MASTERS BY RESEARCH

Species delimitation in Bitis arietans – Are there any cryptic species present?

Settle-Whittaker, Kristian

Award date:
2019

Awarding institution:
Bangor University

Species delimitation in *Bitis arietans* – Are there any cryptic species present?

Masters of Science by Research

Kristian Settle-Whittaker

Bsu237

Student Number: 500347666

Supervisor: Wolfgang Wuster

Bangor University

PRIFYSGOL
BANGOR
UNIVERSITY

Abstract

The genus *Bitis* represents the largest, most widespread and taxonomically diverse group of African snakes within the subfamily Viperinae. The focal species, *Bitis arietans*, is a continually distributed species with a range extending from sub-Saharan Africa to South Africa. *B. arietans*, currently a singular species across its distribution, has shown genetic differentiation, whether this is due to intraspecies genetic population structure or evolutionary isolated lineages is still uncertain. Coalescent methods have become a more popular choice for multi-locus species delimitation studies in recent years. Species delimitation, using these methods, in widespread species with little to no morphological differentiation or previous taxonomic subdivision such as *B. arietans* can be demanding. This study aims to follow on from previous mitochondrial and nuclear work conducted on this species, with a focus on North Africa. Novel data from five nuclear loci was used alongside previous nuclear sequences. Various software following the methodology of previous studies in this area were utilised to achieve a better understanding of the genetics of this species. Seven of the previous mitochondrial clades also proved to be distinct from BPP3 analyses with posterior probabilities of >0.99. Single locus networks revealed two areas of distinct haplotype sharing between clades however visualisation of distance matrices via principal coordinates analyses did not reveal any trends. This study has provided further evidence and clarity into the hypotheses that *B. arietans* displays multiple cryptic species across its range.

CONTENTS

1.0 Introduction

1.1 Systematics and Taxonomy

Species delimitation is achieved by analysing a range of factors such as morphology, genetics and behaviour. In the past, this field has been particularly dominated by morphology and behavioural studies, however in recent years the advance in technology now allows for large genetic studies to be conducted. The reformulation of the species concepts, whereby the term species concept has a more unified approach and takes into consideration a wide range of fields for evidence (De Queiroz 2005, 2007). Previously there have been many separate individual concepts such as morphological, genealogical (Baum & Shaw 1995), biological (Mayr, 1942) and phylogenetic (Cracraft, 1990). For a species to be delimited using any of these concepts they had to show either morphological differentiation, allelic coalescence or reproductive isolation etc.

Species delimitation is the process of reclassification or discovery of new species via the identification of species boundaries and is therefore a large part of systematics (Sites & Marshall 2004). It has long been a controversial area, lacking one known reliable process in which to accurately split species without a large amount of uncertainty. This is shown specifically when populations are allopatric, as there are difficulties in assessing multiple properties based on the biological species concept (Leaché *et al.,* 2009; Ross *et al.,* 2009; Leaché & Fujita, 2010). Furthermore, there have been disagreements as to how far along the species continuum separate lineages can be diagnosed as different distinct species (Hey, 2006; Mallet, 2008; Padial *et al.,* 2010). Despite this, most taxonomists will agree on the separation of a species if evidence is provided for several character states. Obtaining accurate knowledge of these species limits is important for research in a range of evolutionary (Fujita *et al.,* 2012) and ecological studies (Bortolos, 2008). A more unified species concept was proposed by De Queiroz (2007) after thorough review existing species concepts. This approach allows for a more definitive answer as to what qualifies a new species when attempting species delimitation. Relying on whether there is a separately evolving metapopulation lineage as the key factor in determining this. Any other criteria such as intrinsic reproductive isolation or monophyly are to be used as different lines of evidence for the existence of a new species, relevant to assessing lineage separation.

Systematics is a key biological discipline in which the aims focus around the identification of separately evolving lineages of organisms and researching the evolutionary relationships between them. Taxonomy is an area within systematics focused on the classification and naming of organisms (Sites & Marshall 2004). Since the creation of Linnaean nomenclature in 1758, taxonomists from around the globe have been discovering and naming thousands of species each year (Polaszek *et al.,* 2005; Zhang, 2008). Through time this number has increased for many groups including taxa with high levels of cryptic species, this is due to the incorporation of new tools and methods into taxonomy (Agapow *et al.,* 2004; Kohler *et al.,* 2005; Padial & De la Riva, 2006; Sangster, 2009).

In recent years, the field of taxonomy has taken a more integrative approach as it now incorporates a multitude of fields such as population biology, phylogenetics and other evolutionary disciplines when attempting to delimit species (De Queiroz, 2007; Roe & Sterling, 2007; Weins, 2007; Leache *et al.,* 2009; Petit & Excoffer, 2009; Goodman *et al.,* 2009). Despite the apparent benefit of adding information from all these fields and considering it when splitting or lumping species, integrative taxonomy by congruence has a bias for uncovering older species. Recent radiations may be overlooked by a strict consensus approach (Shaffer & Thomson, 2007). Another method of integrative taxonomy is the cumulative approach which allows taxonomists to focus on more relevant characters to the taxa in which they are studying. This method is often more suited to recently diverged species (Shaffer & Thomson, 2007), due to the speciation process and ecological gradients (Streelman & Danley, 2003; Nosil *et al.,* 2009).

The main difference between integrative taxonomy by congruence and by cumulation is that congruence approaches use lineage divergence hypotheses for species discovery. For example, discovery of multiple unlinked genetic loci isolated from any other populations/lineage or a morphological difference would each individually warrant the qualification of a species. Whereas cumulative approaches use the character hypothesis and assume any taxonomic character created by organismal attributes provide evidence for the splitting of species (Padial *et al.,* 2010). The character hypothesis follows the principal of reciprocal illumination whereby each individual character hypothesis is evaluated as to how much is agrees with the overall phylogenetic hypothesis (Lienau *et al.,* 2006). For example, morphological variation alone would provide evidence but not warrant the splitting of species alone. The cumulative method utilises a culmination of factors from multiples sources to warrant the qualification of a new species.

Despite the recent increase in the use of genealogical data when delimiting species, studies still often integrate morphological data (Trape *et al.,* 2009; Wüster & Broadley 2003). Genetic approaches to species delimitation can provide an insight into the history of species that cannot be inferred from morphological data. Morphological differences can not only be associated with inheritance but also environmental factors or convergent evolution. Comparing this to genealogical data where differences in genetic markers can be solely attributed to inheritance (Yang & Rannala 2010). When regarding taxa where there is typically little to no morphological differentiation, utilising genetic approaches is key to revealing cryptic species. Interpretation of morphological characteristics is subjective and could lead to incorrect inferences of species boundaries, whereas genetic data is more objective and easily comparable between a multitude of taxonomic groups (Fujita *et al.,* 2012).

Methods of species identification relying solely on genealogical data are not unheard of due to the cryptic nature of some species. One method, the genealogic concordance method of phylogenetic species recognition or GCPSR (Avise, 1990) proposes that if populations show evidence of distinct lineages or a lack of genetic concordance for multiple unlinked loci then it can be categorised as a new species. Most taxonomists now agree that if evidence is provided from multiple taxonomic

characters and is proven to be unlinked and fixed, then that is sufficient to prove the validity of a new species (Padial *et al.,* 2010). The use of multiple character sets provides taxonomy with stability, therefore due to the indistinction mentioned before with two of the main character sets associated with species delimitation, other areas must be considered. For this study, populations, represented as the previously described mitochondrial clades (Barlow *et al.,* 2013), which provide sufficient evidence for both distinct evolutionary lineages through analysis and display reproductive isolation from other neighbouring populations will be proposed as new species.

The process in which diversity is generated across the world is largely unknown (Dirzo & Raven, 2003; Mora *et al.,* 2011). Even well studied areas, such as the Nearctic have new cryptic species that are being discovered and described every year (Myers *et al.,* 2013; Ruane *et al.,* 2014). New methods have been developed that allow to delimit and assess these diversification processes using coalescent based phylogeography multi locus data sets. These coalescent methods provide a more objective approach to assessing these relationships, however they do not usually consider factors such as gene flow. Despite this they are more effective than single gene trees as they are often a poor representation of species histories due to the same factors.

Up until recently the commonly used method for phylogenetic analyses was a single gene sequence from a single locus of an individual organism (Heled & Drummond 2009). Data sets of multi locus sequences from large number of individuals are now being utilised with coalescent theory. This method is widely used for single species in population genetics to model intraspecific gene trees (Heled & Drummond 2009). This has now been further progressed into generating multispecies gene trees by utilising Bayesian Markov chain Monte Carlo methods which use estimates of the effective population size of extant and extinct species and or populations, this inference is only possible due to multilocus data (Heled & Drummond 2009).

It was first hypothesised by Pamilo and Nei (1988) that combining data from multiple different loci would yield better results than simply increasing the number of samples. Two samples per species or operational taxonomic unit (OTU) with enough loci can be adequate to show coalescent events for an extant species and yield estimates for population size (Heled & Drummond 2009). If only one locus is used it will not be able to detect coalescent events and will have a negative effect on the estimation of speciation dates or species topology (Heled & Drummond 2009).

There are two main types of multi-locus analyses, concatenated and multispecies coalescent. Concatenated analyses is a common approach, however it has recently been found to give wrong averages (Degnan & Rosenberg 2006), leading to poor estimations of species trees (Kubatko, 2007). Despite this, they often have high support values and are still very commonly used (Rokas *et al.,* 2003; Wu & Eisen 2008). Coalescent methods of multi locus analysis revolve around the development of various models which simulate biological events such as speciation. Unlike concatenated analyses, coalescent models can be used to create multiple gene trees using a single species tree (Wilson & Balding 1998; Rannala & Yang 2003;

Wilson *et al.,* 2003; Liu & Pearl 2007; Liu *et al.,* 2008). Coalescent methods assume no horizontal gene flow or admixture of individuals of different species as they create trees that represent the relationships of orthologous genes (Heled & Drummond 2009). However, if horizontal gene flow or admixture occurs, modern methods of phylogenetic inference can counteract this gene tree discordance but only if multiple independent loci are used.

There are many reasons as to why mitochondrial DNA is used for the first marker in many studies (Zink & Barrowclough, 2008; Edwards & Bensch, 2009). Mitochondrial DNA is the main driving force of many phylogeographical studies and therefore is crucial for species delimitation (Hurst & Jiggins, 2005; Zink & Barrowclough, 2008). The gene of choice for studies such as this is Cytochrome B (CytB), this is mainly since there is a lot of existing data already. CytB is chosen for a lot of studies due to having high sequence variability, over many years has contributed to the proposition of new classification schemes. The effective population size of mitochondrial DNA is a lot smaller, approximately 25% of nuclear DNA whereas the mutation rate is significantly larger (Brown *et al.,* 1979), implying more nuclear genes would be required to obtain the same number of polymorphic sites apparent in mitochondrial datasets.

There are further limitations of mitochondrial data that need to be considered regarding species delimitation studies. Mitochondrial DNA is inherited maternally, due to the ecology of some species such as *B.arietans* where females of the species often have a small range, whereas males travel extensively to breed. Due to this data concerning gene flow between populations would be difficult to obtain and cause distinct allopatric genetic clusters to appear (Hamilton *et al.,* 2014; Pentinsaari & Mutanen, 2017). Admixture within populations is impossible to detect from mitochondrial DNA alone as it only contains information from matrilineal ancestors. This would also lead to false results, potentially showing deep divergences of populations despite there being high rates of gene flow. When species or populations are recently diverged, mitochondrial DNA is more informative, however, nuclear DNA has stronger resolving power for species or populations with deep divergence. Rapid radiations can cause noticeable discordance for varying factors such as genetic clustering due to incomplete lineage sorting (Monaghan *et al.,* 2006; Hendrich *et al.,* 2010). When considering mapping species phylogenies, it is important to use nuclear DNA as mitochondrial data is prone to introgression which can cause high levels of gene tree discordance. Introgression has been well recorded between species, for example in *Carabus* species from Japan (Sota *et al.,* 2001). A final limitation of mitochondrial DNA is the possibility of the presence of endosymbiotic microbes causing cytoplasmic incompatibility and in turn displaying deep divergence within species in mitochondrial analyses. Microbes such as *Wolbachia* are known to do this between individuals harbouring different strains of the bacteria (Hurst & Jiggins 2005).

Alongside the advances concerning species delimitation, comes the problems associated with them, for example the models used in delimitation analyses are not exact representations of the ongoing biological processes (Carstens *et al.,* 2013). Various methods have been proposed to counter this, the main one being

conducting simulation runs of the models and software first (Carstens *et al.,* 2013). A study by Camargo *et al.* (2012) is an example of this, by designing a simulation run matching all characteristics they would use for the study, providing an increased confidence for the reliability and accuracy of the results. Furthermore, another method to increase the power of a studies results is to run prior assessments to assign any putative species into groups or operational taxonomic units (OTU), however this does increase the complexity (O'Meara, 2010). Due to this, sample assignment is conducted in many ways, Genetic clustering via Structure (Pritchard *et al.,* 2000) or Structurama (Huelsenbeck *et al.,* 2011), Gaussian clustering (Hausdorf & Hennig 2010) or simply guide trees generated from prior mitochondrial phylogenies as seen in Barlow *et al.* (2013). Bayesian species delimitation also shifts what an OTU constitutes, as described in Rannala & Yang (2003) and further suggested by Knowles & Carstens (2007), representing separate evolutionary lineages as opposed to single individuals used as models (Carstens *et al.,* 2013).

