University of Salford MANCHESTER

Elucidating cryptic diversity in East African frogs: the case of *Arthroleptis francei* Loveridge, 1953

Natasha Louise Woest

This Thesis is Submitted in Fulfilment of the Requirements for the Degree of MSc by Research

School of Science, Engineering and Environment
University of Salford, Salford, UK

2020

"There are as many species as the infinite being created diverse forms in the beginning, which, following the laws of generation, produced many others, but always similar to them: therefore, there are as many species as we have different structures before us today."

~ Carl Linnaeus ~

Table of Contents

i. List of Figures

ii. List of Tables

iii. Acknowledgements

There are many individuals and organisations who made this research possible and all deserve my gratitude. Firstly, I would like to thank my supervisor, Robert Jehle who provided the initial Natural History Museum (NHM) link, for the time to proofread this thesis and the constructive advice along the way. Secondly, a huge thank you to Simon Loader from the Natural History Museum along with Gabriela Bittencourt-Silva for all the advice, laboratory training, time and allowing me to work on the tissue samples in this research. Their help and advice have been invaluable throughout this research, and their first-hand knowledge on anuran from Sub-Saharan Africa provided a helpful background. The NHM provided a connection with Krystal Tolley from the South African Biological Institute (SANBI) who included me on their 2018 expedition which was funded by National Geographic Society. This expedition gave me the first-hand experience in field work and a very helpful insight into frog capture, tissue extraction and preservation, identification and learning the habitats in this very delicate ecosystem. Valuable information and training were provided by Gabriela and Krystal with the additional expert input from Werner Conradie, a novel African herpetologist from Port Elizabeth Museum. Logistics were arranged by Mike Scott (Khangela Safaris) and Dave Langerman which made the ease for a full focus on the research at hand. Avelino Miguel was part of the team from the University of Zambezia provided a helping hand in field work and translation with the local people, along with GBS. I would also like to thank my father, Patrick Woest for being one of the drivers and providing a vehicle for the expedition and accompanying the research team in collecting specimens and being a moral support along the way. The endless vehicle issues were worth it. Further funding was provided by the student allowance and a Santander Grant from the University of Salford which helped fund my travel to London to conduct my laboratory work and for the consumables used. A personal thank you to my partner, Lucy Stevens for the emotional support during this Masters and to all my family and friends who made this possible. Thank you to Christopher Barratt for helping with the phylospatial coding at the eleventh hour. An extended thank you to Arthur Loveridge for the initial collection of the specimens of *A. francei* in 1956 and to David for the DNA sequences on GenBank from his 2008 paper.

iv. Statement of originality:

I declare that, with the exception of any statements to the contrary, the contents of this report/dissertation are my own work, that the data presented herein has been obtained by experimentation and that no part of the report has been copied from previous reports/dissertations, books, manuscripts, research papers or the internet.

Signed..... Print Name.....Natasha Woest.....

Date.....21/07/2020.....

v. Abbreviations

Asl.	-	Above Sea Level
BH	-	Biodiversity Hotspots
Bp	-	Base Pairs
dB	-	Decibels
bGMYC	-	Bayesian Mixed Yule-Coalescent (bGMYC)
BPA	-	Bayesian Phylogenetic Analysis
Bpp	-	Bayesian Posterior Probability
DD3	-	Forearm Third Digit Disk Diameter
DNA	-	Deoxyribonucleic Acid
EA	-	Eastern Arc
EABH	-	East African Biodiversity Hotspot
EABR	-	East African Biodiversity Region
EACF	-	Eastern Arc Coastal Forest
EAMSR	-	Eastern Arc Mountains and Southern Rift
ED	-	Eye Diameter
END	-	Eye-Nostril Distance
ENM	-	Ecological Niche Models
FL	-	Foot Length
FOL	-	Forearm Length
CEPF	-	Critical Ecosystem Partnership Fund
DD	-	Data deficient
EARS	-	East African Rift System
HL	-	Head Length
HPD	-	Highest Posterior Density
HW	-	Head Width
Hz	-	Hertz
IND	-	Internarial Distance
IOD	-	Interorbital Distance
IUCN	-	International Union for the Conservation of Nature
KBH	-	Key Biodiversity Hotspots
kHz	-	Kilohertz
LHU	-	Length of Humerus
ma	-	Millions of years
MAFFT	-	Multiple Alignment using Fast Fourier Transform
NHML	-	Natural History Museum London
nuDNA	-	Nuclear Deoxyribonucleic Acid
PD	-	Phylogenetic Diversity
PE	-	Phylogenetic Endemism
rDNA	-	Ribosomal Deoxyribonucleic Acid
SDM	-	Species Distribution Model
SMI	-	Southern Montane Islands
SR	-	Species Richness
SSD	-	Sexual Size Dimorphism
SVL	-	Snout-vent Length
TD	-	Tympanum Diameter
THL	-	Thigh Length
TL	-	Tibia Length
UNESCO-MAB	-	United Nations Educational, Scientific and Cultural Organization Man and Biosphere
VU	-	Vulnerable
WE	-	Weighted Endemism

Elucidating cryptic diversity in East African frogs: the case of *Arthroleptis francei* Loveridge, 1953.

vi. Abstract

Detailed information on the diversity of species and their distributions is crucial for the implementation of useful conservation measures. Here, I focus on the Afromontane region of Malawi and Mozambique, and use molecular techniques, environmental inferences and species delimitation methods to clarify the phylogenetic position, existing phylogenetic diversity and distribution of the direct-developing leaf litter frog *Arthroleptis francei*, a species which is currently listed as Vulnerable (VU) by the IUCN. This study is based on already available as well as newly collected specimens (n = 52), and also serves to test wider biogeographic hypotheses for Afromontane isolates in Mozambique (Mount Mabu, Mount Socone, Mount Chiperone, Mount Inago, Mount Lico and Mount Namuli) and Malawi (Mount Mulanje). The derived phylogenetic trees were based on sequence data across one nuclear (*rag-1*, 629bp) and two mitochondrial (*12S* and *16S*, 352bp and 455bp, respectively) genes, and suggest the presence of several cryptic taxa linked to mountain clusters. The Bayesian analysis yielded putative species within the mitochondrial gene (*16S*). Mount Inago diverges from the larger clade with a posterior probability of 1. This larger clade can further be split into five smaller clades per mountain however, Mounts Lico and Chiperone form a clade dispite their vast geographic distance. Mount Mabu appears occupied by two separate taxa, which were found in sympatry within the same locality. The putative distinct taxa could however not be separated with morphometric means, or through colouration. A species distribution model based on environmental data suggests the possible presence of *A. francei* further atop the explored mountains, as well as on smaller outcrops nearby. Details of the first recorded call of *A. francei* are presented. I also combined the phylogeny of evolutionary lineages and geographic distribution of the *A. francei* complex to estimate the weighted endemism, phylogenetic endemism, phylogenetic diversity and species richness. Endemism was mainly average whilst species richness was high in seven hotspots. The newly acquired information contributes to a better understanding of the diversity within *A. francei and* will hopefully lead to a revision of conservation management practices and recommendations.

vii. Keywords:

Mozambique, Phylogenetics, Anuran, Sky Islands, *Arthroleptis francei*

1. Introduction

Here, I focus on the Afromontane region of Malawi and Mozambique, and use molecular techniques, environmental inferences and species delimitation methods to clarify the phylogenetic position, phylogenetic endemism, existing phylogenetic diversity and distribution of the direct-developing leaf litter frog *Arthroleptis francei*, a species which is currently listed as Vulnerable (VU) by the IUCN. The study is based on already available as well as newly collected specimens (n = 52), and also serves to test wider biogeographic hypotheses for Afromontane isolates in Mozambique (Mount Mabu, Mount Socone, Mount Chiperone, Mount Inago, Mount Lico and Mount Namuli) and Malawi (Mount Mulanje). The derived phylogenetic trees were based on sequence data across one nuclear (r*ag-1*, 629bp) and two mitochondrial (*12S* and *16S*, 352bp and 455bp, respectively) genes, and suggest the presence of several cryptic taxa linked to mountain clusters.

The Bayesian phylogenetic analysis, species delimitation and haplotype network analysis jointly suggest that *A. francei* comprises of cryptic taxa distributed across the inselbergs of Southern Malawi and Mozambique, including a case of a divergence at the same location, Mount Mabu. The species delimitation model identifies all locations as well supported lineages of distinct species except for the populations located in Mount Socone, Inago and Namuli. The molecular analyses uncovered that speciation events have likely occurred on each mountain. There is evidence for three allopatric speciation events between Mount Lico and Mount Chiperone, between Mount Inago, Mount Socone and Mount Namuli and between Mount Mulanje. Interestingly, the individuals on Mount Mabu indicate potential sympatric speciation as they were located in the same geographic coordinates.

The putative separate taxa could however not be separated by morphometric means, bioacoustically or through skin colouration. This is mainly due to the small sample sizes and is an area for further research. A species distribution model based on environmental data suggests the possible presence of *A. francei* further atop the explored mountains, as well as atop smaller outcrops on Mount Ribáuè, Mozambique and the Zomba mountain range in Malawi amongst others. Details of the first recorded call of *A. francei* are presented. The newly acquired information contributes to a better understanding of the diversity within *A. francei*, and will hopefully lead to a revision of conservation management practices and recommendations.

1.2. Paleogeography of Southern Africa

The continent of Africa is home to a diverse and rich array of vertebrate communities that have undergone substantial changes. During the Paleogene (ca. 66.0 – 23.03 ma) many species assemblages differed from those of the present day (Kappelman *et al.*, 2003) due to significant geological and climatic events. An understanding of the current biogeographical patterns of the African fauna can be revealed by an evaluation of species' evolutionary histories, and how this corresponds to shifting palaeoenvironmental conditions (Tolley, Chase and Forest, 2008). Evolutionary relationships allow us to reconstruct the biogeographic history of regions using either single lineages (such as the brevicipitid frogs in east Africa, Loader *et al.*, 2014), or entire species assemblages (for example those presented by Barratt *et al.* 2018).

There are some key events in the biogeographic history of the African continent that are believed to have had strong influences on the current biota. It is hypothesised that lineage diversification in southern African vertebrates were impacted by climatic shifts, particularly in the Pliocene era (5.3 – 2.6 ma; Bauer and Lamb, 2005; Lee-Thorp, Sponheimer and Luyt, 2007). Furthermore, the topography of Africa went through a major change in the late Neogene (23ma to 2.6ma), and during the Eocene-Oligocene (33.9 ma) the Eastern branch of the East African Rift System (EARS) began to form. This range starts in southern Ethiopia, intersecting the north of Kenya at the Turkana depression to the Aberdare range which is located to the south west and ending in the north of Tanzania where Mount Kilimanjaro is located (Partridge, 1997; Sepulchre *et al.*, 2006; Gordon *et al.*, 2012). The western branch of the EARS formed the central Tanganyika Basin in the middle to late Miocene era (10ma – 12ma) where more recently (5ma – 2ma) the Tanganyika and Malawian rifts formed. These geological processes impacted species distributions and the types of assemblages that diversified or became more or less prominent. The Eastern Arc (EA) and the Southern Rift mountains which run from the Taita Hills in Kenya to southern Tanzania (Gordon *et al.*, 2012) in East Africa are an important geological feature that formed and influenced the diversification of rainforest species which became restricted to these mountains – refuges for rainforest habitats. The area is recognised as a hotspot of high species richness – now called the Eastern Afromontane Biodiversity Hotspot (EABH). This region continued to

change from the middle to the late Miocene period (23ma to 5.3ma) from the Tanganyika basin to Mount Mulanje (3002m asl.) in the south of Malawi (Sepulchre *et al.*, 2006; Kaspar, Prömmel and Cubasch, 2010). Landscapes were evolving, and further mountains such as the ones located in northern Mozambique were formed. These isolated mountains extend eastwards and southwards from Mount Mulanje in Malawi and into Mozambique, such as Mounts Gorongosa, Mabu, Namuli, Chiperone amongst others, including the Mafinga mountains that border Zambia and Malawi (Gordon *et al.*, 2012). These have become significant areas for biodiversity. The EABH is formed of many different mountains of varying geological origin but have likely been connected at various times, linking mountains in Ethiopia, Kenya, Tanzania, Rwanda, Malawi, Mozambique and Zimbabwe at various times (Gordon *et al.*, 2012, Figure 1). It is also likely that the EABH has intermittently been connected to the coastal forests of East Africa during wetter phases where forests have expanded across mountains and to lower elevations, such as the coastal forests of East Africa.

These changes in the topography influenced the climatic regime of eastern and southern Africa (Kaspar, Prömmel and Cubasch, 2010), with forests that once dominated the whole East African landscape becoming mainly restricted to mountains and coastal regions (Griffiths, 2011). Global climate also influenced the distribution of forests in Africa. It was in the mid Miocene when forests started to shrink due to changes in precipitation resulting from global cooling (Lovett, Midgley and Barnard, 2005; Sepulchre *et al.,* 2006; Kissling *et al.*, 2012). The uplift of the Tibetan Plateau led to intensified monsoons, continental aridification and increased erosion also in Africa (Dupont-Nivet *et al.,* 2007; Gupta *et al.,* 2004). With the decline of forests, grasslands began to dominate the landscape, resulting in poor connectivity between the forests (Kissling *et al.,* 2012). More recently, humans have converted a large proportion of forests into croplands (Gordon *et al.*, 2012; Ryan, Berry and Joshi, 2014; Adole, Dash and Atkinson, 2018).

The climate and vegetation of mainland Southern Africa are predominantly influenced by oceanic currents: the Benguela current in the south west of Africa in the Atlantic Ocean, and the Agulhas current south east of Africa in the Indian Ocean (Figure1). South pole air circulation in the southern ocean coupled with the cool Benguela current drove the

aridification in the south west region during the Miocene (Neumann and Bamford, 2015). South Atlantic winter rainfall in the form of cyclones, prevailing winds with high pressure systems and seasonal shifts determine the climatic conditions in southern Africa. These climatic conditions move northwards in the winter months (Sciscio *et al.*, 2013), and the Great Escarpment (GE in Figure 1) and the warm Agulhas current cast a rain shadow over East Africa, influencing the moist environmental conditions (Neumann and Bamford, 2015). This indicates that temporal and spatial climate dynamics and landform evolution have played a large part of the distribution of both flora and fauna.

1.3. Biogeography and Speciation

Macroecologists and biogeographers particularly have an interest in partitioning geographical regions, large or small, into meaningful biological units. This interest dates back to A. R. Wallace more than 100 years ago, when he proposed the evaluation of the world's zoogeographical expanses (Rueda, Rodríguez and Hawkins, 2013). Early biogeographical regions were almost entirely based on the knowledge researchers had gained on species distributions, such as those proposed by Wallace. Current studies enhance past knowledge with latest techniques by information provided on digital databases coupled with various laboratory and statistical methods. Nevertheless, whichever method is used, an assemblage of species can still be created for a determined biogeographic area, and, more than likely, the species would share this distribution with others of similar nature (Carstensen *et al.*, 2013). Carstensen *et al*. (2013) refer to biogeographic regions as operational species pools which hold clues to the geographic distribution in response to biological or physiological influences.

The principal aim of biogeographic studies is to examine how the environment, and its changes, influences species diversification. For example, are closely related species found on two adjacent mountains the product of an ancient fragmentation of these two mountain areas or some other process? To address this question, we need to understand something about both the geology of the mountains (the "geography" in biogeography) and the timing of diversification/speciation (the "bio" in biogeography).

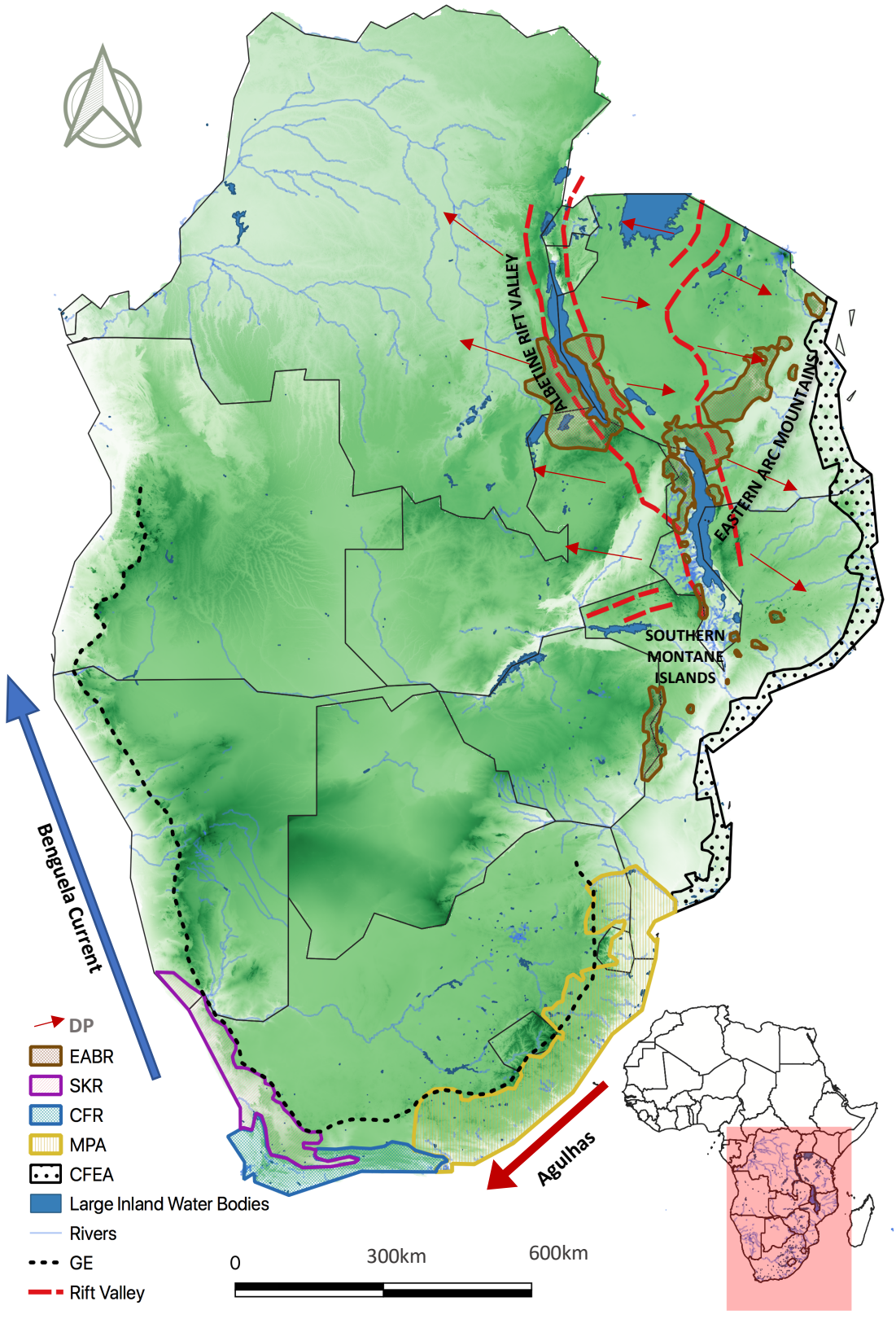


Figure 1. Biodiversity hotspots of sub-Saharan Africa: EABR – East Afromontane Biodiversity Region, SKR – Succulent Karoo Region, CFR – Cape Floristic Region, MPA – Maputa-Pondoland-Albany, CFEA -Coastal Forest of East Africa, GE – Great Escarpment. DP shows the direction of the plate movements to form the great rift valley which is partly shown on this map. Green shades depict elevation where darker is greater and lighter was lower. Adapted from Nielsen et al., *(2018).*

Therefore, biogeography is the coupling of geographic and biological processes. Often, however, knowledge of one set of data is used to infer something about the biological or geographical processes. Different geographic processes often produce predictable outcomes in the evolutionary patterns of organisms.

The fragmentation of once continuous habitat, caused by drying, climatic and vegetation changes, resulted in the creation of isolated refuges. This is referred to as the refuge hypothesis, a term coined by Edward Forbes in 1846 and later by Charles Darwin (Mayr and O'Hara, 1986). Analytical biogeography refers to traits of a species in response to the environment they inhabit, being behaviours, reproduction rates and dispersal methods. Ecological biogeography refers to how a species responds to the abiotic (physical) environment (soil, light, temperature, land formations, fire, water), and the biotic surroundings (parasites, predators, competition, disease, Huggett, 2004). During periods of biotic and abiotic alterations, these changes resulted in the evolution of taxa in refuges such as isolated forests that were surrounded by savannah, followed by forest expansions due to increased precipitation. Landscape changes are hypothesised to have driven vicariant evolution by fragmenting species distributions that were formerly continuous (Barratt *et al.,* 2018). Closely related species arise from natural barriers such as rivers as recognised through the 'river refuge hypothesis' by Wallace (1852) on birds (Mayr & O'Hara, 1986; Voelker *et al.,* 2013) and primates (Harcourt and Wood, 2012), or mountains (Park and Allaby, 2017; López *et al.*, 2018). The spatial partitioning of species across forested to non-forested habitat in light of the vanishing of the original habitat is known as the 'vanishing refuge hypothesis' (Vanzolini and Williams, 1981).

With these abiotic processes in mind, populations can become discontinuous, forming island-like, separated sub-populations becoming endemics to an area. Variations in phenotypes and genotypes are a result of these isolates and can lead to speciation. Speciation is the formation of new species through mechanisms that will maintain them and establish cohesive units of interbreeding individuals that retain a degree of individuality (Huggett, 2004). Various scenarios will result in different speciation models: sympatric, allopatric, stasipatric, parapatric and peripatric. Allopatric speciation is where a geographical barrier divides and isolates a population preventing gene flow. Should this

barrier persist long enough, the two new populations will evolve into two new species. Peripatric speciation is similar to allopatric speciation, yet with the exclusion of a geographic barrier. A small sub-set of populations disconnects from the original population, becoming isolated and evolving into a new species (e.g *Tanysiptera*, the Paradise kingfishers of New Guinea; Mayr, 1942). Parapatric speciation is a result of divergent evolution where two populations are living next to each other, and due to local adaptation natural selection eliminates between-population hybrids. Sympatric speciation occurs in a geographical area where species overlap yet natural selection encourages reproduction isolation.

Other land use changes have taken place due to anthropogenic factors, intensified agriculture practices, livestock (Newbold *et al.*, 2017), infrastructure growth, urbanisation (Smith., *et al*, 2018) deforestation (charcoal industry), mining (Edwards *et al.*, 2014) as well as natural processes (extreme weather and erosion, Sintayehu, 2018). Over large temporal frames, with the influence of climatic conditions and land use changes, the phylogenetic relationships of organisms that have diversified and dispersed can help in identifying historical connectivity (Blackburn, 2008b). Lineages from Africa are old, yet have successfully persevered alongside the evolving modern ones, which could be related to a number of biogeographic influences (Neumann and Bamford, 2015).

With such influences, species become endemic to a specific area which can therefore threaten their existence. Studies of for example endemic amphibian species within the Southern montane isolates are limited to *Nothophryne baylissi, N. inagoensis* and *N. ribauensis* (Conradie *et al.*, 2018a). However, both environmental and phylospatial data have not been considered to validate that they are endemic, only the genetic divergence and call differences between localities. To conclude with certainty, the endemism requires a broad multi-level analysis considering phylogenetic endemism and species richness. Species richness (SR) as measured at the level of biodiversity, can be supplemented with phylogenetic information, which is based on branch lengths which connect each tip to the root of an evolutionary tree, to obtain a measure of phylogenetic diversity (PD: Faith, 1992). Later, Rosauer *et al.* (2009) enhanced the use of PD by combining it with weighted endemism (WE: Crisp *et al.*, 2001) which results in phylogenetic endemism (PE) which indicates where considerable sections of PD are spatially restricted. Because cryptic

diversity can be incorporated in PE, this measure has gained momentum amongst macro ecologists and conservationists and studies within this area have increased (Rosauer and Jetz, 2015; Rosauer *et al.*, 2016; Barratt *et al.*, 2017). Usually areas of high PE are found where geographically restricted relatives are few in numbers and close together on a phylogenetic tree. These species are considered as important in a conservation setting because they indicate refugia where, over time, a large amount of evolutionary history has accumulated (Rosauer *et al.*, 2009; Rosauer and Jetz, 2015). Qualifying biodiversity loss is an urgent priority accompanied with explanations as to why some areas are richer in biodiversity than others, helping in prioritising conservation efforts. Species endemism and high biodiversity at global and regional scales have been shown to correlate with current and historical climatic regimes and topography (Kissling *et al.*, 2012), yet inclusive assessments are limited within biodiversity hotspots.

1.4. Biodiversity Hotspots

Biodiversity hotspots (BHs) were conceptualised to prioritise areas around the world where immediate attention is needed to alleviate biodiversity loss, and distribute funds accordingly. Myers *et al.* (2000) identified 25 biodiversity hotspots globally, as areas experiencing loss of habitat and having high concentrations of endemic plants and vertebrates. For the biodiversity hotspot assessment criteria, 70% or more of the original habitat must have been lost and 0.5% of the world's plant species, mainly consisting of vascular plants, must have been lost within an area. Vertebrates exclude fishes due to insufficient data, and do not have to meet a global total as they act as a backup support for the assessment.

Identifying hotspot areas aids in conservation planning to prevent the mass extinctions we are currently facing. This outlook aids in the prioritisation of locations where the greatest protection of species can be obtained for the amount of resources available. Hotspot boundaries across the globe have been determined by areas sharing 'biological commonalities' from the smallest (New Caledonia, 526.7 km^2) to the largest (Mesoamerica, 138 437 km^2) (Myers *et al.,* 2000). In Africa, one of the largest biodiversity hotspots is the Eastern Afromontane Hotspot EABR, Figure 1). This hotspot spans more than one million km^2, and comprises a) the Eastern Arc Mountains and Southern Rift, b) the Albertine Rift,

and c) the Ethiopian Highlands (Mittermeier *et al.*, 2004). Of the hotspots mentioned, the Eastern Arc Mountains are ranked within the eight hottest globally, in terms of number of both plant and vertebrate endemics and endemics per area to habitat loss. It is to the south of this region where the montane islands of southern Malawi and central Mozambique are found.

1.5. Vegetation of the Southern Montane Islands

Temperate, tropical and boreal forests are among the richest ecosystems in the world, covering an area of 30% of the land surface of the world of which 18% is found in Africa (FAO, 2016). This vital habitat provides home to around 80% of the global terrestrial animals and plants, yet only 50% of the African fauna is found here (Aerts and Honnay, 2011). Central Africa is characterised by 90% of forest cover, whereas other representative numbers are 6% in Western Africa, 2.2% on the island of Madagascar, and 2.4% in Eastern Africa (Malhi *et al.*, 2013). African savannah is a dominant vegetation type in such areas (Eardley, Gikungu and Schwarz, 2009), accounting for over half of the coverage of the continent (Adole, Dash and Atkinson, 2018), with forests having an estimated area of 675 million hectares, yet 7.4 million hectares have been lost since 1990 (FAO, 2016; Mucova *et al.,* 2018). With such rich and varied flora, the biodiversity in Africa is extremely diverse and home to more than 10% of fish, birds and plants, in addition to 6% of mammals and reptiles of the world (IUCN, 2019).

The topography changes from the low-lying coastal region of eastern Mozambique to the west on the Southern Montane Islands of Mozambique and southern Malawi. Much of the vegetation in the lowlands in this area has been shaped by extended and seasonal droughts, where hardy species form the iconic *miombo* woodlands dominate the landscape (e.g. *Brachystegia* spp., *Julbernardia* spp.). Within this generally flat landscape lie scattered inselbergs, which are “steep-sided isolated hills rising relatively abruptly above gently sloping ground” (Young, 1972). These inselbergs are covered with different vegetation types, such as grasslands, heathlands and evergreen forests (Timberlake *et al.,* 2007).

1.6. 'Sky Islands'

The granite inselbergs in southern Malawi and the Zambezia province of Mozambique, also called 'sky islands', reach heights of 3000m asl. at Mount Mulanje and 2500m asl. at Mount Namuli respectively. Many of these inselbergs are crowned with moist, evergreen forests (Bayliss *et al.*, 2014) surrounded by *miombo* woodland (Timberlake *et al.*, 2007). These inselbergs are a result from millions of years of uplifts due to tectonic events and subsequent erosion, surrounded by areas of flat, dry and predominantly subsistence farmed land (Timberlake *et al.*, 2007). They are commonly found at the southern point of Malawi near Mount Mulanje and extend eastwards into the west of Mozambique to form the Southern Montane Islands as seen in Figure 1. This region forms an important link between the Coastal Forests of East Africa and the Eastern Arc Mountains and falls within the Eastern Afromontane Biodiversity Hotspot.

Timberlake *et al.* (2007) compiled a list of vegetation on Mount Namuli, the second highest mountain in this region. Depending on the altitude and composition, these moist forests can be separated into three categories: montane forest, medium-altitude forest and riverine forest, each with canopy, understory and forest floor layers with a wooded shrubby transitional zone. A thick and slow degrading layer of leaf litter covers the forest floor. These delicate and rich habitats are the last remining evergreen forests in this region and are disappearing fast, mainly due to subsistence farming and illegal logging. Slash-and-burn practices undertaken by subsistence farmers, succession of native bracken (*Pteridium aquilinum*) on the barren land dominates after fires or after cultivated land has been left to naturally regenerate (personal observation). This leads to habitat loss and has a catastrophic effect on the biodiversity of this region.

1.7. Frogs of East Africa

The tetrapod vertebrate class Amphibia comprises three modern groups: salamanders (Caudata), frogs (Anura) and caecilians (Gymnophiona). According to the IUCN, amphibians are the most threatened vertebrates on earth, with 40% of species (predominantly terrestrial) classified as threatened with extinction (IUCN, 2019a). In Sub-Saharan Africa, 11.5% are for example jointly listed as threatened for Malawi and Mozambique. Overall threats on a global scale can be linked to habitat loss and fragmentation (Funk *et al.,* 2005;

Green, 2003), disease (Pounds *et al.,* 2006; Stuart *et al.,* 2004), and climate change (Courtois *et al.,* 2016; Pounds *et al.,* 2006; Stuart *et al.,* 2004). According to the IUCN (2019a), the threats to Sub-Saharan African amphibians are predominantly due to land use changes by agriculture practices, and the expansion of residential and commercial development (Tsinda *et al.*, 2016). The ranges of amphibians can be greatly reduced when their vital requirements become restricted, and fragmentation of their habitat has occurred (Cushman, 2006). Local extinctions will result from the reduction of a species' range within their distribution (Pocock *et al.*, 2006). The vulnerability of amphibians will depend on their response to occurring changes. This will involve habitat tracking of the required climatic conditions for survival, adaptation to new climatic conditions (i.e. niche evolution), phenotypic plasticity and an evolution of behaviours to maximise their existence (Sunday *et al.*, 2014; Pacifici *et al.*, 2015).

Frogs are not uniformly distributed across Africa, and have their highest species richness in lowland forest (n = 575 species), savannah habitats (n = 528 species) and montane forests (n = 467 species) (IUCN, 2019). These ecosystems consist of five macro-habitats for amphibians following du Preez & Carruthers (2017):

- Endorheic systems are depressions in the land such as pans, pools, and ponds filled with rainwater or run off, often depleted by evaporation or absorption.
- Riverine systems such as permanent rivers, dry riverbeds, flood plains, temporary streams, perennial streams and mountain torrents.
- Lacustrine systems are large bodies of water, predominantly permanent such as manmade dams and natural lakes.
- Palustrine systems are shallow marshland with a no greater depth than 2m, such as vleis (marshy, grass covered wetland), perched wetland and inundated grassland
- Terrestrial systems with no standing, permanent or semi-permanent water bodies such as a forest floor, rocky outcrops, sand dunes, open fynbos and open grasslands.

Amphibians generally exhibit three main different reproductive modes, with 85% of species being egg laying, 10% direct developers and 1% live bearing (Franco *et al.*, 2008). For direct development, reproduction takes place in terrestrial areas without direct access to water bodies (McDiarmid and Altig, 1999). Tadpole development is completed within the egg,

from which miniature forms of the adult emerge (Altig and Johnston, 1989). This breeding mode relies on moist environmental conditions to enable the survival of small clutches of eggs.

Despite the diversity and distribution of frogs in Africa, they remain poorly studied (Poynton, 1999; Frost *et al.*, 2006; Zimkus and Blackburn, 2008a). Between 2004 and 2019, 18 new species of anurans have been described for Mozambique and Malawi out of a total of 278 species being described in the whole of Africa (AmphibiaWeb, 2019). Of late, there has been an increase of sampling intensity which has led to the discovery of new species, such as *Tomopterna branchi* (Wilson and Channing, 2019), *Nothophryne baylissi, N. inagoensis, N. ribauensis, N. ulilurio* (Conradie *et al.*, 2018a), *Hyperolius stictus* (Conradie *et al.*, 2018b), *Breviceps passmorei* (Minter, Netherlands and Du Preez, 2017) and *Arthroleptella kogelbergensis* (Turner and Channing, 2017). These discoveries highlight the need for further research and an urgent need for the protection of these mountains. Habitats are disappearing before we have the chance of describing their biodiversity. The more we can understand them, the better evidence can be gathered for their conservation.

1.8. Conservation of Amphibians

Impacts on the earth's ecosystems and a decrease of biodiversity have been linked to the expansion of the human population, pollution, spread of invasive species, disease, climate change and overexploitation (Davies, 2015; Mimouni and Beisner, 2016). Current extinction rates are estimated to be several orders of magnitude higher than historical values, leading the natural world towards a sixth mass extinction (Plotnick *et al.,* 2016). However, due to a lack of taxonomic and distribution data for many species, precise extinction rates can currently only be approximated (the "Linnean and Wallacean shortfall", Brown and Lomolino, 1998; Mimouni & Beisner, 2016). Detailed information on species and their distributions are also crucial for assessing their conservation status as listed on the International Union for the Conservation of Nature (IUCN) Red List of Threatened Species (Brito, 2010). Of the 6035 worldwide anurans listed on the IUCN Red List, 1218 are data deficient (DD: information is insufficient to make a conservation status assessment, IUCN, 2018). It is important to improve on the data of these species, to broaden our knowledge on the true extent of the extinct and extant biodiversity (Tapley *et al.*, 2018).

1.9. Study species: Arthroleptis francei

One example of a lack of baseline knowledge for amphibians relates to taxonomic arrangements within the family Arthroleptidae, including the reclassification of the genera *Arthroleptis, Cardioglossa* and *Schoutedenella.* De Witte (1921) described *Schoutedenella* as a monotypic genus, being very similar to *Arthroleptis* but lacking maxillary teeth. Laurent (1954) disagreed with De Witte (1921) as individuals lacking teeth could be juveniles of the larger species. Frost *et al.* (2006) rejected previous suggestions by Laurent (1954), where the smaller species of *Arthroleptis* (snout-vent length [SVL] <25mm) were previously assigned to *Schoutedenella*, and those with a SVL greater than 25mm were placed within *Arthroleptis*. Laurent and Fabrezi (1985) stated that the small *Schoutedenella* are more closely related to *Cardioglossa* than to *Arthroleptis*, which was later rejected by Frost *et al.* (2006). Duellman (1993) grouped all *Schoutedenella* with *Arthroleptis* and altered their names, but again this was disregarded by Frost *et al.* (2006) as they synonymised *Schoutedenella* with *Arthroleptis* based on molecular evidence. Phylogenetic analysis based on molecular data presented by Frost *et al.* (2006) and Blackburn (2008b) suggest that *Schoutedenella* and *Arthroleptis* are monophyletic groups. Along with the discrepancies over the placement of *A. francei*, this species has undergone a number of name changes over the years. It was Loveridge (1953) who originally gave the name *Arthroleptis adolfifriederici francei*, yet later disregarded this stating it was merely related to the central African species, *A. adolfifriederici*. Laurent (1957) suggested the name *Abroscaphus adolfifriederici francei* by implication but it was Poynton and Broadley (1985) who refer to the species by the name known today as *Arthroleptis francei.* Despite some progress with the taxonomy of this group (Blackburn, 2008b; Tolley *et al.,* 2018), there are still some taxonomic conundrums, particularly within the smaller species of *Arthroleptis* (Blackburn, 2009a).

Arthroleptis, also known as African squeakers, is a genus of frogs widespread throughout sub-Saharan Africa, inhabiting a diverse array of habitats and currently comprising more than 40 species (Figure 2). In the last 30 years, new species have been discovered in central (Blackburn, Gvoždík, & Leaché, 2010; Zimkus & Larson, 2009), eastern (Blackburn, 2012; Blackburn, 2009a; Poynton, 2003), and western Africa (Ernst, Agyei and Rödel, 2008; Rödel

et al., 2009). These recent discoveries suggest that there may be further hidden diversity within this genus.

