

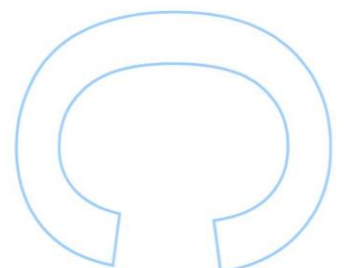
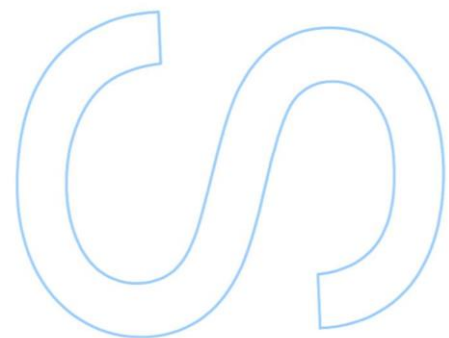
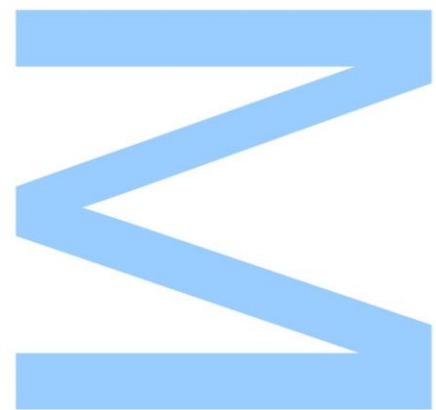


# Biogeography and phylogenetic position of a Sahara-Sahel mountain endemic, *Felovia vae* (Ctenodactylidae)

Fábio Alberto Vieira Sousa  
Mestrado em Biodiversidade, Genética e Evolução  
Departamento de Biologia  
2015

**Orientador**  
Zbyszek Boratyński, Post-doc, University of Jyväskylä

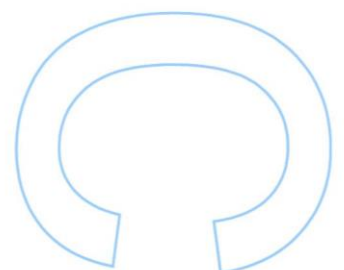
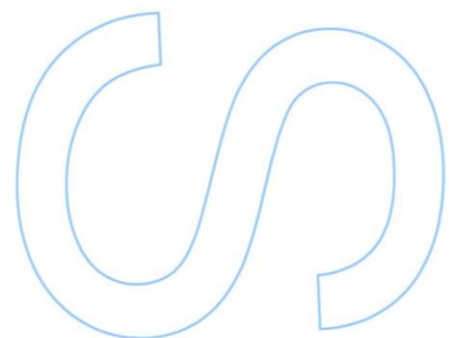
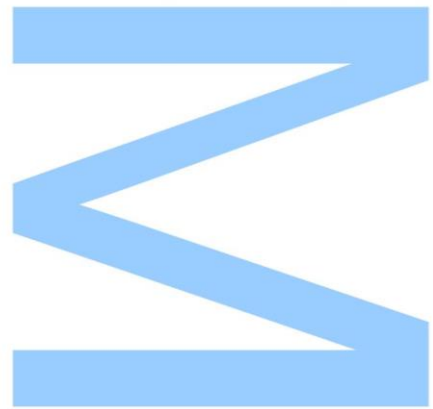
**Coorientador**  
José Carlos Brito, Senior Scientist, Assoc. Researcher, FCUP/CIBIO





Todas as correções determinadas pelo júri, e só essas, foram efetuadas.  
O Presidente do Júri,

Porto, \_\_\_\_/\_\_\_\_/\_\_\_\_



## Agradecimentos

Em primeiro lugar gostaria de agradecer aos meus orientadores, Zbyszek Boratyński e José Carlos Brito, por todo o apoio prestado durante este trabalho.

Um especial obrigado á Teresa Silva pela sua enorme paciência e ajuda no laboratório.

Um obrigado ao Duarte Gonçalves e Paulo Pereira pela disponibilidade e prontidão em ajudar.

Um obrigado a todos os membros dos Bideserts, pelo ambiente descontraido do grupo, que ajuda bastante a troca de conhecimento e que muito contribuiu para a minha aprendizagem.

Mais uma de muitas metas alcançadas ao lado das minha grandes amigas Filipa e Sara. Venham as próximas!

Não poderia faltar o agradecimento á minha familia por todo o seu apoio incondicional, bem como á Carolina pelos seus preciosos conselhos nos bons e maus momentos deste percurso.

## Resumo

A região do Saara-Sael é composta de um mosaico de habitats, montanhas, rios sazonais, áreas rochosas e extensas dunas de areia, que em conjunto definem a distribuição das espécies e a sua estrutura genética. Estas variáveis foram intensamente moldadas pelas flutuações climáticas durante o Pleistoceno, sendo normalmente possível recuperar tal efeito nos padrões genéticos das espécies. Roedores da família Ctenodactylidae são típicos habitantes das montanhas e afloramentos rochosos presentes na África do Norte, constituindo bons modelos para testar tais efeitos climáticos na região. Foi usada uma espécie endémica do Noroeste Africano, *Felovia vae*, em combinação com os restantes membros da família Ctenodactylidae, para proceder a um estudo detalhado da biogeografia que combina análises da estrutura genética e técnicas de modelação ecológica baseados no nicho da espécie. As sequências obtidas para o gene parcial do citocromo b (ADN mitocondrial) e do recetor do fator de crescimento da hormona (ADN nuclear) foram analisados usando métodos filogenéticos e de genética populacional. Fatores topoclimáticos relacionados com a ocorrência da *Felovia vae* foram identificados para as condições do presente e projetados para três tempos do passado (meio do Holoceno, último período glacial e último período interglacial). Os resultados demonstraram: 1) eventos de diversificação da família ocorreram durante o Mioceno, o género *Pectinator* constitui o clado basal da família, e o género *Massoutiera* como sendo mais relacionado com o género *Ctenodactylus* do que com o género *Felovia*, em oposição a estudos morfológicos anteriores; 2) a *Felovia vae* encontra-se subdividida em cinco clados, que correspondem maioritariamente a diferentes montanhas, sugerindo um papel importante de barreiras geográficas, associadas a oscilações climáticas do Pleistoceno, na explicação do padrão filogeográfico observado; 3) Vales arenosos, secos e extensos, assim como rios permanentes constituem importantes barreiras, durante períodos áridos e húmidos respetivamente, contudo habitats tipo savana poderão ter dissolvido estas barreiras durante certos períodos climáticos; 4) Expansões demográficas parecem estar relacionadas com períodos mais húmidos (último interglacial), enquanto períodos mais áridos (presente, meio do Holoceno, último glacial) restringiram a distribuição da espécie às montanhas; 5) o padrão observado para a *Felovia vae* é congruente com o conhecimento

biogeográfico limitado de outros vertebrados da região. Este estudo realça sinais de conectividade no passado entre populações possivelmente isoladas no presente, sugerindo uma história dinâmica do biota regional moldada por oscilações climáticas que ocorreram, e continuam a ocorrer, no Saara-Sael. Informação sobre a biologia da espécie continua a ser muito limitada (estatuto DD na IUCN), tornando-a potencialmente vulnerável a alterações de habitat, dado o seu habitat especializado, com requerimentos de afloramentos rochosos com elevada inclinação e da presença de *gueltas*.

**Palavras-chave:** Ctenodactylidae, *Felovia vae*, biogeografia, Noroeste Africano, Saara, Sael, montanhas, Mauritânia, Mali, filogeografia, modelação ecológica baseada no nicho.

## Abstract

The Sahara-Sahel is composed of a mosaic of habitats, mountains, seasonal rivers, sandy and rocky areas, that can define both species distribution and within species genetic structure. These variables were strongly shaped by the Pleistocene climatic oscillations, and such effects can be often assessed through species genetic variation patterns. Rodents of the Ctenodactylidae family are typical inhabitants of the mountains ranges and rocky outcrops across North Africa, and therefore a suitable model for testing these climatic effects. We use an endemic species from North West Africa, *Felovia vae*, in combination with all four extant Ctenodactylidae family members, to conduct a detailed biogeographical study by combining genetic structure and ecological niche-based modeling. Sequences obtained for the partial cytochrome b (mtDNA) and growth hormone factor receptor genes (nDNA) were analyzed using phylogenetic and population genetic methods. Topoclimatic factors related to the occurrence of *Felovia vae* were identified for the present conditions and projected into the past (Mid-Holocene, Last Glacial Maximum and Last Interglacial). Results showed: 1) diversification events of the family occurred during Miocene, with *Pectinator* genus being a basal clade for the family, while *Massoutiera* being more closely related with *Ctenodactylus* than with *Felovia*, opposite to what have previous morphological studies suggested; 2) five *Felovia vae* genetic clades were identified, corresponding mostly to distinct mountain massifs, suggesting an important role of external geographical barriers, associated with Pleistocene climatic fluctuations, in explaining the phylogeographical patterns observed; 3) Dry and wide sandy valleys and permanent rivers are important barriers, during dry and humid periods respectively, but savannah like habitats might have dissolved these barriers during particular climatic periods; 4) species demographic expansions seems to be related with more humid periods (Last Interglacial), while arid phases (Present, Mid-Holocene and Last Glacial Maximum) restricted the species distribution to mountains; 5) the observed pattern for *Felovia vae* is congruent with limited biogeographical knowledge of other vertebrates from the region. This study highlight past connectivity between putatively isolated contemporary populations, suggesting dynamic history of the regional biota shaped by Sahara-Sahel climatic fluctuations. Moreover, information about this species biology is still very limited (IUCN DD status) making it potentially vulnerable to habitat alteration, given its

strict habitat specialization to rocky patches with higher inclinations and the presence of gueltas.

**Keywords:** Ctenodactylidae, *Felovia vae*, biogeography, North West Africa, Sahara, Sahel, mountains, Mauritania, Mali, phylogeography, ecological niche-based modelling

# Index

<b>Agradecimentos</b> .....	<b>i</b>
<b>Resumo</b> .....	<b>ii</b>
<b>Abstract</b> .....	<b>iv</b>
<b>Index</b> .....	<b>vi</b>
<b>Figure Index</b> .....	<b>viii</b>
<b>Table Index</b> .....	<b>xii</b>
<b>Abbreviations</b> .....	<b>xiv</b>
<b>1. Introduction</b> .....	<b>1</b>
1.1. Deserts and arid regions of North Africa - the Sahara-Sahel.....	1
1.2. Historical conditions and evolution of the Sahara-Sahel .....	2
1.3. Role of Mountains on biodiversity patterns in the Sahara-Sahel .....	5
1.4. Ctenodactylidae - general characteristics .....	8
1.4.1. Ctenodactylidae - origin and evolution .....	10
1.4.2. <i>Felovia vae</i> , a rock-dwelling rodent from North West Africa.....	11
1.5. Assessing biodiversity through an integrative approach combining phylogeography and ecological niche based modeling.....	14
1.5.1. Molecular markers .....	17
1.6. Non-invasive sampling .....	18
1.7. Objectives .....	20
<b>2. Material and Methods</b> .....	<b>21</b>
2.1. Molecular approach.....	21
2.1.1. Study area and Sampling.....	21
2.1.2. DNA extraction, amplification and sequencing .....	23
2.1.3. Phylogenetic analysis .....	24
2.1.4. Time of divergence.....	25
2.1.5. Population genetic analysis.....	27
2.1.6. Historical demography .....	27
2.1.7. Spatial analyses .....	29
2.2. Ecological niche-based modeling approach.....	30
2.2.1. Species occurrences.....	30
2.2.2. Ecogeographical variables .....	30
2.2.3. Ecological niche-based models.....	31
2.2.4. Ecogeographical variables (EGVs) importance .....	32
2.2.5. Present prediction of species occurrence.....	32

2.2.6. Past prediction of species occurrence .....	33
2.2.7. Refugia .....	33
<b>3. Results.....</b>	<b>34</b>
3.1. Molecular approach.....	34
3.1.1. Laboratory overview.....	34
3.1.2. Phylogenetic analysis .....	34
3.1.2.1 “ <i>Felovia vae</i> tree”- Cytb.....	34
3.1.2.2. “Ctenodactylidae tree” - Cytb .....	39
3.1.2.3. “ <i>Felovia vae</i> tree” - GHR .....	41
3.1.3. Time of divergence.....	43
3.1.4. Population genetic analysis.....	45
3.1.5. Historical demography .....	49
3.1.6. Spatial analyses.....	52
3.2. Ecological niche-based modeling approach.....	53
3.2.1. Models performance .....	53
3.2.2. Environmental factors related to species occurrence .....	54
3.2.3. Present conditions .....	57
3.2.4. Past conditions .....	59
3.2.5. Refugia identification.....	59
<b>4. Discussion.....</b>	<b>61</b>
4.1. Ctenodactylidae phylogeny.....	61
4.2. Biogeography of <i>Felovia vae</i> .....	63
4.3. Effects of the past climate change on the demographic pattern of <i>Felovia vae</i> ..	66
4.4. Future research .....	69
4.4.1. Molecular markers .....	69
4.4.2. Sample size .....	70
<b>5. Conclusion .....</b>	<b>72</b>
<b>6. References .....</b>	<b>74</b>
<b>7. Supplementary material .....</b>	<b>91</b>

## Figure Index

- Fig. 1** – Africa climate fluctuations and its influence on vegetation cover, since the Last Glacial Maximum (LGM) until current conditions (adapted from: Adams & Faure, 2004).  
..... 4
- Fig. 2** – Guelta El Barda in Guidimaka, Mauritania (by: Zbyszek Boratyński)..... 6
- Fig. 3** – Known (putative) distributions in North Africa of the members of Ctenodactylidae family (Aulagnier, 2008; Coetzee & Grubb, 2008; both from [www.iucnredlist.org](http://www.iucnredlist.org))..... 9
- Fig. 4** – Principal mountain massifs (light grey background under black text), two relevant escarpments (light grey background with underlined black text) and river valleys (dark grey background under white text) in Mauritania. .... 13
- Fig. 5** – Study area: A) African geographic distribution of *Felovia vae* samples used in phylogenetic analyses for Cytb (n=94; white circles) and GHR genes (n=26; black triangles). In the background is the known distribution of *F. vae* (Aulagnier, 2008: [www.iucnredlist.org](http://www.iucnredlist.org)) B) Small inset map representing the study area in Mauritania and Mali. More than one sample can be found in one dot. Samples from the other Ctenodactylidae member, both amplified in this study and from GenBank are not represented (unidentified country and location for most samples)..... 22
- Fig. 6** – Phylogenetic tree for *Felovia vae* based on mtDNA (Cytb) data (700 bp length; 94 sequences). The tree is rooted with outgroups from the Ctenodactylidae family. Values of Bayesian posterior probabilities and bootstrap values of ML are above or near the branches. Colors represent the main groups obtained and are disposed on map in Fig. 7..... 37
- Fig. 7** – Geographic distribution of *Felovia vae* samples belonging to each main clade (colors) recovered by the phylogenetic tree based on partial Cytb gene (700 bp length; 94 sequences). Numbers represent hypothetical populations delineated using a 30 km distance around each sample (see Tab. S1 for more information)..... 38
- Fig. 8** – Phylogenetic tree for the Ctenodactylidae family based on mtDNA (Cytb) data (813 bp length; 10 sequences). The tree is rooted with outgroups from the sister taxa

Diatomydae and two Hystricognathi. Values of Bayesian posterior probabilities and bootstrap values of ML are above or near the branches. Colors represent the main *Felovia vae* clades obtained in Fig. 6. .... 40

**Fig. 9** – Phylogenetic tree for *Felovia vae* based on non-recombinant nDNA (GHR) data (362 bp length; 24 sequences). The tree is rooted with outgroups from the Ctenodactylidae family. Values of Bayesian posterior probabilities and bootstrap values of ML are above or near the branches .....42

**Fig. 10** – Trees showing mean divergence dates in million years (above the branches/before the branches) and 95 % credible intervals (blue bars) for relevant supported nodes (given by the black dot). Left: “Ctenodactylidae tree” (Cytb) following the approach B (include two calibration constrains: fossil and mutation rate) (see Tab. S2 for detailed information). Right: “*Felovia vae* tree” (Cytb) (see Tab. S3 for detailed information) using the estimated substitution rate by the previous approach B. Colors represent the main *Felovia vae* clades obtained in Fig. 6..... 44

**Fig. 11** – Median-joining network inferred by Network based on *Felovia vae* partial Cytb gene (700 bp length; 94 sequences). Colors represent each main clade, while circles the haplotypes scaled proportionally to their frequency in the sample. Numbers inside the circles represent haplotypes (see Tab. S1 for more information). Mutational steps among haplotypes are indicated by black dots..... 46

**Fig. 12** – Neighbor-net network inferred by Splitstree based on *Felovia vae* partial Cytb gene (700 bp length; 94 sequences). Scale bar represents 1 % sequence divergence and numbers correspond to bootstrap values. .... 47

**Fig. 13** – Non-recombinant median-joining network of *Felovia vae* partial nDNA (GHR) (362 bp length; 24 sequences). Circles represent haplotypes scaled proportionally to their frequency in the sample. Numbers represent haplotype names. Mutational steps among haplotypes are indicated by black dots. Colors represent the main *F. vae* clades identified for mtDNA (Cytb). .... 47

**Fig. 14** – Pairwise mismatch distribution of nucleotide differences of Cytb in each clade of *Felovia vae* (Left). Dashed lines indicate the expected distribution under a population expansion model. Solid lines represent the observed distribution. Bayesian skyline plots of Cytb *F. vae* clades (Right). The median estimate (black line) and 95% highest

posterior density limits (blue background) are indicated. The y-axis represents the scaled population size and the x-axis indicates measures time in substitutions per site per million years. A, B, C and E correspond to the clades in Fig. 6 and 7, while the combined clades are represented by “ALL”..... 51

**Fig. 15** – Geographic genetic variation in *Felovia vae*. Surface of interpolated genetic diversity based on nucleotide diversity, resulting raster file was reclassified into 9 different classes using natural breaks. The circles correspond to samples used for the interpolation, the circle colors to the Cytb clades and the circles with no color were excluded because they had no “neighbor” to merge in a population. Darker green colors in background correspond to areas with the lowest nucleotide diversity while darker red areas are assigned to areas with high nucleotide diversity. .... 52

**Fig. 16** – Geographic genetic differentiation in *Felovia vae*. First axis of the spatial principal component analysis applied to the interpolations of the uncorrected p-distances matrix. The circles correspond to the samples used for the interpolation and the point colors to the Cytb clades. .... 53

**Fig. 17** – Response curves for the three most related environmental factors to the distribution of *Felovia vae*. Curves traduce the average probability of occurrence from 20 model replicates along the environmental gradients..... 55

**Fig. 18** – Plot of the two environmental variables that most described the differences between genetic groups. For each observation was attributed a genetic group, supported with genetic information. Each color corresponded to a Cytb clade (Fig. 6). .... 56

**Fig. 19** – Predicted suitability area of *Felovia vae* in the consensus model for current conditions (different scale) and projections to the Mid-Holocene, Last Glacial Maximum (LGM) and Last Interglacial (LIG). The country borders were left for past projections with an orientation propose and the rivers were removed, since nothing was known about their dynamics in the past..... 58

**Fig. 20** – Potential refugia given by stable .suitable conditions across all periods (present, Mid-Holocene, Last Glacilar Maximum, Last Interglacial). .... 60

**Fig. S1** – Depiction of the three principal components of the PCA, with decreasing relevance from left to right for explaining diversity patterns. On the left the first component clearly separates the Adrar Atar-Tagant from the Assaba Afollé mountains, while in the second image we can see a separation of the Assaba from the other mountains, and on the third component we can observe the isolation of the Felou populations against the other systems. .... 98

**Fig. S2** – Average probability of *Felovia vae* occurrence, standard deviation (small insets) (up) and predicted suitable areas (down) for the current conditions and for projections to the Mid-Holocene, Last Glacial Maximum (LGM) and Last Interglacial (LIG). Different global circulation models were used CCSM4, MIROC-ESM and MPI-ESM for Mid-Holocene and LGM and the NCAR-CCSM for LIG (Mid-Holocene – CC, ME and MR; LGM – CC, ME and MR; LIG – NCAR-CCSM). .... 99

## Table Index

<b>Tab. 1</b> - Environmental factors (coded as BIO in WorldClim website) used for modeling the distribution of <i>Felovia vae</i> in the study area. ....	30
<b>Tab. 2</b> - Summary statistics and population growth for <i>Felovia vae</i> in two genes, Cytb (mtDNA) and non-recombinant GHR (nDNA).....	36
<b>Tab. 3</b> - Genetic net average pairwise distances among members of Ctenodactylidae family and outgroups (Diatomyidae and Histricognathi) for partial Cytb gene. Means and standard deviation values appear below and above the diagonal, respectively. ...	41
<b>Tab. 4</b> - Genetic net average pairwise distances among members of Ctenodactylidae family for partial GHR gene. Means and standard deviation values appear below and above the diagonal, respectively. ....	42
<b>Tab. 5</b> - Spatial genetic structure inferred by analyses of molecular variance. Percentages of variation explained by two different grouping approaches are indicated. Statistical significance was tested using non parametric permutation approach (**P < 0.01, *P < 0.05, NS P > 0.05). See Tab. S1 for number locality codes information. ....	48
<b>Tab. 6</b> - Genetic net average pairwise distances calculated under K2P model between <i>Felovia vae</i> clades (means and standard deviation values appear below and above the diagonal, respectively) and within genetic distance (means followed by standard deviations under brackets in the diagonal) for partial Cytb gene. ....	49
<b>Tab. 7</b> - Average percent contribution (and standard deviation) of each environmental variable for the 20 replicate models. ....	54
<b>Tab. S1</b> - Summary table of the <i>Felovia vae</i> samples. Mitochondrial clades are described along with the resulted haplotype both for Cytb and GHR gene fragments. Recombining haplotypes that have been excluded from further analyses are represented by R*.....	91
<b>Tab. S2</b> - Estimated times to the most recent common ancestor (TMRCA; mean values and 95% HPD) for the three approaches used, regarding the type of calibration. The	

“Ctenodactylidae tree” uses the family members with a set of particular calibration constrains (set = approach) to estimate substitution rate values. Meaning: Eocene (Eoc.); Oligocene (Olig.); Miocene (Mio.); Pliocene (Plio.); Pleistocene (Ple.)..... 95

**Tab. S3** - Estimated times to the most recent common ancestor (TMRCA; mean values and 95% HPD) for the three approaches used, regarding the type of calibration. The “*Felovia vae* tree”, which include only *F. vae* haplotypes, uses the substitution rates estimated in Tab. S2. Meaning: Eocene (Eoc.); Oligocene (Olig.); Miocene (Mio.); Pliocene (Plio.); Pleistocene (Ple.)..... 97

## Abbreviations

μL – Microliter

μM – Micromolar

BI – Bayesian inference

BP – Before present

BPP – Bayesian posterior probability

BSP – Bayesian skyline plot

Cytb – Cytochrome b

DNA - Deoxyribonucleic acid

dNTP – Deoxyribonucleotide triphosphate

EGVs – Ecogeographical variables

Eoc. – Eocene

GCM – General circulation models

GHR – Exon 10 of the growth hormone receptor

GIS – geographic information system

HPD – High posterior density

HPD – Highest posterior density

K2P – Kimura 2-parameter model

km – Kilometer

LGM – Last Glacial Maximum

LIG – Last Interglacial

Mid-Holocene – Middle Holocene

min – Minute

Mio. – Miocene

ML – Maximum Likelihood

mM – Milimolar

mtDNA – Mitochondrial DNA

Mya – Million years ago

n – Number of individuals/ other units

nDNA – nuclear DNA

ng – Nanogram

NI – Non-invasive

° C – Celsius degrees

Olig. – Oligocene

PCA – Principal component analysis

PCR – Polymerase chain reaction

Ple. – Pleistocene

Plio. – Pliocene

rRNA – Ribosomal RNA

s – Second

TMRCA – Time to the most recent common ancestor

tRNA – Transfer RNA

yr – Year

# 1. Introduction

## 1.1. Deserts and arid regions of North Africa - the Sahara-Sahel

The word “deserts” most often evokes a landscape of endless sand dunes, a cloudless blue sky and a blazing sun (Harris, 2003). Deserts are conventionally seen as almost bare of vegetation or wildlife. Although some parts of the Sahara fit quite well this image, it overshadows the remarkably fertile habitats provided for plants, animals, and also humans, each of which adapted to extract the best that deserts as to give (Harris, 2003). This homogeneous view does have an influence on attracting less scientific attention (Durant *et al.*, 2012). Consequently fewer resources are given for nature conservation in these regions, as funding priorities have been mostly directed to global biodiversity hotspots (Durant *et al.*, 2012). One general and important step, to reverse biodiversity loss on arid regions, is to shed light upon their species diversity and distribution and also the evolutionary processes that precedes them. However, sometimes it can be a difficult task in such often remote conditions. For instance, the Sahara-Sahel large size, remoteness and long-term political instability with regional widespread conflicts and high levels of poverty hamper scientific exploration (Brito *et al.*, 2014). The hyperarid (deserts), arid and semi-arid (semi-desert) zones can be identified through an aridity index, which is given by the ratio between mean annual precipitation and mean annual potential evapo-transpiration index (<0.05, between 0.05 and 0.20, between 0.45 and 0.65; for hyperarid, arid and semi-arid areas; UNEP, 1992). The Sahara, the largest and the warmest desert in the world (covering 8% of the earth’s land area), and the bordering semi-arid Sahel constitute two major ecoregions of the African continent. The limits between these two ecoregions mark the transition between the Palearctic and Afro-Tropical biographical realms (Olson *et al.*, 2001). As result the region comprise a great latitudinal variation in species distribution and increased local biodiversity (Dumont, 1982; Le Houérou, 1992; Brito *et al.*, 2014). Together with its particular diversity of topographic features (high mountains, seasonal rivers, extensive sand dunes, bare-rock plateaus, rocky areas and oases) and the heterogeneity of its climate (large temperature and rainfall variation in space and time) make this an area of particular interest for biodiversity and evolutionary research (Le Houérou, 1997; Harris, 2003; Anthelme *et al.*, 2008).

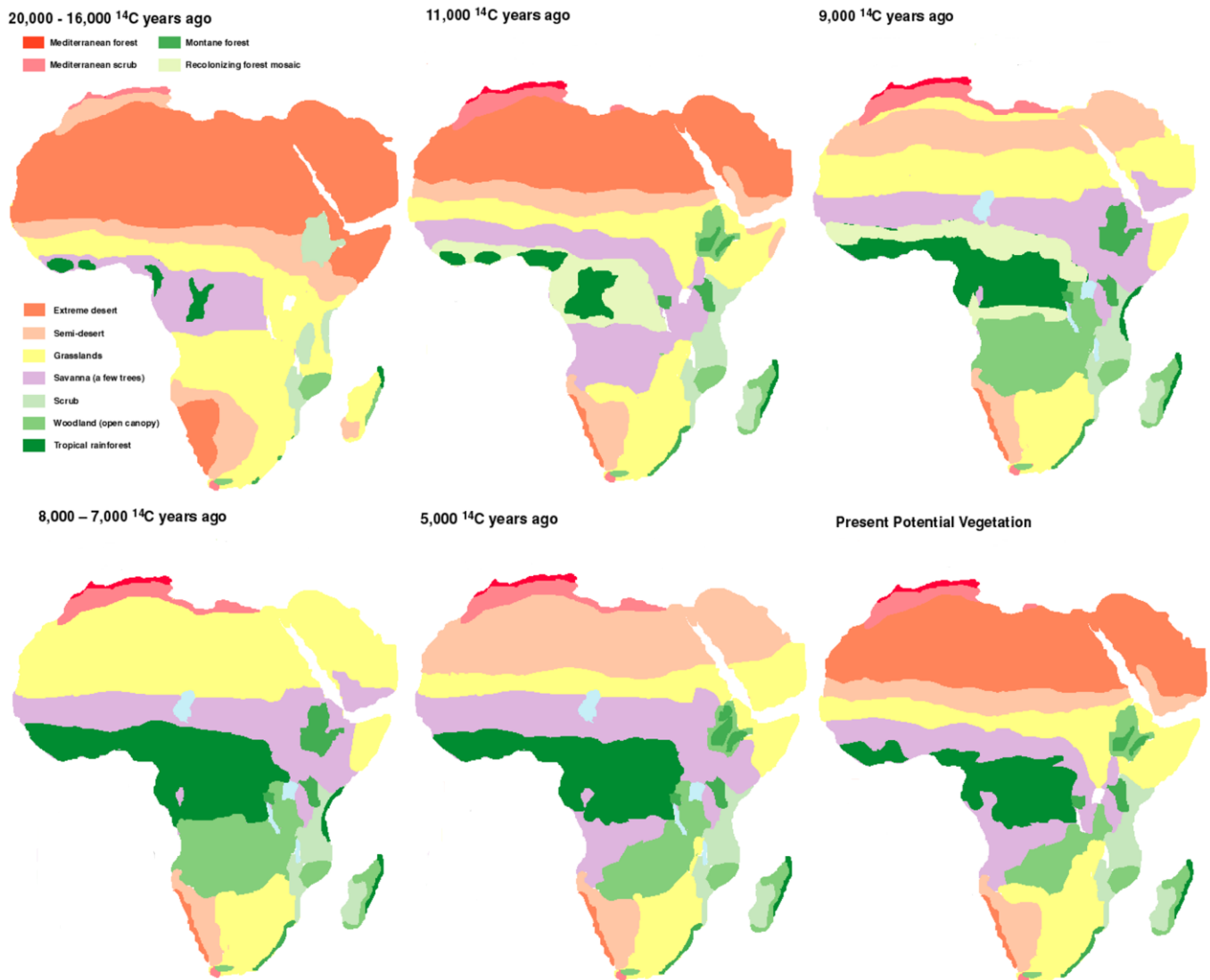
Sharp ecological gradients can be found in these extreme environments (Schulz *et al.*, 2009), with species highly adapted to such harsh conditions (Ward, 2009). These species are exposed, and developed several adaptations, to scarce and unpredictable water and food resources, as well as extreme solar radiation and temperatures (Brito *et al.*, 2014). However, the on-going climate change, which is predicted to have the fastest velocities in the Sahara-Sahel region and other deserts (Loarie *et al.*, 2009), combined with an increase in human pressure can contribute to fast environmental switches. This may cause shifts in arid and semi-arid environments and thus increase the vulnerability to extinction of its already delicate biodiversity (Brown *et al.*, 1997; Sanz, 2002; Davies *et al.*, 2012). At a species level, slight increases in aridity may have negative impacts on habitat suitability. For instance, the loss of the scarce water sources as a consequence of prolonged droughts may induce habitat fragmentation and limit dispersal, consequently increasing chances for local extinction (Ward, 2009).