Gene flow is a process that is common across populations living in sympatry or parapatry especially within species with widespread populations. Integrative taxonomy utilising Bayesian species delimitation is currently the best method to identify new species. Unfortunately, one of the main assumptions of this method is that gene flow ceases following speciation (Burbrink & Guiher, 2014) which is rarely the case. The work carried out by Burbrink & Guiher (2014) hypothesised that these analyses could still be carried out with low to moderate amounts of gene flow being present using empirical datasets. Difficulties may arise when trying to delimit cryptic species, as to first identify if gene flow is present the different lineages must first be discovered. However, to provide reliable evidence of split lineages there needs to be data proving no gene flow. Burbrink & Guiher (2014) confirmed the ability to use Bayesian species delimitation with gene flow. They stressed the need for the use of mitochondrial data to initially determine the presence of putative lineages to then be used as guide trees for Bayesian analyses. Both *Agkistrodon* species are recently diverged, *Agkistrodon piscivorus* having diverged 1.9-2.5 Mya and *Agkistrodon contortrix* a mere 1.38 Mya. Mitochondrial DNA may be more reliable in its representation of the phylogenetic history of recently diverged species as the nuclear genes have not had the time needed to sort or accumulate substitutions (Moore & Aug, 1995; Palumbi *et al.,* 2001). Further problems may include the higher likelihood of gene discordance in younger lineages (Wakely, 2008; Degnan & Rosenberg, 2009). Bayesian species delimitation has also been conducted before with the inclusion of hybrids (Zhang *et al.,* 2011). Typically, admixed individuals that have 85% or less of their alleles associated with one population are classed as hybrids. Speciation probabilities show a trend of decreasing as the proportion of hybrids included in the data set increase (Zhang *et al.,* 2011; Burbrink & Guiher, 2014). Support for speciation also decreased with the removal of the CytB gene, further emphasising the importance of mitochondrial data (Burbrink & Guiher, 2014).

Due to the recent advances in the understanding of the patterns of biodiversity across various taxa (Jetz *et al.,* 2012; Jarvis *et al.,* 2014; Hinchcliff *et al.,* 2015; Sukumaran & Knowles, 2017), coupled with technological advances regarding genomic data and genetic analyses the identification of species is becoming more

common with sufficient evidence (Sukumaran & Knowles, 2017). Early methods of species delimitation required one of a few factors, reciprocal monophyly in gene trees, migratory cut-offs or genetic distance between candidate species or fixed sequence differentiation (Sites and Marshall, 2004; Zhang *et al.,* 2014). Recent methods using the multispecies coalescent models created by Rannala & Yang (2003) avoid the need for these factors. This method using Bayesian Phylogenetics and Phylogeography (BPP3) utilises the transmodel rjMCMC (Reversible-jump Markov chain Monte Carlo) whilst including a likelihood implementation of the multispecies coalescent to give posterior probabilities for different species trees (Flouri *et al.,* 2018). Multiple loci can be used and considerations regarding heterogeneity of gene trees (Flouri *et al.,* 2018), incomplete lineage sorting and conflicts due to ancestral polymorphism are accounted for (Zhang *et al.,* 2014).

Other methods of species delimitations have been used but have some short comings in which BPP3 seemingly does not. Two popular methods include structurama (Huelsenbeck *et al.,* 2011) and Gaussian Clustering (Hausdorf & Hennig 2010). Both do these methods do not require the use of a priori information and are flexible in terms of data which can be analysed however are ineffective in assessing evolutionary and temporal divergence of population clusters (Carstens *et al.,* 2013). SpedeSTEM is a method similar to that of BPP3 that was developed to test species boundaries (Carstens & Dewey 2010) and requires the user to assign all samples to putative lineages. However unlike BPP3, the accuracy of spedeSTEM is dependent on the quality of the gene tree estimates (Carstens *et al.,* 2013).

The advances in methodology regarding species delimitation in turn present a new set of difficulties, determining whether what has been detected is in fact evolutionary isolated lineages or in fact intraspecies genetic population structure (Sukumaran & Knowles, 2017). The population genetic structure of species in all taxa is known to be apparent in situations where gene flow is inhibited by environmental barriers (Avise, 2000). Throughout species delimitation studies this problem of misidentifying populations as different lineages is becoming a large issue (Carstens *et al.,* 2013). The crucial element of this issue is being able to distinguish between species lineages and population genetic structure as it important to help understand speciation (Rosenblum *et al.,* 2012; Dynesdius *&* Jansson, 2014). It is challenging for researchers to come to a decision for either of these conclusions. Speciation theory itself leads to a probability that isolated population lineages will in time evolve into a new species (Nosil *et al.,* 2009). This is dependent on many factors including the duration of the period of isolation and the strength of the selection processes under taken by the species. Therefore, taking an unknown period to fully develop into a new evolutionary distinct species (Rosenblum *et al.,* 2012; Dynesdius *&* Jansson, 2014; Etienne *et al.,* 2014). It is important that for each putative lineage that there are at least ten samples to allow for high a probability that deep coalescent events have been sampled (Saunders *et al.,* 1994)

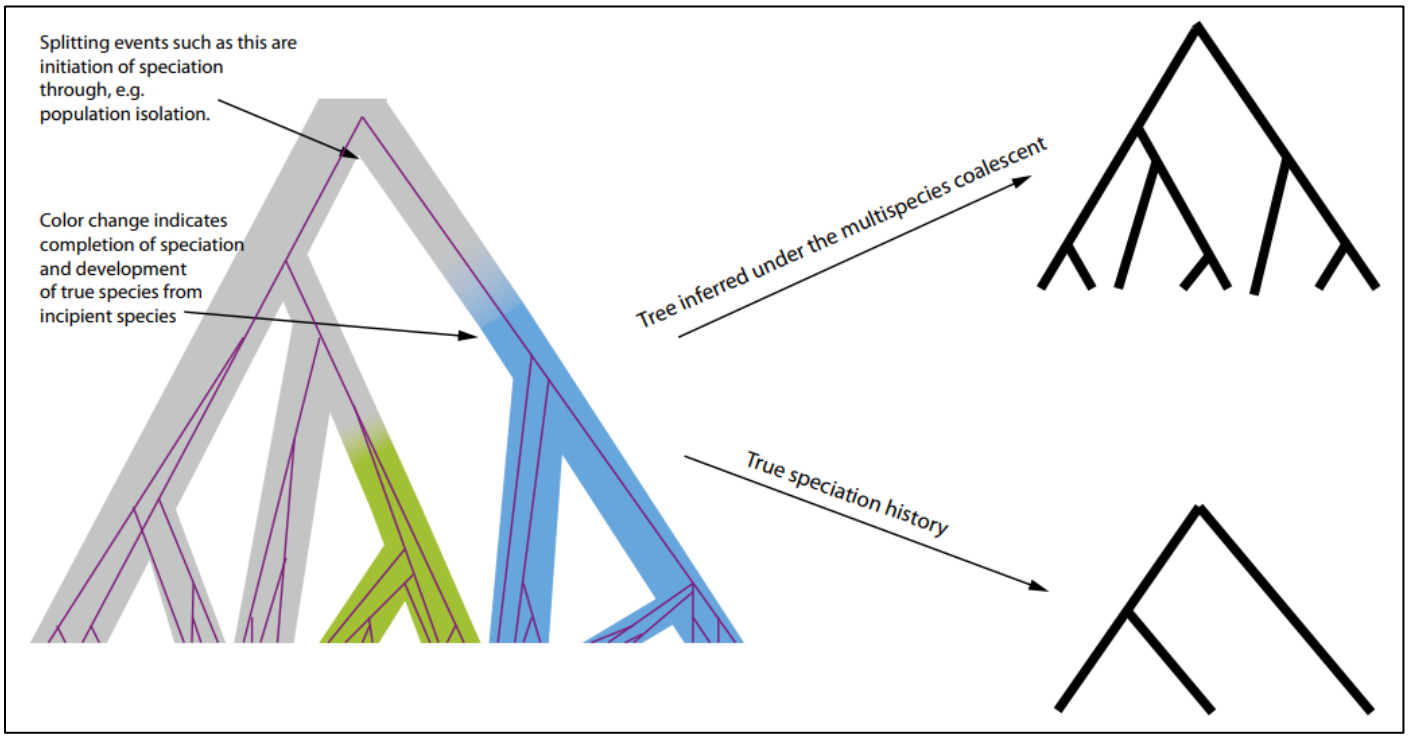

Figure 1: Taken from Sukumaran & Knowles, 2017. These trees highlight the potential issues which arise from Bayesian species delimitation, specifically Bayesian phylogenetics and phylogeography (BPP3). The left tree details the true history, colours being separate species and the purple lines showing gene genealogies, whereas the two trees on the right display a comparison between the corrected speciation history and the results from BPP3 in which population genetic structure has been mistaken for true species.

There are two categories of species delimitation, discovery and validation (Carstens *et al.,* 2013). Studies in which data is input for analyses without a prior partitioning are categorised as discovery methods of delimitation, whereas studies that have clearly delineated populations (Camargo *et al.,* 2012), existing sub specific taxonomy (Carstens & Dewey 2010) or use priors are categorised as validation methods. Assignment of priors in software such as BPP3 can also be classed as an area of difficulty due to the potential lack of knowledge in assigning realistic values for the parameters. This can be mitigated by grouping populations into OTU's as mentioned previously (Carstens *et al.,* 2013).

1.2 Introduction to *Bitis*

The genus *Bitis* represents the largest, most widespread and taxonomically diverse group of African snake within the subfamily Viperinae (Lenk *et al.,* 1999; Branch, 1998; Barlow *et al.,* 2013). The genus displays species which inhabit a multitude of habitats ranging from sand dunes to tropical rainforests. A range that stretches across all of sub Saharan Africa and into northern regions which include south-western Morocco and parts of the Arabian Peninsula (Spawls *et al.,* 1995). The focal species, *Bitis arietans* or African puff adder, favours the more arid areas of Africa, inhabiting lowland biomes such as savannah, grassland and open woodland. This excludes areas such as the equatorial rainforests of central and west Africa, the

extreme arid areas of the Namib desert and any high-altitude zones (Barlow *et al.,* 2013). Despite the broadness of this species range there is near complete morphological and ecological indistinction between the populations of *B.arietans*. Populations also exist in southern Morocco; this is the most northern part of the species range. Despite there being suitable habitat further north along the Mediterranean coast the species does not occur there. This suggests that this is determined by the climate and is the top of the latitudinal range for the species (Barlow *et al.,* 2013).

B. arietans is considered one of the most dangerous and medically significant snakes across Africa. This is due to a variety of factors including the extensive geographical range and common occurrence around people and livestock (Currier *et al.,* 2010). The highly potent haemorrhagic and cytotoxic venom displayed in this species makes it extremely dangerous and without treatment of antivenom, symptoms such as tissue necrosis and blood haemorrhaging can have fatal consequences (Warrell *et al.,* 1975). It is estimated that there are 32,000 deaths per year due to snakebite in Africa, however there are many more victims which are left with permanent disability (Kasturiratne *et al.,* 2008). Many of the fatalities are caused by *Echis ocellatus, Naja nigricollis* and the focal species *B. arietans* (Currier *et al.,* 2010).

Venom variation is a widely-studied area across multiple taxa and in turn is well documented on varying taxonomic levels particularly within species occupying wide geographical ranges (Currier *et al.,* 2010). To fully understand venom variation within species or sub species and the possible medical implications of this variation, species limits must be identified and thus species delimitation is key for this area of research.

1.3 Background of previous work

Over many years there have been many attempts at considering the intrageneric relationships within the genus *Bitis.* Some of which include Boulenger (1896), Parker (1932) and Haacke (1975). There was however one substantial study carried out by Groombridge (1980). This work was undertaken using morphological data focusing on traits such as scalation and dentition (Lenk *et al.,* 1999). A more recent study has been published detailing the phylogeny of all viper species, showing the relationships within *Bitis* (Alencar *et al.,* 2016).

Further work was carried out on the phylogeny of *Bitis* by Lenk *et al* (1999) and they found noticeable differences within *B. arietans*. Samples taken from Rwanda showed genealogical differences in mitochondrial DNA to individuals sampled from South Africa. Phylogenetic trees were generated from the sequences, they displayed these different populations within a monophyletic group, this was the case for maximum parsimony, maximum likelihood and neighbour joining trees. Furthermore, a sample was also taken from Morocco for comparison. This individual displayed less sequence divergence to the Rwanda population than the South African population in which there was considerable difference (3.85% sequence divergence compared to 5.03% of the South African sample).