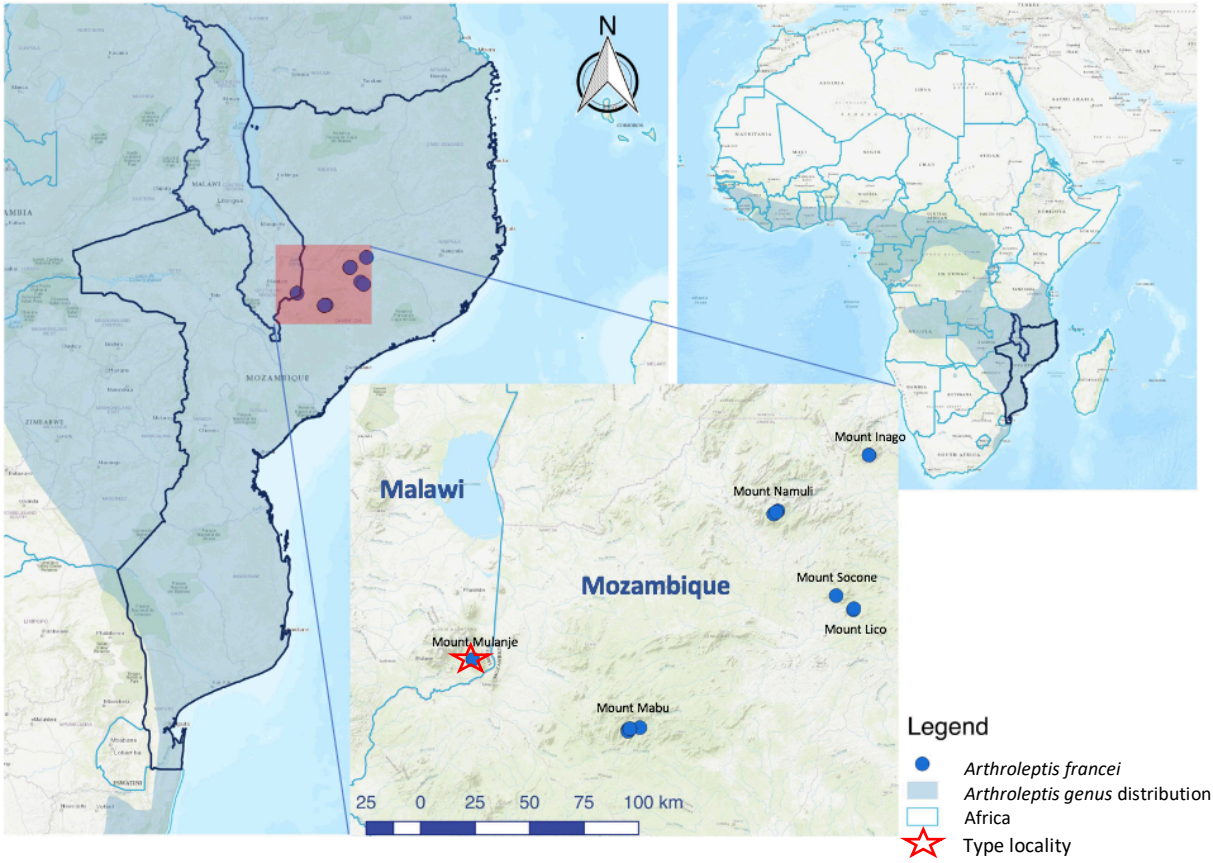


Figure 2. Distribution of the Arthroleptis francei *across Africa in the right inset. Distribution of* Arthroleptis francei *in Mozambique and Malawi. Mount Mulanje is the type locality where the type specimens were collected by Arthur Loveridge..*

Members of the *Arthroleptis* genus are terrestrial leaf litter frogs that feed on arthropods (Blackburn & Moreau, 2006; Loveridge, 1953). They are direct developers, which means that embryo development is completed within the egg and fully metamorphosed froglets emerge (Schweiger *et al.*, 2017). They are narrow range endemics, relatively small in size, live in moist forest environments and are known to be confined to the sky islands which provides delicate requirements for survival (Blackburn, 2009a). Snout-vent length (SVL) ranges from 16 to 54mm (Zimkus and

Figure 3. Leaf litter in which the specimens were collected on Mount Inago, Mozambique, 2018 expedition. The variation in colouration matching the phenotypic morphs.

Blackburn, 2008; right inset, Figure 2). Currently there are 48 species within this genus (Frost, 2019), from the largest, *Arthroleptis nikeae* with an SVL of up to 55mm, to one of the smallest, *Arthroleptis xenodactyloides* with a SVL of up to 22mm when fully grown (du Preez and Carruthers, 2017).

Up until very recently, *Arthroleptis francei* (Loveridge, 1953) was only known to exist on Mount Mulanje, a UNESCO-MAB Biosphere Reserve in Malawi and part of the EABH. Timberlake *et al.* (2009) reported the first record of *A. francei* outside its type locality, on Mount Namuli in northern Mozambique. Later, Conradie *et al.* (2016) recorded this species on Mount Mabu, also in northern Mozambique. Individuals putatively belonging to *A. francei* have further been found on other inselbergs in Mozambique (Bittencourt-Silva pers. comm. 2018; Conradie *et al.,* 2016), and preliminary molecular data suggests the existence of at least one new species within *A. francei* (Bittencourt-Silva pers. comm. 2018).

1.10. Aims and Objectives

Phylogenetic studies in Africa have proved challenging in the past due to its sheer size, political unrest (Timberlake *et al.*, 2007) and the inaccessibility of areas (Jongsma *et al.*, 2018). However, sampling efforts on African taxa increased within the last two decades, covering amphibians (Blackburn, 2008a; Conradie *et al.,* 2016; Jongsma *et al.,* 2018; Tolley et al., 2018), mammals (Neves, Da Luz Mathias and Bastos-Silveira, 2018; Van Berkel *et al.*, 2019), birds (Dowsett-Lemaire, 2010), reptiles (Ceccarelli *et al.*, 2014; Branch *et al.*, 2019) and plants (Van Noort, Gardiner and Tolley, 2007). Due to their low dispersal capacity and sensitivity to habitat and climate changes amphibians constitute a good model to study biogeographic patterns and climatic influences on their genetic structure (Zeisset and Beebee, 2008).

This study uses a multilocus dataset to understand the phylogenetic relationships, haplotype networks, species distribution modelling to construct the bio/phylogeographic history of the complex. I aim to assess:

1. The *A. francei* species complex and eliminate taxonomic uncertainty to reveal possible cryptic species within this taxon. I used two mitochondrial genes (*12S* and

16S ribosomal RNA genes) and one nuclear gene (*recombination activating gene*-1; *rag-1*) to reconstruct the phylogeny of *A. francei*. I investigated the cryptic diversity and the species distribution.

2. To understand where this species has persisted during the periods of climatic instability and changes in the physical environment. Taking into account the climatic oscillations and changes in the forest cover in eastern Africa since the Miocene, we hypothesise that climate and land use changes have had a major influence on the phylogeographic history of *Arthroleptis francei*. It is assumed that due to these habitat changes, this species has undergone various speciation events, are restricted and therefore endemic to each separate mountain forest.
3. Areas of endemism, species richness and phylogenetic diversity through climatic patterns and the phylogenetic analysis.
4. A revision in the conservation of *Arthroleptis francei*. By incorporating all these findings will help identify areas of importance and prioritise conservation efforts by focussing on areas with high endemism. This revision will also be a suggestion to the IUCN status of *A. francei* due to the findings of new lineages.

2. Methods

This study did go through ethical considerations and was classified as Type 1 under the University of Salford ethical procedures. Specimens included in this study were collected by the Natural History Museum London under a permit conceded by the Mozambican Government and the Museu de História Natural de Maputo.

2.1. Field Study

A total of 52 specimens of *Arthroleptis francei* were collected from Northern Mozambique and Southern Malawi between 2008 to 2018 (Figure 2) from the following expeditions: Mount Mulanje (Blackburn, 2008b), Mount Lico and Mount Socone (Bittencourt and Conradie, 2018), Mount Namuli and Mount Ribáuè (National Geographic Expedition arranged by SANBI, 2014) and finally Mount Inago and Mount Ribáuè by myself, alongside the Natural History Museum (NHM), South African National Biodiversity Institute (SANBI) and Pretoria Museum for two weeks in November to December 2018 (Figure 4 and 5).

Figure 4. Mount Inago, Zambezia province in Mozambique. One of the field sites where Arthroleptis francei were collected in November 2018 (Woest, 2019).

Figure 5. Mount Ribáuè, Zambezia province in Mozambique. Field site visited in November 2018 where males were heard calling but not located or collected (Woest, 2019).

The majority of individuals were found on montane forests atop granite rocky outcrops as seen in Figure 4 and 5. From the latest 2018 expedition, tissue samples were collected from livers of 13 male voucher specimens. These males were mostly found elevated <1m off the ground, perched on branches or leaves. Tissue samples from these and all other specimens from the previous expeditions are held at the NHM (Table 2).

2.2. Arthroleptis francei

The Ruo River screeching frog (*Arthroleptis francei*) is endemic to the southern montane sky islands of Malawi and Mozambique. This species is found in elevations ranging from 400m asl. to 1890m asl. (Table 1). Dorsal colour pattern varies from golden yellow to dark brown; the belly is white to light grey, mottled with darker grey; mature males have a darker throat. A distinctive black band running from the nostril and above the eye towards the forearm is prominent in all individuals. They are leaf litter dwellers but can be found elevated on lower understorey vegetation (<1m); fingers and toes lack webbing and the tips are slightly swollen; mature males develop a long third finger (Loveridge, 1957; Zimkus & Blackburn, 2008b). Measurements from Conradie *et al.* (2016) suggest that males are smaller than females, with a snout-urostyle length (SUL) of 22.85 mm and 31.10 mm, respectively. *Arthroleptis francei* is currently listed as Vulnerable (VU) on the IUCN Red List, and their population trend is decreasing (IUCN, 2019).

2.3. Holotype

Specimen number: MCZ 27479 (Figure 6).

A gravid female, *Arthroleptis adolfifriederici francei* was discovered below the Ruo River Falls, Mount Mulanje, Malawi around 1524m elevation by Arthur Loveridge, 4th April 1949 (Appendix 1. Loveridge, 1953). The species was renamed by Laurent (1957) as *Abroscaphus adolfifriderici francei,* is now known as *Arthroleptis francei*, Ruo River screeching frog (Poynton and Broadley, 1985). The name 'France' originates from a young forestry officer, Mr. F. H. France who passed away crossing the Ruo River, Mount Mulanje, Malawi in 1949 near to where the initial specimens were collected in 1949 (Loveridge, 1953).

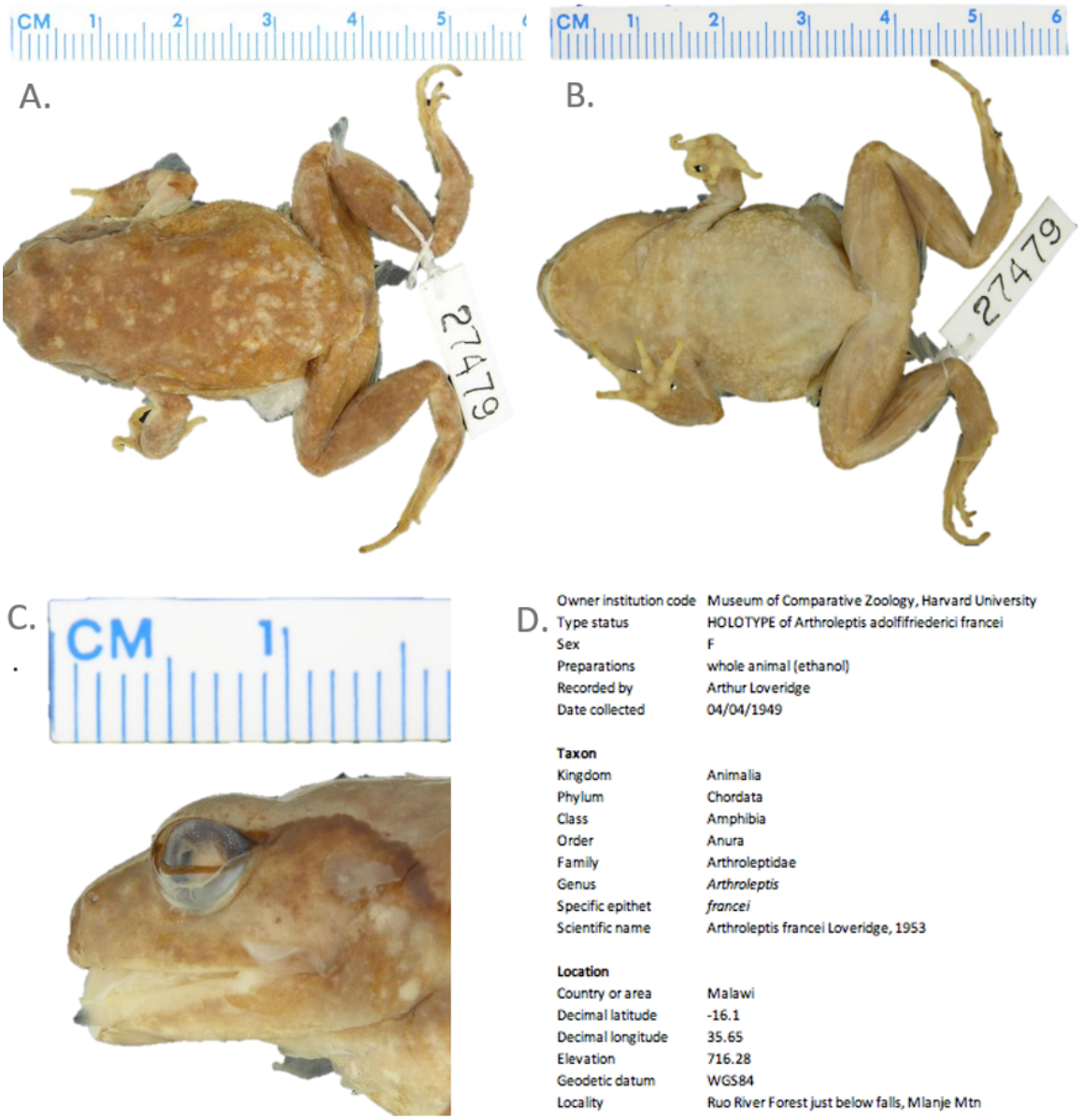


Figure 6. Photographs with size comparison and details of the gravid female holotype Arthroleptis francei (MCZ. 27479), collected by Arthur Loveridge in 1949. A. Dorsal view, C. Side view of the head, B. Lateral view, D. Details of the type specimen. Photos and data taken from gbif.org (2019).

2.4. Morphometrics and Sexual Dimorphism

Seventeen morphometric measurements were taken to compare populations from different mountains (largely following Napoli, 2006 and Tolosa *et al.* 2015): SVL – Snout-vent length, FL – Foot length, HW – Head width, THL – Thigh length, TL – Tibia length, TD – Tympanum diameter, ED – Eye diameter, HL – Head length, FOL – Forearm length, LHU – Length of humerus, IOD – Interorbital distance, IND – Internarial distance, END – Eye-nostril distance and DD3 – Forearm third digit disk width; to characterise possible differences within *A.*

francei, the length of lateral head band, the length of the inner-metatarsal tubercle and the third digit length on the forefoot were further included (Figure 7). Measurements were taken to the nearest 0.02mm three times for each characteristic with digital Vernier callipers (calibrated to 0.1) then averaged to reduce any errors. The SUL measurements from Conradie (2016) cannot be included in this analysis as here, the SVL was taken and eliminating them will allow standardisation.

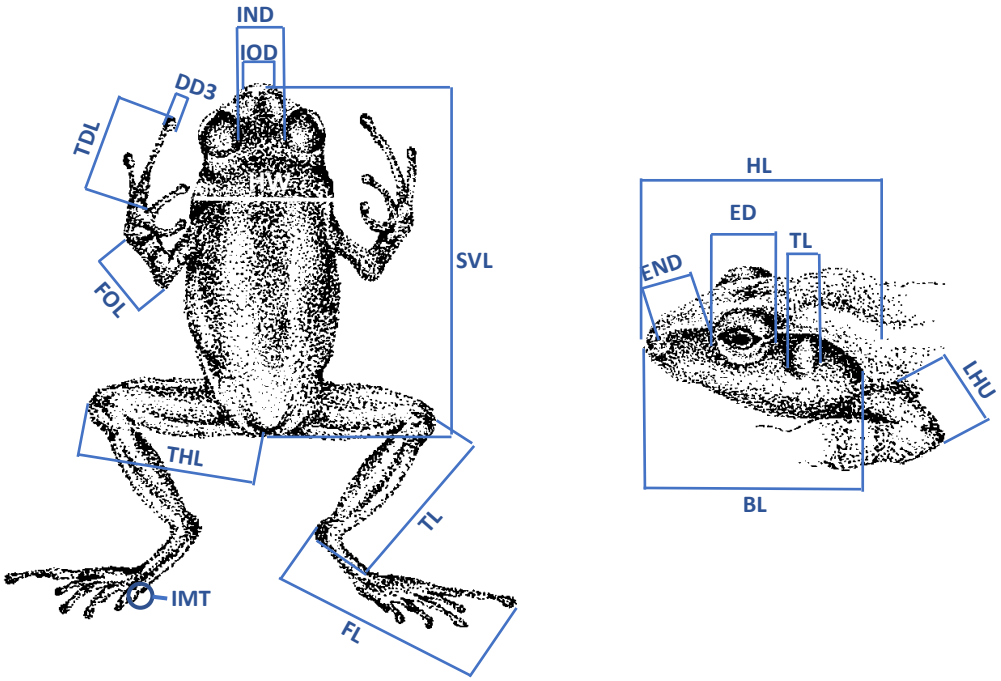


Figure 7. Morphometric measurements adapted from Tolosa et al., (2015). SVL – Snout-vent length, FL – Foot length, HW – Head width, THL – Thigh length, TL – Tibia length, TD – Tympanum diameter, ED – Eye diameter, HL – Head length, FOL – Forearm length, LHU – Length of humerus, IMT – Inner metatarsal length, TDL – Third digit length, BL – Band length, DD3 – Third digit diameter, END - Eye nose distance, IND – Inner nostril distance and IOD – Inner eye diameter. Illustrations and annotations by Natasha Woest (2019).

Due to logistical and timing constraints of accessing the fifty-two specimens, I made the decision to randomly measure representatives from each mountain within the small collection held at the NHM. This limited morphological data results in a basic graphical illustration, however they do give an introduction to a further understanding the species and direction for research. No statistical analysis was performed due to the small sample size (n = 14) but measurements were taken for an initial idea of sizing. Individuals measured are highlighted in grey in Table 2. Sexes were identified by call (males only), throat colouration (darker on males), presence of enlarged third finger (mature males) and size (females are generally larger). The morphological measurements were plotted in R Studio

using the ggplot2 package (Wickham, 2016) for a stacked box and whisker graph. Each individual graph represents the different feature measured, displaying the collective measurements separated by sex or unknown sex per mountain.

2.5. DNA Extractions and Polymerase Chain Reactions (PCRs)

DNA was extracted using a Qiagen DNeasy Blood and Tissue Kit, following the protocol for purification of DNA from animal tissue set out by the manufacturer. Tissue samples in this study based on mostly liver were derived from 52 individuals stored at the Natural History Museum London (Table 2). The choice of the target loci was due to the already available data on GenBank and widely used on amphibian phylogenetics in Africa. Two partial mitochondrial rRNA (*12S*) and rRNA (*16S*) and one nuclear *recombination activating gene*-1 (rag-*1*) were amplified (primers shown in Table 1). Both the mitochondrial genes were sequenced in a forward direction whilst *rag-1* was sequenced in both forward and reverse directions (Table 1).

Table 1. Primer used to amplify the sequence 12S, 16S *mitochondrial genes and a* rag-1 *nuclear gene. Bp – Base pairs*

Gene	Direction	Target	Length (bp)	Source Reference
***12S* L1091**	Forward	AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT	400	(Kocher *et al.*, 1989)
***12S* H1478**	Reverse	TGACTGCAGAGGGTGACGGGCGGTGTGT	400	(Kocher *et al.*, 1989)
***16S*AF**	Forward	CGCCTGTTTATCAAAAACAT	600	(Palumbi *et al.*, 1991)
***16S*AR**	Reverse	CCGGTCTGAACTCAGATCACGT	600	(Palumbi *et al.*, 1991)
rag-1	Forward	AGCTGCAGYCARTAYCAYAARATGTA	1500	(San Mauro, *et al.,* 2004)
rag-1	Reverse	AACTCAGCTGCATTKCCAATRTCA	1500	(San Mauro, *et al.*, 2004)

Polymerase chain reactions were carried out using illustra™ puReTaq Ready-To-Go™ PCR Beads (GE Healthcare UK Limited) using the following conditions for *12S* and *16S*: single cycle for 5min denaturation at 95°C followed by 1 min at 95°C, 1 min annealing at 51°C and a 1.5 min extension at 72°C. This was followed by 35 cycles with a 1 min denaturation at 95°C. A 7min extension at 72°C finishing off with 4°C until removed from the machine (Goebel,

Donnelly and Atz, 1999). The same conditions were used for rag-*1* with only the annealing temperature set to 54°C (San Mauro *et al.*, 2004).

PCR products were run on 1% agarose gels made from 100ml TAE buffer and 1g Agarose powder, which was microwaved until boiling, cooled and placed in the relevant tray with the relevant comb to create the wells. The gel was loaded with 2µM of DNA sample and 4µM of blue dye (bromophenol blue, xylen cyanol, glycerol, dH_2O and 2 gel red) per well. Electrophoresis was performed using a VWR, Powersource 250V set at 100V, Amp 19, 19W to draw the sample from the negative to the positive set on a timer for 30 minutes.

2.6. Data analysis and alignment

This dataset included an outgroup specimen represented by the close relative *Arthroleptis reichei* (specimen voucher MCZ:HERP:A-138365): *rag-1* (GenBank accession number MH744350) and the concatenated *12S* and *16S* genes (GenBank accession number FJ151151) from Blackburn (2008b). There was an inclusion of one additional *A. francei* from GenBank due to the availability of all three loci: Accession number MH744349, MCZ:HERP:A-137038 (*rag-1*), and Accession number FJ151100.1, MCZ:HERP:A-137038 (*12S* and *16S*) (Blackburn, 2008b). The total number of sequences that were included in this analysis per marker was 47 (*12S*: 352bp), 49 (*16S*: 455bp) and 47 (*rag-1*: 629bp) as summarised in Table 2.

Sequences were aligned in Geneious version 7.0.6 (Kearse., *et al.,* 2012) using MAFFT (Multiple Analysis Fast Fourier Transform) v7.017 (Katoh, Kuma & Miyata, 2002) with default settings. First, alignments were checked by eye by two people independantly for a thorough investigation and for obvious mistakes which were corrected manually, then ambiguously aligned blocks were removed using GBlocks (Talavera & Castresana, 2007) set to allow gaps in the final blocks within the non-coding mitochondrial sequences (*12S* and *16S*). For *rag-1*, TranslatorX was used to find the correct reading frame and resolve translation ambiguities when possible (Abascal, Zardoya & Telford, 2010).

Table 2. Taxa samples used for genetic analyses. DNA vouchers with "T" refer to unpublished sequences from the DNA collection from the Natural History Museum (London); others are from GenBank '-' = Missing information, "Y" = used sample. Grey highlighted specimen vouchers numbers refer to those included in the morphometric analysis.

DNA voucher	Specimen voucher	Species	Latitude	Longitude	Elevation (m) asl.	Mountain	Country	*12S*	*16S*	*rag-1*
MH744349	MCZ A-137038	*A.francei*	-16.00028	35.72417	1710	Mount Mulanje	Malawi			Y
FJ151100.1	MCZ A-137038	*A.francei*	-16.00028	35.72417	1710	Mount Mulanje	Malawi	Y	Y	
T0762	M 415	*A.francei*	-16.00028	35.72417	1710	Mount Mulanje	Malawi	Y	Y	Y
T0787	M 101	*A.francei*	-16.00028	35.72417	1710	Mount Mulanje	Malawi	Y	-	-
T0800	M 310	*A.francei*	-16.00028	35.72417	1710	Mount Mulanje	Malawi	Y	-	-
T0802	M 406	*A.francei*	-16.00028	35.72417	1710	Mount Mulanje	Malawi	Y	-	-
T5737	WC 3232	*A.francei*	-16.28622	36.40006	1892	Mount Namuli	Mozambique	Y	Y	Y
T5738	WC 3233	*A.francei*	-15.38986	37.03019	1779	Mount Namuli	Mozambique	Y	Y	Y
T5739	WC 3235	*A.francei*	-15.38986	37.03019	1779	Mount Namuli	Mozambique	Y	Y	Y
T5740	WC 3236	*A.francei*	-15.38986	37.03019	1779	Mount Namuli	Mozambique	Y	Y	Y
T5741	WC 3244	*A.francei*	-15.39797	37.01978	1611	Mount Namuli	Mozambique	Y	Y	Y
T5742	WC 3247	*A.francei*	-15.39797	37.01978	1611	Mount Namuli	Mozambique	Y	Y	Y
T5743	WC 3354	*A.francei*	-15.38986	37.03019	1779	Mount Namuli	Mozambique	Y	Y	Y
T5744	WC 3364	*A.francei*	-15.39797	37.01978	1611	Mount Namuli	Mozambique	Y	Y	Y
T5747	WC 3073	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5748	WC 3074	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5749	WC 3075	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5750	WC 3076	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	-	-
T5751	WC 3110	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5752	WC 3111	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5753	WC 3112	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5754	WC 3138	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5755	WC 3139	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5756	WC 3140	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5757	WC 3141	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5760	WC 3077	*A.francei*	-16.28910	36.39250	1247	Mount Mabu	Mozambique	Y	Y	Y
T5761	WC 3083	*A.francei*	-16.28153	36.44378	430	Mount Mabu	Mozambique	Y	Y	Y
T5763	WC 3174	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5764	WC 3175	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5765	WC 3176	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5766	WC 3177	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5767	WC 3182	*A.francei*	-16.29682	36.39243	1644	Mount Mabu	Mozambique	Y	Y	Y
T5768	WC 3183	*A.francei*	-16.29682	36.39243	1644	Mount Mabu	Mozambique	Y	Y	Y
T5769	WC 3184	*A.francei*	-16.29682	36.39243	1644	Mount Mabu	Mozambique	Y	Y	Y
T5770	WC 3185	*A.francei*	-16.29682	36.39243	1644	Mount Mabu	Mozambique	Y	Y	Y
T5772	WC 3187	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5773	WC 3188	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5774	WC 3220	*A.francei*	-16.28622	36.40006	919	Mount Mabu	Mozambique	Y	Y	Y
T5775	WC 3245	*A.francei*	-15.39797	37.01978	1611	Mount Namuli	Mozambique	Y	Y	Y
T5776	WC 3246	*A.francei*	-15.39797	37.01978	1611	Mount Namuli	Mozambique	Y	Y	Y
T5777	WC 3248	*A.francei*	-15.39797	37.01978	1611	Mount Namuli	Mozambique	Y	Y	Y
T5778	WC 3249	*A.francei*	-15.39797	37.01978	1611	Mount Namuli	Mozambique	Y	Y	Y
T5779	WC 3355	*A.francei*	-15.38986	37.03019	1779	Mount Namuli	Mozambique	Y	Y	Y
T6836	MOZ17-051	*A.francei*	-16.50801	35.73143	997	Mount Chiperone	Mozambique	-	Y	Y
T7132	WC 6415	*A.francei*	-15.73623	37.28815	1159	Mount Socone	Mozambique	Y	Y	Y
T7133	WC 6485	*A.francei*	-15.79470	37.36216	844	Mount Lico	Mozambique	Y	Y	Y
T7166	WC 6453	*A.francei*	-15.78978	37.36457	1039	Mount Lico	Mozambique	Y	Y	Y
T7376	MOZ17-428	*A.francei*	-15.15340	37.42950	1291	Mount Inago	Mozambique	Y	Y	Y
T7377	MOZ17-429	*A.francei*	-15.15340	37.42950	1291	Mount Inago	Mozambique	-	Y	Y
T7378	MOZ17-430	*A.francei*	-15.15340	37.42950	1291	Mount Inago	Mozambique	-	Y	Y
T7379	MOZ17-431	*A.francei*	-15.15340	37.42950	1291	Mount Inago	Mozambique	-	Y	Y
T7381	MOZ17-438	*A.francei*	-15.15340	37.42950	1291	Mount Inago	Mozambique	-	Y	Y
Outgroup										
FJ151151	MCZ A-138365	*A.reichei*	-8.19770	35.56399	1950	Udzunga Mountains	Tanzania	Y	Y	
MH744340	MCZ A-138365	*A.reichei*	-8.19770	35.56399	1950	Udzunga Mountains	Tanzania			Y
								352bp	455bp	629bp

2.7. Bayesian Phylogenetic Analysis

For the phylogenetic relationship analysis, mitochondrial and nuclear genes were gathered with the use of Bayesian inference (BI) approaches. PartitionFinder (Lanfear *et al.,* 2016) was used applying the Bayesian Information Criterion to select best partitioning schemes and models of evolution for each gene. The following models were applied to each partition: GTR+I+G for *12S* and *16S*, and HKY+G for *rag-1*. Phylogenetic trees were reconstructed for each gene and for a concatenated alignment (*12S*, *16S* and *rag-1*) using MrBayes v3.2

(Ronquist and Huelsenbeck, 2003). Two runs were executed (using four chains) for 50 million generations with trees sampled every 10000 generations with the initial 25% discarded as burn-in. Final trees were built in MrBayes and visualised using FigTree v.1.4.4 (Rambaut, 2018). TRACER v1.7.1 (Rambaut *et al.*, 2018) calculated the model convergence and 95% of the highest posterior density (HPD) intervals.

Evolutionary analyses of the distance estimations were conducted in MEGA X (Kumar *et al*., (2018). The number of base differences per site from between sequences were calculated using MEGA X software for MacOS (Stecher *et al.,* 2020). Standard error estimate(s) were obtained by a bootstrap procedure of 500 replicates. This analysis involved 49 nucleotide sequences for the substitution *p*-distance model including d: transitions and transversions. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 437 positions in the final dataset. During this comparison, the number of changes of base changes and the insertion and deletion events are tallied and displayed as a proportion of the overall sequence length.

2.8. Species delimitation

Possible cryptic species were explored within the samples of *Arthroleptis 16S* gene and analysed by using a Bayesian implementation of the General Mixed Yule-Coalescent model (bGMYC, Reid and Carstens, 2012). This gene was chosen due to it being the most widely used and universal DNA barcoding marker in amphibians (Vences *et al*., 2005). The data were initially prepared in BEAUTi v 1.10.3 (Drummond, Rambaut and Marc, 2016) using the HKY model (Hasegawa, Kishino and Yano, 1985) with estimated base frequencies was selected with an uncorrelated relaxed clock. Tree priors were set to a random starting Speciation Yule Process tree (Yule, 1925; Gernhard, 2008). The MCMC chain length was set to 100 000 000, with the echo state and log parameters set to 10000. An ultra-metric tree with Yule and lognormal relaxed molecular clock (Drummond *et al.*, 2006) priors was selected for a species level analysis in BEAST v 1.10.3 (Drummond, Rambaut, & Marc, 2016). The bGMYC model produced in BEAST was expressed in RStudio v 1.2.1335 (RStudioTeam, 2019) where 100 trees were randomly selected with a 10% burn in rate discarded and 5000 bGMYC generation simulations sampling every 100th generation using the APE (Paradis, Claude and Strimmer, 2004) and bGMYC (Noah, 2014) packages (Appendix 4).

2.9. Distribution network

Distribution networks, also known as haplotype networks, were used in the analysis and a visualisation of population genetic data at an intraspecific and interspecific level. These networks can also help understand the biogeographic distribution of populations (Leigh and Bryant, 2015). The software Population Analysis with Reticulated Trees (PopART) was used (Leigh and Bryant, 2015) with the Templeton approach (Templeton, Crandall and Sing, 1992). The TCS network package uses an algorithm where clusters are progressively combined with one or more connecting edges. TCS was used for estimating the gene genealogies which was selected for the visualisation for all three genes separately: *12S*, *16S* and rag-*1* (Clement *et al.,* 2002) with the Nexus file produced in Geneious with the addition of the longitude and latitude of each taxa added. The inbuilt statistical package was used for a basic analysis for the Analysis of Molecular Variation (AMOVA, Excoffier, Smouse and Quattro, 1992). Geographical locations of each individual were used with the k-means algorithm to cluster the sequences according to the six mountain clusters for *12S* and *rag-1* genes, and seven for *16S* gene, Mount Chiperone. Number of samples per cluster are determined by the circle size. The sister group was excluded from the distribution network as this is purely to display the *A. francei* and candidate species.

2.10. Species Distribution Models (SDM)

Species distribution models are also referred to as Ecological Niche Models (ENM). Occurrence co-ordinates were used from the field data stored at the NHM, only for the *A. francei* used in the phylogenetic analysis. Projection for the species distribution against the 19 different climatic conditions (Current: the average for 1960 – 1990) were downloaded from WorldClim (http://www.worldclim.org/) at a spatial resolution of 2.5 arc minutes/km. Bioclim variables are widely used in ENM, where Bio 1 – 11 reflect temperature and Bio 12 – 19 reflect precipitation systems (Booth *et al.*, 2014).

The R Studio packages Raster, Maps and MapData were used to prepare the BioClim data for the following ten variables: BIO1 – Annual mean temperature, BIO2 - Mean diurnal range (mean of monthly maximum temperature − minimum temperature), BIO3 – Isothermy (mean diurnal range temperature range), BIO7 – Temperature annual range, BIO8 – Temperature of the wettest quarter, BIO11 - Mean temperature of the coldest quarter,

BIO12 – Annual Precipitation, BIO15 – Precipitation seasonality, BIO18 – Precipitation of the warmest quarter and BIO19 – Precipitation of the coldest quarter. These variables are identical to a previous study on the closely related *A. whalbergii* (Tolley *et al.*, 2018).

Layers were cut to the focal area in R Studio and standardised to the same resolution, using a mask fitted to the species distribution regions co-ordinates (xmin - 34.72417, xmax - 38.4295, ymin - -17.29682, ymax – -14.1534) set to WGS 1984. The output format for each file was set to ASCII (ESRI .asc format) to fit the requirements of the Maximum Entropy Modelling (MaxEnt). The software MaxEnt 3.4.0 (Dec 2016) was used to calculate the SDMs, using a machine-learning algorithm based on maximum entropy principles (Phillips, Dudík and Schapire, 2019). This approach is suggested to be the most appropriate for modelling presence data (Hijmans and Elith, 2017). The output format was set to logistic, ASC, with a random subset test percentage set to 20 of the occurrences and used the remaining localities for testing at 100 replicates. MaxEnt generates the probability distribution in a pixelated grid format from a uniform distribution and repeating the runs to improve the fit. Output files were visualised and formatted using QGIS (QGIS, 2019).

2.11. Phylospatial distribution of Arthroleptis francei

Here, phylospatial distributions, a term used by Rosauer (2020) incorporates spatial and phylogenetic data from this study which were integrated to investigate phylogenetic endemism (PE), phylogenetic diversity (PD), weighted endemism (WE) and species richness (SR) across the mountain range to the south of Malawi and into Mozambique. Statistical scrips were provided by Rosauer *et al.* (2015 and 2020) and adapted to suit the *A. francei* data within this study and later used to estimate PE, PD, WE and SR (Appendix 9). Firstly, a batch multi method was used to break down the SDM into constituent lineage models for the six cryptic species from the *16S* BI analysis. Branches from the type localities (T0762, GenB 1 and 2 were collapsed as they were identical, yet other branches were left as separate tips to enable a thorough investigation of each individual. These clades referred to as Mount Mabu 1 – M1, Mount Mabu 2 – M2, Mount Mulanje – MUL, Mounts Namuli, Inago and Socone – NIS, Mount Chiperone – CHI and Mount Lico – LIC. The SDM was consequently separated into linear distribution models (LDM) by geological data and known geo-referenced genetic data (Rosauer and Jetz, 2015). The likelihood of species occurrence for

each grid cell and the lineage data were kept as continuous which were later used to estimate the PE, PD, WE and SR. This data was extracted from the spatial occurrences, climatic data from the SDM model and the tree tip lengths from the mitochondrial *16S* phylogeny (Figure 13) and were used at a 1km^2 resolution to align the species distribution. Phylogeny branch lengths were partitioned into grid cells of occurrences based on the SDM and scored accordingly (0 and 1 for all but with PD being scored between 0 and 2). Maps were produced for each output displaying the PE, PD, WE and SR per grid cell where each cell corresponds to the branches on the *16S* phylogeny and analysed across the region as presented in the SDM model.

2.12. Conservation

To identify the diversity within a specific geographic location can be summed by the total PE, a measure of the spatial range of each branch on the phylogeny. Areas with medium to high PE results were retained and used to intersect the shapefiles of protected areas provided by the Critical Ecosystem Partnership Fund (CEPF: http://www.cepf.net) and the World Resources Institute (WRI: http://www.wri.org). This then helps to visualise whether areas harbouring evolutionary history of *A. francei* and the candidate species are protected.

