600	1.39	1.65
2	500	145.00	72500	1.49	1.71
3	500	81.00	40500	1.44	1.69
4	500	46.20	23100	1.45	1.73
5	500	218.20	109100	1.55	1.76
6	500	15.20	7600	1.55	1.90

DNA was also extracted from cattle buccal swabs from three individuals. This was to demonstrate another useful way of sampling animals. We achieved a high DNA yield for all samples, with samples 2 and 3 having around three times as much DNA **(Table 9)**. This may be due to the amount of saliva that was collected from each cattle. On average this resulted in 11,937.17 ng of DNA.

Table 9: Extraction of DNA from cattle buccal swabs.

Sample	Elution Volume (µl)	ng/µl	ng	260/280	260/230
1	50	106.60	5330	1.72	2.11
2	50	302.59	15129.5	1.75	2.03
3	50	307.04	15352	1.48	1.64

Comparison of Sample Types:

All sample types were compared for DNA yield **(Figure 6)**. As expected, DNA extraction from cell culture resulted in the highest median DNA yield (33,550 ng). Tail tissue and buccal swabs both yielded the next highest DNA yield (10,763.5 ng and 15,129.5 ng, respectively). Hair and faecal samples produced the lowest DNA yield (1,521.25 ng and 1,018.25 ng, respectively). An Analysis of Variance (ANOVA) test found a statistically significant difference in average DNA yield by sample type (f(4) = 7.223, p < 0.001).

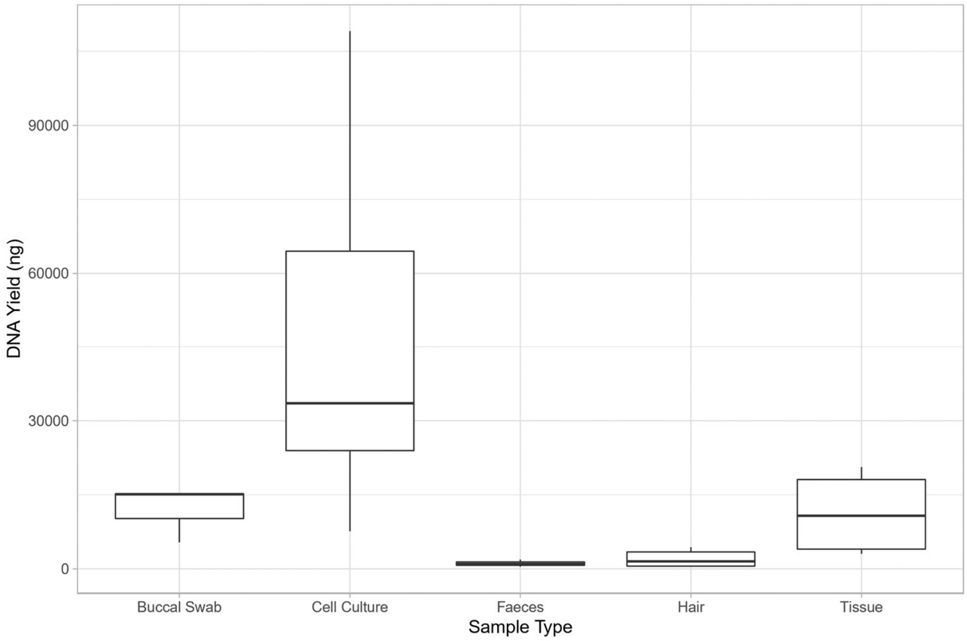


Figure 6: Comparison of the median DNA yields (ng) for each sample type. Calculated from DNA yields from samples using the final optimised protocols.

3.2. Amplification and Sequencing of the Mitochondrial Control Region:

After optimisation of PCR protocols, the mitochondrial DNA control region for all 20 water voles was amplified. Gel electrophoresis was used to identify which samples amplified, as shown by the presence of bands between 750 bp and 1000 bp **(Figure 7)**. Post-PCR, samples were cleaned to remove any PCR products and then sent for Sanger sequencing.

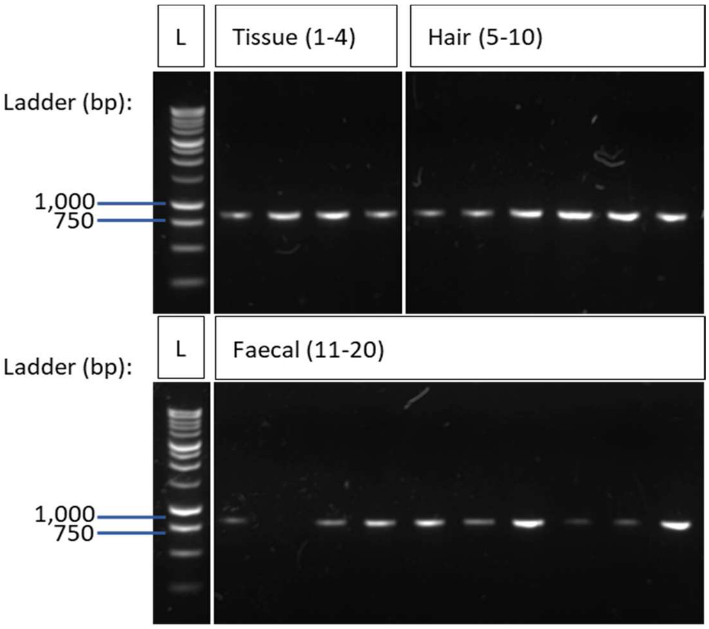


Figure 7: Gel electrophoresis of the amplified mitochondrial DNA control region for 20 water voles. L is the DNA ladder with markings at 750bp and 1000bp. Poorly amplified samples were repeated.

A total of 20 individuals were sequenced, but three had poor chromatogram results. Of these three samples, one was hair sample 5 and two were faecal samples 12 and 18. All three samples also proved difficult to amplify before sequencing, with numerous PCR repeats required to achieve a visible band with gel electrophoresis. These three samples were excluded from further analysis.

Chromatograms of the sequence files showed a region rich with A/T repeats, followed by weak and inhibited peaks, with the sequence unable to be accurately determined. Therefore, the forward and reverse strands were trimmed at this region and aligned to form a consensus sequence. The 17 Wildwood Trust consensus sequences were used for further analysis.

3.3. Population Genetics of the Water Vole:

Individuals from Wildwood Trust:

A haplotype network and ML phylogenetic tree were constructed using the 17 mitochondrial control region DNA sequences from the sampled Wildwood Trust water voles. Out of 17 individuals there were 12 haplotypes **(Figure 8a)** and two main haplogroups. There were six mutational steps between the two haplogroups (i.e. between haplotype 1 and 4). One haplogroup contained haplotypes 1 and 2, and the other contained the 10 remaining haplotypes. Haplotypes 1, 2 and 4 were sampled from tissue and were some of the oldest samples collected from Wildwood Trust. Three haplotypes contained more than one individual. Haplotype 3 contained individuals 3, 6, and 13, haplotype 8 contained individuals 10, 14, and 16, and haplotype 9 contained individuals 11 and 17. Within the largest haplogroup there was a maximum of two mutational steps between each haplotype.

A ML tree was constructed, and this grouped individuals 3, 6, 9, 10, 11, 13, 14, 15, 16, and 17 into the same clade with a polytomy. Therefore, only one sequence from these individuals was used in the following phylogenetic analyses to aid visualisation. The multiple individuals were grouped into a clade with individual 7 **(Figure 8b)** and were the last to diverge. This result conflicted with the haplotype network as these individuals are not grouped into one haplotype, but instead formed haplotypes 3, 7, 8, 9, and 10. Each of the haplotypes had multiple mutational steps between them. Individuals 1 and 2 diverged first in the Wildwood Trust samples and this was supported by haplotypes 1 and 2 forming a separate haplogroup to all other individuals and a 100% bootstrap score for this specific node.

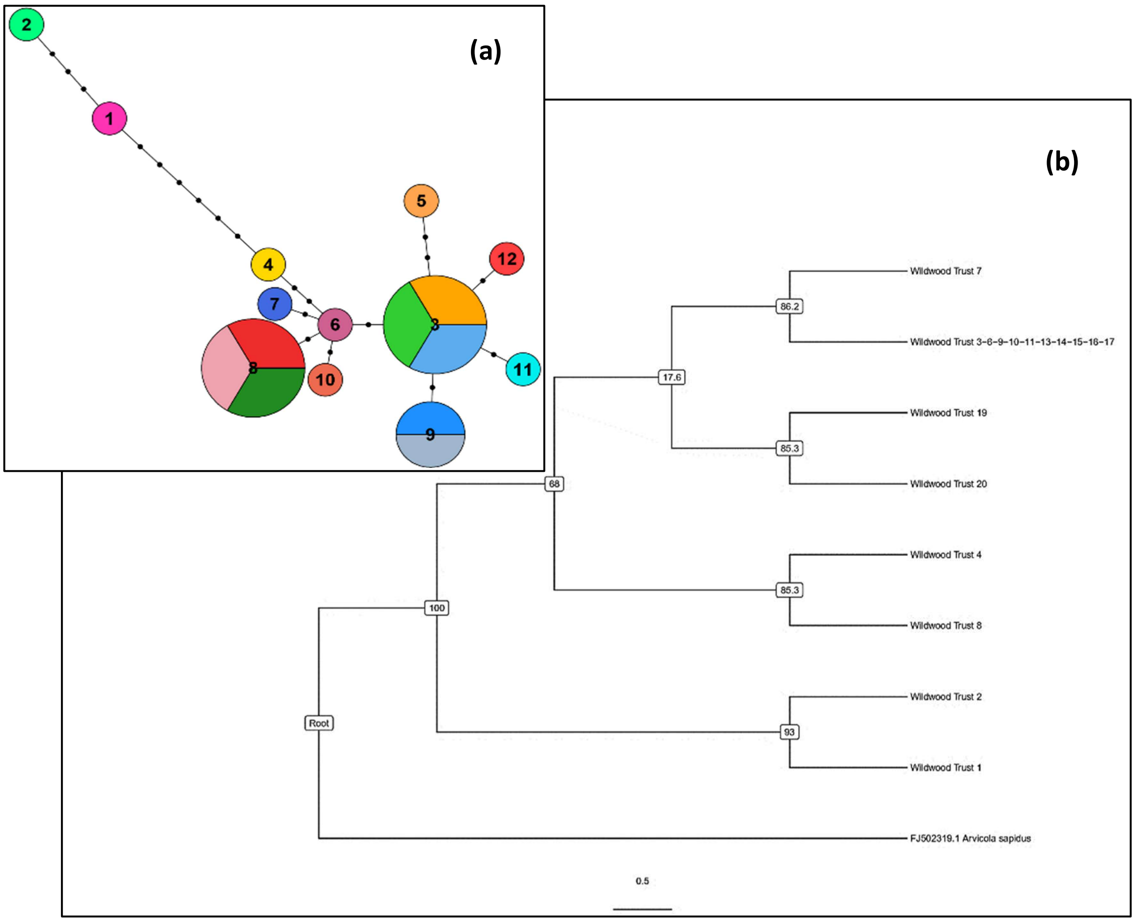


Figure 8**: Analysis of Wildwood Trust water voles using the mitochondrial control region.** (a) Haplotype network of Wildwood Trust samples. Each pie chart represents a haplotype, and each colour represents a different sample. The dotted lines are the number of mutational steps between haplotype sequences. (b) ML phylogenetic tree of Wildwood Trust samples. Bootstrap scores are shown for each node. The tree is rooted on *Arvicola sapidus*.

Comparison of Captive and Natural Populations:

Captive Wildwood Trust sequences, as well as haplotype sequences from Baker, 2015 from natural water vole populations in the South East and East of England were aligned and a haplotype network was constructed. There were 26 haplotypes in total, 12 haplotypes from the Wildwood Trust samples and 15 haplotypes from Baker, 2015 **(Figure 9)**. Haplotypes formed two haplogroups. One contained only one haplotype, the South East of England haplotype 14, which was 17 mutational steps from the other haplogroup, where the remaining samples were clustered. Captive individuals were found in separate haplotypes to natural individuals. Haplotypes 1 and 2, containing captive individuals, were closely clustered with haplotype 25, which contained natural water voles from the East of England. These haplotypes were five mutational steps from other haplotypes (i.e. between haplotype 1 and haplotype 3).

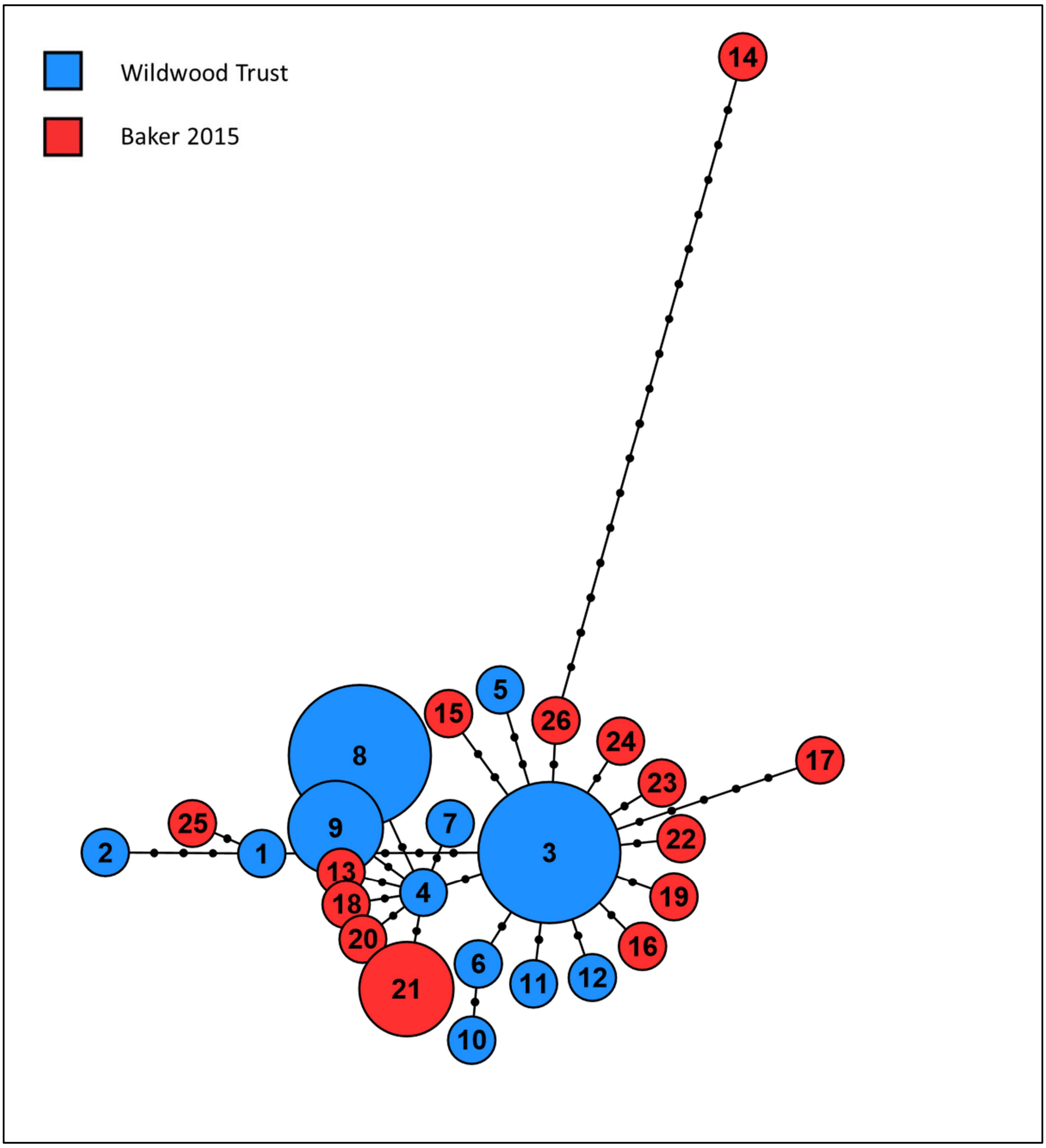


Figure 9: Captive and wild water vole samples from the South East and East of England. Wildwood Trust sequences are individual sequences, whilst Baker 2015 sequences are haplotype sequences.

The captive Wildwood Trust sequences were compared with sequences from the (Piertney *et al.*, 2005) paper which sampled water voles across the British isle from natural populations. A haplotype network was constructed using the alignment of DNA sequences for 17 Wildwood Trust individuals and 57 sequences for natural water voles in Britain and five from mainland Europe **(Figure 10)**.

A total of 45 haplotypes were found for 78 individuals. There were two main haplogroups, one containing individuals from Scotland and mainland Europe, and the other containing individuals from England, Wales, and mainland Europe. This latter haplogroup contained all Wildwood Trust individuals. There were 13 mutational steps between haplotype 14 from the Scotland haplogroup and haplotype 45 from the England/Wales haplogroup. All Wildwood

Trust individuals formed separate haplotypes to the natural individuals and formed the same haplotypes as the previous haplotype network **(Figure 8a)**. In both haplogroups there were multiple haplotypes with more than one individual. Haplotypes 42-45 were individuals sampled from mainland Europe. Mainland Europe samples were found in the Scotland haplogroup, except for haplotype 45 which contained a water vole from Finland and was found in the England/Wales haplogroup.

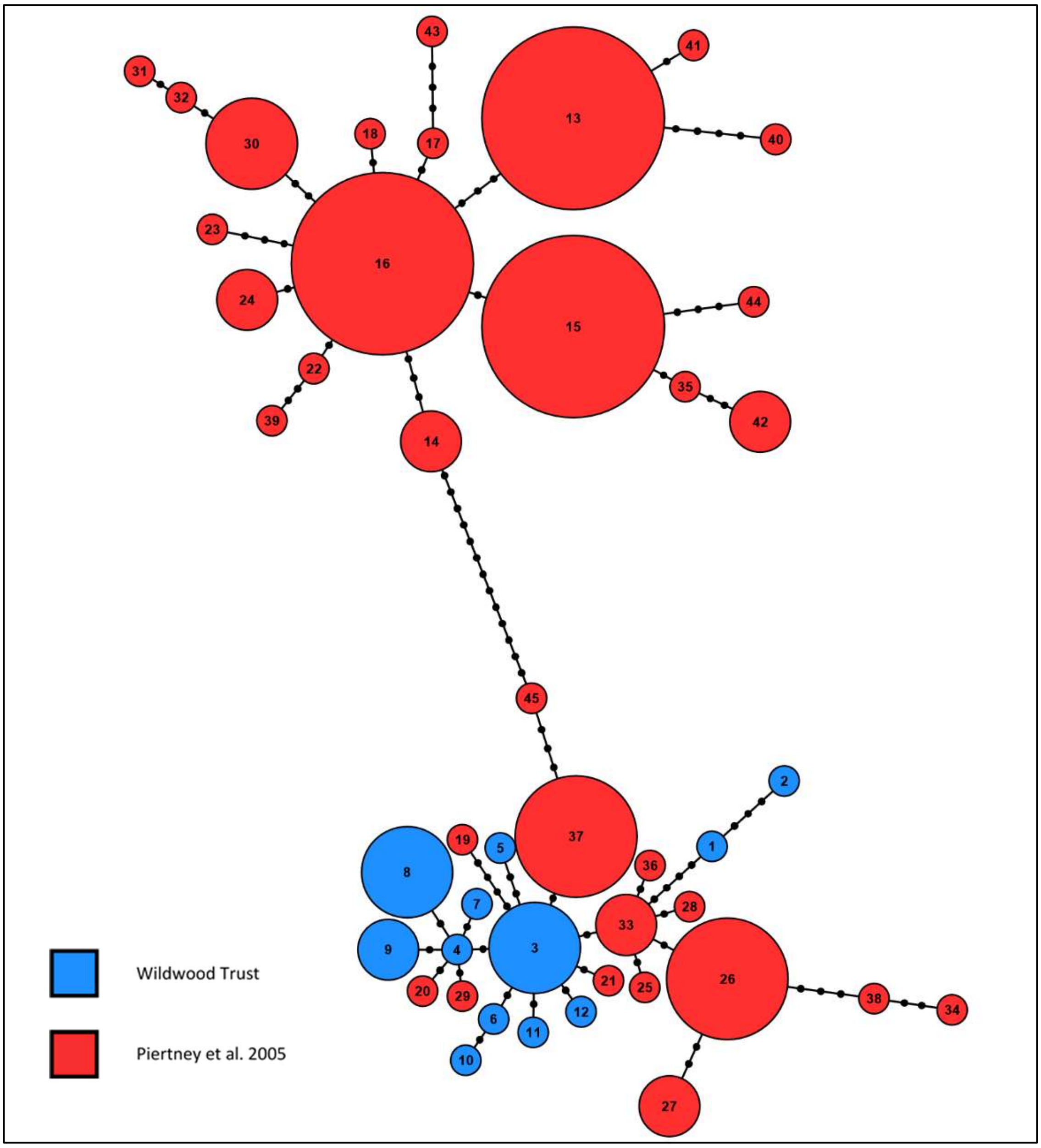


Figure 10: Haplotype network using DNA sequences of the mitochondrial control region for captive and natural water voles. Each pie chart represents a haplotype, and the size represents the number of individuals with the haplotype. The dotted lines are the number of mutational steps between haplotypes.