There is one sub species of *B. arietans* already described, *Bitis arietans somalica* (Parker, 1949) which is located along the horn of Africa. This was described based on the differences in sub caudal scale structure (Branch, 1999), the sub-caudal scales are noticeably narrower and distally keeled (Parker, 1949). This is the only evidence of morphological difference of the species, therefore, there is little evidence for any further morphological taxonomic subdivision.

Seven bioregions within sub-Saharan Africa have been discovered (Linder *et al.,* 2012), they are based on presence and or absence of differing species of multiple taxa, both fauna and flora. These bioregions have been shown to correspond with distributions of varying taxa, examples of savannah mammals have shown distinct mitochondrial clades in either East, West or Southern Africa (Arctander *et al.,* 1999; Birungi & Arctander 2000; Van Hooft *et al.,* 2002; Muwanika *et al.,* 2003; Lorenzen *et al.,* 2007, 2010).

A study by Barlow *et al* (2013) was conducted on the phylogeography of *B. arietans*. Analysis from this work revealed multiple mitochondrial clades along with a small number of nuclear clusters in South Africa. It was hypothesised that there were multiple areas with isolated refugia across Africa, caused by Plio-Pleistocene oscillations, that harboured many species including *B. arietans.* This includes stages such as glacial maxima and the eruption of tropical forests, during warmer and more humid interglacial periods (Hewitt, 2004; Lorenzen *et al.,* 2012). During these periods, the mean winter temperature also decreased making the coastal and northern areas of South Africa more favourable. The emergence of unsuitable habitat across Africa will have isolated populations of *B. arietans* over time, this coupled with the temperature changes have contributed to the current distribution of the species.

Multiple clades were discovered from the mitochondrial DNA in South Africa, of which one was split into two subclades. Most of the clades occur parapatrically, this provides evidence of past refugial isolation as it is concerning a continuously distributed species. The mitochondrial clades revealed in Barlow *et al* (2013) show a distribution which corresponds with the proposed bioregions from Linder *et al.,* (2012). Patterns of clade distribution of *B. arietans* show similar distribution to that of the freshwater terrapin *Pelomedusa subrufa* (Vargas-Ramirez *et al.,* 2010), major clades in the West and two in the East. Another widespread African venomous snake, *Naje haje* displays the same pattern of distribution throughout the species complex (Pook *et al.,* 2009; Trape *et al.,* 2009). Despite some of the *B. arietans* populations overlapping, there was no evidence discovered displaying admixture between the clades on a mitochondrial level. However, nuclear DNA displayed multiple clusters all of which contained individuals from opposing mitochondrial clades including several examples of intermediates between clades. This suggests that there are elaborate patterns of gene flow occurring, often along narrow contact zones between clades. It is highly likely that this gene flow is largely mediated by males of the species. This male mediated gene flow has potentially also reduced or removed any evidence of past isolation from the nuclear genome. There were two more clades found that lacked any close evolutionary affinity with any of the other South African clades discovered. The first of the two clades on the Mozambique and South African border, wherein a single

individual was found to have a distinct lineage separate to any other and an additional population with a central African lineage, located along the northern Botswana border.

1.4 Aim of this project

The purpose of this study is to test if there is any evidence to suggest that there is more than a singular widespread species of *B. arietans* across Africa, using data from mitochondrial sequences and novel sequences generated from fast evolving nuclear genes. As mentioned previously, Barlow *et al.,* (2013) discovered multiple mitochondrial clades in South Africa alone and several others across the rest of the species distribution up to the Arabian Peninsula. Populations over time have become isolated and in turn must have evolved separately. Five anonymous loci will be used to generate nuclear sequences for 20 individuals of *B.arietans* across Africa. Haplotypes will be taken from these sequences and distance matrices generated similarly to Barlow *et al.,* (2013). This study will also utilise single locus networks and software BPP3 in a hope to resolve the hypothesis of whether *B. arietans* displays a singular widespread or multiple cryptic species across its large geographical range.

2.0 Method

2.1 Sampling and laboratory methods

Samples were obtained during a study carried out by Barlow *et al.* (2013) where extensive geographical sampling was carried out throughout the entirety of Africa. Tissue and Blood samples from 20 individuals originating from throughout Northern Africa and the Arabian Peninsula were utilised. 62 existing sequences of *B.arietans* from South Africa were also added into the dataset prior to analysis from Barlow *et al.* (2013).

DNA was extracted using a Qiagen DNeasy Blood and Tissue kit following manufacturers protocol. Five anonymous non-coding nuclear loci were amplified Ba05, Ba34, Ba42, Ba46 and Ba49 using primers described in Barlow *et al.* (2012) see appendix 1.

All PCR reactions consisted of a total volume per reaction of 15µl, 5.6µl of Sigma-Aldrich sterile H2O, 7.5µl of DreamTaq Green PCR Master Mix, 0.45µl of both forward and reverse primer and finally 1µl of the relevant samples DNA which is standardised at 20 ng/µl. All products for each gene were cleaned using two enzymes, TSAP (thermo-sensitive alkaline phosphatase) and Exo1 (Exonuclease 1). Some samples required further mitochondrial sequencing to assign clades. These individuals were sequenced for NADH dehydrodenase sub-unit 4 (ND4). See appendix 1 for all thermocycling conditions.

2.2 Dataset preparation

Chromatograms of the sequences were aligned using CodonCode aligner version 7.0.1. Heterozygous positions were identified through CodonCode aligner via visual inspection, IUPAC codes were then assigned. Due to the loci being non-protein coding, they are highly variable. Haplotypes of samples were estimated from the diploid nuclear sequences using the software PHASE version 2.1.1 (Stephens *et al.,* 2001). Prior to PHASE samples were run through a web tool, SeqPHASE (Flot, 2010), to adjust the file type from FASTA to PHASE input files.

Separate runs of the software PHASE were conducted for each gene, three times, all with separate random seeds to check for consistency. Each run of the programme was conducted with 1000 iterations, a thinning interval of 10 and a burn in of 100 generations. The P-values from all the pairs files were analysed and samples were not taken below a 0.8 confidence probability unless for individuals from a locality with no other samples, in which case these individuals were taken with a confidence probability of >0.6. Any positions which fell below the confidence probability threshold retained IUPAC ambiguity codes (IUPAC Committee, 1985). PHASE can be utilised to infer sections of sequence with missing, however for this study they were left as missing data. Any length heterozygote sequences had haplotypes estimated using INDELLIGENT v.1.2 (Dmitriev & Rakitov 2008).

2.3 Multilocus analyses

Firstly, Bayesian Phylogenetics and Phylogeography (BPP3) v.1.2.2 (Yang & Rannala 2010) was used with all phased sequences for all five of the nuclear loci.

The reverse jump Markov chain Monte Carlo (rjMCMC) analyses was conducted using 100,000 generations with a sampling interval of 2 and a burn in of 20,000. Following equations detailed in Yang & Rannala (2010) species delimitation algorithm 0 and a fine-tuning parameter ϵ = 15 was used, this assigns equal probabilities for each node of the guide tree prior to speciation. Before running BPP3 the prior distributions of ancestral population size and the root age must be input in the form of Theta (θ) and Tau (τ) priors. Four separate tests were conducted across the dataset, each containing a different combination of Theta and Tau priors. The first test assumed large ancestral population size and deep divergence (θ – 1 10/ τ – 1 10), the second assumed large ancestral population size but shallow divergence (θ – 1 10/ τ – 2 2000), thirdly the assumptions were both small ancestral population size and shallow divergence (θ – 2 2000/ τ – 2 2000) and the final test assumed a small ancestral population size but deep divergence (θ – 2 2000/ τ – 1 10). Utilising these varying parameters in separate runs should give a wide range of models. The following priors (θ – 2 2000/ τ – 2 2000) should utilise models favouring fewer species whereas (θ – 1 10/ τ – 1 10) is likely to favour more.

Guide trees are a fully resolved species tree that details the relationships between species which are to be included in analyses, here the mitochondrial phylogeny presented in Barlow *et al.* (2013) was used as the guide tree. Two samples did not have mitochondrial data from this study. To assign them to a clade, I therefore sequenced the ND4 gene for those samples and aligned the sequences alongside others in MEGA 7 (Kumar *et al.,* 2016) for clarification of their clade relationships. Once this was done a Newick tree was created to input into BPP3 to compete the analysis.

Network v.5.0.0.1 (Bandelt *et al.,* 1999) was utilised to show any connections between haplotypes of putative species for each locus, detailing any possible haplotype sharing. Networks of all haplotypes for each locus were then generated. Each proposed haplotype was then split into sections corresponding to putative species.

Multi-locus distance matrices were generated from nuclear gene sequences using POFAD, version 1.0.3 (Joly & Bruneau 2006). Distance matrices for each locus, were first generated from haplotype alignments in PAUP* v 4.0 (Swofford & Douglas 1993). POFAD cannot run with missing data, prior to the generation of the distance matrices, all samples without complete data for all genes were removed, reducing sample size from 81 to 66, the locus Ba46 had many unique sequences missing, therefore it was removed completely from the analysis. Sequences are loaded into the programme in nexus format and set to Kimura 2-parameter distance. Once this was done for all five loci, POFAD was run. Once run, the standardised distance matrix was taken from the output file and utilised in a principal coordinates analyses in Multi-Variate Statistical Package (MVSP) v.2.1 (Kovach, 1993). The aim of this is to generate visual representation to identify any distinct nuclear clusters which have the potential to correspond with a cryptic species.

3.0 Results

Once the sequences were run through PHASE, each sequence was split into two, a and b. Each sequence that was split had varying number of sequence pairs displaying the possible options for each heterozygous position. The highest probability sequence was chosen for each and samples with a confidence rating of less than 0.8 were not taken. Thus, removing all heterozygotes. Each locus had a varying number of sequence length, Ba05 – 594 bp, Ba34 – 517 bp, Ba42 – 544 bp, Ba46 - 764 bp and Ba49 – 496 bp.

The single locus networks provide some evidence for haplotype sharing between various putative species. Three patterns have emerged from this analysis and have proven to some extent consistent across all loci. The Angolan/Zambian (yellow), North-Western South Africa (orange) and North-Eastern South Africa (red) clades are all grouping within the networks and displaying haplotype sharing. The same pattern has emerged in the South-Eastern Africa (light green), South-Western South Africa (dark green) and the Western South Africa (light blue) clades however, this is not as clear across all loci. Finally, as expected the Eastern Kenyan (pink) and North-West African (black) clades have grouped loosely across all loci. Despite these patterns appearing, there are a lot of larger circles with high numbers of samples and affinity to a high number of clades, especially for the nuclear gene Ba34 (fig 3). This is likely to be the retention of ancestral haplotypes prior to the dispersion of the common ancestor.

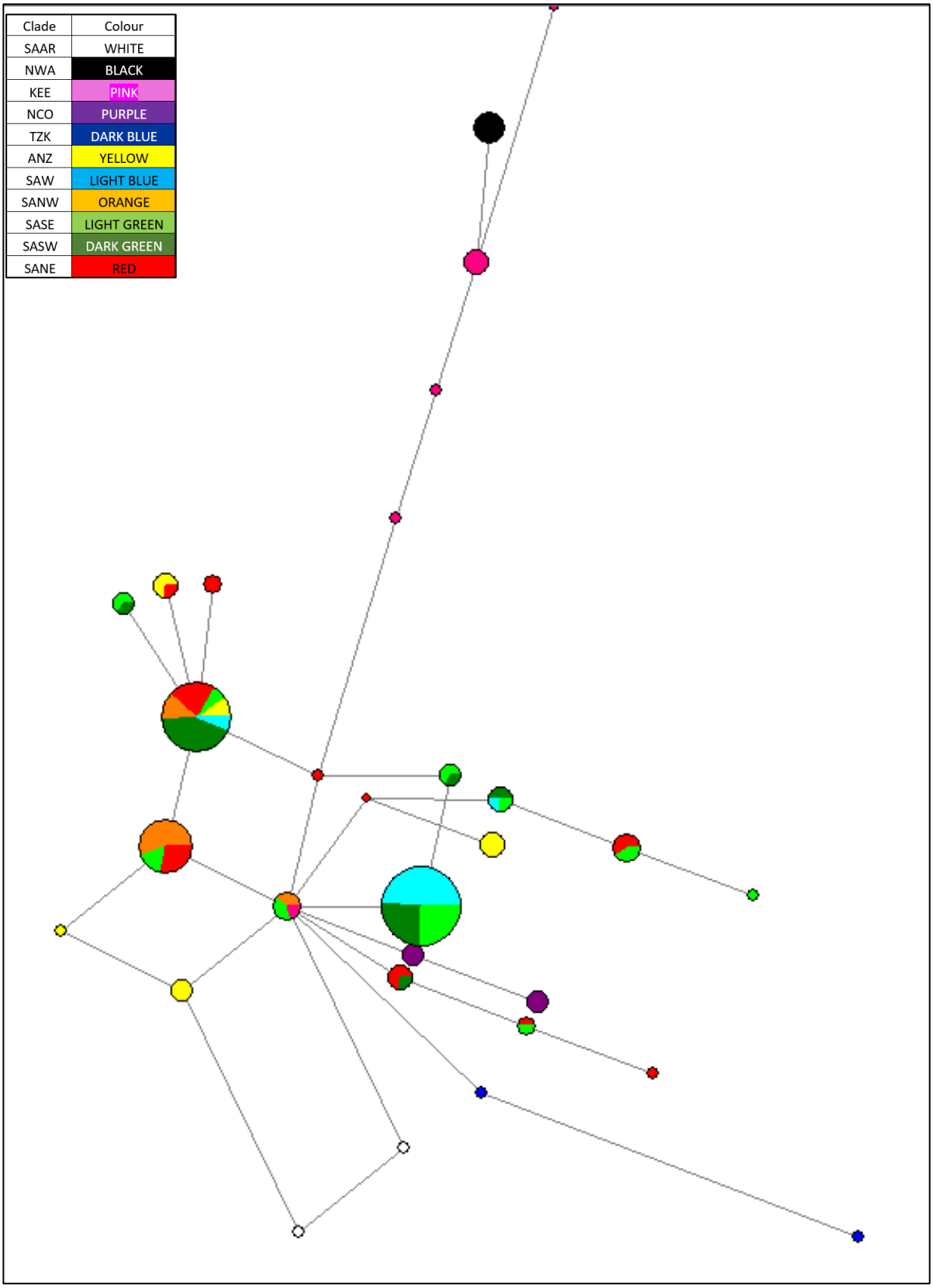

Figure 2: Network of haplotypes generated through the programme PHASE for the nuclear gene Ba05, circle size corresponds to haplotype frequency. Refer to table 2 for details of which colours correspond to putative species.