2.13. Call analysis

As taxonomic complexities are prevalent in *Arthroleptis*, leading to difficulties distinguishing between species on morphologies in the field, calls could prove useful. A detailed analysis of advertisement calls on a species basis could help in the clarification of any taxonomic uncertainties (Köhler *et al.*, 2005). Unfortunately, only a single male courtship call was recorded whilst at the study site on Mount Inago using a Samsung Galaxy S9 (SM-G960F) mobile phone. The call was later visualised with Luscinia 2.25.03.15.01 (Lachlan, 2007) and analysed using Praat 6.1 (Boersma and Weenink, 2019). This is the only known call to have been recorded of *A. francei* to date. Upon approaching the individuals, they terminated calling, even when waiting in silence and all lighting was turned off. Due to only one call recording with a duration of a minute, only standard measurements can be taken: call intensity (dB), call frequency (Hz), call times and syllables per call.

3. Results

3.1. Morphometrics

Morphological measurements were taken from fourteen individuals from five out of seven mountains: Mounts Mabu, Lico, Inago, Namuli and Socone. Specimens were not available for morphometric analysis from Mount Mulanje and Mount Chiperone. According to the museum tags, five specimens were male and female each, and four specimens were listed as unknown. Due to colouration and size, they are presumed male but were excluded in the sexual dimorphism analysis. As the sample size was too small to perform statistical tests, Table 3 displays the measurements and the difference in size by percentage between males and females. Females were larger than males overall (SVL = 30.7%, HW = 40.3%, TL = 31.2%, END = 38.4%, IMT = 34.8% and BL = 33.1%). The third digit length (TDL) is 3.7% larger in females, but in proportion to the body size (SVL) males' TDL is longer (24.5%) compared to females (17.6%). Band lengths (BL) are similar in proportion to the snout-vent-length (SVL) between the sexes, where females averaged 44.52% and males averaged 43.13%.

Only one female was measured from Mount Namuli and four from Mount Mabu, limiting any comparisons between localities. Males were sampled from three mountains (Table 3). It is again evident that females are larger than males in all their features, except for the third digit length (TDL) which was longer in males (Figure 3 and 8). Forearm length (FOL), length of humerus (LHU) and thigh length (THL) are all very similar in size between the sexes, suggesting that females are stocky in built due to shorter limbs, in comparison to their body size. Variation in size is greatest on Mabu, yet this could be due to more specimens measured and collected from this mountain (Figure 8).

Table 3. Morphological features (first column) of Arthroleptis francei *were measured in millimetres (mm). For each sex, the measurements were standardised by displaying the mean and standard deviation (Mean±STD). Individuals were chosen based on having all three genes and random representatives from each mountain. All specimens could not be measured due to logistics and time constraints. Mount Chiperone and Mount Mulanje are missing as these specimens are not held at the NHM. Abbreviations for each feature are explained in Section 4.3. Sex of individuals were stated: ♂ = Males, ♀ = Female and those unknown are left blank.*

Specimen	T5773	T5774	T5743	T5747	T5768	**Mean±STD**	T5757	T7377	T7379	T7132	T5766	**Mean±STD**	**Difference**	T5763	T5765	T7133	T7166	**Mean±STD**
Mountain	Mabu	Mabu	Namuli	Mabu	Mabu		Mabu	Inago	Inago	Socone	Mabu			Mabu	Mabu	Lico	Lico	
	♀	♀	♀	♀	♀		♂	♂	♂	♂	♂		♀ vs ♂					
SVL	32.82	32.36	36.36	35.04	31.83	**33.68±1.93**	22.01	25.71	23.08	21.86	24.04	**23.34±1.59**	**10.34**	26.45	24.94	23.77	24.17	**24.83±1.183**
FL	20.52	19.99	25.40	23.03	21.57	**22.1±2.18**	14.30	16.49	15.23	15.44	21.96	**16.68±3.05**	**5.42**	17.32	19.79	17.62	16.82	**17.89±1.31**
HW	15.09	14.36	16.74	16.25	14.88	**15.46±0.99**	9.34	10.55	8.29	8.64	9.32	**9.23±0.87**	**6.24**	9.69	10.49	10.00	9.54	**9.93±0.42**
THL	16.30	13.48	16.12	14.43	14.13	**14.89±1.25**	10.43	11.49	14.75	10.86	13.90	**12.29±1.92**	**2.61**	11.29	13.80	12.60	12.23	**12.48±1.04**
TL	15.82	15.88	16.56	20.86	14.69	**16.76±2.39**	10.69	11.27	10.44	10.84	14.42	**11.53±1.64**	**5.23**	12.68	14.14	12.95	11.87	**12.91±0.94**
TD	2.63	2.42	2.33	2.06	1.91	**2.27±0.29**	2.13	1.86	1.85	1.23	1.99	**1.81±0.35**	**0.46**	2.39	1.82	1.82	1.62	**1.91±0.33**
ED	3.92	3.84	4.02	3.93	3.24	**3.79±0.31**	2.83	3.40	3.00	2.72	3.29	**3.05±0.29**	**0.74**	3.72	3.75	3.37	3.04	**3.47±0.34**
HL	11.22	12.40	12.32	12.60	11.68	**12.04±0.58**	8.23	9.31	8.75	6.88	9.82	**8.598±1.13**	**3.45**	10.48	9.50	8.84	9.14	**9.49±0.71**
FOL	8.31	7.08	9.20	9.26	7.62	**8.29±0.96**	6.57	7.45	6.80	5.72	7.50	**6.81±0.73**	**1.49**	8.50	6.58	7.25	6.94	**7.32±0.84**
LHU	7.38	6.02	8.57	8.19	6.23	**7.28±1.14**	4.92	5.21	4.55	3.99	7.29	**5.19±1.26**	**2.09**	7.17	5.23	5.81	5.37	**5.9±0.89**
IOD	4.31	3.23	3.64	4.22	2.72	**3.62±0.67**	2.48	2.71	2.75	2.71	2.37	**2.6±0.17**	**1.02**	2.90	2.96	3.12	3.15	**3.03±0.12**
IND	3.79	2.89	3.64	3.40	3.17	**3.38±0.36**	2.39	2.67	2.29	2.42	2.73	**2.5±0.19**	**0.88**	2.77	2.97	2.54	2.55	**2.71±0.21**
END	2.86	2.85	3.25	3.02	3.13	**3.02±0.17**	2.05	2.38	2.01	0.52	2.36	**1.86±0.77**	**1.16**	1.82	2.26	2.78	2.71	**2.4±0.45**
DD3	0.84	0.75	0.64	0.92	0.74	**0.78±0.11**	0.66	0.73	0.48	0.53	0.55	**0.59±0.10**	**0.19**	0.60	0.43	0.35	0.55	**0.48±0.11**
BL	13.97	13.96	16.26	16.62	14.24	**15.01±1.32**	9.97	10.03	9.58	9.70	10.93	**10.04±0.53**	**4.97**	11.86	10.93	9.71	9.75	**10.56±1.03**
TDL	5.77	4.57	6.65	6.20	6.45	**5.93±0.83**	6.47	6.49	4.32	4.58	6.69	**5.71±1.16**	**0.22**	7.79	5.20	5.03	5.41	**5.86±1.30**
IMT	1.82	2.02	2.42	2.29	1.82	**2.07±0.27**	1.31	1.47	1.27	1.19	1.53	**1.35±0.14**	**0.72**	1.51	1.47	1.29	1.12	**1.35±0.18**

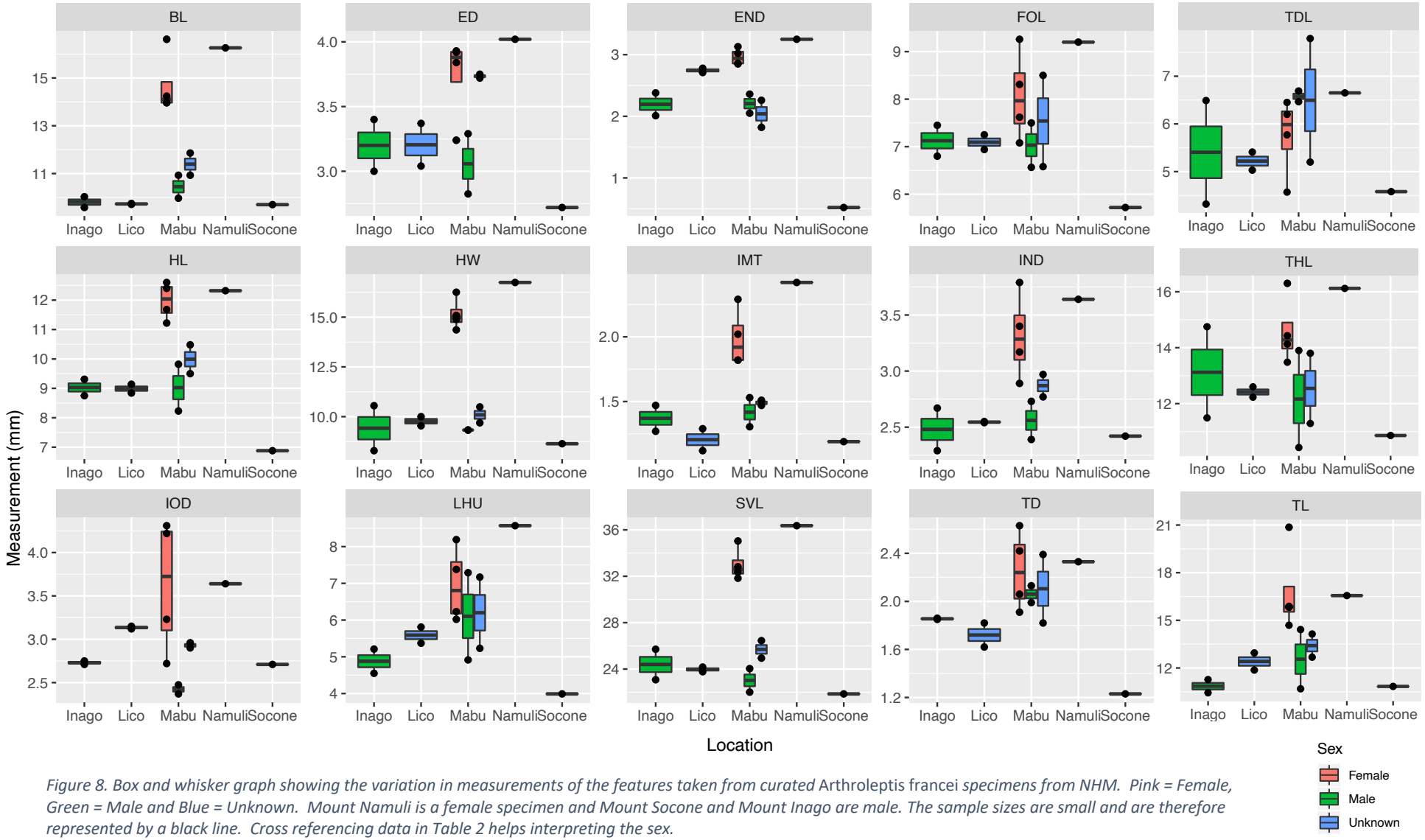
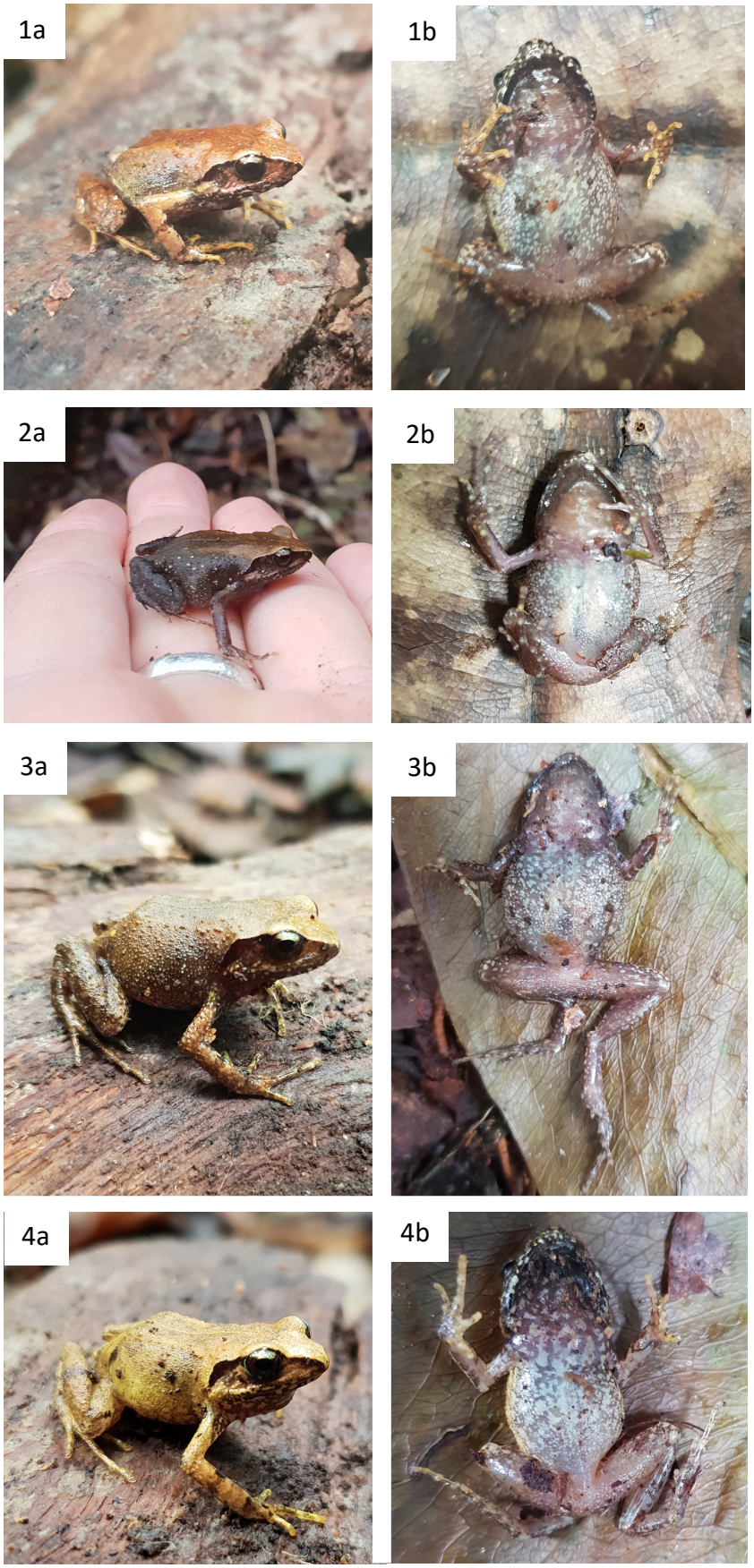


Figure 8. Box and whisker graph showing the variation in measurements of the features taken from curated Arthroleptis francei *specimens from NHM. Pink = Female, Green = Male and Blue = Unknown. Mount Namuli is a female specimen and Mount Socone and Mount Inago are male. The sample sizes are small and are therefore represented by a black line. Cross referencing data in Table 2 helps interpreting the sex.*



1a. *Side profile – Copper colouration, darker sides small epiderma lumps on the skin, distinct band across the eye, darker bands along the forearm. Bicolour iris.*
1b. *Ventral view – Very dark brown with light mottling and a lighter belly area.*

2a*. Side profile – Very dark brown with a lighter brown head and a distinct dark band along the eye. There are small white spots clustered on the smooth torso. Bicolour iris.*
2b. *Ventral view – dark brown vent particularly under the chin and rear with a lighter belly area.*

3a. *Side profile – Rough golden-brown fading to a darker brown on the torso. Distinctive dark band along the eye past the tympanum. Bicolour iris*
3b. *Ventral view – Dark brown under the chin and the rear with lighter mottling to the centre. Very long third digits on the forearm feet.*

4a. *Side profile – Golden on the entire dorsal. Smooth skin with the distinctive band along the eye. Light bands along the forearms and hind legs. Bicolour iris. Short third digit on forefoot.*
4b. *Ventral view – Very dark chin and lighter rear with a white mottling in the centre.*

Figures 1a – 4b: Woest, 2019.

Figure 9. Images of four male calling Arthroleptis collected on Mount Inago, December 2018.

Morphological features of *A. francei* vary in colouration with different shades of brown to golden colours (Figure 9). Specimen 1a in Figure 9 has a copper colouration with a darker torso with small, white epidermal spot clusters. There were also sparsely dispersed white spots along the dorsal. This was the only specimen with such spots out of all the individuals collected on Mount Inago. For all specimens collected, there is a distinct band across the eye and tympanum and a bicolour iris, with a lighter part at the top third and darker on the lower two thirds with a horizontal pupil as also described by Meijden (2006). Specimen 1a, 3a and 4a have darker bands along the forearms. All photographed individuals in the last expedition possessed a bicolour iris and a distinctive band along the eye and tympanum but all four were varying in shades of brown and golden. For all the specimens in the morphological analysis, the skin was not smooth on the back, yet small rough spots could be seen and felt. All the individuals had a dark brown chin and rear which extends to all four limbs, with a white mottling in the centre on the belly. There is very little webbing on the front feet and almost none on the hind feet.

3.2. Bayesian Phylogenetic Analysis

When considering the divergence from the sister taxon *A. reichei*, each tree formed three or four distinctive clades: 1.) Mount Lico and Chiperone (Southern), 2.) Mounts Mabu, Mulanje, Socone, Namuli and Inago (Northern), 3.) Mount Mabu 1 (Central), and 4. Mount Mabu 2 (Sandwich) (Figure 10). Mount Chiperone does not appear in the nuclear tree as sequence data were missing. In all cases (Appendix 2, Figures 12 - 14), Mount Mabu has two sympatric populations, P1 (T5764, T5766, T5765, T5756) and the larger population comprising of all the remaining individuals (P2, P3, P4; Figure 10).

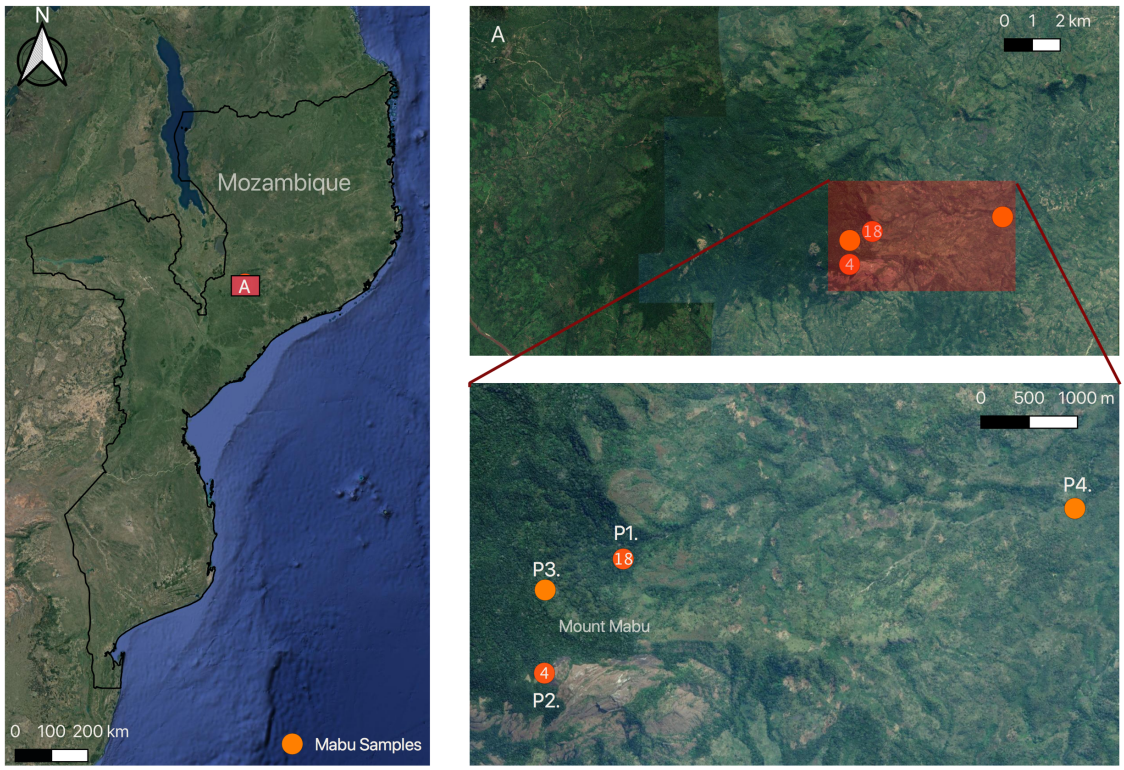


Figure 10. Left figure: Outline of Mozambique and the red box with an 'A' refers to the location of Mount Mabu. Top Right: A closer view of 'A', Mount Mabu samples (n = 24). Bottom right: Closer view of the samples from Mount Mabu. P1. referring to Population 1 of which 18 samples were collected. P2. n = 4, P3. n = 1 and P4. n = 1. Mount Mabu has a two-way sympatric population split where four samples from P1. form a definite clade with the remainder taxa forming a large, monophyletic clade.

To simplify the Bayesian phylogenetic analysis (BPA), a concatenated *12S*, *16S* and *rag-1* pruned tree was created with representatives from each mountain and clade (Figure 11). Taxa with missing genes were excluded from this tree, resulting in the absence of the individual from Mount Chiperone in the *16S* analysis. Despite the low branching support of all the individual genes, a sympatric divergence on Mount Mabu is confirmed, with a questionable additional clade consisting of T5757. Mounts Namuli, Mulanje, Inago and Socone support their own clade. Mount Lico is in the second Mabu clade.

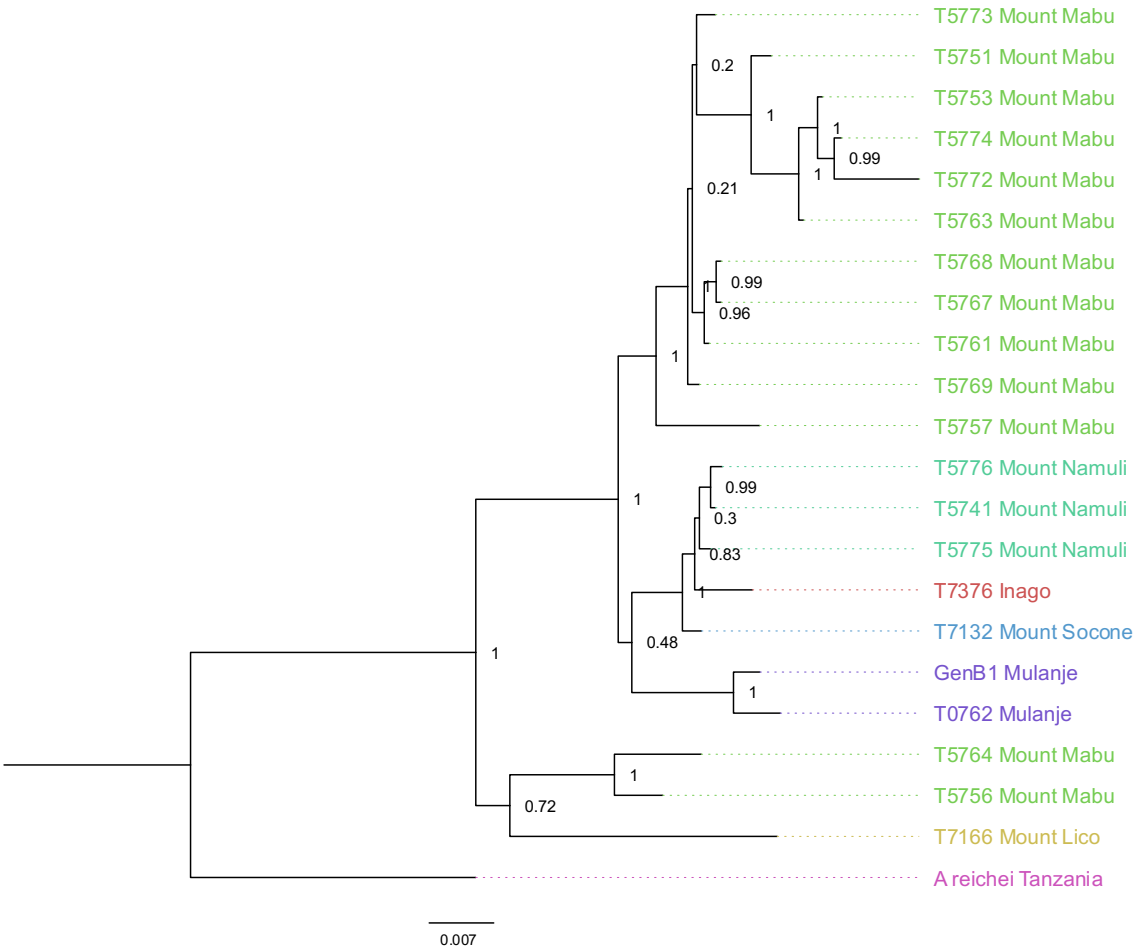


*Figure 11. Phylogenetic tree of the three genes (*12S, 16S *and* rag-*1) of representatives from each mountain displaying the probability of divergence between 0 and 1. Here it maintains the split on Mount Mabu and the cluster of Mount Namuli, Inago and Socone with the inclusion of another cluster containing Mounts Mabu and Lico individuals*

Bayesian phylogenetic analysis produced different topologies between the three markers used. According to all three markers (mtDNA – *12S* and *16S* and nuDNA – *rag-1* in Figure 12 – 14 and Appendix 2) Mount Mabu is occupied by two separate lineages (posterior probability of 0.21), and Mount Lico is occupied by another, distinct lineage. The concatenated mitochondrial *12S* and *16S* tree (Appendix 2) agrees with the nuclear *rag-1* tree (Figure 14), both supporting different posterior probabilities yet producing the same monophyletic relationships. There is strong evidence throughout the trees to support that Mount Lico (Figures 12 - 14) as well as Mount Chiperone (Figure 13) form distinct lineages. Mutations have created a hierarchy of shared substitutions to reconstruct the evolution history of the clade (Figures 12 – 13). Branch lengths represent the rate of substitutions, for example, with individuals from Mount Mabu

(T5772: Figure 12) and Mount Chiperone (T6836: Figure 13) showing longer branch lengths. Comparative differences in substitution rates are higher in *rag-1* (Figure 14).

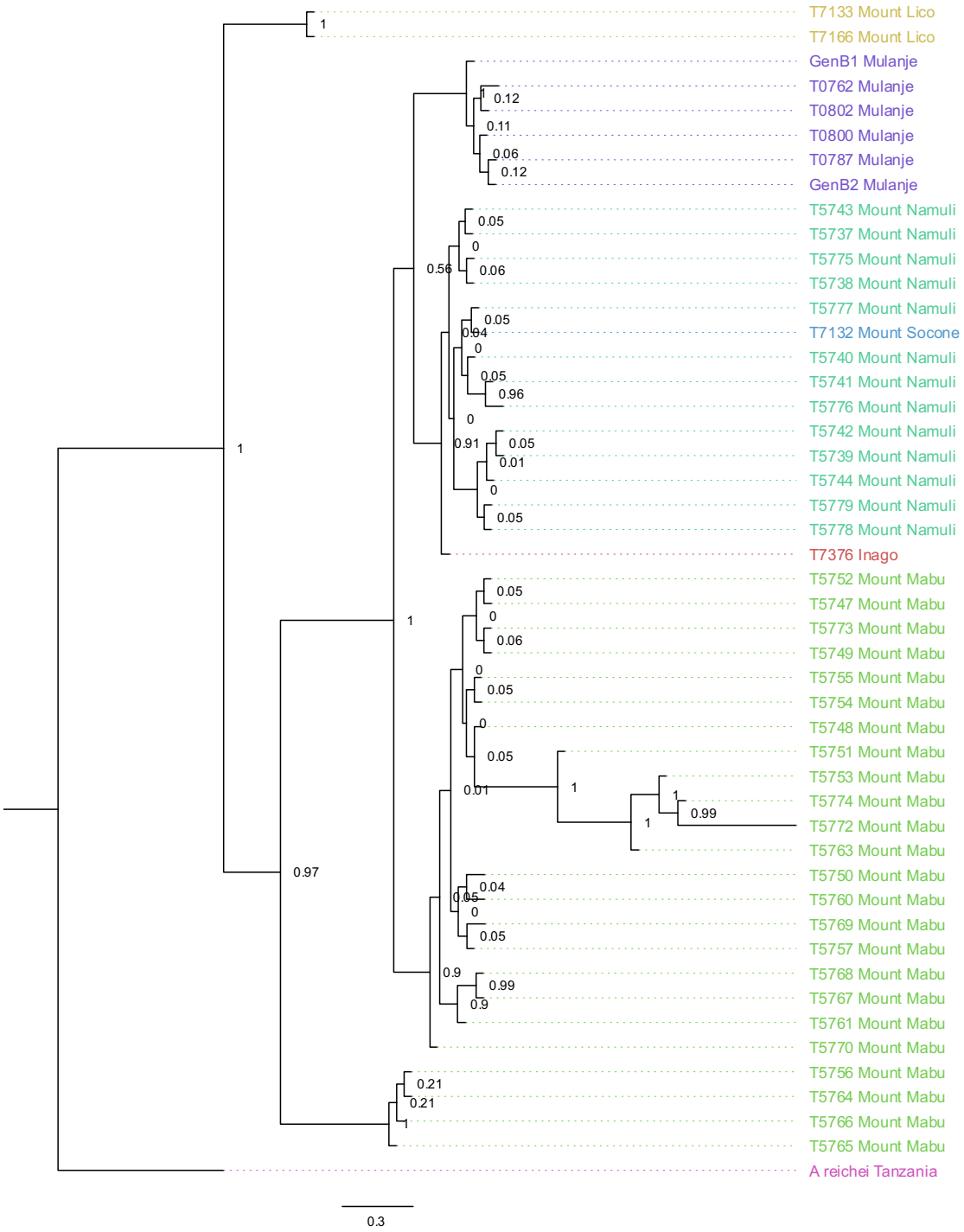


Figure 12. A 12S phylogenetic relationship between Arthroleptis francei *candidate species where* A. reichei *was used to root the tree. Branch support posterior probability values are displayed from 0 to 1. Tip labels represent the individual by the 'T' number and Mountain on which they were collected. Tip colouration is by mountain.*

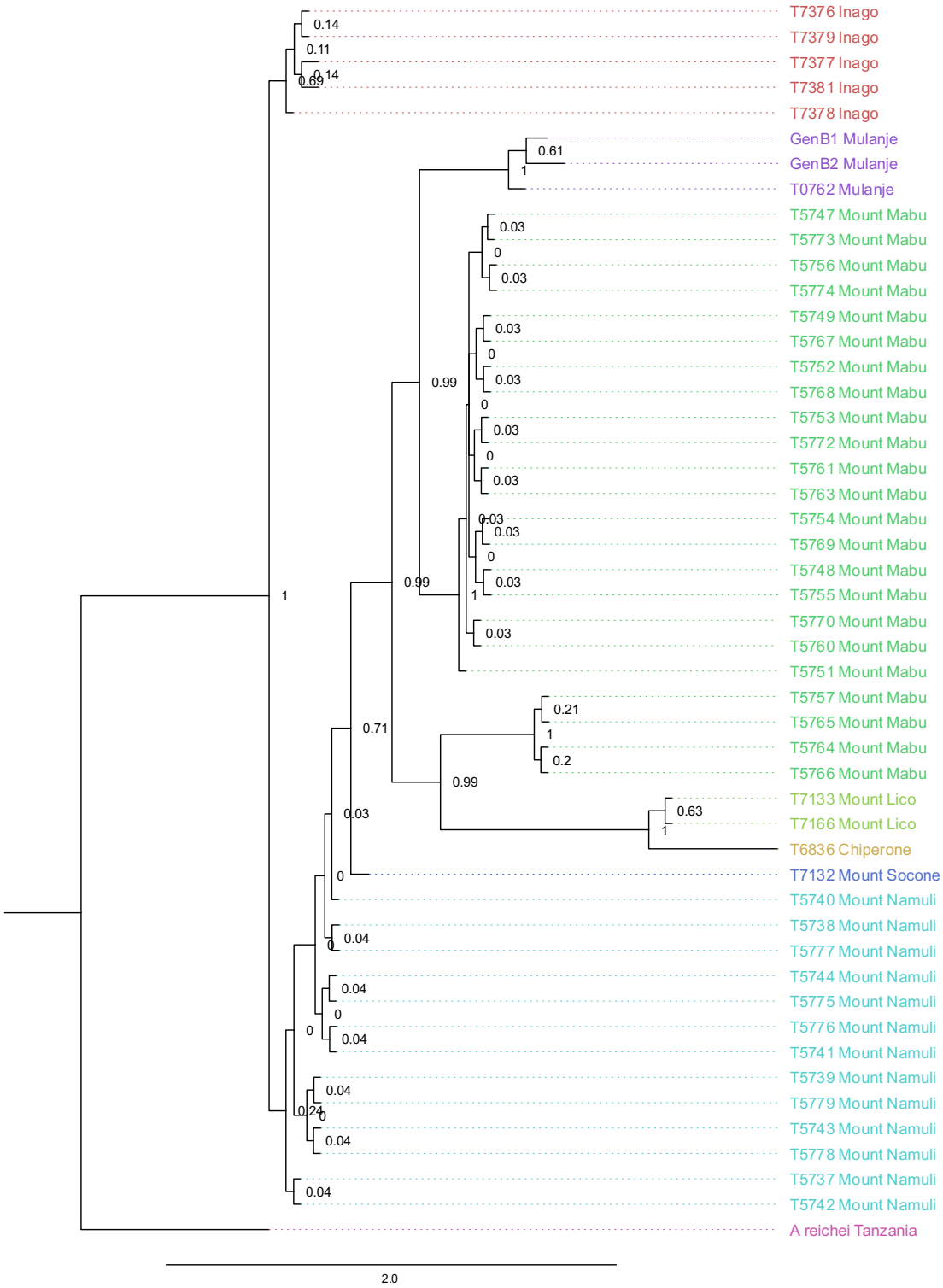


Figure 13. A 16S phylogenetic relationship between Arthroleptis francei candidate species where A. reichei was used to root the tree. Branch support posterior probability values are displayed from 0 to 1. Tip labels represent the individual by the 'T' number and Mountain on which they were collected. Tip colouration is by mountain.

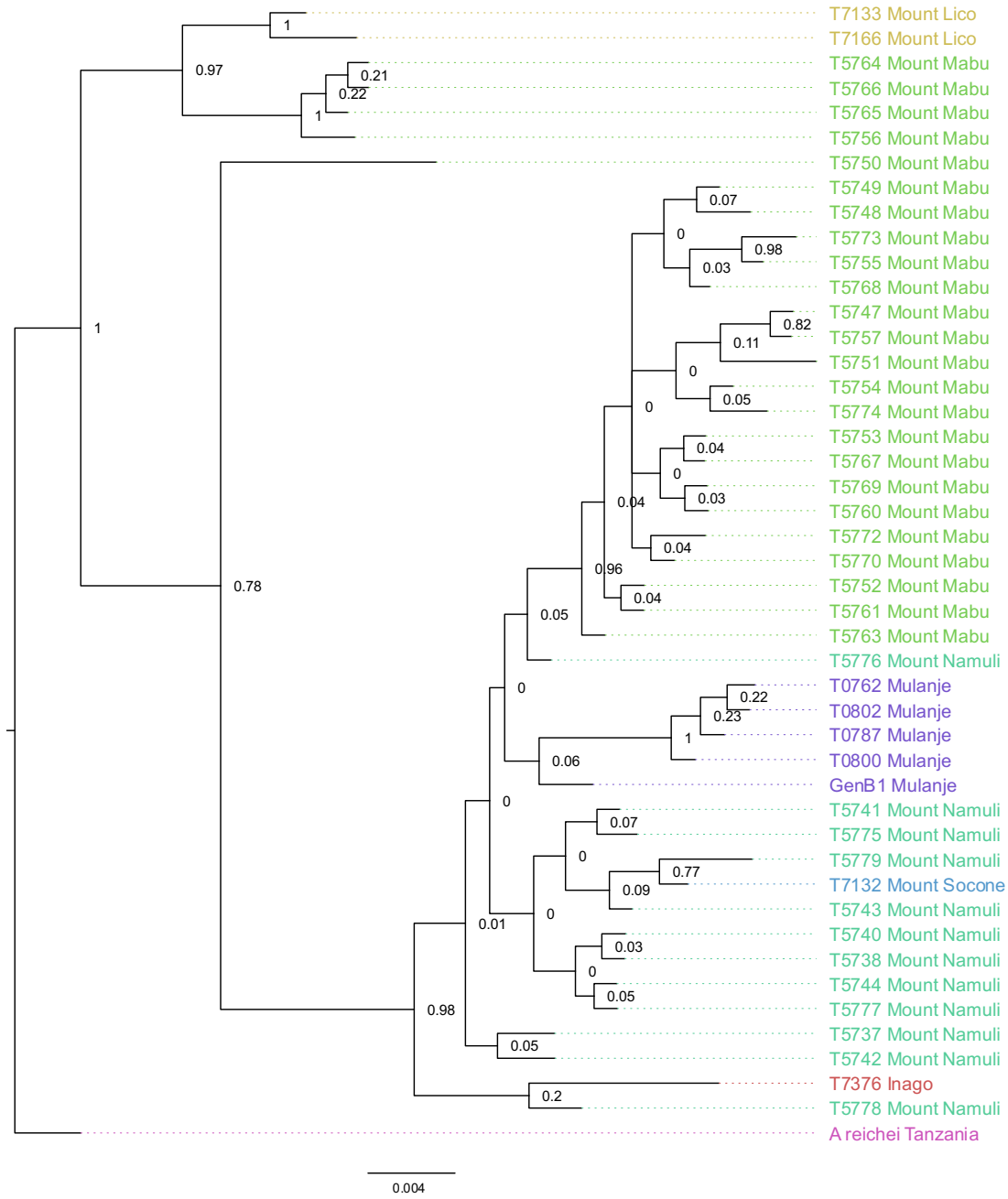


Figure 14. A rag-1 phylogenetic relationship between Arthroleptis francei candidate species where A. reichei was used to root the tree. Branch support posterior probability values are displayed from 0 to 1. Tip labels represent the individual by the 'T' number and Mountain on which they were collected

It was in Blackburn (2008b) that *A. francei* was first sampled, but only represented by populations from the type locality in Mulanje, Malawi. The present study provides evidence that there is extensive variation in intra-species divergence of *A. francei* across the known populations. The proportional (*p*) distances in Table 4 displays a matrix of all the possible sequence pairs from the *16S* gene sequences for those used within this study. Based on a Mount Mulanje specimen (GenB1) in the *16S* analysis (Figure 13), the divergences between individuals from other mountains form monophyletic clusters according to mountain. With the Mount Mulanje specimens being from the type locality, distances between the clusters formed in Figure 13 are as follows with the *p*-distances stated in brackets taken from Appendix 10. Results revealed a low intrapopulational variation of 0 - 0.07 uncorrected pairwise distances in the *16S* gene from the GenB1 individual: GenB1 to Mount Inago individuals - T7378 (0.04), T7381 (0.04), T7377 (0.04), T7379 (0.04), T7376 (0.04). GenB1 to Mount Chiperone T6836 (0.09). GenB1 to Mount Mabu individuals - T5757 (0.05), T5765(0.05), T5764 (0.05), T5766(0.05). GenB1 to Mount Lico – T7133(0.07) and T7166 (0.07). GenB1 to Mount Socone - T7132 (0.03). To simplify, a comparison of the proportional (*p*) distance between the clusters in the *16S* BI analysis are displayed in a matrix in Table 4. The p-distance in Table 4 reveal the Malawi clades (Mulanje) to be c. 3% - 9% divergence from the Mozambiquean clades.