## 1.2. Historical conditions and evolution of the Sahara-Sahel

Since the onset of hyper-arid conditions in the Sahara, approximately 7 to 2,5 million years ago (Mya) (Swezey, 2009), the Sahara has experienced drastic climatic changes between dry and wet periods (Le Houérou, 1997), leading to highly variable boundaries and a strong influence on the Sahel ecoregion (Brito *et al.*, 2014). Some 3 Mya, the African climate started to get generally cooler and drier, along with successive spread and retraction of arid and semi-arid conditions cycles, more or less following the glacial-interglacial ones at higher altitudes (deMenocal, 2004). The frequency of such cycles occurred approximately 100 000 – 20 000 years (Le Houérou, 1997; deMenocal, 2004). It is generally admitted the relation of this phenomenon with the orbital (Milankovitch) cycles and West African monsoon circulation variability and vegetation feedbacks in regional climate (Rognon, 1989; Maley, 2010; Armitage *et al.*, 2015).

On the one hand, in the periods of expansion the area of Sahara was much larger, with the climate much colder and arid than present conditions. The current Sahelian region was a wide desert covered by sand dunes. Aridity is, therefore, the major feature of North Africa during such periods, as for instance in the Last Glacial Maximum (LGM; 24 000 – 18 000 years ago) (Le Houérou, 1997) (Fig. 1). During

humid periods (warmer and moister), many regions that are now dry were more humid, harboring steppe, savannah, even forest conditions and have been covered by dense river networks, lakes and wetlands (Kröpelin *et al.*, 2008; Drake *et al.*, 2011). These periods of greater humidity, were suggested to occur e.g. during Marine Isotope Stage 5 (~130 000 – 75 000 yr) (Drake *et al.*, 2011) and at the Early Holocene (~11 000 – 5 000 yr) (Gasse, 2000; Kröpelin *et al.*, 2008; deMenocal & Tierney, 2012) (Fig. 1). After the last African humid period, between around 6 000 and 5 000 yr onwards, the increase of aridity led to an abrupt end of the so called “Green Sahara” and its mesic vegetation communities and hydrological networks (Kröpelin *et al.*, 2008; Armitage *et al.*, 2015) (Fig. 1). The conditions became more arid and similar to the present. Also recently, severe droughts occurred in the semi-arid Sahel during the 1970s and 1980s (Ahmed *et al.*, 2008; Mahé & Paturel, 2009) and the causes of the persistence of such conditions are the center of a long-standing debate (see Giannini, 2015 for a brief review). However, since mid- 1980s, an increase of the seasonal total amount of precipitation through intense, but less frequent and so predictable, rains are being suggested (Giannini, 2015). Dong and Sutton (2015) attribute this increase to a direct effect of Green House Gases.



**Fig. 1** – Africa climate fluctuations and its influence on vegetation cover, since the Last Glacial Maximum (LGM) until current conditions (adapted from: Adams & Faure, 2004).

The paleoclimate fluctuations may have had a major influence on rates of species diversification, geographic distribution and evolution of animal species in the region (Ward, 2009), leaving traces in the genetic structure of populations (Brito *et al.*, 2014). However, phylogeographical studies concerning the Sahara-Sahel region are scarce (Brito *et al.*, 2014), when compared for instance with Europe and North America where the knowledge about the role of Quaternary climatic oscillations in the history of species is well documented (many examples are given by Hewitt, 2000, 2004). In one way, this could be explained by the weaker impact of such oscillations in deserts when

compared to the ice retraction and expansions in other parts of the Globe (that periodically driven local extinction of entire communities due to physical act of ice sheets), that are described to be related to events of vicariance, expansions, colonization and recolonization of species (Hewitt, 2004).

Nevertheless, available phylogeographic research in North Africa has also revealed significant taxa diversification influenced by these Miocene-Pleistocene oscillations. For instance, studies have suggested that Sahara acted as a barrier preventing individuals exchange between north and central Africa during more arid phases, promoting vicariant speciation (e.g. Carranza *et al.*, 2008; Ben Faleh *et al.*, 2012). During humid phases the dispersal throughout the Sahara appears to be allowed (Douady *et al.*, 2003; Gaubert *et al.*, 2012). Importantly, Sahel-Saharan mountains, that form sky-island-like systems throughout North Africa, were affected in a different way by the climatic fluctuations than surrounding lowland deserts and savannas, affecting mostly their altitudinal distribution and so, connectivity between populations (Messerli & Winiger, 1992; Gonçalves *et al.*, 2012).

### 1.3. Role of Mountains on biodiversity patterns in the Sahara-Sahel

The Sahara-Sahel is marked by a variable orography being occupied by large and smaller mountain ranges scattered along its landscape. The most important in extension and altitude are the Ahaggar (or Hoogar) mountains in Algeria, the Tassili n'Ajjer in Algeria and Acacus Mountains in south-western Libya, the Air Mountains in northern Niger, the Adrar des Ifoghas in Mali and the Adrar Plateau (or Adrar Atar) and Tagant Plateau in Mauritania. These areas have played a key role in the biological diversification by providing dispersal corridors across arid areas and acting as refugia (Brito *et al.*, 2014). They also harbor several isolated populations of endemic species of mammals (e.g. Brito *et al.*, 2010), fishes (e.g. Trape, 2009), amphibians, reptiles (e.g. Geniez *et al.*, 2004; Brito *et al.*, 2011) and birds (e.g. Tellería, 2009), highlighting their role in the diversification of the Saharan vertebrate fauna. In total mountains comprise 41 Sahara-Sahel vertebrate (51% of all Sahara-Sahel endemics) and other 88 vertebrates without a Saharan origin (45 % of all non-Saharan) (Brito *et al.*, 2014). Some of these Saharan mountains offer sufficiently high levels of humidity and

relatively cool temperatures, allowing the survival of Sahelian and Mediterranean species not adapted to the drier and hotter conditions in the surrounding extremely arid lowlands (e.g. Migliore *et al.*, 2013).

Furthermore, recent studies are suggesting that distribution of biodiversity micro-hotspots (high species diversity in a narrow range) throughout the Sahara-Sahel coincides with these mountain regions (Wagner *et al.*, 2011), and around water bodies (Vale *et al.*, 2015). In the particular case of Mauritanian mountains, Vale and colleagues (2015) gave special relevance to the conservation importance of mountain lagoons (locally known as *gueltas*), which holds endemic fauna and range-margin populations (Fig. 2). These permanent water pools are sparsely distributed and surrounded by dry areas, and allow the maintenance of rich communities with several endemic species, acting as refugia to relict populations and potential speciation drivers (Anthelme *et al.*, 2008; Brito *et al.*, 2014).



**Fig. 2** – Guelta El Barda in Guidimaka, Mauritania (by: Zbyszek Boratyński).

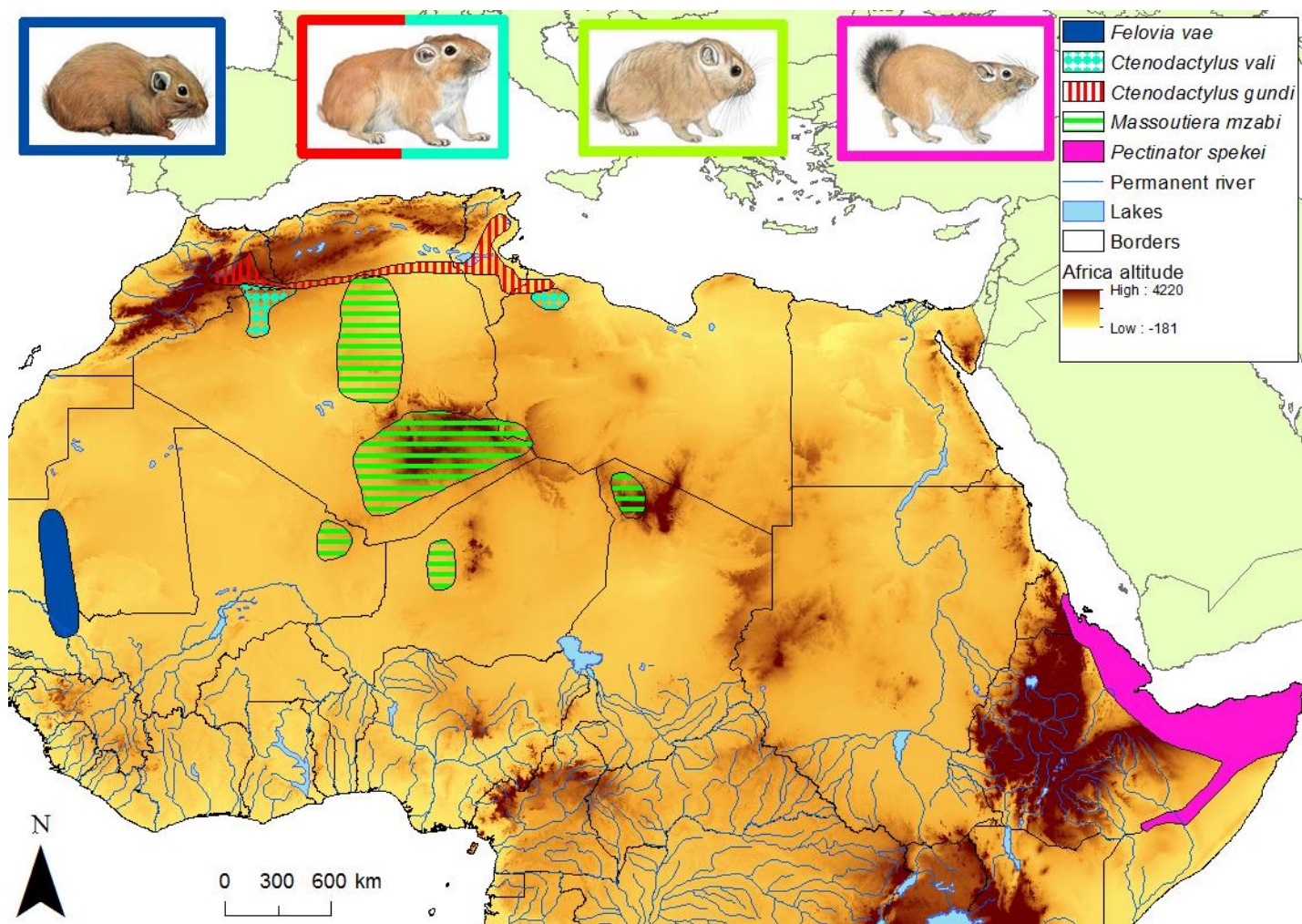
The relatively high number of endemic species and mountain isolated populations found in the Sahara-Sahel may be due to the effect of such Plio-Pleistocene climate fluctuations on connectivity (Le Hou  rou, 1997; Drake, 2011; Brito *et al.*, 2014). During past dry climatic periods mesic (require moisture, although being adapted to arid conditions) and particularly water dependent species suffered severe population contractions with range fragmentation to rocky massifs and mountains that retained humidity (Brito *et al.*, 2014; Velo-Ant  n *et al.*, 2014). Thus, allopatric distribution and long-term interruption of gene flow can initiate evolution of independent lineages, which can under particular circumstances cause speciation. Eventually, population expansion and gene flow between mountainous and isolated populations can be renewed during more favorable and humid periods, where Savannah-like habitats form corridors for species dispersal (Kr  pelin *et al.*, 2008). The reverse scenario, isolation during humid stages of Sahara, also, is expected for desert specialists [e.g. *Jaculus* spp. (Boraty  nski *et al.*, 2012)]. Nicolas and colleagues (2009) in a phylogeographic study of the spiny mice (genus *Acomys*) from the south-Western Sahara, suggested population contractions with population fragmented across rocky areas surrounded by large sandy areas during arid or semi-arid paleoclimate periods, followed by expansion during phases of favorable habitat expansion. Also, Gon  alves and colleagues (2012) gave insights about the colonization process and diversification of North African lizards from the genus *Agama*. In this study the mountain ranges in Sahara-Sahel explained the splitting, genetic diversity and distribution of allopatric lineages, suggesting a colonization of northern Africa during wet periods and population contraction to Mediterranean climates, Sahel and Saharan mountains during dry/unfavorable periods. In the same line but for a narrow range, Velo-Ant  n and co-workers (2014) identified a recent meta-population system of *Crocodylus suchus* in Mauritanian mountains, formed when these mountains were colonized from southern ranges during the last humid phase of the Sahara. The beginning of the posterior arid phase in the Mid-Holocene contributed for a geographic isolation and genetic structuring between the mountain relict populations. However, new barriers to non-aquatic species dispersal, like rivers, can emerge during more humid periods (e.g. Dobigny *et al.*, 2005, 2013; Brouat *et al.*, 2009), where climatic fluctuations may have had an influence on hydrological networks and thus on species distribution (Le Hou  rou, 1997; Drake *et al.*, 2011).

#### 1.4. Ctenodactylidae - general characteristics

The North African endemic rodent family, Ctenodactylidae, encompasses members inhabiting exclusively rocky ecosystems (mountains, cliffs) restricted to Sahara-Sahel. This strong association with particular ecological niches, along with other characteristics like weak dispersal abilities, small size, short generation times, and rapid mtDNA substitution rates are relatively common in rodents (Dobigny *et al.*, 2013). The combination of these characteristics facilitates the detection of phylogeographic patterns, by increasing species sensibility to geographic barriers such as mountain uplifts, valleys or rivers and thus promoting independent evolution of allopatric lineages. Several studies, showing strong phylogeographical signals are, indeed, available for sub-Saharan rodents with very different ecological requirements [arid zones: *Jaculus jaculus* (Boratyński *et al.*, 2012); steppes, shrub and tree savannahs: *Arvicanthis niloticus* (Dobigny *et al.*, 2013), *Lemniscomys striatus* (Nicolas *et al.*, 2008), *Praomys daltoni* (Bryja *et al.*, 2010), *Mastomys erythroleucus* (Brouat *et al.*, 2009), *Mastomys natalensis* (Colangelo *et al.*, 2013); and humid areas: *Mastomys huberti* (Mouline *et al.*, 2008)]. Therefore, rodents are promising biogeographic models to assess the influence of the climatic oscillations and the role of mountains in shaping biodiversity patterns in the region.

The gundis family (Rodentia, Ctenodactylidae) comprises four genera that include five species with mainly non-overlapping (with an exception to two *Ctenodactylus* species) distributions (Aulagnier, 2008; Coetzee & Grubb, 2008; both from [www.iucnredlist.org](http://www.iucnredlist.org)) (Fig. 3). *Ctenodactylus vali* is known from northeastern Morocco and adjacent parts of northwestern Algeria, while the other area of its distribution is in Libya. *Ctenodactylus gundi* ranges from eastern Morocco, through Algeria and Tunisia, to western Libya. In the west of the Saharan Atlas, precipitation levels seem to influence the limit between the two species distributions (Aulagnier, 2008). *Massoutiera mzabi* occurs in the central Sahara regions of Algeria, northeastern Mali, northern Niger and northwestern Chad. *Felovia vae* distribution and ecology is poorly known with records coming from Mauritania, western Mali and in the southeastern Senegal. And finally, *Pectinator spekei* is patchily distributed species with range including Ethiopia, Djibouti and Eritrea, and with an unclear distribution in Somalia. The genera and species of the family may be distinguished by tail length, and

visibility, size of auditory bullae, size of fecal pellets and geographic distribution (Kingdon *et al.*, 2013).



**Fig. 3** – Known (putative) distributions in North Africa of the members of Ctenodactylidae family (Aulagnier, 2008; Coetzee & Grubb, 2008; both from [www.iucnredlist.org](http://www.iucnredlist.org)).

Gundis are considered as locally common species throughout their range and with no major threats associated, except for *Felovia vae* and *Ctenodactylus vali* (Aulagnier, 2008; Coetzee & Grubb, 2008; both from [www.iucnredlist.org](http://www.iucnredlist.org)). The lack of knowledge on these two species, concerning population size, range dimensions, ecology and population trends resulted on its Data Deficient status in IUCN category (Aulagnier, 2008). A proposal for modification of this status for the category to Least Concerned was recently presented for *Felovia vae* (Vale *et al.*, 2012a). However, there are still several points of knowledge that need to be assessed in order to support this alteration. For example, studies with molecular markers are essential to determine if genetic sub-structuring occurs among the distinct subpopulations occupying mountain regions in Mauritania-Mali of this mammal.

#### 1.4.1. Ctenodactylidae - origin and evolution

The systematic position of the African gundis, Ctenodactylidae, has been the center of debate for decades. This taxon includes some of the oldest rodent fossils, therefore it was proposed as the first offshoot in rodent clade (Hartenberger, 1985). It is now widely accepted that Ctenodactylidae and *Hystricognathi* robustly cluster together, forming the Ctenohystrica clade (Huchon *et al.*, 2000).

The discovery of a South-East Asia potential survival member of the extinct family Diatomyidae (*Laonates aenigmamus*) shed new light on the origins of the modern African endemic families. It served as a boost for molecular studies that described it as a sister-group of Ctenodactylidae, and now the Ctenohystrica is divided into two well-supported clades: Ctenodactylidae (gundi) + Diatomyidae (*Laonastes*) and Hystricognathi (Huchon *et al.*, 2007). Huchon and co-workers (2007), through molecular analyses, confirm that the dichotomy *Ctenodactylus-Laonastes* dates around 44.7 Mya ( $\pm 3.5$  Mya) and indicates that these families diverged by vicariance.

Ctenodactylidae fossils have been discovered in early Miocene in North Africa (Monadjem *et al.*, 2015). In this epoch occurred their greatest diversification with the widest distribution, known from West Africa, north western India and Central China (López-Antoñanzas & Knoll, 2011). Currently this family is restricted to dry and arid environments, with tropical forest acting as a putative barrier preventing their dispersal

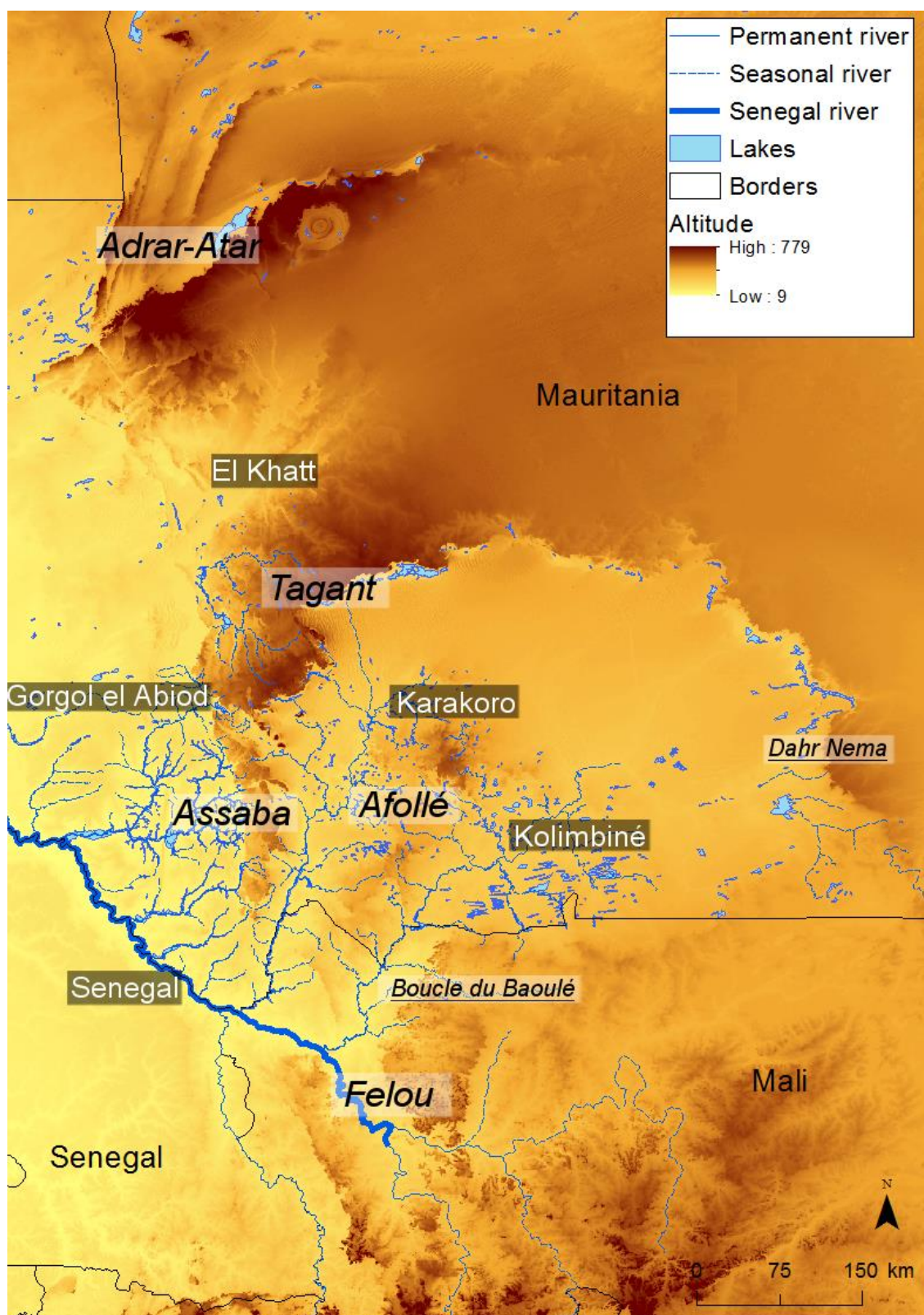
into other parts of Africa (Monadjem *et al.*, 2015). The evolution of this group is quite exciting, being related with paleoecological perturbations such as the collision of Afro-Arabia with Eurasia in the Early Miocene and the birth of the Sahara desert in Late Miocene, as explained by López-Antoñanzas & Knoll (2011). Based on morphological and dental data, the genus *Pectinator* has been placed basal in the family Ctenodactylidae, while the closest related clades seems to be *Felovia* and *Massoutiera* genus (George, 1985; López-Antoñanzas & Knoll, 2011; Gomes *et al.*, 2012). However, a molecular phylogeny for the family is needed to shed light upon the relationships between Ctenodactylidae members.

#### 1.4.2. *Felovia vae*, a rock-dwelling rodent from North West Africa

The Felou gundi (*Felovia vae* Lataste 1886) is a predominantly Sahelian endemic species with a very limited distribution, restricted to north-western Mali and south-central Mauritania. It occurs in the Felou hills along the upper Senegal River in Mali, and in 4 mountain massifs in Mauritania (Adrar Atar, Tagant, Assaba and Afollé) (Brito *et al.*, 2010; Vale *et al.*, 2012a). It is closely associated with the presence of *gueltas* and rocky slopes given by canyon walls, cliffs and large boulders (Padiál & Tellería, 2009; Vale *et al.*, 2012a). *Gueltas* are rocky holes (usually small: 0.01 to 5 ha) with permanent water, fed by seasonal rivers. The effect of water on the occurrence of *Felovia* is believed to be indirect, as gundi adapted to obtain moisture entirely from herbivorous food (grasses, dry leaves, flowers, seeds; Kingdon, 2013). Despite the importance of permanent water presence for food resources, higher primary productivity and thus species richness (Vale *et al.*, 2015), *F. vae* was also observed drinking from *gueltas* (Vale *et al.*, 2012a). Large distances between patches of habitat were suggested to induce population fragmentation (Brito *et al.*, 2010), and ecological models predicted barriers to current gene flow between mountains (Vale *et al.*, 2012a). Considering the species relatively low body weight (177-193 g; Kingdon, 2013), the large unsuitable habitat patches may hamper gene flow.

The principal ecological discontinuities found within Mauritanian mountains can be given by one permanently dry river valley and two seasonal hydrographic sub-basins (Campos *et al.*, 2012) (Fig. 4). The El Khatt (20 km wide), permanently dry valley, between the Adrar Atar and Tagant mountains is covered by dunes and displays lack of significant rocky outcrops. The Gorgol el Abiod (2 km wide) is covered by

patches of rocks that separate the southern Tagant from the northern Assaba mountain, and results from both mountains drained water. The Karakoro river valley (25km wide) between Tagant-Assaba and Afollé lacks rock outcrops and it is relatively populated (Vale *et al.*, 2012a), although being characterized by low water availability (Velo-Antón *et al.*, 2014). The southern Mauritanian mountains and Mali are separated by the Senegal and the Kolombiné rivers. The first river flows between southern Assaba and Felou, while the second is seasonal and flows between Afollé and Felou mountains. The Senegal River also bisects the Felou mountains in Mali, separating the west populations of Felou (west from the Senegal) from all the others. All the mentioned sub-basins within Mauritania and between Mauritania and Mali mountains flows to the lower Senegal river and all include *gueltas* located upstream, including smaller tributaries, at the base of mountains (Campos *et al.*, 2012; Velo-Antón *et al.*, 2014). In general the climate for the region tend to be dry and is characterized by cool (November – January) and hot (March – June) periods.



**Fig. 4** – Principal mountain massifs (light grey background under black text), two relevant escarpments (light grey background with underlined black text) and river valleys (dark grey background under white text) in Mauritania.

The likely isolation of Mauritanian populations of *F. vae*, is suggested to be related with the reduction of the savannah-like ecosystem and expansion of arid conditions after mid-Holocene (Kröpelin *et al.*, 2008; Vale *et al.*, 2012a). *Gueltas* are generally isolated and spatially restricted, being susceptible by droughts and extreme temperatures (Vale *et al.*, 2015). The droughts of the 1970s (Brooks, 2004; Ahmed *et al.*, 2008; Mahé & Paturel, 2009) contributed for the northernmost Mauritania *gueltas* water *run-out* (Trape, 2009). In many *gueltas*, water is only abundant during the rainy period from July to September, when the pools are filled up by abrupt waterfalls (Cooper *et al.*, 2006; Brito *et al.*, 2011). Therefore, past drought events in Sahel region may have affected this species life history (Padial & Tellería, 2009), and vulnerability to climate changes in the future can also be expected, as climate models predict more droughts for the region (Brooks, 2004; Held *et al.*, 2005).

Threats to the species are generally not well known, however it has been considered threatened by the general desertification and deforestation of the Sahel (Schlitter, 1989; Amori & Gippoliti, 2003). Probably there has been a decrease in habitat quality as a consequence of the severe droughts in the region from the 1970s onwards (Ahmed *et al.* 2008; Mahé & Paturel 2009; Vale *et al.*, 2012a). A current problem that might also threaten the species, is the increasing human pressure in *gueltas* after the droughts of 1970's (Vale *et al.*, 2015). These fragile humid habitats surrounded by dry environment, for instance in the Tagant mountain in Mauritania, are exposed to activities as excavating pools, pumping water and fecal contamination of water by local agriculture (Vale *et al.*, 2015).

## 1.5. Assessing biodiversity through an integrative approach combining phylogeography and ecological niche based modeling

Phylogeography is a sub-discipline of biogeography, and is concerned with the evolutionary processes behind the geographical distributions of contemporary genealogical lineages at intraspecific level (Avice, 2000). Its aim is to shed light upon the causal relationship among geographic phenomena, species distributions and the mechanisms driving speciation (Hickerson *et al.*, 2010).

In a conventional way, the phylogeographic patterns are interpreted in a descriptive way, by combining inference from estimated gene genealogies and/or gene networks (estimates of gene flow or divergence times) with summary statistics (partition genetic differences within and between populations) (Hickerson *et al.*, 2010). Permutation tests on various summary statistics were also commonly used to test for demographic expansion and/or geographic structuring of genetic variation (Excoffier *et al.*, 1992; Fu & Li, 1993; Tajima, 1989). This information together gives some light onto the history of populations and is widely used to obtain a general overview of genetic structure. The statistical methods pushed the field from descriptive methods into using coalescent based models for parameter estimation (e.g. historical population sizes, divergence times, and migration rates given the stochastic timing of coalescent events), delineate *a priori* hypothesis, estimate of spatially-explicit demographic histories among many other aspects (Hickerson *et al.*, 2010).

One of the main “advantages” in this field relies on its comparative component between genetic data from multiple co-distributed species (overlapping in time and space). By comparing historical patterns of gene flow and divergence among co-distributed species, comparative phylogeography has helped elucidate the influence of shared historical events (geographic, geological, and/or climatological) on current patterns of biodiversity (Hickerson *et al.*, 2010). However, sometimes different patterns of genetic variation for a particular landscape configuration can be found. This might be the results of different dispersal capabilities, ecologies and/or natural histories that leave different genetic signatures as taxa experience species-specific bottlenecks and/or rates of population expansion as they move across a landscape (e.g. Knowles & Alvarado-Serrano, 2010; Marske *et al.*, 2012). Nonetheless, inferences from such regional studies have helped inform conservation priorities by identifying areas with important effects on the evolutionary histories of co-distributed species.

Phylogeography is rapidly becoming one of the most integrative fields in evolutionary biology, for instance phylogeographic parameter estimates can be potentially combined with ecological niche models and thus increasing perspectives in biogeography and evolutionary biology (Alvarado-Serrano & Knowles, 2014). This integrative approach is showing very promising for understanding how divergence and speciation are linked to abiotic factors. Such approaches have allowed the characterization of species environmental preference and tolerance (e.g. Barata *et al.*,

2012), identification of potential location of stable climatic areas across time (known as refugia) (e.g. Waltari *et al.*, 2007) and understanding the role of past climatic changes in spatial and genetic connectivity (e.g. Velo-Antón *et al.*, 2013; Martínez-Freiría *et al.*, 2015).

Behind the approach of different correlative ecological niche modeling (ENM) methods there is a common statistical procedure that establishes an association between the locations occupied by a species and a set of environmental variables of these locations, in order to identify the environmental niche and its potential distribution (Peterson *et al.*, 2011). This environmental niche can then be projected into a geographic space, which allows tracking down geographic areas environmentally similar to those preferred by the species (considered as suitable), both for the present, past and future. Therefore, constitute a valuable tool to evaluate or develop phylogeographic hypothesis. One of the most common applications of ENMs in phylogeography is to interpret patterns of genetic variation based on the post hoc concordance between patterns of genetic divergence and projections of species distribution (Alvarado-Serrano & Knowles, 2014). For example at a species level, a deep phylogenetic split between geographically proximate populations may be suggestive of long-term isolation. Support for this hypothesis might be obtained by visual inspection of projections of past species distribution, where the distributional disjunctions (areas of low suitability that may act as plausible isolating barriers) correspond to the genetic differentiation observed between regions (Alvarado-Serrano & Knowles, 2014). These analyses are typically qualitative, seeking corroborative evidence visually from the ENMs. Ecological models can also be used to search for potential niche divergence between different phylogeographic lineages (Peterson *et al.*, 2011). For instance, niche divergence across many different environmental factors might be assessed by ordination, in which the differentiation of phylogeographic lineages along environment axes is characterized (Alvarado-Serrano & Knowles, 2014).