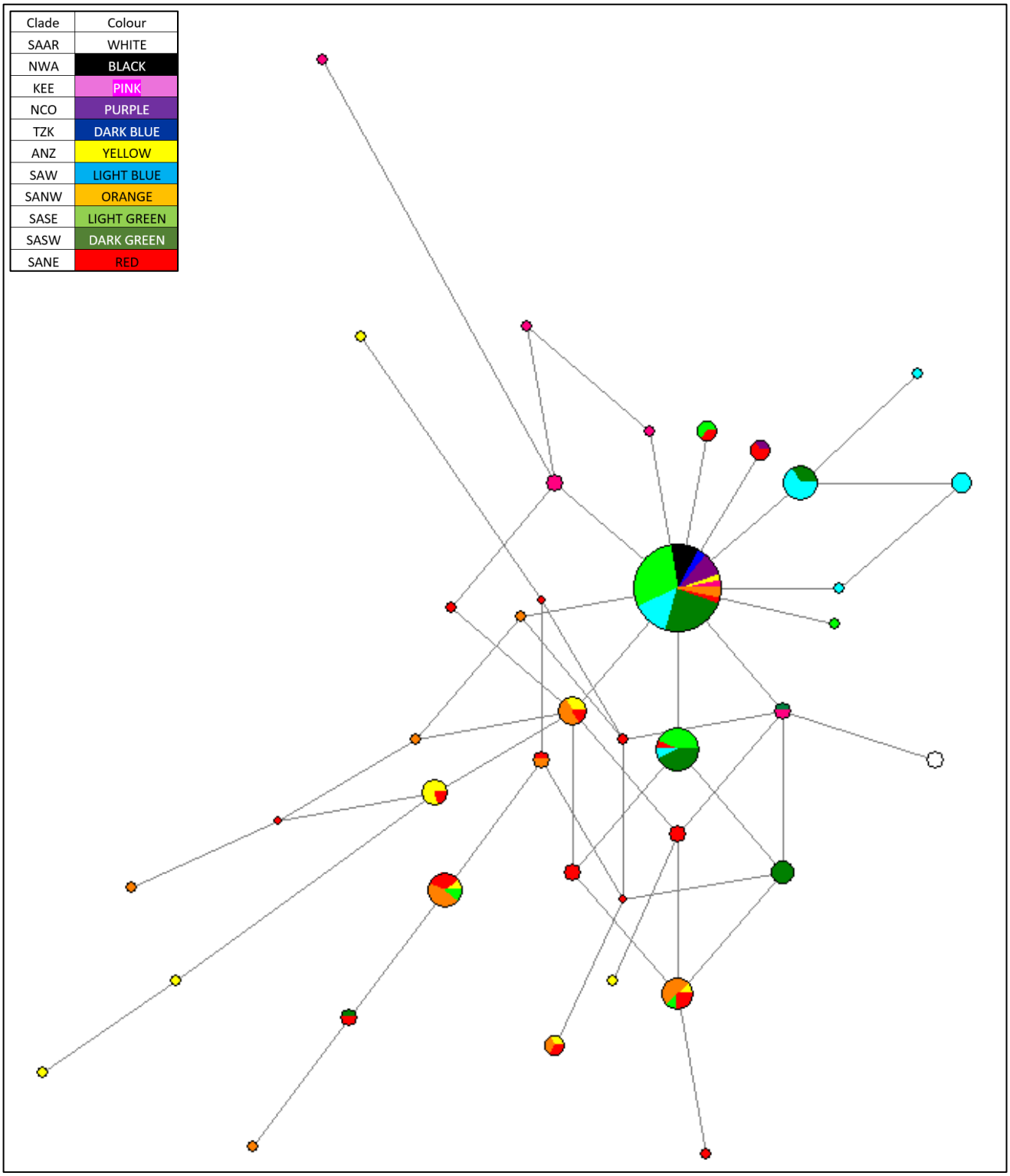

Figure 3: Network of haplotypes generated through the programme PHASE for the nuclear gene Ba34, circle size corresponds to haplotype frequency. Refer to table 2 for details of which colours correspond to putative species.

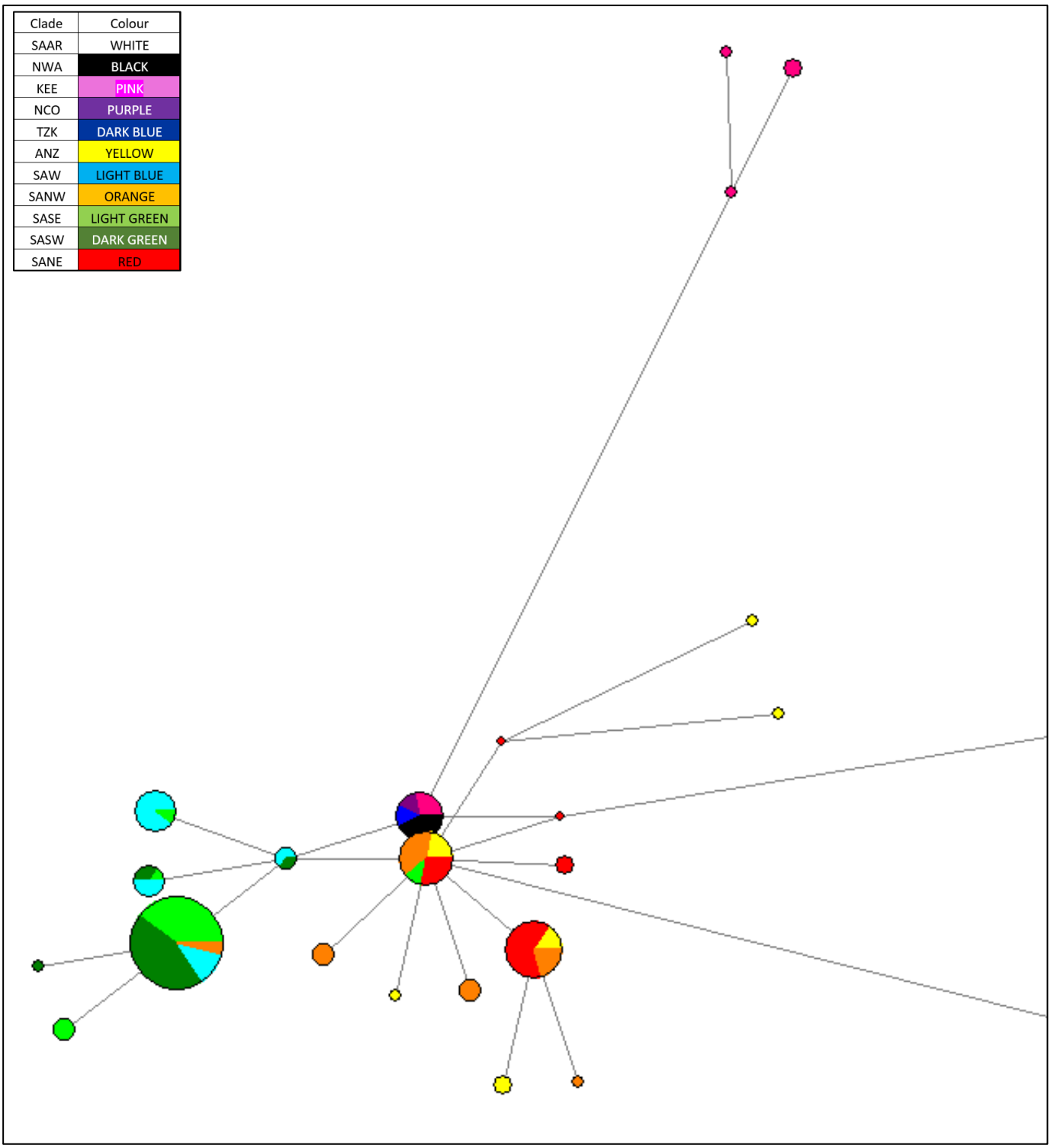

Figure 4: Network of haplotypes generated through the programme PHASE for the nuclear gene Ba42, circle size corresponds to haplotype frequency. Refer to table 2 for details of which colours correspond to putative species. Two samples not displayed in the figure, to the right, are associated with the purple (NCO) clade.

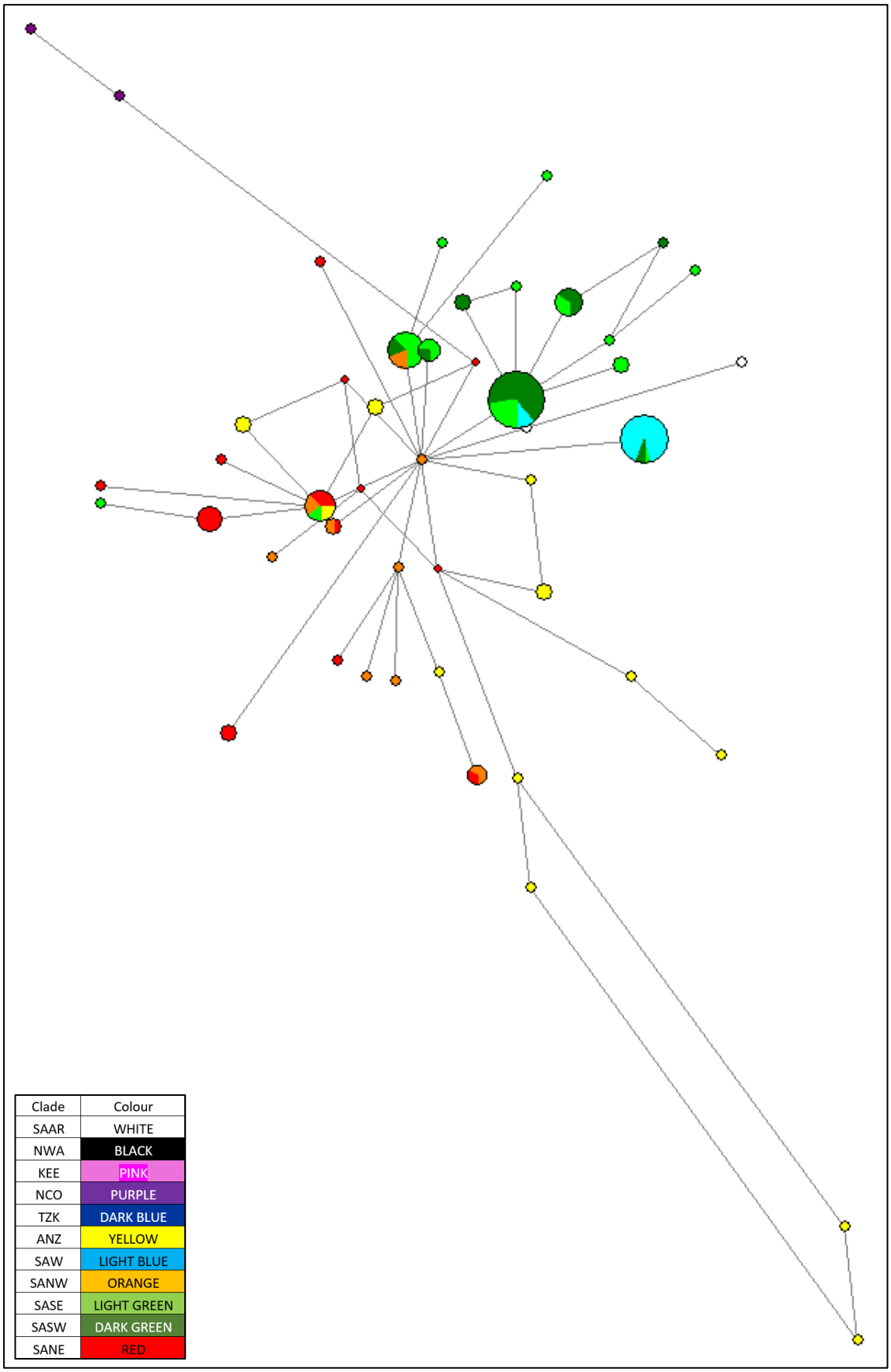

Figure 5: Network of haplotypes generated through the programme PHASE for the nuclear gene Ba46, circle size corresponds to haplotype frequency. Refer to table 2 for details of which colours correspond to putative species.