Table 4. A comparison of proportional (p) distances for the 16S BI Arthroleptis francei clusters. The bottom matrix presents the interclade distances.

	Malawi	Mozambique					
	Mulanje	Inago	Mabu 1	Mabu 2	Lico	Chiperone	Socone	Namuli
Mulanje	0.000							
Inago	0.042	0.000						
Mabu 1	0.054	0.041	0.000					
Mabu 2	0.031	0.025	0.037	0.000				
Lico	0.066	0.062	0.059	0.062	0.000			
Chiperone	0.087	0.083	0.076	0.083	0.028	0.000		
Socone	0.033	0.009	0.034	0.021	0.057	0.078	0.000	
Namuli	0.038	0.005	0.037	0.021	0.057	0.078	0.005	0.000

3.3. Distribution network

There was a clear separation between the main geographic haplotype groups across all three genes; *12S* and *16S* rRNA (Figure 15 and 16 respectively) and *rag-1* nuDNA (Figure 17). The *12S* marker showed an overlap between lineages which only occurred on Mounts Mabu, Socone and Namuli. The same was the case for *rag-1*, with the addition of Mount Inago. All *16S* haplotypes were fully diagnostic for specific mountains. Mount Mabu (green) supports two separate clades, whilst *rag-1* results in a possible three-way split. Individual T5756 has multiple index mutations from the larger Mabu clade as well as T7133 from Mount Lico. Haplotypes from Mounts Namuli, Socone and Inago have fewer index mutations, resulting in a close relationship with those from the type locality, Mount Mulanje. The geographic haplotype clusters in Figures 15 and 16 concur with each other, resulting in Mount Mabu harbouring two populations, and populations from Mounts Inago, Socone and Namuli being closely related to each other. Mount Lico and Mount Chiperone form distinct populations.

The *rag-1* TCS network (Figure 17) follows a similar pattern but forms four main clusters: 1) Mount Mabu, 2) Mounts Mabu and Socone, 3) Mount Lico and 4) all other mountains. One specimen from Mount Socone (T7133) differs from other Mount Socone individuals who share close similarities with Mounts Namuli, Mabu and Inago. One individual from Mount Mabu is positioned closely to those from Namuli, Socone, Mulanje and Inago.

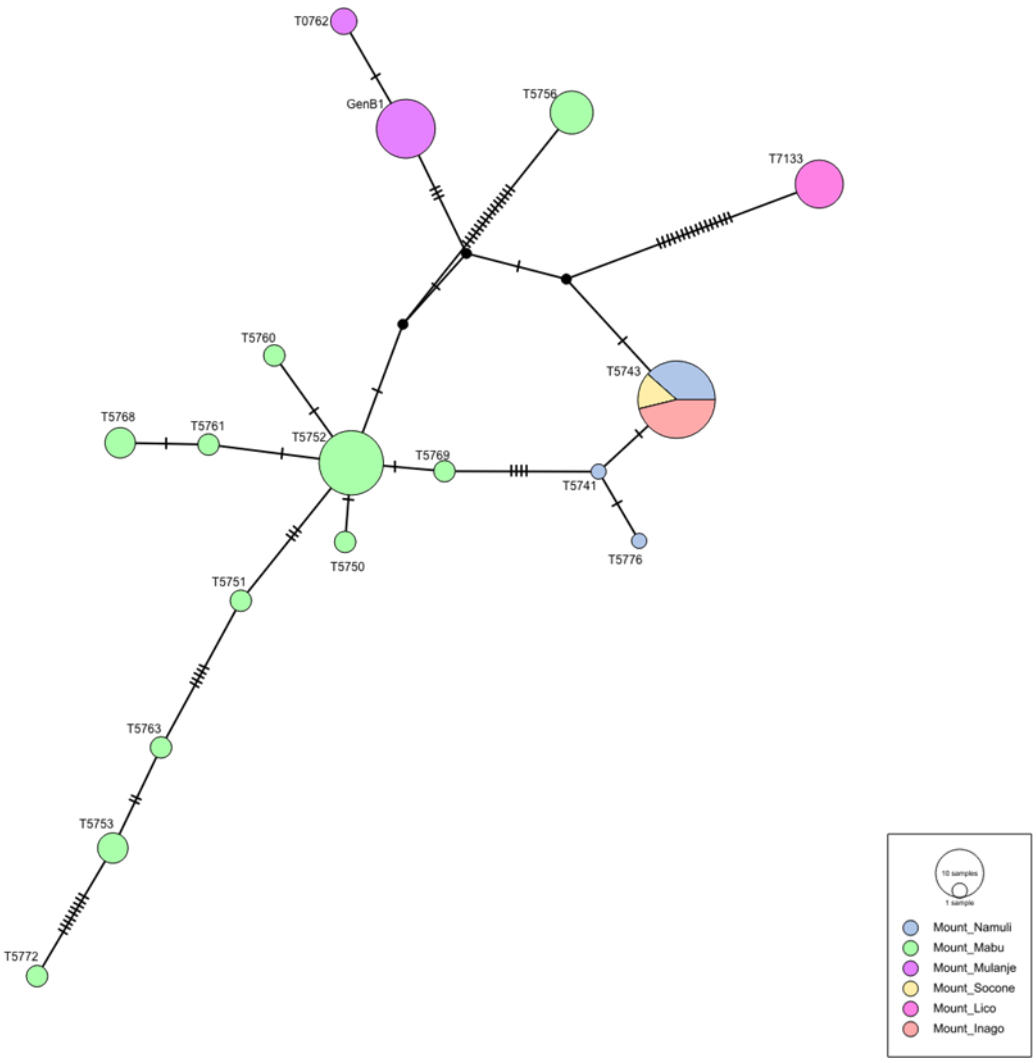


Figure 15. TCS Network for mtDNA 12S with circle colour representing the locality and the circle size depicting the number of individuals. Mutations at index restriction sites are presented with a diagonal strikethrough on the branch which is proportional to the number of mutational steps. These index mutations display the change in haplotypes from the common ancestor

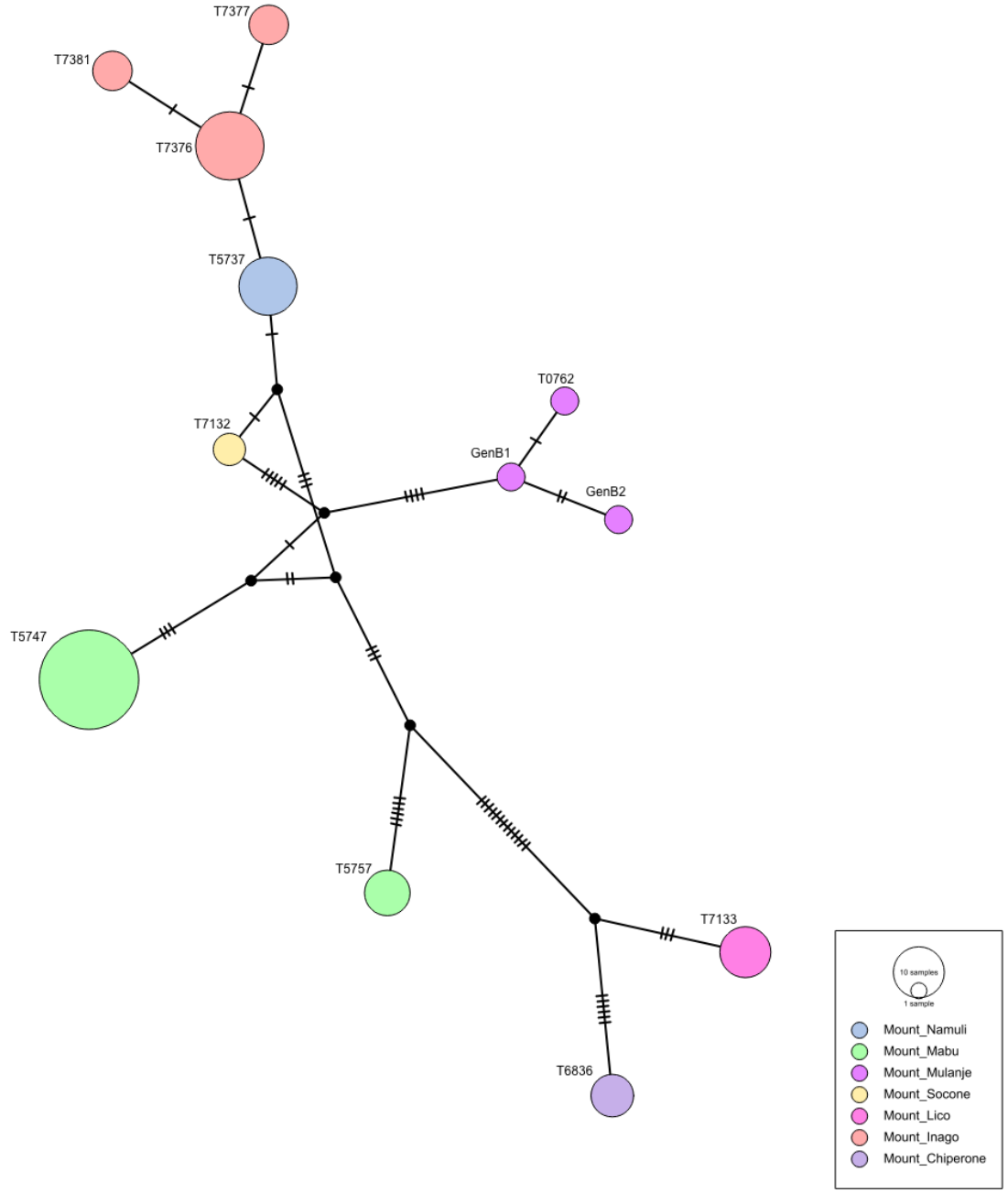


Figure 16. TCS Network for mtDNA 16S *with circle colour representing the locality and the circle size depicting the number of individuals. Mutations at index restriction sites are presented with a diagonal strikethrough on the branch which is proportional to the number of mutational steps. These index mutations display the change in haplotypes from the common ancestor*

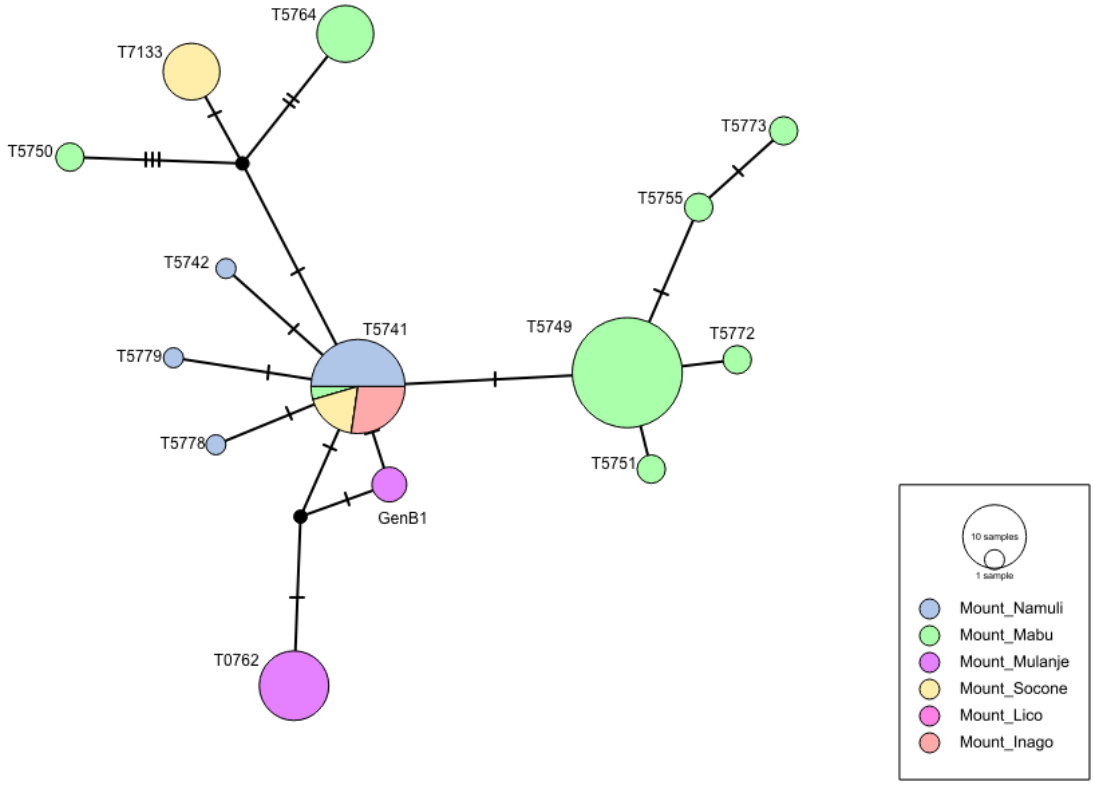


Figure 17. TCS Network for nuDNA rag-1 *with circle colour representing the locality and the circle size depicting the number of individuals. Mutations at index restriction sites are presented with a diagonal strikethrough on the branch which is proportional to the number of mutational steps. These index mutations display the change in haplotypes from the common ancestor*

3.4. Species delimitation

The species delimitation method (*Bayesian General Mixed Yule Coalescent* - bGMYC) revealed six cryptic species of *Arthroleptis* (Figure 18). Mounts Inago, Socone and Namuli formed a joint clade, and Mounts Chiperone, Lico and Mulanje forming a clade each. The previously revealed split into two lineages on Mount Mabu remains and is also supported throughout the *16S* BPA posterior probability (Figure 13). The bGMYC shows more support than the BPA in defining a larger number of lineages. The distinct Mount Inago lineage (green) is supported by the BPA model, whilst bGMYC groups them with Mount Namuli and Mount Socone. The bGMYC posterior probability threshold was set to 0.5 for the lumping of samples into species, also to avoid over lumping or further splitting.

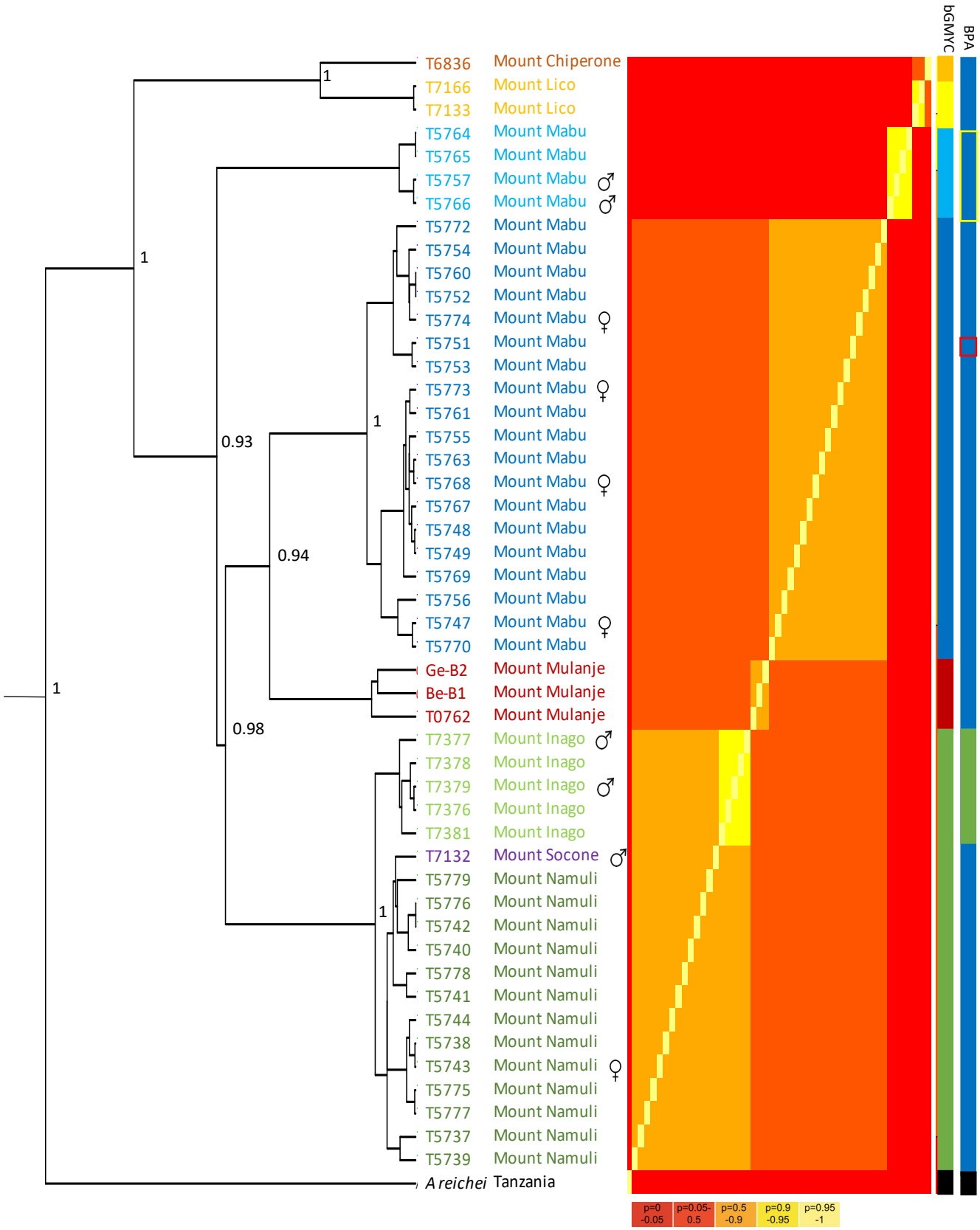


Figure 18. Species delimitation using bGMYC for the 16S gene of Arthroleptis francei displaying posterior probabilities. The heat map shows six putative species of Arthroleptis with A reichei as the outgroup. Specimen numbers and locations are coloured according to a new clade. Probability subsets of the heat map are: Low (red) to High (yellow) with the values of P = 0 – 0.05, 0.05 – 0.5, 0.5 – 0.9, 0.9 – 0.95, 0.96 – 1. Right: Simplification of the 16S bGMYC and BPA models (taken from Figure 13) where each colour corresponds to a new radiation. The hollowed-out shapes represent the different Mabu clades

3.5. Species Distribution Model (SDM)

The species distribution model (Figure 19) reveals the predicted locations suitable for *A. francei*. The total permutation importance variation of the model consisted of annual precipitation (Bio 12, 90.9%), mean temperature of the coldest quarter (Bio 11, 6%), temperature of the wettest quarter (Bio 8, 2.2%) and annual mean temperature (Bio 1, 0.9%). The jacknife test verified that Bio 12 is the most important environmental variable with the highest gain. *Arthroleptis francei* are known to live in moist forest at

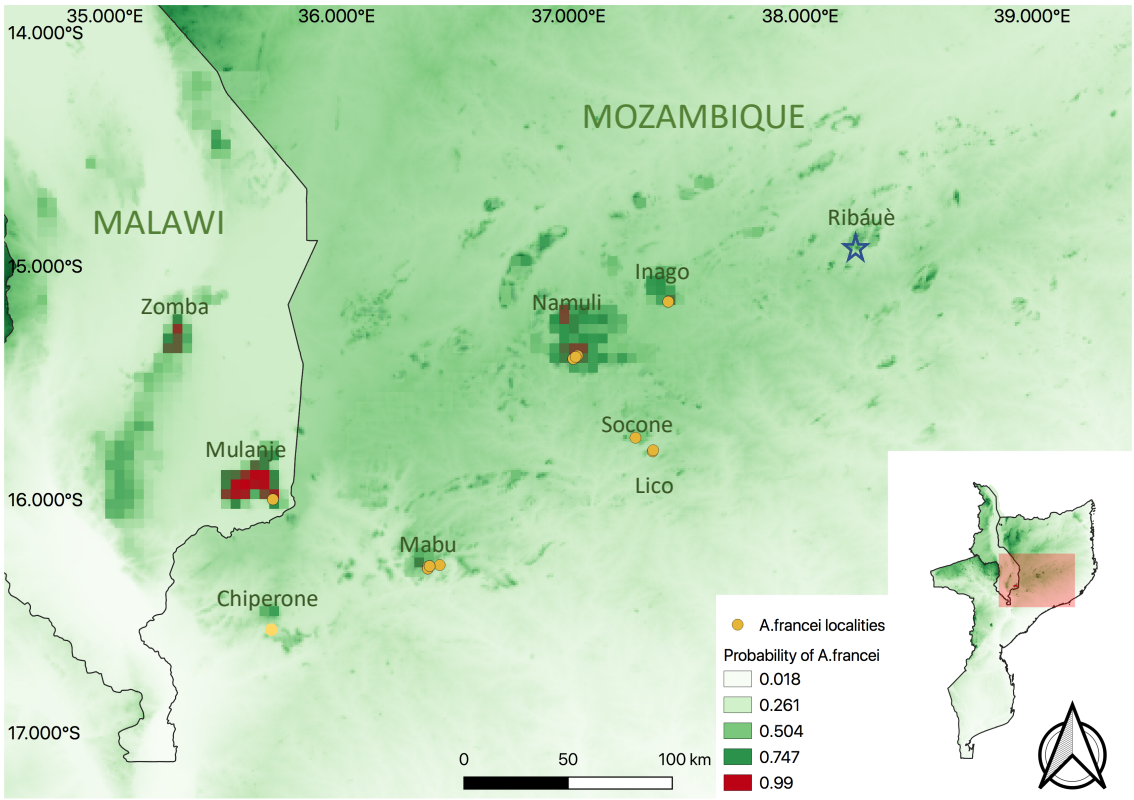


Figure 19. Species distribution model for Arthroleptis francei *for the current climate. The colour ramp in the key indicating the probability of distribution and the yellow dots displaying the known distribution of specimens in the phylogenetic study. The star denotes the location where* Arthroleptis *males were heard calling but not collected. The circles denote the separate clades according to the bGMYC model (Figure 18).*

high elevations, atop granite inselbergs where temperature and precipitation are important. This is confirmed by a model where precipitation dominates, despite the additional influence of temperature (see also Appendix 8). In Figure 19, the probability of distribution is displayed as a colour ramp from white being low (0.018%) and red being high (0.99%). Mount Ribáuè has a probability of *A. francei* occurrence between 0.504 and 0.747, which corresponds with males being heard calling in November 2018 (personal observation). Several locations are suggested to have a high probability of occurrence, including the Zomba mountain in Malawi where the species was not

recorded during past expeditions (Simon Loader, pers. comm.). Mount Chiperone was identified as a suitable location which led to the individual T6836 being included in this study.

3.6. Phylospatial distribution of Arthroleptis francei

The phylospatial analysis resulted in an uneven distribution of PE, PD, WE and SR across the geographic region, with three, four, five and six hotspots, respectively (Figure 20). The highest level of PE (0.020) was found at Mount Chiperone, but this may be due to limited genetic sampling. The PE for the isolates, Mounts Mulanje, Mabu and Lico have highly divergent lineages resulting in a medium PE whilst the remaining isolates resulted in low PE per unit area. Compared to PE, WE measured slightly greater spatial distribution throughout the geographic area, yet the higher levels were the same with Mount Mabu and Mount Namuli resulting in medium endemicity. Sampling here would be beneficial to confirm such results. Mounts Mulanje, Chiperone, Lico, and Mabu all show a medium to high PD across the region which corresponds with the diversity within the BI (Figure 13) and the species delimitation model (Figure 19). The SR is the diversity of species within a given area or ecosystem. Therefore, with the phylogeny and SDM data for *A. francei* candidate species, Mounts Mulanje, Mabu, Chiperone, Namuli, Socone and the area of high ground between Mabu and Namuli have all resulted in high SR (0.6 – 0.8). All the phylospatial data confirms the high endemism and richness rankings of the *A. francei* candidate species on the SMI.

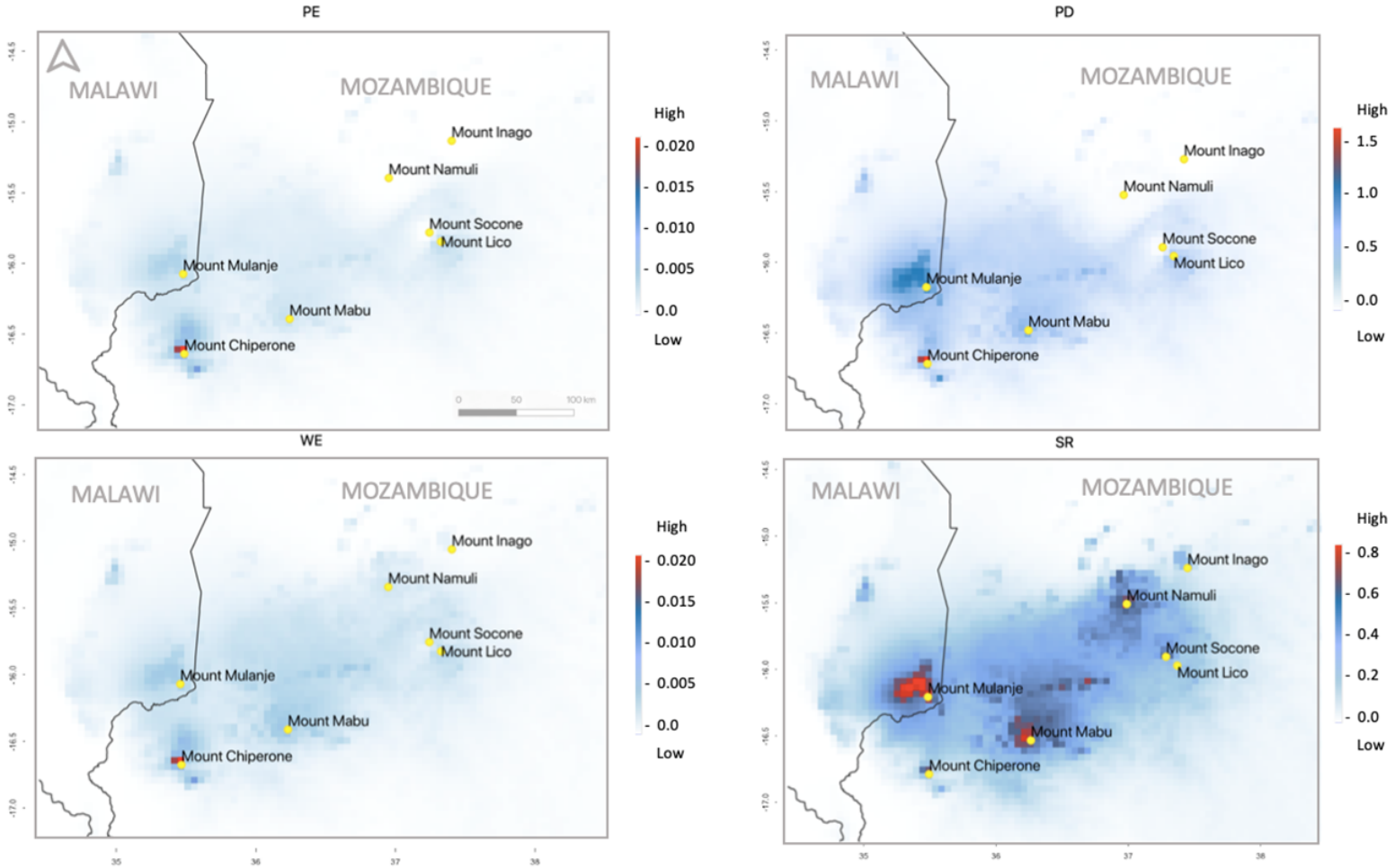


Figure 20. Distribution of phylogenetic endemism (PE), phylogenetic diversity (PD), weighted endemism (WE) and species richness (SR) of A. francei candidate species within this study in Malawi and Mozambique. Models were constructed with 16S phylogenetic data from Figure 13 and distribution data from Figure 20.

3.7. Conservation

Three areas, Mounts Chiperone, Mulanje and Lico were identified as hotspots for PE. Of these montane isolates, only Mount Mulanje is listed as protected on the world database of protected areas. Mounts Namuli, Mabu and Chiperone feature as KBA's by the CEPF, with Mount Morrumbala, to the south of Chiperone being listed as a KBA awaiting assessment (Hoffman *et al.*, 2016). Of the eight sites where specimens were collected, four contain a certain protection status with *A. francei* potentially residing on Mount Morrumbala. So far no known specimens were collected from this isolate, therefore it could be an area requiring further surveying efforts. The PE on Mount Lico measures at a medium level (between 0.010 – 0.015) and along with the phylogenetic analysis, this lineage is separate to that of the initially named *A. francei* (Figure 21). A new lineage endemic to this mountain is sufficient to point to a possible requirement for protection.

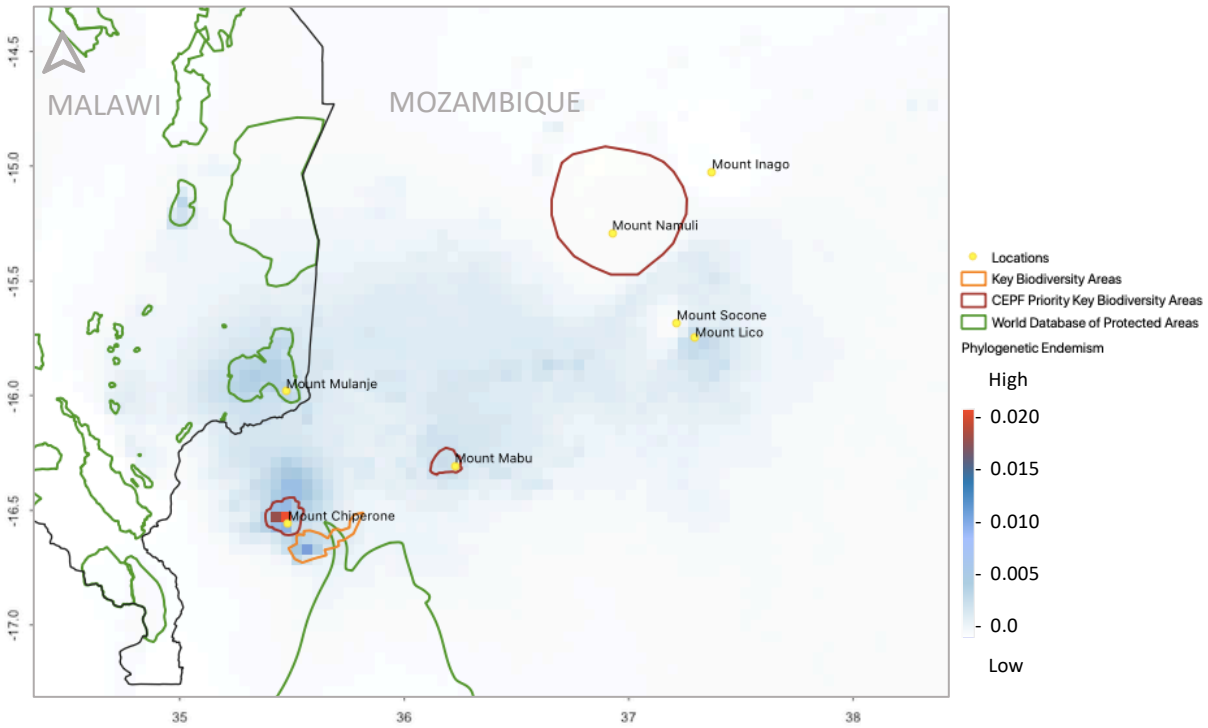


Figure 21. Identified PE hotspots intersecting protected areas in both Malawi and Mozambique. These protected areas are provided by the World Database of Protected Areas and the Critical Ecosystem Partnership Fund. A total of 50% of the locations where specimens were found in this study have received a level of a protection status.

3.8. Call analysis

The documented call (Figure 22) is of a single male *A. francei*. The call displays three parts, with ten syllables in total. Part 1 consists of a short "beep" (# 1), followed by six longer calls, increasing in intensity and sounding like a drawn out, high pitched "beeeep" (# 2-7), with similar time gaps between each. Part three consists of three rapid "beeps" with similar short gaps between them.

The entire call lasted 25.82 seconds (Table 5). The total mean call intensity (measured in decibels - dB) of Part 1 was 54.40dB (between 10.96dB for syllable 1 and 58.34dB for syllable 9). The call intensity of Part 2 was greater than Part 1 at 70.79dB, with the minimum at the start at 48.33dB and a maximum in the centre at 75.72dB. The minimum frequency (measured in Hertz – Hz) was 2975.95Hz at both syllable 3 and 4 with a maximum recording in the last three syllables, in particular the final one at 3034.15Hz. The frequency tends to slightly fluctuate throughout the call ending with a graduating increasing frequency at the end.

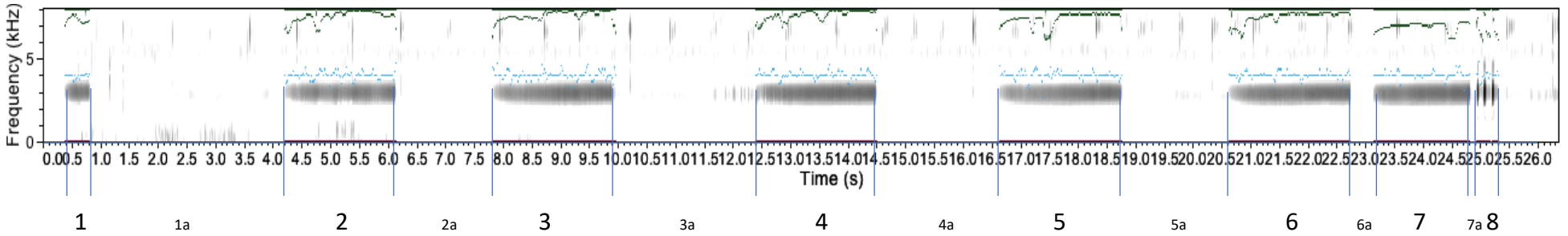


Figure 22. Spectrogram to visualise the call of a male Arthroleptis francei from Mount Inago, Mozambique. Sections with whole numbers 1 – 8 display the intermediate units (syllables) within the song. The gaps are labelled with smaller numbers (1a – 7a). Light blue – frequency change, dark green – amplitude.

Table 5. Acoustic analysis of the call data where the column headings represent the sections visualised in Figure 22. Section 8 in Figure 22 encompasses 8, 9 and 10 in the table below. Abbreviations: dB – Decibels, Hz - Hertz

	1	1a	2	2a	3	3a	4	4a	5	5a	6	6a	7	7a	8	8a	9	9a	10
Time (s)	0.62	4.15	2.01	1.63	2.14	2.54	2.03	2.14	2.2	1.81	2.12	0.35	1.68	0.09	0.06	0.06	0.06	0.06	0.07
Minimum Intensity dB	10.96		32.06		37.31		39.02		33.7		44.49		56.48		48.33		50.18		50.19
Maximum Intensity dB	56.38		58.01		57.22		58.35		57.16		58.34		59.36		73.19		75.72		75.64
Mean Intensity dB	52.68		54.52		53.7		54.59		53.69		54.82		56.82		68.6		71.77		71.99
Frequency Hz	3007.29		2993.86		2975.95		2975.95		2984.91		2980.44		2993.86		3016.24		3029.67		3034.15

4. Discussion

This study has conducted a phylogenetic analysis based on the largest sequence data set of *A. francei* accessible to date. This phylogeny was used as a basis for investigating several phylogenetic, biogeographic and trait-related hypotheses. Furthermore, with the incorporation of distribution modelling, morphometrics and elevation analyses, the study has provided an opportunity to suggest further areas for research. The study framework builds on previous work by Blackburn (2008b), with the addition of further elements to the research and a larger dataset. The trees are based on the largest geographical sampling of *A. francei* to date, propose new relationships among populations, and reveal cryptic diversity suggesting the existence of a species complex.