Wildwood Trust sequences were also aligned with sequences from Brace *et al.*, 2016, a paper which used ancient DNA from museum specimens dating back to the Pleistocene (before the last glacial period). Phylogenetic trees, both ML **(Figure 11)** and Bayesian inference, were constructed using an alignment containing sequences from Wildwood Trust, Piertney *et al.*, 2005, Baker, 2015, and Brace *et al.*, 2016. Both phylogenetic trees showed comparable results. All Holocene individuals, most modern English/Welsh water voles, all South East and East of England water voles, and all Wildwood Trust individuals were grouped into one clade (shaded in dark grey). Wildwood Trust and 14 out of 15 of the South East and East of England water voles were grouped more closely, whilst South East of England 11 was grouped with modern samples from Somerset, Wales, Staffordshire, and Shropshire. The sister clade contained all Pleistocene and Scottish water voles, and three modern English samples (shaded in a lighter grey). Samples from mainland Europe were found in both clades. Water vole from Italy and Switzerland were grouped into a separate, early diverging clade, after the outgroup species (*Arvicola sapidus*). *Arvicola scherman* samples were found in both the English/Welsh clade and the Scottish clade. The phylogenetic tree had relatively high bootstrap scores at nodes separating the major three clades. Polytomies were seen at nodes within the major clades, so relationships between individuals is less clear.

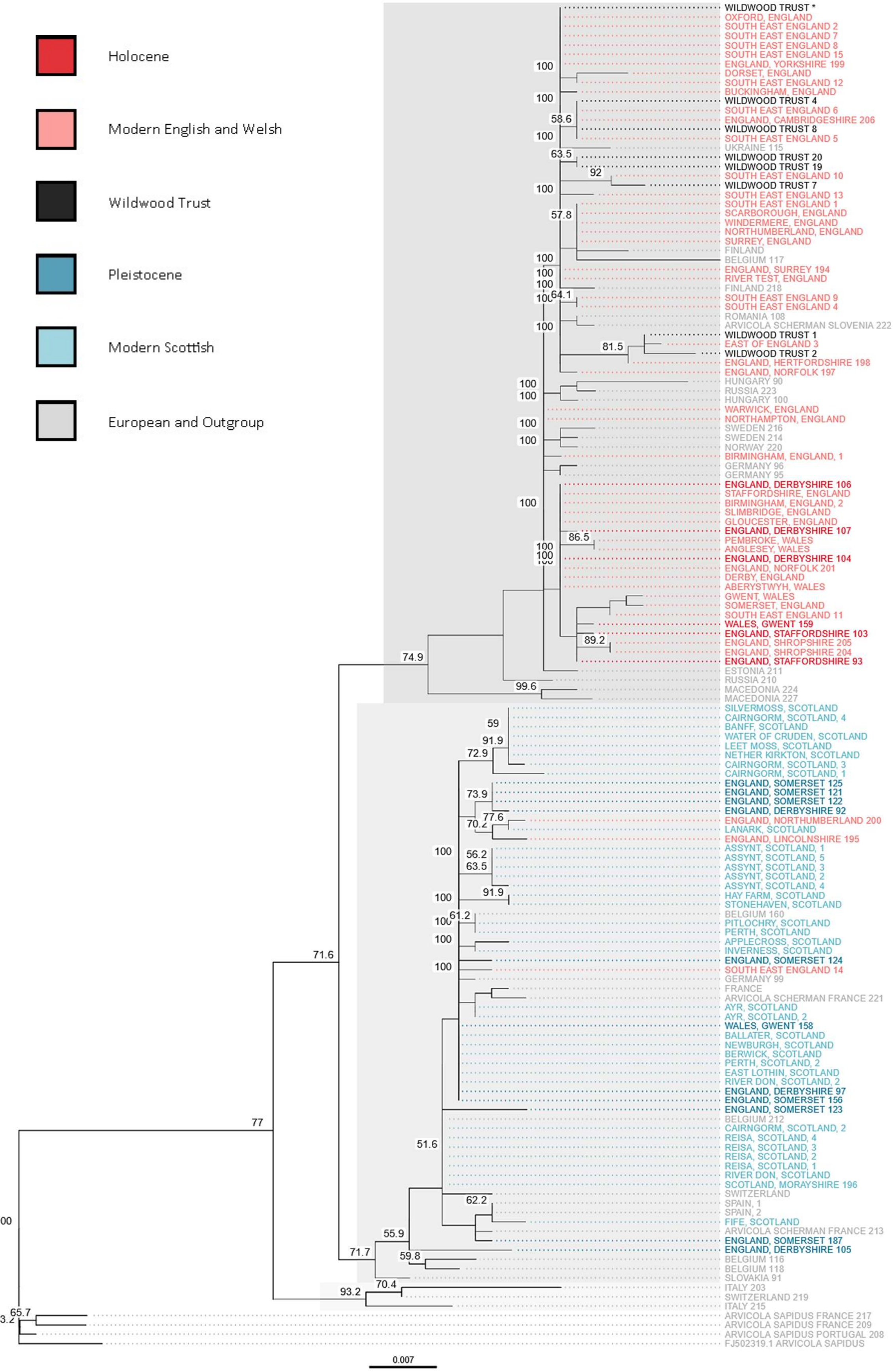


Figure 11: ML phylogenetic tree containing all samples. Taxon colours represent the age or location of sample. The tree is rooted on *Arvicola sapidus*. Taxon "Wildwood Trust*" contains multiple Wildwood Trust samples. Only bootstrap scores greater than 50% are shown.

Population genetics calculations were computed for each of the alignments **(Table 10)**. Wildwood Trust water voles had lower haplotype and nucleotide diversity (0.949 and 0.004, respectively) compared with natural British water voles (0.971 and 0.016, respectively). Tajima's D was lower in Wildwood Trust individuals (-2.186 compared with -0.113). Natural English and Welsh water voles had higher haplotype and nucleotide diversity (0.982 and 0.008, respectively), and lower Tajima's D (-2.164) than natural Scottish water voles which had a haplotype diversity of 0.945, a nucleotide diversity of 0.009, and Tajima's D of -1.857. The natural Scottish population had a lower haplotype diversity than the captive water vole population (Wildwood Trust).

To put British water vole genetic diversity into perspective, we compared population genetics estimates with a sister species, the bank vole (*Myodes glareolus*) which had stable numbers in Britain. Available mitochondrial control region sequences for the bank vole were aligned and population genetics calculations were computed. Focusing on British populations, bank voles had lower haplotype and nucleotide diversity (0.967 and 0.006, respectively) than natural British populations (0.971 and 0.016, respectively). British bank voles had lower haplotype and nucleotide diversity than all European bank vole samples.

Table 10: European water vole (*Arvicola amphibius*) and bank vole (*Myodes glareolus*) population genetics. Number of sequences (n), alignment length in base pairs (bp), number of haplotypes (Hap. No.), haplotype diversity (Hap. Div.), variance (Var.), nucleotide diversity (π), Tajiam's D (D), P-value for a normal distribution (P-value Norm.), and P-value for a beta distribution (P-value Beta). *Arvicola amphibius* sequences from: (1) Baker, 2015, (2) Piertney et al., 2005, and (3) only modern samples from Brace et al., 2016. *Myodes glareolus* sequences were from Filipi *et al.*, 2015 and Marková *et al.*, 2020.

Arvicola amphibius	**n**	**bp**	**Hap. No.**	**Hap. Div.**	**Var.**	**π**	**Var.**	**D**	**P-value Norm.**	**P-value Beta**
Captive Wildwood Trust	17	706	12	0.949	0.001	0.004	0.000	-2.186	0.029	0.007
Natural South East / East of England Haplotypes (1)	15	731	14	N/A	N/A	0.007	0.000	-2.378	0.017	0.000
Natural English and Welsh (2) & (3)	32	644	32	0.982	0.000	0.008	0.000	-2.164	0.030	0.011
Natural Scottish (2) & (3)	25	644	25	0.945	0.000	0.009	0.000	-1.857	0.063	0.041
Natural British (2) & (3)	67	639	39	0.971	0.000	0.016	0.000	-0.113	0.910	0.950
Natural Mainland European (2) & (3)	20	639	19	0.995	0.000	0.024	0.000	-0.753	0.452	0.491
Myodes glareolus										
British	24	940	17	0.967	0.000	0.006	0.000	-1.115	0.265	0.281
All	118	940	97	0.996	0.000	0.009	0.000	-1.775	0.076	0.050

Lastly, we analysed the population structure of all water vole mitochondrial DNA control region sequences **(Figure 12)** to distinguish the groupings of individuals into populations. The sequences were grouped into two populations (K=2), based on the highest delta K value. All modern English/Welsh and Holocene samples were grouped into one population (red) and all modern Scottish and Pleistocene samples were grouped into another population (green). This is comparable with the ML phylogenetic tree, with two distinct clades for British water voles. Some admixture was seen between the two populations, but only in samples from other European countries.

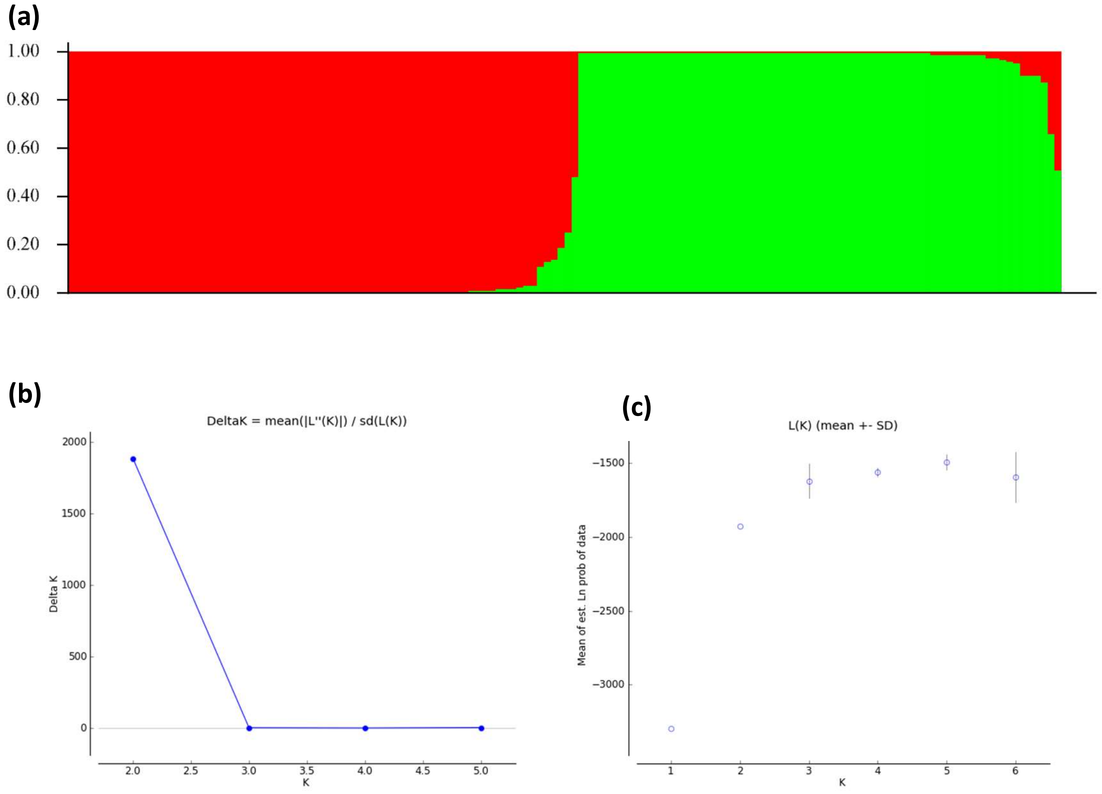


Figure 12: Structure analysis of mtDNA for all sequences (K=2). (a) Run 3/5 for K=2. Red and green subpopulations. (b) Delta K: K=1 to K=6 with 5 runs each. K=2 selected. (c) L(K): K=1 to K=6 with 5 runs each.

3.4. Arvicolinae Phylogenetics:

Using available mitochondrial and nuclear sequences from 'NCBI' for Arvicolinae taxa, we constructed ML and Bayesian inference phylogenetic trees. Several approaches were taken using different genetic markers. (1) whole mitochondrial genomes, (2) individual molecular markers (*Cytb*, *GHR*, and *IRBP*), and (3) concatenated alignment of all three markers.

Mitochondrial Genomes:

All available mitochondrial genomes from 'NCBI' were obtained and aligned, producing a 16,689bp multiple sequence alignment with 7,678 sites with at least one substitution and 34 taxa (30 Arvicolinae taxa). ML and Bayesian inference trees were constructed using the Generalised Time Reversible (GTR) + G + I substitution model, based on the model with the lowest Akaike Information Criterion (AIC) value. Both approaches resulted in comparable topologies. The maximum-likelihood phylogenetic tree had high bootstrap scores for most nodes, with a few nodes having bootstrap scores below 50% **(Figure 13)**. Four outgroup species were used as controls and all diverged before Arvicolinae taxa, with the tree rooted on *Mus musculus*. *Ondatra zibethicus* and *Dicrostonyx* species diverged first within the available Arvicolinae taxa and formed a monophyletic group. *Eothenomys* and *Myodes* species diverged next, forming a separate clade. *Arvicola amphibius* diverged after supported by 100% bootstrap score, followed by *Prodromys liangshanensis* with a bootstrap score below 50%. *Microtus fortis* and *Microtus kikuchii* formed a monophyletic group with *Lasiopodomys* and *Neodon* species, whilst the remaining *Microtus* taxa formed a clade with *Terricola subterraneous*. The Bayesian inference tree had posterior probabilities for all nodes of 1.00/100% **(Appendix A1)**.

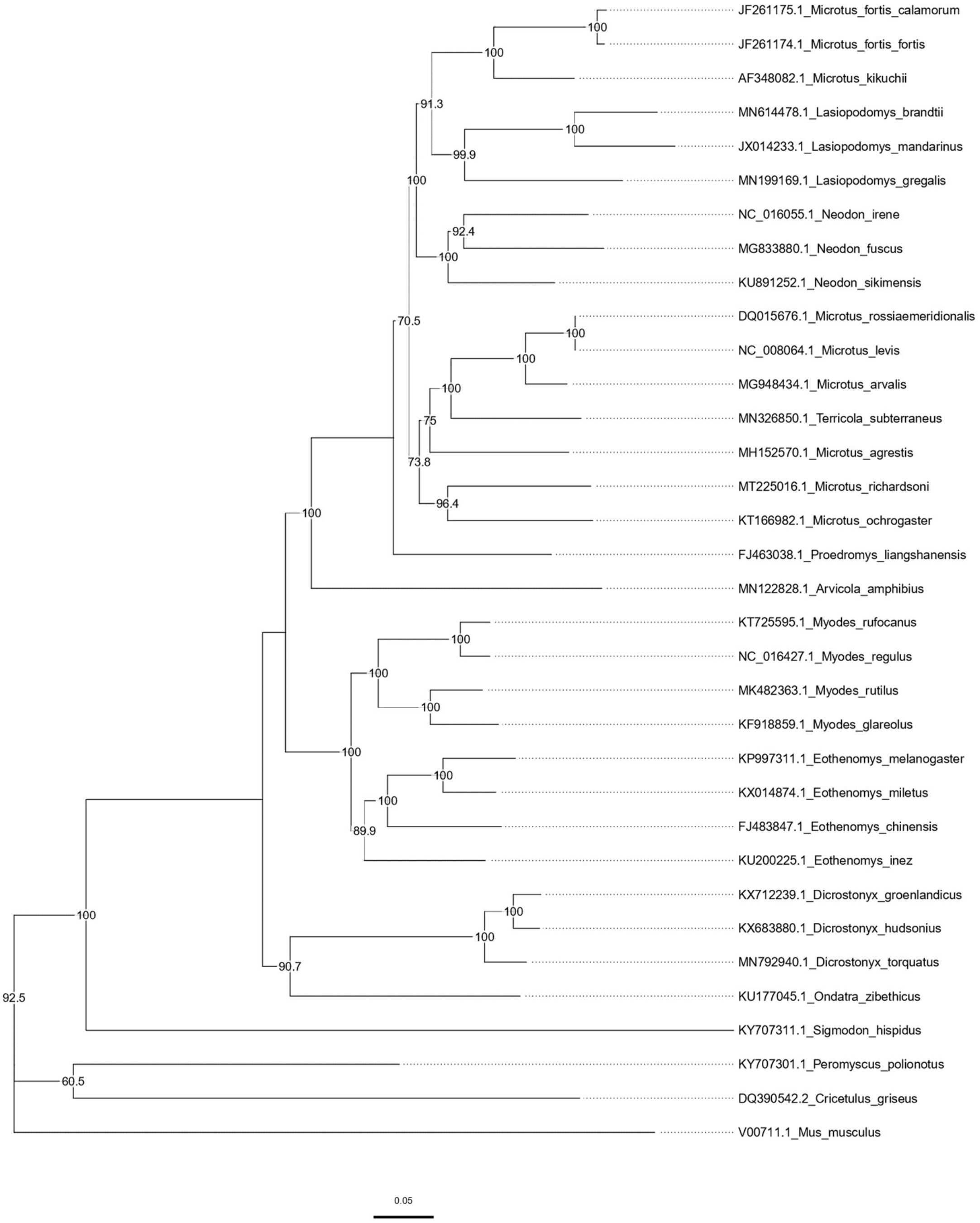

JF261175.1_Microtus_fortis_calamorum
JF261174.1_Microtus_fortis_fortis
AF348082.1_Microtus_kikuchii
MN614478.1_Lasiopodomys_brandtii
JX014233.1_Lasiopodomys_mandarinus
MN199169.1_Lasiopodomys_gregalis
NC_016055.1_Neodon_irene
MG833880.1_Neodon_fuscus
KU891252.1_Neodon_sikimensis
DQ015676.1_Microtus_rossiaemeridionalis
NC_008064.1_Microtus_levis
MG948434.1_Microtus_arvalis
MN326850.1_Terricola_subterraneus
MH152570.1_Microtus_agrestis
MT225016.1_Microtus_richardsoni
KT166982.1_Microtus_ochrogaster
FJ463038.1_Proedromys_liangshanensis
MN122828.1_Arvicola_amphibius
KT725595.1_Myodes_rufocanus
NC_016427.1_Myodes_regulus
MK482363.1_Myodes_rutilus
KF918859.1_Myodes_glareolus
KP997311.1_Eothenomys_melanogaster
KX014874.1_Eothenomys_miletus
FJ483847.1_Eothenomys_chinensis
KU200225.1_Eothenomys_inez
KX712239.1_Dicrostonyx_groenlandicus
KX683880.1_Dicrostonyx_hudsonius
MN792940.1_Dicrostonyx_torquatus
KU177045.1_Ondatra_zibethicus
KY707311.1_Sigmodon_hispidus
KY707301.1_Peromyscus_polionotus
DQ390542.2_Cricetulus_griseus
V00711.1_Mus_musculus
100
100
91.3
100
99.9
100
92.4
100
70.5
100
100
100
75
73.8
96.4
100
100
100
100
100
100
100
89.9
100
100
90.7
100
92.5
60.5
0.05

Figure 13: Maximum-likelihood phylogenetic tree of Arvicolinae mitochondrial genomes. Bootstrap scores above 50% are shown.

Mitochondrial Marker:

A phylogenetic analysis of the mitochondrial *Cytb* was applied as a comparison to the mitogenome analysis. Substitution model GTR + G + I was used to construct a ML **(Figures 14 and A2)** and Bayesian inference **(Figure A3)** phylogenetic tree was constructed. Substantial differences were seen between the two mitochondrial analyses **(Figure 14)**. The position of *Arvicola amphibius* changed when using only the *Cytb* marker and formed a monophyletic group with *Dicrostonyx* species. *Lasipodomys* spp., *Proedromys liangshanensis*, and *Microtus kikuchii* formed a monophyletic group. *Microtus agrestis* formed a monophyletic group with *Neodon* species. *Microtus fortis* was grouped with other *Microtus* species, not *Microtus kikuchii*. Similarities were the grouping of *Eothenomys spp.* and *Myodes spp.*, *Ondatra zibethicus* being the first taxa to diverge, and *Terricola subterraneous* forming a monophyletic group with selected *Microtus* species.