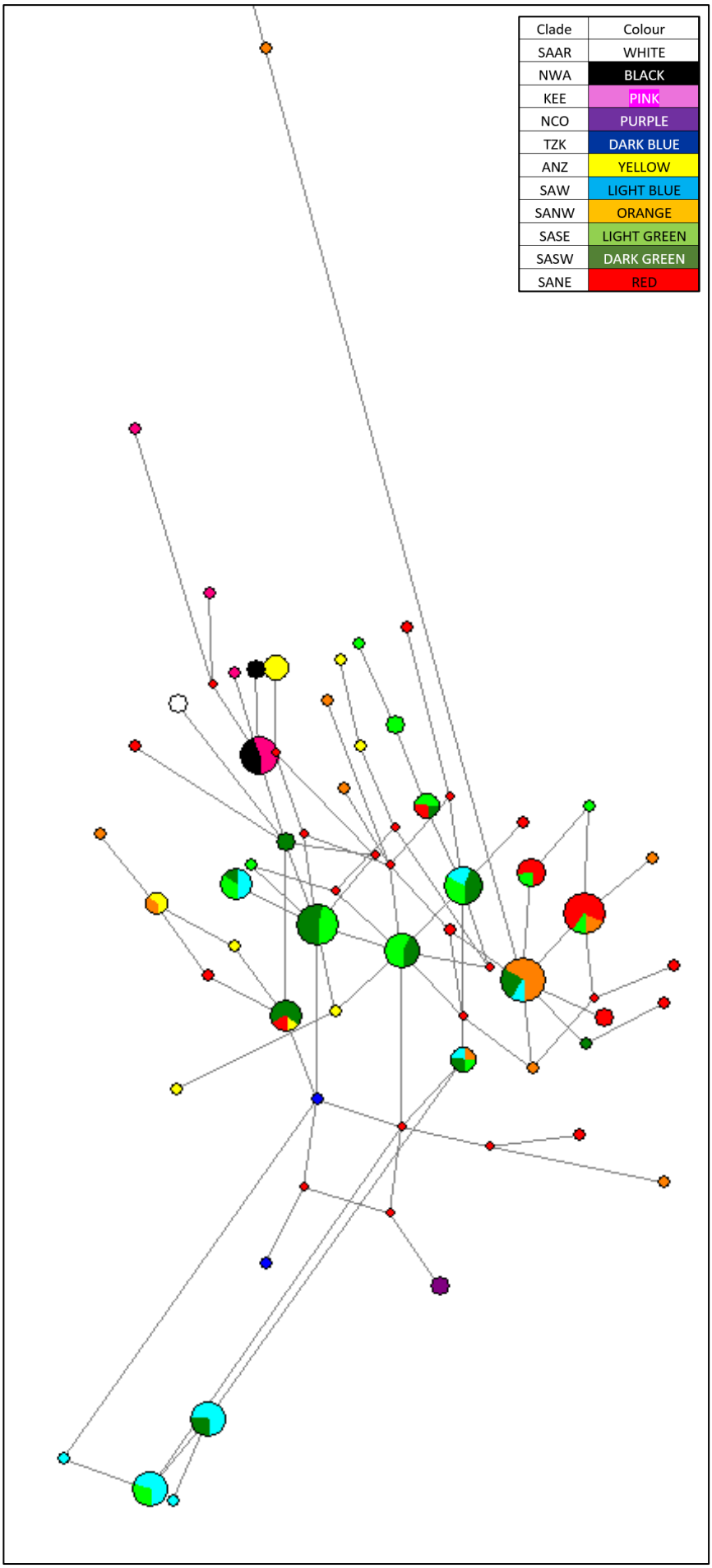

Figure 6: Network of haplotypes generated through the programme PHASE for the nuclear gene Ba49, circle size corresponds to haplotype frequency. Refer to table 2 for details of which colours correspond to putative species. One sample not displayed in the figure, to the left, is associated with the orange (SANW) clade.

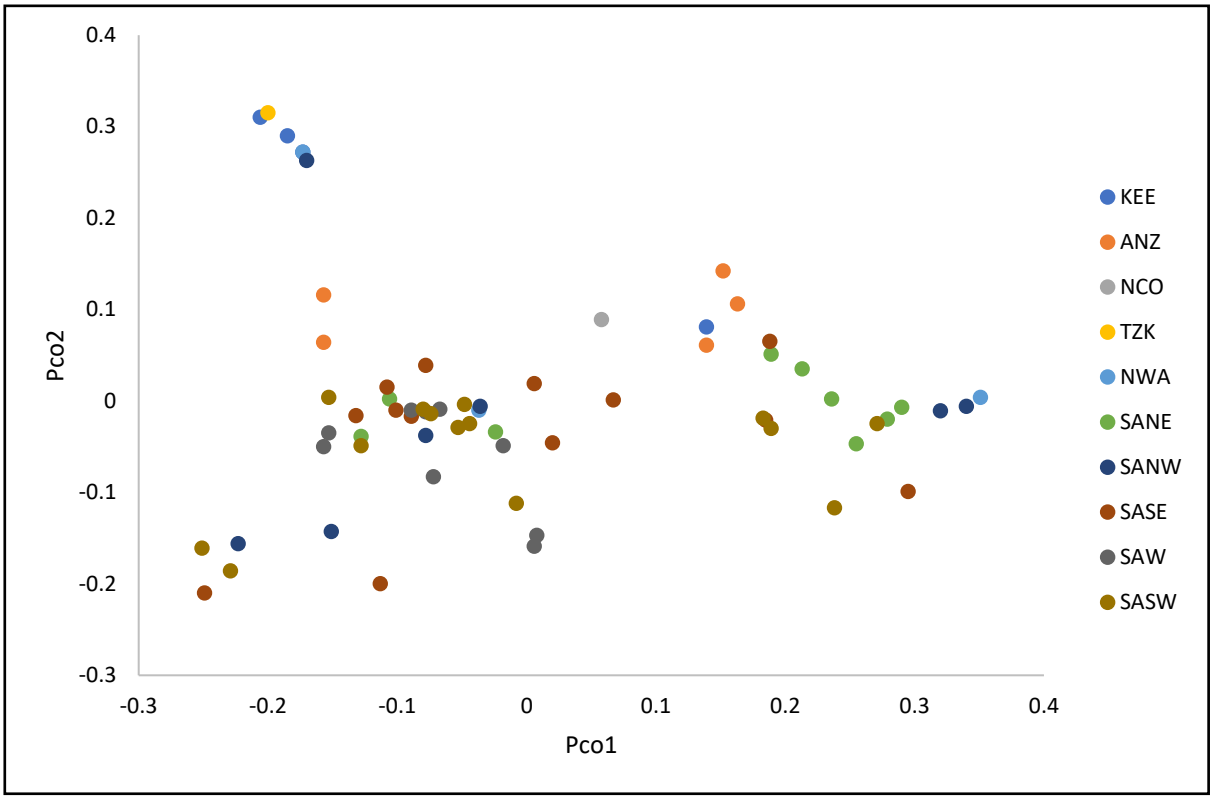

Figure 7: Principal coordinates analyses of four nuclear genes (Ba05, Ba34, Ba42 and Ba49) for 66 individuals from varying localities and clade affinity.

Analyses of distance matrices generated through POFAD were visually represented through MSVP as a principal coordinates analysis (Fig 7). There is no nuclear clustering to any extent shown through the analyses. Data points which showed previous grouping through mitochondrial phylogeny and through BPP3 analyses do not display any further clustering.

Delimited species & their posterior probabilities	Theta Prior = 1 10 taur Prior = 1 10			Theta Prior = 1 10 taur Prior = 2 2000			Theta Prior = 2 2000 taur Prior = 2 2000			Theta Prior = 2 2000 taur Prior = 1 10		
	Large ancestral population, deep divergence			Large ancestral population, shallow divergences			Small ancestral population, shallow divergence			Small ancestral population, deep divergence		
	1st run	2nd run	3rd run	1st run	2nd run	3rd run	1st run	2nd run	3rd run	1st run	2nd run	3rd run
NWA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ANZ	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
KEE	0.999	0.995	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SAAR	0.999	0.999	1.000	0.999	1.000	0.999	0.999	0.999	1.000	1.000	1.000	1.000
NCO	0.513	0.439	0.484	0.413	0.389	0.425	0.927	0.932	0.897	0.890	0.907	0.920
TZK	0.512	0.434	0.484	0.412	0.389	0.425	0.927	0.932	0.897	0.890	0.907	0.920
SAW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SASE	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.975	1.000	1.000	0.914	1.000
SASW	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.975	1.000	1.000	0.914	1.000
SANW	0.600	0.698	0.666	0.753	0.722	0.715	0.023	0.128	0.141	0.492	0.636	0.058
SANE	0.600	0.698	0.666	0.753	0.722	0.715	0.023	0.128	0.141	0.492	0.636	0.058
NCO/TZK	0.487	0.560	0.516	0.587	0.610	0.575	0.074	0.068	0.104	0.110	0.093	0.080
SANW/SANE	0.401	0.302	0.334	0.246	0.277	0.285	0.977	0.872	0.859	0.508	0.278	0.942
SASE/SASW	0.001	-	-	-	-	-	-	0.025	-	-	-	-
KEE/TZK	0.000	0.004	-	-	0.000	-	-	-	-	-	-	-
SAAR/TZK	0.000	0.000	-	0.001	-	0.000	0.000	0.000	-	-	-	-
SAAR/NCO/TZK	-	-	-	0.000	-	-	-	-	-	-	-	-
SASE/SASW/SANW/SANE	-	-	-	-	-	-	-	-	-	-	0.086	-
SASW/SANW/SANE	-	-	-	-	-	-	-	-	-	-	0.000	-

Table 1: Posterior probabilities for species status of putative species of *Bitis arietans* using different Theta and Tau priors for a varying combination of assumptions. Three runs for each assumption were conducted and are shown. All values are shown with 3 decimal places, even for insignificant data, all values >0.9 are highlighted. Data points containing no data are displayed with a '-'.

All results from the various analyses display a varying amount of evidence for the support of lumping and splitting the proposed putative species. The strongest evidence for the separation of a candidate species was for the Saudi Arabian clade.

Both the SANW and SANE clades have extremely close affinity as shown on table 1. Not only this but there seems to be two anomalies when BPP3 attempted to group them. This is likely since evolutionarily they are very closely linked. Barlow *et al.,* (2013) found them to be part of a singular Northern South-Africa clade in which he divided into two subclades.

```
((((NWA,KEE),SAAR),((NCO,TZK),ANZ)),(SAW,(SASE,(SASW,(SANW,SANE)))))
```

Figure 8: Newick tree used as a guide tree for BPP3 analyses based off previous mitochondrial work by Barlow *et al.*, (2013).