4.1. Morphology

This study revealed marked variation in both size and colour in *A. francei*, however without consistent patterns across populations, suggesting that further research is required. Due to data missing for specimens from the type locality Mount Mulanje, a thorough comparison of both colour and size between specimens from other locations cannot be made. Four of the five females measured belong to the largest Mabu clade with reference to the SDM (Figure 19). Three males and one female were from the Inago/Socone/Namuli clade, and two males were from the small Mabu clade. The female-biased sexual size dimorphism (SSD) revealed in the present study conforms to previous findings, and is exhibited by around 90% of anurans around the world (Shine, 1979; Fairbairn, Blanckenhorn and Székely, 2007; Bagaturov, 2018). Female-biased SSD is generally consistent with Shine's (1979) sexual selection hypothesis which predicts that females are on average 30% larger than the males. Male-biased SSD is only exhibited by 3% of frogs, and examples include the *Limonectes* genus (fanged frogs) from South East Asia (Emerson, 1994). Larger males are usually linked to territorial behaviours, display aggressive male to male combat, and have developed secondary adaptations such as fangs or tusks (Tsuji, 2004). Differences in age structure between the sexes result in sexual bi-maturity, whereby females can be older than males and also mature at a smaller size (Halliday and Verrell, 1986; Monnet and Cherry, 2002). Many other factors can affect SSD in *A. francei*, such as life histories, fecundity, sexual

and correlational selection as well as ecological and energetic constraints (Liao *et al.*, 2013). Being direct developers, size could be positively correlated with investment into offspring production, resulting in males being smaller (Zhang and Lu, 2013).

Males secondary sexual traits of *Arthroleptis* and *Cardioglossa* include the possession of an elongated third finger, which is reflected in their name, meaning 'long-fingered'. Within these genera the males' third finger (TDL) can reach lengths of 30-40% of their snout vent length (SVL), whilst respective values for females are generally <20% (Blackburn, 2008b). In the present study, males' TDL is longer at an average of 24.5% compared to females, reaching only 17.6% of SVL. Despite the female's overall larger body size, the males' TDL are absolutely longer than those of females. There have been no comprehensive studies regarding secondary sexual traits within *Arthroleptis*, and only brief mentions of male-biased TDL within *Cardioglossa* (Herrmann *et al.*, 2004; Blackburn, 2008a). As there are no other documented differences between the sexes, it is likely that the TDL and SVL are characteristics of sexual selection processes as first mentioned by Darwin (1879). This is then later confirmed by Blackburn (2009b), who concluded that a positive relationship between SVL and finger length is a result of allometric growth.

The use of these longer digits still remains unclear and raise many questions for their uses: Are they used as a warning signal? Are they for attracting females? Are they useful in amplexus? Or, do they help in male to male combat as seen with *A. stenodactylus* (Bittencourt-Silva, Langerman and Tolley, 2020)? Blackburn (2009) looked into various hypotheses of TDL with the inclusion of digital and inguinal spines on the toes to understand the diversity and evolution of secondary sexual characters in *Arthroleptis* and *Cardioglossa*. Male traits were present in the most recent common ancestor, whilst the reduction or loss occurred later on. Changes in sexual selection pressures are important in understanding mail trait diversity and whether it contributes to reproductive success. However, without in-depth behavioural studies, it is impossible to understand the exact purpose of the third digit length and other morphologies such as the spines as mentioned by Blackburn (2009b). Further research is needed on *Arthroleptis* candidate species in regard to SSD on a per-species basis, and phylogenetic

comparative analyses would be beneficial in understanding the complex evolutionary patterns of anuran SSD and mating systems.

As data on the calls of *A. francei* are limited to the present study, more research is needed to determine the two sympatric clades identified on Mount Mabu exhibit any niche partitioning. To the ear, the call on Mount Ribáuè was indiscernible from the call which was recorded on Mount Inago. None of the measured morphological characteristics enabled a differentiation between the candidate species, despite differences in the length of the humerus and the third digit between Mount Mabu and Inago. Further measurements of specimens in collections and a statistical analysis would aid in the identification of diagnostic morphological characteristics between taxa.

Selection promotes the evolution of exaggerated secondary sexual traits (Darwin, 1879; Shuster and Wade, 2003), and in combination with multiple biotic and abiotic factors impacts on colour phenotypes of animals (Rudh and Qvarnström, 2013). Colour and patterns have evolved for predator avoidance (Segami Marzal *et al.*, 2017), attracting a mate (Stuart-Fox and Ord, 2004), thermoregulation (Tracy *et al.,* 2010), intra- and interspecific communication (Caro, 2005), and UV-B protection (Garcia, Stacy and Sih, 2004). Sexual dimorphism amongst anuran species is visually evident in morphological traits such as colouration and body size (Hoffman and Blouin, 2000). The colouration and patterns of the candidate species in this study vary greatly from golden to dark brown, some presenting patterns on their dorsal side, and none being uniform in colour. The forest floor habitat the species reside in contains a variation in the colouration and sizes of leaves which form the leaf litter (Figure 3). This plays a large part on the evolution of colouration which would aid in camouflage for predator avoidance such as *Rhinella margaritifera* with a leaf litter colouration (Duellman and Trueb, 1994), similar to that of *A. francei*. Local adaptations often induce geographical colouration differences, camouflaging to the backgrounds inhabited in particularly when there is a higher risk of predation (Krohn and Rosenblum, 2016). Within the candidate species in this study, there are different colour and pattern phenotypes within the same location (Figure), yet all individuals within the same or different populations are different to each other. Variation of colour and size between specimens leads to taxonomic

implications when in the field identifying between specimens. To make a thorough analysis on whether colour and/or size plays a defining factor in distinguishing between populations will help solve the taxonomic uncertainties.

During adaptive radiation, selection results in rapid phenotypic diversification within a species (Glor, 2010). A large body of literature focussed on adaptive radiation for example in the famous Darwin's finches (Grant and Grant, 2002), cichlids in Lake Malawi (Seehausen, 2006), and a range of mammals (Baldwin & Sanderson, 1998; Campagna *et al.,* 2015; Couzens & Prideaux, 2018). However, this area is less explored in amphibians (Blackburn *et al.,* 2013; Bossuyt & Milinkovitch, 2000; Stuckert, Linderoth, MacManes, & Summers, 2019), and could be an area for further research. There is a close correlation between amphibian phenotypes and habitat types (Duellman and Trueb, 1994), where there is substantial opportunity for dispersal limitations and high philopatry leading to population divergence (Vences and Wake, 2007). The adaptive radiation in this *Arthroleptis francei* clade could be enabled by advances in life histories such as being direct developers or their morphologies. Studies on relationships between habitat utilisation and morphological variation which are linked to adaptive radiations are limited (Kozak *et al.*, 2005), yet this has been seen in the tropical plethodontid salamanders (Wake, 2009). Could the colouration of the Arthroleptis candidate species be a trait of adaptive radiation or a characteristic of niche segregation? With niche segregation applying to different species inhabiting the same locality, yet adapting to the use of different habitats, geographic areas, diets and much more to successfully co-inhabit. This may only be applicable to the two populations on Mount Mabu, as phylogenetics have resulted two populations within the same locality. Without knowing the full extent of their ecologies, could the phenotypic traits of skin colour be the adaptation to perhaps occupying different niches?

As most amphibians are nocturnal, the importance of colour in communication has largely been neglected. However, there are studies stating the that colouration plays an important factor in mating (Rudh and Qvarnström, 2013), and it is not only the vocalisations that attract females (Jacobs *et al.,* 2017). Colouration and vocalisation, in the purpose of mate choice, may convey different signals (Gomez *et al.*, 2011) where

the vocal sacs may be a visual cue rather than for sound proliferation (Starnberger, Preininger and Hödl, 2014). For *A. francei*, the vocal sacs of males are darker than the rest of their underside, and therefore could act as a visual attractant. Female squirrel tree frog (*Hyla squirella*), have been found to prefer males with a larger lateral body stripe at given calls displayed (Taylor, Buchanan and Doherty, 2007), whilst female European tree frogs (*Hyla aborea*) have shown to prefer noticeably coloured males (Gomez *et al.*, 2009). Here, colour may not just be an adaptation to the environment but also a display to attract a mate.

4.2. Phylogeny and taxonomy

Phylogenies based on molecular data enable an understanding of the taxonomy of a group as well as diversification patterns. By combining existing and new molecular data this is the first detailed, phylogenetic analysis of *A. francei*. The Bayesian phylogenetic analysis, species delimitation and haplotype network analysis jointly suggest that *A. francei* comprises of cryptic taxa distributed across the inselbergs of Southern Malawi and Mozambique, including a case of a divergence at the same location, Mount Mabu. The species delimitation model identifies all locations as well supported lineages of distinct species except for the populations located in Mount Socone, Inago and Namuli. The multilocus analysis supports a split amongst the Mabu specimens, which is independently evidenced by the three investigated genes (Figures 12 - 14), and the concatenated results (Appendix 2). There are six well supported clades from the seven mountains investigated in the SDM, confirming a high probability of two taxa on Mount Mabu.

With these new findings, an updated phylogeny of *Arthroleptis* from Blackburn (2008b) can be created. The sister species of *A. francei* are *A. nikeae, A. reichei* and *A. tanneri*, all larger *Arthroleptis* found in the Eastern Arc Mountains in Tanzania (Blackburn, 2008b). The only larger species of *Arthroleptis* found in Mozambique is currently *A. francei*, suggesting that the *Arthroleptis* species in this region could have originally radiated from the northern part of the EAR. Recent studies from the sky islands of this region are helping in the identification of cryptic vertebrates such as in reptiles (Ceccarelli *et al.*, 2014; Branch *et al.*, 2019), and amphibians (Conradie, Bittencourt-Silva

et al., 2018) (Conradie, Verburgt *et al.*, 2018) and all pertain a similarity in their distribution between the inselbergs of Malawi and Mozambique. It was in Blackburn (2008b) that *A. francei* was first sampled, but only represented by populations from the type locality in Mulanje, Malawi. The present study provides evidence that there is extensive variation in intra-species divergence of *A. francei* across the known populations. It was in Blackburn (2008b) that *A. francei* was first sampled, but only represented by populations from the type locality in Mulanje, Malawi. The present study provides evidence that there is extensive variation in intra-species divergence of *A. francei* across the known populations.

Research on how much molecular difference constitutes whether a population can be considered a new species or not has been looked at extensively (Fregin *et al.*, 2012). Pairwise distances in *16S*rRNA among populations of over 5% provides a good support for thinking they represent distinct and separate lineages. For *A. francei*, we see almost all populations >5% which suggests a strong possibility of multiple cryptic species. While these results are the first to document the phylogenetic relationships among clades within *A. francei.* Nuclear loci retain ancestral polymorphisms, particularly in more recently divergent species. With the species delimitation model, Mount Mabu was intensely sampled whilst all other locations had a low effort with Mount Chiperone and Socone with only one specimen for each. These phylogenetic relationships give an interesting insight into the biogeographic patterns of *Arthroleptis* specimens on the SMI. The genetic divergence suggests that these populations were once connected by a widespread continuous forest which are now restricted to higher altitudes atop these isolated inselbergs. The significant population-level genetic divergence of *Arthroleptis* candidate species and the fact that they are specialist species in regard to habitat and breeding mode could be linked to their limited dispersal abilities. Given the limited number of biological surveys in the SMI region and the vast expanse across which the inselbergs are distributed, there is a possibility that more species await discovery. Further sampling will determine the support for the addition of new taxa.

4.3. Biogeographic Distribution

The phylogenetic relationships of the taxa in this study, along with those of other species (Conradie, Bittencourt-Silva *et al.*, 2018), help build an understanding of biogeographic distributions in Mozambique and Malawi. The SDM suggests that populations of *A. francei* might be present on further mountains than those investigated in this study (Figure 19), as well as at other locations of surveyed mountains. There is still a lot more to be discovered on the distribution of forest-restricted taxa. During field work for the present study, the calls of *A. francei* were heard on Mount Ribáuè, and only due to time constraints we were unable locate any individuals. This however confirms the predictions of the SDM that Mount Ribáuè is a suitable habitat for *A. francei*. According to the SDM, Mount Chiperone was also a potential site to find *A. francei*, and a specimen at NHM was indeed confirmed to be *A. francei* and included in the *16S* phylogenetic analysis. Further expeditions to the scattered inselbergs in this mountain range in Mozambique, in particular Mounts Namuli and Mabu and other parts of Mount Mulanje, Zomba mountains are needed to compile a more comprehensive distribution map of *A. francei*. An individual (BM 93.10.26.80) was collected allegedly from the Zomba region in the nineteenth century with no exact locality and was later rejected as *A. francei* by Loveridge (1953). Although this specimen is macerated, the description does resemble that being *A. francei* yet Loveridge leaves no explanation why the location was rejected and that the specimen being *A. macrodactyla*, a synonym (female) of *A. poecilonotus* (male) from central Africa. This however does not rule out the fact that the Zomba mountains are unsuitable for the *A. francei* complex, and further exploration here would be beneficial.

Further sampling effort across Mount Namuli and Mount Mulanje will confirm the predictions of 0.747% – 0.99% likelihood of occurrence in such locations (see Figure 19), similar values compared to Mount Chiperone, Socone and Lico where *A. francei* was recorded. Care however needs to be taken when interpreting the model, as the central area between the Zomba Mountain and the southern tip to the east of Mount Chiperone (Figure 19) represents a low elevation, non-forested area which likely would not be suitable. The SDM in this study utilised climatic variables which do not account for areas that might have previously been forest but due to anthropogenic change no

longer have forest habitats. The SDM should not be used entirely on its own when predicting occurrences but used as a guide in conjunction with ground-truthing field surveys and current forest distribution maps. As all occurrences except Mount Mabu (n = 24) are based on a sample size of less than ten records, the inferences made should only be taken as a guidance.

4.4. Taxanomic issues

This genetic analysis reveals that the populations on the Mozambiquean sky island forests represent candidate species. The Mount Inago clade is well supported as distinct, and is differentiated from the collective other mountain clades in the barcoding (BI) analysis. However, the clade containg individuals from Mounts Namuli, Socone is only moderately supported as monophyletic with a low 0 – 0.24 Bayesian posterior probability (Bpp). There is a further divergence within this larger clade with a 0.71 Bpp forming the smaller Mounts Mabu, Lico and Chiperone clade with a further 0.99 Bpp dividing Mount Mabu and jointly Mounts Lico and Chiperone. The largest Mount Mabu clade has a 0.99 Bpp from the type location, Mount Mulanje. Within the *rag-1* analysis, there is a common allele shared with the smaller Mount Mabu and Mount Lico clades unlike those within the mitochondrial *16S* analysis. This *rag-1* allele could be the result of being ancestory stored within the larger clade giving that *16S* haplotypes are extremely divergent. Currently, with the *16S* mitochondrial divergence and barcoding results and the species delimitation of the *16S* all lead to supporting the recognition of six distinct lineages (Figure 13).

The species delimitation method is a further analysis to understand divergence by grouping individuals into putative lineages. This method does not test whether the identified groups are independently evolving to one another. This analysis in conjunction with the understanding of any reproductive barriers can give clarity in understanding the species boundaries. As these populations are forest restricted species, the habitat fragmentation between the mountains and the vast distances would give confidence in accepting the results of species delimitation. Mount Socone and Mount Lico have the least distance between them within the whole mountain range

however, according to Figure 18, there is an obvious divergence between the two clades.

The phylogenetic relationships revealed for *A. francei* resemble other Sub-Saharan African taxa found within *Nothophryne* (Bittencourt-Silva *et al.*, 2016; Conradie, Bittencourt-Silva, *et al.*, 2018), *Dipsadaboa* (Branch *et al.*, 2019) and *Hyperolius* (Conradie, Verburgt, *et al.*, 2018), representing groups of closely related or identical species found atop mountain isolates. A thorough molecular analysis by Bittencourt-Silva *et al.* (2016) on the specialist *Nothophryne* species revealed a substantial genetic divergence between populations restricted to isolated moist habitats in the sky islands in Mozambique and Malawi. A similar pattern can also be seen in the *Arthroleptis* candidate species within this study (Figure 23). The mountain chain between Mount Mulajne and Mount Ribáuè spans across approximately 464 km, with distances between mountains ranging between approximately 63 km (Mount Mulanje to Mount Chiperone) and 14 km (Mount Socone and Mount Lico). Forest dominated this landscape in the late Oligocene-Miocene eras (Sepulchre *et al.*, 2006), providing a suitable habitat for *Arthroleptis* to disperse and inhabit a wider range and connecting the populations, perhaps at all elevations. Climatic fluctuations have played a large factor in the ever expanding and contracting of forest habitats in similar environments (Lawson *et al.*, 2015), and this physical process could be the influence of the distribution of these *Arthroleptis* candidate species. More recently, aridification and anthropogenic changes have resulted in further habitat fragmentation, therefore restricting these forest dwellers to isolated populations in higher elevations. Further paleo-geographical data will aid in the understanding of the biogeography and biodiversity of these SMI.

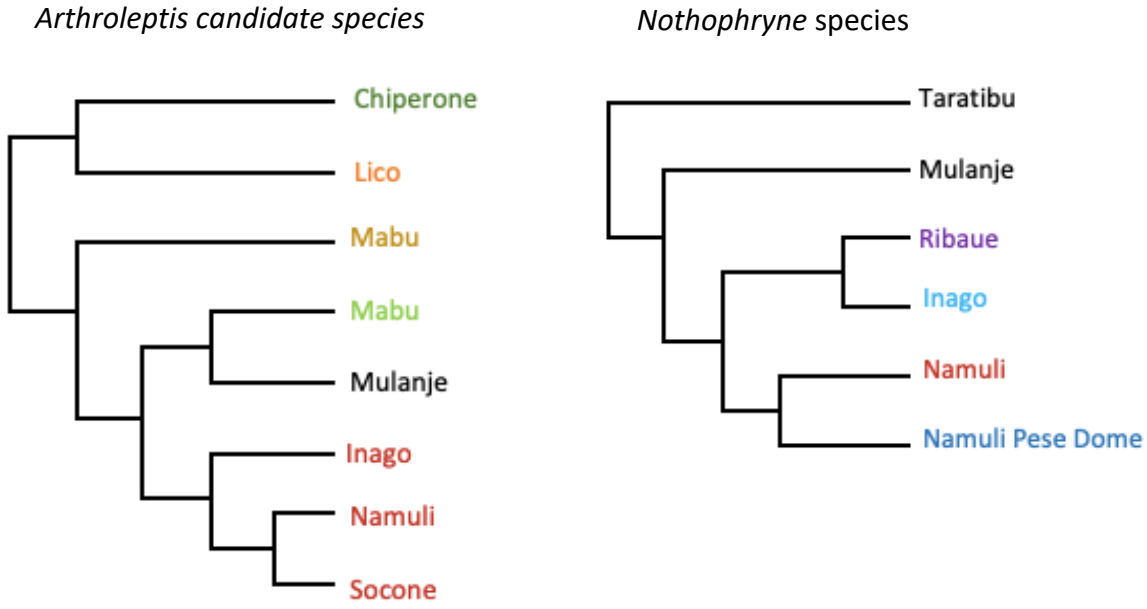


Figure 23. A simplified comparison between the SDM of Arthroleptis *candidate species in this study and* Nothophryne *from Bittencourt-Silva (2016).*

I understand that this study did not include a large morphological analysis due to the availability of specimens however this would only complement the phylogenetic results which strongly support my decision to recognise the cryptic species within *Arthroleptis francei*. In addition, the call analysis is an addition as with the morphological analysis. This was chosen to be included in this study as a result of a personal discussion with Werner Conradie and Gabriella Bittencourt on the field trip in November 2018. The reason being due to there being limited studies in this region on amphibian bioacoustics and it will be beneficial to build up a database for future research. This being the hypothesis that there could be slight call differentiations between the newly diverged species. Additional collection of specimens from locations where the sampling effort was poor such as Mounts Inago, Socone and Chiperone would help in a further phylogenetic analysis along with expeditions to mountain presented in the SDM model. Behavioural studies would help in understanding the ecologies of this species and inparticular, their breeding habits, habitat use, colour morphologies and the differences between sexes.

4.5. Speciation

The predominant divergence of populations found amongst the SMI suggests that mountain formations and retracting forest act as a significant barrier for dispersal. Without dated divergence, we can only assume the current driving forces contributing to the speciation of the candidate taxa within *Arthroleptis francei*. The molecular analyses uncovered that speciation events have likely occurred on each mountain. There is evidence for three allopatric speciation events between Mount Lico and Mount Chiperone, between Mount Inago, Mount Socone and Mount Namuli and between Mount Mulanje. Interestingly, the individuals on Mount Mabu indicate potential sympatric speciation as they were located in the same geographic coordinates. Further research will help in understanding if the two Mount Mabu taxa are evolving independently, and if niche separation allows the separate Mabu clades to co-exist in the same location. There could be a possibility of the species from the Mabu clades being in secondary contact with each other and had diverged earlier. These patterns of spatial segregation are associated with adaptations to the environments and limited or lack of gene flow. Ancestral species accumulate independent genetic changes (mutations) through time, often after the populations have become reproductively isolated.

With the forest habitat retreating on each mountain, it would be beneficial to map the last remaining forests to monitor and predict the rate of loss of species. As the *Arthroleptis* candidate species are known to inhabit forest isolates, the minimum range between the points of collection to the maximum elevation on Mount Mabu and Mount Mulanje are 56m and 1292m, respectively (Appendix 6b). Only one individual from Mount Mabu was located at 430m asl, which appears to be an outlier whilst all the other individuals were collected at elevations above 844m. These species are specialists and rely on higher, forested habitats. This reliance links to the potential conservation implications and restricted gene flow of already isolated populations.

Our knowledge of amphibian diversification mechanisms on montane isolates in the EABR is limited. Breviceptid frogs appear to have arisen from allopatric speciation within forest and grassland habitats (Loader *at al.* 2014), whereas for spiny-throated

reed frogs' divergence predominantly through a combination of peripatric and allopatric speciation resulted in an increase of species richness (Lawson *et al.*, 2015). Within the habitats on the SMI, there are high concentrations of old and new endemics with fragmented distributions. An in-depth analysis of speciation amongst amphibians in these habitats is lacking therefore further studies on an individual species or community approach are necessary to clarify speciation mechanisms, and to gain a broader knowledge of its rich biota. From this study, it seems there are two main speciation processes occurring, yet a further analysis would help in clarifying this. Knowing more about the drivers of these events would be beneficial to the conservation of this species complex. If elevation is the driving factor, taxa at higher elevations may be prone to genetic bottlenecks and reduced habitat ranges, which puts the survival of the species at risk. Fragmentation, habitat destruction and climate change pose major threats and force slightly lower elevation species to even higher elevations, where extinctions and speciation events might occur at a faster rate.

4.6. Conservation

This multi-analysis approach of phylogenetic, spatial data and distribution modelling with the discovery of cryptic species results in an in-depth insight into a species complex and confirms where efforts for conservation are needed. These montane isolates are known to harbour a rich diversity of species, and the *A. francei* complex adds to the need to increase the protection of these critical habitats. According to the CEPF and the world database of protected areas, only 50% of these mountains within this study are protected, with a greater investment into Mounts Mulanje, Mabu and Namuli. Other mountains such as Lico and Chiperone have proved to be endemic hotspots, and Mount Morrumbala remains to be surveyed. The phylospatial modelling highlights areas of conservation importance, but there are certainly gaps in other areas which need a species-wide assessment and protection planning. Protected areas in this study region are few, and with the rapid loss of these forests perhaps strengthening the protection of areas that fall within the PE hotspots will be a strategic move.

With molecular-based analyses showing a genetic difference between lineages, it is evident there are multiple species distributed across the isolated inselbergs. The likely

presence of cryptic species within the *A. francei* complex will affect the future species' conservation status on the IUCN red list, where it is currently listed as Vulnerable. The revision of this status would result in a possible elevation to Endangered due to a more restricted range of *A. francei*, jointly with an independent assessment of newly emerging taxa. As these species are currently known to only occur on single mountain blocks each, they effectively have a narrow range. Coupled with this range, their specific habitat needs, and breeding biology suggest that any changes to the habitat will be detrimental to the species. The isolated forested sky island inselbergs are an important ecosystem in Africa, providing a vital habitat for many species of flora and fauna. These forests are disappearing entirely or have reduced in size due to the local practices of slash and burn for agriculture purposes (Figure 24). These practices are generally conducted on the outskirts of the forest with a gradual encroachment towards the centre (Ryan *et al.*, 2014). Forests are also used for the extraction of trees for firewood and bushmeat hunting, where fires are set to capture animals and gin traps can be found for trapping (personal observations). The surrounding boarder of the forests are miombo woodlands which act as a buffer zone but are too being lost for firewood.

Forests are known to be vital terrestrial habitats that act as a carbon sink (Pan *et al.*, 2011). A study by Bayliss *et al.* (2014) on the carbon storage of the forest of Mount Mabu resulted in a value of 3 634 539 Mg (3.6 Tg), including the live vegetation, woody debris, soil, and the below-ground live matter. If the total forest was lost to bushland and agriculture with sparse crops, 117 Mg ha^{-1} or 2.7 Tg of carbon in total will be released into the atmosphere, reducing to 2.2 Tg of carbon If converted into agriculture combined with woodland (Willcock *et al.*, 2012). The forested habitat on the surrounding mountains in Southern Malawi and Northern Mozambique (north of the Zambezi river) show similar ecological characteristics, underlining their importance for carbon sequestration and as a habitat for people and animals alike.

Figure 24. Deforested forest which has been converted into farmland (currently maize) with a temporary thatched hut for shelter when the land is being managed. This location was a five hour walk from the nearest village which once was thick forest as seen by the expedition two years prior. Photo: Woest (2019).

The increased interest in research on Mount Mulanje lead to the formation of the non-governmental Mulanje Mountain Conservation Trust (Bayliss *et al*., 2014), with the aim to, as a joint venture between funding and educational organisations, initiate an ongoing biodiversity assessment also for other inselbergs of Mozambique. The presentation of data collected by Julian Bayliss and the research expedition team where a number of new specimens were discovered has indeed resulted in the protection of Mount Mabu by the government of Mozambique (Windsor, 2009). The discovery of new species, and the gaining of new knowledge of already known species has more recently resulted in the protection of Mount Ribáuè, where any person caught deforesting the last remaining forest will receive a monetary fine and imprisonment (pers. comm.). With enforcement being low and these anthropogenic processes, this SMI region is a cause for conservation concern. Government co-operation and community-based projects on sustainable farming practices and reforestation would be a vital investment to the protection of these forests, and to protect the biodiversity it harbours.

5. Conclusion

The phylogenetic results from this study display diversity amongst *Arthroleptis francei* distributed on different mountains, including new candidate species from southern montane sky islands. There is a substantial amount of genetic variation between populations, with a similar pattern arising between the three genes and the delimitating approaches (BI model and bGMYC). In total, *A. francei* appears to comprise of six to seven candidate species, which are mostly endemic (Mounts Mabu, Chiperone, Lico, Namuli and Mulanje) and isolated to single mountains except for individuals from Mount Socone which are grouped with Mount Namuli and Mount Mabu, which harbours two taxa. According to species delimitation methods, populations from Mount Namuli, Mount Socone and Mount Inago are diverged but belong to the same taxon, whereas populations from the remainder of the mountains display divergence at the level of species. The overall distribution of *Arthroleptis* candidate species is determined by vicariance due to geological and climatic events and, more recently, anthropogenic land use changes, in particular in Malawi and to the east of Mount Mabu. With more intense sampling effort, understanding the complex nature of *Arthroleptis francei* will be better understood in the future. The present work is an insight into a both phenotypically and genetically diverging species endemic to isolated small areas of the last remaining rainforest pockets in East Africa. The integrative approach in this study merits a thorough species delimitation, providing a first step to uncover previously unrecognised diversity. This mixed method approach is applicable to discovering the distributions of biodiversity within any given area and scale and the phylospatial analysis can include cryptic diversity. This allows a broad approach to identifying critical areas and the prioritisation of conservation efforts.

6. References

Abascal, F., Zardoya, R., Telford, M. (2010) 'TranslatorX : multiple alignment of nucleotide sequences guided by amino acid translations', *Nucleic Acids Research*, 38(April), pp. 7–13. doi: 10.1093/nar/gkq291.

Adole, T., Dash, J. and Atkinson, P. M. (2018) 'Large-scale prerain vegetation green-up across Africa', *Global Change Biology*, 24(9), pp. 4054–4068. doi: 10.1111/gcb.14310.

Aerts, R. and Honnay, O. (2011) 'Forest restoration, biodiversity and ecosystem functioning', *BMC Ecology*. doi: 10.1186/1472-6785-11-29.

Altig, R. and Johnston, G. F. (1989) 'Guilds of Anuran Larvae: Relationships among Developmental Modes, Morphologies, and Habitats', *Herpetological Monographs*, 3, pp. 89–109. doi: 10.2307/1466987.

AmphibiaWeb. (2019) *<https://amphibiaweb.org> University of California, Berkeley, CA, USA.*

Bagaturov, M. F. (2018) 'Overview of Amphibian Sexual Dimorphism, with Description of New Secondary Sexual Dimorphic Character in Rhacophorid Frogs', *International Journal of Zoology and Animal Biology*, 1(1), p. 105. doi: 10.23880/izab-16000105.

Baldwin, B. G. and Sanderson, M. J. (1998) 'Age and rate of diversification of the Hawaiian silversword alliance (Compositae)', *Proceedings of the National Academy of Sciences of the United States of America*, 95, pp. 9402–9406. doi: 10.1073/pnas.95.16.9402.

Barratt, C. D. *et al.* (2017) 'Environmental correlates of phylogenetic endemism in amphibians and the conservation of refugia in the Coastal Forests of Eastern Africa', *Diversity and Distributions*, 23(8), pp. 875–887. doi: 10.1111/ddi.12582.

Barratt, C. D. *et al.* (2018) 'Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians', *Molecular Ecology*, 27, pp. 4289–4308. doi: 10.1111/mec.14862.

Bauer, A. M. and Lamb, T. (2005) 'Phylogenetic relationships of southern african geckos in the pachydactylus group (squamata: Gekkonidae)', *Journal of the Herpetological Association of Africa*, 52, pp. 105–129. doi: 10.1080/21564574.2005.9635525.

Bayliss, J. *et al.* (2014) 'The discovery, biodiversity and conservation of Mabu forest—the largest medium-altitude rainforest in southern Africa', *Oryx*, 48(2), pp. 177–185. doi: 10.1017/s0030605313000720.

Van Berkel, T. *et al.* (2019) 'A Mammal Survey of the Serra Jeci Mountain Range, Mozambique, with a Review of Records from Northern Mozambique's Inselbergs', *African Zoology*, 54(1), pp. 31–42. doi: 10.1080/15627020.2019.1583081.

Bittencourt-Silva, G. B. *et al.* (2016) 'The phylogenetic position and diversity of the enigmatic

mongrel frog Nothophryne Poynton, 1963 (Amphibia, Anura)', *Molecular Phylogenetics and Evolution*, 99, pp. 89–102. doi: 10.1016/j.ympev.2016.03.021.

Bittencourt-Silva, G., Langerman, D. and Tolley, K. (2020) 'Why the long finger? Observation of male–male combat in African bush squeaker frogs, Arthroleptis stenodactylus (Anura: Arthroleptidae)', *Herpetological Bulletin*, (151), p. 45. doi: 10.33256/hb151.45.

Blackburn, D (2008a) 'A new species of *Cardioglossa* (Amphibia: Anura: Arthroleptidae) endemic to Mount Manengouba in the Republic of Cameroon, with an analysisof morphological diversity in the genus.', *Zoological Journalof the Linnean Society*, 154, pp. 611–630.

Blackburn, D. (2008b) 'Molecular phylogenetics and evolution biogeography and evolution of body size and life history of African frogs : Phylogeny of squeakers (*Arthroleptis*) and long-fingered frogs (*Cardioglossa*) estimated from mitochondrial data', *Molecular Phylogenetics and Evolution*. Elsevier Inc., 49(3), pp. 116–136. doi: 10.1016/j.ympev.2008.08.015.

Blackburn, D. (2009a) 'Description and phylogenetic relationships of two new species of miniature *Arthroleptis* (Anura: Arthroleptidae) from the Eastern Arc Mountains of Tanzania', *Breviora*, 517(1), pp. 1–17. doi: 10.3099/0006-9698-517.1.1.

Blackburn, D. (2009b) 'Diversity and evolution of male secondary sexual characters in African squeakers and long-fingered frogs', *Biological Journal of the Linnean Society*, 96, pp. 553–573. doi: 10.1111/j.1095-8312.2008.01138.x.

Blackburn, D. (2012) 'New species of *Arthroleptis* (Anura: Arthroleptidae) from Ngozi Crater in the Poroto Mountains of Southwestern Tanzania', *Journal of Herpetology*, 46, pp. 129–135. doi: 10.1670/10-322.

Blackburn, D. *et al.* (2013) 'An adaptive radiation of frogs in a southeast asian island archipelago', *Evolution*, 67(9), pp. 2631–2646. doi: 10.1111/evo.12145.

Blackburn, D., Gvoždík, V. and Leaché, A. D. (2010) 'A new squeaker Frog (Arthroleptidae: *Arthroleptis)* from the Mountains of Cameroon and Nigeria', *Herpetologica*, 66, pp. 335–348. doi: 10.1655/herpetologica-d-10-00015.1.

Blackburn, D. and Moreau, C. S. (2006) 'Ontogenetic diet change in the arthroleptid frog *Schoutedenella xenodactyloides*', *Journal of Herpetology*, 40(3), pp. 388–39. doi: 10.1670/0022-1511(2006)40[388:odcita]2.0.co;2.

Boersma, P. and Weenink, D. (2019) 'Praat: Doing Phonetics by Computer', *Accessed August 2019*. doi: 10.1097/aud.0b013e31821473f7.

Booth, T. H. *et al.* (2014) 'Bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies', *Diversity and Distributions*,

20, pp. 1–9. doi: 10.1111/ddi.12144.

Bossuyt, F. and Milinkovitch, M. C. (2000) 'Convergent adaptive radiations in Madagascan and Asian ranid frogs reveal covariation between larval and adult traits', *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp. 6585–6590. doi: 10.1073/pnas.97.12.6585.

Branch, W. R. *et al.* (2019) 'A new species of tree snake (*Dipsadoboa*, Serpentes: Colubridae) from "sky island" forests in northern Mozambique, with notes on other members of the Dipsadoboa werneri group', *Zootaxa*, 4646(3), pp. 541–563. doi: 10.11646/zootaxa.4646.3.6.

Brito, D. (2010) 'Overcoming the Linnean shortfall : Data deficiency and biological survey priorities', *Basic and Applied Ecology*. Elsevier GmbH, 11(8), pp. 709–713. doi: 10.1016/j.baae.2010.09.007.

Brown, J. H. and Lomolino, M. (1998) *Biogeography*. 2nd ed., *Sunderland, MA:Sinauer Press*. 2nd ed.

Campagna, L. *et al.* (2015) 'Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation', *Molecular Ecology*, 24, pp. 4238–4251. doi: 10.1111/mec.13314.

Caro, T. (2005) 'The adaptive significance of coloration in mammals', *BioScience*, 55, pp. 125–136. doi: 10.1641/0006-3568(2005)055[0125:tasoci]2.0.co;2.

Carstensen, D. W. *et al.* (2013) 'Introducing the biogeographic species pool', *Ecography*, 36, pp. 1–9. doi: 10.1111/j.1600-0587.2013.00329.x.

Ceccarelli, F. S. *et al.* (2014) 'Evolutionary relationships, species delimitation and biogeography of Eastern Afromontane horned chameleons (Chamaeleonidae: Trioceros)', *Molecular Phylogenetics and Evolution* 80(*1*), pp. 125–136. doi: 10.1016/j.ympev.2014.07.023.

Clement, M. *et al.* (2002) 'TCS: Estimating gene genealogies', in *Proceedings - International Parallel and Distributed Processing Symposium, IPDPS 2002*, p. 184. doi: 10.1109/IPDPS.2002.1016585.

Conradie, W. *et al.* (2016) 'Exploration into the hidden world of Mozambique's sky island forests: new discoveries of reptiles and amphibians', *Zoosystematics and Evolution*, 92(2), pp. 163–180. doi: 10.3897/zse.92.9948.

Conradie, W., Verburgt, L., *et al.* (2018) 'A new Reed Frog (Hyperoliidae: Hyperolius) from coastal northeastern Mozambique', *Zootaxa*, 4379, pp. 177–198. doi: 10.11646/zootaxa.4379.2.2.