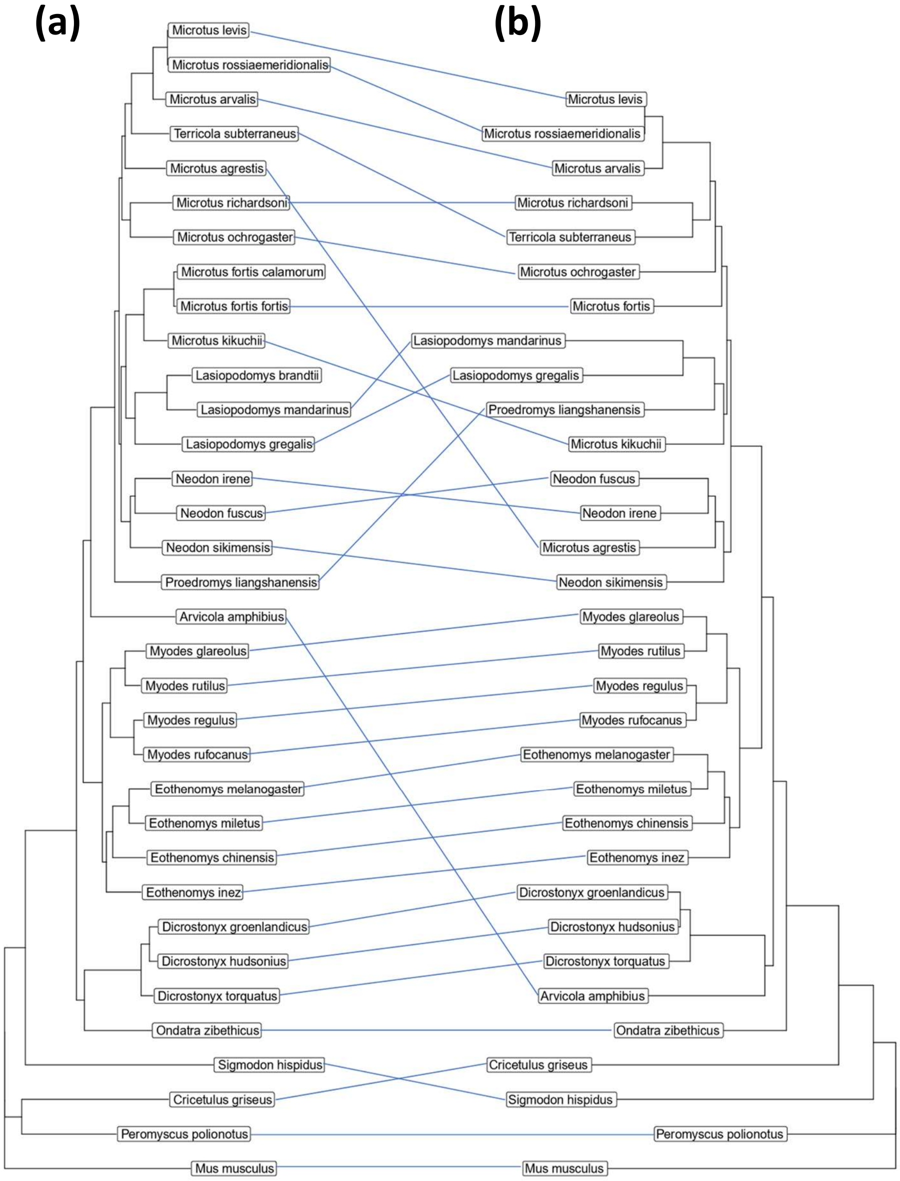


Figure 14: Maximum-likelihood trees of (a) mitogenomes and (b) mitochondrial *Cytb* sequences. Rooted on *Mus musculus*. Bootstrap scores for *Cytb* ML phylogenetic tree can be found in the Appendix.

All available *Cytb* sequences were also used to construct phylogenetic trees. This allowed for a more extensive taxon sampling, with more species for the genus *Microtus* and more genera included. The *Cytb* alignment contained 147 sequences with 1,143 sites. The GTR + G + I substitution model was used to construct ML and Bayesian inference phylogenetic trees **(Figures A4 and A5)**. A significant finding was the position of *Cricetulus migratorius*, an outgroup taxon, which was found in a clade with *Eolagurus* and *Lagurus*, suggesting the unreliability of this tree. *Dicrostonyx* diverged first within *Arvicolinae*, but with a bootstrap score of less than 50%. This was followed by *Ondatra* and *Neofiber*. Other notable findings that were not seen in the previous *Cytb* tree, with only selected taxa that had mitogenomes sequenced, was *Arvicola* grouped in a clade with *Dinaromys* and *Lemmiscus*, as well as the paraphyly of *Volemys*. Overall, there was very low bootstrap scores (less than 50%) for many of the major nodes, therefore this showed a lack support for using the *Cytb* gene alone to determine the phylogeny of this subfamily, even with extensive taxon sampling.

Nuclear Markers:

Nuclear sequences for the *GHR* and *IRBP* genes were obtained and sequences for each were aligned. The *GHR* alignment contained 74 sequences with 725 sites. The GTR + G + I substitution model was used to construct ML **(Figures 15a and A6)** and Bayesian inference phylogenetic trees **(Figure A7)**. The *IRBP* alignment contained 84 sequences with 1,198 sites (482 of which had at least one substitution). The HKY + G + I substitution model was used, based on the lowest AIC value, and a ML **(Figures 15b and A8)** and a Bayesian inference **(Figure A9)** phylogenetic tree was constructed.

There was little support for the tree produced using the *GHR* marker because there were high levels of polytomy and low bootstrap scores at many nodes. Low bootstrap scores (below 50%) were also seen for many nodes in the *IRBP* phylogenetic tree, but less polytomy than when using the *GHR* gene. Similarities included the grouping of the genera *Alticola*, *Eothenomys*, and *Myodes*, the grouping of *Arborimus*, *Phenacomys*, and *Dicrostonyx*, *Arvicola* branching separately, and the paraphyly of *Microtus*. Differences were *Prometheomys schaposchnikovi* diverged first out of the Arvicolinae taxa in the *GHR* tree, but support for this node was below 50%, whereas in the IRBP tree *Phenacomys*, *Arborinus*, and *Dicrostonyx* genera were the first to diverge. *Microtus* was paraphyletic with genera *Terricola* and *Blanfordimys* in the *GHR* tree, but in the *IRBP* tree *Microtus* was paraphyletic with genera *Terricola*, *Lasiopodomys*, *Proedromys*, and *Alexandromys*.

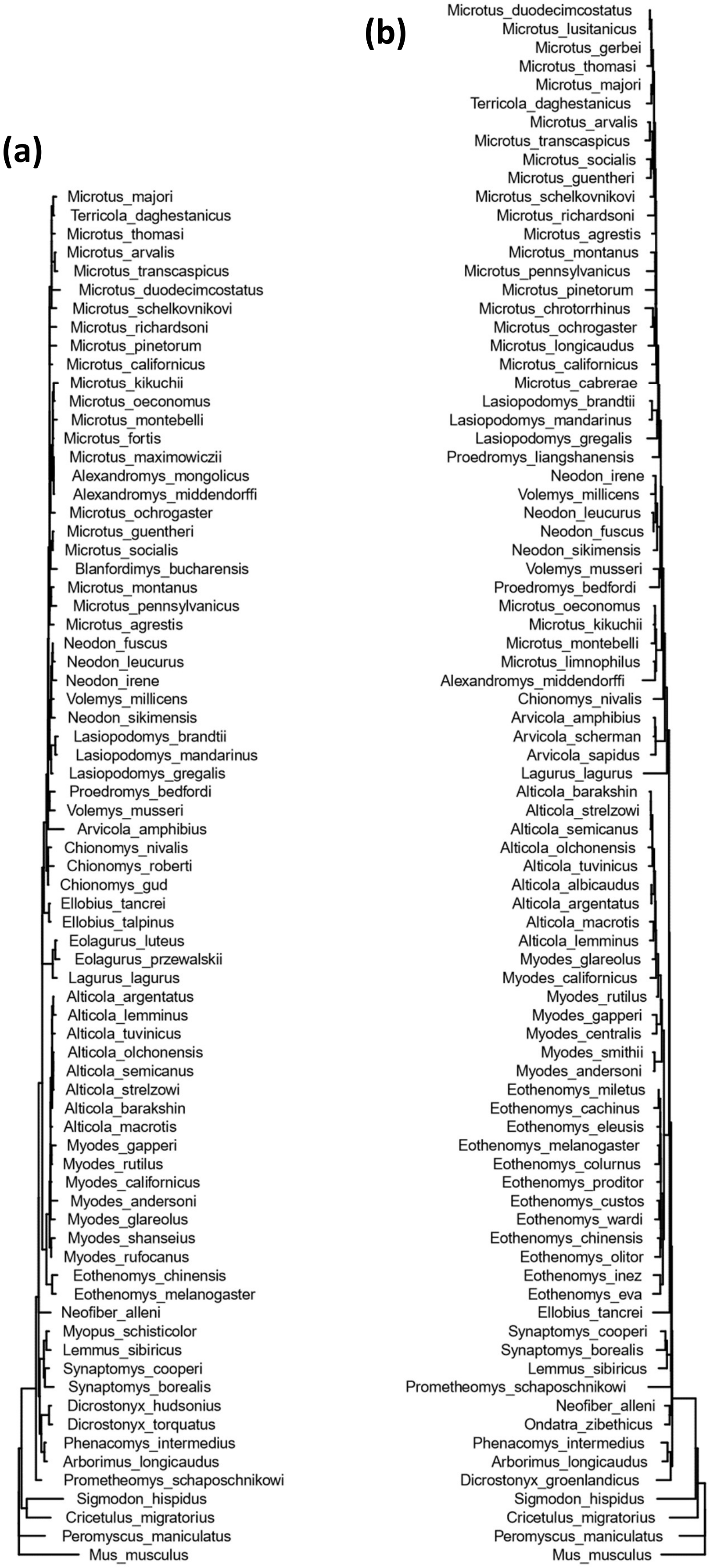


Figure 15: Maximum-likelihood phylogenetic tree of Arvicolinae using (a) *GHR* and (b) *IRBP* nuclear markers. Both trees are rooted on Mus musculus. Bootstrap scores can be found in the Appendix.

Supermatrix Approach:

Taxa which had available sequences for mitochondrial *Cytb*, nuclear *GHR*, and nuclear *IRBP* genes were selected, and each individual marker was aligned. The three alignments were then concatenated. The concatenate contained 3,101 sites (1,268 sites with at least one substitution) and 57 sequences (including four outgroup taxa). This alignment was then used to construct the supermatrix tree. ML **(Figure 16)** and Bayesian inference **(Figure A10)** phylogenetic trees were constructed, based on the substitution model GTR + G + I. Using a supermatrix approach improved the support for the topology, with no polytomies and a greater number of nodes with bootstrap scores higher than 50%. *Neofiber alleni* diverged first in the Arvicolinae species, with a bootstrap score of 65.6%. *Lemmus*, *Synaptomys*, *Arborimus*, *Phenacomys*, and *Prometheomys* were grouped into a monophyletic group and were the next to diverge. *Alticola*, *Myodes*, and *Eothenomys* formed a monophyletic group. *Microtus* species were grouped into one monophyletic clade with *Terricola daghestanicus* and *Alexandromys middendorffi*. *Neodon spp.*, *Lasiopodomys spp.*, and *Volemys milicens* were grouped together into a monophyletic group, whilst *Volemys musseri* diverged separately.

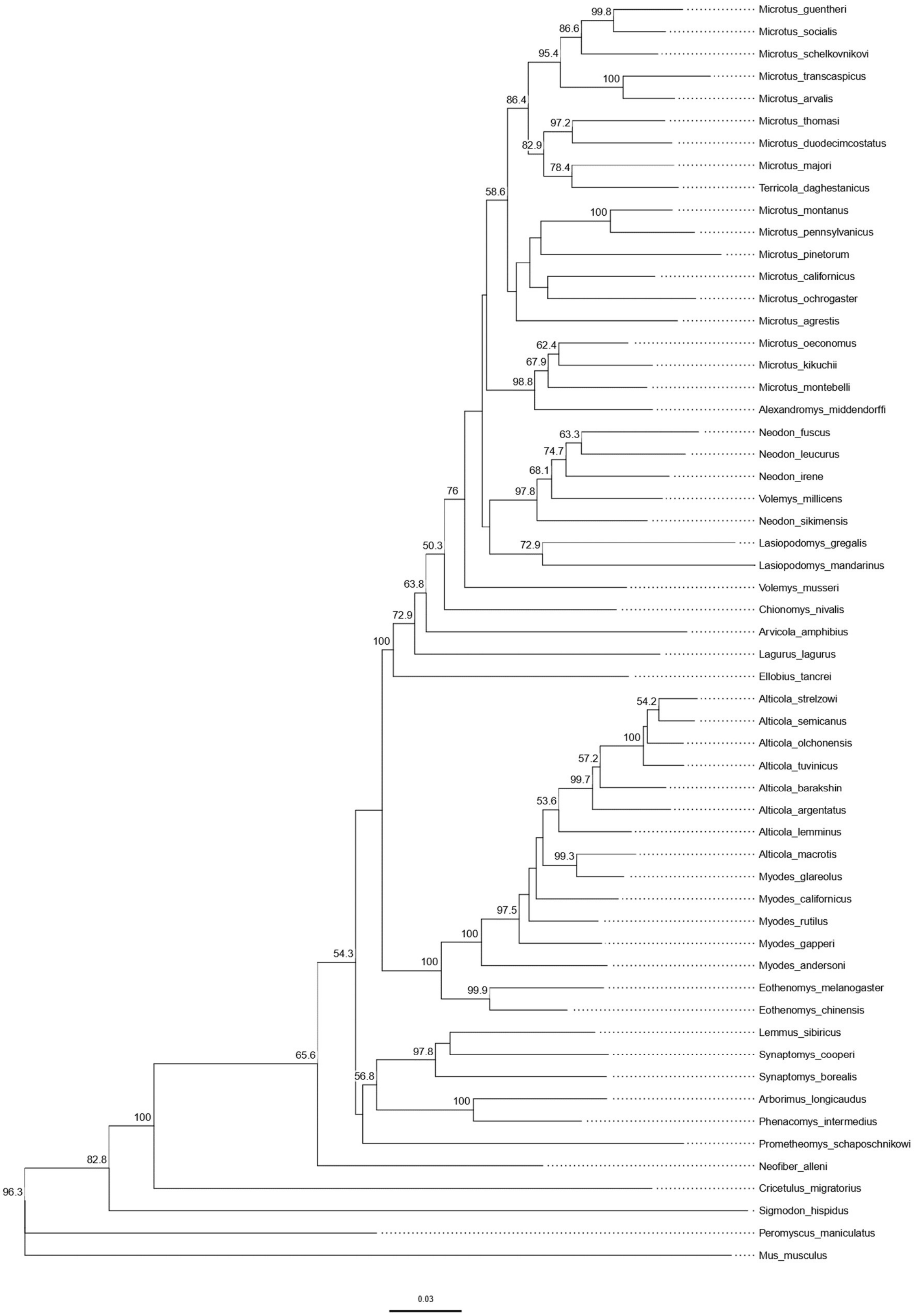


Figure 16: Maximum likelihood supermatrix for Cytb, GHR, and IRBP genes. Rooted on *Mus* musculus. Only bootstrap scores over 50% are shown.

4. Discussion:

4.1. DNA Extraction and Amplification:

All sample types used in this project resulted in variable amounts of extracted DNA, with a statistically significant difference between average DNA yield and sample type. Cell culture, as to be expected, produced the highest median quantity of DNA (33,550 ng; **Figure 6)**. This demonstrates the usefulness of this sample type. It is widely used in genomic sequencing studies for obtaining high molecular weight DNA and high concentrations of DNA required for next generation sequencing. The starting DNA requirements for 3rd generation sequencing platforms, such as PacBio and Illumina, require 50-1,000 ng of HMW DNA (Quail *et al.*, 2012). Our study found buccal swabs from cattle and tail tissue from water voles, to be an alternative to cell culture as the they produced substantial quantities of DNA (a median of 15,129.5 ng and 10,763.5 ng, respectively). All three sample types produced high molecular weight DNA when measured using gel electrophoresis (results only shown for tail tissue in **Figure 5**).

DNA extraction from hair and faecal samples achieved lower DNA yields (median of 1521.25 ng and 1018.25 ng, respectively; **Figure 6)**, but enough to sequence the mitochondrial control region and within the range of DNA yields required for next generation sequencing. One hair sample (sample 5) and two faecal samples (samples 12 and 18) produced poor sequencing results and were omitted. All samples amplified as viewed on the gel electrophoresis, but PCR clean-up and remaining contaminants in the DNA sample may have caused sequencing inhibition because only 83% of hair and 80% of faecal samples successfully sequenced, suggesting that non-invasive sampling is less reliable than tissue samples for DNA sequencing. The purity of the sample may also limit the success of amplifying longer reads for these sample types, and the quantity of nuclear DNA to mitochondrial DNA and non-host DNA was unknown.

We successfully optimised tail tissue, hair, and faecal DNA extractions, to achieve higher DNA yields and a purer DNA sample. For tail tissue samples buffer 3 (Wang and Storm, 2006) resulted in the highest DNA yield and improved purity scores. This buffer contained a higher concentration of Tris-HCl but lower concentrations of EDTA, NaCl, and SDS than the other two buffers. Using this buffer and the phenol-chloroform extraction achieved better results than ethanol or isopropanol extractions. A longer centrifugation step during DNA precipitation may have improved the binding of DNA to the tube, resulting in less DNA lost when isopropanol was discarded. The toxic nature of phenol-chloroform makes this extraction method less favourable than others, but our study shows it remains the gold-standard for obtaining high quantities of DNA.

For hair samples, the addition of $CaCl_2$ and DTT appears to improve the lysis of hair and hair follicles. It has been previously shown that the addition of Ca^{2+} to hair samples activates the enzyme proteinase K (Walsh, Metzger and Higuchi, 1991; Pfeiffer *et al.*, 2004). DTT is used as an alternative to physical digestion, by acting as both a reducing agent and an anionic detergent (Ghatak, Muthukumaran and Nachimuthu, 2013). The quantities of DNA can vary in hair samples due to the quantity of hair follicles and hair shafts obtained during sampling, with the recommended number of hairs per sample of at least 25 (Henry and Russello, 2011). This is because the hair shaft predominantly contains large quantities of mitochondrial DNA because the nuclei and nuclear DNA in the hair shaft are degraded during keratinization (Linch, Whiting and Holland, 2001; Graffy and Foran, 2005). Due to keratinization, hair shafts therefore contain large amounts of protein that need to be broken down to prevent PCR or sequencing inhibition. Phenol-chloroform extraction performed better when compared to a silica-bead extraction kit, with regards to both DNA yield, but more significantly purity.

Nuclear DNA can be recovered from hair, but this is more successful in hair in the anagen phase, the growth period, when there is more likely to be a root or sheath cells still attached that contain nuclei (Graham, 2007). Often non-invasive sampling obtains hair in the telogen phase, a phase where nuclear DNA is broken down into small fragments. These are often less than 200bp in length and would not be useful for sequencing nuclear markers.

Faecal samples produced the lowest DNA yield, with both extraction kits achieving similar DNA yields. On the other hand, the 'Qiagen QIAamp DNA Stool Mini Kit' resulted in a purer DNA sample, whilst absorbance 260/230 values were extremely low when using the 'Qiagen QIAamp PowerFecal DNA Kit'. Both kits required further DNA precipitation using ethanol in order to increase DNA yield and improve purity. In most samples, DNA yield increased, whilst some samples decreased due to loss of DNA during the process. We show that non-invasive faecal sampling is useful to obtain more samples and is less invasive than other sampling methods, with success at amplifying the mtDNA control region in 80% of faecal samples. But using faecal samples for extracting nuclear DNA, for short tandem repeat (STR) loci, has been shown to result in mismatches between faecal and blood samples of the same individual in a minority of cases (Forgacs *et al.*, 2019). This was mainly caused by allelic dropout in faecal samples. Consequently, care should be taken when using faecal samples in nuclear DNA analyses to undertake rigorous marker selection and to extract large quantities of DNA.

Although we obtained enough DNA to perform follow-up analysis, the percentage of host DNA is still unknown. Contaminants, such as the presence of microorganisms and human DNA on the sample, may have increased the DNA yield measured. Further screening using quantitative polymerase chain reaction (qPCR) and specific host primers to quantify the amount of host DNA in each sample (techniques used in Baker, 2015), will attain more accurate quantifications of DNA in each sample type, and further improve optimisation of DNA extraction specifically for host DNA.

4.2. Population Genetics of British Water Voles:

The 17 captive water vole samples from Wildwood Trust overall showed considerable genetic diversity **(Figure 8; Table 10)**. A total of 12 haplotypes were found, equating to a high haplotype diversity of 0.949, but a low nucleotide diversity of 0.004. Of the two haplogroups the largest showed considerably low genetic diversity, with a maximum of two mutational steps between each haplotype and three haplotypes with multiple individuals. Whilst the other haplogroup, containing two haplotypes, differed substantially with at least six mutational steps between the haplogroups. Water voles from these two haplotypes may have been captured from different populations than the others sampled. Overall, we find that there is maintenance of genetic diversity in the captive population at Wildwood Trust.

When comparing the captive population with natural populations from the South East and East of England we found tight clustering of haplotypes into one major haplogroup **(Figure 9)**. Only one haplotype (haplotype 14) diverged considerably by 17 mutational steps. This haplotype consisted of samples from Dartford, Kent. The reason for this divergence from all other haplotypes has been hypothesized to be due to breeding programs in the area which may have caused admixture of captive stock and release of this admixed population into the wild (Baker, 2015). The study found that this haplotype clustered with Scottish haplotypes. Admixing between England and Scotland populations will lead to the loss of local genetic heritage. The authors sampled 58 individuals from 13 populations, finding 14 unique haplotypes, with a haplotype diversity of 0.80. This is considerably lower than the captive population at Wildwood Trust. Unfortunately, we were only able to obtain haplotype sequences, so we could not calculate our own haplotype diversity, to allow for a comparison between captive and natural populations in the South East of England.

