Comparative phylogeography of five swallowtail butterfly species (Lepidoptera: Papilionidae) in South Africa: ecological and taxonomic implications

A thesis submitted in fulfilment of
the requirements of the degree of

MASTER OF SCIENCE
of
RHODES UNIVERSITY

By Götz-Georg Neef

March 2014

Abstract

With current biota under constant threat of extinction, it is important to ascertain where and how biological diversity is generated and partitioned. Phylogeographic studies can assist in the identification of places and processes that indicate the origin and maintenance of biodiversity. Forest fragmentation has a big effect on local extinction and loss of genetic diversity of forest-restricted taxa, along with divergence and speciation of forest biota. This study aims to understand the effects of these processes on a number of forest-dwelling butterflies using a comparative phylogeographic approach.

Mitochondrial DNA of five different *Papilio* species with different degrees of forest specificity was analysed using phylogenetic methods. In addition, the subspecific taxonomy of *P. ophidicephalus* was investigated using morphometrics of discal spots on the wings and nuclear DNA analysis along with mitochondrial DNA analysis.

The results show that the forest-restricted species (*P. ophidicephalus* and *P. echerioides*) have more genetic structure and less genetic diversity than the more generalist species (*P. dardanus*, *P. demodocus* and *P. nireus*). This could be due to inbreeding depression and bottlenecks caused by forest fragmentation. As forest patches become smaller, the population size is affected and that causes a loss in genetic diversity, and increasing habitat fragmentation disrupts gene flow.

The intraspecific taxonomy of *P. ophidicephalus* is far from revealed. However, this study shows there is evidence for the different subspecies when comparing morphological results and genetic results. From the evidence provided here it is suggested that *P. ophidicephalus* should be divided into two separate species rather than five subspecies.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

bp	base pairs
BI	Bayesian Inference
CI	Consistency Index
COI	Cytochrome Oxidase I
DFA	Discriminant Function Analysis
dH20	Distilled Water
DNA	Deoxyribonucleic acid
DNAsp	DNA Sequence Polymorphism
dNTP	Deoxyribonucleotide
GTR	Generalised time-reversible
H	Number of haplotypes
Hd	Haplotype diversity
IBD	Isolation By Distance
IOCB	Indian Ocean Coastal Belt
ITS	Internal Transcribed Spacer
KYR	Thousand years
KZN	Kwazulu-Natal
LDG	Latitudinal Diversity Gradient
LGM	Last Glacial Maximum
MEGA	Molecular Evolutionary Genetics Analysis
MgCl2	Magnesium Chloride
mtDNA	Mitochondrial DNA
MY	Million Years
nDNA	Nuclear DNA
NMB	Northern Mistbelt
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
RI	Retention Index
S	Number of polymorphic sites
SC	Southern Coastal
SMB	Southern Mistbelt
π	Nucleotide diversity

Acknowledgements

There are many people I would like to thank for the assistance and support. First of all I would like to thank my supervisors Prof. Martin Villet and Prof. Nigel Barker for their endless support and guidance for this research. Prof Barker, thank you for all the support in the laboratory, and helping me to plan my sampling trips. Martin, I would like to thank you for all your input to the research. Some great ideas have come up in discussion.

I would also like to thank my family for their endless support to pursue my studies in the field I have always been passionate about. Growing up on a farm allowed me to start of this passion about nature at a young age and ever since I started collecting insects and trying to collect plants and press them as a small boy, my parents told me that is what I have to study one day, and so I did!

My sincere gratitude to my lab partners: Garreth Keevey who thought me all there is to know about extracting DNA and PCRs, which didn't always work and provided endless frustration; Alistair Barker for his time to do all my sequencing of all my samples, without which this research would not have been possible; and Justin Nicolau for helping me to catch some of the butterflies, which was not an easy job.

I would like to thank all my housemates, Richard Danckwerts, Richard Moss and my cousin Alex Winkler for their support and distractions especially in the final stages of the thesis. The fishing sessions on the dam some afternoons were needed to clear my head and keep me sane. Thank you to my friends Sean Gallagher, Devin Ayers, Claire Adams, Matt Janks and Duncan Stodardt for the coffee breaks and discussions about my work. I hope I didn't bore you with all of the butterfly talk.

Partial funding was provided by the National Research Foundation (NRF) of South Africa. Any opinion, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Research Foundation. I would also like to thank TUCSIN (The University Centre for Studies in Namibia) for their funding in the second year of my masters; I am truly grateful for their support.

Disclaimer

Any opinion, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Research Foundation.

Declaration

The following thesis has not been submitted to any university other than Rhodes University, Grahamstown, South Africa. The work presented here is that of the author.

_____ Date: _____

1 General Introduction

Forests in South Africa

Indigenous forest is the smallest biome in South Africa. Despite its size, this biome contains a high proportion of the country's biodiversity; only the Fynbos biome has a higher biodiversity (Lawes *et al.*, 2000a). Forests cover about 3000 km^2 or about 0.1% of the country's land surface, and are highly fragmented, some forest patches being smaller than one square kilometre (Mucina and Rutherford, 2006). The first comprehensive country-wide classification of indigenous forest was presented in a report by von Maltitz *et al*. (2003). This classification contains seven forest groups, comprising 24 forest types. This classification system was updated by Mucina *et al*. (2006).