Conradie, W., Bittencourt-Silva, G. B., *et al.* (2018) 'New species of Mongrel Frogs

(Pyxicephalidae: Nothophryne) for northern Mozambique inselbergs', *African Journal of Herpetology*. doi: 10.1080/21564574.2017.1376714.

Courtois, E. A. *et al.* (2016) 'Taking the lead on climate change: modelling and monitoring the fate of an Amazonian frog', *Oryx*, 50(3), pp. 450–459. doi: 10.1017/s0030605315000083.

Couzens, A. M. C. and Prideaux, G. J. (2018) 'Rapid Pliocene adaptive radiation of modern kangaroos', *Science*, 362(6410), pp. 72–75. doi: 10.1126/science.aas8788.

Crisp *et al.* (2001) 'Endemism in the Australian flora', *Journal of Biogeography*, 28, pp. 183–198. doi: 10.1046/j.1365-2699.2001.00524.x.

Cushman, S. A. (2006) 'Effects of habitat loss and fragmentation on amphibians: A review and prospectus', *Biological Conservation*, 128, pp. 231–240. doi: 10.1016/j.biocon.2005.09.031.

Darwin, C. (1879) *The Descent of Man, and Selection in Relation to Sex*. 2nd. London: Penguin Classics. doi: 10.1038/011305a0.

Davies, T. J. (2015) 'Losing history : how extinctions prune features from the tree of life', *Philisophical Transactions*, 370(1), p. 20140006. doi: http://dx.doi.org/10.1098/rstb.2014.0006.

Dowsett-Lemaire, F. (2010) 'Further ornithological exploration of Namuli and Mabu Mountains (northern Mozambique), and the urgent need to conserve their forests.', *African Bird Club Bulletin*, 1(17), pp. 159–177.

Drummond, A. J. *et al.* (2006) 'Relaxed phylogenetics and dating with confidence', *PLoS Biology*, 4088, p. https://doi.org/10.1371/journalpbio0040088. doi: 10.1371/journal.pbio.0040088.

Drummond, A. J., Rambaut, A. and Marc, S. (2016) 'Bayesian Evolutionary Analysis Sampling Trees', in *Bayesian Evolutionary Analysis with BEAST 1.8.3*. doi: 10.1017/CBO9781139095112.007.

Duellman, W. E. (1993) *Amphibian Species of the World: Additions and Corrections*, *University of Kansas Museum of Natural History. Special Publication.* Kansas: Lawrence : University of Kansas.

Duellman, W. E. and Trueb, L. (1994) *Biology of Amphibians*, *The Quarterly Review of Biology*. New York.: McGraw-Hill Publishing Company. doi: 10.1086/418798.

Dupont-Nivet, G. *et al.* (2007) 'Tibetan plateau aridification linked to global cooling at the Eocene-Oligocene transition', *Nature*, 455(635), p. 638. doi: 10.1038/nature05516.

Eardley, C. D., Gikungu, M. and Schwarz, M. P. (2009) 'Bee conservation in Sub-Saharan Africa

and Madagascar: diversity, status and threats', *Apidologie*, 40(3), pp. 355–366. doi: 10.1051/apido/2009016.

Edwards, D. P. *et al.* (2014) 'Mining and the African environment', *Conservation Letters*. doi: 10.1111/conl.12076.

Ehlers Smith, Y. C. *et al.* (2018) 'Forest habitats in a mixed urban-agriculture mosaic landscape: patterns of mammal occupancy', *Landscape Ecology*. doi: 10.1007/s10980-017-0580-1.

Emerson, S. B. (1994) 'Testing pattern predictions of sexual selection: A frog example', *American Naturalist*, 143, pp. 848–869. doi: 10.1086/285636.

Ernst, R., Agyei, A. C. and Rö̈del M. (2008) 'A new giant species of Arthroleptis (Amphibia: Anura: Arthroleptidae) from the Krokosua Hills Forest Reserve, south-western Ghana', *Zootaxa*, 1697, pp. 58–68.

Excoffier, L., Smouse, P. E. and Quattro, J. M. (1992) 'Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data', *Genetics*, 131, pp. 479–491.

Fairbairn, D. J., Blanckenhorn, W. U. and Székely, T. (2007) *Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism*, *Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism*. doi: 10.1093/acprof:oso/9780199208784.001.0001.

Faith, D. P. (1992) 'Conservation evaluation and phylogenetic diversity', *Biological Conservation*, 61, pp. 1–10. doi: 10.1016/0006-3207(92)91201-3.

FAO (2016) *State of the World's Forests 2016. Forests and agriculture: land-use challenges and opportunities*, *State of the world's forests and agriculture:Land use challenges and opportunities.* doi: 10.1146/annurev-environ-020411-130608.

Franco, A. *et al.* (2008) *Chapter 5. Threatened Amphibians of the Afrotropical Realm*. 1st edn, *Threatened Amphibians of the World*. 1st edn. Edited by S. Stuart et al. Barcelona: Lynx Edicions.

Fregin, S. *et al.* (2012) 'Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae', *Molecular Phylogenetics and Evolution*, 62, pp. 319–328. doi: 10.1016/j.ympev.2011.10.003.

Frost, D. R. *et al.* (2006) 'The amphibian tree of life', *Bulletin of the American Museum of Natural History*, 297, pp. 1–291. doi: 10.1206/0003-0090(2006)297[0001:tatol]2.0.co;2.

Frost, D. R. (2019) *Amphibian Species of the World: an Online Reference.*, *Version 6.0*.

Funk, W. C. *et al.* (2005) 'High dispersal in a frog species suggests that it is vulnerable to habitat fragmentation', *Biology Letters*, 1(1), pp. 13–16. doi: 10.1098/rsbl.2004.0270.

Garcia, T. S., Stacy, J. and Sih, A. (2004) 'Larval salamander response to UV radiation and predation risk: Color change and microhabitat use', *Ecological Applications*, 14, pp. 1055–1064. doi: 10.1890/02-5288.

Gernhard, T. (2008) 'The conditioned reconstructed process', *Journal of Theoretical Biology*, 253, pp. 769–778. doi: 10.1016/j.jtbi.2008.04.005.

Glor, R. E. (2010) 'Phylogenetic insights on adaptive radiation', *Annual Review of Ecology, Evolution, and Systematics*, 41, pp. 251–270. doi: 10.1146/annurev.ecolsys.39.110707.173447.

Goebel, A. M., Donnelly, J. M. and Atz, M. E. (1999) 'PCR primers and amplification methods for *12S* ribosomal DNA, the control region, cytochrome oxidase I, and cytochromebin bufonids and other frogs, and an overview of PCR primers which have amplified DNA in amphibians successfully', *Molecular Phylogenetics and Evolution*, 11(1), pp. 163–199. doi: 10.1006/mpev.1998.0538.

Gomez, D. *et al.* (2009) 'The role of nocturnal vision in mate choice: Females prefer conspicuous males in the European tree frog (*Hyla arbore*a)', *Proceedings of the Royal Society B: Biological Sciences*, 276, pp. 2351–2358. doi: 10.1098/rspb.2009.0168.

Gomez, D. *et al.* (2011) 'Multimodal signals in male European treefrog (Hyla arborea) and the influence of population isolation on signal expression', *Biological Journal of the Linnean Society*, 103, pp. 633–647. doi: 10.1111/j.1095-8312.2011.01662.x.

Gordon, I. *et al.* (2012) *Biodiversity Hotspots: Eastern Afromontane*. Available at: http://www.biodiversityhotspots.org/xp/hotspots/afromontane/Pages/default.aspx.

Grant, P. and Grant, B. R. (2002) 'Adaptive radiation of Darwin's Finches', *American Scientist*, 90(2), pp. 130–139. doi: 10.1511/2002.2.130.

Green, D. M. (2003) 'The ecology of extinction: Population fluctuation and decline in amphibians', *Biological Conservation*, 111, pp. 331–343. doi: 10.1016/S0006-3207(02)00302-6.

Griffiths, C. J. (2011) 'The geological evolution of East Africa', in Lovett, J. C. and Wasser, S. K. (eds) *Biogeography and ecology of the rain forests of eastern Africa*. 1st edn. Cambridge, UK: Cambridge University Press, pp. 9–21. doi: 10.1017/cbo9780511895692.002.

Gupta, A. K. *et al.* (2004) 'Indian Ocean high-productivity event (10-8 Ma): Linked to global cooling or to the initiation of the Indian monsoons?', *Geology*, 32, pp. 753–756. doi: 10.1130/G20662.1.

Halliday, T. and Verrell, P. (1986) 'Sexual selection and body size in amphibians', *Herpetological*

Journal, 1, pp. 86–92.

Harcourt, A. H. and Wood, M. A. (2012) 'Rivers as barriers to primate distributions in Africa', *International Journal of Primatology*, 33, pp. 168–183. doi: 10.1007/s10764-011-9558-z.

Hasegawa, M., Kishino, H. and Yano, T. aki (1985) 'Dating of the human-ape splitting by a molecular clock of mitochondrial DNA', *Journal of Molecular Evolution*, 22(2), pp. 160–174. doi: 10.1007/BF02101694.

Herrmann, H. *et al.* (2004) 'A new frog species of the genus *Cardioglossa* from the Tchabal Mbabo Mtns, Cameroon', *Herpetozoa*, 17, pp. 119–125.

Hijmans, R. J. and Elith, J. (2017) 'Species distribution modeling with R Introduction', *R Cran Tutorial*, pp. 1–78.

Hoffman, E. A. and Blouin, M. S. (2000) 'A review of colour and pattern polymorphisms in anurans', *Biological Journal of the Linnean Society*, 70, pp. 633–665. doi: 10.1006/bijl.1999.0421.

Hoffman, M. *et al.* (2016) 'Hotspots (Version 2016.1)', *Zenodo*. doi: http://doi.org/10.5281/zenodo.3261807.

Huggett, R. J. (2004) *Fundamentals of Biogeography*, *Fundamentals of Biogeography*. doi: 10.4324/9780203356586.

IUCN (2018) *The IUCN Red List of Threatened Species*. Available at: www.iucnredlist.org (Accessed: 17 August 2018).

IUCN (2019) *Habitat types in East Africa.* Available at: https://www.iucnredlist.org/search/stats?habitats=4&searchType=species (Accessed: 22 May 2019).

IUCNa. (2019) *The IUCN Red List of Threatened Species. Global Amphibian Statistics*. Available at: https://www.iucnredlist.org (Accessed: 29 March 2019).

Jacobs, L. E. *et al.* (2017) 'Local not vocal: Assortative female choice in divergent populations of red-eyed treefrogs, Agalychnis callidryas (Hylidae: Phyllomedusinae)', *Biological Journal of the Linnean Society*, 120, pp. 170–178. doi: 10.1111/bij.12861.

Jongsma, G. F. M. *et al.* (2018) 'Diversity and biogeography of frogs in the genus *Amnirana* (Anura: Ranidae) across sub-Saharan Africa', *Molecular Phylogenetics and Evolution*, 120(April 2017), pp. 274–285. doi: 10.1016/j.ympev.2017.12.006.

Kappelman, J. *et al.* (2003) 'Oligocene mammals from Ethiopia and faunal exchange between Afro-Arabia and Eurasia', *Nature*, 426, pp. 549–552. doi: 10.1038/nature02102.

Kaspar, F., Prömmel, K. and Cubasch, U. (2010) 'Impacts of tectonic and orbital forcing on East

African climate: A comparison based on global climate model simulations', *International Journal of Earth Sciences*, 99(7), pp. 1677–1686. doi: 10.1007/s00531-010-0538-x.

Katoh, K., Kuma, K., Miyata, T. (2002) 'MAFFT v7.017: a novel method for rapid multiple sequence alignment based on fast Fourier transform.', *Nucleic Acids Research*, pp. 3059–3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A. (2012) 'Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.', *Bioinformatics*, pp. 1647–1649. doi: http://dx.doi.org/10.1093/bioinformatics/ bts199.

Kissling, W. D. *et al.* (2012) 'Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide', *Proceedings of the National Academy of Sciences*, 109, pp. 7379–7384. doi: 10.1073/pnas.1120467109.

Kocher, T. D. *et al.* (1989) 'Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers', *Proceedings of the National Academy of Sciences of the United States of America*, 86, pp. 6196–6200. doi: 10.1073/pnas.86.16.6196.

Köhler, J. *et al.* (2005) 'New amphibians and global conservation: A boost in species discoveries in a highly endangered vertebrate group', *BioScience*, 55, pp. 693–696. doi: 10.1641/0006-3568(2005)055[0693:naagca]2.0.co;2.

Kozak, K. H. *et al.* (2005) 'Phylogenetic analysis of ecomorphological divergence, community structure, and diversification rates in dusky salamanders (Plethodontidae: Desmognathus)', *Evolution*, 59(9), pp. 2000–2016. doi: 10.1111/j.0014-3820.2005.tb01069.x.

Krohn, A. R. and Rosenblum, E. B. (2016) 'Geographic color variation and physiological color change in Eastern collared lizards (Crotaphytus collaris) from Southern New Mexico, USA', *Herpetologica*, 72, pp. 318–323. doi: 10.1655/herpetologica-d-15-00074.1.

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35**:1547-1549.

Lachlan, R. F. (2007) 'Luscinia: a bioacoustics analysis computer program (Version 1.0)'.

Lanfear, R. *et al.* (2016) 'PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular biology and evolution.' doi: dx.doi.org/10.1093/molbev/msw260.

Laurent, R. (1954) 'Remarques sur le genre Schoute- denella Witte.', *Annales du Musée Royal du Congo Belge. Sciences Zoologiques*, 1, pp. 34–40.

Laurent, R. F. (1957) 'Notes sur les Hyperoliidae', *Revue de Zoologie et de Botanique Africaines. Tervuren*, 56, pp. 274–282.

Lawson, L. P. *et al.* (2015) 'Divergence at the edges: Peripatric isolation in the montane spiny throated reed frog complex', *BMC Evolutionary Biology*, 15, p. 128. doi: 10.1186/s12862-015-0384-3.

Lee-Thorp, J. A., Sponheimer, M. and Luyt, J. (2007) 'Tracking changing environments using stable carbon isotopes in fossil tooth enamel: an example from the South African hominin sites', *Journal of Human Evolution*, 53, pp. 595–601. doi: 10.1016/j.jhevol.2006.11.020.

Leigh, J. and Bryant, D. (2015) 'POPART : full-feature software for haplotype network construction', *Methods in Ecology and Evolution*, 6, pp. 1110–1116. doi: 10.1111/2041-210X.12410.

Liao, W. B. *et al.* (2013) 'Sexual size dimorphism in anurans fails to obey Rensch's rule', *Frontiers in Zoology*, 10. doi: 10.1186/1742-9994-10-10.

Loader, S. P. *et al.* (2014) 'Persistence and stability of Eastern Afromontane forests: Evidence from brevicipitid frogs', *Journal of Biogeography*, 41, pp. 1781–1792. doi: 10.1111/jbi.12331.

López, J. Á. R. *et al.* (2018) 'Mountain barriers and trans-Saharan connections shape the genetic structure of Pimelia darkling beetles (Coleoptera: *Tenebrionidae)*', *Biological Journal of the Linnean Society*, 124(3), pp. 547–556. doi: 10.1093/biolinnean/bly053.

Loveridge, A. (1953) 'Zoological results of a fifth expedition to East Africa. I-Mammals from Nyasaland and Tete', *Bulletin of the Museum of Comparative Zoology, Harvard*, 110, pp. 325–406.

Loveridge, A. (1957) 'Check List of the Reptiles and Amphibians of East Africa (Uganda; Kenya; Tanganyika; Zanzibar). Arthur Loveridge', *Bulletin of the Museum of Comparative Zoology*, 117(i-xxxvi.), pp. 153–362.

Lovett, J. C., Midgley, G. F. and Barnard, P. (2005) 'Climate change and ecology in Africa', *African Journal of Ecology*, 43, pp. 279–281. doi: 10.1111/j.1365-2028.2005.00598.x.

Malhi, Y. *et al.* (2013) 'African rainforests: Past, present and future', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1625), p. 20120312. doi: 10.1098/rstb.2012.0312.

Mayr, E. (1942) *Systematics and the origin of species from the viewpoint of a zoologist*, *1*. New

York, Columbia University Press.

Mayr, E. and O'Hara, R. J. (1986) 'The biogeographic evidence supporting the Peistocene forest refuge hypothesis', *Evolution*, 40(1), pp. 55–67. doi: 10.2307/2408603.

McDiarmid, R. W. and Altig, R. (1999) *Tadpoles*. Chicago: University of Chicago Press. doi: 10.1643/0045-8511(2000)000[1125:br]2.0.co;2.

Meijden, A. Van Der (2006) 'Molecular phylogeny and biogeography of ranoid frogs', *Unpublished*.

Mimouni, E. and Beisner, B. (2016) 'Phylogenetic diversity and its conservation in the presence of phylogenetic uncertainty : a case study of cladoceran communities in urban waterbodies', *Biodiversity and Conservation*, 25(11), pp. 2113–2136. doi: 10.1007/s10531-016-1181-z.

Minter, L. R., Netherlands, E. C. and Du Preez, L. H. (2017) 'Uncovering a hidden diversity: Two new species of *Breviceps* (Anura: Brevicipitidae) from northern KwaZulu-Natal, South Africa', *Zootaxa*, 4300, pp. 195–216. doi: 10.11646/zootaxa.4300.2.3.

Mittermeier, R. A. *et al.* (2004) 'Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions', *CEMEX*, p. 392.

Monnet, J. M. and Cherry, M. I. (2002) 'Sexual size dimorphism in anurans', *Proceedings of the Royal Society B: Biological Sciences*, 269, pp. 2301–2307. doi: 10.1098/rspb.2002.2170.

Mucova, S. A. R. *et al.* (2018) 'Assessment of land use and land cover changes from 1979 to 2017 and biodiversity & land management approach in Quirimbas National Park, Northern Mozambique, Africa', *Global Ecology and Conservation*, 16. doi: 10.1016/j.gecco.2018.e00447.

Napoli, M. F. (2006) 'A new species allied to *Hyla circumdata* (Anura: Hylidae) from Serra da Mantiqueira, Southeastern Brazil', *Herpetologica*, 61(1), pp. 63–69. doi: 10.1655/03-41.

Neumann, F. H. and Bamford, M. K. (2015) 'Shaping of modern southern African biomes: Neogene vegetation and climate changes', *Transactions of the Royal Society of South Africa*, 70, pp. 195–212. doi: 10.1080/0035919X.2015.1072859.

Neves, I. Q., Da Luz Mathias, M. and Bastos-Silveira, C. (2018) 'The terrestrial mammals of Mozambique: Integrating dispersed biodiversity data', *Bothalia*, 48(1), p. a2330. doi: 10.4102/abc.v48i1.2330.

Newbold, T. *et al.* (2017) 'The present and future effects of land use on ecological assemblages in tropical grasslands and savannas in Africa', *Oikos*. doi: 10.1111/oik.04338.

Noah, R. (2014) 'bGMYC: A Bayesian MCMC implementation of the general mixed Yule-coalescent model for species delimitation.. R package version 1.0.2.'

Van Noort, S., Gardiner, A. J. and Tolley, K. A. (2007) 'New records of Ficus (Moraceae) species emphasize the conservation significance of inselbergs in Mozambique', *South African Journal of Botany*, 73(4), pp. 642–649. doi: 10.1016/j.sajb.2007.04.063.

Norman Myers *et al.* (2000) 'Biodiversity hotspots for conservation priorities', *Nature*, 403, pp. 853–859.

Pacifici, M. *et al.* (2015) 'Assessing species vulnerability to climate change', *Nature Climate Change*, 5, pp. 215–224. doi: 10.1038/nclimate2448.

Palumbi, S. *et al.* (1991) 'The Simple Fool's Guide To PCR', *Department of Zoology, University of Hawaii, Honolulu*. doi: 10.1186/s13620-015-0060-3.

Pan, Y. *et al.* (2011) 'A large and persistent carbon sink in the world's forests', *Science*, 33, pp. 988–993. doi: 10.1126/science.1201609.

Paradis, E., Claude, J. and Strimmer, K. (2004) 'APE: analyses of phylogenetics and evolution in R language.', *Bioinformatics*, 20, pp. 289–290.

Park, C. and Allaby, M. (2017) 'A Dictionary of Environment and Conservation (3rd. Edition)', *Oxford University Press*. doi: 10.1093/acref/9780198609957.001.0001.

Partridge, T. (1997) 'Tectonic Uplift and Climate Change', *W. F. Ruddiman, Edition*, pp. 63–70.

Phillips, S. J., Dudík, M. and Schapire, R. E. (2019) 'Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/'.

Pocock, M. J. O. *et al.* (2006) 'Ecological correlates of range structure in rare and scarce British plants', *Journal of Ecology*, 94, pp. 581–596. doi: 10.1111/j.1365-2745.2006.01123.x.

Pounds, J. A. *et al.* (2006) 'Widespread amphibian extinctions from epidemic disease driven by global warming', *Nature*, 439, pp. 161–167. doi: 10.1038/nature04246.

Poynton, J.C and Broadley, D. (1985) 'Amphibia Zambesiaca 1. Scolecomorphidae, Pipidae, Microhylidae, Hemisidae, Arthroleptidae', *Annual Natal Museum*, 26(2), pp. 503–553.

Poynton, J. (1999) 'Distribution of amphibians in sub-Saharan Africa, Madagascar and Seychelles.', *Patterns of Distribution of Amphibians: A Global Perspective*, pp. 483–539.

Poynton, J. (2003) 'Arthroleptis troglodytes and the content of Schoutedenella (Amphibia: Anura: Arthroleptidae)', *African Journal of Herpetology*, 52(1), pp. 49–51.

du Preez, L. and Carruthers, V. (2017) *Frogs of Southern Africa A Complete Guide*. 2nd edn. Cape Town: Random House Struik.

QGIS Geographic Information System. Open Source Geospatial Foundation Project (2019) 'QGIS Development Team'.

Rambaut, A. (2018) 'FigTree 1.4.4 2006-2018'.

Rambaut, A. *et al.* (2018) 'Posterior summarization in Bayesian phylogenetics using Tracer 1.7', *Systematic Biology*, 67(5), pp. 901–904. doi: 10.1093/sysbio/syy032.

Reid, N. M. and Carstens, B. C. (2012) 'Phylogenetic estimation error can decrease the accuracy of species delimitation: A Bayesian implementation of the general mixed Yule-coalescent model', *BMC Evolutionary Biology*, 12, p. 196. doi: 10.1186/1471-2148-12-196.

Rö del, M.-O. *et al.* (2009) 'A new small *Arthroleptis* (Amphibia: Anura: Arthroleptidae) from the Liberian part of Mount Nimba, West Africa.', *Zootaxa2*, 2302, pp. 19–30.

Ronquist, F. and Huelsenbeck, J. P. (2003) 'MRBAYES 3: Bayesian phylogenetic inference under mixed models.', *Bioinformatics*, 19, pp. 1572–1574.

Rosauer, D. *et al.* (2009) 'Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history', *Molecular Ecology*, 18, pp. 4061–4072. doi: 10.1111/j.1365-294X.2009.04311.x.

Rosauer, D. (2020) *Phylospatial*, *Phylospatial*. Available at: https://github.com/DanRosauer/phylospatial. (Accessed March 2020).

Rosauer, D. F. *et al.* (2016) 'Phylogeography, hotspots and conservation priorities: an example from the Top End of Australia', *Biological Conservation*, 204, pp. 83–93. doi: 10.1016/j.biocon.2016.05.002.

Rosauer, D. F. and Jetz, W. (2015) 'Phylogenetic endemism in terrestrial mammals', *Global Ecology and Biogeography*, 24, pp. 168–179. doi: 10.1111/geb.12237.

RStudioTeam (2019) 'RStudio: Integrated development environment for R', *Inc. .Boston.* doi: 10.1002/jwmg.232.

Rudh, A. and Qvarnström, A. (2013) 'Adaptive colouration in amphibians', *Seminars in Cell and Developmental Biology*, (6–7), pp. 553–561. doi: 10.1016/j.semcdb.2013.05.004.

Rueda, M., Rodríguez, M. Á. and Hawkins, B. A. (2013) 'Identifying global zoogeographical regions: Lessons from Wallace', *Journal of Biogeography*, 40, pp. 2215–2225. doi: 10.1111/jbi.12214.

Ryan, C. M., Berry, N. J. and Joshi, N. (2014) 'Quantifying the causes of deforestation and degradation and creating transparent REDD+ baselines: A method and case study from central Mozambique', *Applied Geography*, 53, p. 45. doi: 10.1016/j.apgeog.2014.05.014.

San Mauro, D. *et al.* (2004) 'Phylogeny of caecilian amphibians (*Gymnophiona*) based on complete mitochondrial genomes and nuclear rag-*1*', *Molecular Phylogenetics and Evolution*, 33, pp. 413–427. doi: 10.1016/j.ympev.2004.05.014.

Schweiger, S. *et al.* (2017) 'Direct development in African squeaker frogs (Anura: Arthroleptidae: *Arthroleptis)* reveals a mosaic of derived and plesiomorphic characters.', *Organisms Diversity & Evolution*, 17(3), pp. 693–707.

Sciscio, L. *et al.* (2013) 'Fluctuations in Miocene climate and sea levels along the southwestern South African coast: Inferences from biogeochemistry, palynology and sedimentology', *Palaeontologia Africana*, 48, pp. 2–18.

Seehausen, O. (2006) 'African cichlid fish: A model system in adaptive radiation research', *Proceedings of the Royal Society B: Biological Sciences*, 273, pp. 1987–1998. doi: 10.1098/rspb.2006.3539.

Segami Marzal, J. C. *et al.* (2017) 'Cryptic female Strawberry poison frogs experience elevated predation risk when associating with an aposematic partner', *Ecology and Evolution*, 7(2), pp. 744–750. doi: 10.1002/ece3.2662.

Sepulchre, P. *et al.* (2006) 'Tectonic uplift and Eastern Africa aridification', *Science, New York*, 313(5792), pp. 1419–1423. doi: 10.1126/science.1129158.

Shine, R. (1979) 'Sexual Selection and Sexual Dimorphism in the Amphibia', *Copeia*, 1972(2), pp. 279–306. doi: 10.2307/1443418.

Shuster, S. and Wade, M. (2003) 'Mating systems and strategies', *Princeton University Press*.

Sintayehu, D. W. (2018) 'Impact of climate change on biodiversity and associated key ecosystem services in Africa: a systematic review', *Ecosystem Health and Sustainability*, 4(9), pp. 225–239. doi: 10.1080/20964129.2018.1530054.

Slater, G. J. *et al.* (2010) 'Diversity versus disparity and the radiation of modern cetaceans', in *Proceedings of the Royal Society B: Biological Sciences*, pp. 3097–3104. doi: 10.1098/rspb.2010.0408.

Starnberger, I., Preininger, D. and Hödl, W. (2014) 'The anuran vocal sac: A tool for multimodal signalling', *Animal Behaviour*, 97, pp. 281–288. doi: 10.1016/j.anbehav.2014.07.027.

Stecher G., Tamura K., and Kumar S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* (https://doi.org/10.1093/molbev/msz312).

Stuart-Fox, D. M. and Ord, T. J. (2004) 'Sexual selection, natural selection and the evolution of dimorphic coloration and ornamentation in agamid lizards', *Proceedings of the Royal Society B: Biological Sciences*, 271, pp. 2249–2255. doi: 10.1098/rspb.2004.2802.

Stuart, S. N. *et al.* (2004) 'Status and trends of amphibian declines and extinctions worldwide', *Science*, 306, pp. 1783–1786. doi: 10.1126/science.1103538.

Stuckert, A. *et al.* (2019) 'Differential gene expression and gene variants drive color and pattern

development in divergent color morphs of a mimetic poison frog', *bioRxiv*. doi: 10.1101/706671.

Sunday, J. M. *et al.* (2014) 'Thermal-safety margins and the necessity of thermoregulatory behavior across latitude and elevation', *Proceedings of the National Academy of Sciences*, 111, pp. 5610–5615. doi: 10.1073/pnas.1316145111.

Talavera, G., Castresana, J. (2007) 'Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Accessed http://molevol.cmima.csic.es/castresana/Gblocks_server.html in 2019', *Systematic Biology*, 56, pp. 564–577. Available at: http://molevol.cmima.csic.es/castresana/data/syst_biol_2007_appendix.pdf.

Tapley, B. *et al.* (2018) 'The disparity between species description and conservation assessment: A case study in taxa with high rates of species discovery', *Biological Conservation*, 220, pp. 209–214. doi: 10.1016/j.biocon.2018.01.022.

Taylor, R. C., Buchanan, B. W. and Doherty, J. L. (2007) 'Sexual selection in the squirrel treefrog *Hyla squirella*: the role of multimodal cue assessment in female choice', *Animal Behaviour*, 74, pp. 1753–1763. doi: 10.1016/j.anbehav.2007.03.010.

Templeton, A. R., Crandall, K. A. and Sing, C. F. (1992) 'A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation', *Genetics*, 132, pp. 619–633.

Timberlake, J. *et al.* (2007) *The biodiversity and conservation of Mount Chiperone, Mozambique*, *Report produced under the Darwin Initiative Award*.

Timberlake, J. *et al.* (2009) 'Mt Namuli, Mozambique: biodiversity and conservation.', *Darwin Initiative Award*, 15, p. 36.

Tolley, K. *et al.* (2018) 'Molecular phylogenetics reveals a complex history underlying cryptic diversity in the Bush Squeaker Frog *Arthroleptis wahlbergii* in southern Africa', *African Zoology*, 53(3), pp. 83–97. doi: https://doi.org/10.1080/15627020.2018.1517608.

Tolley, K. A., Chase, B. M. and Forest, F. (2008) 'Speciation and radiations track climate transitions since the Miocene Climatic Optimum: A case study of southern African chameleons', *Journal of Biogeography*, 35, pp. 1402–1414. doi: 10.1111/j.1365-2699.2008.01889.x.

Tolosa, Y. *et al.* (2015) 'Sexual maturity and sexual dimorphism in a population of the rocket-frog *Colostethus aff. fraterdanieli* (Anura : Dendrobatidae) on the northeastern Cordillera Central of Colombia', *Actualidades Biológicas*, 37(102).

Tracy, C. R., Christian And, K. A. and Richard Tracy, C. (2010) 'Not just small, wet, and cold:

Effects of body size and skin resistance on thermoregulation and arboreality of frogs', *Ecology*, 91(5), pp. 1477–1484. doi: 10.1890/09-0839.1.

Tsinda, A. *et al.* (2016) 'Biodiversity informatics in Eastern Africa: Status, drivers and barriers', *Journal for Nature Conservation*, 32, pp. 67–80. doi: 10.1016/j.jnc.2016.05.002.

Tsuji, H. (2004) 'Reproductive ecology and mating success of male *Limnonectes kuhlii,* a fanged frog from Taiwan.', *Herpetologica*, 60, pp. 155–167. doi: 10.1655/02-10.

Turner, A. and Channing, A. (2017) 'Three new species of *Arthroleptella Hewitt*, 1926 (Anura: Pyxicephalidae) from the Cape Fold Mountains, South Africa', *African Journal of Herpetology*, 66, pp. 53–78. doi: 10.1080/21564574.2017.1324918.

Vanzolini, P. E. and Williams, E. F. (1981) 'The vanishing refuge: A mechanism for ecogeographic speciation', *Papéis Avulsos de Zoologia*, 34, pp. 251–255.

Vences, M., Thomas, M., van der Meijden, A., Chiari, Y., & Vieites, D. R. (2005). Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Frontiers in zoology*, *2*(1), 5. https://doi.org/10.1186/1742-9994-2-5

Vences, M. and Wake, D. B. (2007) 'Speciation, species boundaries and phylogeography of amphibians', in *Amphibian Biology, Vol.7, Systematics*, pp. 2613–2669.

Voelker, G. *et al.* (2013) 'River barriers and cryptic biodiversity in an evolutionary museum', *Ecology and Evolution*, 3(3), pp. 536–545. doi: 10.1002/ece3.482.

Wake, D. B. (2009) 'What salamanders have taught us about evolution', *Annual Review of Ecology, Evolution, and Systematics*, 40, pp. 333–352. doi: 10.1146/annurev.ecolsys.39.110707.173552.

Wallace, A. R. (1852) 'On the monkeys of the Amazon', *Proceedings of the Zoological Society of London*, 20, pp. 107–110. doi: 10.1080/037454809494374.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Willcock, S. *et al.* (2012) 'Towards regional, error-bounded landscape carbon storage estimates for data-deficient areas of the world', *PLoS ONE*, 7, p. e44795. doi: 10.1371/journal.pone.0044795.

Wilson, L. and Channing, A. (2019) 'A new sand frog from Namaqualand, South Africa (Pyxicephalidae: *Tomopterna*)', *Zootaxa*, 4609, pp. 225–246.

Windsor, A. (2009) 'Mozambique agrees to protect lost rainforest of Mount Mabu', *The Guardian*.

Woest, N. (2019) *Photos from my private collection.*

Young, A. (1972) *Slopes*. Edinburgh: Oliver and Boyd. doi: 10.1111/j.1745-7939.1975.tb00803.x.

Yule, G. U. (1925) 'A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, F.R.S.', *Journal of the Royal Statistical Society*, 213, pp. 21–87. doi: 10.2307/2341419.

Zeisset, I. and Beebee, T. J. C. (2008) 'Amphibian phylogeography: A model for understanding historical aspects of species distributions', *Heredity*, 101, pp. 109–119. doi: 10.1038/hdy.2008.30.

Zhang, L. and Lu, X. (2013) 'Sexual size dimorphism in anurans: Ontogenetic determination revealed by an across-species comparison', *Evolutionary Biology*, 40, pp. 84–91. doi: 10.1007/s11692-012-9187-2.

Zimkus, B. M. and Blackburn, D. C. (2008) 'Distinguishing features of the Sub-Saharan frog genera *Arthroleptis* and *Phrynobatrachus*: A short guide for field and museum researchers', *Breviora*. doi: 10.3099/0006-9698(2008)513[1:dfotsf]2.0.co;2.

Zimkus, B. M. and Larson, J. G. (2009) 'External morphology and osteology support the placement of *Phrynobatrachus nlonakoensis* Plath, Herrmann & Böhme, 2006 within the genus arthroleptis', *African Journal of Herpetology*, 58, pp. 36–38. doi: 10.1080/21564574.2009.9635578.

7. Appendix

Appendix 1

Arthroleptis francei directly described by Arthur Loveridge (1953):

Holotype. M.C.Z. No. 27479, a gravid female from the forested banks of the Ruo River just below the Ruo Falls on Mlanje Mountain, about 5000feet. Collected by Arthur Loveridge, April 4, 1949.

Paratypes
M.C.Z. Nos. 27470-8 and twenty uncatalogued duplicates taken at the same time and place as the type.

Diagnosis
Obviously related to adolfifriederici Nieden of Central Africa, from which it differs in the much less developed dilations of fingers and toes (strongly dilated in adolfifriederici), the shorter hind limbs (which when adpressed in adolfifriederici reach from between eye and nostril to beyond end of snout), and general ground color (varying shades of nut brown in adolfifriederici).

Description
Type female. Head not wider than body (also in paratypes); tympanum distinct, half the orbital diameter (in entire series); first finger slightly shorter than second which extends as far as fifth when pressed together, fourth (on its outer side) less than twice as long as the fifth (twice as long in the male paratypes); tips of fingers slightly, of toes strongly, swollen, their bases without webbing; tibio-tarsal articulation of the adpressed hind limb reaches the eye (as is the case with 26 of the paratypes) or between eye and nostril (in 3 instances) ; an inner, but no outer, metatarsal tubercle whose length is shorter than the first toe.