As shown by our results, the water voles from Wildwood Trust clustered closely together with South East of England haplotypes, with a maximum of three mutational steps between natural and captive individuals, excluding natural water vole haplotypes 14 and 17, which were at least 17 and five mutational steps from other haplotypes, respectively. Only two captive haplotypes (haplotypes 1 and 2) showed divergence from the rest, clustering more closely with the East of England haplotype (haplotype 25) by at least five mutational steps. This may indicate that these water voles were captured from the East of England, rather than more locally in the South East, or that water voles from the East of England had been translocated to this area in the past. When haplotypes 1 and 2 were removed for the alignment, haplotype diversity was slightly lower at 0.933. This is still considerably high for a declining population. This again demonstrates that genetic diversity is being maintained in the sampled captive population at Wildwood Trust.

Our study finds that all captive individuals are part of the English/Welsh group **(Figure 10)** when compared with wild populations from England, Wales, and Scotland (Piertney *et al.*, 2005). This is important for the maintenance of the two lineages, as the two populations should be treated as separate ESUs. The release of these captive water voles or their offspring into the South East will not cause any considerable admixture. In all our analyses that include samples from both modern and ancient British water voles, there is support for the English/Welsh and Scottish divergence, as well as two colonisation events into Britain which have shaped current phylogeographic structure (Piertney *et al.*, 2005; Brace *et al.*, 2016).

Constructing a phylogenetic tree of all available sequences showed all Scottish water voles to diverge first, whilst English and Welsh water voles diverge later, forming a separate monophyletic clade **(Figure 11)**. Three outliers were present in the tree. South East of England (haplotype 14) which as mentioned above and in Baker, 2015 has probably been relocated from Scotland and released into the South East of England. Samples collected from Northumberland and Lincolnshire are again likely to be reintroductions, or in the case of Northumberland, geographically close to Scottish populations (Piertney *et al.*, 2005). In the phylogenetic tree the Italian and Swiss samples diverged first and were separate to all other European water voles. The three *Arvicola scherman* samples were paraphyletic and found in both the English/Welsh and Scottish clades. Based on mitochondrial DNA *A. scherman* should not be classified as a separate species (Brace *et al.*, 2016), but further analysis is needed. Further analysis is also need for the Italian and Swiss water voles, which may be classified as a separate evolutionary distinct unit with additional molecular analysis.

Baker *et al.*, 2020 found that the diversity of South East of England water voles mtDNA is structured by watersheds. They suggested that conservation management should take place at this local level, with reintroductions ensuring local genetic heritage is maintained. They found evidence using both microsatellite data and mitochondrial haplotypes that reintroductions in the area had left a significant genetic footprint, and this is also demonstrated in our results. Our results suggest that if captive water voles are released back into the South East of England, we expect there to be little outbreeding depression and genetic footprint on local populations, because Wildwood Trust and South East populations are closely related, in terms of the clustering of haplotypes. If water voles are released back into watersheds in the local area, then care should be taken that the populations are not too closely related to the released water voles, to avoid inbreeding and subsequently inbreeding depression, decreasing genetic diversity further. This demonstrates the importance of considering genetic and evolutionary relationships in conservation management.

4.3. Phylogenetics of Arvicolinae:

Phylogenies of the Genera of Arvicolinae:

The vast amount of genetic data available today provides taxonomists and systematists with a wealth of information for discovering how species have evolved. Our results show the importance of selecting the appropriate genetic marker in phylogenetic studies, as different markers and phylogenetic approaches can result in different phylogenies. By using several markers, both nuclear and mitochondrial, as well as mitochondrial genomes, it allows the comparison of tree topologies and the support for phylogenies that are identical in all phylogenetic constructions.

The availability of DNA sequences is vital to fully resolve phylogenies of groups of species, however the quantity of phylogenetic markers for all Arvicolinae genera is limited to date. Our mitogenome analysis used 10 genera, with up to 30 proposed, resulting in missing information for the 20 other possible genera. Individual markers provided considerably more genera, but only 18 genera were available in all three of the selected markers for the supermatrix tree. However, our analyses do reveal new information and support previous publications on the phylogenetics of Arvicolinae. We use the proposed phylogeny of Arvicolinae by Robovský, Řičánková and Zrzavý, 2008 as a basis for discussion of our results as well as other publications.

Basal Arvicolinae – The phylogeny of the basal arvicolines has long been debated, with no resolved topology. *Hyperacrius*, *Prometheomys*, *Ellobius*, *Eolagurus*, and *Lagurus* have been proposed as the basal arvicoline, but the relationships between them are unknown (Robovský, Řičánková and Zrzavý, 2008). In our study mitochondrial genomes were unavailable for all the proposed basal arvicolines, so *Ondatra* and *Dicrostonyx* diverged first forming a monophyletic clade **(Figures 13 and A1)**. Whereas in our supermatrix trees **(Figures 16 and A10)** *Neofiber* diverged first within Arvicolinae, followed by *Prometheomys*, while *Lagurus* and *Ellobius* diverged much later (Figure). This is consistent with other studies which proposed *Lagurus*, *Eolagurus*, and *Ellobius* diverged later in Arvicolinae (when sampled) and *Prometheomys* was the first to diverge (Galewski *et al.*, 2006; Abramson *et al.*, 2009; Steppan and Schenk, 2017). *Hyperacrius* was not sampled in any of these studies, so its phylogeny remains unknown.

'Dicrostonychini' – The next clade to diverge has been proposed as 'Dicrostonychini' and contains *Dicrostonyx*, *Phenacomys*, and *Arborimus* (Robovský, Řičánková and Zrzavý, 2008).

Mitochondrial genomes were unavailable for *Phenacomys* and *Arborimus*. *Dicrostonyx* grouped with *Ondatra* **(Figures 13 and A1)**. Our supermatrix trees supported the grouping of *Phenacomys* and *Arborimus* but sequences for *Dicrostonyx* were unavailable to fully support this clade **(Figures 16 and A10)**. Both the *GHR* **(Figures A6 and A7)** and *IRBP* **(Figures A8 and A9)** phylogenetic trees supported the proposed grouping, with all genera accounted for. Our results are also supported by a more recent publication using multiple mitochondrial and nuclear phylogenetic markers (Steppan and Schenk, 2017).

Dinaromys*, *Neofiber*, *Ondatra*, and the clade 'Lemmini' containing *Synaptomys*, *Lemmus*, and *Myopus – These two clades formed a polytomy in the proposed phylogeny by Robovský, Řičánková and Zrzavý, 2008. Our mitogenome analysis proved ineffective in resolving the phylogeny of this polytomy, due to only *Ondatra* being sampled **(Figures 13 and A1)**. In the supermatrix trees *Neofiber* diverged first in Arvicolinae, before the 'Dicrostonychini' clade, disagreeing with the proposed phylogeny **(Figure 16 and A10)**. The grouping of 'Lemmini' is supported, with *Lemmus* and *Synaptomys* sampled and grouped together. They were a sister clade to 'Dicrostonychini'. *Neofiber* and *Ondatra* were grouped together in both the *Cytb* **(Figures A4 and A5)** and *IRBP* **(Figures A8 and A9)** trees. *Dinaromys* sequences were only found for *Cytb* gene trees, placing *Dinaromys* as a sister clade to *Arvicola*. All our analyses supported the grouping of 'Lemmini', in agreement with Steppan and Schenk, 2017. However, the relationship between 'Lemmini' and the other genera (*Dinaromys*, *Neofiber*, and *Ondatra*) is unresolved and not consistent in the literature, and our results do not shed any light on this.

'Clethrionomyini' – This is a highly-supported clade containing genera *Eothenomys*, *Myodes*, and *Alticola*, with the clade name proposed in (Robovský, Řičánková and Zrzavý, 2008). All our analyses highly support the grouping of *Eothenomys* and *Myodes*, and *Alticola,* when sampled, and is supported by more extensive molecular studies (Steppan and Schenk, 2017).

'Arvicolini' – The placement of *Arvicola*, *Lemmiscus*, *Stenocranius*, and *Chionomys* within this clade has not been fully resolved. The four genera have been proposed to diverge after 'Clethrionomyini' and first within the 'Arvicolini' clade, before all other genera (Robovský, Řičánková and Zrzavý, 2008). *Stenocranius* was not sampled in any of our trees and is underrepresented in the literature. In our mitogenome trees, only *Arvicola* is sampled, which diverged after 'Clethrionomyini' taxa and before *Proedromys* **(Figures 13 and A1)**. The supermatrix trees contain both *Arvicola* and *Chionomys*, with the former of these genera diverging first and separately. Each of our gene trees showed different topologies for these four genera **(Figures 16 and A10)**. In the *Cytb* trees *Arvicola*, *Dinaromys*, and *Lemmiscus* formed a monophyletic group, after 'Clethrionomyini' taxa, whilst *Chionomys* diverged later **(Figures A4 and A5)**. Whereas in the *GHR* trees *Chionomys* diverged first, followed by *Arvicola* **(Figures A6**

and A7). *Arvicola* diverged first in the *IRBP* tree, then *Chionomys* **(Figures A8 and A9)**. We suggest, based on our results and the literature, that *Arvicola* (including water voles) may be the basal 'Arvicolini' and diverges separately from other genera.

The remaining proposed genera pertaining to 'Arvicolini' are *Microtus*, *Neodon*, *Alexandromys*, *Mynomes*, *Lasiopodomys*, and *Terricola* (Robovský, Řičánková and Zrzavý, 2008). It has been found that *Microtus* is paraphyletic in this group and therefore the nomenclature of genera in this group should be changed to *Microtus* to reflect monophyletic relationships, while others propose that the remaining genera should be re-classified as sub-genera within the genera *Microtus* (discussed in Robovský, Řičánková and Zrzavý, 2008). In all our trees *Microtus* was paraphyletic and *Terricola* was always grouped within *Microtus* taxa. We therefore support the reclassification of *Terricola* to *Microtus*.

In the mitogenome trees *Lasiopodomys* and *Neodon* taxa were also grouped with *Microtus* **(Figures 13 and A1)**. In the supermatrix tree *Alexandromys* is grouped with *Microtus* **(Figures 16 and A10)**. *Neodon*, *Volemys*, and *Lasiopodomys* formed a separate sister clade and this was also true for the *GHR* gene trees **(Figures A6 and A7)**. Phylogenies of this group in individual gene trees differed due to the presence or absence of genera. In our *Cytb* trees **(Figures A4 and A5)** *Microtus* taxa were in a clade alongside *Terricola*, *Blanfordimys*, *Alexandromys*, *Neodon*, and *Volemys* taxa and in our *IRBP* gene trees **(Figures A8 and A9)** *Terricola*, *Lasiopodomys*, *Proedromys*, *Neodon*, *Volemys*, *Proedromys*, and *Alexandromys* taxa formed a clade with the available *Microtus* taxa. Our results were not consistent enough to make any further conclusions about this group.

Other Genera – *Proedromys*, *Volemys*, and *Blanfordimys*. The position of *Proedromys* was unresolved in our analyses, with different phylogenies for each of the different markers. The mitogenome trees placed it between *Arvicola* and the remaining sampled 'Arvicolini' genera **(Figures 13 and A1)**. Robovský, Řičánková and Zrzavý, 2008 proposed that the genus was grouped with *Lasiopodomys* within 'Arvicolini', whereas another study (Steppan and Schenk, 2017) found that the genus groups with selected *Volemys* species.

For *Volemys* the nuclear makers showed the genus to be paraphyletic, with *Volemys musseri* grouping with *Proedromys bedfordi*, whereas *Volemys millicens* grouped with the remaining sampled *Neodon spp*. The latter grouping is also supported by the supermatrix trees **(Figures 16 and A10)**, but *Proedromys bedfordi* was not sampled to support the former clade. A recent publication showed *Volemys* to be paraphyletic, with some *Volemys* species grouping with *Neodon* and others grouping with *Proedromys* (Steppan and Schenk, 2017). We support the notion that *Volemys* is paraphyletic, and more support may be achieved with sequencing the mitochondrial genome.

The phylogeny of *Blanfordimys* within 'Arvicolini' is relatively supported in the literature when the genus is sampled. Our mitogenome **(Figures 13 and A1)** and supermatrix **(Figures 16 and A10)** phylogenetic trees do not provide any information regarding this genus due to a lack of sequences. More extensive sequencing of *Blanfordimys* is needed to resolve its phylogeny.

Summary – Using both the mitochondrial genome and supermatrix approach for all available genera, we support previous studies and propose new phylogenies for genera within Arvicolinae. Our results showed support for *Prometheomys* as a basal arvicoline, along with *Neofiber*. Our results disagree with the basal positioning of *Ellobius*, *Eolagurus*, and *Lagurus* by Robovský, Řičánková and Zrzavý, 2008. Instead we support the phylogenies of Abramson *et al.*, 2009 and Steppan and Schenk, 2017. We fully support the grouping of *Dicrostonyx*, *Arborimus*, and *Phenacomys* into the clade 'Dicrostonychini', as well as the grouping of *Eothenomys*, *Myodes*, and *Alticola* into the clade 'Cletherionomyini'. We support the positioning of *Arvicola* within the clade 'Arvicolini' and our mitogenome analyses suggest it may be the earliest to diverge within this clade, followed by *Proedromys*. There was not sufficient sampling to resolve the phylogenies of the remaining genera.

Nomenclature and Classification – Our analysis of the mitochondrial genome of Arvicolinae species, although limited by the low numbers of available taxa, provides support for the paraphyly of *Microtus*. We demonstrate high support, based on both high bootstrap scores and posterior probabilities, for *Terricola*, *Neodon*, and *Lasiopodomys* diverging with *Microtus* species. Therefore, based on the mitochondrial analysis alone, the nomenclature of these genera should be reflected by their evolutionary history at the genus or sub-genus level. Further support is provided for renaming of the genus *Terricola*, which was found to group with *Microtus* species in all phylogenetic trees. The supermatrix trees also support the paraphyly of *Microtus* with *Alexandromys*, but *Neodon*, *Volemys*, and *Lasiopodomys* group into a monophyletic clade, providing less support that the nomenclature of these genera should be change. An increased sampling of taxa is needed to resolve this.

Comparison of Phylogenetic Approaches:

The use of molecular markers has revolutionised the field of phylogenetics, greatly improving our understanding of the relationships between species. Approaches include using individual gene markers, using mitochondrial or nuclear genomes (phylogenomics), and approaches which combine multiple markers (i.e. concatenating sequences and constructing supermatrix trees).

However, choosing which marker or approach to take remains pivotal for achieving resolved species phylogeny.

Our results demonstrate the importance of selecting appropriate phylogenetic markers as we have considerable differences in phylogeny between markers and approaches. The main issue was the availability of genetic markers for all genera and species within Arvicoline. Sampling more taxa often greatly improves phylogenetic accuracy (Zwickl and Hillis, 2002), and provides further information. The other issue is the choice of phylogenetic marker and the use of the supermatrix approach. Differences between individual gene trees and the actual species tree can be caused by hybridisation, incomplete lineage sorting, and gene duplications (Maddison, 1997). This can often result in inaccurate phylogenies when individual genes are used.

The supermatrix approach allows for all characters to be included in a phylogenetic analysis, increasing the information available. For our concatenated sequence we only used taxa with sequences available for all three phylogenetic markers. This was to avoid missing data during the phylogenetic tree construction. However, missing data might not negatively impact the supermatrix approach (Wiens, 2006), therefore increasing the number of taxa available might have improved phylogenetic support and provided more information from our supermatrix tree.

We demonstrate that constructing phylogenies using Arvicolinae mitochondrial genomes greatly improves bootstrap and posterior probabilities, compared with using individual genes or concatenated sequences of a handful of markers. All posterior probabilities were 1.00 in the Bayesian phylogenetic tree and relatively high (>70) for most nodes in the ML phylogenetic tree. This shows the importance of sequencing the mitochondrial genome for all Arvicolinae species and may prove useful in further resolving the family's phylogeny.

4.4. Future Work:

The future of water voles in Britain relies on the maintenance of population numbers and genetic diversity. Future work could include more extensive molecular studies, such as exploring the genetic diversity of whole mitochondrial genomes with additional samples from both captive and natural populations throughout Britain. Moreover, using nuclear DNA of captive water voles, such as microsatellites, will allow for other genetic diversity indicators to be determined, such as heterozygosity and fixation indices. This will provide a more thorough investigation into the genetic diversity of water voles in captivity. If funds were available, using population genomics to extensively assess population structure of both captive and natural water voles in Britain using techniques such as RADSeq or whole-genome sequencing. This would increase the number of markers in the study and increase confidence in the results, which would then improve the management of the species. Using these techniques would also help to uncover the relationships between *Arvicola* species and their classification as ESUs. Our study provides an excellent framework for future work in this field.

To resolve the phylogeny of genera and species within Arvicolinae a more extensive sampling of taxa and genetic markers is required. The reduction in sequencing costs will allow for more species to be sequenced in the future, as well as additional genetic markers for each species. Phylogenomics is becoming more widely used with the further sequencing of non-model animals. This would increase the number of informative sites to infer phylogeny, which would prove useful to fully support relationships in this subfamily. Sequencing mitogenomes for all species would be a good start and following this whole-genome sequencing. This would also allow for comparative and functional genomics of the subfamily.

5. Conclusion:

DNA was successfully extracted from various animal sample types and DNA extraction protocols were optimised to achieve high DNA yields and pure DNA samples. We found that cell culture achieved the highest DNA yield, demonstrating its usefulness in providing high quantities of high-quality DNA. Our optimised protocol for buccal swabs and tail tissue also provided considerably high quantities of DNA. We show that DNA from all three samples types can be obtained relatively quickly, easily, and are within the quantity range needed for next-generation sequencing. Hair and faecal samples provided adequate amounts of DNA for amplifying the mitochondrial control region for water voles at Wildwood Trust. But it appears that these samples were more prone to sequencing inhibition, as they contain more contaminants. Their use in sequencing nuclear DNA was not tested in this study but is an area that needs to be explored further.

We show that using non-invasive or non-destructive sampling can provide adequate quantities of DNA for sequencing mitochondrial DNA for population genetic studies. Captive water voles at Wildwood Trust were randomly sampled to better understand their genetic diversity and population structure. We found that the captive population had considerable genetic diversity, in terms of the clustering of haplotypes and relatively high haplotype diversity. When comparing with natural populations, we show that the captive population had maintained genetic diversity, with only a slight decrease in haplotype diversity. Water vole haplotypes closely clustered with natural water voles from the South East and East of England. We proposed that the release of captive water voles back into the South East of England would not cause significant loss of local genetic heritage. Wildwood Trust water voles were all found within the English and Welsh clade, with Scottish water voles diverging separately. Care should be taken to maintain the genetic distinctiveness of these two groups of water voles in Britain, and as others have found, a few water vole populations in England are more closely related to Scottish populations. This could lead to admixture between the England/Wales and Scotland water voles, which is concerning and should be prevented.

The complete phylogenetics of species within Arvicolinae remains unresolved. Our study provides support for the phylogeny of several genera, such as the paraphyly of *Microtus* and the grouping of *Microtus* with other arvicoline genera, such as *Terricola*, *Lasiopodomys*, and *Neodon,* in our mitogenome analyses. We therefore support the renaming of these genera to reflect evolutionary relationships. We also propose that the genus *Arvicola*, containing the European water vole, is the basal 'Arvicolini', with evidence from the mitogenome and supermatrix analyses. However, lack of extensive taxon sampling for all markers and mitochondrial genomes resulted in our analyses lacking crucial genera to fully resolve the phylogeny of this subfamily. Moreover, we show support for using the mitochondrial genome as a marker for Arvicolinae phylogenetics, with a fully resolved tree and high bootstrap scores and posterior probabilities that individual markers and propose that sequencing more mitogenomes will help to fully understand the evolutionary relationships of Arvicolinae.

References:

Abramson, N. I. *et al.* (2009) ‘Supraspecies relationships in the subfamily Arvicolinae (rodentia, cricetidae): An unexpected result of nuclear gene analysis’, *Molecular Biology*, 43(5), pp. 834–846. doi: 10.1134/S0026893309050148.

Alqahtani, F. *et al.* (2020) ‘Complete mitochondrial genome of the water vole, Microtus richardsoni (Cricetidae, Rodentia)’, *Mitochondrial DNA Part B: Resources*. Taylor & Francis, 5(3), pp. 2498–2499. doi: 10.1080/23802359.2020.1780640.

Baker, B. R. (2015) *Demographic and genetic patterns of water voles in human modified landscapes : implications for conservation By Rowenna Baker*. University of Brighton.

Baker, R. J. *et al.* (2020) ‘Genetic structure of regional water vole populations and footprints of reintroductions: a case study from southeast England’, *Conservation Genetics*. Springer Netherlands, 21(3), pp. 531–546. doi: 10.1007/s10592-020-01268-4.

Barreto, G. R. *et al.* (1998) ‘The role of habitat and mink predation in determining the status and distribution of water voles in England’, *Animal Conservation*, 1(2), pp. 129–137. doi: 10.1111/j.1469-1795.1998.tb00020.x.