Human impact has greatly affected these natural forests because they are poorly managed and mostly ignored for conservation purposes. In South Africa it is suggested that about 43% of the forest biome has been transformed and most of that has happened fairly recently (over the last 100 years) (Eeley *et al*., 1999). Many of these forests are also invaded by introduced plants that frequently have a deleterious effect on the biodiversity of the forests; the greater the area invaded, the greater the loss of indigenous biodiversity. However, fragmentation is not only due to anthropogenic effects; natural events also play a part. Climatic change plays the biggest role in the fragmentation of forests (Roy *et al*., 2001), and the biological diversity of forests decreases with increasing fragmentation. This raises the question of how much fragmentation and loss forests can tolerate without losing characteristic ecological functions (Danielsen, 1996).

African Forest Biodiversity

During the last glacial maximum there were more moisture sources closer to the Atlantic in comparison to the present day, where most of the moisture comes from the Indian Ocean by easterly winds. From the map reconstructed in Partridge *et al.* (1999), most of Southern Africa was covered by dessert on the western side and by xerophytic woodland on the eastern side in the LGM (21-18 kyr). The eastern coastline was covered by temperate

evergreen forests. Comparing that to the Holocene altithermal (9-7 kyr), there is extensive change in vegetation. By that time, the western coastline's vegetation became tropical rain forests and some patches of xerophytic woodland occurred further inland on the eastern side of South Africa (Partridge *et al.*, 1999).

Lawes (1990) suggested that many forest-restricted mammals follow the extent of the development of Afrotemperate and Indian Ocean Coastal Belt (IOCB) forests. Lawes *et al*. (2000) compared species richness against regional richness and found that local richness of butterflies and mammals is dependent on regional richness with some degree of variation.

Role of phylogeography in elucidating forest histories

Phylogeography is biogeography on a smaller temporal scale that focuses on population-level response to environmental change operating on an evolutionary time scale (Avise, 2000). It is also a link between microevolutionary disciplines (like ethology, demography and population genetics) and macroevolutionary disciplines (like palaeontology, historical geography and population biology) (Avise, 2000).

A study done by (Lawes, 1990) on the distribution of samango monkeys suggests that the radiation of southern African forest mammals followed to two possible routes, which relate directly to the development of Afromontane and IOCB forests in the southern African subregion.

A study on *Sheppardia* forest birds in tropical Africa suggested that all of the species of robin evolved from a common ancestor around the time of the Miocene/Pliocene transition Roy *et al*. (2001). Intraspecific speciation indicates a more recent population expansion in the upper Pliocene correlated with major climate variation and vegetation change. This was caused by tectonic uplift in central Africa, which changed the rainfall patterns and isolated the eastern forests.

Hughes *et al*. (2005) studied the speciation of two *Streptocarpus* species across forests in southeastern South Africa, investigating the population genetic structure and phylogenetic patterns within a species complex. They invoked Pleistocene habitat expansion and contraction cycles, dispersal and adaptation to lower temperatures to explain patterns of genetic diversity. The results show that the costal species are Pleistocene relicts and have a higher genetic diversity than the species occurring at higher altitudes and lower temperatures.

A study on dwarf chameleons by Tolley *et al*. (2006) suggested that both vicariance and dispersal shaped the extant of the dwarf chameleons of the Cape Floristic region.

Chameleons are highly dependent on vegetation, as their survival depends on camouflage for protection from predators and to obtain food. Thus, a change in vegetation structure has a direct consequence on survival. Not all chameleons are restricted to the fynbos biome; some are generalist species. The study also suggested that the most recent lineages diversified about 3-6 million years ago, which means that the origin of most present-day lineages were in the Plio-Pleistocene. This corresponds to the reduction of forests in the region and the establishment of the fynbos biome.

McDonald and Daniels (2012) studied the impacts of climatic oscillations on historical Afromontane forests in the Western Cape with the use of a phylogeographic study of the Cape velvetworm. They inferred that forest expansion and contraction of forests led to the diversification of the velvetworm in the Pliocene. Their results suggest that that there are three geographically discreet, genetically distinct putative species.

The studies above all dealt with the diversification of different forest taxa across Africa. It is evident that forest expansions and contractions in the Plio-Pleistocene have shaped the diversification of a variety of different organisms including mammals, birds, reptiles and invertebrates.

Aim

Managing and conserving forests is made difficult by the fact that the processes that determine their community structure are poorly understood (Lawes *et al*., 2000a). Large scale processes such as species radiations (Lawes, 1990) and climate-driven extinction events (Balmford, 1996) can determine the species pool from which local communities are assembled (Lawes *et al*., 2000a). It is therefore possible that community structure depends on biogeography rather than on local processes (Cornell and Lawton, 1992). To determine if local ecological or regional biogeographic factors explain forest biodiversity, one needs to determine if forest communities are saturated with species (Lawes *et al*., 2000a).