Colour of female type in life
Above, dark brownish red; from nostril through eye above (and on upper portion of) tympanum to above forearm, a deep black band edged with lighter above, especially pale on eyelid; from eyelid to eyelid an obsolescent crossbar; similar dusky markings

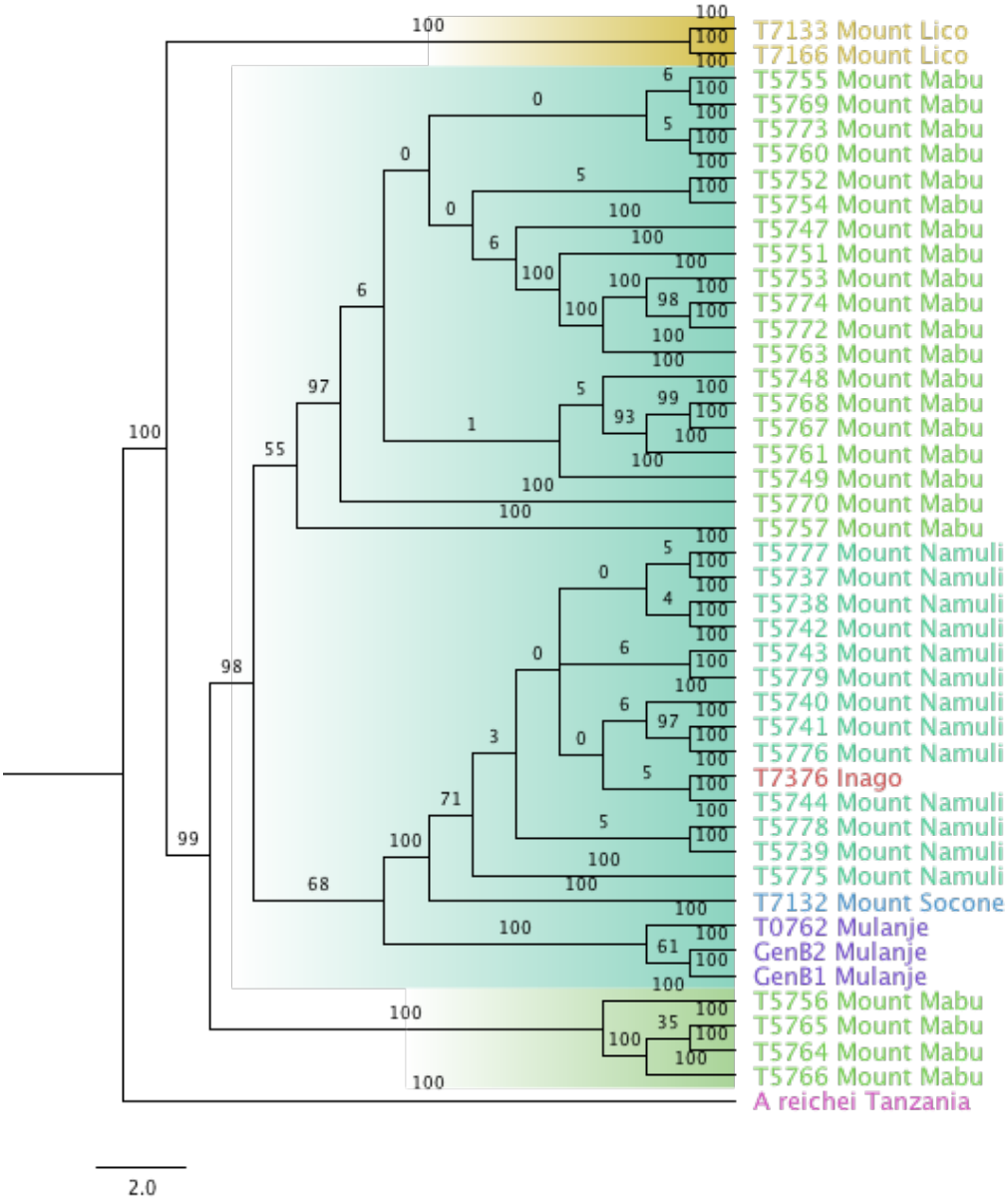
occur as marblings on back, flanks, limbs and around anus. Below, pinkish white with underlying dusky markings and silvery white flecks on chest, sides of abdomen, posterior aspects of thighs, and on the almost blackish soles of hind feet; palms, fingers and toes more reddish. A slightly smaller female had the snout and anterior half of head pale pinkish buff, the black interorbital crossbar merging into the general black of the vertebral region which is dorso-laterally bounded by the same shade as the snout; on each flank, especially posteriorly, and on each hind limb, are a score of cream colored spots about whose edges are superposed small red dots that enhance this frog's striking appearance. In alcohol the entire series is predominantly gray, but in life the ground color was gray, buff, fawn, pale green, or rich brick red. The vertebral hour-glass pattern characteristic of Arthroleptis, conspicuous in some, appears to be absent in others, but can usually be detected with the aid of a lens; dark dorso-lateral lines are present on two young frogs causing them to look a little like pinkish Hyperolius. Even more imposing is a 20 mm. juvenile that is black except for the tip of its snout, which is buffy, an almost whitish interorbital bar, and numerous white spots on back and limbs. Below, all are substantially the same as described for the type except that grayish, not pinkish, predominates.

Size
Snout to anus of paratype cf (M.C.Z. 27476), 32 mm.; of type female, 46 (49 just after death) mm.; the entire series ranges from 10 to 46 mm., but only nine frogs are over 30 mm. Breeding. In April the ova were only moderately large. Habitat. Unlike the equatorial forests of Tanganyika, Nyasaland forests at high altitudes apparently become too cold for amphibians owing to the drop in temperature when it comes on to rain. In a three-hour (7 to 10 am) walk up through the forest from the Power House under conditions that appeared ideal, I did not see a single frog. During the first two hours the enveloping cloud cap rendered it dull but afterwards intermittent shafts of sunlight illuminated stretches of the leaf-strewn path. It was on our way down from the Ruo Plateau that I caught the first frog among moss-grown boulders on the east bank, then, after recrossing the Ruo immediately below the Falls, a large one among leaves on the west bank. By this time it was again overcast and dark with rain threatening. Ten minutes later we took the other twenty-eight frogs within an area of a 100 square yards

between the path and the river bank. It began to rain but not another frog was seen during the rest of the way down the mountain to the Power House.

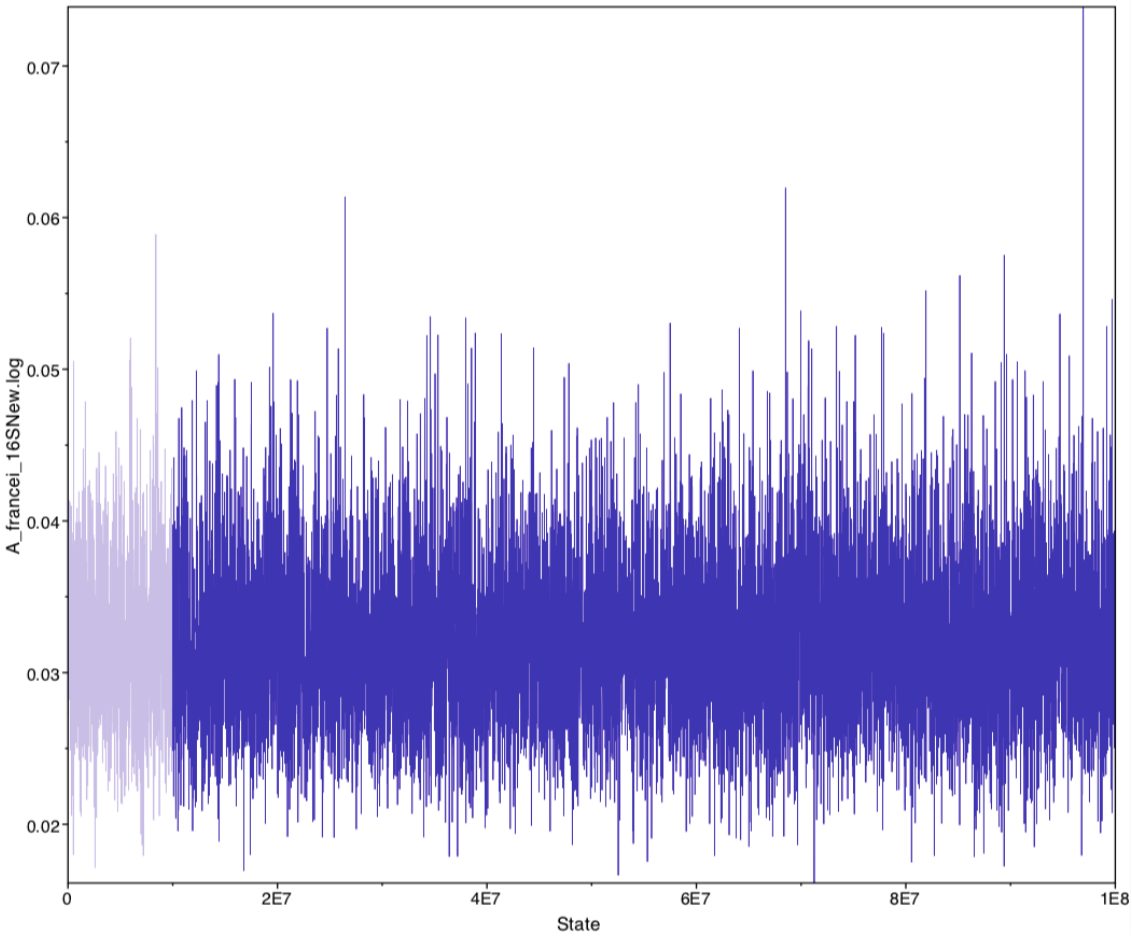
Appendix 2

*Appendix 2. mtDNA (*12S *and* 16S*) concatenated tree of* Arthroleptis francei *candidate species.*



Appendix 3

Appendix 3. 16S *Caterpillar*



Appendix 4.

R Script for bGMYC

```
#set directory
#Load packages
library(ape)
library(bGMYC)

#Read file
All_trees<-read.nexus("A_francei_16SNew.trees")

#Subset of 100 trees from my mcc tree file
subset<-sample(All_trees, 100)
```

#Using the subset of trees. t1 and t2 need to be set to possible values (e.g. between 2 and the number of leaves in the tree); the starting value for that parameter (the third value in the "start" vector) needs to be t1<start<t2 or else there will be an error.

# t2: must be less than or equal to the number of tips in the trees being analyzed. Gives the maximum number of species for which the prior probability is greater than 0. If it is greater than the number of tips in the tree, an error will result.
result.multi<-bgmyc.multiphylo(subset, mcmc=50000, burnin=40000, thinning=100, t2=49, start=c(1,1,4))

#Plot
plot(result.multi)

#to check the behavior of the MCMC. These plots may look a little funny, however, because they are the results of several sequential MCMC runs, each with their own burn-in. You may find that one or two of the trees fails to reach stationarity or winds up in a weird area of parameter space. This may suggest that you should use a longer burn-in.

#If things look fine, then you can run
result.spec<-bgmyc.spec(result.multi)

#If you want to visualize the posterior distribution in the context of the tree, you should do:
result.probmat<-spec.probmat(result.multi)

beast.tree<-read.nexus("A_francei_*16S*New.trees")

#Ladderize trees
beast.tree<-ladderize(beast.tree, right = TRUE)
right ladderized, do right right = TRUE
plot(result.probmat,beast.tree)

#Plot
plot(result.probmat, subset[[1]])

#I have recently added another function to visualize the results in a way that will hopefully be helpful. The GMYC model is likely to be successful when the rate of branching for the coalescent process is much higher (approaching an order of magnitude higher) than for the Yule process. Accordingly, there is a function "checkrates" that will output all parameter values from the run in a table. If we look at the distribution of ratios of the Coalescence to Yule rates sampled in the analysis, we would like to see that it is well above 0, with no negative values. If this is true, then the model may be a decent approximation to the reality of the data. If the ratio of rates overlaps or is close to 0, then the transition between branching processes (if it exists) is likely to be indistinct and estimates of species limits may be misleading. Log ratios of less than 0 would indicate a higher speciation rate than coalescence rate. If such samples ever appear in the MCMC this should raise huge red flags and you should consider the possibility that the model may not be a good fit for your data. The usage for these functions is:
bgmycrates<-checkrates(result.multi)

#Plot
plot(bgmycrates)

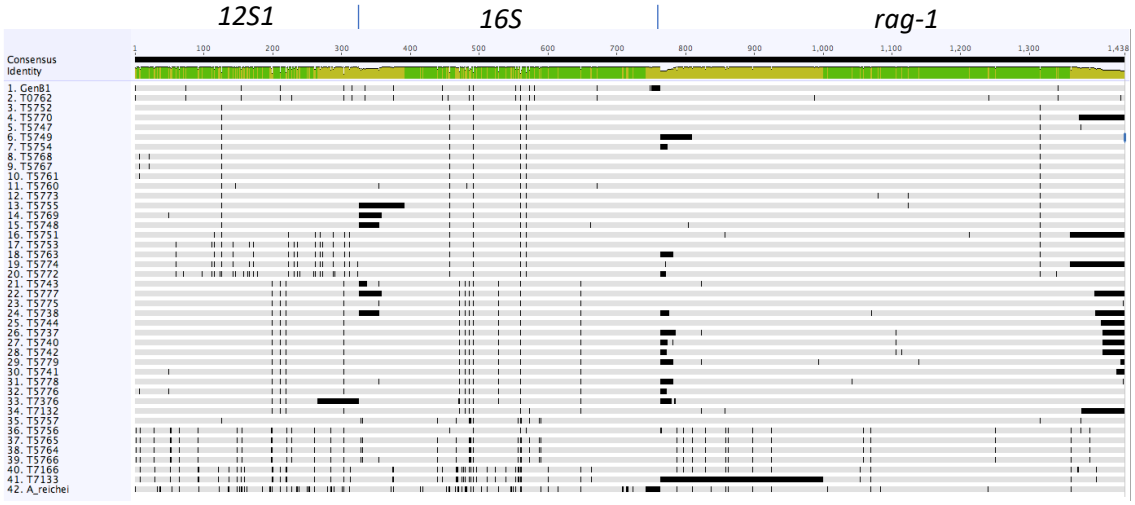
#bgmyc.point(probmat, ppcutoff)
#probmat: output from function spec.probmat of class "bgmycprobmat"
#ppcutoff: a posterior probability threshold for lumping samples into species. e.g. if 0.05 were selected, all individuals having greater than 0.05 posterior probability of conspecificity will be lumped into species returned.

species.list<-bgmyc.point(result.probmat, 0.5)

```
#Add labels to the mcc tree tips
species.df<-
data.frame(Species=rep(c(1:length(species.list)),lapply(species.list,length)),individuals=unlist(speci
es.list))
rownames(species.df)<-species.df$individuals
species.df<-species.df[beast.tree$tip,]
rownames(species.df)==beast.tree$tip
plot(beast.tree, tip.col=as.numeric(species.df$Species), cex=0.5)

#Add labels to the tree tips
species.df<-
data.frame(Species=rep(c(1:length(species.list)),lapply(species.list,length)),individuals=unlist(speci
es.list))
rownames(species.df)<-species.df$individuals
species.df<-species.df[subset[[1]]$tip,]
rownames(species.df)==subset[[1]]$tip
plot(subset[[1]], tip.col=as.numeric(species.df$Species), cex=0.5)

#Save
save.image("bgmyc_out.Rdata")

#get the posterior probabilities of each possible species cluster
#optionally writes to an output file.
bgmyc.spec(res=result.multi)->result.spec
```

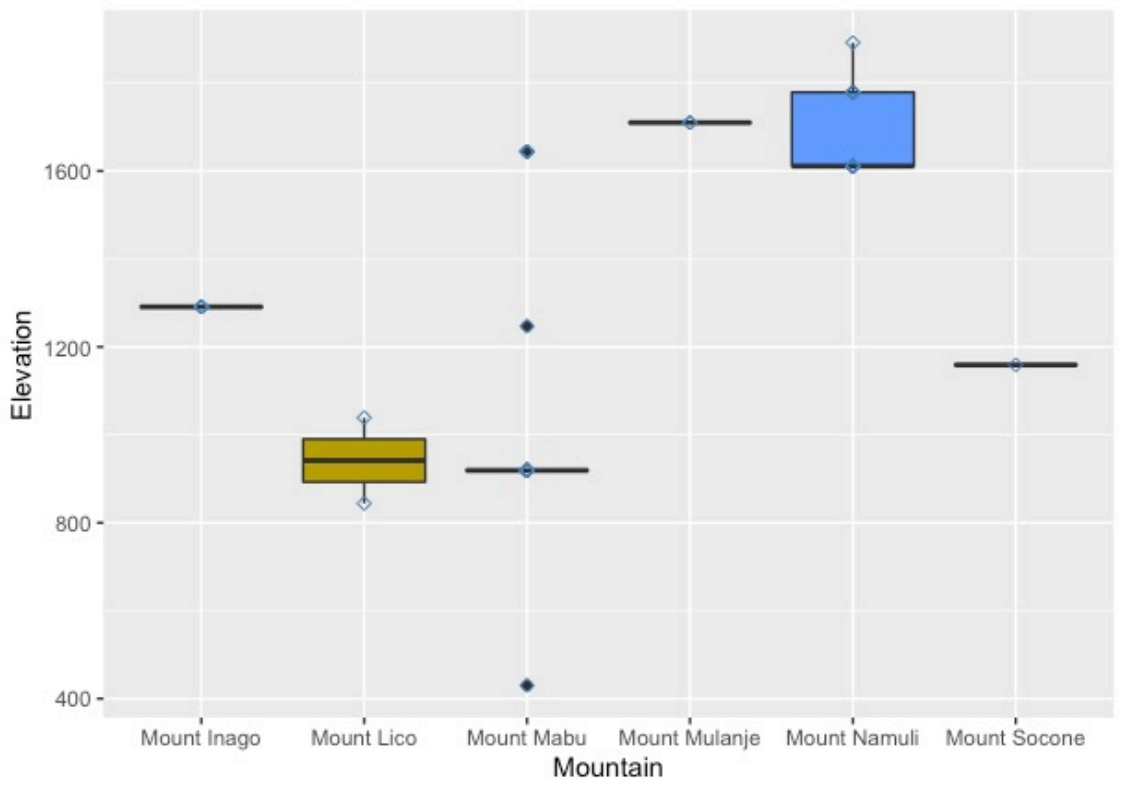
Appendix 5

Appendix 5. Mined concatenated alignment of 12S, 16S and rag-1 where taxa with missing genes or those with short sequences have been excluded. Large black bars represent missing nucleotides replaced as 'N' whilst smaller black lines are changes of nucleotides between sequences.

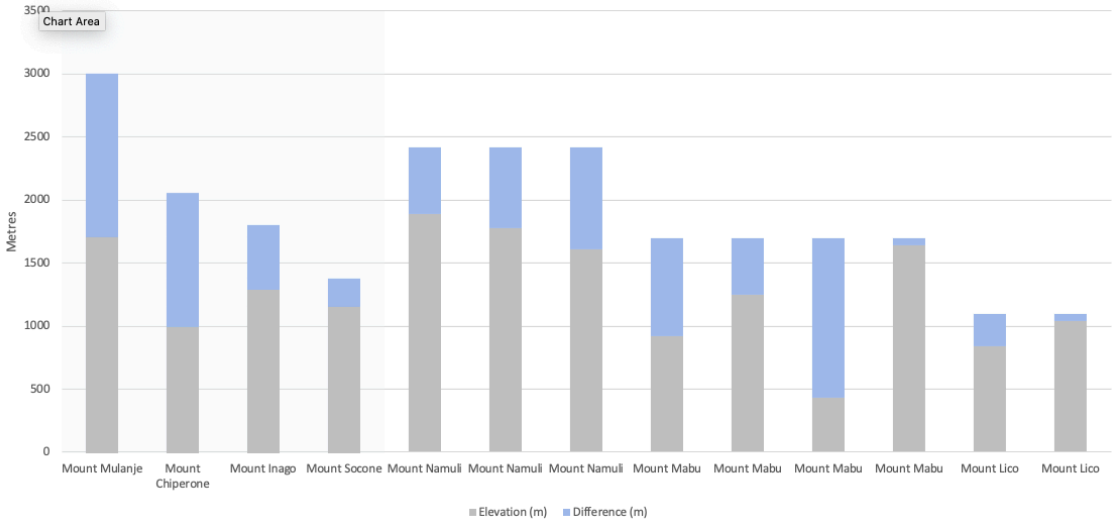


Appendix 6

Appendix 6a. Elevation of where individuals were collected on each mountain.



Appendix 6b. Elevation where individuals were collected (grey) and difference from located to the total elevation of the mountain (blue). Repetitions were omitted. Pale blue background denotes data of mountains containing a single elevation point.



Appendix 7

Appendix 7. Elevation in meters (m) for the specimens in this study

Code	Mountain	Elevation (m)	Max Elevation (m)	Difference (m)	Difference (%)
T5761	Mount Mabu	430	1700	1270	74.7
T7133	Mount Lico	844	1100	256	23.3
T5747	Mount Mabu	919	1700	781	45.9
T5748	Mount Mabu	919	1700	781	45.9
T5749	Mount Mabu	919	1700	781	45.9
T5751	Mount Mabu	919	1700	781	45.9
T5752	Mount Mabu	919	1700	781	45.9
T5753	Mount Mabu	919	1700	781	45.9
T5754	Mount Mabu	919	1700	781	45.9
T5755	Mount Mabu	919	1700	781	45.9
T5757	Mount Mabu	919	1700	781	45.9
T5763	Mount Mabu	919	1700	781	45.9
T5764	Mount Mabu	919	1700	781	45.9
T5765	Mount Mabu	919	1700	781	45.9
T5766	Mount Mabu	919	1700	781	45.9
T5772	Mount Mabu	919	1700	781	45.9
T5773	Mount Mabu	919	1700	781	45.9
T5774	Mount Mabu	919	1700	781	45.9
T7166	Mount Lico	1039	1100	61	5.5
T7132	Mount Socone	1159	1379	220	16.0
T5760	Mount Mabu	1247	1700	453	26.6
T7376	Inago	1291	1804	513	28.4
T7377	Inago	1291	1804	513	28.4
T7378	Inago	1291	1804	513	28.4
T7379	Inago	1291	1804	513	28.4
T7381	Inago	1291	1804	513	28.4
T5741	Mount Namuli	1611	2419	808	33.4
T5742	Mount Namuli	1611	2419	808	33.4
T5744	Mount Namuli	1611	2419	808	33.4
T5775	Mount Namuli	1611	2419	808	33.4
T5776	Mount Namuli	1611	2419	808	33.4
T5777	Mount Namuli	1611	2419	808	33.4
T5778	Mount Namuli	1611	2419	808	33.4
T5767	Mount Mabu	1644	1700	56	3.3
T5768	Mount Mabu	1644	1700	56	3.3
T5769	Mount Mabu	1644	1700	56	3.3
T5770	Mount Mabu	1644	1700	56	3.3
GenB1	Mount Mulanje	1710	3002	1292	43.0
T0762	Mount Mulanje	1710	3002	1292	43.0
T5738	Mount Namuli	1779	2419	640	26.5
T5739	Mount Namuli	1779	2419	640	26.5
T5740	Mount Namuli	1779	2419	640	26.5
T5743	Mount Namuli	1779	2419	640	26.5
T5779	Mount Namuli	1779	2419	640	26.5
T5737	Mount Namuli	1892	2419	527	21.8

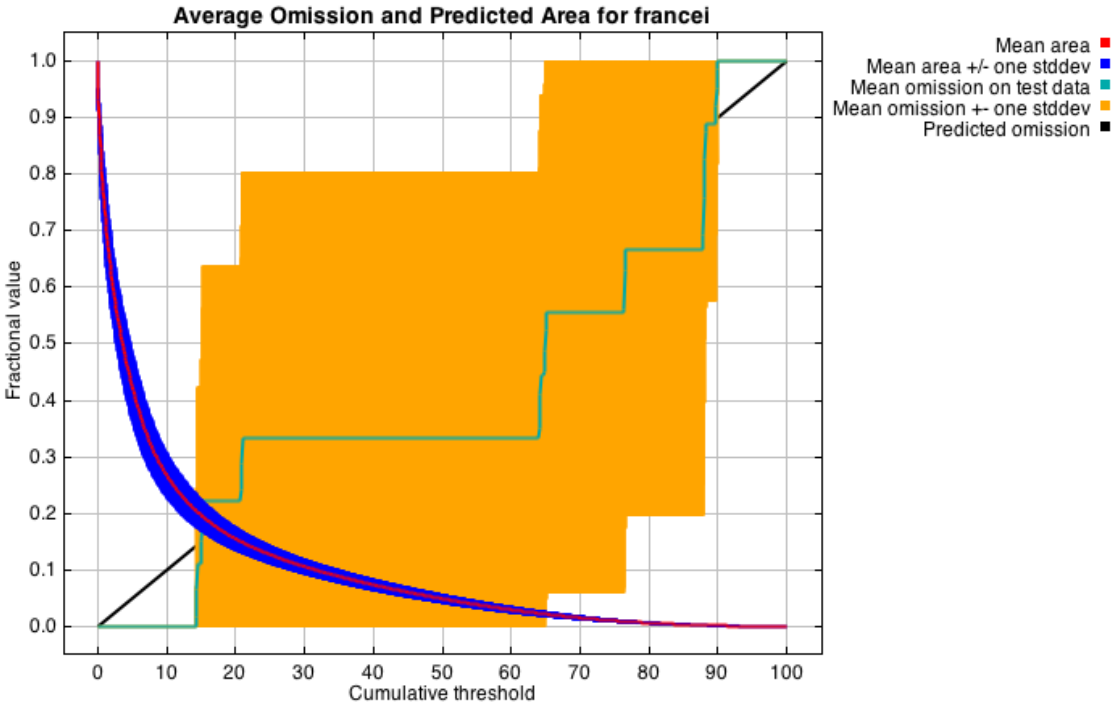
Range (m)
0 - 450
451 - 850
851 - 1250
1251 - 1650
1651 - 2050

Appendix 8

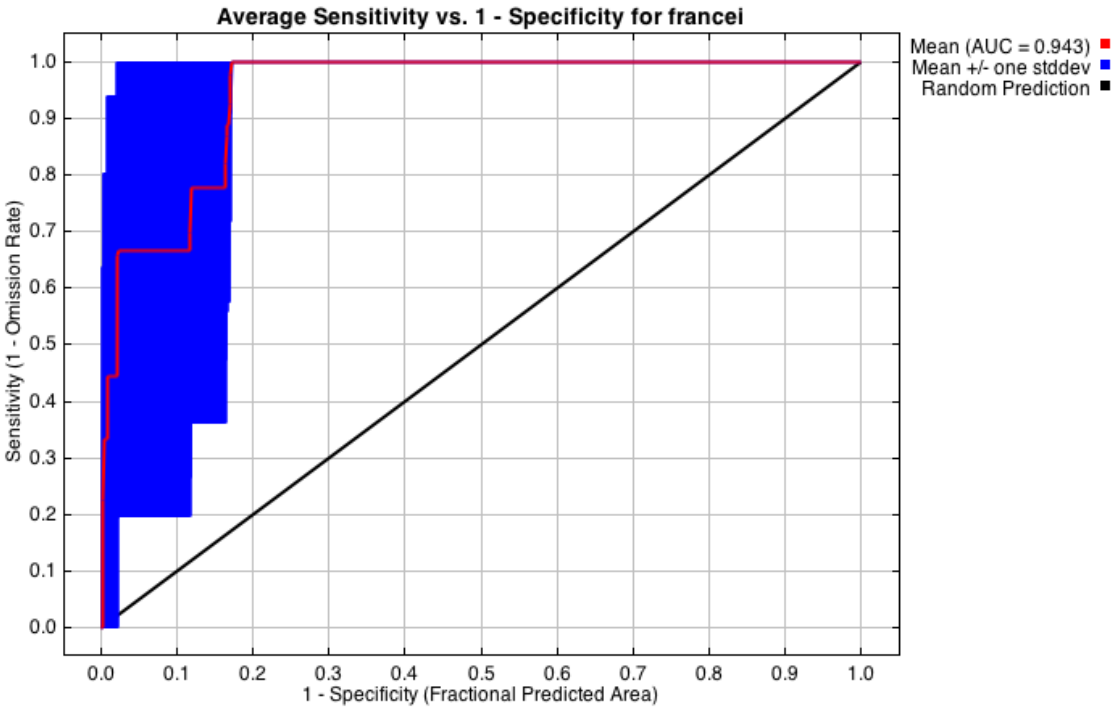
The following are results from the replicated maxent model for *Arthroleptis francei* for the BioClim variables:
BIO1 – Annual mean temperature,
BIO2 - Mean diurnal range (mean of monthly maximum temperature – minimum temperature),
BIO3 – Isothermy (mean diurnal range temperature range),
BIO7 – Temperature annual range,
BIO8 – Temperature of the wettest quarter,
BIO11 - Mean temperature of the coldest quarter,
BIO12 – Annual Precipitation,
BIO15 – Precipitation seasonality,
BIO18 – Precipitation of the warmest quarter and
BIO19 – Precipitation of the coldest quarter

This page summarizes the results of 9-fold cross-validation for francei, created Thu Aug 01 16:23:32 BST 2019 using Maxent version 3.4.1.

Analysis of omission/commission
The following picture shows the test omission rate and predicted area as a function of the cumulative threshold, averaged over the replicate runs. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.

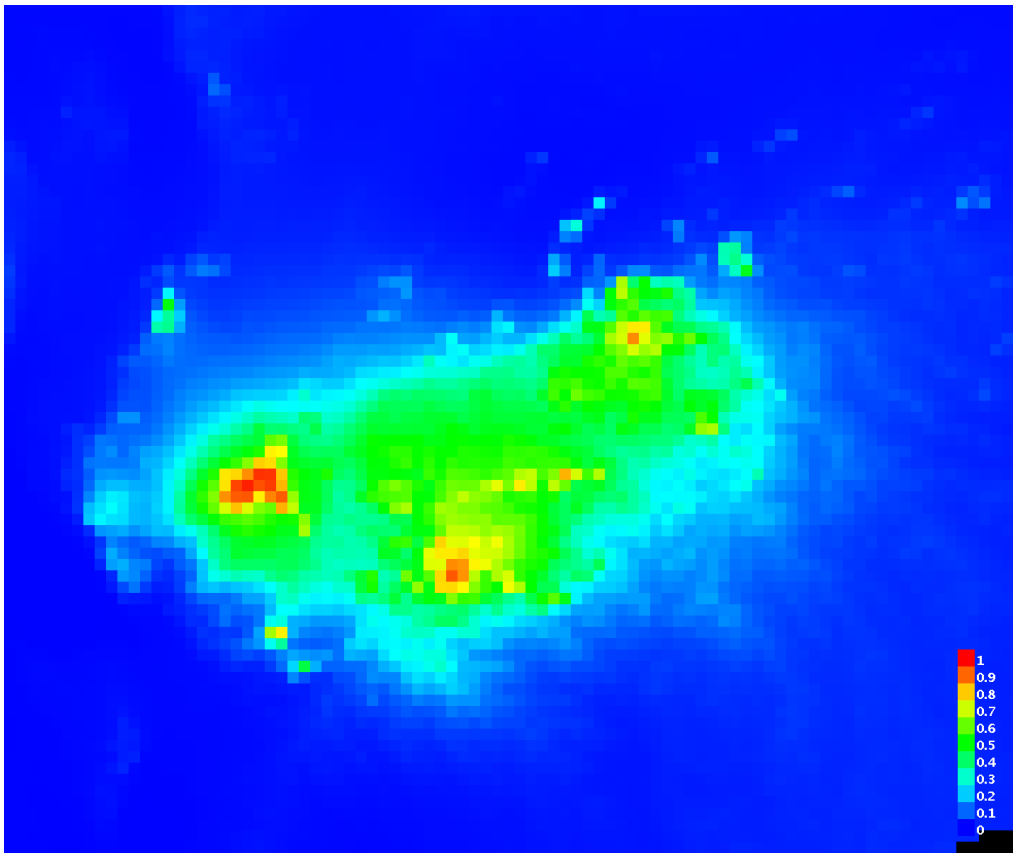


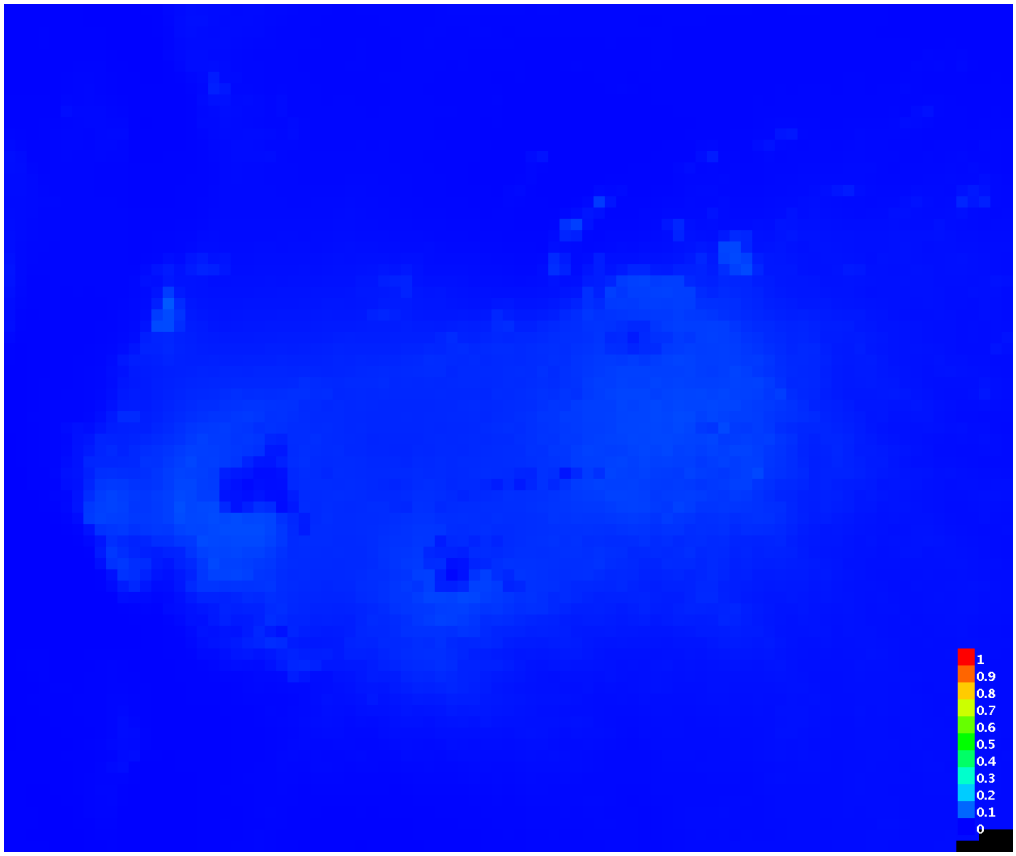
The next picture is the receiver operating characteristic (ROC) curve for the same data, again averaged over the replicate runs. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). The average test AUC for the replicate runs is 0.943, and the standard deviation is 0.068.



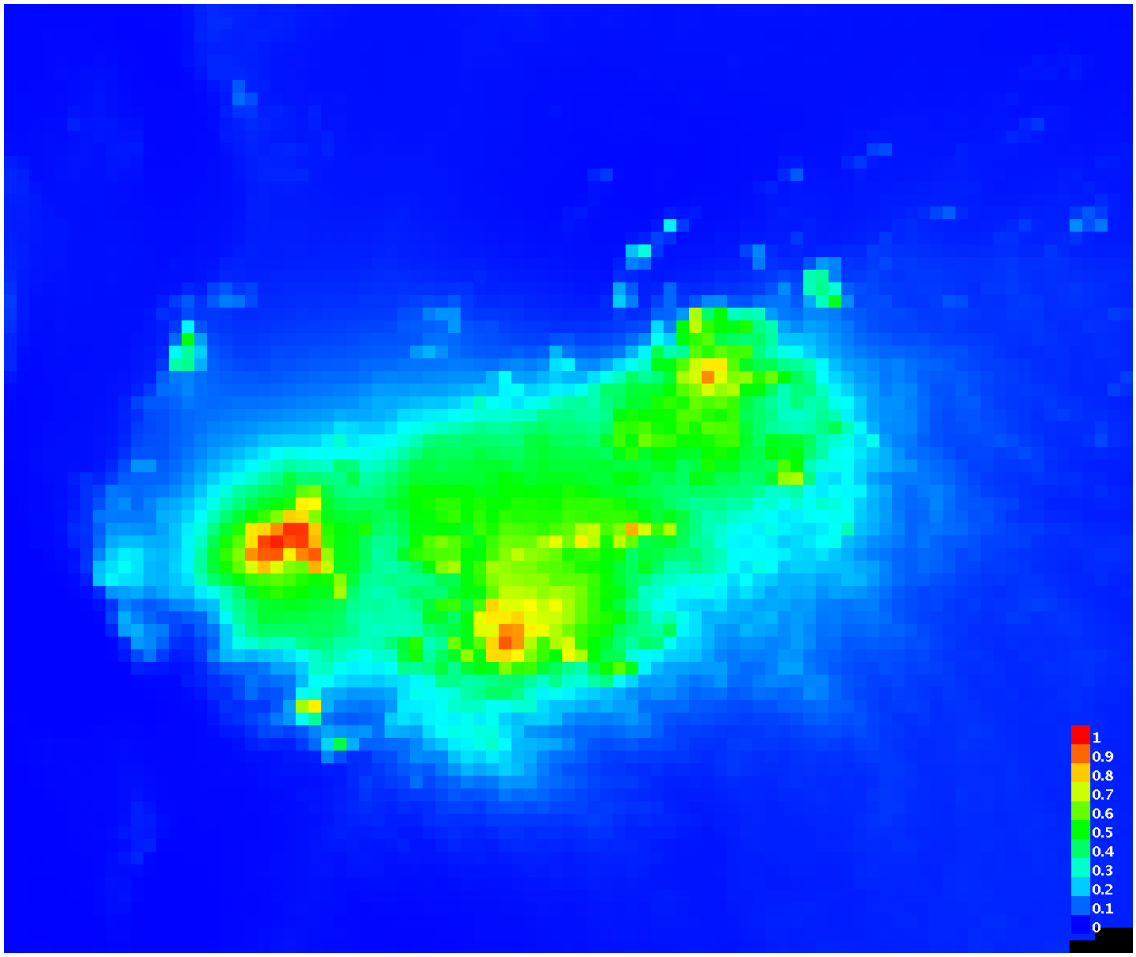
Pictures of the model

The following two pictures show the point-wise mean and standard deviation of the 9 output grids. Other available summary grids are min, max and median.





The following two pictures show the point-wise mean and standard deviation of the 9 models applied to the environmental layers in Documents. Other available summary grids are min, max and median.