Batsaikhan, N. *et al.* (2016) ‘Arvicola Amphibius’, *The IUCN Red List of Threatened Species 2016*. doi: http://dx.doi.org/10.2305/IUCN.UK.2016- 3.RLTS.T2149A22358646.en.

Bernatchez, L. and Wilson, C. C. (1998) ‘Comparative phylogeography of Nearctic and Palearctic fishes’, *Molecular Ecology*, 7(4), pp. 431–452. doi: 10.1046/j.1365-294x.1998.00319.x.

Blanga-Kanfi, S. *et al.* (2009) ‘Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades’, *BMC Evolutionary Biology*, 9(1), p. 71. doi: 10.1186/1471-2148-9-71.

Bondareva, O. V. and Abramson, N. I. (2019) ‘The complete mitochondrial genome of the common pine vole Terricola subterraneus (Arvicolinae, Rodentia)’, *Mitochondrial DNA Part B: Resources*. Taylor & Francis, 4(2), pp. 3925–3926. doi: 10.1080/23802359.2019.1687026.

Brace, S. *et al.* (2016) ‘The colonization history of British water vole (Arvicola amphibius (Linnaeus, 1758)): Origins and development of the Celtic fringe’, *Proceedings of the Royal Society B: Biological Sciences*, 283(1829). doi: 10.1098/rspb.2016.0130.

Burgin, C. J. *et al.* (2018) ‘How many species of mammals are there?’, *Journal of Mammalogy*, 99(1), pp. 1–14. doi: 10.1093/jmammal/gyx147.

Canu, A. *et al.* (2013) ‘Influence of management regime and population history on genetic diversity and population structure of brown hares (Lepus europaeus) in an Italian province’, *European Journal of Wildlife Research*, 59(6), pp. 783–793. doi: 10.1007/s10344-013-0731-x.

Carleton, M. D. and Musser, G. G. (2005) ‘Order rodentia’, in *Mammal species of the world*. 3rd edn. Baltimore: The John Hopkins University Press, pp. 745–2142.

Corbet, G. B. *et al.* (1970) ‘The taxonomic status of British Water voles, genus Arvicola’, *Journal of Zoology*, 161(3), pp. 301–316. doi: 10.1111/j.1469-7998.1970.tb04515.x.

DeBry, R. W. and Sagel, R. M. (2001) ‘Phylogeny of Rodentia (Mammalia) Inferred from the Nuclear-Encoded Gene IRBP’, *Molecular Phylogenetics and Evolution*, 19(2), pp. 290–301. doi: 10.1006/mpev.2001.0945.

Earl, D. A. and vonHoldt, B. M. (2012) ‘STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method’, *Conservation Genetics Resources*, 4(2), pp. 359–361. doi: 10.1007/s12686-011-9548-7.

Filipi, K. *et al.* (2015) ‘Mitogenomic phylogenetics of the bank vole Clethrionomys glareolus, a model system for studying end-glacial colonization of Europe’, *Molecular Phylogenetics and Evolution*, 82(PA), pp. 245–257. doi: 10.1016/j.ympev.2014.10.016.

Folkertsma, R. *et al.* (2018) ‘The complete mitochondrial genome of the common vole, Microtus arvalis (Rodentia: Arvicolinae)’, *Mitochondrial DNA Part B: Resources*. Informa UK Ltd., 3(1), pp. 446–447. doi: 10.1080/23802359.2018.1457994.

Forgacs, D. *et al.* (2019) ‘Evaluation of fecal samples as a valid source of DNA by comparing paired blood and fecal samples from American bison (Bison bison)’, *BMC Genetics*. BMC Genetics, 20(1), pp. 1–8. doi: 10.1186/s12863-019-0722-3.

Gagneux, P., Boesch, C. and Woodruff, D. S. (1997) ‘Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair’, *Molecular Ecology*, 6(9), pp. 861–868. doi: 10.1111/j.1365-294X.1997.tb00140.x.

Galewski, T. *et al.* (2006) ‘The evolutionary radiation of Arvicolinae rodents (voles and lemmings): Relative contribution of nuclear and mitochondrial DNA phylogenies’, *BMC Evolutionary Biology*, 6(i), pp. 1–17. doi: 10.1186/1471-2148-6-80.

Ghatak, S., Muthukumaran, R. B. and Nachimuthu, S. K. (2013) ‘A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis’, *Journal of Biomolecular Techniques*, 24(4), pp. 224–231. doi: 10.7171/jbt.13-2404-001.

Graffy, E. A. and Foran, D. R. (2005) ‘A Simplified Method for Mitochondrial DNA Extraction from Head Hair Shafts’, *Journal of Forensic Sciences*, 50(5), pp. 1–4. doi: 10.1520/JFS2005126.

Graham, E. A. M. (2007) 'DNA reviews: hair', *Forensic Science, Medicine, and Pathology*, 3(2), pp. 133–137. doi: 10.1007/s12024-007-9005-9.

Green, M. R., Hughes, H., Sambrook, J. and MacCallum, P. (2012) 'Molecular Cloning. A Laboratory Manual', in *Biochemical Education*. 4th edn. New York: Cold Spring Harbor Laboratory Press,U.S., pp. 58–60.

Guo, W. *et al.* (2009) 'DNA extraction procedures meaningfully influence qPCR-based mtDNA copy number determination', *Mitochondrion*, 9(4), pp. 261–265. doi: 10.1016/j.mito.2009.03.003.

Henry, P. and Russello, M. A. (2011) 'Obtaining high-quality DNA from elusive small mammals using low-tech hair snares', *European Journal of Wildlife Research*, 57(3), pp. 429–435. doi: 10.1007/s10344-010-0449-y.

Hewitt, G. M. (1996) 'Some genetic consequences of ice ages, and their role in divergence and speciation', *Biological Journal of the Linnean Society*, 58(3), pp. 247–276. doi: 10.1111/j.1095-8312.1996.tb01434.x.

Hewitt, G. M. (1999) 'Post-glacial re-colonization of European biota', *Biological Journal of the Linnean Society*, 68(1–2), pp. 87–112. doi: 10.1006/bijl.1999.0332.

Huelsenbeck, J. P. and Ronquist, F. (2001) 'MRBAYES: Bayesian inference of phylogenetic trees ', *Bioinformatics*, 17(8), pp. 754–755. doi: 10.1093/bioinformatics/17.8.754.

Jain, M. *et al.* (2017) 'Nanopore sequencing and assembly of a human genome with ultra-long reads', *bioRxiv*, p. 128835. doi: 10.1101/128835.

Jefferies, D. J., Morris, P. A. and Mulleneux, J. E. (1989) 'An enquiry into the changing status of the Water Vole Arvicola terrestris in Britain', *Mammal Review*, 19(3), pp. 111–131. doi: 10.1111/j.1365-2907.1989.tb00406.x.

Laikre, L. *et al.* (2010) 'Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals', *Trends in Ecology & Evolution*, 25(9), pp. 520–529. doi: 10.1016/j.tree.2010.06.013.

Larkin, M. A. *et al.* (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23(21), pp. 2947–2948. doi: 10.1093/bioinformatics/btm404.

Linch, C. A., Whiting, D. A. and Holland, M. M. (2001) 'Human Hair Histogenesis for the Mitochondrial DNA Forensic Scientist', *Journal of Forensic Sciences*, 46(4), p. 15056J. doi: 10.1520/JFS15056J.

Maddison, W. P. (1997) ‘Gene Trees in Species Trees’, *Systematic Biology*. Edited by J. J. Wiens, 46(3), pp. 523–536. doi: 10.1093/sysbio/46.3.523.

Marková, S. *et al.* (2020) ‘High genomic diversity in the bank vole at the northern apex of a range expansion: The role of multiple colonizations and end-glacial refugia’, *Molecular Ecology*, 29(9), pp. 1730–1744. doi: 10.1111/mec.15427.

Mathews, F. *et al.* (2018) *A Review of the Population and Conservation Status of British Mammals*. A report by the Mammal Society under contract to Natural England, Natural Resources Wales and Scottish Natural Heritage.

McNeely, J. A. *et al.* (1990) *Conserving the world’s biological diversity*. International Union for conservation of nature and natural resources.

Miller, G. S. (1912) *Catalogue of the mammals of Western Europe (Europe exclusive of Russia) in the collection of the British museum, by Gerrit S. Miller.* Edited by B. M. of N. History. London,: Printed by order of the Trustees,. doi: 10.5962/bhl.title.8830.

Paradis, E. (2010) ‘pegas: an R package for population genetics with an integrated–modular approach’, *Bioinformatics*, 26(3), pp. 419–420. doi: 10.1093/bioinformatics/btp696.

Paradis, E. and Schliep, K. (2018) ‘ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R’, *Bioinformatics*, 35(3), pp. 526–528. doi: 10.1093/bioinformatics/bty633.

Pfeiffer, I. *et al.* (2004) ‘Forensic DNA-typing of dog hair: DNA-extraction and PCR amplification’, *Forensic Science International*, 141(2–3), pp. 149–151. doi: 10.1016/j.forsciint.2004.01.016.

Piertney, S. B. *et al.* (2005) ‘Phylogeographic structure and postglacial evolutionary history of water voles (Arvicola terrestris) in the United Kingdom’, *Molecular Ecology*, 14(5), pp. 1435–1444. doi: 10.1111/j.1365-294X.2005.02496.x.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000) ‘Inference of population structure using multilocus genotype data.’, *Genetics*, 155(2), pp. 945–59. doi: 10.1016/0730-725x(92)90453-7.

Quail, M. *et al.* (2012) ‘A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers’, *BMC Genomics*, 13(1), p. 341. doi: 10.1186/1471-2164-13-341.

Robovský, J., Řičánková, V. and Zrzavý, J. (2008) ‘Phylogeny of Arvicolinae (Mammalia, Cricetidae): Utility of morphological and molecular data sets in a recently radiating clade’, *Zoologica Scripta*, 37(6), pp. 571–590. doi: 10.1111/j.1463-6409.2008.00342.x.

Schiebelhut, L. M. *et al.* (2017) ‘A comparison of DNA extraction methods for high-throughput DNA analyses’, *Molecular Ecology Resources*, 17(4), pp. 721–729. doi: 10.1111/1755-0998.12620.

Schliep, K. *et al.* (2017) ‘Intertwining phylogenetic trees and networks’, *Methods in Ecology and Evolution*. Edited by R. Fitzjohn, 8(10), pp. 1212–1220. doi: 10.1111/2041-210X.12760.

Spielman, D., Brook, B. W. and Frankham, R. (2004) ‘Most species are not driven to extinction before genetic factors impact them’, *Proceedings of the National Academy of Sciences*, 101(42), pp. 15261–15264. doi: 10.1073/pnas.0403809101.

Steppan, S. J. and Schenk, J. J. (2017) ‘Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates’, *PLOS ONE*. Edited by D. Huchon, 12(8), p. e0183070. doi: 10.1371/journal.pone.0183070.

Strachan, R. (2004) ‘Conserving water voles: Britain’s fastest declining mammal’, *Water and Environment Journal*, 18(1), pp. 1–4. doi: 10.1111/j.1747-6593.2004.tb00483.x.

Strachan, R. and Jefferies, D. J. (1993) *The water vole Arvicola terrestris in Britain 1989-1990: its distribution and changing status*, *London: Vincent Wildlife Trust*.

Taberlet, P. *et al.* (1998) ‘Comparative phylogeography and postglacial colonization routes in Europe’, *Molecular Ecology*, 7(4), pp. 453–464. doi: 10.1046/j.1365-294x.1998.00289.x.

Taberlet, P., Waits, L. P. and Luikart, G. (1999) ‘Noninvasive genetic sampling: look before you leap’, *Trends in Ecology & Evolution*, 14(8), pp. 323–327. doi: 10.1016/S0169-5347(99)01637-7.

Telfer, S. *et al.* (2003) ‘Demographic and genetic structure of fossorial water voles (Arvicola terrestris) on Scottish islands’, *Journal of Zoology*. University of Kent, 259(1), pp. 23–29. doi: 10.1017/S0952836902003321.

Triant, D. A. and DeWoody, J. A. (2008) ‘Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodents’, *Genetica*, 132(1), pp. 21–33. doi: 10.1007/s10709-007-9145-6.

Waits, Lisette, P. and Paetkau, D. (2005) ‘Noninvasive Genetic Sampling Tools for Wildlife Biologists: a Review of Applications and Recommendations for Accurate Data Collection’, *Journal of Wildlife Management*, 69(4), pp. 1419–1433. doi: 10.2193/0022-541x(2005)69[1419:ngstfw]2.0.co;2.

Walker, M. J. C. *et al.* (2012) ‘Formal subdivision of the Holocene Series/Epoch: a Discussion Paper by a Working Group of INTIMATE (Integration of ice-core, marine and terrestrial records) and the Subcommission on Quaternary Stratigraphy (International Commission on Stratigraphy)', *Journal of Quaternary Science*, 27(7), pp. 649–659. doi: 10.1002/jqs.2565.

Walsh, P. S., Metzger, D. A. and Higuchi, R. (1991) 'Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material.', *BioTechniques*, 10(4), pp. 506–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1867860.

Wang, Z. and Storm, D. R. (2006) 'Extraction of DNA from mouse tails', *BioTechniques*, 41(4), pp. 410–412. doi: 10.2144/000112255.

Wiens, J. J. (2006) 'Missing data and the design of phylogenetic analyses', *Journal of Biomedical Informatics*, 39(1), pp. 34–42. doi: 10.1016/j.jbi.2005.04.001.

Yu, G. (2020) 'Using ggtree to Visualize Data on Tree-Like Structures', *Current Protocols in Bioinformatics*, 69(1). doi: 10.1002/cpbi.96.

Zhu, L. *et al.* (2019) 'The complete mitochondrial genome of Microtus fortis pelliceus (Arvicolinae, Rodentia) from China and its phylogenetic analysis', *Mitochondrial DNA Part B: Resources*. Taylor & Francis, 4(1), pp. 2039–2041. doi: 10.1080/23802359.2019.1618212.

Zwickl, D. J. and Hillis, D. M. (2002) 'Increased Taxon Sampling Greatly Reduces Phylogenetic Error', *Systematic Biology*. Edited by K. Crandall, 51(4), pp. 588–598. doi: 10.1080/10635150290102339.

Appendix:

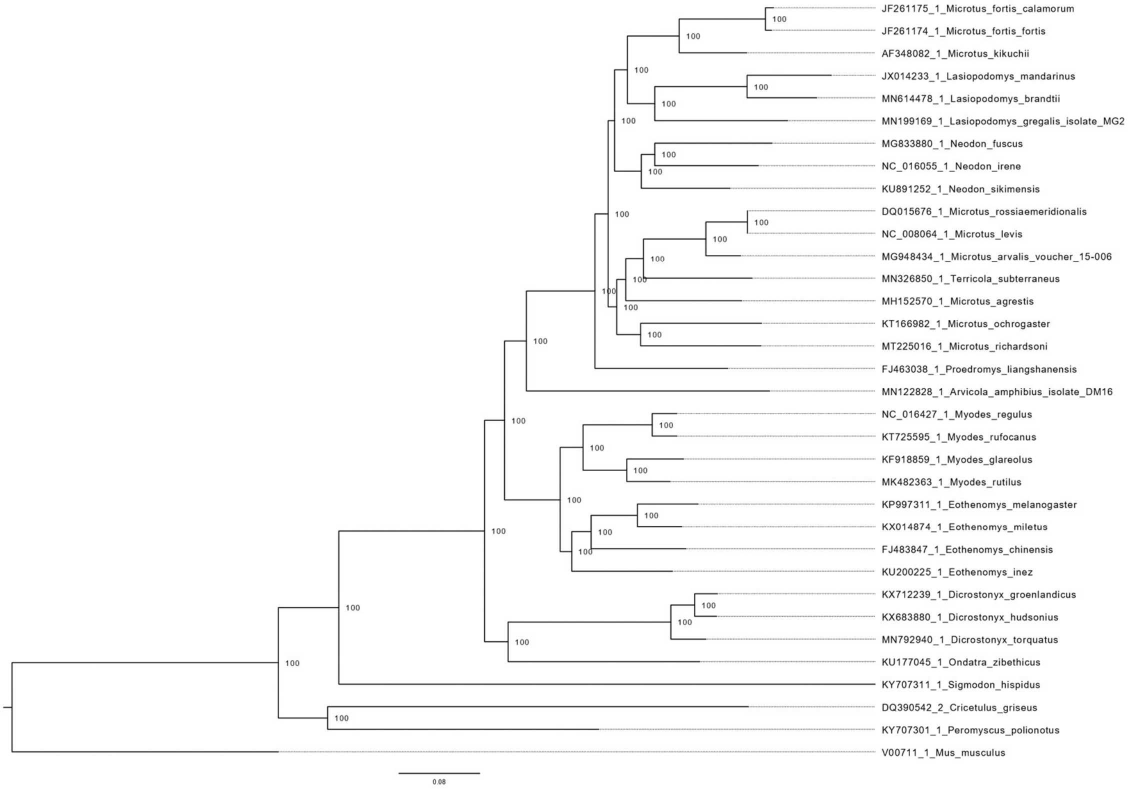

JF261175_1_Microtus_fortis_calamorum
100
JF261174_1_Microtus_fortis_fortis
100
AF348082_1_Microtus_kikuchii
100
JX014233_1_Lasiopodomys_mandarinus
100
MN614478_1_Lasiopodomys_brandtii
100
100
MN199169_1_Lasiopodomys_gregalis_isolate_MG2
MG833880_1_Neodon_fuscus
100
NC_016055_1_Neodon_irene
100
KU891252_1_Neodon_sikimensis
100
DQ015676_1_Microtus_rossiaemeridionalis
100
NC_008064_1_Microtus_levis
100
MG948434_1_Microtus_arvalis_voucher_15-006
100
100
MN326850_1_Terricola_subterraneus
100
100
MH152570_1_Microtus_agrestis
KT166982_1_Microtus_ochrogaster
100
100
MT225016_1_Microtus_richardsoni
FJ463038_1_Proedromys_liangshanensis
MN122828_1_Arvicola_amphibius_isolate_DM16
NC_016427_1_Myodes_regulus
100
100
KT725595_1_Myodes_rufocanus
100
KF918859_1_Myodes_glareolus
100
MK482363_1_Myodes_rutilus
100
KP997311_1_Eothenomys_melanogaster
100
KX014874_1_Eothenomys_miletus
100
100
100
FJ483847_1_Eothenomys_chinensis
KU200225_1_Eothenomys_inez
KX712239_1_Dicrostonyx_groenlandicus
100
100
KX683880_1_Dicrostonyx_hudsonius
100
MN792940_1_Dicrostonyx_torquatus
100
KU177045_1_Ondatra_zibethicus
100
KY707311_1_Sigmodon_hispidus
DQ390542_2_Cricetulus_griseus
100
KY707301_1_Peromyscus_polionotus
V00711_1_Mus_musculus
0.08

Figure A1: Bayesian inference phylogenetic tree of Arvicolinae mitogenomes. Posterior probabilities shown as percentages.