### 1.5.1. Molecular markers

Analysis of mitochondrial DNA (mtDNA) data at the species level was the starting point to the empirical development of phylogeography and, currently, is still among the most used data source for most initial phylogeographic research (Weiss & Ferrand, 2007; Hickerson *et al.*, 2010). The preference for this type of marker is mainly due to its lack of recombination, uniparental inheritance (maternal), small effective population size (shorter expected time to reciprocal monophyly between geographic regions) (Hickerson *et al.*, 2010) and higher rates of PCR amplification success, which is quite valuable for non-invasive samples (mtDNA is found in hundreds to thousands of copies per cell, compared to the 2 copies of the nuclear DNA) (Waits & Paetkau, 2005).

The general slow rate of evolution of nuclear DNA, compared to mtDNA (like protein-coding genes as cytochrome b), granted its use for deeper phylogenetic history among genera or families. However, this is not the general rule, since depending on the region (coding or non-coding gene) and if the gene is under selection or not, it can have enough resolution at population levels (Zhang & Hewitt, 2003). For more recent events (the last 10 thousand years) highly variable markers are more useful, such as microsatellites, in revealing fine-scale population structure (Zhang & Hewitt, 2003). Noteworthy, each gene of the mitochondria (37 genes coding for two rRNAs, 22 tRNAs and 13 polypeptides) and the control region have different evolutionary rates, with the most conservative ones allowing studies at categorical levels as phyla. Therefore, depending on the objective to reach, different regions like coding, non-coding regions or the type of DNA as nuclear or mitochondrial allow drawing conclusions at different time scales and assess different taxonomic levels.

The individual history of any particular locus is highly stochastic, which means that each gene can have limited confidence in reconstructing a species history, since they represent a single window (among many others) of evolution (Zhang & Hewitt, 2003). Reliable inferences on population structure and evolutionary history should therefore be drawn from genealogical data of several independent loci (Hare, 2001). It is also important to have the notion that the genetic outputs from mitochondrial DNA and nuclear DNA might be discordant. Besides the stochastic processes of lineage divergence, the occurrence of gene flow, introgression and incomplete lineage sorting

are among the most well documented cases behind these contrasting patterns (Machado & Hey, 2003; Pinho *et al.* 2008; Boratyński *et al.*, 2011; Toews & Brelford, 2012; Boratyński *et al.*, 2014). Moreover, we must not forget that mtDNA markers are restricted to female historical processes (Hurst & Jiggins, 2005).

## 1.6. Non-invasive sampling

Non-invasive (NI) genetic sampling have become increasingly more popular since their first use in wild animals around two decades ago. This is mostly attributed to its great advantage in study wildlife without handling, disturbing or even observing individual animals. Thus, causing zero stress for the animals and allowing the study of rare, elusive and endangered species. One of the most commonly used sources of DNA for NI studies is scat samples. Fecal material can yield, among other things, information about population genetic parameters and phylogeography, as genetic diversity, population structure, gene flow, demographic history (see Waits & Paetkau, 2005 and Beja-Pereira *et al.*, 2009 for studies references). Its attractiveness is because: 1) animals defecate regularly; 2) it provides more information (e.g. diet, stress hormone status, reproductive hormones, etc.) (Beja-Pereira *et al.*, 2009); 3) relatively easy detection; 4) sampling, conservation and transportation are rather inexpensive (Fernando *et al.*, 2003); and 5) genetic analyses have been found to yield comparable results (Boston *et al.*, 2012).

The DNA from the target species is obtained from the sloughed intestinal epithelial cells present in the surface of the scats. These cells remain relatively intact, since the surface of the scat was the last portions to contact the intestine and is among the first parts to dry after deposition (Fernando *et al.* 2003). Therefore, several sampling strategies are focus in obtaining the most of these cells by peeling off, scraping or washing the outer layer of the scat before DNA extraction. However, extracting DNA from scats is usually more laborious when compared to blood and tissues because of the presence of large amounts of PCR (polymerase chain reaction) inhibitors mixed in fecal material (Waits & Paetkau, 2005). For example, inhibitors can be complex polysaccharides, products from food degradation (e.g. secondary plant metabolites, acids, enzymes, lipids and proteins), RNA and bacteria (Beja-Pereira *et*

*al.*, 2009). In addition, both quality and quantity of scat DNA is normally low and negatively affected by the exposure to local ambient conditions, the presence of digested food items, and natural degradation processes (Piggot, 2004; Hájková *et al.*, 2006). These factors dramatically reduce the amplification success of fecal samples and impose the need for pilot optimization analyses and multiple replicates, which overall increase the cost of such studies (Fernando *et al.*, 2003). Contamination is another major problem in NI samples, due to the small quantities of DNA involved. Thus special precautions must be taken during field collection and in laboratory. For the first case the use of gloves and sterilization of materials between samples is required while for the second the DNA extraction and PCR should be spatially separated, a minimum of one negative control should be included and the “good lab techniques” should be followed (as described in Waits & Paetkau, 2005 and Beja-Pereira *et al.*, 2009). There is still no optimal method both for sample preserving and DNA extraction and thus amplification success may vary with species, geographic region, diet, season and PCR primers (Waits & Paetkau, 2005; Hájková *et al.*, 2006). For instance, in herbivores animals living in arid and semi-arid habitats, it is quite normal to produce pellet-form scats that dry quickly after deposition (Wehausen, 2004). The natural environmental conditions experienced in the Saharan region helps to dry the outermost layer of the pellet, reducing the microbial degradation of DNA and increasing its extraction success (maximize DNA and minimize inhibitors) (Wehausen, 2004). There are many protocols for scats preservation by inhibit enzymes that degrade DNA (e.g. removal of water, elimination of cations, low temperatures), for scat DNA extraction where commercial kits are widely used (easy to use and adaptability to a wide range of biological samples) and there are also, many hints to improve PCR amplification of DNA. For instance, PCR treatments can include diluting the extract (overcome PCR inhibitors), amplifications of smaller fragments (overcome DNA degradation and fragmentation), double amplification with internal primers and increase the number of annealing steps in the PCR (overcome low DNA quantity) (see Beja-Pereira *et al.*, 2009 for a review). However, in order to increase the application to genetic studies it needs constantly new methods less time-consuming and that increase the cost-effective in order to decrease the costs given mostly by the intrinsic limitations of the samples. Thus, finding more efficient methods of collection, preservation and DNA extraction that are viable in the field and laboratory are the major goal in this field and the technological advances are giving an important boost to it (e.g. Ramón-Laca *et al.*, 2015).

## 1.7. Objectives

The main goal of this study is, through non-invasive samples and molecular markers (mitochondrial and nuclear) understand the phylogeographic structure in *Felovia vae*, which is known from the often remote mountains of Mauritania and western Mali. We attempted to identify barriers to dispersal and gene flow between *F. vae* populations. We pursued to understand the influence of Pleistocene climatic oscillations on *F. vae* corridors for dispersal by integrating phylogeographic and ecological niche-modeling approaches, and to compare it to currently known phylogeographic patterns described for other taxa in North West Africa. Additionally, we assessed the preliminary molecular phylogenetic relationships among all the members of the Ctenodactylidae family.

Specifically the questions that we want to address using the two approaches are as it follows:

1. Molecular approach: Which are the phylogenetic relationships among the members of Ctenodactylidae family? Is there genetic sub-structuring within *Felovia vae*? Which is the distribution of the recovered clades? Where are located putative landscape barriers between populations? How has population size changed throughout time?
2. Ecological niche-based modeling approach: Which are the most important environmental factors related to the distribution of *F. vae*? Are there any differences in the ecological niches among clades? Where are located potentially suitable areas in the present and where were they located in the past? Where are the main refugia located?

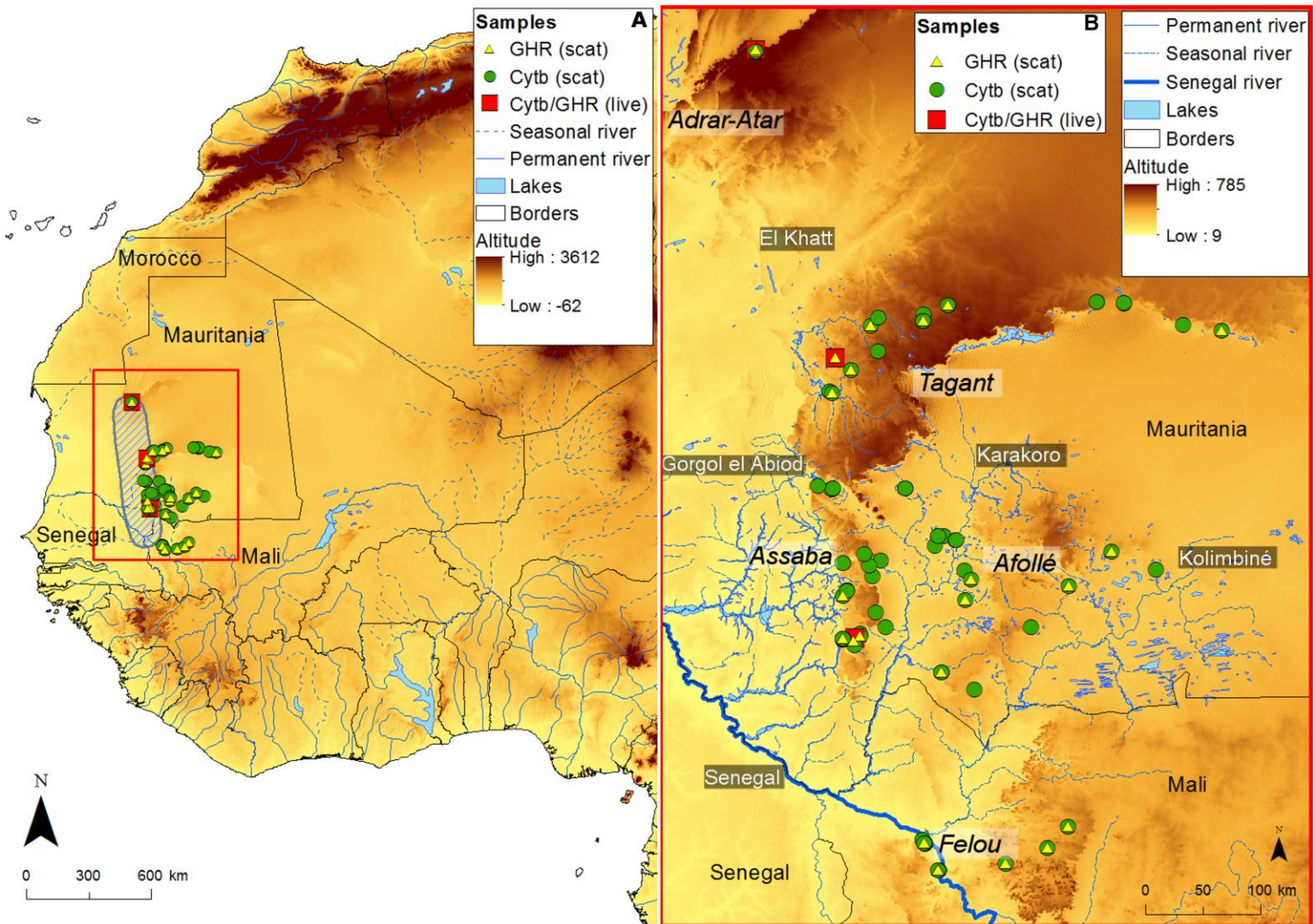
## 2. Material and Methods

### 2.1. Molecular approach

#### 2.1.1. Study area and Sampling

The study area includes the known distribution of the West Africa endemic *Felovia vae*, being apparently restricted to four mountain massifs in Mauritania (Adrar Atar, Tagant, Assaba and Afollé) and one western Mali mountain (Felou) (Vale *et al.*, 2012a) (Fig. 5). The Mauritanian mountains are separated by one permanently dry river valley and two seasonal hydrographic sub-basins: the El Khatt (20 km wide); the Gorgol el Abiod (2 km wide); the Karakoro (25 km wide). The southern Mauritanian mountains and Mali are separated by the Senegal (permanent) and the Kolombiné (seasonal) rivers.

A total of 148 samples from *Felovia vae* comprising 143 scats, two museum samples (dry skins), and three tissue samples (pieces of ears) of live-captured *F. vae* were used for DNA extraction (Tab. S1). The scats and tissue samples were collected during expeditions to Africa (November, 2010-2014) and their location was recorded with a Global Positioning System (GPS) on the WGS84 datum and represented on map with Geographical Information ArcMap 10.0 (Fig. 5). The museum samples did not have geographical information. The live tissue samples were preserved in ethanol (96%) until DNA extraction. The fecal pellets were collected from communal latrine sites in the field, approximately 20-50 per location with variable deposit age and sizes. The samples were dry preserved and sealed in plastic bags and stored in the facility at host institution at room temperature and without sunlight (CIBIO, University of Porto). A total of 23 samples (6 scats and 17 museum specimens) from the other Ctenodactylidae family members were included in DNA extraction. The 11 samples from London Natural History Museum (dry skin) were sampled and provided by museum curator. The 6 samples from the Vienna Natural History museum were collected in person (2014) from specimens conserved in absolute ethanol and dry mounted skin.



**Fig. 5** – Study area: A) African geographic distribution of *Felovia vae* samples used in phylogenetic analyses for Cytb (n=94; white circles) and GHR genes (n=26; black triangles). In the background is the known distribution of *F. vae* (Aulagnier, 2008: [www.iucnredlist.org](http://www.iucnredlist.org)) B) Small inset map representing the study area in Mauritania and Mali. More than one sample can be found in one dot. Samples from the other Ctenodactylidae member, both amplified in this study and from GenBank are not represented (unidentified country and location for most samples).

### 2.1.2. DNA extraction, amplification and sequencing

One scat pellet from one location was used to pursue the “isolation of genomic DNA from forensic case work samples” protocol, with half of the reagent volumes as described by Barbosa and co-workers (2013), using other commercial extraction Kit (E.Z.N.A.® Tissue DNA Kit). The DNA of the museum samples was extracted using the QIAamp® DNA Micro Kit (QIAGEN) and following the “Isolation of Total DNA from Nail Clippings and Hair, (QIAamp®)” protocol. All pre-PCR procedures were performed in a physically isolated sterile room with conditions optimized for the manipulation of low quality DNA in order to reduce risk of DNA contamination (Gilbert *et al.*, 2005). The fresh tissue DNA was extracted using the QIAamp® DNA Micro Kit (QIAGEN), following the “Stool Tubes for Isolation of DNA from Stool for Human DNA Analysis, (QIAamp®)” protocol. For both non-invasive and invasive samples a negative control was used in all extractions to check for contaminations. Extracted DNA was stored at -20 °C. The whole cytochrome b (Cytb) gene (1140 bp) was amplified using the primers *L7* and *H6* from Montgelard and colleagues (2001). The exon 10 of the growth hormone receptor (GHR) gene (around 364 bp) was amplified using the primers *GH7* and *GH8* from Lecompte and colleagues (2008). The Cytb was amplified with two universal primers described by Palumbi and colleagues (1991), *GLUDGL* and *CB3H*, resulting in a fragment with 800 bp length. Shorter DNA fragments were selected for Cytb to increase the amplification success in non-invasive samples (and also in degraded tissue samples). Primers *cb2F* and *cb2R* are internal to the fragment amplified with *GLUDGL* and *CB3H* primer pair, producing two fragments with 400 bp each (*GLUDGL* – *cb2R*; *cb2F* – *CB3H*) (Palumbi *et al.*, 1991).

Amplifications were performed in 10 µL, containing: 5 µL of QIAGEN PCR MasterMix [consisting of QIAGEN Multiplex PCR buffer with a final concentration of 3 mM MgCl<sub>2</sub>, dNTP mix, Q solution and HotStart Taq DNA polymerase (QIAGEN, Valencia, CA)]; 0.2 µM of each primer and 1 µL of DNA extraction (approximately 10 ng of genomic DNA). PCR reactions were performed in MyCycler (BIO-RAD) and carried out with a first denaturation step at 95 °C for 15 min, followed by 45 cycles at 95 °C for 30 s, 53 °C (for partial Cytb) / 56 °C (for GHR) for 30 s, 72 °C for 30 s and then a final extension step at 60 °C for 10 min. For the complete Cytb fragment, the annealing and extension time were increased to 1 min and 45 seconds, respectively. Also the

annealing temperature was different, around 56 °C. The PCR programs, although worked for the majority of samples, were subjected to re-adjustments for other samples with lower quantities of DNA (e.g. often by decreasing the annealing temperature by 2 °C). Pre- and post-PCR manipulations were conducted in physically separated rooms, always including negative controls, which allowed confirming absence of contaminations. PCR products were purified using ExoSAP-IT® PCR clean-up Kit (GE Healthcare) and sequenced for both strands in 3130XL Genetic Analyzer following the BigDye® Terminator 3.1 Cycle protocol (AppliedBiosystems). Sequences were aligned in SeqScape v2.6 (AppliedBiosystems, [www.appliedbiosystems.com](http://www.appliedbiosystems.com)) and alignments were manually proofed and any discrepancies reassessed. All the DNA sequences were re-sequenced on both directions (and ultimately re-amplified) to reduce the effect of background “noise” and double peaks uncertainties. Sequences obtained for the *GHR* gene fragment were phased using the software PHASE v.2.1.1 (Stephens *et al.*, 2001), to reconstruct the phase of each haplotypes at heterozygous positions. The heterozygous states were commonly identified as strong double peaks of similar height for both strands. The analysis were run three times for 10000 iterations (10000 burn-in steps), and the haplotypes selected had a minimum base call probability of 80%. Moreover, evidence of recombination events was assessed in the partial exon 10 of the *GHR* using DnaSP v.5.0 (Librando & Rozas, 2009) since it might influence phylogenetic analysis (Schierup & Hein, 2000). All the sequences were collapsed into unique haplotypes using DnaSP v.5.0 software, to facilitate the phylogenetic analysis.

### 2.1.3. Phylogenetic analysis

The phylogenetic trees for *Cytb* and *GHR* were rooted with members of Ctenodactylidae family from sequences available in GenBank (*Ctenodactylus vali*: AJ389532; *Massoutiera mzabi*: AJ389533) and generated in this study (*Ctenodactylus vali*, *Massoutiera mzabi*, *Pectinator spekei* and *Ctenodactylus gundi*) – a “*Felovia* tree”. Additionally, in order to infer the phylogenetic relationships among members of the Ctenodactylidae family a sub set of *Felovia vae* and other members of family sequences were choose, along with one member from a sister taxa Diatomyidae

(*Laonastes aenigmamus*: HQ687474) and two Hystricognathi (*Heterocephalus glaber*: JN242813 and *Thryonomys swinderianus*: NC002658) – a “Ctenodactylidae tree”.

Phylogenetic relationships were assessed with Bayesian Inference (BI) and Maximum Likelihood (ML) methods. The most appropriate substitution models of sequences evolution for each locus alignment were selected with jModelTest v.2.1.4 (Posada, 2008), under an Akaike information criterion. The BI was implemented using MrBayes v.3.2 (Ronquist *et al.*, 2012), which uses a Markov Chain Monte Carlo (MCMC) sampling approach. All analyses started with randomly generated trees and ran for 10 million generations, with a sampling frequency that provided a total of 10 000 samples for each run, excluding the first 25% sampled trees as burn-in. The stationary phase was guaranteed when the standard deviation of split frequencies were below 0.01 (Ronquist *et al.*, 2012). The Maximum likelihood (ML) phylogenetic analyses were performed using PHYML online webserver (Guindon *et al.*, 2005). The program estimated the base frequencies, the ts/tv ratio, the gamma distribution parameter and the proportion of invariable sites. The input tree was determined with the BIONJ algorithm, tree topology was optimized and 1000 bootstrap re-sampling was used for each run. The stability of nodes of the phylogenetic trees was evaluated with Bayesian posterior probabilities (bpp) for BI and by bootstrapping (pseudo-replicates) for ML. For both methods, three independent runs were generated for each locus to compare results. The resulting trees were visualized and edited using the software FIGTREE v.1.3.1 (Rambaut, 2009) and other image editors (e.g. GIMP).

To compare genetic divergence between members of the Ctenodactylidae family and outgroups used in this study, net average pairwise distances were calculated in MEGA v.6.0 (Tamura *et al.*, 2011) under the Kimura 2-parameter model (K2P) with 10000 bootstrap replicates.

#### 2.1.4. Time of divergence

Time to most recent common ancestor (TMRCA) for the Ctenodactylidae family members (“Ctenodactylidae tree”) and for the main *F. vae* mtDNA clades (“*Felovia vae* tree”) was estimated using BEAST v.1.7.5. with three different approaches (Drummond *et al.*, 2012). For the “Ctenodactylidae tree”, three calibration constrains were applied:

(1) setting a prior distribution for the TMRCA of *Laonastes* and *Ctenodactylus* following a normal distribution with a mean of 44.3 Mya and a standard deviation of 3.5 Mya based on divergence estimates given by Huchon et al. (2007); (2) based on the divergence between *Massoutiera* and *Ctenodactylus* fossil lineages (9.7 – 13.0 Myr) (Jaeger, 1977; Huchon *et al.*, 2000), in which the age of the preceding branch in the MRCA node was taken into account (“Include stem” option, because the fossil could be older than the MRCA of the sampled extant terminal taxa) and a lognormal distribution with an offset of 9.7 Mya (both mean and standard deviation = 1.5) was followed; and (3) the general value for Cytb evolutionary rate of 2% per million years with a normally distributed prior (mean=0.02 and standard deviation=0.005), based on rodent literature (Triant & DeWoody, 2006; Nabholz *et al.*, 2008; Ben Faleh *et al.*, 2012). Three approaches were used to compare the estimates for the TMRCA: A) included all the three calibration constraints; B) included just the inside family fossil and the general evolutionary rate for Cytb; and C) included only the 2% evolutionary rate for Cytb.

For the “*Felovia vae* tree”, estimated substitution rate for each of the three previous approaches (A, B and C), was applied to a dataset including only *F. vae* haplotypes without outgroups.

A Yule (pure-birth) model, most suitable for species-level processes, was used for the “Ctenodactylidae tree”, while a coalescence constant population size model, more adapted to population-level processes, was used for the “*Felovia vae* tree”. Preliminary analyses were carried out to determine whether the Cytb fragment evolved according to a strict molecular clock or a log-normal relaxed clock, by comparing the posterior distribution of the standard deviation. As a result an uncorrelated lognormal relaxed-clock model was used as tree prior to allow rate variation among branches (Drummond *et al.*, 2006). For each approach three independent runs of 100 million generations were implemented, sampling every 10000 generations and 10% of the trees discarded as burnin (1000 trees). TRACER v.1.6 (Rambaut & Drummond, 2007) was used to verify that the sampling achieved stationarity (ESSs higher than 300 for all parameters). Samples from each run were combined using the software LogCombiner v.1.7.5. (available in the Beast package), and a consensus chronogram with node height distribution was generated and visualized using TreeAnnotator v.1.7.5. (available in the BEAST package) and FigTree v.1.3.1.

### 2.1.5. Population genetic analysis

Diversity indices, namely haplotype number (H) and diversity ( $H_d$ ), nucleotide diversity ( $\pi$ ), number of polymorphic sites (S) were calculated for *Felovia vae* in both genes using DnaSP v.5.0 (Librado & Rozas, 2009). Two haplotype networks were constructed to visualize genetic relationships between haplotypes within *Felovia vae*. One haplotype network was constructed with the median-joining method implemented in NETWORK v.4.6.1.1. (Bandelt *et al.*, 1999), and the other network with a neighbor-net method based on uncorrected patristic distances and bootstrap analysis with 10 000 replicates, using SPLITSTREE v.4.6. (Huson & Bryant, 2006).

The partition of genetic variability was assessed by performing a three-level hierarchical analysis of molecular variance (AMOVA) using ARLEQUIN v.3.5.1. (Excoffier & Lischer, 2010). AMOVA divides the total variance into additive components, that is, variation within populations, among populations within groups and among groups. A population was defined as all individuals coming from one group of localities with less than 30 km apart. This 30 km distance was used to reduce the number of populations given by mostly one individual per locality. Groups of populations were defined according to genetic criteria (clades) and putative geographic barriers (El Khatt, Karakoro and Kolombiné river basins; Djouk valley; Senegal river). For each recovered lineage, summary diversity statistics (haplotype diversity and nucleotide diversity) were calculated using DnaSP v.5.0. The net average pairwise distances between and within clades were calculated in MEGA v5.05 (Tamura *et al.*, 2011) under the Kimura 2-parameter model (K2P) with 10000 bootstrap replicates.

### 2.1.6. Historical demography

Demographic analyses of *Felovia vae* were conducted by the Cytb discriminated clades through the phylogenetic reconstruction, haplotype networks, and AMOVA. DnaSP v.5.0 was used to calculate an array of statistics known as neutrality tests, which have variable power to detect departure from neutrality caused by population size expansion and background selection (Fu, 1997). Tajima's D test

principle is that the number of nucleotide differences between sequences from a random sample should be equal to the number of differences between the polymorphic sites only. Significant negative values suggest fit to a model of sudden population expansion. Fu's  $F_s$  (Fu, 1997) and Ramos's  $R_2$  statistics are among the most powerful test available to detect selection and/or demographic expansions (Ramos-Onsins & Rozas, 2002). The robustness of these tests relies on their statistical power to detect significant population changes in small sample sizes (Ramos-Onsins & Rozas, 2002).  $F_s$  represents the probability of observing a similar or higher number of haplotypes in a random neutral population, while  $R_2$  stands for the difference between the number of singleton mutations and the average number of pairwise differences. Therefore, the excess of rare alleles contained in populations that undergone recent population expansion, will lead to negative values of these statistics. The significance of all these tests was calculated using 10,000 coalescent simulations (bootstrap method).

Mismatch distributions were also calculated in DnaSP v.5.0., to compare the observed frequency distribution of pairwise differences between haplotypes with that expected under a model of sudden population expansion. A unimodal distribution is consistent with sudden demographic expansion of population. A bimodal distribution might be related with past constant effective population size or admixture of distinct populations (Hamilton, 2011). Finally multimodal patterns can indicate strong population subdivision that persisted through small populations or a population bottleneck (Hamilton, 2011). One statistical test relying on the raggedness index (Harpending *et al.*, 1993; Harpending, 1994) was calculated and compared with data simulated under an expansion model. Significant  $P$  values indicate rejection of the recent expansion hypothesis and were calculated using 10,000 coalescent simulations (bootstrap method).

The maximum-likelihood (ML) method implemented in the software fluctuate (Kuhner *et al.*, 1998) was used to estimate the growth parameter  $g$  and theta  $\theta$  ( $= 2N_f\mu$  - where  $N_f$  represent the female effective population size and  $\mu$  the mutation rate; not shown) for each lineage. An initial growth factor of 100 and Watterson's (1975) estimate of theta were used as starting values. A random tree was chosen to start the analysis, that were ran for 50 short (10 000 steps for each short chain) and 15 long chains (50 000 steps for each long chain) with sampling every 5 steps. Five runs were performed with random seeds to check for convergence. Significant values of  $g$  had to

have standard deviations lower than one-third of the  $g$  value, while very significant values are those in which  $g > 6 \times \text{SD}(g)$  [according with Lessa *et al.*, (2003)].

A coalescent-based Bayesian skyline plot (BSP) approach was used to provide a temporal perspective on the historical changes in population size for *Felovia vae*, as implemented in BEAST 1.7.5. No outgroup was included in the analysis for each and all clades combined together, as BEAST estimates the root of the individual gene trees (Heled & Drummond, 2010). jModelTest v.2.1.4 was used to determine the best-fit prior model of sequence evolution using the Akaike Information Criterion (AIC). The estimated substitution rate for the “best” approach of the dating process was implemented and the same clock model was used. The final MCMC simulation ran for 100 000 000 generations, sampling every 10 000 generations and a 10% burnin was used. TRACER v.1.6 was used to check for convergence and stationarity of the effective population sizes.

### 2.1.7. Spatial analyses

To visualize geographical patterns of genetic structure and variability, both genetic distances (uncorrected  $p$ -distance) and nucleotide diversity ( $\pi$ ) were spatially interpolated using a kriging interpolation method (Oliver, 1990). Putative barriers to gene flow were assessed by calculating inter sample genetic distances (Veríssimo, 2014). The genetic distances between each point to all the others, given by the genetic matrix columns, were treated independently and interpolated using the kriging interpolation method. A Principal Components Analysis was used to summarize the results, using the Principal Components Analysis implemented in the R package “raster”.

Genetic diversity was calculated by sub-sampling locations contained in a buffer of  $0.243^\circ$  (approximately 25 km) around each sample location. Buffer value was arbitrary established. The given populations were only considered in case at least one more individual (neighbor point) was contained inside the buffer of the original sample. Nucleotide diversity values were interpolated as a continuous surface through the kriging interpolation method, implemented in the “Geostatistical Analyst” extension of GIS ArcMap 10.0. The final raster was reclassified into 9 classes, following Natural Jenks as division criteria.

## 2.2. Ecological niche-based modeling approach

### 2.2.1. Species occurrences

A total of 178 occurrences at 1 x 1 km grid were used to develop ecological models. The duplicates, given by two occurrences in the same 1km grid, were excluded from the initial dataset (n = 120). The individuals detection was based on directly observed individuals (n = 52) and scats (n = 68) gathered in the field from the period between 2003 and 2014.

### 2.2.2. Ecogeographical variables

Ecogeographical variables (hereafter EGVs) related to the ecology and distribution of *Felovia vae* were used in ecological models (George, 1974; Padial & Tellería, 2009; Brito *et al.*, 2010; Vale *et al.*, 2012a). The EGVs for current and past conditions were the same and included: one topographical grid (USGS, 2006) that was used to derive Slope, with the “Slope” function of Geographical Information System ArcGIS (ESRI, 2010); and six climatic grids from WorldClim (version 1.4, see [worldclim.org/paleo-climate](http://worldclim.org/paleo-climate) for variable descriptions) (Tab. 1).

**Tab. 1** - Environmental factors (coded as BIO in WorldClim website) used for modeling the distribution of *Felovia vae* in the study area.