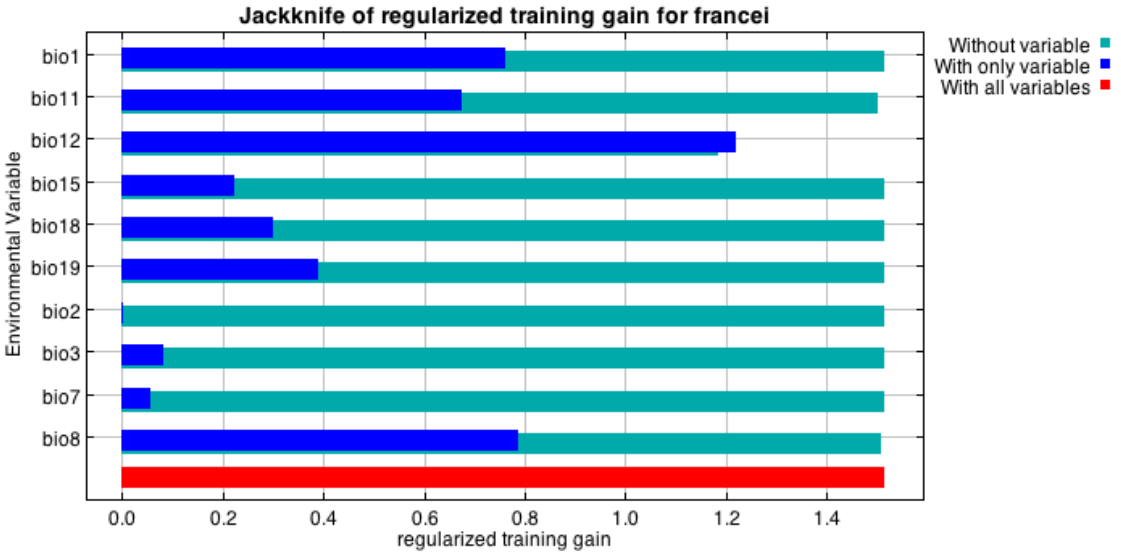


Analysis of variable contributions

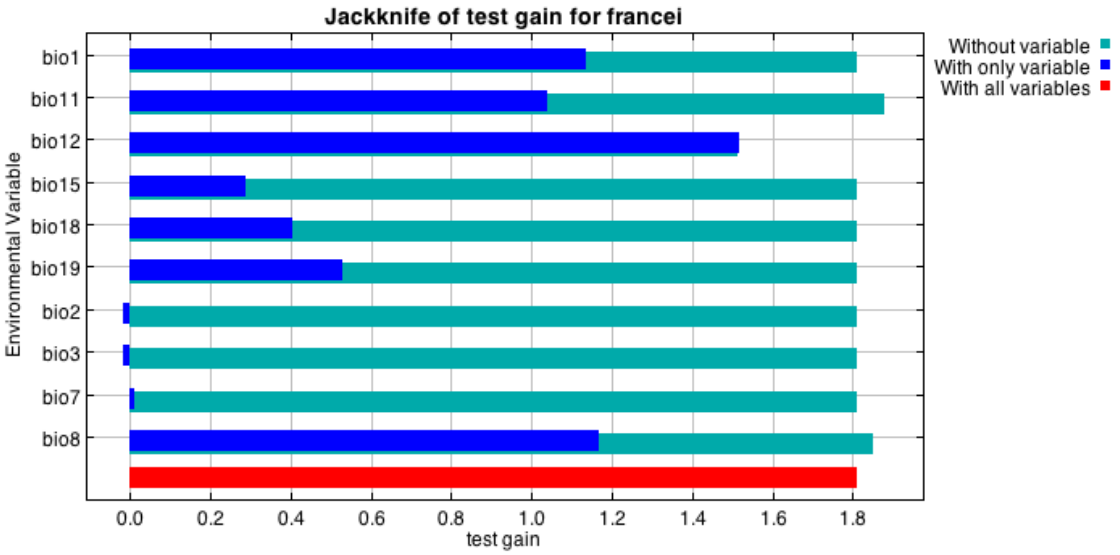
The following table gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is re-evaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated. Values shown are averages over replicate runs.

Variable	Percent contribution	Permutation importance
bio12	76.3	90.9
bio11	14.7	6
bio8	7.4	2.2
bio1	1.5	0.9
bio19	0	0
bio18	0	0
bio15	0	0
bio3	0	0
bio7	0	0
bio2	0	0

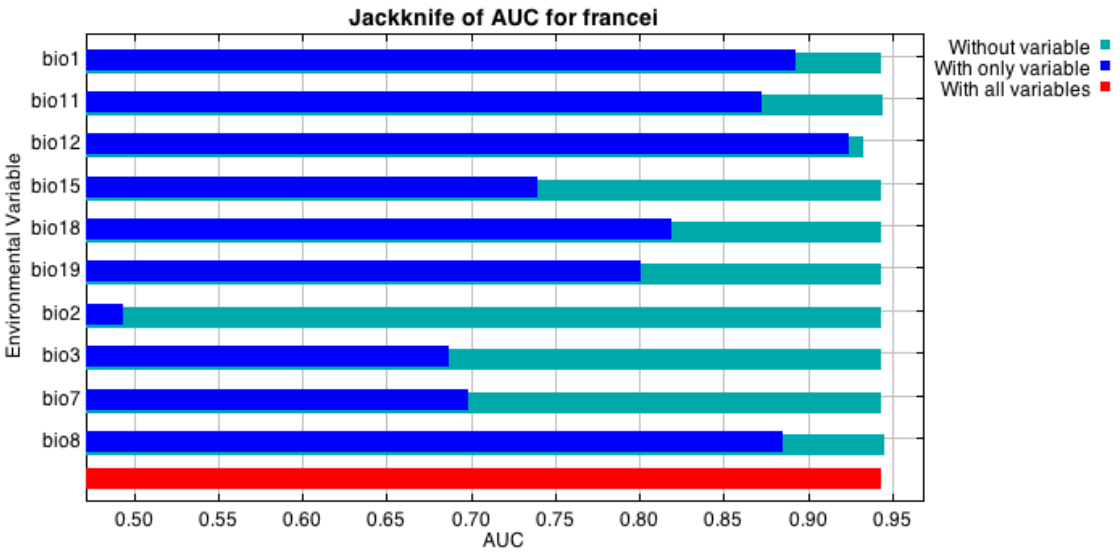
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is bio12, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is bio12, which therefore appears to have the most information that isn't present in the other variables. Values shown are averages over replicate runs.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Appendix 9

R Script for the Phylospatial models.

a. **For intraspecific lineages to generate lineage distribution models (LDM).**

```
#### Copyright Dan Rosauer 2016       ####
#### Australian National University    ####
#### September 2012 - November 2016    ####
#### dan.rosauer@anu.edu.au            ####

#############################################################################
# This program is free software: you can redistribute it and/or modify #
# it under the terms of the GNU General Public License as published by #
# the Free Software Foundation, either version 3 of the License, or    #
# (at your option) any later version.                                  #
```

```
# This program is distributed in the hope that it will be useful,      #
# but WITHOUT ANY WARRANTY; without even the implied warranty of       #
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the        #
# GNU General Public License for more details.                         #

# You should have received a copy of the GNU General Public License    #
# along with this program.  If not, see <http://www.gnu.org/licenses/> #
#########################################################################

## This script uses a set of species distribution models and a set of points for intraspecific
lineages
## to generate lineage distribution models.

## It was rewritten to run in R, from an earlier version with used ArcGIS functions via Python.

## STEPS WHICH THE CODE DOES
## 1. import the points for the whole species
##
## 2. to bound the whole analysis, use euclidian distance to create a grid to define a boundary
at a specified distance
##
## 3. load a species distribution model for the whole species, generated before running this
script
##
## 4. loop through all of the lineages in the species
##    5a. generate a euc distance layer from sequenced locations for each lineage, bounded by
the total species euclidean
##        distance layer from (2)
##     or
##    5b. generate a cost distance layer from sequenced locations for each lineage, using the
maxent model to define the cost.
##        Cost = 1 - suitability
##
##    6a. generate a weight layer for each lineage as 1 / distance  from (5)
##     or
##    6b. generate a weight layer for each lineage as 1 / distance^n  from (5) where n = 2, 3 or 4
##
##    7.  set all weights below a threshold to 0, to reduce the effect of distant lineages
##
## 8. sum all of the lineage weight layers
##
## 9. divide each lineage weight layer by the sum of weights (8) so that the weights for each
pixel
##    sum to 1 An option in this step, is to exclude lineages from a pixel where they have a low
probablility
##    of occurring. Initial models found lineages predicted over a wide area beyond their
primary range,
##    but with very low values.  To use this option set the parameters:
##         handle_minor = "threshold"
##         omit_minor_threshold = 0.1
```

```
##    as an example, 0.1 means that a lineage with less that 10% of the total of all potential
lineages for
##    that species, in the pixel, would be omitted, with the model scaled across the lineages
more likely
##    to occur in that pixel.
##    To not use this option, simply set
##        handle_minor = ""
##
## 10. multiply each lineage weight layer by the model likelihood so that the weights for each
pixel sum to the original
##    SDM model value for that pixel.

library(raster)
library(sp)
library(stringr)
library(gdistance)

rm(list=ls())

############### START OF PARAMETERS ###############

genera            <- "Arthroleptis" # allows script to run for one or more genera

base_dir          <- "/Users/natasha/Desktop/A_francei_PE/"
target_dir        <- paste(base_dir, "asc/", sep="")  # where the lineage model grids and
working data go

buffer_dist       <- 2.5    # the buffer distance in map units (presumably decimal degrees)
additional_buffer  <- 0     # how much (as a proportion) the output grids should extend
beyond the buffered points
grid_resolution    <- 0.01
spRef             <- "+proj=longlat +datum=WGS84 +ellps=WGS84 +towgs84=0,0,0"


output_model_extent       <- extent(34.7083333333333002,38.4166666666666359,-
17.2916666666666998,-14.1666666666666980)    # the maximum extent for all lineage
models
lineage_field_name  <- "lineage_from_mtDNA"   # the column for lineage name in the site
data
distance_method    <- "model-cost"          # determines whether distance is calculated as
euclidean or model-weighted cost distance
## so far, can be "euclidian" or "model-cost"
weight_function    <- "inverse_cube"    ## determines whether lineage weight is calculated as
1/distance or 1/(distance^2), or simply closest distance
## so far, can be "inverse" or "inverse_square" or "cost_allocation"
min_dist_value     <- grid_resolution/2  ## remove as a parameter, once working
min_SDM_value      <- 0.005  # values below this for the SDM are set to zero, to simplify
calculation in areas of essentially unsuitable habitat
##   but keep current value for consistency in this study
min_weight_threshold <- 0.02       ## weights below this for any layer are set to 0.  If the value
here is 0, then no threshold is applied
```

```
scale_to          <- "model"     ## determines whether lineage weights within a model group sum to the SDM value or to 1
## can be "model" or "one"

handle_minor      <- "threshold"  # if handle_minor = 'threshold', then lineages with < than the specified proportion of the lineage sum
omit_minor_threshold <- 0.1       # for that cell, are set to 0

skip_distance_layers <- FALSE      # skip creating the distance layers - they are already done.
# THIS OPTION IS ONLY TO SAVE TIME DURING DEBUGGING

named_species     <- c("francei")        # a changeable list to allow for species in the dataset to be skipped
use_list          <- "do"          # specify what to do with the named species:
#do - the named species (use_list="do")
#skip - the named species (use_list="skip")
#do all the species in the data and ignore the named species list (use_list="" or anything else);

lin_exclude_list   <- c()   # this list allows for skipping at the lineage level

################ END OF PARAMETERS ################

cat ("\n\n************************************* \n Lineage Distribution Estimation Tool ")
cat ("\n    Dan Rosauer \n    September 2012 - November 2016")
cat ("\n*************************************\n")

SDM_model_base    <- paste(base_dir, "full_maxent_model", sep="")

time_start <- Sys.time()

# loop through each genus in the vector genera
for (genus in genera) {

cat("\nGenus:", genus, "\n")

# Load the sequence site data
cat ("Loading the lineage locations...\t")
lineage_site_filename <- paste(base_dir, "species_sites/", genus, "_sites.csv", sep="")

all_lineages_sites <- read.csv(lineage_site_filename)
orig_rowcount <- nrow(all_lineages_sites)

# filter the records
not_sequenced <- grep("not_sequenced", all_lineages_sites[, lineage_field_name])  # remove records with 'not_sequenced' in the lineage name
if (length(not_sequenced >0)) {all_lineages_sites <- all_lineages_sites[-not_sequenced, ]}
all_lineages_sites <- all_lineages_sites[which(all_lineages_sites$Use %in% c(-1,1)), ]    # remove records where 'use' <> -1 or 1
all_lineages_sites <- all_lineages_sites[which(is.finite(all_lineages_sites[, "latitude"]) & is.finite(all_lineages_sites$longitude)), ]    # remove without a numeric latitude and longitude
```

```
all_lineages_sites$model_group <- str_trim(all_lineages_sites$model_group)
all_lineages_sites[, lineage_field_name] <- str_trim(all_lineages_sites[, lineage_field_name])

groupLineageList  <- unique(all_lineages_sites[, 1:2])
groups            <- unique(all_lineages_sites$model_group)
groupLineages     <- unique(all_lineages_sites[, lineage_field_name])

cat (orig_rowcount, "rows read,", nrow(all_lineages_sites), "valid records loaded\n")

# turn lineage points into a SpatialPoints object
allLineagePoints <- SpatialPointsDataFrame(all_lineages_sites[, c("longitude", "latitude")],
data=all_lineages_sites, proj4string = CRS(spRef))
allLineagePoints <- crop(allLineagePoints, output_model_extent)
# python writes to file at this point, but prob not needed

# restrict groups to particular species based on the named_species parameter
if (use_list == "do") {
groups <- intersect(groups, named_species)
} else if (use_list == "skip") {
groups <- setdiff(groups, named_species)
  }

 # print a list of model groups
 cat ("\nModel groups to do in", genus, "\n")
 for (group in groups) {
 cat("\t", group, "\n")
  }

  # start looping through the model groups for this genus
  for (group in groups) {
  group <- as.character(group)
  if (group == "0" | group == "") {next}

  cat ("\nStarting group", genus, group, "on", date(), "\n")

  # load the SDM model raster
  #SDM_model <- paste(SDM_model_base, genus, "/", str_replace(group," ","_"),
   "_median.asc", sep="")
   SDM_model <-
   "/Users/natasha/Desktop/A_francei_PE/full_maxent_model/Arthroleptis/Arthroleptis_franc
   ei_median.asc"
    SDM.ras <- raster(SDM_model)
   projection(SDM.ras) <- CRS(spRef)
   group_extent <- extent(SDM.ras)

   # get a list of the lineages in this group
   lineages <- groupLineageList[which(groupLineageList$model_group==group),
   lineage_field_name]
    if (length(lin_exclude_list) > 0) {
      lineages <- setdiff(lineages,lin_exclude_list)
```

```
  }

  cat ("\nLineages in", group, ":")
  for (lineage in lineages) {
    cat ("\n   ", lineage)
  }

  # # set the environment
  # env.snapRaster  = maxent_model
  # env.mask        = maxent_model
  # env.extent      = maxent_model

  if (length(lineages) > 1) { # proceed with lineage models if there are multiple
  # lineages - otherwise just copy the SDM for the model group

  thisGroupPoints <- allLineagePoints[allLineagePoints$model_group == group, ]
  thisGroupPoints <- crop(thisGroupPoints, group_extent)

  # calculate a new extent
  group_points_extent <- extent(thisGroupPoints)
  buffer_ratio  <- 1 + additional_buffer
  extent_buffer <- buffer_dist * buffer_ratio

  # new extent is the same as points layer + a buffer, but where the extended buffer
  # goes beyond the extent of the maxent model, limit to the output model extent.
  xmin <- max(group_points_extent@xmin - extent_buffer,   output_model_extent@xmin)
    ymin <- max(group_points_extent@ymin - extent_buffer, output_model_extent@ymin)
    xmax <- min(group_points_extent@xmax + extent_buffer, output_model_extent@xmax)
    ymax <- min(group_points_extent@ymax + extent_buffer, output_model_extent@ymax)
    thisGroupExtent <- extent(xmin, xmax, ymin, ymax)

 ### generate a weight grid for each lineage  START OF STEP 4

 #SDM.ras[SDM.ras < min_SDM_value] <- min_SDM_value
 if (distance_method == "model-cost") {
 cat("\n\nCreating a transition matrix based on the SDM for", group, "\n")

 model_cost.ras  <- -1 * log(SDM.ras)     # this is the original version of model cost
 model_trans.ras <- 1 / model_cost.ras  # but using the inverse here, to fit the  accCost
     # function which is based on a transition matrix. For accCost() the cost of moving between
     # cells = 1 / transition value.
     trans     <- transition(model_trans.ras, transitionFunction=mean, directions = 8)
     trans     <- geoCorrection(trans, type="c", multpl=F)
   }

   cat ("\nLooping through the lineages in group", group, "to generate weight grids\n")
   count <- 0

for (lineage in lineages) {

count <- count +1
```

```
thisLineagePoints <- thisGroupPoints[thisGroupPoints@data[, lineage_field_name] ==
 lineage, ]

# create a cost distance layer for the current lineage
if (lineage == "0") {
cat ("\nCreating distance layer for sequenced locations of unknown lineage")
} else {
cat ("\nCreating distance layer for lineage", lineage)
  }

    if (distance_method == "model-cost") {                          ## STEP 5b
     # calculates the least cost distance to the nearest lineage point
     # the result is written directly to lineage_dist_gridname
     lin_dist.ras = accCost(trans, thisLineagePoints)

    } else {
     thisLineagePoints.ras <- rasterize(thisLineagePoints, SDM.ras, 1)
     lin_dist.ras <- distance(thisLineagePoints.ras) / 1000     ## STEP 5a
    }

    # change zero values to a very small non-zero value, to avoid nodata in division
    lin_dist.ras[lin_dist.ras < min_dist_value] <- min_dist_value

    if (weight_function == "inverse_square") {              ## STEP 6b
     lin_weight.ras <- 1 / (lin_dist.ras ^ 2)
    } else if (weight_function == "inverse_cube") {
     lin_weight.ras <- 1 / (lin_dist.ras ^ 3)
    } else if (weight_function == "inverse_quad") {
     lin_weight.ras <- 1/(lin_dist.ras ^ 4)
    } else {
     lin_weight.ras <- 1/lin_dist.ras
    }

    # add the results to stack of distance layers for this group
    if (count == 1) {
     weight.stack <- stack(lin_weight.ras)
    } else {
     weight.stack <- stack(weight.stack, lin_weight.ras)
    }
    names(weight.stack)[count] <- lineage
   }

   weight_sum.ras <- sum(weight.stack)
   model.stack <- weight.stack / weight_sum.ras
   model.stack <- model.stack * SDM.ras
   names(model.stack) <- names(weight.stack)

  } else {
   SDM.ras[SDM.ras < min_SDM_value] <- min_SDM_value
   model.stack <- stack(SDM.ras)
```

```
    names(model.stack)[1] <- lineages[1]
   }

   # write results to file for this group
   model_names <- names(model.stack)

   cat("\n")

   for (i in 1:nlayers(model.stack)) {
    model.ras <- model.stack[[i]]
    model_filename <- paste("lin_model_", genus, "_", group, "_", model_names[i], ".asc",
sep="")
    model_path <- paste(target_dir, model_filename, sep="")
    cat("writing model ascii for:", model_filename, "\n")
    writeRaster(model.ras, model_path, overwrite=T)
   }
   cat ("\nLineage models done for", genus, "\n")
   time_diff <- difftime(Sys.time(), time_start, units='mins')
   cat ("\nTime elapsed:", round(time_diff,2), "minutes\n")

  }

}
```

b. <u>Lineage Modelling</u>

```
rm(list=ls())
library(raster)

#define directories
base.dir <- '/Users/natasha/Desktop/A_francei_PE/'   # modify to the base directory for your
lineage modelling

input.dir   = paste(base.dir, '/asc/', sep ='')       # location of existing lineage distribution models
output.dir    = paste(base.dir, '/asc_aligned/', sep='')    # location to save aligned lineage
distribution models
template_ext                                                                                    =
paste('/Users/Chris/Desktop/A_francei_PE/full_maxent_model/Arthroleptis/Arthroleptis_fran
cei_median.asc')  # an .asc grid with the extent to which all models will be cropped and aligned.

new_only <- TRUE  # if true, skip grids which are already in the output directory
file.pattern   <- '*.asc$'  #regex

setwd(input.dir)

input_files = list.files(path=input.dir, pattern=file.pattern, full.names=FALSE, recursive=FALSE,
ignore.case=TRUE, include.dirs=FALSE)
output_files=          list.files(path=output.dir,          pattern=file.pattern,          recursive=FALSE,
ignore.case=TRUE, include.dirs=FALSE)
```

```
template.ras = raster(template_ext)
new_extent = extent(template.ras)

raster_names <- ""

for (tfile in input_files) {
 filepath=paste(input.dir,tfile,sep='')
 outname = paste(output.dir,tfile,sep="")

 if ((!tfile %in% output_files) | (! new_only)) {
 grid.ras = raster(tfile)
 grid_ext.ras = extend(grid.ras,new_extent,value=0) # extend to the union of current grid and
  new extent
 grid_ext.ras = crop(grid_ext.ras,new_extent) # crop back to new extent
 writeRaster(grid_ext.ras,outname,overwrite=TRUE, NAflag=-9999)
 cat("\nExtended asc written for",tfile)

 # make a vector of the layer names
 if (raster_names == "") {
 raster_names <- outname
 } else {
 raster_names <- c(raster_names,outname)
   }

 } else {
  cat("\nSkipped",tfile)
 }
}

# now make a raster stack
lin.stack <- stack(raster_names)
writeRaster(lin.stack, "lin_models.stack")

setwd(output.dir)

maxval  <- stackApply(lin.stack,rep(1,nlayers(lin.stack)),fun=max,filename="max_val.asc")
sum     <- stackApply(lin.stack,rep(1,nlayers(lin.stack)),fun=sum,filename="sum.asc")
maxprop <- maxval / sum
writeRaster(maxprop,"max_prop.asc")

maxlin  <- which.max(lin.stack)
writeRaster(maxlin,"max_lin.asc")
stack_names <- data.frame(cbind(1:nlayers(lin.stack),names(lin.stack)))
names(stack_names) <- c("layer_num","layer_name")
write.csv(stack_names,"layer_names")
############################
```

c. <u>Script for calculating richness and endemism</u>

```
rm(list=ls())
```

```
library(SDMTools)
library(raster)
library(ape)
library(phylobase)
library(foreach)
library(doParallel)
library(ggplot2)

 source('/Users/natasha/Desktop/35_all_lineages/phylogenetic endemism.r')
################################################################################
###
#first define some functions

map_raster = function(raster, output_file, title) {

 p       <- rasterToPoints(raster)
 p       <- data.frame(p)
 names(p) <- c("x", "y", "Model")
 colour_gradient <- scale_fill_gradientn(colours = rainbow(15), values=p$model)
 colour_gradient <- scale_fill_gradient2(low="white", mid="yellow", high="red",
                          limits=c(min(p$Model),max(p$Model)),    midpoint=quantile(p$Model,
0.75), space='Lab')
 m <- ggplot(data=p) + geom_tile(aes(x, y, fill=Model)) + coord_equal() + labs(x=NULL, y=NULL)
+ colour_gradient

 # delete a previous file if needed
 if (file.exists(output_file)) {
  file.remove(output_file)
  cat("Previous", output_file, "removed\n")
 }

 m <- m + ggtitle(title)
 m <- m + theme(axis.title=element_text(face="bold", size="18"))
 m <- m + theme(axis.text=element_text(face="bold", size="14"))
 m <- m + theme(plot.title=element_text(face="bold", size="24"))
 m <- m + xlab("longitude") + ylab("latitude")

 png(output_file, width=image.width, height=image.height)
 print(m)
 dev.off()
 m <- NULL
}
################################################################################

max.rows       <- 10000000
core_count     <- 4 # number of cores to use for parallel steps
write_matrices  <- TRUE

# size in pixels for maps
image.width=1400
image.height=1400
```

```
#define directories
base.dir        <- '/Users/natasha/Desktop/A_francei_PE/'      # modify to the base directory for your lineage modelling
input.dir       <- '/Users/natasha/Desktop/A_francei_PE/asc_aligned/'
output.dir      <- '/Users/natasha/Desktop/A_francei_PE/PE_output/'  # output location where diversity results and maps will be saved
file_pattern    <- 'lin_model_'                 # modify this to a match the start of the name of all lineage model asc files

group_lin_file  <- '/Users/natasha/Desktop/A_francei_PE/All_lineage_list.csv'

#tree details  - this works for one genus at a time
tree.file       <- '/Users/natasha/Desktop/A_francei_PE/A_francei16S.con.tre'
outgroup        <- 'A_reichei'
preface         <- ""

genus           <- ''
output_prefix   <- 'A_francei'
threshold       <- 0.0000000000001  # this is not a species level threshold, but one used for each lineage model

####  end of parameters  ####

setwd(base.dir)
files <- list.files(path = input.dir, pattern = file_pattern, recursive = FALSE,ignore.case = TRUE, include.dirs = FALSE)

setwd(input.dir)
template.asc = read.asc('../full_maxent_model/Arthroleptis/Arthroleptis_francei_median.asc')

model_rows=nrow(template.asc)
model_cols=ncol(template.asc)

# the original version, excluding NA cells
pos <- as.data.frame(which(is.finite(template.asc),arr.ind=TRUE)) #get all points that have data

cat("\nLoading model rasters in parallel\n")

cl <- makeCluster(core_count)
registerDoParallel(cl)

pos_par <- foreach (j=1:length(files), .combine=cbind, .packages='SDMTools') %dopar% {
  tfile <- files[j]
  pos_temp <- pos
  checkname = unlist(strsplit(tfile,".",fixed=T))
  if (checkname[length(checkname)]=="asc") {   # only accept filenames ending in .asc

    cat(j)

    tasc = read.asc(tfile)                      #read in the data
```

```
   dataname=gsub(".asc",'',tfile)
   newname <- tolower(gsub(preface, "", dataname))

   cat("About to do pos\n")

   pos_temp[newname] <- tasc[cbind(pos_temp$row, pos_temp$col)]
   pos_temp[(which(pos_temp[newname]< threshold)), newname] <- 0   # set values below the
threshold to 0
   pos_temp[(which(is.na(pos_temp[newname]))), newname]    <- 0   # set the nulls to 0
   cat("\n", j, newname, "loaded")
   newcol <- data.frame(pos_temp[, newname])
   newcol <- round(newcol,3)
   names(newcol) <- newname
   newcol
  }
}

pos <- cbind(pos,pos_par)
rm(pos_par)

setwd(base.dir)
group_lin_list <-read.csv(group_lin_file, stringsAsFactors=F)
group_lin_list$lineage <- tolower(group_lin_list$lineage)
group_lin_list <- group_lin_list[group_lin_list$genus == genus,]

# read in the tree
tree <- read.nexus('/Users/natasha/Desktop/A_francei_PE/A_francei16S.con.tre')
plot(tree)
# Your tree was not in format needed (per lineage) so I dropped tips for everything except one
representative per lineage as that's what is needed
tree <- drop.tip(tree, c('A_reichei','T5748', 'T5749', 'T5751', 'T5752', 'T5753', 'T5754', 'T5755',
'T5756', 'T5760', 'T5761', 'T5763', 'T5767', 'T5768', 'T5769', 'T5770', 'T5772', 'T5773', 'T5774',
'T5775', 'GenB2', 'T0762', 'T5738', 'T5739', 'T5740', 'T5741', 'T5742', 'T5743', 'T5744', 'T5764',
'T5765', 'T5766', 'T5776', 'T5777', 'T5778', 'T5779', 'T7132', 'T7166', 'T7376', 'T7377', 'T7378',
'T7379', 'T7381'))
plot(tree)
# also renamed the tips to match the models
tree$tip.label[tree$tip.label=="T5747"] <- "lin_model_arthroleptis_francei_m1"
tree$tip.label[tree$tip.label=="GenB1"] <- "lin_model_arthroleptis_francei_mul"
tree$tip.label[tree$tip.label=="T5737"] <- "lin_model_arthroleptis_francei_nis"
tree$tip.label[tree$tip.label=="T5757"] <- "lin_model_arthroleptis_francei_m2"
tree$tip.label[tree$tip.label=="T6836"] <- "lin_model_arthroleptis_francei_chi"
tree$tip.label[tree$tip.label=="T7133"] <- "lin_model_arthroleptis_francei_lic"
plot(tree)

tree <- phylo4(tree)
plot(tree)
labels(tree) <- tolower(labels(tree))
plot(tree)

# ensure that the tree tips match the model names
```

```
model.names <- names(pos)
model.names <- model.names[-(1:2)] # names of all columns except the 1st two which are row, col
model.names <- tolower(gsub(preface,"",model.names))

model.groups <- data.frame(model.groups=vector("character",nTips(tree)),stringsAsFactors=F)

for (i in 1:nTips(tree)) {
  cat(labels(tree)[i],"\n")
  row <- group_lin_list[group_lin_list$lineage==labels(tree)[i],]
  if (nrow(row) > 0) {
    new_tip_name <- tolower(paste(row$genus, row$model_group, row$lineage, sep="_"))
    labels(tree)[i] <- new_tip_name
    model.groups[i,1] <- as.character(row$model_group)
  }
}

plot(tree)
tree <- phylo4d(tree,tip.data=model.groups)
tree

tree.names  <- as.character(labels(tree)[1:nTips(tree)]) ## original line

matched.names <- intersect(model.names,tree.names)
matched.tips  <- which(labels(tree,"tip") %in% matched.names)
cat("\nNot in tree names:",setdiff(model.names,tree.names),"\n")
cat("\nNot in model names:",setdiff(tree.names,model.names),"\n")

# a subtree containing only the tips for which there is a corresponding model
subtree <- subset(tree,tips.include=matched.tips)
subtree <- tree


if (write_matrices) {
  cat("\nWriting the site x lineage matrix to file\n")
  write.csv(pos,paste(output.dir,"sites_x_lineage.csv",sep=''))
}

# limit the occurrence table to lineages which match the tree
matching_pos_columns <- which(names(pos) %in% matched.names)
matching_pos_columns <- unique(c(1, 2, matching_pos_columns))  # ensure that row and column are included
pos <- pos[, matching_pos_columns]

cat("\n\nRemoving unoccupied cells\n")
cat("Before:",nrow(pos),"\n")
rowsums <- apply(pos[,3:ncol(pos)],1,sum,na.rm=T)
pos <- pos[which(rowsums>0),]
rm(rowsums)
cat("After:",nrow(pos),"\n")
max.rows <- min(max.rows,nrow(pos))
```

```
gc()
result <- calc_PE_from_models(subtree,pos[1:max.rows,which(names(pos) %in% matched.names)], core_count = core_count)
gc()

cat("\nDiversity calculations completed. Now writing outputs to file.")

pos_output <- cbind(pos[1:max.rows,1:2],result)
pos_output <- pos_output[,-3] # omit the site column

# add lat and long columns
cellsize <- attr(template.asc,"cellsize")
ymin <- attr(template.asc,"yll")
xmin <- attr(template.asc,"xll")

x <- ((pos_output$row - 1) * cellsize) + xmin
y <- ((pos_output$col - 1) * cellsize) + ymin
pos_output <- cbind(pos_output[,1:2],x,y,pos_output[,-(1:2)])

# add residual columns
PE_WE_mod <- lm(pos_output$PE~pos_output$WE)
pos_output$PE_WE_resid <- PE_WE_mod$residuals
PE_WE_loglog_mod <- lm(log(pos_output$PE)~log(pos_output$WE),subset=which(!is.infinite(log(pos_output$WE))))
pos_output_log <- cbind(pos_output[which(!is.infinite(log(pos_output$WE))),],PE_WE_loglog_mod$residuals)

dataframe2asc(pos_output_log[,c(4,3,10)],paste(output_prefix,"PE_WE_loglog_resid.asc",sep=""),output.dir)

pos_output$logPE <- log(pos_output$PE)
filenames <- c(paste(output_prefix,"PE.asc",sep=""),paste(output_prefix,"PD.asc",sep=""),paste(output_prefix,"WE.asc",sep=""),paste(output_prefix,"SR.asc",sep=""),paste(output_prefix,"PE_WE_resid.asc",sep=""),paste(output_prefix,"logPE.asc",sep=""))
dataframe2asc(pos_output[,c(4,3,5:10)],filenames,output.dir)

write.csv(pos_output,paste(output_prefix,"scores.csv",sep=""),row.names=FALSE)

stopCluster(cl)

setwd(output.dir)

# make some output images
PE.ras <- raster(filenames[1])
map_filename <- paste(output_prefix, "PE.png", sep="")
map_raster(PE.ras, map_filename, paste(output_prefix, "PE"))

# PE without the top 0.3%
quant_top <- quantile(pos_output$PE,0.997,na.rm=T)
```

```
PE_top.ras <- PE.ras
PE_top.ras[PE_top.ras > quant_top] <- NA
map_filename <- paste(output_prefix, "PE_without_top_0.3.png", sep="")
map_raster(PE_top.ras, map_filename, paste(output_prefix, "PE without top 0.3%"))

# log PE
PElog.ras <- raster(filenames[6])
map_filename <- paste(output_prefix, "log_PE.png", sep="")
map_raster(PElog.ras, map_filename, paste(output_prefix, "log PE"))

PD.ras <- raster(filenames[2])
map_filename <- paste(output_prefix, "PD.png", sep="")
map_raster(PD.ras, map_filename, paste(output_prefix, "PD"))

WE.ras <- raster(filenames[3])
map_filename <- paste(output_prefix, "WE.png", sep="")
map_raster(WE.ras, map_filename, paste(output_prefix, "WE"))

WElog.ras <- log(WE.ras)
map_filename <- paste(output_prefix, "logWE.png", sep="")
map_raster(WElog.ras, map_filename, paste(output_prefix, "logWE"))

SR.ras <- raster(filenames[4])
map_filename <- paste(output_prefix, "SR.png", sep="")
map_raster(SR.ras, map_filename, paste(output_prefix, "SR"))

PE_WE_resid.ras <- raster(filenames[5])
map_filename <- paste(output_prefix, "PE_WE_resid.png", sep="")
map_raster(PE_WE_resid.ras, map_filename, paste(output_prefix, "PE_WE_resid"))


map_filename <- paste(output_prefix, "4maps.png", sep="")
png(map_filename, 1600, 1200)
par(mfrow=c(2,2),mar=c(3,4,3,2))
plot(PE.ras,main="PE",col=rainbow(25,start=0.1,end=1))
plot(PD.ras,main="PD",col=rainbow(25,start=0.1,end=1))
plot(WE.ras,main="WE",col=rainbow(25,start=0.1,end=1))
plot(SR.ras,main="SR",col=rainbow(25,start=0.1,end=1))
dev.off()
```

Appendix 10

Appendix 10. Table displaying the estimates of evolutionary divergence between Arthroleptis francei 16S *sequences (n=49).*

	GenB1	T0762	GenB2	T5747	T5749	T5751	T5752	T5753	T5761	T5763	T5767	T5768	T5772	T5773	T5756	T5774	T5754	T5770	T5748	T5769	T5755	T5760	T5737	T5739	T5740	T5744	T5779	T5776	T5741	T5742	T5738	T5743	T5775	T5778	T5777	T7376	T7378	T7379	T7377	T7381	T7132	T5757	T5764	T5765	T5766	T7133	T7166	T6836	A_reichei
GenB1																																																	
T0762	0.01																																																
GenB2	0.01	0.01																																															
T5747	0.03	0.03	0.03																																														
T5749	0.03	0.03	0.03	0.00																																												
T5751	0.03	0.03	0.03	0.00	0.00																																										
T5752	0.03	0.03	0.03	0.00	0.00	0.00																																									
T5753	0.03	0.03	0.03	0.00	0.00	0.00	0.00																																								
T5761	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00																																							
T5763	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00																																						
T5767	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00																																					
T5768	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																																				
T5772	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																																			
T5773	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																																		
T5756	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																																	
T5774	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																																
T5754	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																															
T5770	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																														
T5748	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																													
T5769	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																												
T5755	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																											
T5760	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																										
T5737	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02																									
T5739	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02																								
T5740	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00																							
T5744	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00																						
T5779	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00																						
T5776	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00																					
T5741	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00																				
T5742	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00																			
T5738	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00																		
T5743	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																	
T5775	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																
T5778	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00															
T5777	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00													
T7376	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00												
T7378	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00											
T7379	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00										
T7377	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00									
T7381	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00								
T7132	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01			
T5757	0.05	0.05	0.06	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.03					
T5764	0.05	0.05	0.06	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.00				
T5765	0.05	0.05	0.06	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.00	0.00			
T5766	0.05	0.05	0.06	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.00	0.00	0.00		
T7133	0.07	0.06	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	
T7166	0.07	0.06	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.00
T6836	0.09	0.08	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.09	0.09	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.03	0.03
A_reichei	0.10	0.10	0.10	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.10	0.10	0.10	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.09	0.08	0.08	0.08	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.10	0.10	0.10	0.10	0.10	0.10	0.13