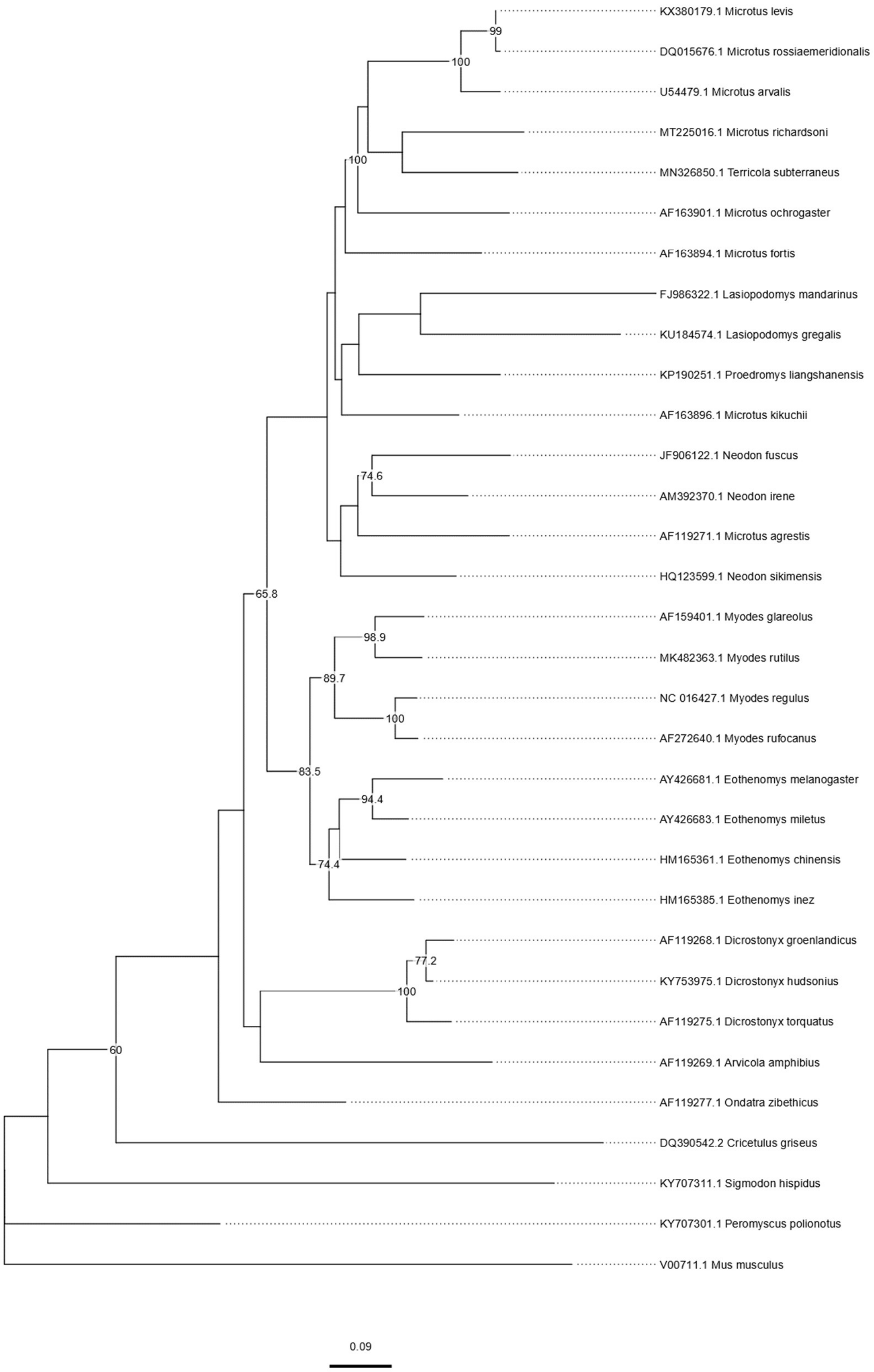

KX380179.1 Microtus levis
DQ015676.1 Microtus rossiaemeridionalis
U54479.1 Microtus arvalis
MT225016.1 Microtus richardsoni
MN326850.1 Terricola subterraneus
AF163901.1 Microtus ochrogaster
AF163894.1 Microtus fortis
FJ986322.1 Lasiopodomys mandarinus
KU184574.1 Lasiopodomys gregalis
KP190251.1 Proedromys liangshanensis
AF163896.1 Microtus kikuchii
JF906122.1 Neodon fuscus
AM392370.1 Neodon irene
AF119271.1 Microtus agrestis
HQ123599.1 Neodon sikimensis
AF159401.1 Myodes glareolus
MK482363.1 Myodes rutilus
NC 016427.1 Myodes regulus
AF272640.1 Myodes rufocanus
AY426681.1 Eothenomys melanogaster
AY426683.1 Eothenomys miletus
HM165361.1 Eothenomys chinensis
HM165385.1 Eothenomys inez
AF119268.1 Dicrostonyx groenlandicus
KY753975.1 Dicrostonyx hudsonius
AF119275.1 Dicrostonyx torquatus
AF119269.1 Arvicola amphibius
AF119277.1 Ondatra zibethicus
DQ390542.2 Cricetulus griseus
KY707311.1 Sigmodon hispidus
KY707301.1 Peromyscus polionotus
V00711.1 Mus musculus
99
100
100
74.6
65.8
98.9
89.7
100
83.5
94.4
74.4
77.2
100
60
0.09

Figure A2: ML phylogenetic tree of selected taxa for the Cytb gene. Bootstrap score shown.

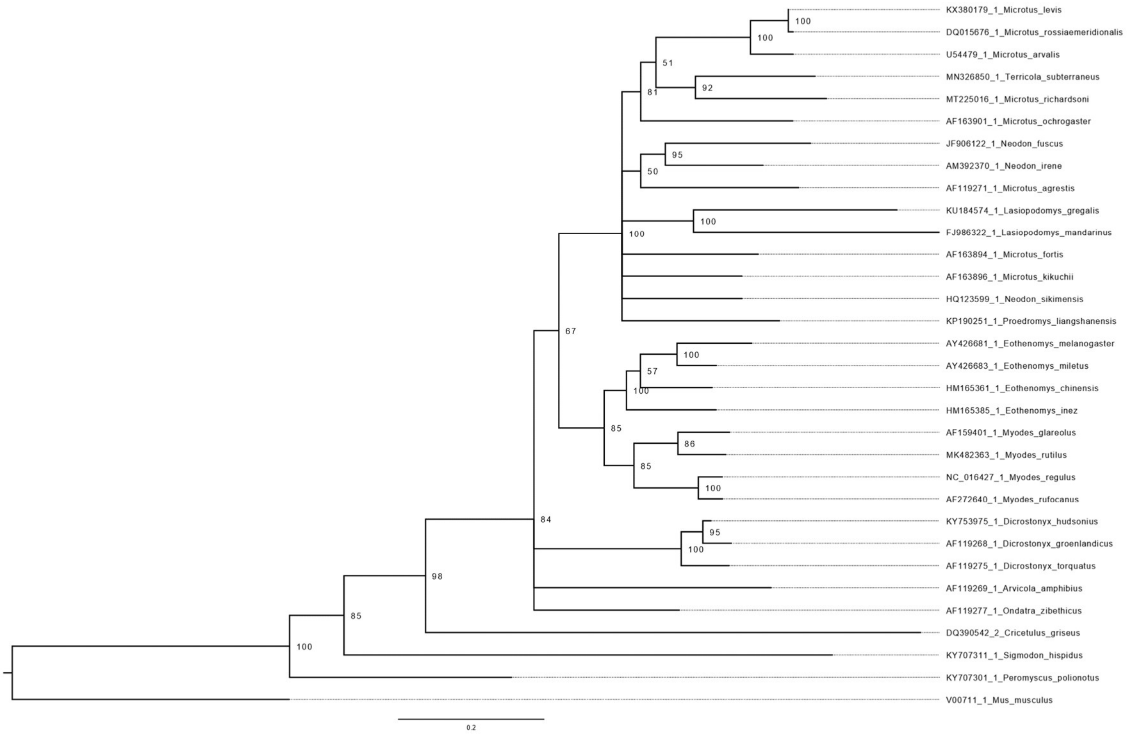

KX380179_1_Microtus_levis
100
DQ015676_1_Microtus_rossiaemeridionalis
100
U54479_1_Microtus_arvalis
51
MN326850_1_Terricola_subterraneus
92
81
MT225016_1_Microtus_richardsoni
AF163901_1_Microtus_ochrogaster
JF906122_1_Neodon_fuscus
95
AM392370_1_Neodon_irene
50
AF119271_1_Microtus_agrestis
KU184574_1_Lasiopodomys_gregalis
100
100
FJ986322_1_Lasiopodomys_mandarinus
AF163894_1_Microtus_fortis
AF163896_1_Microtus_kikuchii
HQ123599_1_Neodon_sikimensis
KP190251_1_Proedromys_liangshanensis
67
AY426681_1_Eothenomys_melanogaster
100
AY426683_1_Eothenomys_miletus
57
100
HM165361_1_Eothenomys_chinensis
HM165385_1_Eothenomys_inez
85
AF159401_1_Myodes_glareolus
86
MK482363_1_Myodes_rutilus
85
NC_016427_1_Myodes_regulus
100
AF272640_1_Myodes_rufocanus
84
KY753975_1_Dicrostonyx_hudsonius
95
AF119268_1_Dicrostonyx_groenlandicus
100
AF119275_1_Dicrostonyx_torquatus
98
AF119269_1_Arvicola_amphibius
AF119277_1_Ondatra_zibethicus
85
DQ390542_2_Cricetulus_griseus
100
KY707311_1_Sigmodon_hispidus
KY707301_1_Peromyscus_polionotus
V00711_1_Mus_musculus
0.2

Figure A3: Bayesian inference phylogenetic tree of selected Arvicolinae taxa for Cytb gene.

Posterior probabilities shown as percentages.

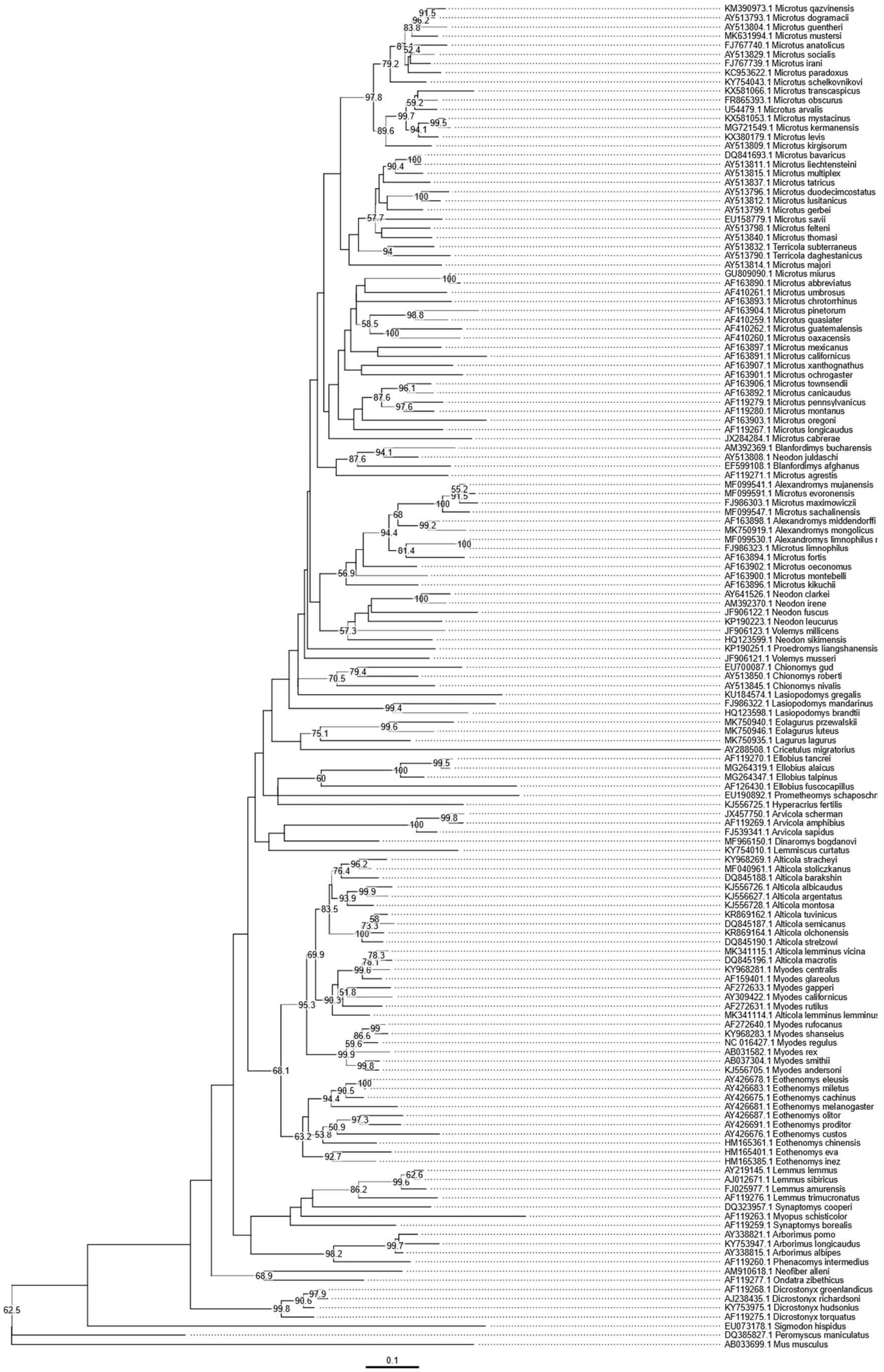

KM390973.1 Microtus qazvinensis
AY513793.1 Microtus dogramacii
AY513804.1 Microtus guentheri
MK631994.1 Microtus mustersi
FJ767740.1 Microtus anatolicus
AY513829.1 Microtus socialis
FJ767739.1 Microtus irani
KC953622.1 Microtus paradoxus
KY754043.1 Microtus schelkovnikovi
KX581066.1 Microtus transcaspicus
FR865393.1 Microtus obscurus
U54479.1 Microtus arvalis
KX581053.1 Microtus mystacinus
MG721549.1 Microtus kermanensis
KX380179.1 Microtus levis
AY513809.1 Microtus kirgisorum
DQ841693.1 Microtus bavaricus
AY513811.1 Microtus liechtensteini
AY513815.1 Microtus multiplex
AY513837.1 Microtus tatricus
AY513796.1 Microtus duodecimcostatus
AY513812.1 Microtus lusitanicus
AY513799.1 Microtus gerbei
EU158779.1 Microtus savii
AY513798.1 Microtus felteni
AY513840.1 Microtus thomasi
AY513832.1 Terricola subterraneus
AY513790.1 Terricola daghestanicus
AY513814.1 Microtus majori
GU809090.1 Microtus miurus
AF163890.1 Microtus abbreviatus
AF410261.1 Microtus umbrosus
AF163893.1 Microtus chrotorrhinus
AF163904.1 Microtus pinetorum
AF410259.1 Microtus quasiater
AF410262.1 Microtus guatemalensis
AF410260.1 Microtus oaxacensis
AF163897.1 Microtus mexicanus
AF163891.1 Microtus californicus
AF163907.1 Microtus xanthognathus
AF163901.1 Microtus ochrogaster
AF163906.1 Microtus townsendii
AF163892.1 Microtus canicaudus
AF119279.1 Microtus pennsylvanicus
AF119280.1 Microtus montanus
AF163903.1 Microtus oregoni
AF119267.1 Microtus longicaudus
JX284284.1 Microtus cabrerae
AM392369.1 Blanfordimys bucharensis
AY513808.1 Neodon juldaschi
EF599108.1 Blanfordimys afghanus
AF119271.1 Microtus agrestis
MF099541.1 Alexandromys mujanensis
MF099591.1 Microtus evoronensis
FJ986303.1 Microtus maximowiczii
MF099547.1 Microtus sachalinensis
AF163898.1 Alexandromys middendorffi
MK750919.1 Alexandromys mongolicus
MF099530.1 Alexandromys limnophilus r
FJ986323.1 Microtus limnophilus
AF163894.1 Microtus fortis
AF163902.1 Microtus oeconomus
AF163900.1 Microtus montebelli
AF163896.1 Microtus kikuchii
AY641526.1 Neodon clarkei
AM392370.1 Neodon irene
JF906122.1 Neodon fuscus
KP190223.1 Neodon leucurus
JF906123.1 Volemys millicens
HQ123599.1 Neodon sikimensis
KP190251.1 Proedromys liangshanensis
JF906121.1 Volemys musseri
EU700087.1 Chionomys gud
AY513850.1 Chionomys roberti
AY513845.1 Chionomys nivalis
KU184574.1 Lasiopodomys gregalis
FJ986322.1 Lasiopodomys mandarinus
HQ123598.1 Lasiopodomys brandtii
MK750940.1 Eolagurus przewalskii
MK750946.1 Eolagurus luteus
MK750935.1 Lagurus lagurus
AY288508.1 Cricetulus migratorius
AF119270.1 Ellobius tancrei
MG264319.1 Ellobius alaicus
MG264347.1 Ellobius talpinus
AF126430.1 Ellobius fuscocapillus
EU190892.1 Prometheomys schaposchn
KJ556725.1 Hyperacrius fertilis
JX457750.1 Arvicola scherman
AF119269.1 Arvicola amphibius
FJ539341.1 Arvicola sapidus
MF966150.1 Dinaromys bogdanovi
KY754010.1 Lemmiscus curtatus
KY968269.1 Alticola stracheyi
MF040961.1 Alticola stoliczkanus
DQ845188.1 Alticola barakshin
KJ556726.1 Alticola albicaudus
KJ556627.1 Alticola argentatus
KJ556728.1 Alticola montosa
KR869162.1 Alticola tuvinicus
DQ845187.1 Alticola semicanus
KR869164.1 Alticola olchonensis
DQ845190.1 Alticola strelzowi
MK341115.1 Alticola lemminus vicina
DQ845196.1 Alticola macrotis
KY968281.1 Myodes centralis
AF159401.1 Myodes glareolus
AF272633.1 Myodes gapperi
AY309422.1 Myodes californicus
AF272631.1 Myodes rutilus
MK341114.1 Alticola lemminus lemminus
AF272640.1 Myodes rufocanus
KY968283.1 Myodes shanseius
NC 016427.1 Myodes regulus
AB031582.1 Myodes rex
AB037304.1 Myodes smithii
KJ556705.1 Myodes andersoni
AY426678.1 Eothenomys eleusis
AY426683.1 Eothenomys miletus
AY426675.1 Eothenomys cachinus
AY426681.1 Eothenomys melanogaster
AY426687.1 Eothenomys olitor
AY426691.1 Eothenomys proditor
AY426676.1 Eothenomys custos
HM165361.1 Eothenomys chinensis
HM165401.1 Eothenomys eva
HM165385.1 Eothenomys inez
AY219145.1 Lemmus lemmus
AJ012671.1 Lemmus sibiricus
FJ025977.1 Lemmus amurensis
AF119276.1 Lemmus trimucronatus
DQ323957.1 Synaptomys cooperi
AF119263.1 Myopus schisticolor
AF119259.1 Synaptomys borealis
AY338821.1 Arborimus pomo
KY753947.1 Arborimus longicaudus
AY338815.1 Arborimus albipes
AF119260.1 Phenacomys intermedius
AM910618.1 Neofiber alleni
AF119277.1 Ondatra zibethicus
AF119268.1 Dicrostonyx groenlandicus
AJ238435.1 Dicrostonyx richardsoni
KY753975.1 Dicrostonyx hudsonius
AF119275.1 Dicrostonyx torquatus
EU073178.1 Sigmodon hispidus
DQ385827.1 Peromyscus maniculatus
AB033699.1 Mus musculus
91.5
96.2
83.8
87
52.4
79.2
97.8
59.2
99.7
99.5
94.1
89.6
100
90.4
100
57.7
94
100
98.8
58.5
100
96.1
87.6
97.6
94.1
87.6
55.2
91.5
100
68
99.2
94.4
100
81.4
56.9
100
57.3
79.4
70.5
99.4
99.6
75.1
99.5
100
60
99.8
100
96.2
76.4
99.9
93.9
83.5
58
73.3
100
69.9
78.3
78.1
99.6
51.8
90.3
95.3
99
86.6
59.6
99.9
99.8
68.1
100
90.5
94.4
97.3
50.9
53.8
63.2
92.7
62.6
99.6
86.2
99.7
98.2
68.9
97.9
90.6
99.8
62.5
0.1

Figure A4: ML phylogenetic tree of all available Arvicolinae taxa for the Cytb gene. Bootstrap scores shown.