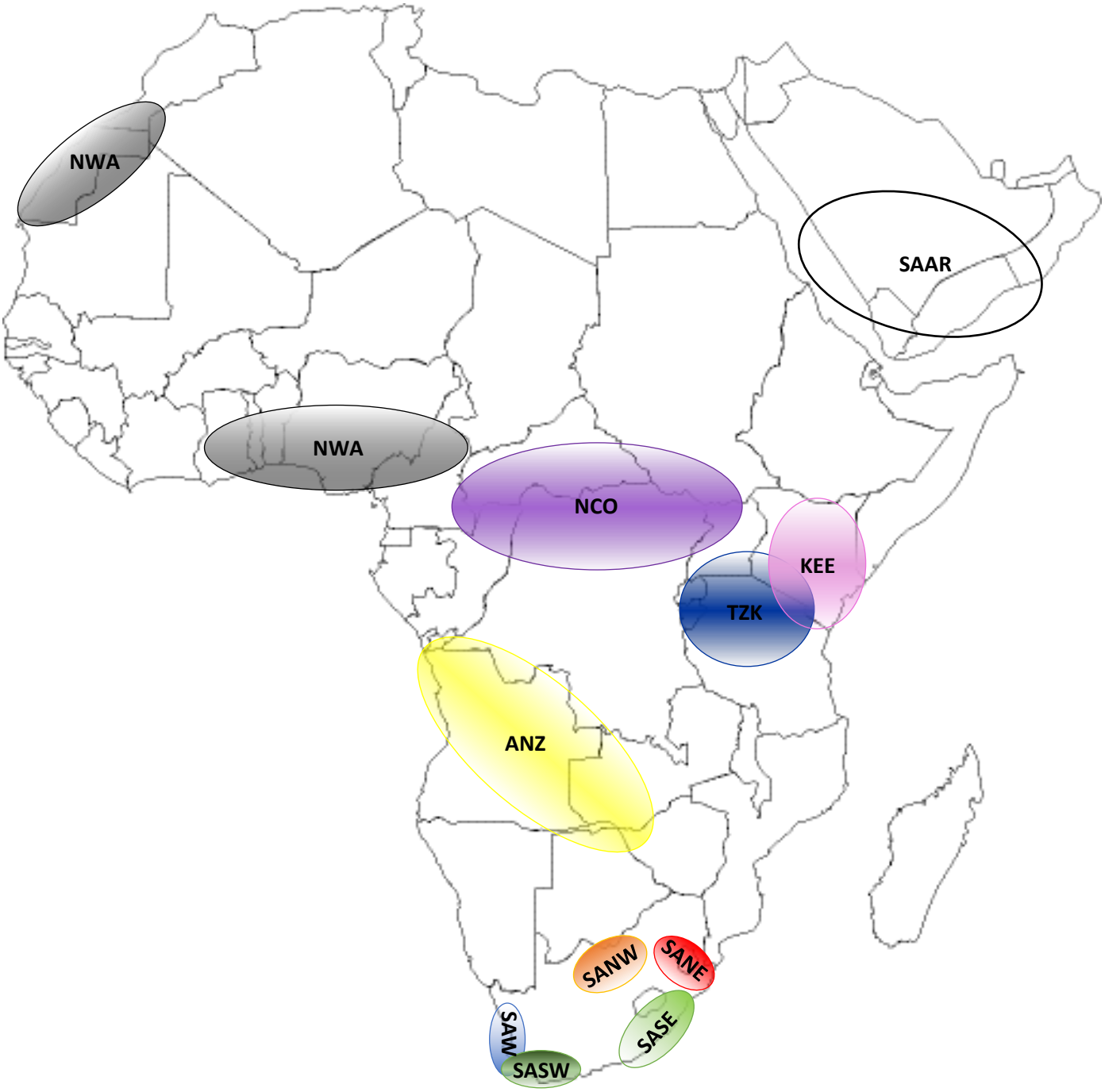

Figure 9: Distribution map of all localities/mitochondrial clades utilised from Barlow *et al.*, (2013) that have been used for this study. See table 2 below for details regarding colours and abbreviations.

Table 2: Details of the colour coding used across all five nuclear gene networks.

Locality/Clade		Colour
Saudi Arabia	SAAR	WHITE
North-West Africa	NWA	BLACK
Eastern Kenya	KEE	PINK
Northern Congo	NCO	PURPLE
Tanzanian/Kenyan Border	TZK	DARK BLUE
Angola/Zambia	ANZ	YELLOW
Western South Africa	SAW	LIGHT BLUE
North-Western South Africa	SANW	ORANGE
South-Eastern South Africa	SASE	LIGHT GREEN
South-Western South Africa	SASW	DARK GREEN
North-Eastern South Africa	SANE	RED

Table 3: Posterior probabilities for number of species displayed in dataset for each combination of Theta and Tau priors. Three runs for each assumption were conducted and are shown. All values are shown with 3 decimal places, even for insignificant data, all values >0.9 are highlighted.

	Theta Prior = 1 10			Theta Prior = 1 10			Theta Prior = 2 2000			Theta Prior = 2 2000		
BPP3 for	taur Prior = 1 10			taur Prior = 2 2000			taur Prior = 2 2000			taur Prior = 1 10		
Number of Species	1st run	2nd run	3rd run	1st run	2nd run	3rd run	1st run	2nd run	3rd run	1st run	2nd run	3rd run
P[8]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.086	0.000
P[9]	0.202	0.161	0.179	0.138	0.159	0.183	0.071	0.086	0.086	0.060	0.030	0.078
P[10]	0.486	0.545	0.492	0.558	0.571	0.495	0.907	0.791	0.790	0.497	0.311	0.868
P[11]	0.311	0.293	0.329	0.304	0.270	0.322	0.021	0.122	0.124	0.442	0.573	0.055

Results from BPP3 (Table 1) show that across all runs of the four different tests using different assumptions the North-West African (NWA), Angolan/Zambian (ANZ), Eastern Kenyan (KEE), Saudi Arabian (SAAR), Western South Africa (SAW), South-Eastern South Africa (SASE) and South-Western South Africa (SASW) clades show significant posterior probabilities of >0.99, strongly supporting nearly all the mitochondrial clades discovered by Barlow *et al.,* (2013) (Fig 9). The status of the species limits for the rest of the clades remains uncertain due to the low probability values. The only grouping of clades that displayed adequate support was between the SANW and SANE clades although there is little evidence as it was only for a singular run for two assumptions, both of which assuming small ancestral population size. Results remain inconclusive regarding the Northern Congo (NCO) and the Tanzanian/Kenyan border (TZK). Every test of all assumptions attempted to group these two clades together, none of which provided significant probability values. When separated and small ancestral population size assumed all values were >0.89 irrespective of shallow or deep divergence.

Over all the runs of BPP3, there is only a singular data point which has a significant speciation probability of above 0.9 for 10 species (table 3), assuming small ancestral population and shallow divergence, no other data points were significant.

4.0 Discussion

4.1 Species delimitation in *B. arietans* – Are there any cryptic species present?

As mentioned previously, multi-locus analysis displayed evidence of differentiation between a multitude of the pre-described mitochondrial clades. The North-Western African, Angolan/Zambian, Eastern Kenyan, Saudi Arabian, Western South Africa and both South-West and South-East South Africa clades all displayed >0.99 Bayesian posterior probability for recognition as species across all four of the different tests run through BPP3. Two of which also recognised the North-Western and North-Eastern South African clades as one species with high probability for conspecificity. Differentiation of other clades varied but were below >0.95 and therefore not inferred as distinct evolutionary lineages. Despite the recognition of these lineages as species through BPP3, there are many limitations that must be taken into consideration. Overall sample size was small for some clades, both the Eastern Kenyan and the North-western African clades only have three individuals and Arabian clade only has a single individual. Due to the fact that it is required to have two or more individuals for coalescent based analysis, no information can be inferred from the high posterior probability of differentiation in the Saudi Arabia clade.

It can be inferred from distribution alone that the Arabian clade is likely to show evidence of a distinct evolutionary lineage. The Arabian clade is isolated from all other populations of *B. arietans* due to the Red sea. It is most likely that the separation of the Arabian and African populations of *B. arietans* occurred during the Pliocene. Haq *et al.* (1987) proposed that during the Pliocene, sea levels in this area dropped periodically therefore allowing migration of populations over to the Arabian Peninsula. Previous mitochondrial work has also given accurate estimates of when the two populations diverged. Analysis conducted by Pook *et al.* (2009) set the date for divergence 4 million years ago, placing it in the early Pliocene.

Each separate single locus network generated from sequences run through PHASE showed no haplotype sharing with any other localities. This provides evidence for evolutionary separation and that over time they have evolved in isolation despite examples of exchange of species between Africa and Arabia in varanids (Portik & Papenfuss 2012) and other viper species (Pook *et al.,* 2009). Due to the possibility of full isolation of this population, it cannot be assumed without further sampling of areas along the coast of the red sea.

Leading on from this the Moroccan individuals of *B. arietans* are another example of a population evolving in potential isolation, separated from other sub-saharan populations by hundreds of kilometres. Despite this separation mitochondrial analysis showed them to group with samples from other North-West African countries, Ghana, Nigeria and Cameroon. This suggests that the separation of the two populations is recent and probably a result of vicariance. This study contained no individuals from these localities so evidence cannot be provided for further separation of this North-West African clade. Multi-locus analyses did provide further evidence of genetic distinction. However single locus networks did reveal haplotype sharing with other clades for several genes.

Alongside these two clades a further five were recognised as distinct evolutionary lineages through all runs of the BPP3 analyses. The Angolan/Zambian, Western South African, eastern Kenyan, south-east South African and south-west South African clades all identified as distinct with a >0.95 confidence probability. The singular grouping proposed by BPP3 of the south Africa North-East and-North West clades is likely due to the fact that previous mitochondrial phylogeny displayed the two clades as monophyletic (Barlow *et al.,* 2013). The results of this analysis must be taken with caution as BPP3 often overestimates number of species (Sukumaran & Knowles, 2017) despite using a range of tests. The number of novel samples sequenced for this study was also low and the range of them sparse. There is not enough mitochondrial or nuclear data around clade contact zones to fully recognise these populations as distinct species. Regarding northern clades, specifically the Saudi Arabian clade which is estimated to have diverged 4 MYA (Pook *et al.,* 2009) the runs of BPP3 with a Tau prior of 2/2000 or shallow divergence may show a more accurate representation of the species history.

The single locus networks generated for the five nuclear genes detail a few patterns of grouping and haplotype sharing between clades. Firstly, the haplotype sharing between The Angolan/Zambian (yellow), North-Western South Africa (orange) and North-Eastern South Africa (red) clades. The range of these clades meets around the Zambian and Zimbabwean border therefore the haplotype sharing has the potential to be recent and not a case of retaining ancestral haplotypes, however more sampling is required around the proposed contact zone. Secondly the South-Eastern Africa (light green), South-Western South Africa (dark green) and the Western South Africa (light blue) clades share haplotypes. Sampling in this region from Barlow *et al* (2013) displays enough evidence showing that the South-Western South Africa (dark green) and the Western South Africa (light blue) occur in sympatry and that the South-Eastern South Africa (light green) clade displays a contact zone with both these clades along the far eastern part of their range. A calibrated mitochondrial MCC tree (Barlow *et al.,* 2013) displayed the Eastern Kenyan (pink) and the North-West Africa (black) clades in a monophyletic group with high posterior probabilities despite the geographical distance between them, leading to the assumption that the dispersal of their common ancestor was recent. Despite the Northern Congo (purple) clade being geographically in between, regarding the mitochondrial MCC tree they are not closely related, and the only evidence of haplotype sharing has been in the larger circles indicating that it is likely ancestral.

A final more general pattern was made apparent after visual analysis of the haplotype networks. The northern clades in most networks appear to cluster together, more specifically the North West African (Black), Northern Congo (Purple), Tanzanian/Kenyan (Dark Blue) and Eastern Kenyan (Pink). Not only do they often cluster together but aside from one example in figure 3 they have not shared a haplotype with any of the South African clades. This may be a potential indicator of significant genetic difference between individuals from the north and south of Africa. Despite this however, more samples from the northern region of Africa would need to be analysed and compared before any assumption can be made.

4.2 Multi-locus coalescent species delimitation

As mentioned previously, the definition and identification of species and species boundaries is a highly controversial topic (Bauer *et al.,* 2011; Fujita *et al.,* 2011). When considering the splitting, or lumping of any species, results must contain evidence from a range of analyses. Multi-locus coalescent species delimitation is often met with queries as to the reliability as it often overestimates the number of species present in a dataset (Sukuraman & Knowles, 2017). Further consideration must also be used as to whether these coalescent methods alone provide sufficient evidence to back up the separation of species.

Methods have been developed that allow for the use of the programme BPP3 to delimit species without the need for guide trees. When a grossly wrong guide tree is used it can cause the programme to considerably over split (Leache & Fujita 2010; Yang & Rannala 2014). It has also shown through simulation runs that using BPP3 to infer the species tree does not cause a significant amount of in correct delimitations (Zhang *et al.,* 2014) and allows for phylogenetic uncertainty to be accounted for. However, not initially implementing a guide tree into the programme severely increases the number of potential models and species phylogenies, therefore, is only feasible for moderate sample sizes (Yang & Rannala 2014).

Simulation studies for both phylogenetic and species delimitation studies have been conducted regarding multi-locus coalescent based analyses and concluded that they are more dependent on the number of loci utilised as opposed to the number of samples used within the study (Heled & Drummond 2008; Lui *et al.,* 2009; Zhang *et al.,* 2011). However, studies that sacrifice volume of samples for number of loci often do not sample over the whole distribution including vital areas such as contact zones between candidate species. With this in mind they also often disregard spatial patterns of variation. Work by Kubatko *et al.,* (2011) is a perfect example of this, sampling 24 individuals of pygmy and Massasauga rattlesnake for 19 genetic loci. Statistical power with coalescent analyses can be found using only a few loci if sample number per population is sufficient (Yang & Rannala 2010; Zhang *et al.,* 2011). This was shown by Brown *et al.* (2012) when a clear split of species was achieved with four loci and 192 samples from 16 sites. Providing evidence that this project has statistical power as five loci were analysed for 82 individuals from 11 clades all of which had multiple separate locations within.

4.3 Further evidence for the assignment of *B. arietans* to a singular or multiple species

Species delimitation in modern day taxonomy uses a variety of approaches and forms of evidence to provide evidence for the splitting and lumping of species. To understand if *B. arietans* is in fact a singular widespread species or multiple cryptic species a range of factors must be considered, not only the genealogy. Any form of variation between different populations of *B. arietans* could be utilised as evidence.