Type	Variable	Range and units
Topographical	Slope	0-23 %
	Annual mean temperature (BIO1)	22.0-30.8 °C
	Maximum temperature of warmest month (BIO5)	32.1-47.3 °C
Climatic	Minimum temperature of coldest month (BIO6)	9.0-18.8 °C
	Temperature annual range (BIO7)	19.5-38.2 °C
	Annual precipitation (BIO12)	15-1428 mm
	Precipitation of wettest month (BIO13)	6-415 mm

For current conditions the variables were at 30 arc sec (~1 x 1 km) resolution (see Hijmans *et al.*, 2005). For past conditions, the same seven EGVs were selected for the Mid-Holocene (~ 6000 years BP), for the Last Glacial Maximum (LGM; ~21 000 years BP) and for the Last Interglacial (LIG; ~120 000-140 000 years BP). Mid-Holocene and LGM variables were at 2.5 arc minutes (~5 x 5 km) resolution and the predictions of three general circulation models (GCM) were used for each: the Community Climate System Model (CCSM4), the Model for Interdisciplinary Research on Climate (MIROC-ESM), and the Max-Planck-Institute Earth System Model (MPI-ESM) ([worldclim.org/paleo-climate](http://worldclim.org/paleo-climate)). LIG variables were at 30 arc seconds (~1 x 1 km) resolution and integrated one GCM: NCAR-CCSM (Community Climate System Model; Otto-Bliesner *et al.*, 2006). The main difference between these GCM relies on the temperature and precipitation patterns and depends on the assumed atmospheric concentration of greenhouse gases. Variables for current conditions, Mid-Holocene, and LGM were upscaled to 30 arc seconds (~1 x 1 km) to match the resolution of LIG variables.

### 2.2.3. Ecological niche-based models

Ecological niche-based models were performed under a maximum entropy algorithm using MAXENT (Philips *et al.*, 2006). MAXENT only requires species occurrence records and background data (given by the study area) for calculating the suitable conditions of the species, calibrating a model to differentiate between the species niche and the background (Philips *et al.*, 2006). Additionally, MAXENT appears to have a good performance in comparison to other methods (Elith *et al.*, 2006), especially when low sample size is available (Hernandez *et al.*, 2006; Wisz *et al.*, 2008). The study area was delineated with a 200 km buffer around each detected occurrence.

Model parameters were calibrated from 20 replicate model runs with random seed. Species records for each replicate were randomly chosen by cross-validation, using 24 occurrences for testing the model predictions (20%), 96 for training the model (80%) and 10 096 as background points. Observations for each replicate were chosen by bootstrap allowing sampling with replacement. Models were run with default

parameters and logistic output (Philips *et al.*, 2006). Model performance was measured using the area under the curve (AUC) of the receiving operator characteristics (ROC) plot, which ranges from 0.5 (completely random) to 1 (strongly predicted) (Fielding and Bell, 1997).

#### 2.2.4. Ecogeographical variables (EGVs) importance

Variable importance for explaining the species distribution was determined from the average percent contribution and permutation importance of each EGV to the model. The relationship between species occurrence and the most important EGV was assessed by univariate response curve profiles, representing how the logistic prediction changes as each EGV is varied, keeping all other EGVs at their average sample value (Philips *et al.*, 2006). The most important EGVs were plotted together with the genetic clades previously identified by molecular studies.

#### 2.2.5. Present prediction of species occurrence

The 20 individual model replicates were added to obtain a consensus prediction model of species presence probability under current conditions (Marmion *et al.*, 2009). Moreover, since predictions can be variable between model replicates, the standard deviation (SD) was used to assess prediction uncertainty (Buisson *et al.*, 2010). Suitable areas were determined by reclassifying the previous average model to a grid map where each cell has a probability of presence and absence associated (suitability model). Analyses were developed using the minimum training presence (assumes all the observation as certain) and the 10 percentile training presence (assumes that 10% of the observation with the lowest probability as wrong) thresholds given by Maxent. However, these commonly used thresholds (see Marmion *et al.*, 2009) produced over-predictions of the suitable area. Therefore, the threshold of minimum probability for species presence ( $p=0.215$ ) was adjusted in order that all observation ( $n=120$ ) occurred in grid cells of probable presence, and each cell had an index of suitability

above 0.0. Values close to zero indicate unsuitable conditions for the species to occur, while values close to one represent suitable conditions.

### 2.2.6. Past prediction of species occurrence

The model was calibrated using the current range of the species, and then projected into seven past scenarios: three for Mid-Holocene (CCSM4, MIROC-ESM and MPI-ESM), the same three for LGM and one for LIG (NCAR-CCSM). Projections were made for a larger area, to cover potential historic places and refuges outside the species “home range”, reducing bias. Similarly to consensus model for current conditions, consensus projections were used to map the species presence probability under each past period, with the respective standard deviation. The thresholds for reclassification were the same (suitability model) and the different general circulation models (GCM) were posteriorly added for each past period.

### 2.2.7. Refugia

Both current and past conditions consensus models were intersected to identify possible refugia, defined as stable climatic areas over time where the species might have persisted. The current models had to be readjusted to a coarse pixel size, compatible with the past condition models, before making refugia maps.

## 3. Results

### 3.1. Molecular approach

#### 3.1.1. Laboratory overview

Out of 148 samples initially available for *Felovia vae*, 94 (63.5 %) were successfully amplified for the Cytb partial fragment (700 bp length). From the 148 samples, three were from fresh tissue (100% success), two from museum dry skin (100%) and the others 143 from scats (65.7%). The sequenced individuals represent all the area sampled, which is known to cover the species current range. Moreover, the 3 fresh tissue samples and one scat (high quality) were amplified for the complete Cytb gene with success. In this case the Eastern Afollé mountain was not covered. For the chosen Cytb sub-set of 34 samples, 26 (76.5 %) were successfully amplified for the GHR partial fragment.

Out of 23 samples (6 scats and 17 museum dry skin) available for the other members of Ctenodactylidae family, 6 (0% scats and 35% museum dry skin) were successfully amplified for Cytb. For the nuclear gene, only 4 samples (17.4 %) were amplified. Representative sequences of Ctenodactylidae family were obtained for Cytb, while for GHR *Pectinator spekei* was not successfully amplified. Detailed information of all amplified sequences can be found in Tab. S1.

#### 3.1.2. Phylogenetic analysis

##### 3.1.2.1 “*Felovia vae* tree”- Cytb

Cytochrome b sequences from *Felovia vae* were aligned and edited to achieve a final dataset of 94 sequences of 700 bp length with 69 polymorphic positions (Tab. 2). Fifty four different haplotypes were identified among the sequences analyzed (Tab. 2 and see also supporting information, Tab. S1). Haplotype diversity for *Felovia vae* was high (0.975), and its nucleotide diversity was low (0.019) (Tab. 2). The *GTR+G+I* model was selected as the most appropriate substitution model and it was subsequently used

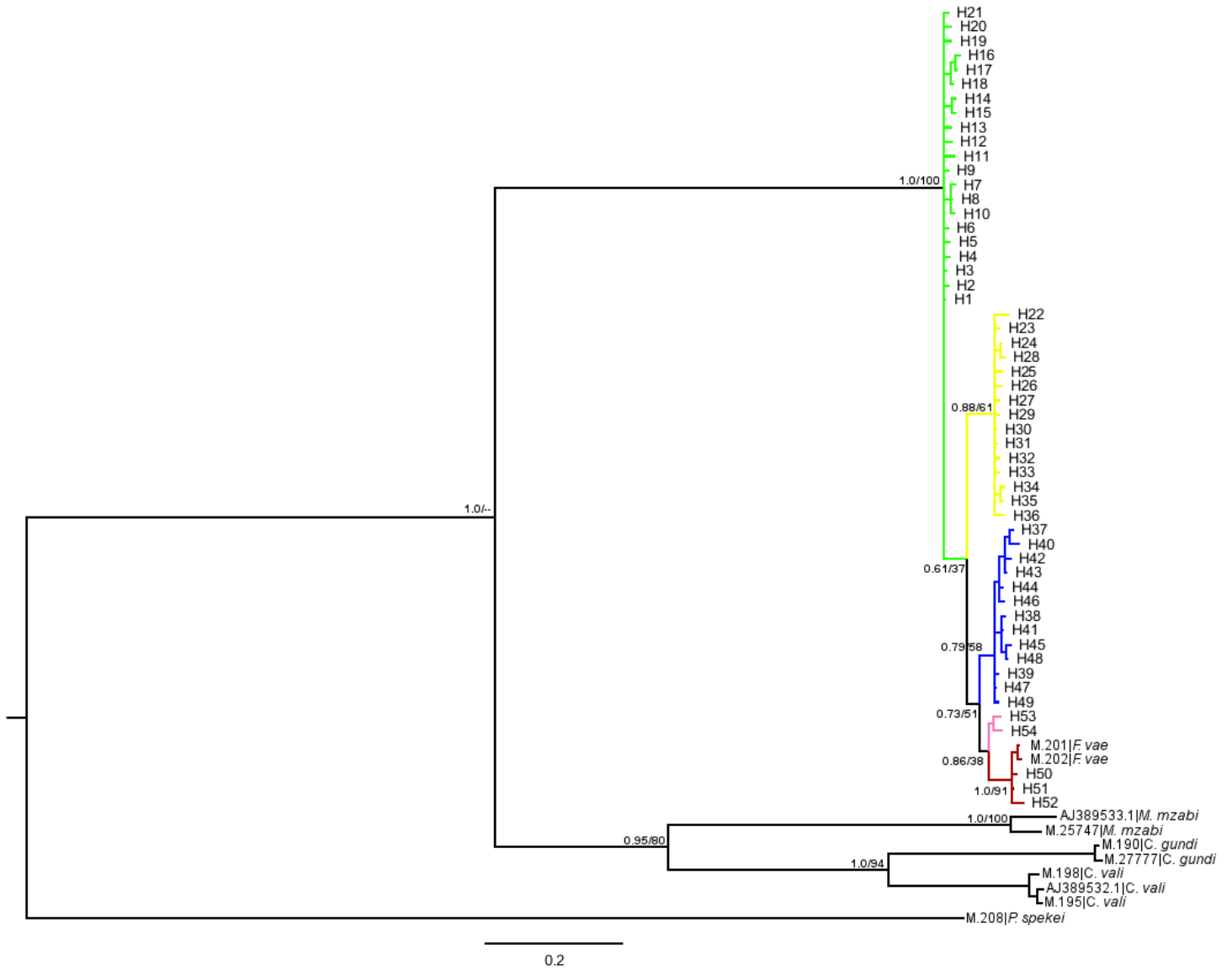
to reconstruct the evolutionary relationships between haplotypes. Both ML and BI phylogenetic trees showed the same topology, although the support in each node was always much inferior for bootstrap compared to BPP (Fig. 6). A highly supported clade includes all haplotypes of *Felovia vae*. The further clades although not so well supported (posterior probability values are generally higher than 0.7 but lower than 0.9) revealed a strong phylogeographical structure. The lowest bootstrap and posterior probability (0.61 and 37, respectively) value found in the node supporting clades B, C, D and E, made it difficult to infer the basal lineage for the species. However, five main clades (A to E) with mostly allopatric geographical distributions were identified, corresponding mainly to distinct mountain massifs (Fig. 7, Image 1):

- Haplotypes of clade A were distributed throughout the north, including the mountains from Adrar Atar and Tagant. However it was also present, in potential geographic sympatry with clade B, in one locality in the south of Assaba mountain (Guelta Goumbel; Fig. 7). Six scats were sequenced for this locality, and four resulted to be from clade A (Haplotype 21, Tab. S1) while the other two from clade B (Haplotype 36, Tab. S1).
- Haplotypes of clade B were restricted to the Central region, in Assaba mountain. In the Djouk Valley region clade A and B occur parapatrically distributed. Haplotypes of clade C were distributed in Center East region, Afollé mountains.
- Haplotypes of clades D and E were distributed in east and west side of the Senegal River, respectively, which crosses the Southern most area, Felou region. Two museum sequences from *Felovia vae* with unknown locations were placed into clade E located in the West of Senegal river in Felou mountain.

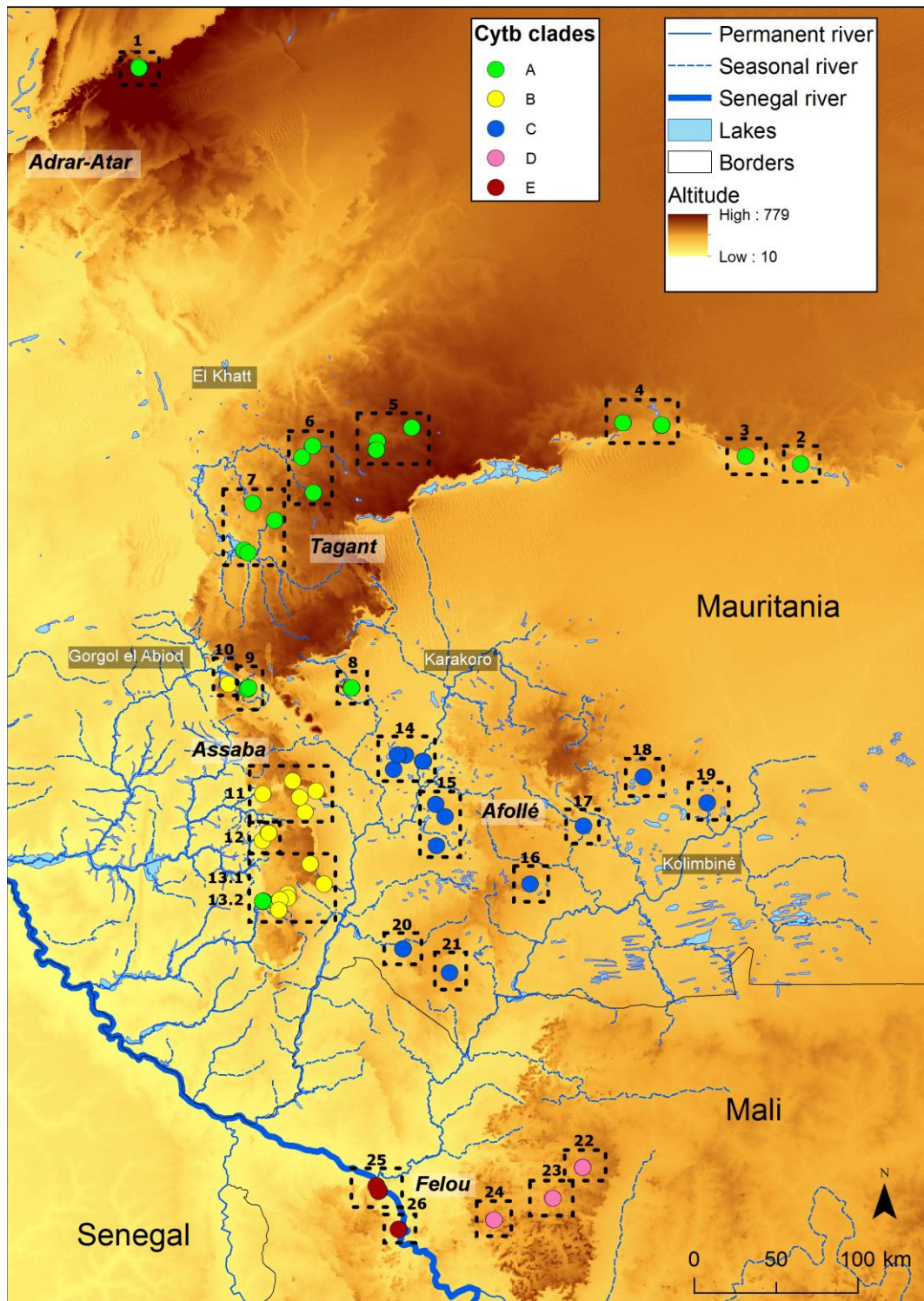
**Tab. 2** - Summary statistics and population growth for *Felovia vae* in two genes, Cytb (mtDNA) and non-recombinant GHR (nDNA).

Location	Polymorphism					Population growth					
	N	S	H	Hd	$\pi$	Tajima's D	Fs	R2	r	g $\pm$ SD	
<b>Cytb</b>											
All	Mauritania-Mali	94	69	54	0.975	0.019	-0.15434 NS	-20.520**	0.0932 NS	0.0119 NS	202.268 $\pm$ 32.568 ††
A	Atar-Tagant	35	25	21	0.951	0.004	-1.77867*	-14.941**	0.0558**	0.0365 NS	1574.22 $\pm$ 208.748††
B	Assaba	29	19	15	0.855	0.004	-1.57391*	-7.924**	0.0635*	0.0981 NS	937.645 $\pm$ 196.073†
C	Afollé	18	13	13	0.961	0.005	-0.08561 NS	-6.340**	0.1315 NS	0.0207*	1102.96 $\pm$ 229.364†
D	Felou East	3	4	2	0.667	0.004	---	---	---	---	---
E	Felou West	9	4	3	0.639	0.002	0.19871 NS	2.309 NS	0.1615 NS	0.3603 NS	19.1537 $\pm$ 95.6895 NS
<b>GHR</b>											
All	Mauritania-Mali	24	6	6	0.703	0.003	---	---	---	---	---

N, number of individuals; S, number of polymorphic sites; H, number of haplotypes detected; Hd, haplotype diversity;  $\pi$ , nucleotide diversity; neutrality test indexes (Tajima's D; Fs, Fu's; R2, Rozas's); r, raggedness statistics; g, growth rate parameter; \*\*significant at  $P < 0.01$ ; \*significant at  $P < 0.05$ ; NS Non-significant at  $P > 0.05$ ; †† g over six standard deviations (SD) above 0; † g between three and six SD above 0.



**Fig. 6** – Phylogenetic tree for *Felovia vae* based on mtDNA (Cytb) data (700 bp length; 94 sequences). The tree is rooted with outgroups from the Ctenodactylidae family. Values of Bayesian posterior probabilities and bootstrap values of ML are above or near the branches. Colors represent the main groups obtained and are disposed on map in Fig. 7.

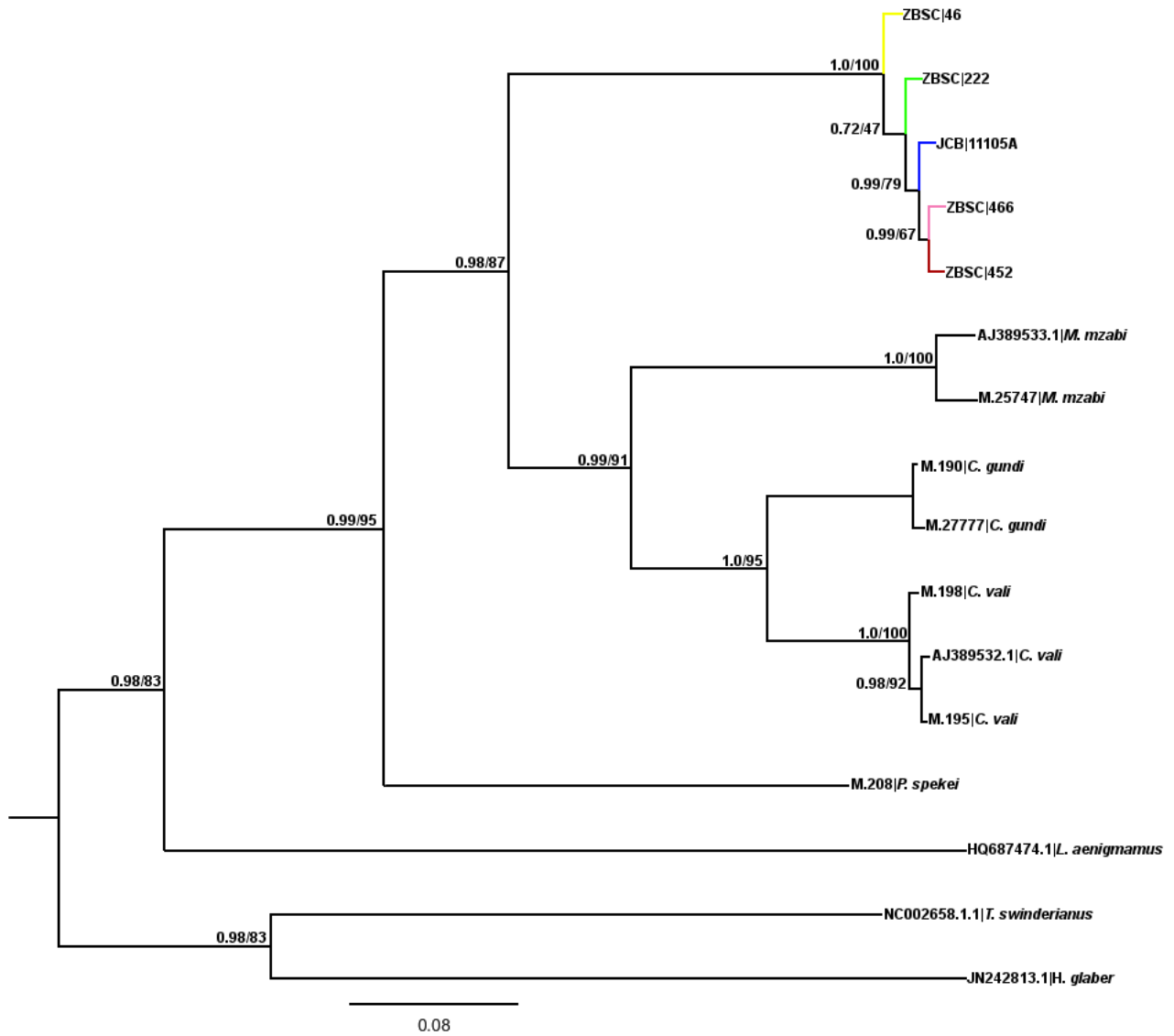


**Fig. 7** – Geographic distribution of *Felovia vae* samples belonging to each main clade (colors) recovered by the phylogenetic tree based on partial Cytb gene (700 bp length; 94 sequences). Numbers represent hypothetical populations delineated using a 30 km distance around each sample (see Tab. S1 for more information).

### 3.1.2.2. “Ctenodactylidae tree” - Cytb

Phylogenetic analyses for the Ctenodactylidae family were based on a final alignment of 10 sequences with 813 bp length. The *GTR+I+G* model was also selected. Both ML and BI phylogenetic trees showed the same topology and similar support in each node (Fig. 8). The family is highly supported and the overall node supports within the family were high (> 90 bootstrap and > 0.9 bpp). The basal position in the phylogenetic tree was occupied by *Pectinator spekei*. A monophyletic group incorporated *Massoutiera mzabi* and *Ctenodactylus* (*C. vali* and *C. gundi*), while *Felovia* formed a distinct monophyletic group. Within *Felovia vae*, the nodes between clades were highly supported, however, and once again, the relation between some clades was still shallow (Fig. 8). The same node as the previous tree had a lower support influencing the position fluctuation of clade A and B.

Net average percentages of K2P pairwise differences between species varied from 23 % between *P. spekei* and the other members to 11% between *Ctenodactylus* species (Tab. 3). Genetic distances between *M. mzabi* and *F. vae* were around 3% superior then *C. gundi* and 4% then *C. vali* (Tab. 3).



**Fig. 8** – Phylogenetic tree for the Ctenodactylidae family based on mtDNA (Cytb) data (813 bp length; 10 sequences). The tree is rooted with outgroups from the sister taxa Diatomydae and two Hystricognathi. Values of Bayesian posterior probabilities and bootstrap values of ML are above or near the branches. Colors represent the main *Felovia vae* clades obtained in Fig. 6.

**Tab. 3** – Genetic net average pairwise distances among members of Ctenodactylidae family and outgroups (Diatomyidae and Histicognathi) for partial Cytb gene. Means and standard deviation values appear below and above the diagonal, respectively.

Species	<i>C.vali</i>	<i>C.gundi</i>	<i>M.mzabi</i>	<i>F.vae</i>	<i>P.spekei</i>	<i>L.aenigmamus</i>	<i>H.glaber</i>	<i>T.swinderianus</i>
<i>C.vali</i>	*	0.013	0.017	0.018	0.020	0.023	0.023	0.021
<i>C.gundi</i>	0.108	*	0.017	0.018	0.021	0.023	0.023	0.023
<i>M.mzabi</i>	0.149	0.163	*	0.018	0.021	0.023	0.022	0.023
<i>F.vae</i>	0.178	0.185	0.188	*	0.020	0.025	0.024	0.023
<i>P.spekei</i>	0.216	0.230	0.224	0.229	*	0.023	0.023	0.022
<i>L.aenigmamus</i>	0.265	0.261	0.261	0.287	0.273	*	0.023	0.025
<i>H.glaber</i>	0.271	0.284	0.270	0.291	0.284	0.262	*	0.022
<i>T.swinderianus</i>	0.255	0.271	0.257	0.269	0.266	0.286	0.256	*

### 3.1.2.3. “*Felovia vae* tree” - GHR

A total of 24 specimens, including representatives from all the defined mtDNA clades, were sequenced for the *GHR* gene. The obtained fragments have a length of 362 bp with 6 polymorphic positions, and some heterozygous specimens (Tab. 2). The overall nucleotide variation for the phased and non-recombinant sequences (44 alleles) was much lower (0.00273) when compared to the mitochondrial marker, which resulted in unresolved phylogenetic trees due to the low level of sequence divergence (Fig. 9). The best-fit evolutionary model used was *HKY+I*, and the tree topology and node support were very similar for both ML and BI analysis (Fig. 9). This nuclear marker was not used to reconstruct the family tree, since the most basal group given by the *Pectinator spekei* for Cytb gene was missing from the dataset (Fig. 9). However, the position of *Massoutiera mzabi* in the phylogenetic tree and the lower genetic distance were closer to *Ctenodactylus* than with *Felovia* species (Fig. 9 and Tab. 4).

Recombination was detected at one location within the exon 10 of GHR in site (212, 273). The recombination event was detected in four haplotypes within *Felovia vae*, in which three of them were constituted by individuals from the Affolé mountains (R2; R8) and the other one from one individual located in Guelta Goumbel locality, in south Assaba mountain (R4) (Tab. S1).



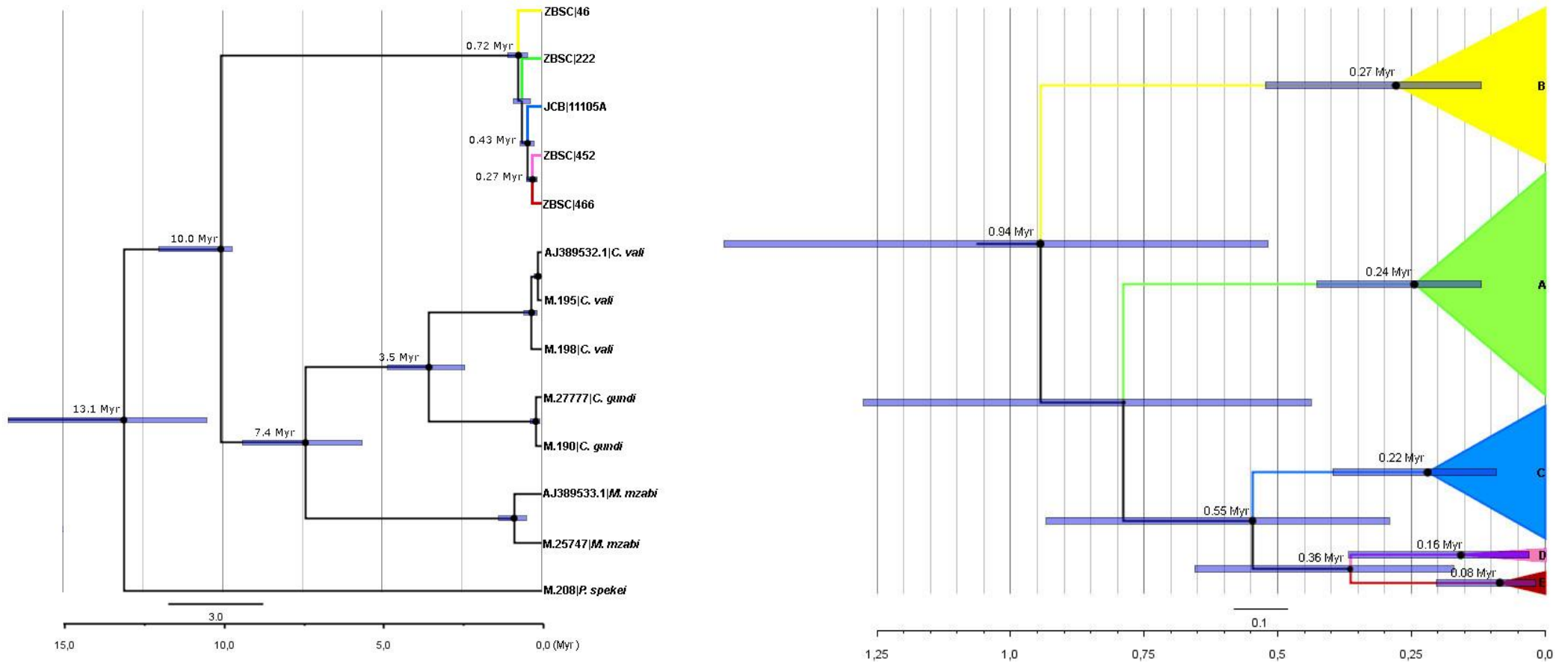
### 3.1.3. Time of divergence

For the approach B in the “Ctenodactylidae tree”, BEAST estimations showed that the most recent common ancestor (TMRCA) for the Ctenodactylidae family occurred during the early to middle Miocene [95% highest posterior density (HPD) between 10.49 and 16.73 Myr], around 13.10 Myr ago (Fig. 10, “left”; Tab. S2). The most recent common ancestor between the most *Felovia vae* and the other members of *Massoutiera* and *Ctenodactylus* genus, and between the last two was estimated to occur during the late Miocene (9.70-12.01; 5.63-9.37), around 10.01 and 7.40 Myr, respectively. The two species of *Ctenodactylus* were estimated to share an ancestor during the early and middle Pliocene (2.41-4.81), around 3.53 Myr ago. The most recent common ancestor for all clades of *Felovia vae* was estimated to live around 0.72 Myr ago during the early and middle Pleistocene (0.44-1.05). The phylogenetic relationships were still not resolved between clades A, B, “C, D and E” to support TMRCA estimation. However, the divergence estimations for clades C, D and E and also D and E were well supported and suggested their appearance during the middle Pleistocene, around 0.43 and 0.27 Myr ago (0.23-0.66; 0.12-0.45). The *GTR+G+I* model ( $\delta=0$ ) was replaced by a simpler one in BEAST, *HKY+I+G* ( $\delta=31$ ), due to consecutive low values of ESSs (below 100) in posterior and prior parameters.

For the same approach but in the “*Felovia vae* tree” the mean for the TMRCA of *Felovia* sequences was very similar with the family tree, but the credibility intervals were doubled (but remained acceptable) (Fig. 10, “right”; Tab. S3). All clades were supported by posterior probabilities higher than 0.95 and, in general, their confidence overlapped with each other and the TMRCA for most of them was estimated to occur in the middle and late Pleistocene. The oldest MRCA estimation goes for clades A (0.24; 0.12-0.52 Myr) and B (0.27; 0.12-0.52), while clade E was the most recent one (0.08; 0.016-0.20). The best-fit evolutionary model used was *HKY+I+G*.