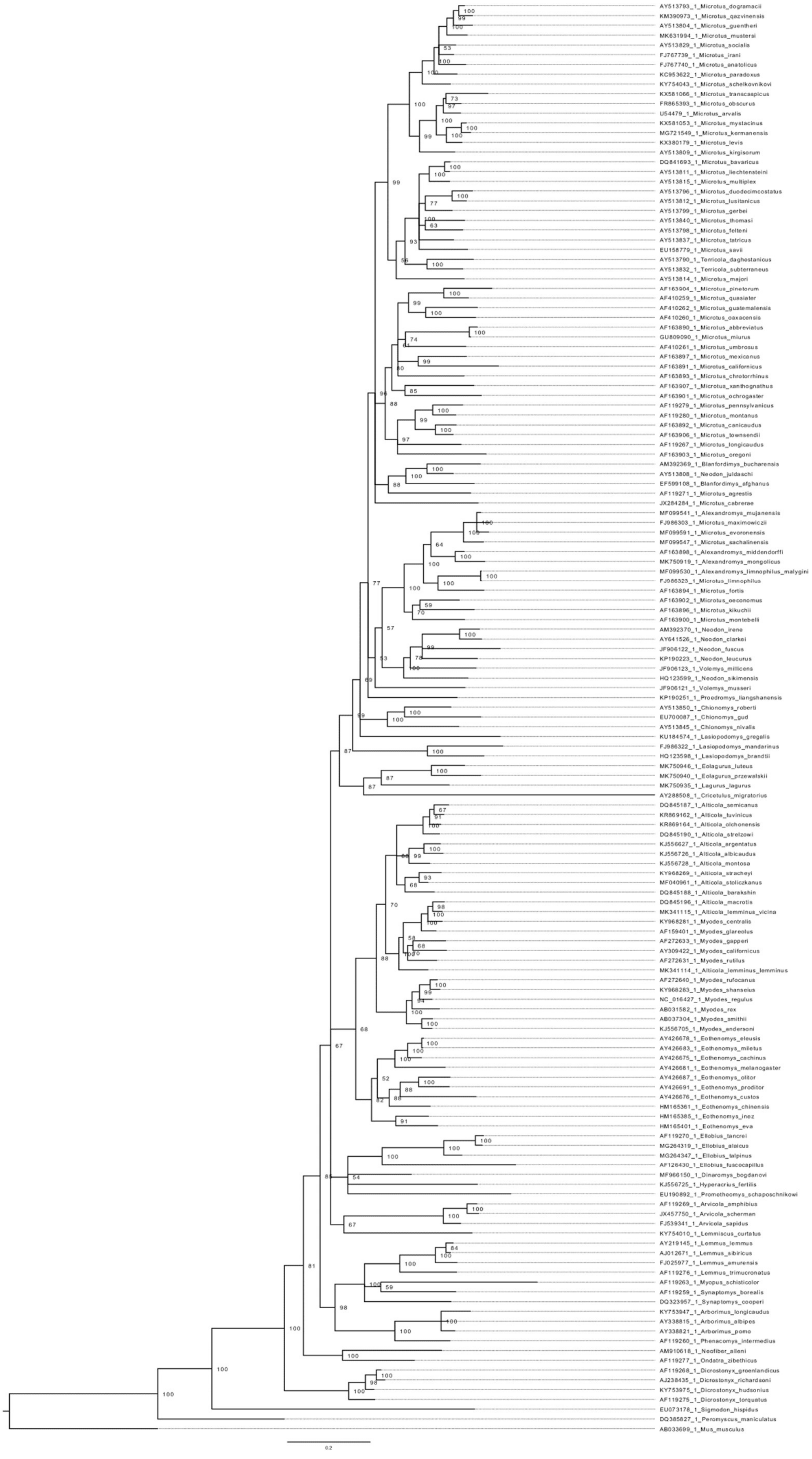

AY513793_1_Microtus_dogramacii
KM390973_1_Microtus_qazvinensis
AY513804_1_Microtus_guentheri
MK631994_1_Microtus_mustersi
AY513829_1_Microtus_socialis
FJ767739_1_Microtus_irani
FJ767740_1_Microtus_anatolicus
KC953622_1_Microtus_paradoxus
KY754043_1_Microtus_schelkovnikovi
KX581066_1_Microtus_transcaspicus
FR865393_1_Microtus_obscurus
U54479_1_Microtus_arvalis
KX581053_1_Microtus_mystacinus
MG721549_1_Microtus_kermanensis
KX380179_1_Microtus_levis
AY513809_1_Microtus_kirgisorum
DQ841693_1_Microtus_bavaricus
AY513811_1_Microtus_liechtensteini
AY513815_1_Microtus_multiplex
AY513796_1_Microtus_duodecimcostatus
AY513812_1_Microtus_lusitanicus
AY513799_1_Microtus_gerbei
AY513840_1_Microtus_thomasi
AY513798_1_Microtus_felteni
AY513837_1_Microtus_tatricus
EU158779_1_Microtus_savii
AY513790_1_Terricola_daghestanicus
AY513832_1_Terricola_subterraneus
AY513814_1_Microtus_majori
AF163904_1_Microtus_pinetorum
AF410259_1_Microtus_quasiater
AF410262_1_Microtus_guatemalensis
AF410260_1_Microtus_oaxacensis
AF163890_1_Microtus_abbreviatus
GU809090_1_Microtus_miurus
AF410261_1_Microtus_umbrosus
AF163897_1_Microtus_mexicanus
AF163891_1_Microtus_californicus
AF163893_1_Microtus_chrotorrhinus
AF163907_1_Microtus_xanthognathus
AF163901_1_Microtus_ochrogaster
AF119279_1_Microtus_pennsylvanicus
AF119280_1_Microtus_montanus
AF163892_1_Microtus_canicaudus
AF163906_1_Microtus_townsendii
AF119267_1_Microtus_longicaudus
AF163903_1_Microtus_oregoni
AM392369_1_Blanfordimys_bucharensis
AY513808_1_Neodon_juldaschi
EF599108_1_Blanfordimys_afghanus
AF119271_1_Microtus_agrestis
JX284284_1_Microtus_cabrerae
MF099541_1_Alexandromys_mujanensis
FJ986303_1_Microtus_maximowiczii
MF099591_1_Microtus_evoronensis
MF099547_1_Microtus_sachalinensis
AF163898_1_Alexandromys_middendorffi
MK750919_1_Alexandromys_mongolicus
MF099530_1_Alexandromys_limnophilus_malygini
FJ986323_1_Microtus_limnophilus
AF163894_1_Microtus_fortis
AF163902_1_Microtus_oeconomus
AF163896_1_Microtus_kikuchii
AF163900_1_Microtus_montebelli
AM392370_1_Neodon_irene
AY641526_1_Neodon_clarkei
JF906122_1_Neodon_fuscus
KP190223_1_Neodon_leucurus
JF906123_1_Volemys_millicens
HQ123599_1_Neodon_sikimensis
JF906121_1_Volemys_musseri
KP190251_1_Proedromys_liangshanensis
AY513850_1_Chionomys_roberti
EU700087_1_Chionomys_gud
AY513845_1_Chionomys_nivalis
KU184574_1_Lasiopodomys_gregalis
FJ986322_1_Lasiopodomys_mandarinus
HQ123598_1_Lasiopodomys_brandtii
MK750946_1_Eolagurus_luteus
MK750940_1_Eolagurus_przewalskii
MK750935_1_Lagurus_lagurus
AY288508_1_Cricetulus_migratorius
DQ845187_1_Alticola_semicanus
KR869162_1_Alticola_tuvinicus
KR869164_1_Alticola_olchonensis
DQ845190_1_Alticola_strelzowi
KJ556627_1_Alticola_argentatus
KJ556726_1_Alticola_albicaudus
KJ556728_1_Alticola_montosa
KY968269_1_Alticola_stracheyi
MF040961_1_Alticola_stoliczkanus
DQ845188_1_Alticola_barakshin
DQ845196_1_Alticola_macrotis
MK341115_1_Alticola_lemminus_vicina
KY968281_1_Myodes_centralis
AF159401_1_Myodes_glareolus
AF272633_1_Myodes_gapperi
AY309422_1_Myodes_californicus
AF272631_1_Myodes_rutilus
MK341114_1_Alticola_lemminus_lemminus
AF272640_1_Myodes_rufocanus
KY968283_1_Myodes_shanseius
NC_016427_1_Myodes_regulus
AB031582_1_Myodes_rex
AB037304_1_Myodes_smithii
KJ556705_1_Myodes_andersoni
AY426678_1_Eothenomys_eleusis
AY426683_1_Eothenomys_miletus
AY426675_1_Eothenomys_cachinus
AY426681_1_Eothenomys_melanogaster
AY426687_1_Eothenomys_olitor
AY426691_1_Eothenomys_proditor
AY426676_1_Eothenomys_custos
HM165361_1_Eothenomys_chinensis
HM165385_1_Eothenomys_inez
HM165401_1_Eothenomys_eva
AF119270_1_Ellobius_tancrei
MG264319_1_Ellobius_alaicus
MG264347_1_Ellobius_talpinus
AF126430_1_Ellobius_fuscocapillus
MF966150_1_Dinaromys_bogdanovi
KJ556725_1_Hyperacrius_fertilis
EU190892_1_Prometheomys_schaposchnikowi
AF119269_1_Arvicola_amphibius
JX457750_1_Arvicola_scherman
FJ539341_1_Arvicola_sapidus
KY754010_1_Lemmiscus_curtatus
AY219145_1_Lemmus_lemmus
AJ012671_1_Lemmus_sibiricus
FJ025977_1_Lemmus_amurensis
AF119276_1_Lemmus_trimucronatus
AF119263_1_Myopus_schisticolor
AF119259_1_Synaptomys_borealis
DQ323957_1_Synaptomys_cooperi
KY753947_1_Arborimus_longicaudus
AY338815_1_Arborimus_albipes
AY338821_1_Arborimus_pomo
AF119260_1_Phenacomys_intermedius
AM910618_1_Neofiber_alleni
AF119277_1_Ondatra_zibethicus
AF119268_1_Dicrostonyx_groenlandicus
AJ238435_1_Dicrostonyx_richardsoni
KY753975_1_Dicrostonyx_hudsonius
AF119275_1_Dicrostonyx_torquatus
EU073178_1_Sigmodon_hispidus
DQ385827_1_Peromyscus_maniculatus
AB033699_1_Mus_musculus
0.2

Figure A5: Bayesian inference phylogenetic tree of all available Arvicolinae taxa for Cytb gene. Posterior probabilities shown as percentages.

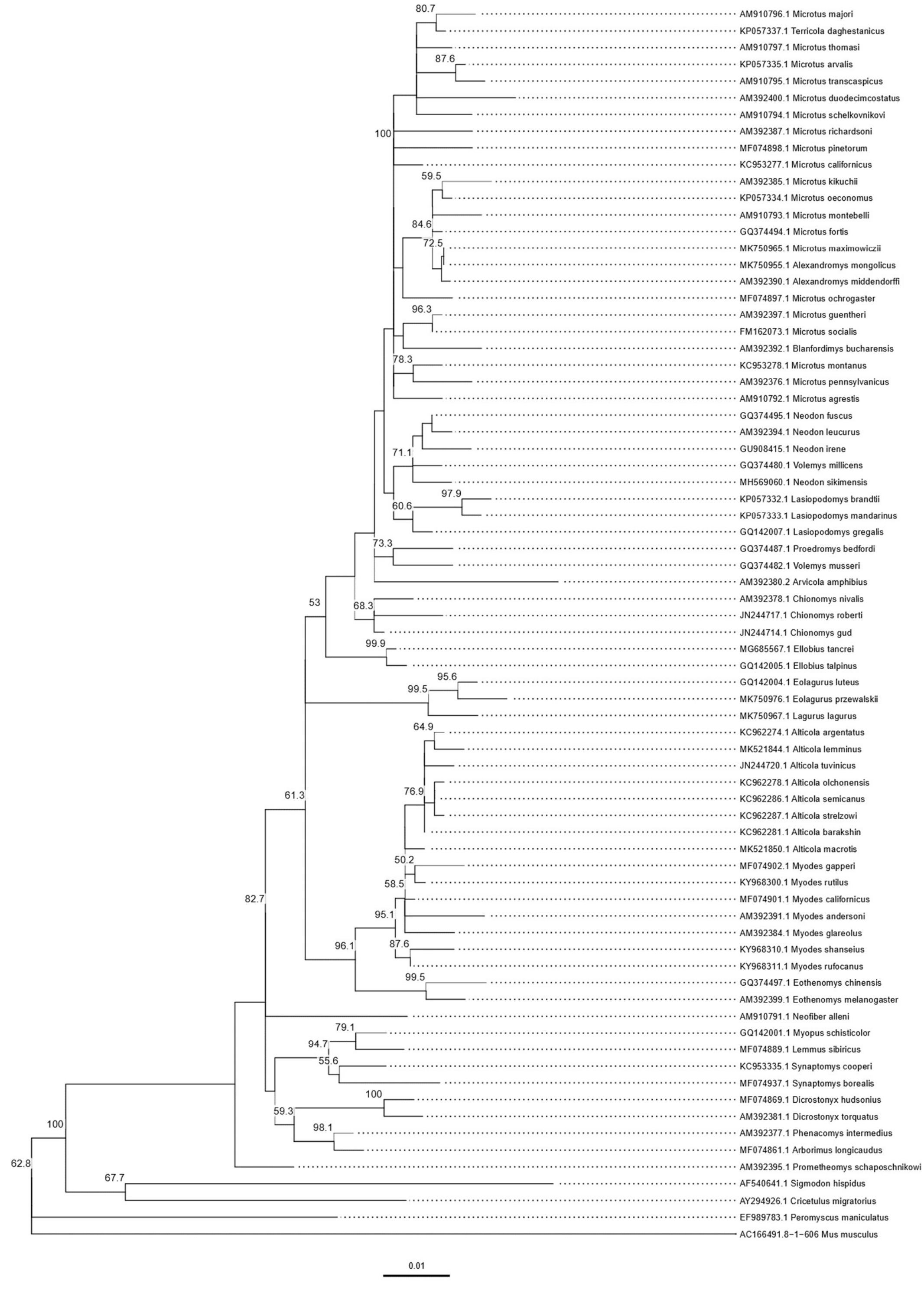

80.7
AM910796.1 Microtus majori
KP057337.1 Terricola daghestanicus
AM910797.1 Microtus thomasi
87.6
KP057335.1 Microtus arvalis
AM910795.1 Microtus transcaspicus
AM392400.1 Microtus duodecimcostatus
AM910794.1 Microtus schelkovnikovi
100
AM392387.1 Microtus richardsoni
MF074898.1 Microtus pinetorum
KC953277.1 Microtus californicus
59.5
AM392385.1 Microtus kikuchii
KP057334.1 Microtus oeconomus
AM910793.1 Microtus montebelli
84.6
GQ374494.1 Microtus fortis
72.5
MK750965.1 Microtus maximowiczii
MK750955.1 Alexandromys mongolicus
AM392390.1 Alexandromys middendorffi
MF074897.1 Microtus ochrogaster
96.3
AM392397.1 Microtus guentheri
FM162073.1 Microtus socialis
AM392392.1 Blanfordimys bucharensis
78.3
KC953278.1 Microtus montanus
AM392376.1 Microtus pennsylvanicus
AM910792.1 Microtus agrestis
GQ374495.1 Neodon fuscus
AM392394.1 Neodon leucurus
GU908415.1 Neodon irene
71.1
GQ374480.1 Volemys millicens
MH569060.1 Neodon sikimensis
97.9
KP057332.1 Lasiopodomys brandtii
60.6
KP057333.1 Lasiopodomys mandarinus
GQ142007.1 Lasiopodomys gregalis
73.3
GQ374487.1 Proedromys bedfordi
GQ374482.1 Volemys musseri
AM392380.2 Arvicola amphibius
53
AM392378.1 Chionomys nivalis
68.3
JN244717.1 Chionomys roberti
JN244714.1 Chionomys gud
99.9
MG685567.1 Ellobius tancrei
GQ142005.1 Ellobius talpinus
95.6
GQ142004.1 Eolagurus luteus
99.5
MK750976.1 Eolagurus przewalskii
MK750967.1 Lagurus lagurus
64.9
KC962274.1 Alticola argentatus
MK521844.1 Alticola lemminus
JN244720.1 Alticola tuvinicus
61.3
KC962278.1 Alticola olchonensis
76.9
KC962286.1 Alticola semicanus
KC962287.1 Alticola strelzowi
KC962281.1 Alticola barakshin
MK521850.1 Alticola macrotis
50.2
MF074902.1 Myodes gapperi
KY968300.1 Myodes rutilus
58.5
82.7
MF074901.1 Myodes californicus
95.1
AM392391.1 Myodes andersoni
AM392384.1 Myodes glareolus
96.1
87.6
KY968310.1 Myodes shanseius
KY968311.1 Myodes rufocanus
99.5
GQ374497.1 Eothenomys chinensis
AM392399.1 Eothenomys melanogaster
AM910791.1 Neofiber alleni
79.1
GQ142001.1 Myopus schisticolor
94.7
MF074889.1 Lemmus sibiricus
55.6
KC953335.1 Synaptomys cooperi
MF074937.1 Synaptomys borealis
100
MF074869.1 Dicrostonyx hudsonius
59.3
AM392381.1 Dicrostonyx torquatus
100
98.1
AM392377.1 Phenacomys intermedius
MF074861.1 Arborimus longicaudus
62.8
AM392395.1 Prometheomys schaposchnikowi
67.7
AF540641.1 Sigmodon hispidus
AY294926.1 Cricetulus migratorius
EF989783.1 Peromyscus maniculatus
AC166491.8-1-606 Mus musculus
0.01

Figure A6: ML phylogenetic tree of Arvicolinae for GHR gene.

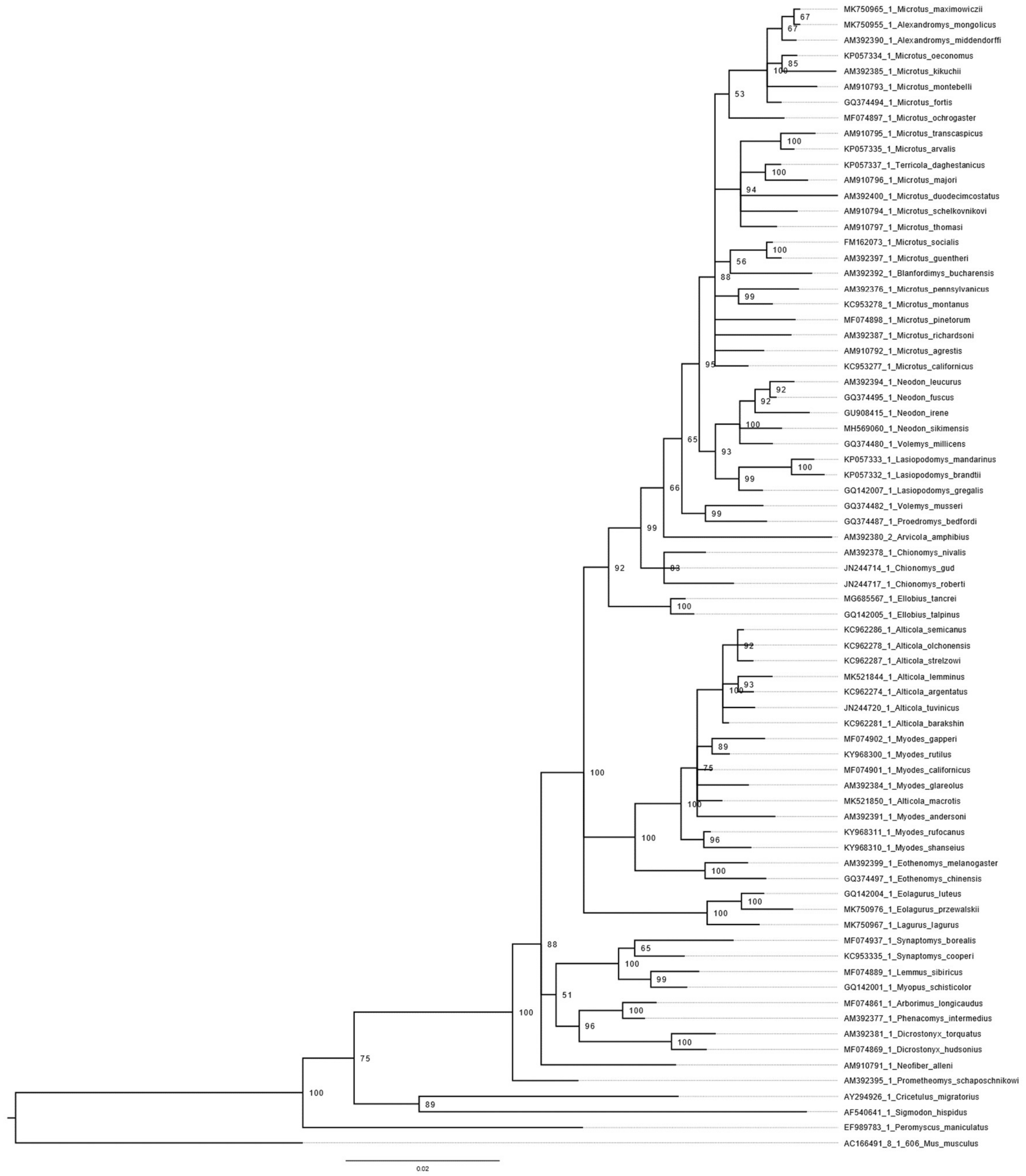

MK750965_1_Microtus_maximowiczii
MK750955_1_Alexandromys_mongolicus
AM392390_1_Alexandromys_middendorffi
KP057334_1_Microtus_oeconomus
AM392385_1_Microtus_kikuchii
AM910793_1_Microtus_montebelli
GQ374494_1_Microtus_fortis
MF074897_1_Microtus_ochrogaster
AM910795_1_Microtus_transcaspicus
KP057335_1_Microtus_arvalis
KP057337_1_Terricola_daghestanicus
AM910796_1_Microtus_majori
AM392400_1_Microtus_duodecimcostatus
AM910794_1_Microtus_schelkovnikovi
AM910797_1_Microtus_thomasi
FM162073_1_Microtus_socialis
AM392397_1_Microtus_guentheri
AM392392_1_Blanfordimys_bucharensis
AM392376_1_Microtus_pennsylvanicus
KC953278_1_Microtus_montanus
MF074898_1_Microtus_pinetorum
AM392387_1_Microtus_richardsoni
AM910792_1_Microtus_agrestis
KC953277_1_Microtus_californicus
AM392394_1_Neodon_leucurus
GQ374495_1_Neodon_fuscus
GU908415_1_Neodon_irene
MH569060_1_Neodon_sikimensis
GQ374480_1_Volemys_millicens
KP057333_1_Lasiopodomys_mandarinus
KP057332_1_Lasiopodomys_brandtii
GQ142007_1_Lasiopodomys_gregalis
GQ374482_1_Volemys_musseri
GQ374487_1_Proedromys_bedfordi
AM392380_2_Arvicola_amphibius
AM392378_1_Chionomys_nivalis
JN244714_1_Chionomys_gud
JN244717_1_Chionomys_roberti
MG685567_1_Ellobius_tancrei
GQ142005_1_Ellobius_talpinus
KC962286_1_Alticola_semicanus
KC962278_1_Alticola_olchonensis
KC962287_1_Alticola_strelzowi
MK521844_1_Alticola_lemminus
KC962274_1_Alticola_argentatus
JN244720_1_Alticola_tuvinicus
KC962281_1_Alticola_barakshin
MF074902_1_Myodes_gapperi
KY968300_1_Myodes_rutilus
MF074901_1_Myodes_californicus
AM392384_1_Myodes_glareolus
MK521850_1_Alticola_macrotis
AM392391_1_Myodes_andersoni
KY968311_1_Myodes_rufocanus
KY968310_1_Myodes_shanseius
AM392399_1_Eothenomys_melanogaster
GQ374497_1_Eothenomys_chinensis
GQ142004_1_Eolagurus_luteus
MK750976_1_Eolagurus_przewalskii
MK750967_1_Lagurus_lagurus
MF074937_1_Synaptomys_borealis
KC953335_1_Synaptomys_cooperi
MF074889_1_Lemmus_sibiricus
GQ142001_1_Myopus_schisticolor
MF074861_1_Arborimus_longicaudus
AM392377_1_Phenacomys_intermedius
AM392381_1_Dicrostonyx_torquatus
MF074869_1_Dicrostonyx_hudsonius
AM910791_1_Neofiber_alleni
AM392395_1_Prometheomys_schaposchnikowi
AY294926_1_Cricetulus_migratorius
AF540641_1_Sigmodon_hispidus
EF989783_1_Peromyscus_maniculatus
AC166491_8_1_606_Mus_musculus
0.02

Figure A7: Bayesian inference phylogenetic tree of Arvicolinae for GHR gene. Posterior probabilities shown as percentages.