Venom variation has been noted intra-specifically within the species *B. arietans* from different geographical origins (Currier *et al.,* 2010). Notable variation has also been

discovered in other members of the same genus as the focal species, *B. gabonica gabonica, B. gabonica rhinoceros, B. nasicornis* and *B. caudalis* (Calvete *et al.,* 2006; Calvete *et al.,* 2007). Variation in venom proteins is generally attributed to factors such as ontogeny, diet and gender. Examples of which include ontogentic shifts in coagulant activity within both *Bothrops atrox* and *Crotalus atrox* (Lopez-Lozano *et al.,* 2002; Reid & Theakson, 1978). *Echis* (Barlow *et al.,* 2009) and *Sistrurus* (Sanz *et al.,* 2006) species have been proven to recruit new toxins to account for their different prey preferences. Gender venom specificity has also been noted in *Bothrops jararaca* (Menezes *et al.,* 2006). The diversity of toxins and changes in venom composition in species such as these is a result of gene duplication (Kordis *et al.,* 2002).

Currier *et al.* (2010) hypothesised that there was the potential for these varying venoms to have a multitude of different clinical symptoms, however despite this the currently available polyvalent antivenom neutralises all *B. arietans* venoms. This work carried out by Currier *et al.,* (2010) investigated the proteome with a focus on major toxin groups to assess the possibility of variation within *B. arietans.* Samples were taken from five countries within Africa and from Saudi Arabia, from the initial results it showed the individuals from Saudi Arabia to have a significant and distinct difference in composition to all the Africa samples. Tanzanian samples have the only noticeable differences within the African countries. Saudi Arabian individuals showed substantially less SVMP (Snake venom metalloproteinases) reactivity whereas the Tanzanian samples displayed a unique intense immunoreactivity towards them. When considering other components of the venom there are more differences, the substrate degrading activity within the SVMP's varied extensively between all of the tested localities however there was no trend regarding geographical origins. A novel undescribed snake venom protein was also discovered in the venom of some Nigerian individuals. Due to the identification of this 23kDa molecular weight protein they observed the potential for two separate clades of *B. arietans* within Nigeria. More investigation of the low molecular weight proteins within the proteome of *B. arietans* has the potential to reveal more novel toxins, this is subject of further study for the laboratory in which this study was conducted. The variation that has already been noted in *B. arietans* could be used as evidence to further back up the splitting of this widespread cryptic species. Not only this but it opens avenues of research into the other locales shown as different clades in this study to see if there is any further variation or novel toxins.

It has also been hypothesised that *B. arietans* has over time developed multiple venom phenotypes and it isn't as simple as widespread pan African random genetic drift which has been previously thought. It could be regarded as a collection of multiple populations or sub-populations that have arisen over time due to varying evolutionary pressures such as the last glacial maxima and the emergence of tropical forests during inter glacial periods.

As previously mentioned in the Linder *et al.* (2012) paper, bioregions were proposed across Africa based on the presence and absence of flora and fauna of varying taxa. They show some correlation with the previous mitochondrial phylogeny produced by Barlow *et al.* (2013) in all areas apart from South Africa. It was suggested that this

was a singular region across all taxa, whereas mitochondrial work provided evidence of five clades for *B. arietans* (Barlow *et al.,* 2013). The nuclear results that have come from BPP3 in this analysis show that nearly all the proposed distinct evolutionary lineages or populations or the evolutionary relationships shown by haplotype networks do not overlap with the bioregions for *B. arietans.* Indicating that the current populations or OTU's are not influenced by the current bioregions and are a product of past refugial isolation as detailed by Barlow *et al.* (2013).

5.0 Conclusion

Multispecies coalescent methods of species delimitation used in this study and many before have proven their suitability regarding the discovery of cryptic species. Species delimitation using these methods in widespread species, with little to no morphological differentiation or previous taxonomic subdivision such as the focal species here can be demanding. This study has given further clarity into the theory that *B. arietans* displays multiple cryptic species across its range in which all are morphologically indistinct. Results provide evidence of evolutionary distinct lineages in seven previously discovered mitochondrial clades of the widespread African puff adder. These lineages can be further analysed in studies to come with an increased sampling intensity and focus on contact zones, potentially leading to definitive evidence of distinct species. Due to the lack of morphological differentiation that has been found in *B. arietans,* the inclusion of ecological and behavioural data alongside genealogical methods may be beneficial.

6.0 Reference list

Agapow, P. M., Bininda-Emonds, O. R., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C., & Purvis, A. (2004). The impact of species concept on biodiversity studies. *The quarterly review of biology*, *79*, 161-179.

Alencar, L. R., Quental, T. B., Grazziotin, F. G., Alfaro, M. L., Martins, M., Venzon, M., & Zaher, H. (2016). Diversification in vipers: Phylogenetic relationships, time of divergence and shifts in speciation rates. *Molecular phylogenetics and evolution*, *105*, 50-62.

Arctander, P., Johansen, C., & Coutellec-Vreto, M. A. (1999). Phylogeography of three closely related African bovids (tribe Alcelaphini). *Molecular biology and evolution*, *16*, 1724-1739.

Avise, J. C. (1990). Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford surveys in evolutionary biology*, *7*, 45-67.

Avise JC (2000) Phylogeography: The History and Formation of Species (Harvard Univ Press, Cambridge, MA).

Barlow, A., Baker, K., Hendry, C. R., Peppin, L., Phelps, T., Tolley, K. A., & Wuster, W. (2013). Phylogeography of the widespread African puff adder (*Bitis arietans*) reveals multiple Pleistocene refugia in southern Africa. Molecular Ecology, 22, 1134-1157.

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, *16*(1), 37-48.

Barlow, A., Pook, C. E., Harrison, R. A., & Wüster, W. (2009). Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*, 2443-2449.

Baum, D. A., & Shaw, K. L. (1995). Genealogical perspectives on the species problem. *Experimental and molecular approaches to plant biosystematics*, *53*(289-303), 123-124.

Birungi, J., & Arctander, P. (2000). Large sequence divergence of mitochondrial DNA genotypes of the control region within populations of the African antelope, kob (Kobus kob). *Molecular Ecology*, *9*, 1997-2008.

Bortolos A (2008) Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. AMBIO: A Journal of the Human Environment, 37: 114–118.

Boulenger, G. A. (1896). *Catalogue of the Snakes in the British Museum (Natural History).* (Vol. 3). order of the Trustees.

Branch B (1998). Field guide to snakes and other reptiles of Southern Africa. 3rd edn. Struik, Cape Town.

Branch, W. R. (1999). *Bitis arietans peghullae* Stewart 1973 (Serpentes: Viperidae): a valid race of puff adder?. *African Journal of Herpetology*, *48*, 15-19.

Brown, R. P., Tejangkura, T., El Mouden, E. H., Baamrane, A., & Znari, M. (2012). Species delimitation and digit number in a North African skink. *Ecology and evolution*, *2*(12), 2962-2973.

Burbrink, F. T., & Guiher, T. J. (2015). Considering gene flow when using coalescent methods to delimit lineages of North American pitvipers of the genus Agkistrodon. *Zoological Journal of the Linnean Society*, *173*, 505-526.

Calvete, J. J., Escolano, J., & Sanz, L. (2007). Snake venomics of *Bitis* species reveals large intragenus venom toxin composition variation: application to taxonomy of congeneric taxa. *Journal of proteome research*, *6*, 2732-2745.

Calvete, J. J., Marcinkiewicz, C., & Sanz, L. (2006). Snake venomics of *Bitis gabonica gabonica*. Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins bitisgabonin-1 and bitisgabonin-2. *Journal of proteome research*, *6*, 326-336.

Camargo, A., Morando, M., Avila, L. J., & Sites, J. W. (2012). Species delimitation with abc and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the Liolaemus darwinii complex (Squamata: Liolaemidae). *Evolution*, *66*(9), 2834-2849.

Carstens, B. C., & Dewey, T. A. (2010). Species delimitation using a combined coalescent and information-theoretic approach: an example from North American Myotis bats. *Systematic Biology*, *59*(4), 400-414.

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular ecology*, *22*(17), 4369-4383.

Cracraft, J. (1990). Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. *Speciation and its consequences.*, 28-59.

Currier, R. B., Harrison, R. A., Rowley, P. D., Laing, G. D., & Wagstaff, S. C. (2010). Intra-specific variation in venom of the African Puff Adder (*Bitis arietans*): Differential expression and activity of snake venom metalloproteinases (SVMPs). *Toxicon*, *55*, 864-873.

De Queiroz, K. (2005). A unified concept of species and its consequences for the future of taxonomy. Proceedings-California Academy of Sciences, *56*, 196.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology*, *56*, 879-886.

Degnan, J. H., & Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *PLoS genetics*, *2*(5), e68.

Dmitriev, D. A., & Rakitov, R. A. (2008). Decoding of superimposed traces produced by direct sequencing of heterozygous indels. *PLoS computational biology*, *4*(7), e1000113.

Dynesius, M., & Jansson, R. (2014). Persistence of within-species lineages: a neglected control of speciation rates. *Evolution*, *68*(4), 923-934.

Etienne, R. S., Morlon, H., & Lambert, A. (2014). Estimating the duration of speciation from phylogenies. *Evolution*, *68*(8), 2430-2440.

Flot, J. F. (2010). SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources*, *10*(1), 162-166.

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2018). Species tree inference with bpp using genomic sequences and the multispecies coalescent. *Molecular biology and evolution*, *35*(10), 2585-2593.

Frankham, R., Ballou, J. D., Dudash, M. R., Eldridge, M. D., Fenster, C. B., Lacy, R. C., ... & Ryder, O. A. (2012). Implications of different species concepts for conserving biodiversity. *Biological Conservation*, *153*, 25-31.

Fujita, M. K., & Leaché, A. D. (2011). A coalescent perspective on delimiting and naming species: a reply to Bauer et al. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1705), 493-495.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology and Evolution, 27: 480–488.

Goodman, S. M., Maminirina, C. P., Weyeneth, N., Bradman, H. M., Christidis, L., Ruedi, M., & Appleton, B. (2009). The use of molecular and morphological characters to resolve the taxonomic identity of cryptic species: the case of *Miniopterus manavi* (Chiroptera, Miniopteridae). *Zoologica Scripta*, *38*, 339-363.

Groombridge, B. C. (1980). *A phyletic analysis of viperine snakes* (Doctoral dissertation, City of London Polytechnic).

Haacke, W. D. (1975). *Description of a new adder (Viperidae, Reptilia) from southern Africa, with a discussion of related forms*. Staatsmuseum.

Hambäck, P. A., Weingartner, E., Ericson, L., Fors, L., Cassel-Lundhagen, A., Stenberg, J. A., & Bergsten, J. (2013). Bayesian species delimitation reveals generalist and specialist parasitic wasps on Galerucella beetles (Chrysomelidae): sorting by herbivore or plant host. *BMC evolutionary biology*, *13*(1), 92.

Hamilton, C. A., Hendrixson, B. E., Brewer, M. S., & Bond, J. E. (2014). An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: a case study of the North American tarantula genus Aphonopelma (Araneae, Mygalomorphae, Theraphosidae). *Molecular Phylogenetics and Evolution*, *71*, 79-93.

Haq, B. U., Hardenbol, J., & Vail, P. R. (1987). Chronology of fluctuating sea levels since the Triassic. *Science*, *235*(4793), 1156-1167.

Hausdorf, B., & Hennig, C. (2010). Species delimitation using dominant and codominant multilocus markers. *Systematic Biology*, *59*(5), 491-503.

Heled, J., & Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, *8*(1), 289.

Heled, J., & Drummond, A. J. (2009). Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, *27*(3), 570-580.

Hendrich, L., Pons, J., Ribera, I., & Balke, M. (2010). Mitochondrial cox1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. *PLoS One*, *5*(12), e14448.

Hey, J. (2006). On the failure of modern species concepts. *Trends in Ecology & Evolution*, *21*, 447-450.

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., ... & Gude, K. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*, *112*(41), 12764-12769.

Huelsenbeck, J. P., Andolfatto, P., & Huelsenbeck, E. T. (2011). Structurama: Bayesian inference of population structure. *Evolutionary bioinformatics online*, *7*, 55.

Hurst, G. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*(1572), 1525-1534.

Isaac, N. J., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution*, *19*, 464-469.

IUPAC Committee. (1985). Nomenclature committee of the international union of biochemistry (nciub). nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. *Biochemistry*, *229*, 75-88.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., ... & Suh, A. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), 1320-1331.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444.

Joly, S. and Bruneau. (2006). Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. *Systematic Biology*. 55(4):623-636.

Kasturiratne, A., Wickremasinghe, A. R., de Silva, N., Gunawardena, N. K., Pathmeswaran, A., Premaratna, R., ... & de Silva, H. J. (2008). The global burden

of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. *PLoS Med*, *5*, e218.

Knowles, L. L., & Carstens, B. C. (2007). Delimiting species without monophyletic gene trees. *Systematic biology*, *56*(6), 887-895.

Köhler, J., Vieites, D. R., Bonett, R. M., García, F. H., Glaw, F., Steinke, D., & Vences, M. (2005). New amphibians and global conservation: a boost in species discoveries in a highly endangered vertebrate group. *BioScience*, *55*, 693-696.