For approaches A and C the 95 % HPD doubled for both trees when compared to approach B (Tab. S2 and S3). For approach A, these values might be explained by the incorporation of considerable uncertainties given by the divergence estimates between Ctenodactylidae and *Laonastes* calculated by Huchon *et al.* (2007). Moreover, points of calibration from outside (“Ctenodactylidae tree”) the tree of interest (“*Felovia vae* tree”) might be misleading.

Biogeography and phylogenetic position of a Sahara-Sahel mountain endemic, *Felovia vae* (Ctenodactylidae)



**Fig. 10** – Trees showing mean divergence dates in million years (above the branches/before the branches) and 95 % credible intervals (blue bars) for relevant supported nodes (given by the black dot). Left: “Ctenodactylidae tree” (Cytb) following the approach B (include two calibration constrains: fossil and mutation rate) (see Tab. S2 for detailed information). Right: “*Felovia vae* tree” (Cytb) (see Tab. S3 for detailed information) using the estimated substitution rate by the previous approach B. Colors represent the main *Felovia vae* clades obtained in Fig. 6.

### 3.1.4. Population genetic analysis

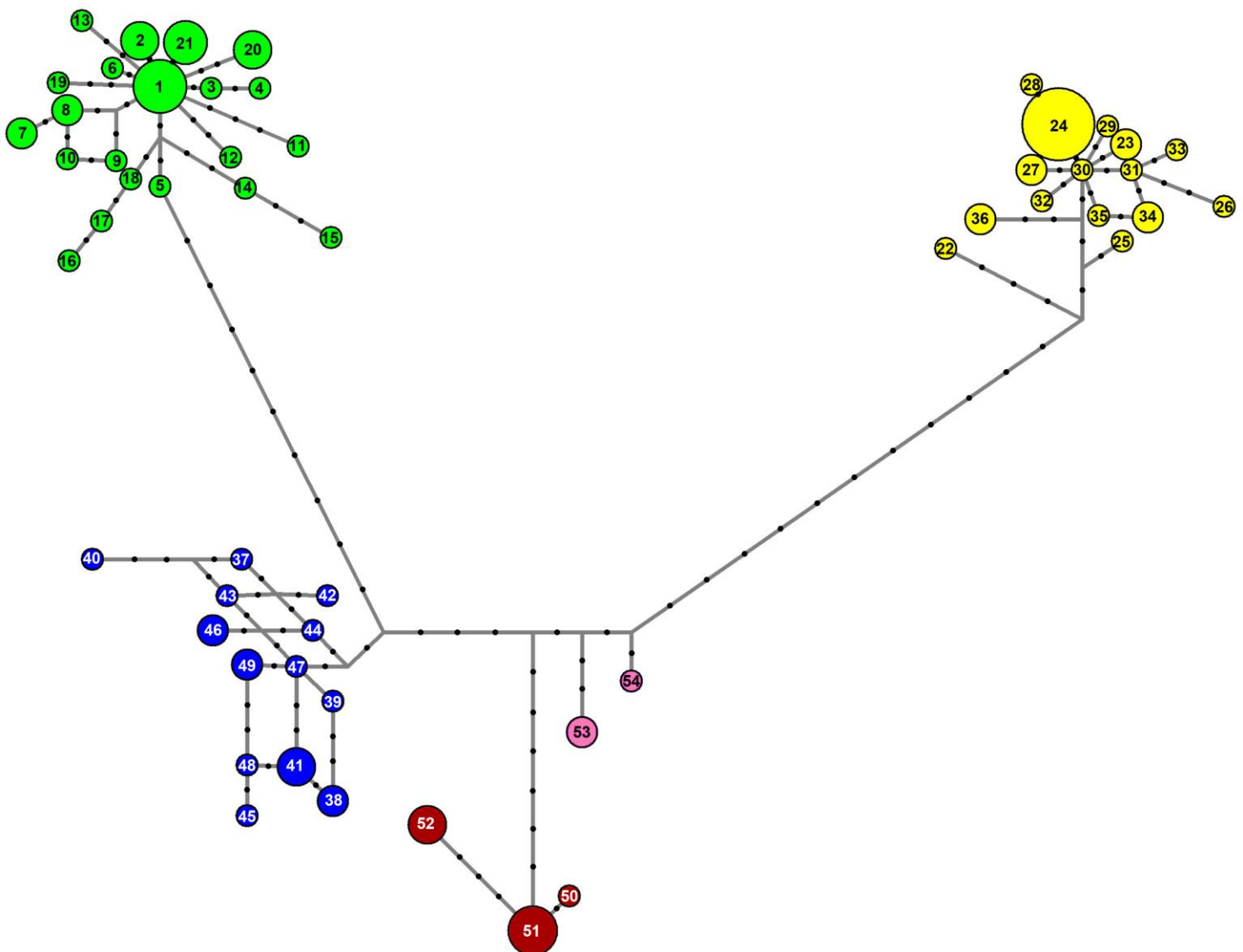
The Cytb median-joining network confirmed split to five main haplotype clusters, separated by fourteen (B-E) to eight mutational steps (C-D) (Fig. 11). Most of the samples from clade C are represented by several haplotypes (number 37-49, Tab. S1), with low sample sizes. Clade E samples share common haplotypes (number 53-54; 50-52), while for clade D it was difficult to see any pattern, due to low sample size (3 samples). Samples from clades A and B are grouped in a star-like networks, characteristic from population expansion scenarios, with a common haplotype (number 1; number 24, respectively) and a few derived haplotypes (numbers 2-21; numbers 25-36, respectively) differing in one, two or three mutational steps (Fig. 11; and Tab. S1 for haplotype information). The same relations to the above described were recovered in the Splitstree network (Fig. 12).

The GHR median-joining network was congruent with the phylogenetic tree and showed not enough resolution to recover any of the mtDNA groups (Fig. 13). A frequent and widespread haplotype was found (number 1). All other 8 haplotypes detected differ from this by one, two or three mutational steps. A second most frequent haplotype, number 6, was found to be fixed in individuals sampled from the southern area in Felou (Fig.13 and Tab. S1). In this case no Splitstree network was done.

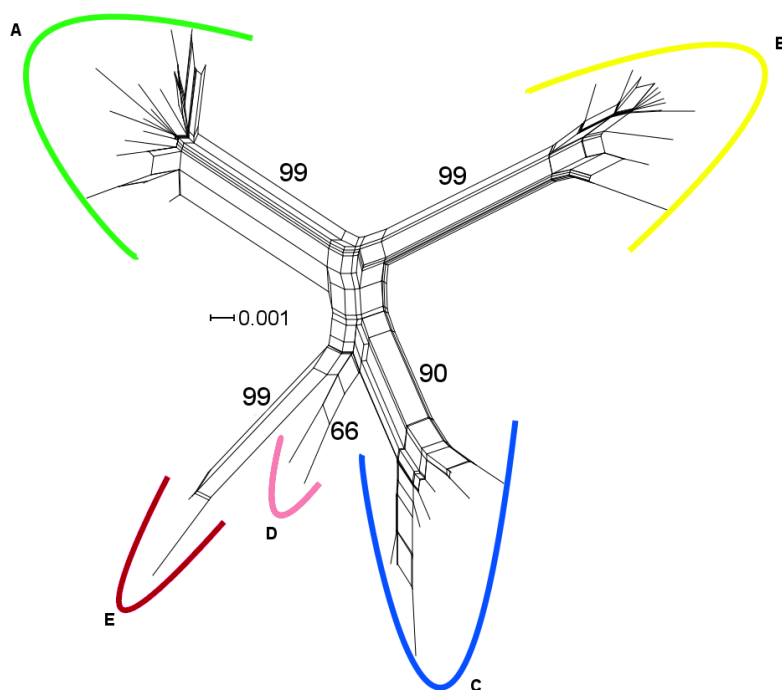
The AMOVA identified highly significant values for the genetic variance at the two partitions tested ( $P < 0.01$ ), using three hierarchical levels. Genetic structure was strong among the five groups of populations corresponding to clades A, B, C, D and E, which means that most of the genetic variability was partitioned among groups (80.89 %), and only a small proportion of the variability was partitioned among populations within groups (7.14 %) or within populations (11.97 %) (Tab. 5). There is a clear geographical pattern in the distribution of haplotypes, with the exception of two haplotypes shared between distant regions (between Atar-Tagant and Tagant-Assaba, in clade A). All subsequent analyses regarded this species as five independent evolutionary units.

The highest within-clade estimates of haplotype and nucleotide diversity was found in the Afollé population (0.96 and 0.0053, respectively) and the lowest in both Felou clades D and E (0.039 and 0.0025; clade D values not shown due to low sample

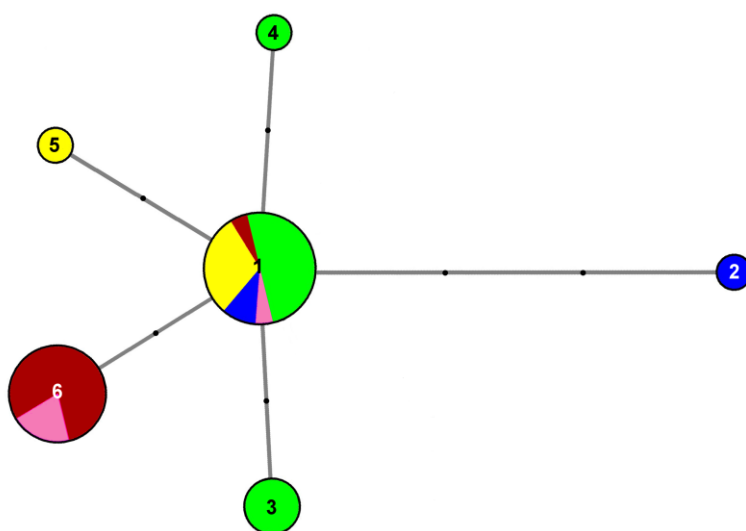
size) (Tab. 6). Average percentages of pairwise differences (Kimura 2-parameter) between clades varied from 2.6 % between clades B and E to 0.9% between clades C and D (Tab. 6). Within clades, average percentages of pairwise differences vary from 0.2% to 0.5% (Tab. 6).



**Fig. 11** – Median-joining network inferred by Network based on *Felovia vae* partial Cytb gene (700 bp length; 94 sequences). Colors represent each main clade, while circles the haplotypes scaled proportionally to their frequency in the sample. Numbers inside the circles represent haplotypes (see Tab. S1 for more information). Mutational steps among haplotypes are indicated by black dots.



**Fig. 12** – Neighbor-net network inferred by Splitstree based on *Felovia vae* partial Cytb gene (700 bp length; 94 sequences). Scale bar represents 1 % sequence divergence and numbers correspond to bootstrap values.



**Fig. 13** – Non-recombinant median-joining network of *Felovia vae* partial nDNA (GHR) (362 bp length; 24 sequences). Circles represent haplotypes scaled proportionally to their frequency in the sample. Numbers represent haplotype names. Mutational steps among haplotypes are indicated by black dots. Colors represent the main *F. vae* clades identified for mtDNA (Cytb).

**Tab. 5** - Spatial genetic structure inferred by analyses of molecular variance. Percentages of variation explained by two different grouping approaches are indicated. Statistical significance was tested using non parametric permutation approach (\*\*P < 0.01, \*P < 0.05, NS P > 0.05). See Tab. S1 for number locality codes information.

Partition (numbers indicate locality codes/mountain/geographic position)	Number of groups	F-statistics		
		Among groups	Among populations within groups	Within populations
<b>By putative geographic barriers</b>				
[1] Adrar Atar in upper North				
[2 → 9] Tagant in North				
[10 → 13.1 + 13.2] Assaba in Center	6	68.86**	18.45**	12.69**
[14 → 21] Afollé in Center East				
[22 → 24] Felou in South East				
[25 → 26] Felou in South West				
<b>By Clades</b>				
[1 → 9 + 13.2] Atar-Tagant in North + Guelta Goumbel				
[10 → 13.1] Assaba in Center				
[14 → 21] Afollé in Center East	5	80.89 **	7.14**	11.97**
[22 → 24] Felou in South East				
[25 → 26] Felou in South West				

**Tab. 6** - Genetic net average pairwise distances calculated under K2P model between *Felovia vae* clades (means and standard deviation values appear below and above the diagonal, respectively) and within genetic distance (means followed by standard deviations under brackets in the diagonal) for partial Cytb gene.

Group	A	B	C	D	E
A	0.004(0.001)	0.005	0.005	0.005	0.005
B	0.022	0.004(0.001)	0.005	0.005	0.006
C	0.019	0.023	0.005(0.002)	0.003	0.004
D	0.018	0.021	0.009	0.004(0.002)	0.004
E	0.021	0.026	0.016	0.012	0.002(0.001)

### 3.1.5. Historical demography

The non-genealogical coalescent methods and more conservative statistics (Tajima's D, Fu's Fs, Rozas's R2, and the mismatch distribution) showed significant values in *Felovia vae* except for lineage E (Felou West, South). Lineage D (Felou East, Southern) was not included in all the demographic analysis due to small sample size (only 3 sequences available) (Tab. 2).

Clades A and B, from north and central mountains (Atar-Tagant and Assaba), were supported by significant large negative values of D ( $D = -1.78$ ,  $P < 0.05$ ;  $D = -1.57$ ,  $P < 0.05$ , respectively) and Fs statistics ( $F_s = -14.94$ ,  $P < 0.01$ ;  $F_s = -7.92$ ,  $P < 0.01$ ) and also by significant and close to zero values of R2 statistics ( $R_2 = 0.056$ ,  $P < 0.01$ ;  $R_2 = 0.064$ ,  $P < 0.05$ ). Consistent with previous results, relatively unimodal patterns with non-significant low raggedness values for the two clades were supported by the mismatch distributions (A, B;  $r = 0.037$  and  $0.098$ , respectively; Fig. 14 and Tab. 2).

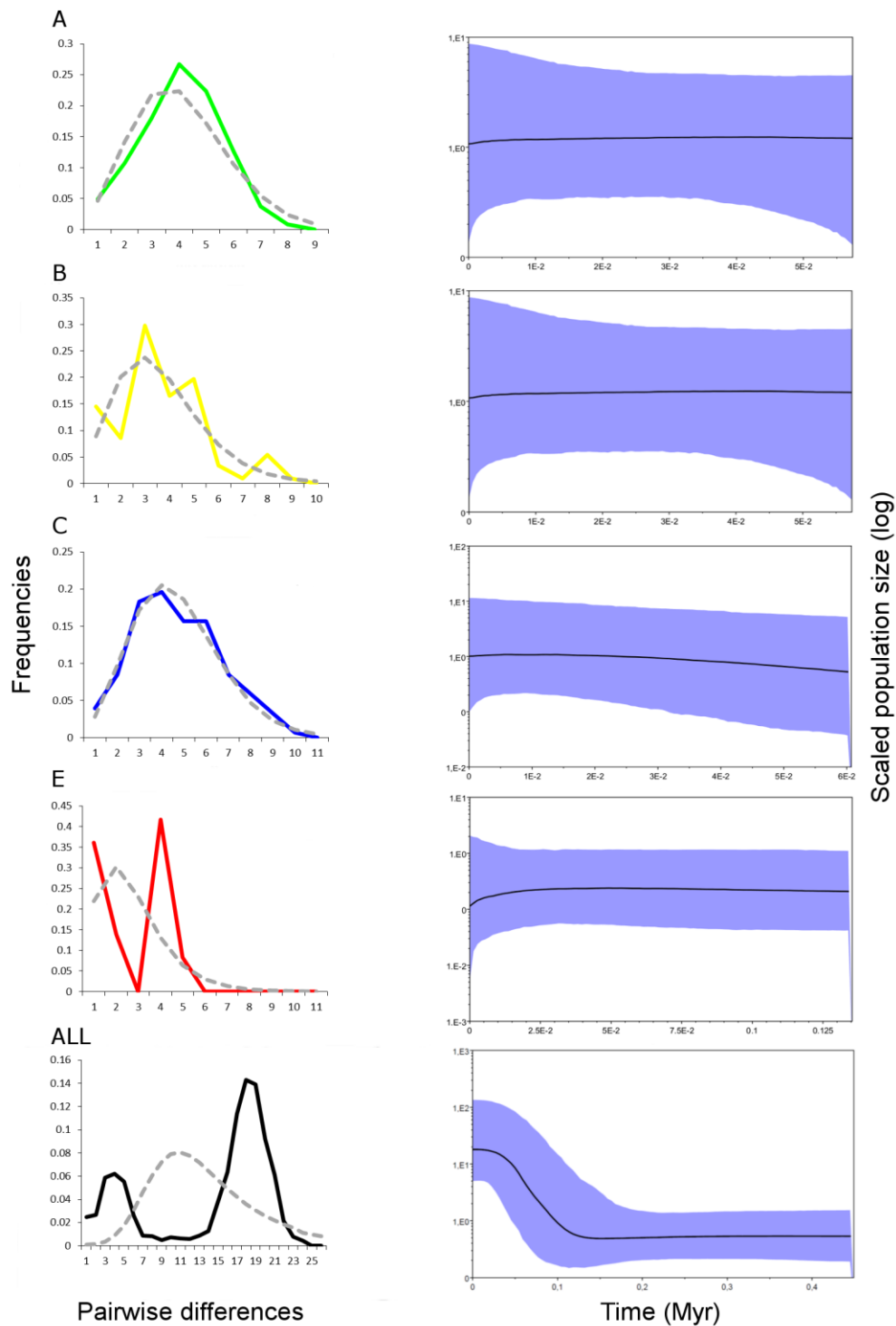
For lineage C, present exclusively in Afollé mountains (East), the neutrality test were not consensual. Only the most sensitive statistics, Fs, captured negatively significant values ( $F_s = -6.34$ ,  $P < 0.01$ ). The mismatch distributions appeared to follow

a perfect unimodal pattern although the R2 statistic and raggedness index supported the opposite pattern (Tab. 2). Low negative and not significant values of D were also found (Tab. 2).

In the same line of lineage C, when all the clades were considered, only the overall value of Fs was significant ( $F_s = -20.520$ ,  $P < 0.01$ ) (Tab. 2). For this dataset a bimodal distribution of pairwise differences was observed (Fig. 14).

Maximum Likelihood growth parameter estimates obtained using FLUCTUATE were positive and significant in all cases (*sensu* Lessa *et al.*, 2003; Pinho *et al.*, 2007), except for lineage E (Felou West, South) (Tab. 2). Large estimates (*sensu* Lessa *et al.*, 2003; Pinho *et al.*, 2007) of this parameter were found among the clades from north and central mountains (A, B: mean = 1574.22 and 937.65, respectively) and in the Eastern lineage (C; mean = 1102.96). When all clades were considered, lower, but still significant, estimates for the growth parameter were found (mean = 202.27).

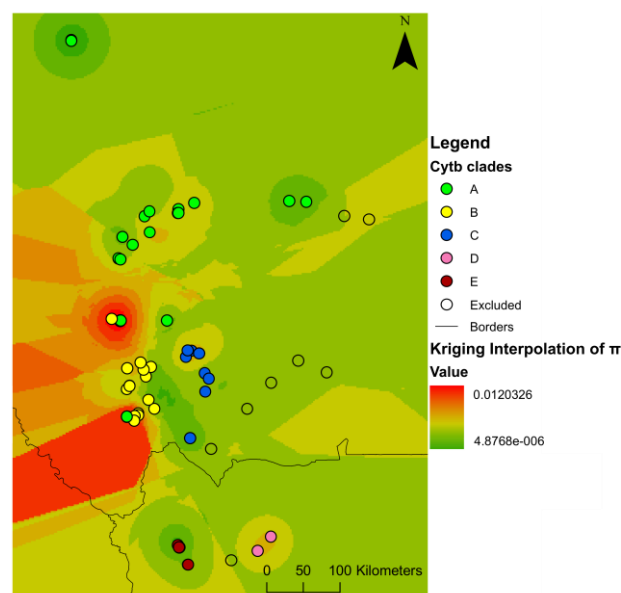
The Bayesian Skyline plot (BSP), showing effective population size through time, failed to detect any sign of expansion for separate analyses for each of the *Cytb* clades (Fig. 14). Although, it revealed one expansion event when all of mitochondrial clades were merged, starting around 120 000 years ago (Fig. 14). This expansion was preceded by a population stabilization, during the last 20 000 years (Fig. 14). The BSP for the combined clades ran under an HKY+G+I model.



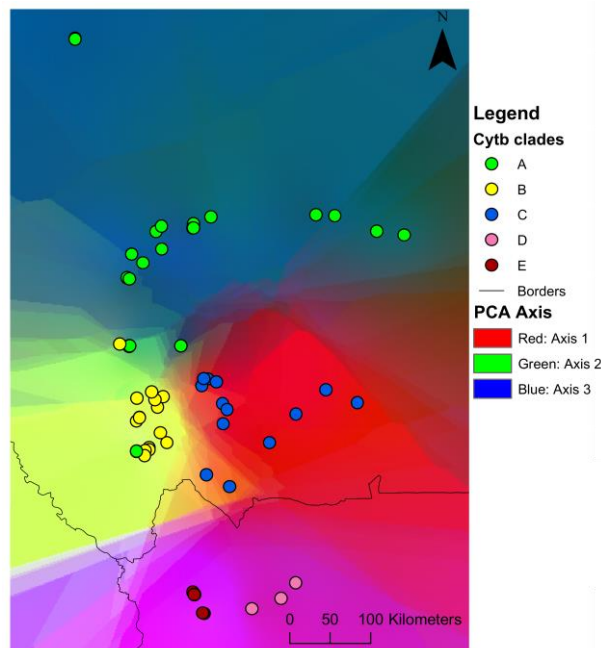
**Fig. 14** – Pairwise mismatch distribution of nucleotide differences of Cytb in each clade of *Felovia vae* (Left). Dashed lines indicate the expected distribution under a population expansion model. Solid lines represent the observed distribution. Bayesian skyline plots of Cytb *F. vae* clades (Right). The median estimate (black line) and 95% highest posterior density limits (blue background) are indicated. The y-axis represents the scaled population size and the x-axis indicates measures time in substitutions per site per million years. A, B, C and E correspond to the clades in Fig. 6 and 7, while the combined clades are represented by “ALL”.

### 3.1.6. Spatial analyses

The kriging interpolation produced a continuous surface of nucleotide diversity that shows the highest genetic diversity in Assaba mountain, particularly in both sides of the Gorgol el Abiod river valley where haplotypes from the two different clades were found in possible parapatry and in one sympatric location in south Assaba (Fig. 15). The lowest levels of genetic diversity were found in the margins of the distribution as for example the Adrar Atar and Felou (west of Senegal river) mountains (Fig. 15). The interpolated genetic distances surface detected an abrupt change between Assaba – Afollé and Afollé – Felou mountains (Fig. 16). A gradual change was found between Tagant-Assaba mountains and, less pronounced, between Felou West – Felou East. A homogeneous surface reflecting the lack of genetic divergence was found between Adrar Atar – Tagant mountains (Fig. 16).



**Fig. 15** – Geographic genetic variation in *Felovia vae*. Surface of interpolated genetic diversity based on nucleotide diversity, resulting raster file was reclassified into 9 different classes using natural breaks. The circles correspond to samples used for the interpolation, the circle colors to the Cytb clades and the circles with no color were excluded because they had no “neighbor” to merge in a population. Darker green colors in background correspond to areas with the lowest nucleotide diversity while darker red areas are assigned to areas with high nucleotide diversity.



**Fig. 16** – Geographic genetic differentiation in *Felovia vae*. First axis of the spatial principal component analysis applied to the interpolations of the uncorrected p-distances matrix. The circles correspond to the samples used for the interpolation and the point colors to the Cytb clades.

## 3.2. Ecological niche-based modeling approach

### 3.2.1. Models performance

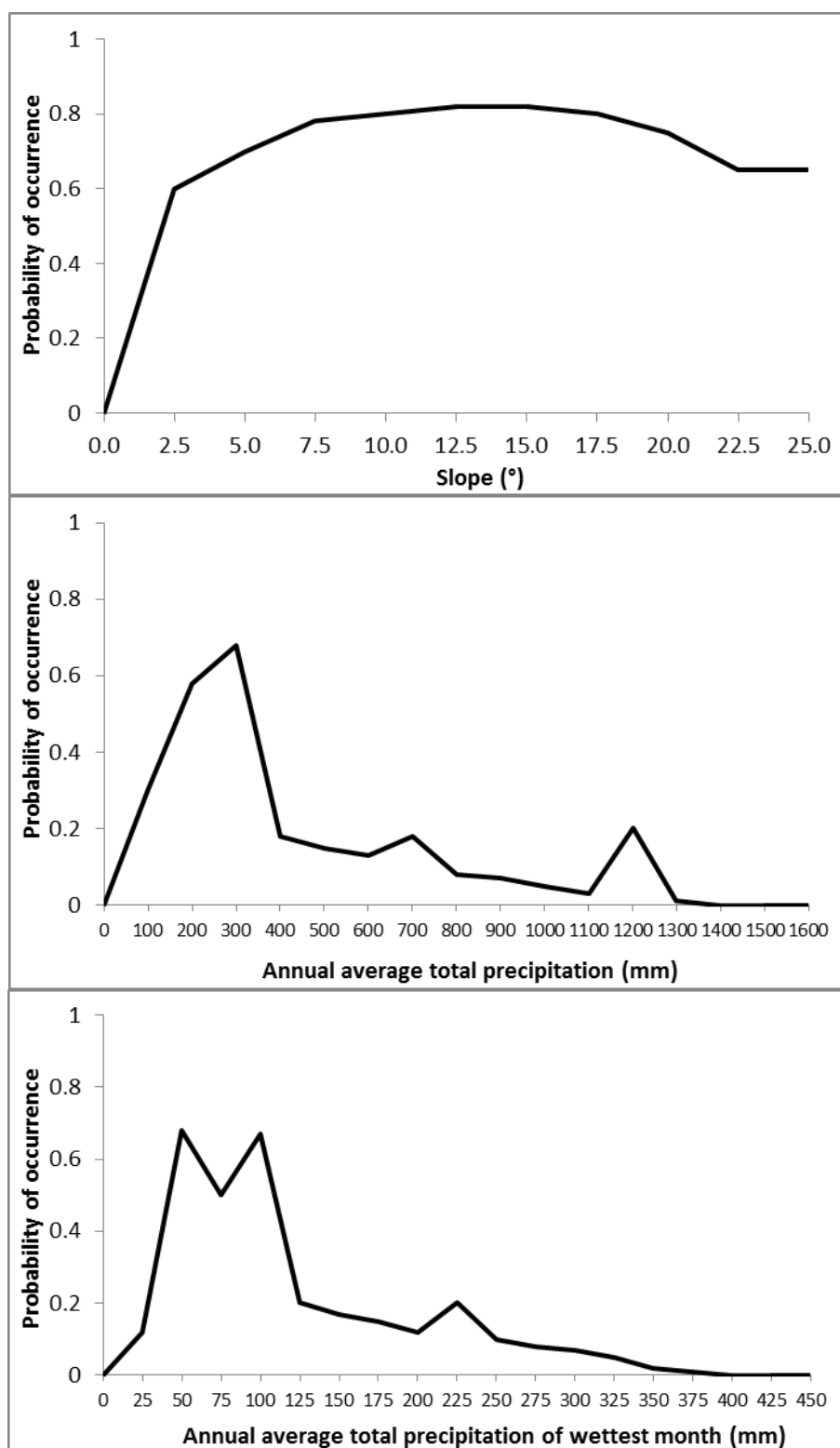
The ROC plots showed high average AUCs in both training and test data sets with low standard deviations ( $0.964 \pm 0.007$ ;  $0.940 \pm 0.025$ , respectively). Meaning that the model was highly accurate (96% of the observations were correctly predicted) and correctly predicted the validation data set (94%; indicates no model over-fitting). Therefore, the model was good to discriminate between species records and background. Average highest standard deviation of individual models occurred mostly in mountain.

### 3.2.2. Environmental factors related to species occurrence

The EGVs identified as mostly contributing to individual models were slope (average  $47.3 \pm 4.8$  % SD), annual average total precipitation ( $23.5 \pm 7.3$  %) and annual average total precipitation of the wettest month ( $13.2 \pm 7.8$  %). The other EGVs also had average contribution higher than 2%, except for annual average temperature ( $1.3 \pm 1.4$  SD) (Tab. 7). The average permutation importance was in agreement (Tab. 7). The average profiles of response curves for the EGVs suggested that the species occurs more frequently in areas with high slope and low annual average total precipitation (around 200 – 300 mm) (Fig. 17). In the other extreme, although with less probability, the species can also occur in areas with higher annual average total precipitation (1200 mm) (Fig. 17).

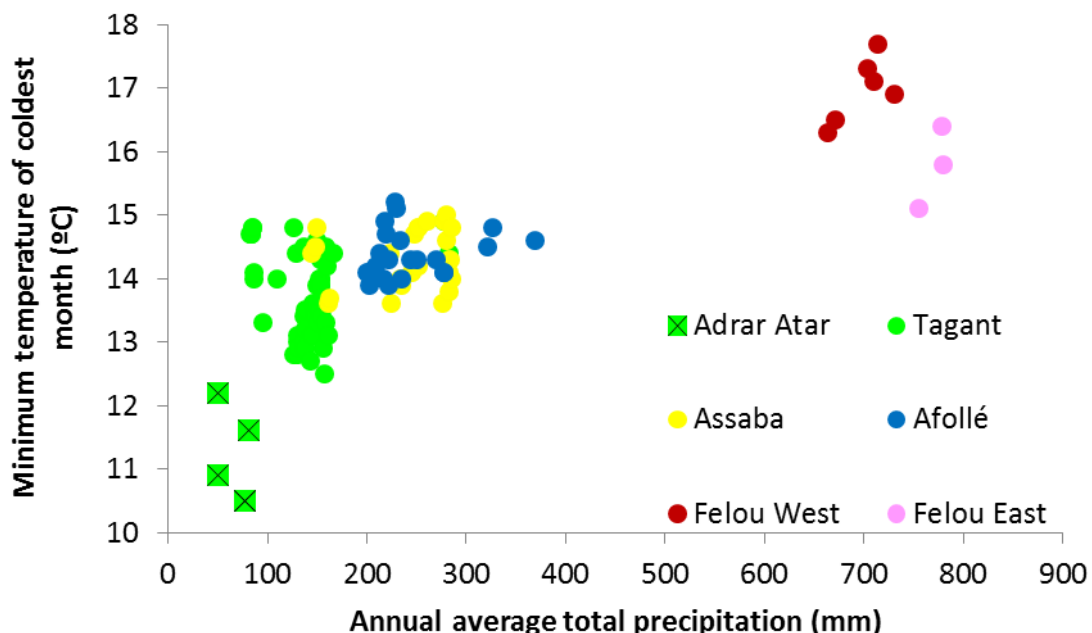
**Tab. 7** – Average percent contribution (and standard deviation) of each environmental variable for the 20 replicate models.