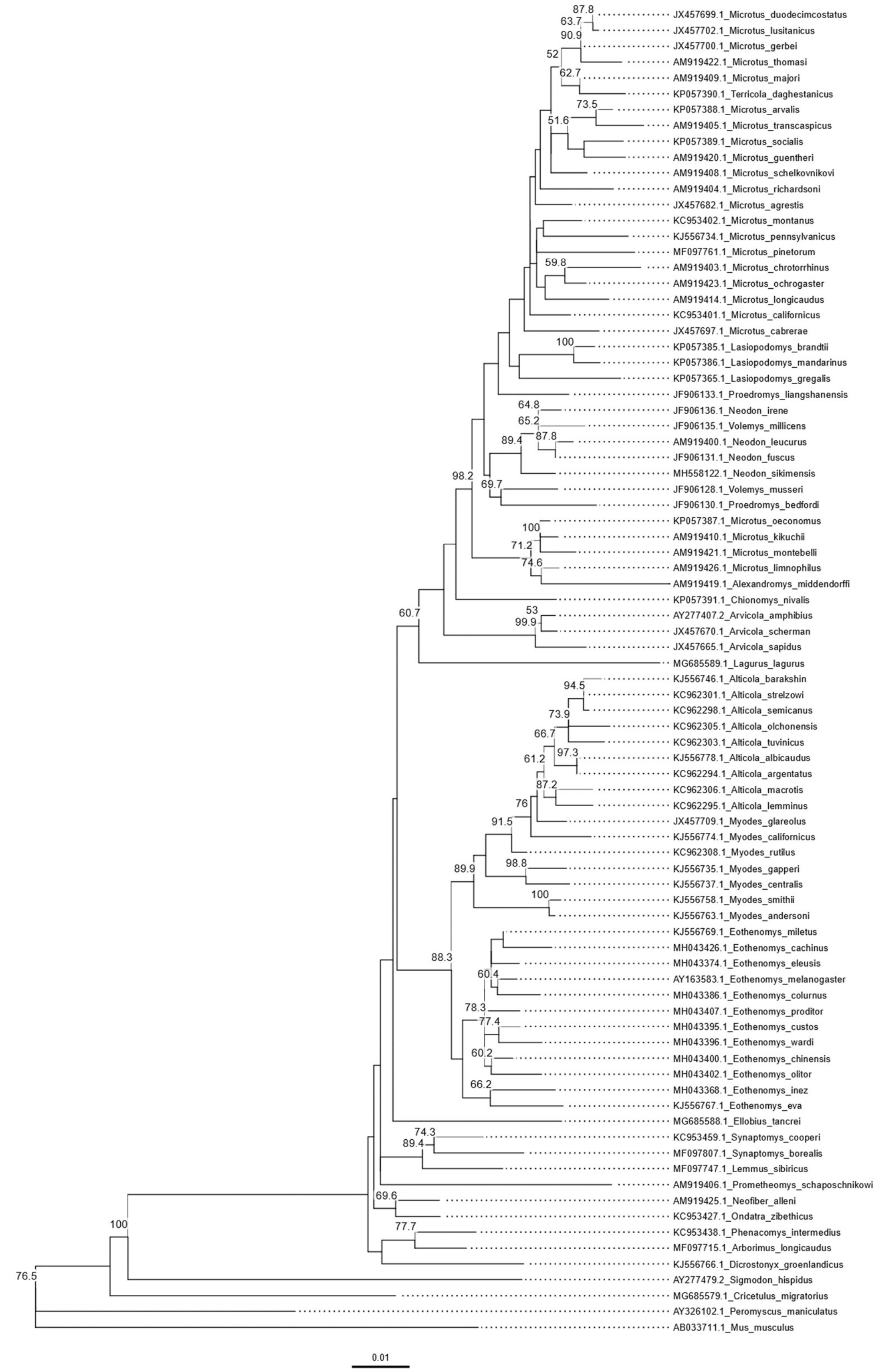

87.8
JX457699.1_Microtus_duodecimcostatus
63.7
JX457702.1_Microtus_lusitanicus
90.9
JX457700.1_Microtus_gerbei
52
AM919422.1_Microtus_thomasi
62.7
AM919409.1_Microtus_majori
KP057390.1_Terricola_daghestanicus
73.5
KP057388.1_Microtus_arvalis
51.6
AM919405.1_Microtus_transcaspicus
KP057389.1_Microtus_socialis
AM919420.1_Microtus_guentheri
AM919408.1_Microtus_schelkovnikovi
AM919404.1_Microtus_richardsoni
JX457682.1_Microtus_agrestis
KC953402.1_Microtus_montanus
KJ556734.1_Microtus_pennsylvanicus
MF097761.1_Microtus_pinetorum
59.8
AM919403.1_Microtus_chrotorrhinus
AM919423.1_Microtus_ochrogaster
AM919414.1_Microtus_longicaudus
KC953401.1_Microtus_californicus
JX457697.1_Microtus_cabrerae
100
KP057385.1_Lasiopodomys_brandtii
KP057386.1_Lasiopodomys_mandarinus
KP057365.1_Lasiopodomys_gregalis
JF906133.1_Proedromys_liangshanensis
64.8
JF906136.1_Neodon_irene
65.2
JF906135.1_Volemys_millicens
89.4
87.8
AM919400.1_Neodon_leucurus
JF906131.1_Neodon_fuscus
98.2
MH558122.1_Neodon_sikimensis
69.7
JF906128.1_Volemys_musseri
JF906130.1_Proedromys_bedfordi
KP057387.1_Microtus_oeconomus
100
AM919410.1_Microtus_kikuchii
71.2
AM919421.1_Microtus_montebelli
74.6
AM919426.1_Microtus_limnophilus
AM919419.1_Alexandromys_middendorffi
KP057391.1_Chionomys_nivalis
60.7
53
AY277407.2_Arvicola_amphibius
99.9
JX457670.1_Arvicola_scherman
JX457665.1_Arvicola_sapidus
MG685589.1_Lagurus_lagurus
KJ556746.1_Alticola_barakshin
94.5
KC962301.1_Alticola_strelzowi
KC962298.1_Alticola_semicanus
73.9
KC962305.1_Alticola_olchonensis
66.7
KC962303.1_Alticola_tuvinicus
61.2
97.3
KJ556778.1_Alticola_albicaudus
KC962294.1_Alticola_argentatus
87.2
KC962306.1_Alticola_macrotis
76
KC962295.1_Alticola_lemminus
91.5
JX457709.1_Myodes_glareolus
KJ556774.1_Myodes_californicus
KC962308.1_Myodes_rutilus
98.8
KJ556735.1_Myodes_gapperi
89.9
KJ556737.1_Myodes_centralis
100
KJ556758.1_Myodes_smithii
KJ556763.1_Myodes_andersoni
KJ556769.1_Eothenomys_miletus
MH043426.1_Eothenomys_cachinus
88.3
MH043374.1_Eothenomys_eleusis
60.4
AY163583.1_Eothenomys_melanogaster
MH043386.1_Eothenomys_colurnus
78.3
MH043407.1_Eothenomys_proditor
77.4
MH043395.1_Eothenomys_custos
MH043396.1_Eothenomys_wardi
60.2
MH043400.1_Eothenomys_chinensis
MH043402.1_Eothenomys_olitor
66.2
MH043368.1_Eothenomys_inez
KJ556767.1_Eothenomys_eva
MG685588.1_Ellobius_tancrei
74.3
KC953459.1_Synaptomys_cooperi
89.4
MF097807.1_Synaptomys_borealis
MF097747.1_Lemmus_sibiricus
AM919406.1_Prometheomys_schaposchnikowi
69.6
AM919425.1_Neofiber_alleni
KC953427.1_Ondatra_zibethicus
100
77.7
KC953438.1_Phenacomys_intermedius
MF097715.1_Arborimus_longicaudus
KJ556766.1_Dicrostonyx_groenlandicus
76.5
AY277479.2_Sigmodon_hispidus
MG685579.1_Cricetulus_migratorius
AY326102.1_Peromyscus_maniculatus
AB033711.1_Mus_musculus
0.01

Figure A8: ML phylogenetic tree of Arvicolinae for IRBP gene.

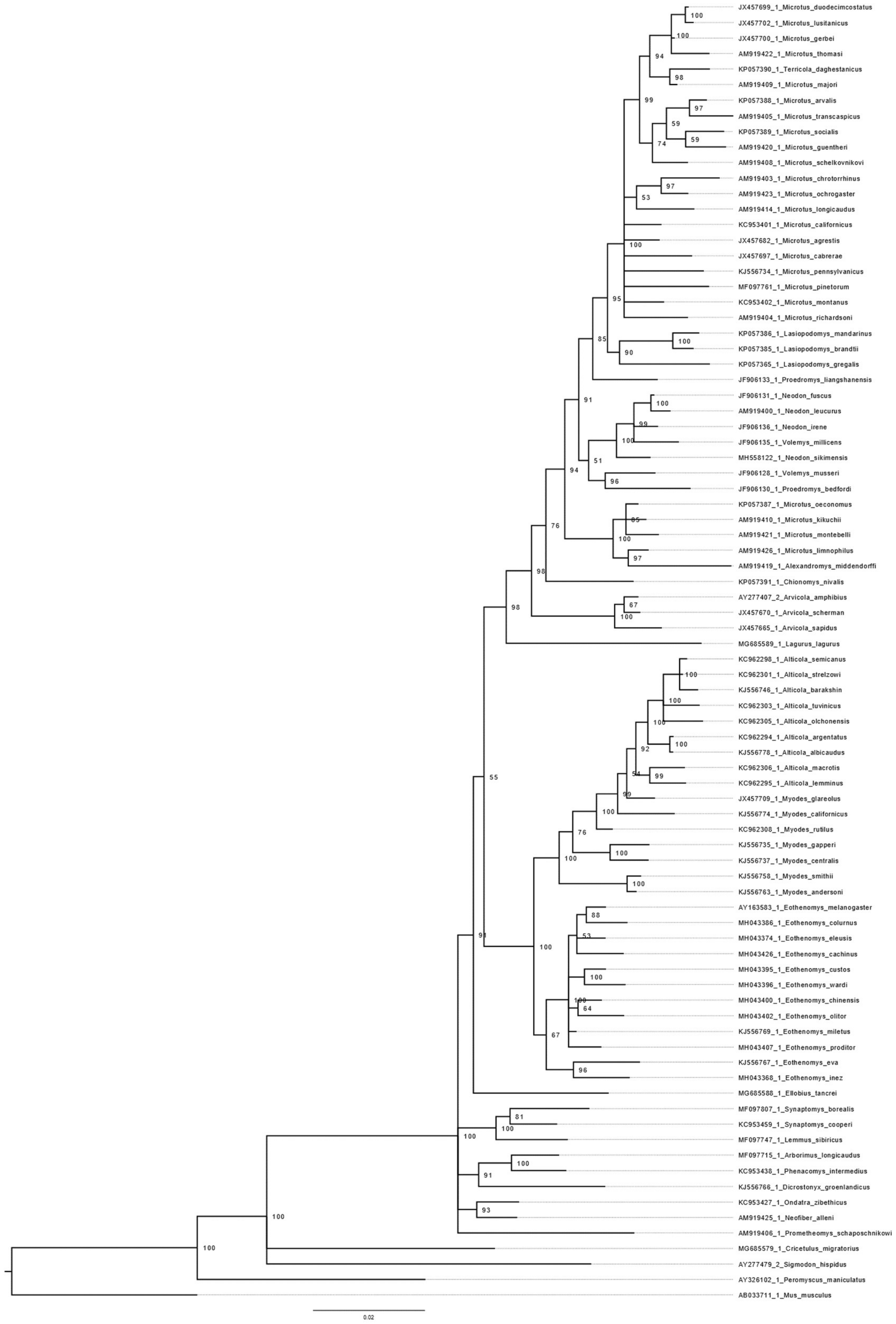

JX457699_1_Microtus_duodecimcostatus
JX457702_1_Microtus_lusitanicus
JX457700_1_Microtus_gerbei
AM919422_1_Microtus_thomasi
KP057390_1_Terricola_daghestanicus
AM919409_1_Microtus_majori
KP057388_1_Microtus_arvalis
AM919405_1_Microtus_transcaspicus
KP057389_1_Microtus_socialis
AM919420_1_Microtus_guentheri
AM919408_1_Microtus_schelkovnikovi
AM919403_1_Microtus_chrotorrhinus
AM919423_1_Microtus_ochrogaster
AM919414_1_Microtus_longicaudus
KC953401_1_Microtus_californicus
JX457682_1_Microtus_agrestis
JX457697_1_Microtus_cabrerae
KJ556734_1_Microtus_pennsylvanicus
MF097761_1_Microtus_pinetorum
KC953402_1_Microtus_montanus
AM919404_1_Microtus_richardsoni
KP057386_1_Lasiopodomys_mandarinus
KP057385_1_Lasiopodomys_brandtii
KP057365_1_Lasiopodomys_gregalis
JF906133_1_Proedromys_liangshanensis
JF906131_1_Neodon_fuscus
AM919400_1_Neodon_leucurus
JF906136_1_Neodon_irene
JF906135_1_Volemys_millicens
MH558122_1_Neodon_sikimensis
JF906128_1_Volemys_musseri
JF906130_1_Proedromys_bedfordi
KP057387_1_Microtus_oeconomus
AM919410_1_Microtus_kikuchii
AM919421_1_Microtus_montebelli
AM919426_1_Microtus_limnophilus
AM919419_1_Alexandromys_middendorffi
KP057391_1_Chionomys_nivalis
AY277407_2_Arvicola_amphibius
JX457670_1_Arvicola_scherman
JX457665_1_Arvicola_sapidus
MG685589_1_Lagurus_lagurus
KC962298_1_Alticola_semicanus
KC962301_1_Alticola_strelzowi
KJ556746_1_Alticola_barakshin
KC962303_1_Alticola_tuvinicus
KC962305_1_Alticola_olchonensis
KC962294_1_Alticola_argentatus
KJ556778_1_Alticola_albicaudus
KC962306_1_Alticola_macrotis
KC962295_1_Alticola_lemminus
JX457709_1_Myodes_glareolus
KJ556774_1_Myodes_californicus
KC962308_1_Myodes_rutilus
KJ556735_1_Myodes_gapperi
KJ556737_1_Myodes_centralis
KJ556758_1_Myodes_smithii
KJ556763_1_Myodes_andersoni
AY163583_1_Eothenomys_melanogaster
MH043386_1_Eothenomys_colurnus
MH043374_1_Eothenomys_eleusis
MH043426_1_Eothenomys_cachinus
MH043395_1_Eothenomys_custos
MH043396_1_Eothenomys_wardi
MH043400_1_Eothenomys_chinensis
MH043402_1_Eothenomys_olitor
KJ556769_1_Eothenomys_miletus
MH043407_1_Eothenomys_proditor
KJ556767_1_Eothenomys_eva
MH043368_1_Eothenomys_inez
MG685588_1_Ellobius_tancrei
MF097807_1_Synaptomys_borealis
KC953459_1_Synaptomys_cooperi
MF097747_1_Lemmus_sibiricus
MF097715_1_Arborimus_longicaudus
KC953438_1_Phenacomys_intermedius
KJ556766_1_Dicrostonyx_groenlandicus
KC953427_1_Ondatra_zibethicus
AM919425_1_Neofiber_alleni
AM919406_1_Prometheomys_schaposchnikowi
MG685579_1_Cricetulus_migratorius
AY277479_2_Sigmodon_hispidus
AY326102_1_Peromyscus_maniculatus
AB033711_1_Mus_musculus
100
100
94
98
99
97
59
59
74
97
53
100
95
100
85
90
91
100
99
100
51
94
96
76
85
100
97
98
67
98
100
55
100
100
100
100
92
54
99
99
100
76
100
100
100
88
91
53
100
100
100
64
67
96
81
100
100
100
91
93
100
100
0.02

Figure A9: Bayesian inference phylogenetic tree of Arvicolinae for IRBP gene. Posterior probabilities shown as percentages.

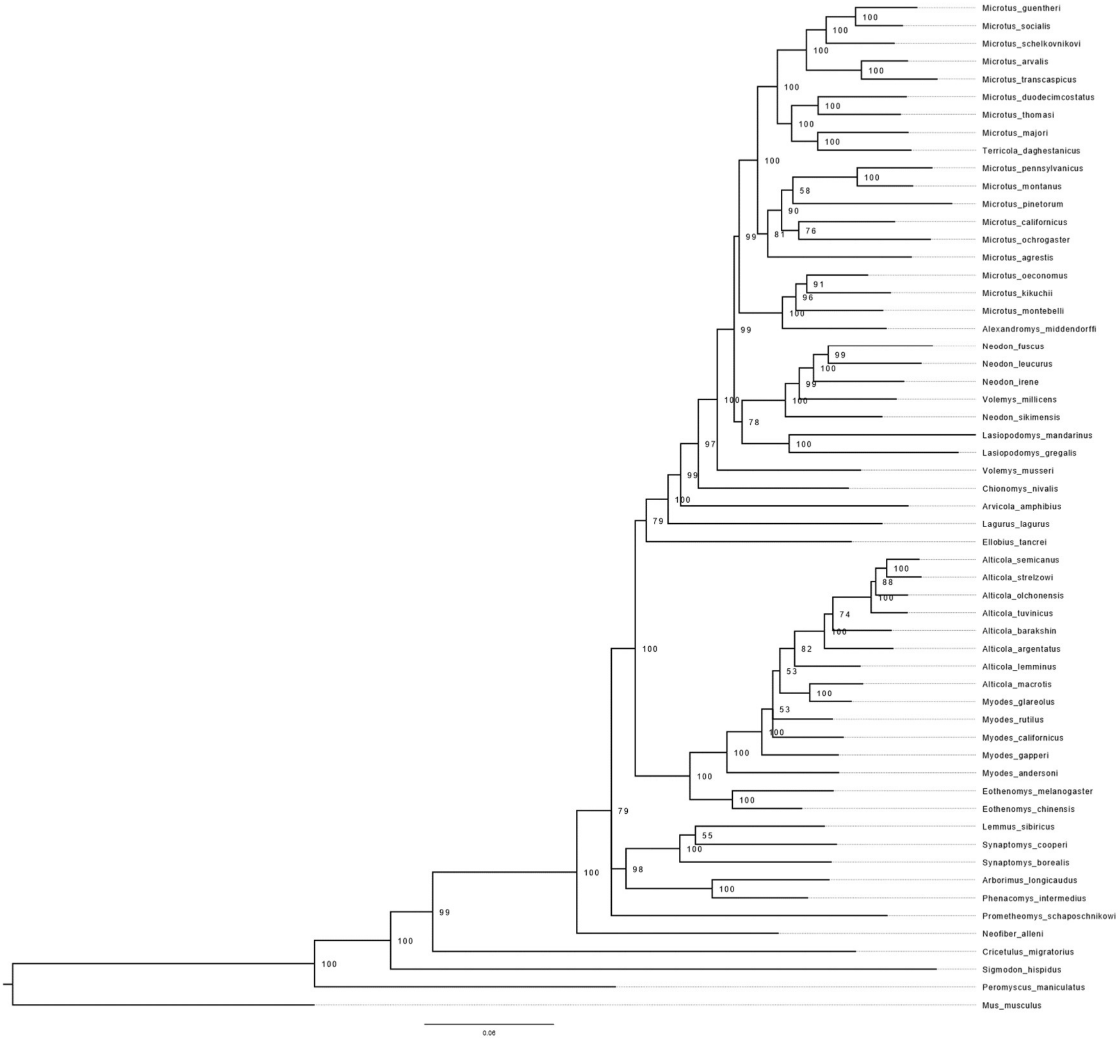


Figure A10: Bayesian inference phylogenetic tree of the concatenate for Cytb, GHR, and IRBP genes. Posterior probabilities shown as percentages.