Kordiš, D., Križaj, I., & Gubenšek, F. (2002). *Functional diversification of animal toxins by adaptive evolution*. pp. 401-419

Kovach, W. L. (1993). MVSP A Multi Variate Statistical Package for IBM-PCs, ver 2.1. *Kovach Computing Services, Pentraeth, Wales, UK*.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, *56*(1), 17-24.

Kubatko, L. S., Gibbs, H. L., & Bloomquist, E. W. (2011). Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in Sistrurus rattlesnakes. *Systematic Biology*, *60*(4), 393-409.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, *33*(7), 1870-1874.

Leaché, A. D., & Fujita, M. K. (2010). Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London B: Biological Sciences*, *277*, 3071-3077.

Leaché, A. D., Koo, M. S., Spencer, C. L., Papenfuss, T. J., Fisher, R. N., & McGuire, J. A. (2009). Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proceedings of the National Academy of Sciences*, *106*, 12418-12423.

Lenk, P., Herrmann, H. W., Joger, U., & Wink, M. (1999). Phylogeny and taxonomic subdivision of *Bitis* (Reptilia: Viperidae) based on molecular evidence. *Kaupia*, *8*, 31-38.

Lienau, E. K., DeSalle, R., Rosenfeld, J. A., & Planet, P. J. (2006). Reciprocal illumination in the gene content tree of life. *Systematic biology*, *55*(3), 441-453.

Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldså, J., & Rahbek, C. (2012). The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography*, *39*, 1189-1205.

Liu, L., & Pearl, D. K. (2007). Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic biology*, *56*(3), 504-514.

Liu, L., Pearl, D. K., Brumfield, R. T., & Edwards, S. V. (2008). Estimating species trees using multiple-allele DNA sequence data. *Evolution*, *62*(8), 2080-2091.

Liu, L., Yu, L., Kubatko, L., Pearl, D. K., & Edwards, S. V. (2009). Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution*, *53*(1), 320-328.

López-Lozano, J. L., de Sousa, M. V., Ricart, C. A. O., Chávez-Olortegui, C., Sanchez, E. F., Muniz, E. G., ... & Morhy, L. (2002). Ontogenetic variation of metalloproteinases and plasma coagulant activity in venoms of wild *Bothrops atrox* specimens from Amazonian rain forest. *Toxicon*, *40*, 997-1006.

Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. Molecular Ecology, 21: 3656–3670

Lorenzen ED, Masembe C, Arctander P, Siegismund HR (2010) A long-standing Pleistocene refugium in southern Africa and a mosaic of refugia in East Africa: insights from mtDNA and the common eland antelope. Journal of Biogeography, 37, 571–581.
Lorenzen ED, Neergaard RD, Arctander P (2007) Phylogeography, hybridization and Pleistocene refugia of the kob antelope (Kobus kob). Molecular Ecology, 16, 3241–3252.

Mallet, J. (2008). Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*, 2971-2986.

Mayr, E. (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Menezes, M. C., Furtado, M. F., Travaglia-Cardoso, S. R., Camargo, A. C., & Serrano, S. M. (2006). Sex-based individual variation of snake venom proteome among eighteen *Bothrops jararaca* siblings. *Toxicon*, *47*, 304-312.

Monaghan, M. T., Balke, M., Pons, J., & Vogler, A. P. (2006). Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1588), 887-893.

Muwanika, V. B., Nyakaana, S., Siegismund, H. R., & Arctander, P. (2003). Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity*, *91*, 361-372.

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*, 145-156.

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in ecology & evolution*, *24*(3), 145-156.

O'meara, B. C. (2009). New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, *59*(1), 59-73.

Padial, J. M., & De la Riva, I. (2006). Taxonomic inflation and the stability of species lists: the perils of ostrich's behavior. *Systematic Biology*, *55*, 859-867.

Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in zoology*, *7*, 1.

Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular biology and evolution*, *5*(5), 568-583.

Parker HW (1949) The snakes of Somaliland and the Sokotra islands. Zoologische Verhandelingen, 6: 1–115.

Parker, H. W. (1932). Scientific results of the Cambridge Expedition to the East African Lakes, 1930-1.—5. Reptiles and Amphibians. *Journal of the Linnean Society of London, Zoology*, *38*, 213-229.

Payo, D. A., Leliaert, F., Verbruggen, H., D'hondt, S., Calumpong, H. P., & De Clerck, O. (2013, February). Extensive cryptic species diversity and fine-scale endemism in the marine red alga *Portieria* in the Philippines. In *Proc. R. Soc. B* (Vol. 280, No. 1753, p. 20122660). The Royal Society.

Pentinsaari, M., Vos, R., & Mutanen, M. (2017). Algorithmic single-locus species delimitation: effects of sampling effort, variation and nonmonophyly in four methods and 1870 species of beetles. *Molecular ecology resources*, *17*(3), 393-404.

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology & evolution*, *24*, 386-393.

Polaszek, A., Agosti, D., Alonso-Zarazaga, M., Beccaloni, G., Place Bjørn, P. D., Bouchet, P., & Brothers, D. J. Earl of Cranbrook, Evenhuis, N., Godfray, HCJ, Johnson, NF, Krell, F. T., Lipscomb, D., Lyal, CHC, Mace, GM, Mawatari, S., Miller, SE, Minelli, A., Morris, S., Ng, PKL, Patterson, DJ, Pyle, RL, Robinson, N., Rogo, L., Taverne, J., Thompson, FC, van Tol, J., Wheeler, QD & Wilson, EO (2005) Commentary: A universal register for animal names. Nature, 437, 477.

Pook, C. E., Joger, U., Stümpel, N., & Wüster, W. (2009). When continents collide: phylogeny, historical biogeography and systematics of the medically important viper genus *Echis* (Squamata: Serpentes: Viperidae). *Molecular Phylogenetics and Evolution*, *53*, 792-807.

Portik, D. M., & Papenfuss, T. J. (2012). Monitors cross the Red Sea: the biogeographic history of *Varanus yemenensis*. *Molecular phylogenetics and evolution*, *62*(1), 561-565.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*(4), 1645-1656.

Reid, H. A., & Theakston, R. D. (1978). Changes in coagulation effects by venoms of *Crotalus atrox* as snakes age. *The American journal of tropical medicine and hygiene*, *27*, 1053-1057.

Roe, A. D., & Sperling, F. A. (2007). Population structure and species boundary delimitation of cryptic Dioryctria moths: an integrative approach. *Molecular Ecology*, *16*, 3617-3633.

Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, *425*(6960), 798.

Rosenblum, E. B., Sarver, B. A., Brown, J. W., Des Roches, S., Hardwick, K. M., Hether, T. D., ... & Harmon, L. J. (2012). Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. *Evolutionary Biology*, *39*(2), 255-261.

Ross, K. G., Gotzek, D., Ascunce, M. S., & Shoemaker, D. D. (2009). Species delimitation: a case study in a problematic ant taxon. *Systematic biology*, syp089.

Sangster, G. (2009). Increasing numbers of bird species result from taxonomic progress, not taxonomic inflation. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*, 3185-3191.

Sanz, L., Gibbs, H. L., Mackessy, S. P., & Calvete, J. J. (2006). Venom proteomes of closely related Sistrurus rattlesnakes with divergent diets. *Journal of proteome research*, *5*, 2098-2112.

Saunders, I. W., Tavaré, S., & Watterson, G. A. (1984). On the genealogy of nested subsamples from a haploid population. *Advances in Applied probability*, *16*(3), 471-491.

Shaffer, H. B., & Thomson, R. C. (2007). Delimiting species in recent radiations. *Systematic Biology*, *56*, 896-906.

Sites JW, Marshall JC (2004) Operational criteria for delimiting species. Annual Review of Ecology, Evolution and Systematics, 35: 199–227.

Spawls, S., & Branch, W. R. (1995). *The dangerous snakes of Africa: natural history, species directory, venoms, and snakebite*. Ralph Curtis Pub.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics, 68, 978-989

Sota, T., Ishikawa, R., Ujiie, M., Kusumoto, F., & Vogler, A. P. (2001). Extensive trans-species mitochondrial polymorphisms in the carabid beetles Carabus subgenus Ohomopterus caused by repeated introgressive hybridization. *Molecular Ecology*, *10*(12), 2833-2847.

Streelman, J. T., & Danley, P. D. (2003). The stages of vertebrate evolutionary radiation. *Trends in Ecology & Evolution*, *18*, 126-131.

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, *114*(7), 1607-1612.

Swofford, D., & Douglas P.. Begle. (1993). PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1, March 1993. Center for Biodiversity, Illinois Natural History Survey.

Trape, J. F., Chirio, L., Broadley, D. G., & Wüster, W. (2009). Phylogeography and systematic revision of the Egyptian cobra (Serpentes: Elapidae: *Naja haje*) species complex, with the description of a new species from West Africa. *Zootaxa*, *2236*, 1-25.

Van Hooft, W. F., Groen, A. F., & Prins, H. H. T. (2002). Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. *Molecular Ecology*, *11*, 267-279.

Vargas-Ramírez, M., Vences, M., Branch, W. R., Daniels, S. R., Glaw, F., Hofmeyr, M. D., ... & Vieites, D. R. (2010). Deep genealogical lineages in the widely distributed African helmeted terrapin: evidence from mitochondrial and nuclear DNA (Testudines: Pelomedusidae: Pelomedusa subrufa). *Molecular Phylogenetics and Evolution*, *56*, 428-440.

Vodă, R., Dapporto, L., Dincă, V., & Vila, R. (2015). Cryptic matters: overlooked species generate most butterfly beta-diversity. *Ecography*, *38*(4), 405-409.

Warrell, D.A., Ormerod, L.D., Davidson, N.McD, 1975. Bites by puff adder (*Bitis arietans*) in Nigeria, and value of antivenom. British Medical Journal 4, 697–700.

Wiens, J. J. (2007). Species delimitation: new approaches for discovering diversity. *Systematic Biology*, *56*, 875-878.

Wilson, I. J., & Balding, D. J. (1998). Genealogical inference from microsatellite data. *Genetics*, *150*(1), 499-510.

Wilson, I. J., Weale, M. E., & Balding, D. J. (2003). Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *166*(2), 155-188.

Wu, M., & Eisen, J. A. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome biology*, *9*(10), R151.

Wüster W, Broadley DG (2003) A new species of spitting cobra (*Naja*) from north eastern Africa (Serpentes: Elapidae). J. Zool., Lond. 259: 345–359.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. Proc. Nat. Acad. Sci. USA, 107: 9264–9269.

Yang, Z., & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, *31*(12), 3125-3135.

Zhang, C., Zhang, D. X., Zhu, T., & Yang, Z. (2011). Evaluation of a Bayesian coalescent method of species delimitation. *Systematic biology*, *60*(6), 747-761.

Zhang, Z. Q. (2008). Contributing to the progress of descriptive taxonomy. *Zootaxa*, *1968*, 65-68.

Zhang, C., Rannala, B., & Yang, Z. (2014). Bayesian species delimitation can be robust to guide-tree inference errors. *Systematic biology*, *63*(6), 993-1004.

7.0 Supplementary Information

7.1 Appendix 1

Table 3: Nuclear gene primer details.

	Primer Sequences	
Primer Name	Forward	Reverse
Ba05	CATGTGACCAACCCGATTTT	CAAGTGACTGATCTGCCATGA
Ba34	CGAGGAGTGATGGGAGTTGT	CTAGCACCTCGTGGCAGAT
Ba42	CAATTAAACAAAGTTGTCCTTTGG	GAGTGAAAGGCAGGAGGTTG
Ba46	TACCATGAGTCCCCCTTCAG	GGTCCGTCAGGTGTTGAAAG
Ba49	CACCAAGACTCACAGGCAGA	TCCACGCTGCTTTAATGTTG

Table 4: ND4 primer details.

Primer Name	Primer Sequence
NADH4	CACCTATGACTACCAAAAGCTCATGTAGAAGC

Table 5: Thermocycling conditions used for all nuclear genes amplified.

Thermocycling conditions			
Step	Time (Minutes)	Temperature (°)	Repetitions
Denaturing	02:00	94˚	0
	00:30	94˚	40
Annealing	00:30	58˚	
Extending	01:00	72˚	
	05:00	72˚	0

Table 6: Thermocycling conditions used to amplify the ND4 gene.

Thermocycling conditions			
Step	Time (Minutes)	Temperature (°)	Repetitions
Denaturing	02:00	94˚	0
	00:30	94˚	40
Annealing	00:45	50˚	
Extending	01:00	72˚	
	05:00	72˚	0

Table 7: Number of samples used by Barlow *et al.* (2013) to define the mitochondrial clades used as a guide tree to infer putative species for this study.

Clade	Number of Samples
Saudi Arabia	7
North-West Africa	6
Eastern Kenya	6
Northern Congo	3
Tanzanian/Kenyan Border	3
Angola Zambia	12
Western South Africa	24
North-Western South Africa	16
South-Eastern South Africa	31
South-Western South Africa	31
North-Eastern South Africa	30