<b>Variables</b>	<b>% Contribution</b>	<b>Permutation importance</b>
Slope	47.3 (4.8)	30.4 (11.3)
Annual average total precipitation	23.5 (7.3)	33.3 (10.2)
Annual average total precipitation of wettest month	13.2 (7.8)	14.0 (14.1)
Annual average temperature	1.3 (1.4)	1.3 (1.1)
Temperature annual range	3.8 (4.6)	6.1 (3.4)
Maximum temperature of warmest month	6.0 (5.1)	5.8 (5.1)
Minimum temperature of coldest month	4.9 (2.2)	9.1 (5.6)



**Fig. 17** – Response curves for the three most related environmental factors to the distribution of *Felovia vae*. Curves traduce the average probability of occurrence from 20 model replicates along the environmental gradients.

The two most important EGVs, slope and annual average total precipitation, were initially selected to generate a plot depicting the location of genetic clades along environmental axis, but no differentiation was detected along the slope axis (data not shown). Thus, Slope was replaced by the minimum temperature of the coldest month in order to check for differences between genetic clades along a temperature gradient. The annual average total precipitation variable discriminated four main groups (Fig. 18): Atar-Tagant (< 100 mm); Atar-Tagant and Assaba (100 < mm < 200); mainly Assaba and Afollé (200 < mm < 400); Felou West and East (650 < mm < 800). The Felou group showed the highest degree of differentiation, while the others were separated by small differences. The minimum temperature of the coldest month variable discriminated three main groups (Fig. 18): Atar-Tagant (10.5 – 12.2 °C); Atar-Tagant, Assaba and Afollé (12.5 – 15.2 °C); Felou (15.1 – 17.7 °C). In general the model showed that the species climatic requirements vary between two extremes, one present in the northernmost areas (Atar mountain) and the other in the southernmost areas (Felou region).



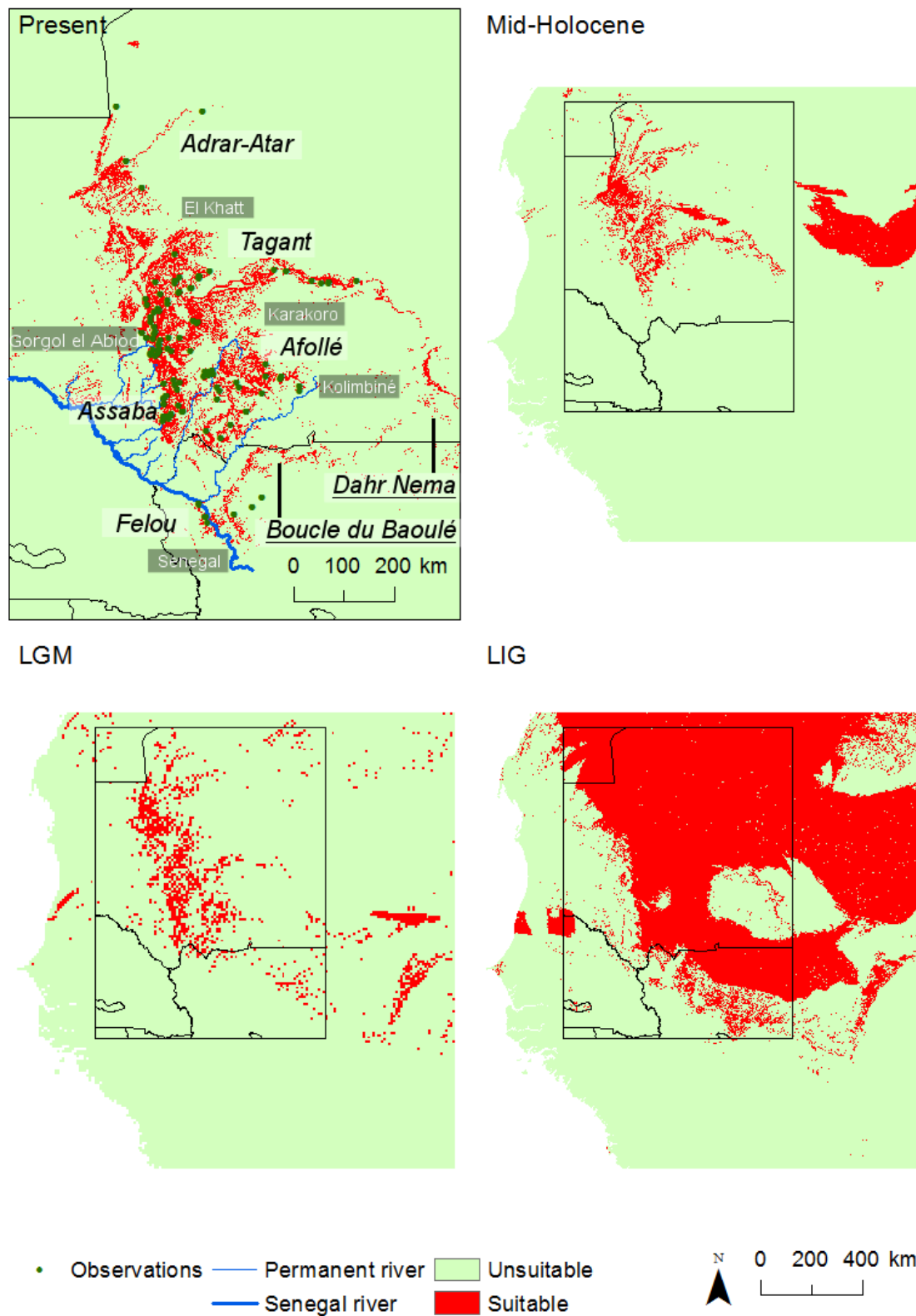
**Fig. 18** – Plot of the two environmental variables that most described the differences between genetic groups. For each observation was attributed a genetic group, supported with genetic information. Each color corresponded to a Cytb clade (Fig. 6).

### 3.2.3. Present conditions

The consensus model for current conditions identified potential areas of the occurrence of *Felovia vae* consistent with the observed current range. These areas were spread across all Mauritanian mountain regions and also in southwestern Mali (Fig. 19). The model was, however, less efficient in identifying the species niche in range margins (3 records fell within unsuitable grids in Felou area, Mali).

The results showed that part of El Khatt river valley extent was identified with suitable environmental conditions for the species to occur (Fig. 19). The Djouk valley was identified as perfectly suitable for the species, giving a continuum of good conditions between Tagant and Assaba mountains (Fig. 19). The Karakoro and Kolombiné river valleys were identified as unsuitable for species occurrence (Fig. 19).

Environmental suitability was high in some areas where the species has not been recorded, like the Dahr Nema escarpment in south-eastern Mauritania and in the northern Boucle du Baoulé area in Mali.



**Fig. 19** – Predicted suitability area of *Felovia vae* in the consensus model for current conditions (different scale) and projections to the Mid-Holocene, Last Glacial Maximum (LGM) and Last Interglacial (LIG). The country borders were left for past projections with an orientation propose and the rivers were removed, since nothing was known about their dynamics in the past.

### 3.2.4. Past conditions

The estimated topoclimatic suitability for the occurrence of *Felovia vae* for past conditions varied considerably among the three different scenarios (Mid-Holocene, LGM and LIG). The consensus model for the Mid-Holocene retrieved the main area of suitable cells confined to Adrar Atar, Tagant and Assaba mountains (Fig. 19). LGM added suitable areas in the Afollé mountain isolated from the other mountains, while Mali remained unsuitable. LIG reported the maximum connectivity, in which suitable areas for species occurrence could be found across all the study area and outside of it, in direction to Morocco and central Mali (Fig. 19). A great variability was also found for the GCM in general (Fig. S2).

### 3.2.5. Refugia identification

Finally, the climatically stable areas for *Felovia vae* occurrence were mainly restricted to Atar, Tagant and Assaba mountains (Fig. 20). However, the Afollé mountain constitutes also a good candidate for refugia, although represent by one high suitable cell is filled by intermediate to high suitable areas. Some suitable grids were also found between Atar and Tagant mountains and in central Mali.

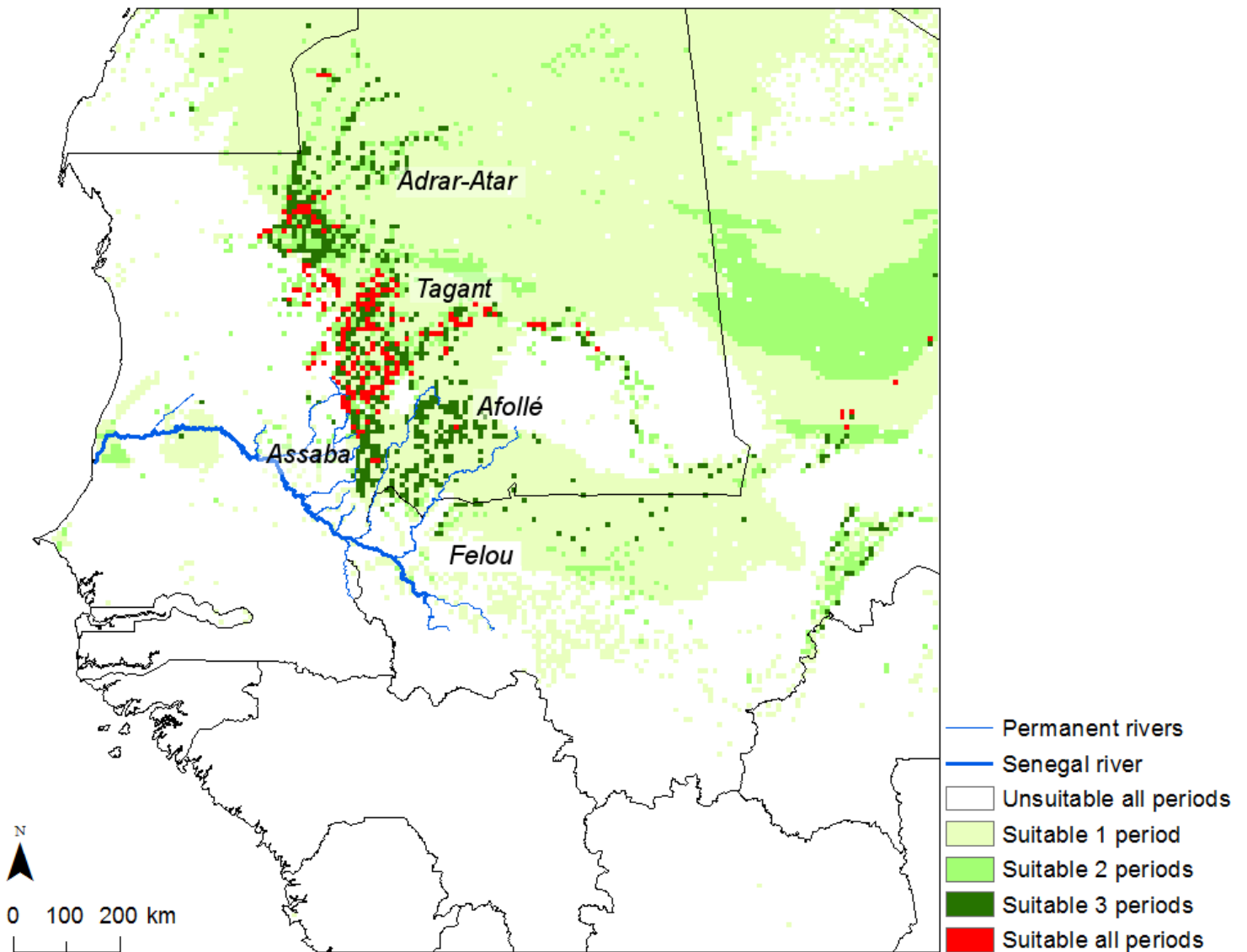


Fig. 20 – Potential refugia given by stable .suitable conditions across all periods (present, Mid-Holocene, Last Glacial Maximum, Last Interglacial).

## 4. Discussion

The benefits of an integrative approach using phylogeography and ecological niche models are known to increase the strength of the inference process about the history of species. In this thesis, using mostly non-invasive samples, we assessed the evolutionary history of *Felovia vae*, an endemic rock-dwelling rodent known from the Mauritania mountains and Western Mali categorized as data deficient by IUCN. Overall, our results indicate the possible presence of five divergent clades confined mostly to different mountain massifs. The TMRCA was estimated to live in the early-middle Pleistocene. External barriers across the species distribution range were identified and compared with co-distributed organisms. Demographic estimates suggested that the species suffered a population expansion around the Last Interglacial (LIG) and remained stable since the Last Glacial Maximum (LGM). Additionally, we found different molecular phylogenetic relationships between Ctenodactylidae members compared with paleontological and morphological studies. The most recent common ancestor of the family was estimated to occur during the early-middle Miocene.

### 4.1. Ctenodactylidae phylogeny

Maximum likelihood and Bayesian analyses for mtDNA (Cytb), supports the monophyly of *Pectinator spekei*, *Felovia vae*, *Massoutiera mzabi*, *Ctenodactylus gundi* and *Ctenodactylus vali* (for species distribution see Fig. 3). The *P. spekei* occupied the basal position in the phylogeny (Fig. 8). This in line with its great geographical isolation from all the members of the family (longer period of independent evolution) and its resemblance with Miocene ancestors of the Ctenodactylidae family (conservative) based on internal and external morphological characters (mostly teeth) and chromosomes (Jaeger, 1971; George, 1979; George, 1985; López-Antoñanzas & Knoll, 2011; Gomes *et al.*, 2012). George (1985) suggested a closer relationship between *Massoutiera* and *Felovia*, as López-Antoñanzas and Knoll (2011) supported. However, George (1985) also observed that *Ctenodactylus* species (*C. vali* and *C. gundi*) separated comparatively early from *Massoutiera-Felovia* group, but suggests that

*Felovia* was the most conservative and, thus, resembled *Pectinator* the most. Our results give relevance to this last observation, because *Massoutiera* appeared more related with *Ctenodactylus* than with *Felovia*, sharing lower genetic distances (Fig. 8; Tab. 3). Finally, *C. gundi* and *C. vali* are the last species to have diverged and share the lowest genetic distances (Fig. 8; Tab. 3). As in most mammals, the Genetic Species Concept (Baker & Bradley, 2006) attribute genetic differentiation often higher than 5 % between sister taxa, inferred from Cytb gene sequences. Therefore, since the genetic distances are always higher than 5 %, differing by 11-23% (within *Ctenodactylus* and between the most basal species, respectively) (Tab. 3), we can assume to be dealing with genetically different species. Calibration estimates derived from mitochondrial markers suggest that the origin of the family took place during the early and middle Miocene (10.49-16.73) (Fig. 10; Tab. S2). These divergence estimates are congruent with temporally active land bridges especially between North Africa, Arabia and Pakistan due to short term sea level oscillations of the Tethys (17-10.5 Mya) (Monadjem *et al.*, 2015). The opening of these terrestrial connections allowed fauna exchange between Africa and Eurasia, as for instance is the case of *Muridae* rodents (Monadjem *et al.*, 2015). The Ctenodactylidae might have followed the same process, as supported by numerous fossils of extinct species belonging to the Ctenodactylinae subfamily found across North-Western Africa to India and China (López-Antoñanzas, 2010; López-Antoñanzas & Knoll, 2011; López-Antoñanzas *et al.*, 2015). The other divergence estimates fall within the estimated period of hyper-arid cycles, and therefore hypothesis for the separation of species generated by the vicariant effect of the Sahara desert cannot be ruled out (for instance the two *Ctenodactylus* species are the only occurring north of Sahara). The vicariant effect generated by the appearance and oscillations of Sahara (and Sahel) boundaries are described for many other species (Douady *et al.*, 2003; Fonseca *et al.*, 2009; Guillaumet *et al.*, 2008; Gonçalves *et al.*, 2012). Other geological events as the formation of the Rift Valley in the early-mid Miocene (Partridge *et al.*, 1995), development of the mega-Lake Chad (much larger during the Quaternary; Leblanc *et al.*, 2006) and the spread of cool-season grasses might have had also their influence in Ctenodactylidae diversification as suggested for other rodent faunas (Mouline *et al.*, 2008; Nicolas *et al.*, 2008; Brouat *et al.*, 2009; Colangelo *et al.*, 2013; Monadjem *et al.*, 2015).

## 4.2. Biogeography of *Felovia vae*

The ML, Bayesian, Network, Splitstree analyses for mtDNA Cytb haplotypes (Fig. 6, 11 and 12) showed a strong genetic structure within Mauritania and western Mali *Felovia vae* populations, with five main clades separated by high to moderate (intraspecific) genetic distances (K2P distances= 0.9 % to 2.6%; Tab. 6). The general genetic structure suggested a coherent geographic distribution within *Felovia vae* range. In contrast the ML, Bayesian and network analyses for nDNA GHR haplotypes (exon 10) did not recover any genetic structure. The Bayesian phylogenetic trees for mtDNA were always more supported than the ML ones, and although revealing a strong phylogeographical structure the nodes were never highly supported (values of posterior probability higher than 0.7 but lower than 0.9) (Fig. 6). The node with the lowest support (Fig. 6; node supporting all the clades except clade A), made it impossible to infer the basal clade for the species and thus the relationships between clades were compromised. However, networks with fourteen to eight mutational steps differences and Splitstree with high bootstrap values (except for clade D) supported well the division of the haplotypes into five main clusters (Fig. 11 and 12). The presence of mountain specific haplogroups separated by relatively large to moderate mutational distances fits in the middle of category I and III for phylogeographical patterns observed in small- or medium-sized mammals (Avice, 2000). The category I explains deep lineages (strong genetic structure) while category III represent shallow lineages (weaker genetic structure), both for allopatric distributed species. In *Felovia vae* case, not sufficient time might have passed to accumulate enough genetic differentiation (accumulation of new mutations in isolation by genetic drift, and/or lineage sorting from a polymorphic ancestor), for the mostly allopatric populations, to occupy recognizable deeply separated branches in a gene tree. However, results suggested that *F. vae* have already started this process, being more close to achieve the category I. A likely explanation for these phylogeographic discontinuities is the presence of long-term extrinsic barriers to dispersal and gene flow between mountain massifs.

To identify potential extrinsic barriers to gene flow it is necessary to have a good knowledge about the habitat preferences of the species. *Felovia vae* was considered to be strictly confined to rocky habitats, in which variables as slope and *gueltas* have

played a major role in defining their patchy distribution (Vale *et al.*, 2012a). In this study, using topoclimatic variables (without landcover variables) we recovered the similar general patterns on the *Felovia vae* current distribution, and thus detected potential barriers of unsuitable habitat, and investigated their possible evolution over time (Fig. 19). Species affinity with rocky patches, which usually occur in areas with higher slope, was the strongest effect explaining species distribution (Tab. 7; Fig. 17). Nonetheless, climatic variables as annual average total precipitation had also high relevance and contributed to explain the niche differentiation of the southern clades from all the rest (clade D and E) (Tab. 7; Fig. 17 and 18). The species was found to occur in two environmental extremes with different precipitation regimes, one present in the northernmost areas (Adrar Atar mountain, clade A; lowest values) and the other in the southernmost areas (Felou mountains, clade D and E; highest values) (Fig. 18). River valleys, with lack of significant rock outcrops and covered by sand, do appear to be unfavorable to *Felovia*, and this could explain the genetic distance observed between most clades. AMOVA analysis indicated high variation among groups limited by river valleys (68.9%), although, and as expected, higher values were obtained with partition by clades (Tab. 5).

The influence of the Karakoro and Kolombiné river valleys for restricting historical gene flow in *F. vae* is inferred by the haplotype distributions. In the first case, haplotypes from Assaba and Afollé are found in different clades (clade B: yellow; clade C: blue, respectively) (Fig. 6 and 7). In the second case, the southern regions in Felou mountains are distributed in two different clades, different from the previous ones (clade D: pink; clade E: red, respectively) (Fig. 6 and 7). The genetic distances geographically interpolated also marked these two potential barriers in space (Fig. 16). The paleoclimate models, more or less accurately, corroborate these two barriers formed by unsuitable conditions both for Mid-Holocene and LGM (Fig. 19). As for the geographical limit between clade D and E, the Senegal river might have had a greater role (Fig. 6, 7 and 16 – spatial demarcation not so clear). Since this southern region of Felou was not always suitable for the species (just in the present and LIG) it suggests that *Felovia* populations went through local extinctions from time to time and eventual recolonizations might come from the closest populations, located in Afollé (genetically closer; Tab. 6). The ecological discontinuity between Adrar-Atar and Tagant mountains formed by, nowadays permanently dry, El Khatt river basin seems not to have been a barrier for gundis dispersal/colonizations in the past (mid-Holocene, LGM and LIG; Fig.

19) as it is suggested for the contemporary populations (Fig. 17; Vale *et al.*, 2012a), since haplotypes from the same clade (A: green) were found in both mountains (Fig. 6, 7 and 16). In contrast, the 3 km wide Gorgol el Abiod river valley, which separates populations of Tagant and Assaba (clades A and B, respectively) (Fig. 6 and 7), although it seems to constitute a barrier for dispersal/gene flow it did not appear as unsuitable for the species (Fig. 18) and was marked by a moderate transition in the spatial analyses (Fig. 16). The detected sympatric location in the southern Assaba mountains (Guelta Goumbel) suggests that the barrier along the Djouk valley was porous, supporting the ecological models connection through suitable habitat (Fig. 7). The presence of both haplotypes in Guelta Goumbel may be a remnant of a wider distribution of *Felovia vae* in the past. Other explanations might be related with possible human translocations and/or more unlikely contaminations or exchange of identification labels. Testing that possibilities will require more population dedicated sampling.

The existence of different, more suitable for gundi, environmental conditions in the relatively recent past might explain for instance the sharing of the same haplotype in mountain of Adrar and Tagant. Although currently mountains are surrounded by sandy and gravel areas, they were connected by more continuous savannah-like habitats during past humid periods (Gasse, 2000; Brito *et al.*, 2014). However, seasonal rivers during prolonged wet climatic phases, as Gorgol el Abiod, Karakoro and Kolombiné, could have been more effective barriers to restrain the dispersal/ gene flow between *Felovia* populations as described for other sub-Saharan rodents (e.g. Brouat *et al.*, 2009; Bryja *et al.*, 2010). While the El Khatt river valley that is known to be permanently dry, could be less effective for the species dispersal and thus explain the observed pattern. The arid periods and the progressive aridification in the region after mid-Holocene (Kröpelin *et al.*, 2008), and the occurrence of severe droughts in the 1970s (Ahmed *et al.*, 2008; Mahé & Paturel, 2009), likely induced more restricted to mountains occurrence of *F. vae*. Moreover, the dependence of *Felovia* occurrence on the presence of water (Vale *et al.*, 2012a) and its relatively small body size, that might hamper long distance dispersal, can reinforce isolation between contemporary populations.

### 4.3. Effects of the past climate change on the demographic pattern of *Felovia vae*

Coalescence within all of the major clades of *Felovia* occurred during Middle Pleistocene and/or the beginning of Late Pleistocene (Fig. 10). These results corroborates those obtained for other sub-Saharan rodents (Mouline *et al.*, 2008; Nicolas *et al.*, 2008; Brouat *et al.*, 2009; Bryja *et al.*, 2010; Colangelo *et al.*, 2013) and reptiles (Gonçalves *et al.*, 2012). During the Pleistocene, the Sahara has experienced drastic climatic changes, rendering its boundaries highly variable. It could at least partly explain the geographical structure of *Felovia* mitochondrial clades. Populations of *Felovia* could have been restricted to mountains during phases of unfavorable periods of the Pleistocene, and have experienced subsequent expansion during phases of favorable habitat expansion. In agreement with this hypothesis, signs of population expansion were recorded.

The Bayesian skyline plots for pooled clades show signs of demographic expansion dating back to 120 000 (Fig. 14), similar with Colangelo and colleagues (2013) estimations. This date approximately coincides with the Last Interglacial (~ 120 000 – 140 000 years ago) in which the climatic conditions were close to the “green Sahara”. During this time the paleoclimatic model show the maximum connectivity between all the mountains, and thus possibly favoring the species expansion after a long period of considerably reduced populations (Fig. 19). Aridity in the region reached its peak during the last glacial maximum (23 000 – 18 000 years ago; Gasse, 2000) and started to decrease at about 14 800 years ago at the beginning of the “African Humid Period” which approximately extended from 14 8000 to 55000 years ago (Gasse, 2000). The early mid-Holocene “green” Sahara and Sahel are evidenced by numerous palaeolakes, paleodrainages and vegetation (Gasse, 2000). Both of these periods might explain, through barriers of unsuitable conditions for species dispersal, given by sandy river valleys in arid phases or dense flowing river in humid phases, the general demographic stabilization that started at approximately 25 000 years ago. Looking for both LGM and Mid-Holocene consensus paleoclimatic models we can see that possible barriers of unsuitable topoclimatic areas begin to rise. For instance, the ecological discontinuity between Assaba and Afollé in LGM and the lack of suitable cells for Afollé and southern regions in Mali possibly by to humid conditions for the

species to occur during Mid-Holocene. Moreover, this expansion pattern is similar for clade A and B when analyzed separately, as shown by neutrality tests, unimodal mismatch distributions and star-like shape in the network (Tab. 2; Fig. 7; Fig. 14). The demographic pattern in clade C, was not so clear with only the F statistic (most adapted for low sample sizes) and the unimodal mismatch distribution, supporting demographic expansion. The analyses conducted within lineage E detects a relative demographic stability, probably derived from limited sample size (Tab. 2; Fig. 14).

This expansion-contraction model in *Felovia vae* populations could be explained by the Quaternary cycling events. Considering that *gueltas* may have dried out (case of Adrar; Vale *et al.*, 2015) during arid phases or conditions for species occurrence might have become unfavourable because of competitive exclusion, predation and/or high humidity in humid phases (e.g. rotting nests inside rochy caves that might have excluded southern populations of Mali; Fig. 7), the isolated populations might have experienced repeated population bottlenecks or have become extinct from time to time. The high levels of haplotype diversity and low levels of nucleotide diversity may suggest population bottleneck followed by rapid population growth and accumulation of mutations. As a genealogical consequence of this long term population isolation, an erratic and ragged distribution of the pairwise differences among DNA sequences is expected (Rogers & Harpending, 1992). Noteworthy, such a bimodal pattern was recovered by the mismatch analysis of clade E (Fig. 14) and by the overall values of the neutrality tests (Tab. 2).

The overall areas pointed as potential refugia are nearly overlapping the current distribution of the species, showing a great stability of the species range throughout time (Fig. 20). The most suitable conditions for all periods were mainly present in the mountains of Tagant and Assaba. Our spatial analyses, in part supported these areas as refugia by identifying the highest values of genetic diversity between Tagant and Assaba mountains and in one location in the south of Assaba mountain (Fig. 15; Tab. 2). However, these values might be biased because the clade A and clade B occur in potential parapatry and sympatry in those locations (Fig. 7).

Comparison of our results with other taxa that cover a geographical range similar to our target species is fundamental in phylogeography. *Agama boulengeri* is a rock-dwelling reptile with the same environmental requirements and geographic distribution as *Felovia*. Therefore a similar distribution of predicted suitable areas (Vale

*et al.*, 2012a) and genetic structure (Gonçalves, 2010; Gonçalves *et al.*, 2012) was found. Three clades were found throughout the Mauritania mountains, which allow us to compare the strength of our identified potential extrinsically barriers. The El Khatt river valley seems to not constitute an ecological barrier for both species, since haplotypes from the same clade are found in both sides. The Gorgol el Abiod river valley is predicted as suitable by the ecological models (Vale *et al.*, 2012b), but limit the ranges of the northern and southern clades (Gonçalves, 2010; Gonçalves *et al.*, 2012). The Karakoro river valley also seems to have been an effective barrier for both species dispersal/gene flow (; Gonçalves, 2010; Vale *et al.*, 2012b). For *A. boulengeri* the population in Assaba was described as the most basal group in the phylogenetic tree and thus it was hypothesized that Mauritania colonization started from this area. However, for *F. vae* it was impossible to make such inferences, since relationships between clades were shallow in the tree. Vicariance through increased aridity was the main driver to explain the genetic differentiation. The TMRCA (excluding the eastern clades from Afollé) was estimated to occur at 5.4 Myr and together with *Felovia* TMRCA (0.7 Myr), both approximately coincide with hyper-arid cycles (Gonçalves *et al.*, 2012). For aquatic organisms in the region, as *Crocodylus suchus* and *Tilapia* spp., fast and slower evolving markers (microsatellites and mitochondrial, among others) were applied for studying the genetic structure. For *C. suchus* (Velo-Antón *et al.*, 2014) it was found high genetic structure for the microsatellites, coherent with *Felovia* and *Agama* species and possibly prompted by geographic isolation, low population size and genetic drift. For *Tilapia* spp. it was also found a population of Sarotherodon B fishes, from Afollé mountain (in Karakoro sub-basin), genetically isolated from Western Mauritania (Dilyte, 2014). Once again arid phases were suggested to induce geographic isolation in small mountain areas (or even local extinction), with the last arid phase having a major role for the observed genetic structure given by microsatellites. In other hand, humid periods before Mid-Holocene might have facilitated population expansions from southern ranges as described for *C. suchus* and also observed in other species (Brito *et al.*, 2014). After Mid-Holocene rainy seasons may contribute for occasional connection between isolated populations, through seasonal rivers and brooks that interconnect *gueltas*. *Felovia vae* during these humid periods/rainy seasons might, in contrast, become disconnected by rivers as suggested for other sub-Saharan rodent species with low swimming capabilities (e.g. Brouat *et al.*, 2009; Bryja *et al.*, 2010).

## 4.4. Future research

### 4.4.1. Molecular markers

The ML, Bayesian and network analyses for nDNA GHR haplotypes (exon 10) did not recover any genetic structure as in mtDNA Cytb. This lack of genetic differentiation observed in nDNA may be explained by incomplete lineage sorting and/or gene flow. Also, these incongruences between mtDNA and nDNA gene genealogies is expected due to the higher evolutionary rate of mtDNA and its lower effective population, which make this molecular marker more likely to recover patterns of recent historical events (lower genetic distances for the nDNA, between Ctenodactylidae species; Tab. 4). Other possibility might be related to a male-biased gene flow, common in mammals, in which the males tend to disperse more than females that remain more or less in the same locations. It has been suggested from recapture data, that *Ctenodactylus gundi* result from both male and female natal philopatry in conjunction with male-biased dispersal (Nutt, 2005), therefore this hypothesis cannot be ruled out for *Felovia*. This tendency might have profound influence in nucleotide diversities. However, none of these possibilities might be taken alone to explain the contrasting pattern of genetic differentiation between the two markers.

An increase of the number of nuclear gene sequences and their length, may clarify the species genetic structure. Once that the individual history of any particular locus is highly stochastic, which means that each gene can have limited confidence in reconstructing a species history (Zhang & Hewitt, 2003). Reliable inferences on population structure and evolutionary history should therefore be drawn from genealogical data of several independent loci (Hare, 2001). It is also important to have the notion that the genetic outputs from mitochondrial DNA and nuclear DNA might be discordant. Besides the stochastic process (higher in mtDNA) of lineage divergence, the occurrence of gene flow, introgression and incomplete lineage sorting are among the most well documented cases (Machado & Hey 2003; Pinho *et al.* 2008; Bryja *et al.*, 2010; Boratyński *et al.*, 2011; Toews & Brelsford, 2012; Boratyński *et al.*, 2014). A good example is the case of other rock habitat specialist rodent, *Laonastes aenigmamus*, for which two nuclear genes (GHR and BFIBR) did not recovered the genetic structure given by one mitochondrial gene (Cytb) showing low nucleotide variation (Nicolas *et al.*,

2012). Having this in mind we have tried to amplify other nuclear genes, including: the intron 7 of the beta-fibrinogen (primers *BFIBR1-BFIBR2* from Seddon *et al.*, 2001); interphotoreceptor retinoid-binding protein (primers *IRBP: F1S-R1S; F2S-R2S; F5-R5; F4-R4; R.SM1-R.SM2* from Barbosa *et al.*, 2013); vWF (4, 5, 6, 7); chromosome x (Hellborg & Ellegren, 2004). More detailed investigation will also be required (with e.g. fast evolving molecular markers) to test if the populations continued in geographic isolation for more current conditions, if there is reproductive isolation between sympatric/parapatric populations and access dispersal capabilities in such harsh environments. Nutt (2003) already developed eleven highly polymorphic microsatellites loci for *Ctenodactylus gundi*, which might be useful for *Felovia vae*. Also telemetry would be a reliable way to test for dispersal capacity of the species. Noteworthy are the four sequences for complete Cytb gene that we obtained from fresh tissues samples and one scat, which should be considered as good quality references for further genetic analyses.

#### 4.4.2. Sample size

Sample sizes of our study greatly affect the power to reconstruct historical demographies using Bayesian Skyline Plot (BSP) analysis, since small sample sizes do not appear to diminish the usefulness of Mismatch distributions (Grant, 2015). A small sample size, most critical for the marginal populations in Mali (Felou mountains) in this study, is more prone to detect deep partitions in a population (Felsenstein, 2006) but it cannot detect the effects of recent population shifts given by the outer branches of a haplotype tree (Grant, 2015). For populations with sample size less than 50 Grant (2015) often obtained flat BSP that instead of suggesting population stability it represented a lack of power to detect population expansion. Therefore, to infer more robust historical demographic patterns there is a urgent need to increase the number of samples.

Suitable areas unknown for the present were identified by our ecological niche models, mostly in Dahr Nema escarpment in south-eastern Mauritania and in the northern Boucle du Baoulé area in Mali (Fig. 4 for localization). Modeling validating work is needed in these areas. However, several problems were identified in the

construction of those models that need to be assessed in further studies. Uncertainties associated with individual model techniques should be reduced, by incorporating several regression-based and machine-learning-based algorithms (like Maxent). Increase the sample size in marginal populations (Vale *et al.*, 2014), in the northern extreme (Adrar Atar) and in Mali, in order to reduce the bias towards central areas that probably affected our prediction for the present conditions, propagating for the past projections and ultimately refugee identification. Also, the quality of the environmental variables might have resulted in several artifacts in our estimation, like bands of suitable habitat in central Mali and overestimations during LIG given by a big red stain (Fig. 19). However, since this does not depend directly from the researcher the only solution is to wait for new climatic data. Other approaches, as using presence-absence data instead of only presence, might also improve the models. The absence of landcover variables for the past, hamper greatly the efficiency of the ecological models. However, this is a problem that, for now, it is impossible to overpass. Therefore, ecological niche models are mainly based on the realized ecological niche concept. This can lead to spurious correlations between species occurrence and climate and thus hinder model validation as well as casting doubts on model accuracy. For example, a species may not respond to climate only because other factors, such as competitive exclusion or predation, are confounding the response (see for example Pearson *et al.* 2006; Wiens *et al.*, 2009; Alvarado-Serrano, 2014; Pacifici *et al.*, 2015 for detailed descriptions).

## 5. Conclusion

This work presented, for the first time, molecular phylogenetic reconstruction of the all five extant members of North African endemic family, Ctenodactylidae, of mountain desert rodents. Moreover, with molecular genetic analysis (of non-invasively collected samples from scats) and ecological niche-modeling approaches, we recovered an mtDNA genetic structure in *Felovia vae*, mostly congruent with the geographic distribution of mountain ranges and constrained by topoclimatic factors.

Divergence between Ctenodactylidae family members dated to 13.1 (10.49-16.73) Myr (early to middle Miocene), whereas variability within *Felovia vae* dated to 0.72 (0.44-1.05) Myr, coinciding with major Pleistocene repeated and dramatic climatic changes. Geographic barriers (and probably habitat vicariance for the southern populations) given by large sandy areas during arid or semi-arid paleoclimatic periods, and wide/strong flowing rivers during more humid conditions may have played a key role on the evolutionary history of *Felovia vae*. These geographic barriers might have been cyclically dissolved, when savannah like habitats were present (e.g. humid periods), which might have able the species to disperse between mountain massifs. This signal was observed in the northern populations (between Adrar Atar and Tagant mountains) where the species seems to have been connected in the past, although currently their population might be isolated, as the field observations suggest. Demographic expansion was identified for the pooled clades and dated back to the Last Interglacial, with posterior population stabilization since Last Glacial Maximum until present, therefore supporting the importance of humid periods for dispersal and gene flow to occur. Potential past refugia areas were nearly overlapping with the current distribution of the species, suggesting a great stability of the species range throughout the time. The observed genetic pattern for *Felovia vae* is congruent with other vertebrates from the region, highlighting the common effect of past climatic events as the main driver for generating barriers to dispersal and, therefore, evolution of biological communities.

Information about this species biology is still very limited (IUCN DD status) making it potentially vulnerable to habitat alteration, given its strict habitat specialization to rocky patches with higher inclinations and the presence of *gueltas*. Further studies

will distinguish ongoing and historical gene flow between populations, and verify lineage affiliations of peripheral and yet unsampled populations. Additional sampling (especially in southern and northern distribution, in Adrar Atar mountains and Mali region) and increased sample sizes in each locality for detailed population study, as well as sequencing regions from multiple nuclear genome (e.g. nuclear genes and microsatellites) will provide further detailed information about genetic structuring, that certainly will find continuity in the current processes of global change experienced in the Sahara-Sahel. The current study gave important insights for a better understanding on the importance of past climatic and environmental oscillations on shaping biodiversity in the North West Africa region that should be further investigated.

## 6. References

- Ahmed AOC, Yasuda H, Hattori K, *et al.* (2008) Analysis of rainfall records (1923 – 2004) in Atar-Mauritania. *Geofizika* **25**, 53 – 64.
- Alvarado-Serrano DF, Knowles LL (2014) Ecological niche models in phylogeographic studies: Applications, advances and precautions. *Molecular Ecology Resources* **14**, 233 – 248.
- Amori G, Gippoliti S (2003) A higher-taxon approach to rodent conservation priorities for the 21st century. *Animal Biodiversity and Conservation* **26**, 1 – 18.
- Anthelme F, Waziri Mato M, Maley J (2008) Elevation and local refuges ensure persistence of mountain specific vegetation in the Nigerien Sahara. *Journal of Arid Environments* **72**, 2232 – 2242.
- Armitage SJ, Bristow CS, Drake NA (2015) West African monsoon dynamics inferred from abrupt fluctuations of Lake Mega-Chad. *Proceedings of the National Academy of Sciences* **112**, 8543 – 8548.
- Aulagnier S (2008) *Felovia vae*, *Massoutiera mzabi*, *Ctenodactylus gundi*, *Ctenodactylus vali*. In: IUCN 2009. IUCN red list of threatened species. Version 2009.1. Available at: <http://www.iucnredlist.org/>.
- Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard university press.
- Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy* **87**, 643 – 662.
- Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution* **16**, 37– 48.
- Barata M, Perera A, Martínez-Freiría F, *et al.* (2012) Cryptic diversity within the Moroccan day geckos *Quedenfeldtia* (Squamata: Gekkonidae): a multidisciplinary

approach using genetic, morphological and ecological data. *Biological Journal of the Linnean Society* **106**, 828 – 850.

Barbosa S, Pauperio J, Searle JB, *et al.* (2013) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. *Molecular Ecology Resources* **13**, 43 – 56.

Beja-Pereira A, Oliveira R, Alves PC, *et al.* (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular ecology resources* **9**, 1279 – 301.

Ben Faleh A, Granjon L, Tatard C, *et al.* (2012) Phylogeography of the greater Egyptian Jerboa (*Jaculus orientalis*) (Rodentia: *Dipodidae*) in Mediterranean North Africa. *Journal of Zoology* **286**, 208 – 220.

Boratyński Z, Alves PC, Berto S, *et al.* (2011) Introgression of mitochondrial DNA among *Myodes* voles: consequences for energetics? *BMC Evolutionary Biology* **11**, 355.

Boratyński Z, Brito JC, Mappes T (2012) The origin of two cryptic species of African desert jerboas (*Dipodidae*: *Jaculus*). *Biological Journal of the Linnean Society* **105**, 435 – 445.

Boratyński Z, Melo-Ferreira Z, Alves PC, *et al.* (2014) Molecular and ecological signs of mitochondrial adaptation: consequences for introgression? *Heredity* **113**, 277 – 86.

Boston ES, Puechmaille SJ, Scott DD, *et al.* (2012) Empirical assessment of non-invasive population genetics in bats: comparison of DNA quality from faecal and tissue samples. *Acta Chiropterologica* **14**, 45 – 52.

Brito C, Godinho R, Martínez-Freiría F, *et al.* (2014) Unravelling biodiversity, evolution and threats to conservation in the Sahara-Sahel. *Biological Reviews* **89**, 215 – 231.

Brito JC, Álvares F, Martínez-Freiría F, *et al.* (2010) Data on the distribution of mammals from Mauritania, West Africa. *Mammalia* **74**, 449 – 455.

- Brito JC, Fahd S, Geniez P, *et al.* (2011) Biogeography and conservation of viperids from North-West Africa: an application of ecological niche-based models and GIS. *Journal of Arid Environments* **75**, 1029 – 1037.
- Brooks N (2004) *Drought in the African Sahel: Long Term Perspectives and Future Prospects*. Tyndall Centre for Climate Change Research Working Paper 61. Available at: <http://www.tyndall.ac.uk/content/drought-african-sahel-long-term-perspectives-and-future-prospects>. Accessed 07.01.2010.
- Brouat C, Tatard C, Ba K, *et al.* (2009) Phylogeography of the Guinea multimammate mouse ( *Mastomys erythroleucus* ): a case study for Sahelian species in West Africa. *Journal of Biogeography* **36**, 2237 – 2250.
- Brown JH, Valone TJ, Curtin CG (1997) Reorganization of an arid ecosystem in response to recent climate change. *Proceedings of the National Academy of Science of USA* **94**, 9729 – 9733.
- Bryja J, Granjon L, Dobigny G, *et al.* (2010) Plio-Pleistocene history of West African Sudanian savanna and the phylogeography of the *Praomys daltoni* complex (Rodentia): The environment/geography/genetic interplay. *Molecular Ecology* **19**, 4783 – 4799.
- Buisson L, Thuiller W, Casajus N, *et al.* (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology* **16**, 1145 – 1157.
- Campos JC, Sillero N, Brito JC (2012) Normalized Difference Water Indexes have dissimilar performances in detecting seasonal and permanent water in the Sahara-Sahel transition zone. *Journal of Hydrology*, 464 – 465: 438 – 446.
- Carranza S, Arnold EN, Geniez P, *et al.* (2008) Radiation , multiple dispersal and parallelism in the skinks , *Chalcides* and *Sphenops* ( Squamata : *Scincidae* ), with comments on *Scincus* and *Scincopus* and the age of the Sahara Desert. *Molecular phylogenetics and evolution* **46**, 1071 – 1094.
- Coetzee N, Grubb P (2008) *Pectinator spekei*. In: IUCN 2009. IUCN red list of threatened species. Version 2009.1. Available at: <http://www.iucnredlist.org/>.

- Colangelo P, Verheyen E, Leirs H, *et al.* (2013) A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biological Journal of the Linnean Society* **108**, 901 – 916.
- Cooper A, Shine T, McCann T, *et al.* (2006) An ecological basis for sustainable land use of Eastern Mauritanian wetlands. *Journal of Arid Environments* **67**, 116 – 141.
- Davies J, Poulsen L, Schulte-Herbrüggen B, *et al.* (2012) *Conserving Dryland Biodiversity*. IUCN, Nairobi, Kenya.
- deMenocal PB (2004) African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters* **220**, 3 – 24.
- deMenocal PB, Tierney JE (2012) Green Sahara: African Humid Periods Paced by Earth's Orbital Changes. *Nature Education Knowledge* **3**, 12.
- Dilyte J (2014) Population structure and gene flow in desert environment: an application of molecular tools to isolated fish populations in West Africa. MSc thesis. University of Porto.
- Dobigny G, Aniskin V, Granjon L, *et al.* (2005) Recent radiation in West African *Taterillus* (Rodentia, Gerbillinae): the concerted role of chromosome and climatic changes. *Heredity* **95**, 358 – 368.
- Dobigny G, Tatard C, Gauthier P, *et al.* (2013) Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and sub-Saharan open habitats pleistocene history. *PLoS ONE* **8**, e77815. doi:10.1371/journal.pone.0077815.
- Dong B, Sutton R (2015) Dominant role of greenhouse-gas forcing in the recovery of Sahel rainfall. *Nature Climate Change*.
- Douady CJ, Catzeflis F, RamanJ, *et al.* (2003) The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order *Macroscelidea* (elephant shrews). *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8325 – 8330.

- Drake N, Blench RM, Armitage SJ, *et al.* (2011) Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 458 –62.
- Drummond AJ, Ho SYW, Phillips MJ, *et al.* (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**, 699. doi: 10.1371/journal.pbio.0040088.
- Drummond AJ, Suchard MA, Xie D, *et al.* (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* **29**, 1969 – 1973.
- Dumont HJ (1982) Relict distribution patterns of aquatic animals: Another tool in evaluating late Pleistocene climate changes in the Sahara and Sahel. *Palaeoeco Africa* **14**,1 – 24.
- Durant SM, Pettorelli N, Bashir S, *et al.* (2012) Forgotten Biodiversity in Desert Ecosystems. *Science* **336**, 1379 – 1380.
- Elith J, Graham CH, Anderson RP, *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129 – 151.
- ESRI (2010) Arc GIS-Arcmap 10.0. *Environmental Systems Research Institute Inc, USA.*
- Excoffier L, Lischer HEL (2010) Arlequin suite version 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564 – 567.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479 – 491.
- Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular biology and evolution* **23**, 691 – 700.
- Fernando P, Vidya T, Rajapakse C, *et al.* (2003) Reliable Noninvasive Genotyping: Fantasy or Reality? *Journal of Heredity* **94**, 115 – 123.

- Fonseca MM, Brito JC, Paulo OS, *et al.* (2009) Systematic and phylogeographical assessment of the *Acanthodactylus erythrurus* group (Reptilia: *Lacertidae*) based on phylogenetic analyses of mitochondrial and nuclear DNA. *Molecular phylogenetics and evolution* **51**, 131 – 42.
- Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915 – 925.
- Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693 – 709.
- Gasse F (2000) Hydrological changes in the African tropics since the Last Glacial Maximum. *Quaternary Science Reviews* **19**, 189 – 211.
- Gaubert P, Bloch C, Benyacoub S, *et al.* (2012) Reviving the African wolf *Canis lupus lupaster* in North and West Africa: a mitochondrial lineage ranging more than 6,000 kmwide. *PLoS One* **7**, 1 – 10. doi: 10.1371/journal.pone.0042740.
- Geniez P, Mateo JA, Geniez M, *et al.* (2004) *The Amphibians and Reptiles of Western Sahara*. Editions Chimaira, Frankfurt am Main.
- George W (1974) Note on the ecology of Gundis (*Ctenodactylidae*). *Symposium of the Zoological Society of London* **34**, 143 – 160.
- George W (1979) The chromosomes of the hystricomorphous family Ctenodactylidae (Rodentia: *Sciuromorpha*) and their bearing on the relationships of the four living genera. *Zoological Journal of the Linnean Society* **65**, 261 – 280.
- George W (1985) Cluster analysis and phylogenetics of five species of *Ctenodactylidae* (Rodentia). *Mammalia* **49**, 53 – 63.
- Giannini A (2015) Climate change comes to the Sahel. *Nature Publishing Group*, **5**, 720–721.
- Gilbert MTP, Bandelt H-J, Hofreiter M, *et al.* (2005) Assessing ancient DNA studies. *Trends in Ecology & Evolution* **20**, 541 – 544.

- Gomes Rodrigues H, Solé F, Charles C, et al. (2012) Evolutionary and biological implications of dental mesial drift in rodents: the case of the Ctenodactylidae (Rodentia, Mammalia). *PloS one* **7**, e50197. doi: 10.1371/journal.pone.0050197
- Gonçalves DV (2010) Evolution of the genus *Agama* in North-West Africa: combining phylogeographic and ecogeographical information for integrative taxonomical review. MSc thesis. University of Porto.
- Gonçalves DV, Brito JC, Crochet P-A, et al. (2012) Phylogeny of North African *Agama* lizards (Reptilia: *Agamidae*) and the role of the Sahara desert in vertebrate speciation. *Molecular phylogenetics and evolution* **64**, 582 – 591.
- Grant WS (2015) Problems and Cautions With Sequence Mismatch Analysis and Bayesian Skyline Plots to Infer Historical Demography. *Journal of Heredity*, 333 – 346.
- Guillaumet A, Crochet P, Pons J (2008) Climate-driven diversification in two widespread *Galerida larks*. *BMC Evolutionary Biology* **8**, 32.
- Guindon S, Lethiec F, Duroux P, et al. (2005) PHYML Online- a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research* **33**, 557 – 559.
- Hájková P, Zemanová B, Bryja J, et al. (2006) Factors affecting success of PCR amplification of microsatellite loci from otter faeces. *Molecular Ecology Notes* **6**, 559 – 562.
- Hamilton M (2011) *Population genetics*. John Wiley & Sons.
- Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution* **16**, 700 – 706.
- Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial-DNA mismatch distribution. *Human Biology* **66**, 591 – 600.
- Harpending HC, Sherry ST, Rogers AR, et al. (1993) Genetic structure of ancient human populations. *Current Anthropology* **34**, 483 – 496.

- Harris N (2003) *Atlas of the World's Deserts* (No. 467). Taylor & Francis.
- Hartenberger JL (1985) The order Rodentia: major questions on their evolutionary origin, relationships and suprafamilial systematics. In *Evolutionary relationships among rodents* **92**, 1 – 33. Springer US.
- Held IM, Delworth TL, Lu J, *et al.* (2005) Simulation of Sahel drought in the 20th and 21st centuries. *Proceedings of the National Academy of Sciences USA* **102**, 17891 – 17896.
- Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**, 570 – 580.
- Hellborg L, Ellegren H (2004) Low Levels of Nucleotide Diversity in Mammalian Y Chromosomes. *Molecular Biology and Evolution* **21**, 158 – 163.
- Hernandez PA, Graham CH, Master LL, *et al.* (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* **29**, 773 – 785.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **359**, 183 – 195.
- Hickerson MJ, Carstens BC, Cavender-Bares J, *et al.* (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution* **54**, 291 – 301.
- Hijmans RJ, Cameron SE, Parra JL, *et al.* (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965 – 1978.
- Huchon D, Catzeflis FM, Douzery EJ (2000) Variance of molecular datings, evolution of rodents and the phylogenetic affinities between *Ctenodactylidae* and *Hystriognathi*. *Proceedings of the Royal Society of London B: Biological Sciences* **267**, 393 – 402.

- Huchon D, Chevret P, Jordan U, *et al.* (2007) Multiple molecular evidences for a living mammalian fossil. *Proceedings of the National Academy of Sciences* **104**, 7495 – 7499.
- Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London B* **272**, 1525 – 1534.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**, 254 – 67.
- Jaeger JJ (1971) Un cténodactylidé (Mammalia, Rodentia) nouveau, *Irhoudia bohlini* n.g. n. sp. du Pléistocène inférieur du Maroc, rapports avec les formes actuelles et fossiles. *Notes et Memoires du Service Géologique du Maroc* **31**, 113–140.
- Jaeger JJ (1977) Les roungeurs du Miocène moyen et supérieur du Maghreb. *Palaeovertebrata* **8**, 1 – 66.
- Kingdon J, Happold D, Butynski T, *et al.* (2013) *Mammals of Africa* (Vol. 3). A&C Black.
- Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology* **19**, 3727 – 3745.
- Kröpelin S, Verschuren D, Lézine A, *et al.* (2008) Climate-Driven Ecosystem Succession in the Sahara : The Past 6000 Years. *Science* **320**, 765 – 768.
- Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**, 429 – 434.
- Le Houérou HN (1992) Outline of the biological history of the Sahara. *Journal of Arid Environments* **22**, 3 – 30.
- Le Houérou HN (1997) Climate, flora and fauna changes in the Sahara over the past 500 million years. *Journal of Arid Environments* **37**, 619 – 647.

- Leblanc MJ, Leduc C, Stagnitti F, *et al.* (2006) Evidence for Megalake Chad, north-central Africa, during the late Quaternary from satellite data. *Palaeogeography, Palaeoclimatology, Palaeoecology* **230**, 230 – 242.
- Lecompte E, Aplin K, Denys C, *et al.* (2008) Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. *BMC evolutionary Biology* **8**, 199.
- Lessa E, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences, USA* **100**, 10331 – 10334.
- Librando P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451 – 1452.
- Loarie SR, Duffy PB, Hamilton H, *et al.* (2009) The velocity of climate change. *Nature* **462**, 1052 – 1055.
- López-Antoñanzas R, Knoll F (2010) The oldest known Massoutiera (Rodentia: Ctenodactylinae). *Mammalia* **74**, 461 – 464.
- López-Antoñanzas R, Knoll F (2011) A comprehensive phylogeny of the gundis (Ctenodactylinae, Ctenodactylidae, Rodentia). *Journal of Systematic Palaeontology* **9**, 379 – 398.
- López-Antoñanzas R, Knoll F, Maksoud S, *et al.* (2015) First Miocene rodent from Lebanon provides the “missing link” between Asian and African gundis (Rodentia: Ctenodactylidae). *Scientific Reports* **5**, 12871.
- Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society B* **270**, 1193 – 202.
- Mahé G, Paturel JE (2009) 1896 – 2006 Sahelian annual rainfall variability and runoff increase of Sahelian Rivers. *Comptes Rendus Geoscience* **341**, 538 – 546.
- Maley J (2010) Climate and Palaeoenvironment evolution in north tropical Africa from the end of the Tertiary to the Upper Quaternary. *Palaeoecology of Africa* **30**, 227–278.

- Marmion M, Parviainen M, Luoto M, *et al.* (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions* **15**, 59 – 69.
- Marske KA, Leschen RAB, Buckley TR (2012) Concerted versus independent evolution and the search for multiple refugia: comparative phylogeography of four forest beetles. *Evolution* **66**, 1862 – 1877.
- Martínez-Freiría F, Velo-Antón G, Brito JC (2015) Trapped by climate: interglacial refuge and recent population expansion in the endemic Iberian adder *Vipera seoanei*. *Diversity and Distributions*, **21**, 331–344.
- Messerli B, Winiger M (1992) Climate, environmental change, and resources of the African Mountains from the Mediterranean to the Equator. *Mountain Research and Development* **12**, 315 – 336.
- Migliore J, Baumel A, Juin M, *et al.* (2013) Surviving in mountain climate refugia: new insights from the genetic diversity and structure of the relict shrub *Myrtus nivellei* (*Myrtaceae*) in the Sahara Desert. *PloS one* **8**, e73795. doi: 10.1371/journal.pone.0073795
- Miles L, Segurado P, Dawson TP, *et al.* (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography* **33**, 1704 – 1711.
- Monadjem A, Taylor PJ, Denys C, *et al.* (2015) *Rodents of Sub-Saharan Africa. A Biogeographic and Taxonomic Synthesis*. Walter de Gruyter, Berlin, Germany.
- Montgelard C, Bentz S, Tirard C, *et al.* (2002) Molecular systematics of *Sciurognathi* (Rodentia): the mitochondrial cytochrome b and 12S rRNA genes support the *Anomaluroidea* (*Pedetidae* and *Anomaluridae*). *Molecular phylogenetics and evolution* **22**, 220 – 233.
- Mouline K, Granjon L, Galan M, *et al.* (2008) Phylogeography of a Sahelian rodent species *Mastomys huberti*: a Plio-Pleistocene story of emergence and colonization of humid habitats. *Molecular ecology* **17**, 1036 – 53.

- Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals the longevity hypothesis. *Molecular Biology and Evolution* **25**, 120 – 130.
- Nicolas V, Granjon L, Duplantier J, *et al.* (2009) Phylogeography of spiny mice ( genus *Acomys* , Rodentia : Muridae ) from the south-western margin of the Sahara with taxonomic implications. *Biological Journal of the Linnean Society* **98**, 29 – 46.
- Nicolas V, Herbreteau V, Couloux A, *et al.* (2012) A remarkable case of micro-endemism in *Laonastes aenigmamus* (Diatomyidae, Rodentia) revealed by nuclear and mitochondrial DNA sequence data. *PloS one* **7**, e48145. doi: 10.1371/journal.pone.0048145
- Nicolas V, Mboumba J-F, Verheyen E, *et al.* (2008) Phylogeographic structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. *Journal of Biogeography* **35**, 2074 – 2089.
- Nutt K (2005) Philopatry of both sexes leads to the eormation oe multemale, multieemale groups in *ctenodactylus gundi* (Rodentia: Ctenodactylidae). *Journal of Mammalogy* **86**, 961 – 968.
- Nutt KJ (2003) Characterization of microsatellite markers for the North African gundi, *Ctenodactylus gundi* (Rodentia: Ctenodactylidae). *Molecular Ecology Notes* **3**, 511 – 513.
- Oliver MA, Webster R (1990) Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Systems* **4**, 313 – 332.
- Olson DM, Dinerstein E, Wikramanayake ED, *et al.* (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* **51**, 933 – 938.
- Otto-Bliesner BL, Marshall SJ, Overpeck JT, *et al.* (2006) Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**, 1751 – 1753.
- Pacifici M, Foden WB, Visconti P, *et al.* (2015) Assessing species vulnerability to climate change. *Nature Climate Change* **5**, 215 – 224.

- Padial JM, JL Tellería (2009) Observations of the Felou Gundi ( *Felovia vae* Lataste, 1886) in central Mauritania (Rodentia: Ctenodactylidae). *Mammalia* **73**, 153 – 154.
- Palumbi SR, Martin A, Romano S, *et al.* (1991) *The Simple Fool's Guide to PCR*, Version 2.0, privately published document compiled by S. Palumbi. Department of Zoology at University of Hawaii, Honolulu, HI, 96822.
- Partridge TC, Wood BA, deMenocal PB (1995) The influence of global climate changes and regional uplift on large- mammalian evolution in East and Southern Africa. In: Vrba E, Denton G, Partridge TC, Burckle L, eds. *Paleoclimate and evolution with emphasis on human evolution*. New Haven: Yale University Press, 331 – 356.
- Pearson RG, Thuiller W, Araújo MB, *et al.* (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography* **33**, 1704 – 1711.
- Peterson AT, Sober on J, Pearson RG, *et al.* (2011) *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton, New Jersey.
- Phillips S J, Anderson RP, Schapire RE (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling* **190**, 231 – 259.
- Piggott MP (2004) Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. *Wildlife Research* **31**, 485 – 493.
- Pinho C, Harris DJ, Ferrand N (2007) Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation. *Molecular Ecology* **16**, 1191 – 1205.
- Pinho C, Harris DJ, Ferrand N (2008) Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis* spp.) are an assemblage of incipient species. *BMC Evolutionary Biology* **8**, 63.
- Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular biology and evolution* **25**, 1253 – 6.

- Rambaut A (2009) *FigTree, a graphical viewer of phylogenetic trees*. Institute of Evolutionary Biology University of Edinburgh.
- Rambaut A, Drummond A (2007) Tracer v1. 5. <http://beast.bio.ed.ac.uk/Tracer>.
- Ramón-Laca A, Soriano L, Gleeson D, *et al.* (2015) A simple and effective method for obtaining mammal DNA from faeces. *Wildlife Biology*, **21**, 195–203.
- Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**, 2092 – 2100.
- Rogers AR, Harpending H (1992) Population-growth makes waves in the distribution of pairwise genetic-differences. *Molecular Biology and Evolution* **9**, 552 – 569.
- Rognon PP (1989) *Biogeographie d'un desert*. Plon, Paris.
- Ronquist F, Teslenko M, van der Mark P, *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539 – 542.
- Sanz JJ (2002) Climate change, birds: have their ecological consequences already been detected in the Mediterranean region? *Ardeola* **49**, 109 –120.
- Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 879 – 891.
- Schlitter DA (1989) African rodents of special concern: a preliminary assessment. In (W.Z. Lidicker, ed.) *Rodents. A world survey of species of conservation concern*. IUCN, Gland, Switzerland. pp. 33 – 39.
- Schulz E, Abichou A, Adamou A, *et al.* (2009) The desert in the Sahara. Transitions and boundaries. In *Palaeoecology of Africa and the Surrounding Islands*, Volume 29 (eds R. Braumhauer and J. Runge), pp. 63 – 89.
- Seddon JM, Santucci F, Reeve NJ, *et al.* (2001) DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*. Pleistocene refugia, postglacial expansion and colonization routes. *Molecular Ecology* **10**, 2187 – 2198.

- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics* **68**, 978 – 989
- Swezey CS (2009) Cenozoic stratigraphy of the Sahara, Northern Africa. *Journal of African Earth Sciences* **53**, 89 – 121.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585 – 595.
- Tamura K, Stecher G, Peterson D, *et al.* (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725 – 2729.
- Tellería, JL (2009) Biodiversidad y Conservación en la Meseta de Tagant. Universidad Complutense de Madrid. Available at: <http://www.tagant.org>. Accessed 04.09.2009.
- Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21**, 3907 – 30.
- Trape S (2009) Impact of climate change on the relict tropical fish fauna of Central Sahara: threat for the survival of Adrar mountains fishes, Mauritania. *PLoS One* **4**, 1 – 10. doi: 10.1371/journal.pone.0004400.
- Triant DA, DeWoody JA (2006) Accelerated molecular evolution in *Microtus* (Rodentia) as assessed via complete mitochondrial genome sequences. *Genetica* **128**, 95 – 108.
- UNEP (1992) *World Atlas of Desertification*. 1st ed. N.Middleton and D.S.G. Thomas (eds.). London, Arnold.
- USGS (2006) Shuttle Radar Topography Mission (SRTM): Mapping the World in 3NDimensions. United States Geological Survey. <http://srtm.usgs.gov/index.html>.
- Vale CG, Álvares F, Brito JC (2012a) Distribution, suitable areas and conservation status of the Felou gundi (*Felovia vae* Lataste 1886). *Mammalia* **76**, 201 – 207.

- Vale CG, Pimm SL, Brito JC (2015) Overlooked Mountain Rock Pools in Deserts Are Critical Local Hotspots of Biodiversity. *Plos One* **10**, e0118367. doi: 10.1371/journal.pone.0118367
- Vale CG, Tarroso P, Brito JC (2014) Predicting species distribution at range margins: Testing the effects of study area extent, resolution and threshold selection in the Sahara-Sahel transition zone. *Diversity and Distributions* **20**, 20 – 33.
- Vale CG, Tarroso P, Campos JC, *et al.* (2012b) Distribution, suitable areas and conservation status of the Boulenger's agama (*Agama boulengeri*, Lataste 1886). *Amphibia-Reptilia* **33**, 526–532.
- Velo-Antón G, Godinho R, Campos JC, *et al.* (2014) Should I stay or should I go? Dispersal and population structure in small, isolated desert populations of West African crocodiles. *PloS one* **9**, e94626. doi: 10.1371/journal.pone.0094626
- Veríssimo J (2014) Phylogeography and current patterns of genetic diversity and structure of the Mediterranean pond turtle. MSc thesis. University of Porto.
- Wagner P, Melville J, Wilms TM, *et al.* (2011) Opening a box of cryptic taxa—the first review of the North African desert lizards in the *Trapelus mutabilis* Merrem, 1820 complex (Squamata: *Agamidae*) with descriptions of new taxa. *Zoological Journal of the Linnean Society* **163**, 884 – 912.
- Waits LP, 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**, 1419 – 1433.
- Waltari E, Hijmans RJ, Peterson AT, *et al.* (2007) Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS ONE* **2**, e563. doi: 10.1371/journal.pone.0000563
- Ward D (2009) *Biology of Deserts*. Oxford University Press, Oxford. Wezel,
- Wehausen (2004) Experiments in DNA Extraction and PCR Amplification from Bighorn Sheep Feces : the Importance of DNA Extraction Method. *Heredity* **95**, 503 – 509.

- Weiss S, Ferrand N (2007) Current perspectives in phylogeography and the significance of South European refugia in the creation and maintenance of European biodiversity. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 341–357. Springer Netherlands.
- Wiens JA, Stralberg D, Jongsomjit D, *et al.* (2009) Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences* **106**, 19729 – 19736.
- Wisz MS, Hijmans RJ, Li J, *et al.* (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distribution* **14**, 763 – 773.
- Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular ecology* **12**, 563 – 584.

## 7. Supplementary material

**Tab. S1** - Summary table of the *Felovia vae* samples. Mitochondrial clades are described along with the resulted haplotype both for Cytb and GHR gene fragments. Recombining haplotypes that have been excluded from further analyses are represented by R\*.

Code	Species	Country, province, locality	Locality groups(30 km)	Mountain	Latitude	Longitude	Cytb clade	Cytb Hp	GHR Hp
ZBSC0220	<i>Felovia vae</i>	Mauritania, Adrar Atar, Chinguetti	1	Atar	20.547	-12.690	A	1	
ZBSC0220	<i>Felovia vae</i>	Mauritania, Adrar Atar, Chinguetti	1	Atar	20.547	-12.690	A	1	
ZBSC0220	<i>Felovia vae</i>	Mauritania, Adrar Atar, Chinguetti	1	Atar	20.547	-12.690	A	1	5/5
ZBSC0221	<i>Felovia vae</i>	Mauritania, Adrar Atar, Chinguetti	1	Atar	20.550	-12.690	A	1	
ZBSC0222	<i>Felovia vae</i>	Mauritania, Adrar Atar, Chinguetti	1	Atar	20.555	-12.692	A	1	5/5
ZBSC0502	<i>Felovia vae</i>	Mauritania	2	Tagant	18.365	-9.051	A	7	1/1
ZBSC0510	<i>Felovia vae</i>	Mauritania	3	Tagant	18.409	-9.354	A	7	
ZBSC0511	<i>Felovia vae</i>	Mauritania	4	Tagant	18.579	-9.816	A	8	
ZBSC0512	<i>Felovia vae</i>	Mauritania	4	Tagant	18.581	-9.815	A	8	
ZBSC0516	<i>Felovia vae</i>	Mauritania	4	Tagant	18.592	-10.027	A	9	
ZBSC0516	<i>Felovia vae</i>	Mauritania	4	Tagant	18.592	-10.027	A	10	
ZBSC0525	<i>Felovia vae</i>	Mauritania	5	Tagant	18.566	-11.190	A	11	1/1
ZBSC0527	<i>Felovia vae</i>	Mauritania	5	Tagant	18.490	-11.381	A	12	
ZBSC0528	<i>Felovia vae</i>	Mauritania	5	Tagant	18.443	-11.387	A	1	
ZBSC0532	<i>Felovia vae</i>	Mauritania	5	Tagant	18.443	-11.383	A	13	1/1
ZBSC0539	<i>Felovia vae</i>	Mauritania	6	Tagant	18.403	-11.794	A	15	1/1
ZBSC0552	<i>Felovia vae</i>	Mauritania	6	Tagant	18.207	-11.731	A	2	
ZBSC0538	<i>Felovia vae</i>	Mauritania	6	Tagant	18.465	-11.735	A	14	
ZBSC0257	<i>Felovia vae</i>	Mauritania, Tagant, Matmâta (guelta)	7	Tagant	17.888	-12.115	A	3	
ZBSC0260	<i>Felovia vae</i>	Mauritania, Tagant, Matmâta (guelta)	7	Tagant	17.891	-12.115	A	4	
ZBSC0261	<i>Felovia vae</i>	Mauritania, Tagant, Tartêga (guelta)	7	Tagant	17.877	-12.091	A	5	1/6
ZBSC0263	<i>Felovia vae</i>	Mauritania, Tagant, Tin Waadine (guelta)	7	Tagant	18.055	-11.943	A	6	1/6
ZBSC0562	<i>Felovia vae</i>	Mauritania	7	Tagant	18.151	-12.066	A	2	

Biogeography and phylogenetic position of a Sahara-Sahel mountain endemic, *Felovia vae*  
(Ctenodactylidae)

ZBSC0563	<i>Felovia vae</i>	Mauritania	7	Tagant	18.151	-12.066	<b>A</b>	2	1/5
ZBSC0374	<i>Felovia vae</i>	Mauritania, Assaba, Nouâmléin	8	Tagant	17.132	-11.523	<b>A</b>	18	
ZBSC0372	<i>Felovia vae</i>	Mauritania, Assaba, Nouâmléin	8	Tagant	17.132	-11.522	<b>A</b>	16	
ZBSC0373	<i>Felovia vae</i>	Mauritania, Assaba, Nouâmléin	8	Tagant	17.132	-11.522	<b>A</b>	17	
ZBSC0375	<i>Felovia vae</i>	Mauritania, Assaba, Sassamba	9	Tagant	17.134	-12.091	<b>A</b>	19	
ZBSC0376	<i>Felovia vae</i>	Mauritania, Assaba, Sassamba	9	Tagant	17.126	-12.095	<b>A</b>	20	
ZBSC0377	<i>Felovia vae</i>	Mauritania, Assaba, Sassamba	9	Tagant	17.125	-12.096	<b>A</b>	20	
ZBSC0379	<i>Felovia vae</i>	Mauritania, Assaba, Sassamba	9	Tagant	17.133	-12.087	<b>A</b>	20	
ZBSC0053	<i>Felovia vae</i>	Mauritania, Assaba, Aouînet Nanâga	10	Assaba	17.152	-12.199	<b>B</b>	25	
ZBSC0049	<i>Felovia vae</i>	Mauritania, Assaba, Fom Goussas	11	Assaba	16.547	-12.010	<b>B</b>	23	
ZBSC0050	<i>Felovia vae</i>	Mauritania, Assaba, Fom Goussas	11	Assaba	16.547	-12.010	<b>B</b>	22	
ZBSC0350	<i>Felovia vae</i>	Mauritania, Assaba, Tarf et Mzeimmet	11	Assaba	16.447	-11.777	<b>B</b>	33	
ZBSC0352	<i>Felovia vae</i>	Mauritania, Assaba, El Fouz	11	Assaba	16.565	-11.717	<b>B</b>	34	
ZBSC0353	<i>Felovia vae</i>	Mauritania, Assaba, El Fouz	11	Assaba	16.530	-11.807	<b>B</b>	35	
ZBSC0355	<i>Felovia vae</i>	Mauritania, Assaba, Tarf Tentahra	11	Assaba	16.619	-11.845	<b>B</b>	34	
ZBSC0048	<i>Felovia vae</i>	Mauritania, Assaba, Aouînet Gdah	12	Assaba	16.292	-12.011	<b>B</b>	24	1/1
ZBSC0315	<i>Felovia vae</i>	Mauritania, Assaba, Guelta Mousse	12	Assaba	16.335	-11.978	<b>B</b>	29	
ZBSC0318	<i>Felovia vae</i>	Mauritania, Assaba, Guelta Mousse	12	Assaba	16.333	-11.977	<b>B</b>	23	
ZBSC0054	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta goumbel	13.2	Assaba	15.957	-12.010	<b>A</b>	21	
ZBSC0054	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta goumbel	13.2	Assaba	15.957	-12.010	<b>A</b>	21	
ZBSC0054	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta goumbel	13.2	Assaba	15.957	-12.010	<b>A</b>	21	R4*/R4*
ZBSC0054	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta goumbel	13.2	Assaba	15.957	-12.010	<b>A</b>	21	
ZBSC0309	<i>Felovia vae</i>	Mauritania, Assaba, Aroueiji	13.1	Assaba	15.958	-12.010	<b>B</b>	26	
ZBSC0310	<i>Felovia vae</i>	Mauritania, Assaba, Aroueiji	13.1	Assaba	15.957	-12.010	<b>B</b>	24	
ZBSC0311	<i>Felovia vae</i>	Mauritania, Assaba, Aroueiji	13.1	Assaba	15.954	-12.010	<b>B</b>	27	
ZBSC0312	<i>Felovia vae</i>	Mauritania, Assaba, Aroueiji	13.1	Assaba	15.959	-12.008	<b>B</b>	28	
ZBSC0313	<i>Felovia vae</i>	Mauritania, Assaba, Aroueiji	13.1	Assaba	15.959	-12.007	<b>B</b>	24	
ZBSC0323	<i>Felovia vae</i>	Mauritania, Assaba, Diandiourou	13.1	Assaba	15.999	-11.872	<b>B</b>	24	
ZBSC0330	<i>Felovia vae</i>	Mauritania, Assaba, Diandiourou	13.1	Assaba	15.975	-11.873	<b>B</b>	30	7/7
ZBSC0332	<i>Felovia vae</i>	Mauritania, Assaba, Diandiourou	13.1	Assaba	15.942	-11.928	<b>B</b>	24	1/1
ZBSC0336	<i>Felovia vae</i>	Mauritania, Assaba, Diandiourou	13.1	Assaba	15.947	-11.932	<b>B</b>	24	
ZBSC0341	<i>Felovia vae</i>	Mauritania, Assaba, Tarf el Mgueissem	13.1	Assaba	16.051	-11.673	<b>B</b>	31	
ZBSC0347	<i>Felovia vae</i>	Mauritania, Assaba, Akabane	13.1	Assaba	16.163	-11.750	<b>B</b>	32	
ZBSC0041	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta Goumbel	13.1	Assaba	15.957	-12.010	<b>B</b>	24	

Biogeography and phylogenetic position of a Sahara-Sahel mountain endemic, *Felovia vae*  
(Ctenodactylidae)

ZBSC0042	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta Goumbel	13.1	Assaba	15.957	-12.010	<b>B</b>	24	
ZBSC0043	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta Goumbel	13.1	Assaba	15.957	-12.010	<b>B</b>	24	
ZBSC0044	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta Goumbel	13.1	Assaba	15.957	-12.010	<b>B</b>	27	
ZBSC0046	<i>Felovia vae</i>	Mauritania, Guidimaka, Oued el Barda	13.1	Assaba	15.965	-11.916	<b>B</b>	24	
ZBSC0047	<i>Felovia vae</i>	Mauritania, Guidimaka, Oued el 'Adam	13.1	Assaba	15.908	-11.921	<b>B</b>	24	
ZBSC0054	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta Goumbel	13.1	Assaba	15.957	-12.010	<b>B</b>	36	1/1
ZBSC0054	<i>Felovia vae</i>	Mauritania, Guidimaka, Guelta Goumbel	13.1	Assaba	15.957	-12.010	<b>B</b>	36	
ZBSC0368	<i>Felovia vae</i>	Mauritania, Afollé, Kiffa	14	Afollé	16.762	-11.223	<b>C</b>	40	
ZBSC0359	<i>Felovia vae</i>	Mauritania, Afollé, Kiffa	14	Afollé	16.684	-11.289	<b>C</b>	37	
ZBSC0362	<i>Felovia vae</i>	Mauritania, Afollé, Kiffa	14	Afollé	16.730	-11.129	<b>C</b>	38	
ZBSC0364	<i>Felovia vae</i>	Mauritania, Afollé, Kiffa	14	Afollé	16.728	-11.128	<b>C</b>	38	
ZBSC0365	<i>Felovia vae</i>	Mauritania, Afollé, Kiffa	14	Afollé	16.730	-11.127	<b>C</b>	39	
ZBSC0371	<i>Felovia vae</i>	Mauritania, Afollé, Kiffa	14	Afollé	16.764	-11.267	<b>C</b>	41	
ZBSC0371	<i>Felovia vae</i>	Mauritania, Afollé, Kiffa	14	Afollé	16.764	-11.267	<b>C</b>	41	
ZBSC0435	<i>Felovia vae</i>	Mauritania	15	Afollé	16.262	-11.053	<b>C</b>	42	R8*/R8*
ZBSC0437	<i>Felovia vae</i>	Mauritania	15	Afollé	16.490	-11.058	<b>C</b>	44	
ZBSC0436	<i>Felovia vae</i>	Mauritania	15	Afollé	16.423	-11.007	<b>C</b>	43	R8*/R8*
JCB11267	<i>Felovia vae</i>	Mauritania, Hodh el Gharbi, Djafarat	16	Afollé	16.051	-10.539	<b>C</b>	48	
JCB11347	<i>Felovia vae</i>	Mauritania, Hodh el Gharbi, Tintane	17	Afollé	16.371	-10.245	<b>C</b>	49	1/R2*
JCB11317	<i>Felovia vae</i>	Mauritania, Hodh el Gharbi, Dmouch Telli	18	Afollé	16.642	-9.915	<b>C</b>	49	3/3
ZBSC0479	<i>Felovia vae</i>	Mauritania	19	Afollé	16.500	-9.565	<b>C</b>	45	
JCB11105	<i>Felovia vae</i>	Mauritania, Assaba, Oued Lemhara	20	Afollé	15.695	-11.238	<b>C</b>	46	1/R2*
JCB11106	<i>Felovia vae</i>	Mauritania, Assaba, Oued Lemhara	20	Afollé	15.695	-11.238	<b>C</b>	46	
JCB11107	<i>Felovia vae</i>	Mauritania, Assaba, Oued Lemhara	20	Afollé	15.695	-11.238	<b>C</b>	47	
JCB11158	<i>Felovia vae</i>	Mauritania, Assaba, Sibar	21	Afollé	15.564	-10.982	<b>C</b>	41	
ZBSC0470	<i>Felovia vae</i>	Mali	22	Felou	14.492	-10.249	<b>D</b>	54	1/9
ZBSC0467	<i>Felovia vae</i>	Mali	23	Felou	14.321	-10.414	<b>D</b>	53	9/9
ZBSC0466	<i>Felovia vae</i>	Mali	24	Felou	14.203	-10.737	<b>D</b>	53	9/9
ZBSC0454	<i>Felovia vae</i>	Mali	25	Felou	14.359	-11.371	<b>E</b>	51	
ZBSC0456	<i>Felovia vae</i>	Mali	25	Felou	14.150	-11.261	<b>E</b>	52	9/9
ZBSC0451	<i>Felovia vae</i>	Mali	25	Felou	14.390	-11.384	<b>E</b>	50	
ZBSC0452	<i>Felovia vae</i>	Mali	25	Felou	14.389	-11.387	<b>E</b>	51	
ZBSC0453	<i>Felovia vae</i>	Mali	25	Felou	14.363	-11.369	<b>E</b>	51	
ZBSC0455	<i>Felovia vae</i>	Mali	25	Felou	14.362	-11.373	<b>E</b>	51	



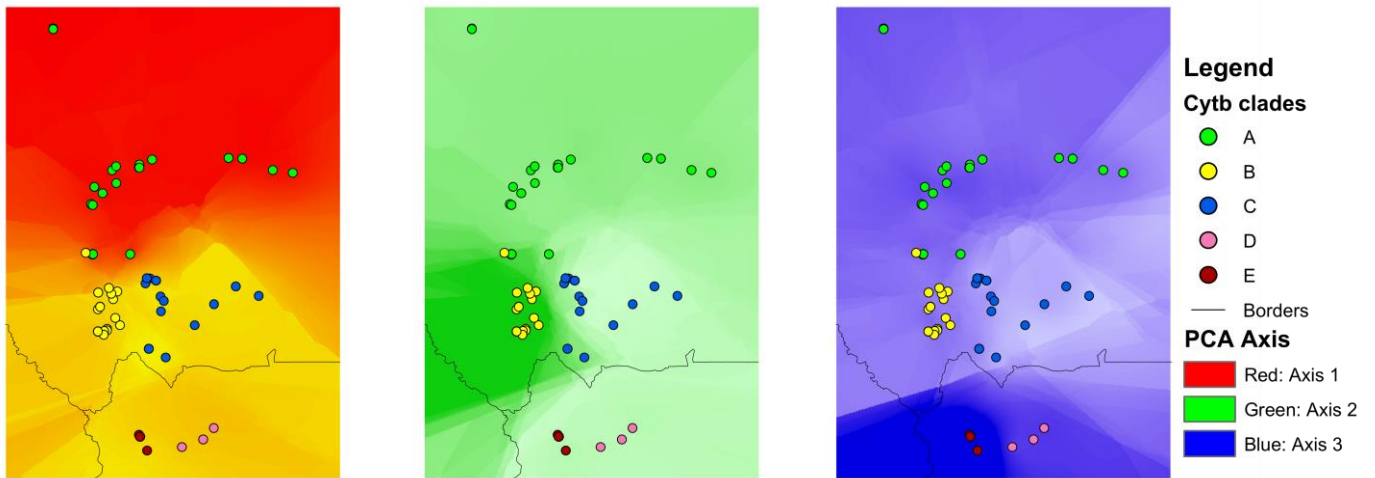
**Tab. S2** - Estimated times to the most recent common ancestor (TMRCA; mean values and 95% HPD) for the three approaches used, regarding the type of calibration. The “Ctenodactylidae tree” uses the family members with a set of particular calibration constrains (set = approach) to estimate substitution rate values. Meaning: Eocene (Eoc.); Oligocene (Olig.); Miocene (Mio.); Pliocene (Plio.); Pleistocene (Ple.).

Approach	Dating		Type of Calibration			
	Mean (Myr)	95% HPD (Myr)	Epoch	Mutation rate (2%)	<i>Laonastes – Ctenodactylidae</i> (44.3 Myr ± 3.5 Myr)	<i>M. mzabi - Ctenodactylus</i> (9.7-13.0 Myr)
<b>Ctenodactylidae members (A)</b>						
TMRCA ( <i>Laonastes + Ctenodactylidae</i> )	40.78	33.69-48.18	Eoc-Olig.			
TMRCA ( <i>P. spekei + M. mzabi + C. vali + C. gundi + F. vae</i> )	26.93	19.83-34.94	Olig-Mio.			
TMRCA ( <i>M. mzabi + C. vali + C. gundi + F. vae</i> )	20.00	13.94-26.68	Mio.	Yes	Yes	Yes
TMRCA ( <i>M. mzabi + C. vali + C. gundi</i> )	14.50	9.78-19.65	Mio-Plio.			
TMRCA ( <i>C. vali + C. gundi</i> )	6.85	4.35-9.77	Plio-Ple			
TMRCA ( <i>F. vae: A + B + C + D + E</i> )	1.36	0.74-2.05	Ple.			
TMRCA ( <i>F. vae: C + D + E</i> )	0.82	0.42-1.33				
TMRCA ( <i>F. vae: D + E</i> )	0.51	0.22-0.88				
<b>Ctenodactylidae members (B)</b>						
TMRCA ( <i>P. spekei + M. mzabi + C. vali + C. gundi + F. vae</i> )	13.10	10.49-16.73	Mio.	Yes	No	Yes
TMRCA ( <i>M. mzabi + C. vali + C. gundi + F. vae</i> )	10.01	9.70-12.01				
TMRCA ( <i>M. mzabi + C. vali + C. gundi</i> )	7.40	5.63-9.37	Plio.			
TMRCA ( <i>C. vali + C. gundi</i> )	3.53	2.41-4.81				
TMRCA ( <i>F. vae: A + B + C + D + E</i> )	0.72	0.44-1.05	Ple.			
TMRCA ( <i>F. vae: C + D + E</i> )	0.43	0.23-0.66				
TMRCA ( <i>F. vae: D + E</i> )	0.27	0.12-0.45				
<b>Ctenodactylidae members (C)</b>						
TMRCA ( <i>P. spekei + M. mzabi + C. vali + C. gundi + F. vae</i> )	13.07	7.58-22.49	Ple.	Yes	No	No
TMRCA ( <i>M. mzabi + C. vali + C. gundi + F. vae</i> )	10.19	5.73-17.33				
TMRCA ( <i>M. mzabi + C. vali + C. gundi</i> )	7.36	4.03-12.53				
TMRCA ( <i>C. vali + C. gundi</i> )	3.52	1.95-6.17				
TMRCA ( <i>F. vae: A + B + C + D + E</i> )	0.71	0.36-1.26				
TMRCA ( <i>F. vae: C + D + E</i> )	0.43	0.19-0.80				
TMRCA ( <i>F. vae: D + E</i> )	0.27	0.09-0.52				

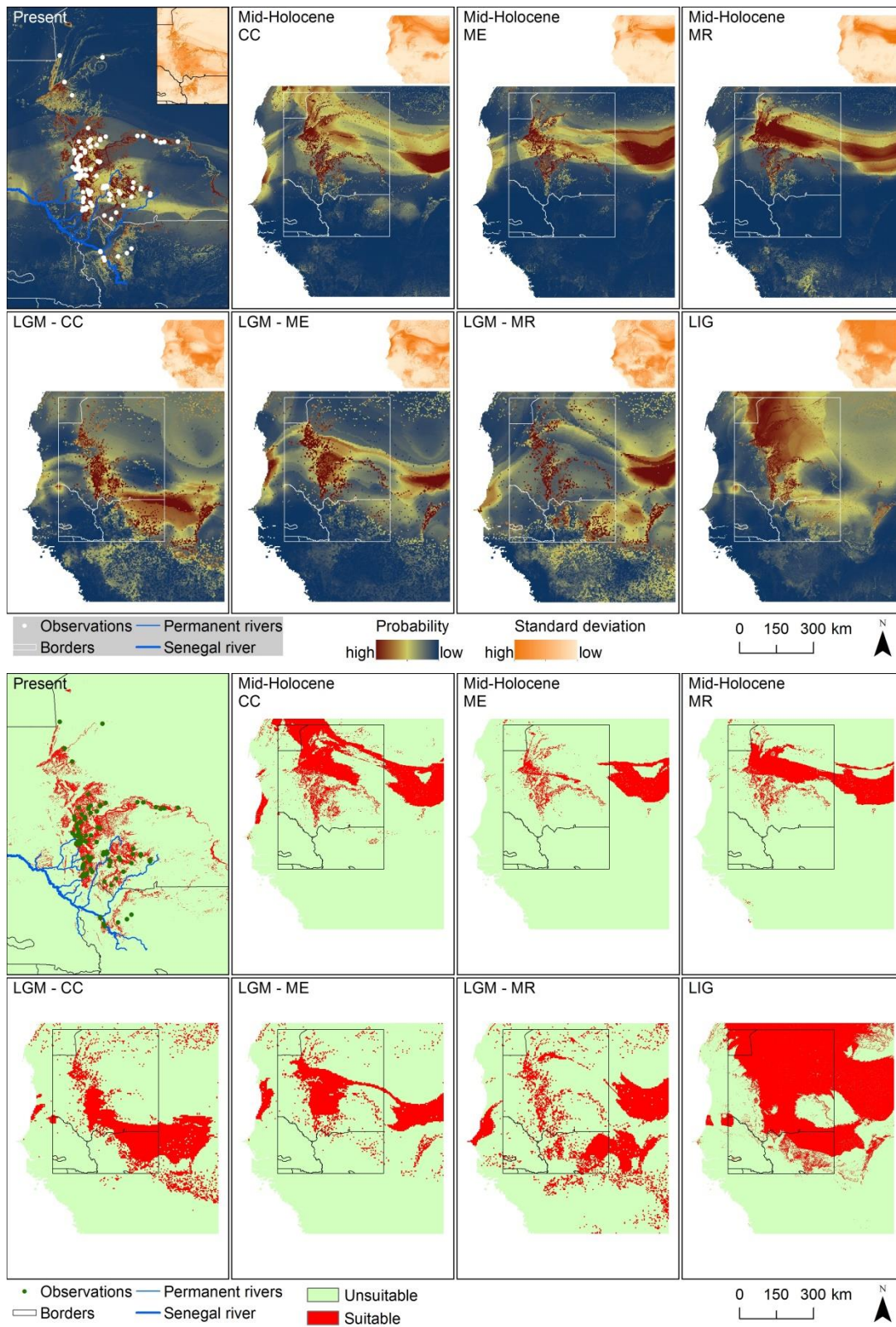
Biogeography and phylogenetic position of a Sahara-Sahel mountain endemic, *Felovia vae*  
(Ctenodactylidae)

**Tab. S3** - Estimated times to the most recent common ancestor (TMRCA; mean values and 95% HPD) for the three approaches used, regarding the type of calibration. The “*Felovia vae* tree”, which include only *F. vae* haplotypes, uses the substitution rates estimated in Tab. S2. Meaning: Eocene (Eoc.); Oligocene (Olig.); Miocene (Mio.); Pliocene (Plio.); Pleistocene (Ple.).

Approach	Dating		Type of Calibration		
	Mean (Myr)	95% HPD (Myr)	Epoch	Substitution rate (substitution/site/Myr)	95% HPD (substitution/site/Myr)
<b><i>Felovia vae</i> clades (A)</b>					
TMRCA ( <i>F. vae</i> : A + B + C + D + E)	1.77	0.83-3.23	Plio.	0.0115	0.0076 - 0.0160
TMRCA ( <i>F. vae</i> : C + D + E )	1.03	0.43-1.93	Plio-Ple.		
TMRCA ( <i>F. vae</i> : D + E )	0.69	0.25-1.34	Ple.		
TMRCA ( <i>F. vae</i> : A)	0.47	0.19-0.90			
TMRCA ( <i>F. vae</i> : B)	0.54	0.19-1.01			
TMRCA ( <i>F. vae</i> : C)	0.43	0.43-1.93			
TMRCA ( <i>F. vae</i> : D)	0.32	0.045-0.74			
TMRCA ( <i>F. vae</i> : E)	0.18	0.026-0.43			
<b><i>Felovia vae</i> clades (B)</b>					
TMRCA ( <i>F. vae</i> : A + B + C + D + E)	0.94	0.52-1.54	Ple.	0.0211	0.0158 – 0.0265
TMRCA ( <i>F. vae</i> : C + D + E )	0.55	0.29-0.93			
TMRCA ( <i>F. vae</i> : D + E )	0.36	0.17-0.65			
TMRCA ( <i>F. vae</i> : A)	0.24	0.12-0.43			
TMRCA ( <i>F. vae</i> : B)	0.27	0.12-0.52			
TMRCA ( <i>F. vae</i> : C)	0.22	0.09-0.39			
TMRCA ( <i>F. vae</i> : D)	0.16	0.03-0.37			
TMRCA ( <i>F. vae</i> : E)	0.08	0.016-0.20			
<b><i>Felovia vae</i> clades (C)</b>					
TMRCA ( <i>F. vae</i> : A + B + C + D + E)	1.00	0.49-1.95	Plio-Ple.	0.0206	0.0118 – 0.0308
TMRCA ( <i>F. vae</i> : C + D + E )	0.58	0.26-1.14	Ple.		
TMRCA ( <i>F. vae</i> : D + E )	0.38	0.15-0.78			
TMRCA ( <i>F. vae</i> : A)	0.26	0.10-0.52			
TMRCA ( <i>F. vae</i> : B)	0.29	0.11-0.63			
TMRCA ( <i>F. vae</i> : C)	0.23	0.09-0.49			
TMRCA ( <i>F. vae</i> : D)	0.17	0.03-0.44			
TMRCA ( <i>F. vae</i> : E)	0.09	0.01-0.23			



**Fig. S1** – Depiction of the three principal components of the PCA, with decreasing relevance from left to right for explaining diversity patterns. On the left the first component clearly separates the Adrar Atar-Tagant from the Assaba Afollé mountains, while in the second image we can see a separation of the Assaba from the other mountains, and on the third component we can observe the isolation of the Felou populations against the other systems.



**Fig. S2** – Average probability of *Felovia vae* occurrence, standard deviation (small insets) (up) and predicted suitable areas (down) for the current conditions and for projections to the Mid-Holocene, Last Glacial Maximum (LGM) and Last Interglacial (LIG). Different global circulation models were used CCSM4, MIROC-ESM and MPI-ESM for Mid-Holocene and LGM and the NCAR-CCSM for LIG (Mid-Holocene – CC, ME and MR; LGM – CC, ME and MR; LIG – NCAR-CCSM).