The consequences of forest fragmentation include local extinction of forest dwelling taxa, loss of genetic diversity (through genetic drift), and (with sufficient time and isolation) divergence and speciation within the forest biota. The latter process could explain the highly diverse biota of forests. However, this has neither yet been investigated in a structured and strategic manner, nor across multiple taxa. The research proposed here investigates the effects of these processes on a number of forest-dwelling butterflies using a comparative phylogeographical approach. Studies of individual, strategically chosen target species will

illuminate their evolutionary and population history, and collectively these studies will help us to unravel the history and mechanisms underlying the richness of biodiversity in South Africa's forest biome.

The over-arching hypothesis to be tested is that different forest-restricted taxa will show similar patterns and depth of divergence in gene trees, and that divergence times will reflect the evolution of the forest biome in southern Africa. In this study, the taxon selected to be a proxy for determining forest history are butterflies of the genus *Papilio.*

Papilio (swallowtail butterflies)

Large, brightly-coloured swallowtail butterflies are one of the most taxonomically well known and broadly studied insect groups. They are also recognised model organisms in evolutionary and conservation biology, genetics and ecology (Collins and Morris, 1985; Zakharov *et al*., 2004a; Condamine *et al*., 2013). There is a strong scientific foundation for research on the genus *Papilio* because the taxonomy, life history, larvae and plant associations of about 200 species worldwide have been fairly well documented (Caterino and Sperling, 1999).

Up until 2004, the phylogeny of this diverse group was not highly resolved and many hypotheses involving the phylogeny were poorly supported. To resolve questions around the genus *Papilio*, additional effort needs to be invested in accurate phylogenetic studies (Zakharov *et al*., 2004a).

Calibrating a molecular clock for a given phylogenetic hypothesis can be contentious. There are two ways in which such a clock can be calibrated. One uses the fossil record and tries to assign minimum ages to the first common ancestor. The second method uses vicariance events. However, there are several setbacks with each of the methods. Calibrating the clock using vicariance events presents a problem for the absolute dating of a specific event. The method involving fossils is considered reliable; there are some hidden problems such as incorrect systematic placement of the fossil (Sanderson, 1997). Fossils are also constrained to a minimum age and cannot fix dates of internal nodes. Molecular dating of swallowtail butterflies have been done in a number of studies (Clark and Vogler, 2009; Condamine *et al*., 2013; Zakharov *et al*., 2004a). Two previous studies done on molecular dating of swallowtail butterflies include species which have been included in this study.

In the study by Zakharov *et al*. (2004), 206 *Papilio* species were sampled, which represents about 25 percent of all species in the genus. The phylogeny of *Papilio* is far from

resolved and many biogeographic hypotheses are poorly supported. To resolve this issue, an accurate phylogenetic resolution is needed as well as approximate dates for some of the significant events. Of relevance to this study is the split between *P. demodocus* and *P. demoleus*, which is dated between 13 and 15 million years ago. Furthermore, they date the split between *P. dardanus* and *P. constantinus* to between 19 and 23 million years ago. The substitution rate for the COI gene in genus *Papilio* is estimated to be $7.8\text{-}10.2\text{x}10^{-9}$ per site per year which is 2-4 times slower than the standard rates for *Drosophila.* The calibration was based on *Papilio* fossils and geology (Zakharov *et al*., 2004a).

Clark and Vogler (2009) undertook a study on the African *P. dardanus* and its relatives, and calibrated their molecular clock based on the age of the formation of Grand Comore Island 0.5 million years ago. They found the split between *P. dardanus* and *P. constantinus* to be around 7 million years ago; much more recent than estimates by Zakharov *et al*. (2004). Clark and Vogler (2009) suggest that the internal splits between subspecies in *P. dardanus* occurred about 0.72 million years ago. This age estimate for these two species is perhaps more plausible as it corresponds to diversification of other taxa in Southern Africa (McDonald and Daniels, 2012; Tolley *et al*., 2006) along with the forest expansions and contraction in the Plio-Pleistocene.

Thesis structure

This thesis consists of two data chapters. One deals with phylogeography and haplotype diversity of two forest-restricted, one forest-associated, and two generalist *Papilio* butterflies, and attempts to use these species as proxies for assessing forest history. The other chapter deals with intraspecific taxonomy of a forest-restricted butterfly (*Papilio ophidicephalus*) which has previously been split into ten subspecies, five of which occur in South Africa and another one in Zimbabwe. More specific hypotheses are outlined within each of the chapters.

2 Afromontane forest history as indicated by the phylogeography of *Papilio* species in South Africa

Introduction

Past climatic changes have altered the sizes and distribution of forests in southern Africa (Lawes, 1990). South African forests have been broadly classified into Afromontane forests and Indian Ocean Costal Belt forests. Afromontane forest is considered the more ancient of the two and therefore the more persistent (Eeley *et al*., 1999). Indian Ocean Coastal Belt forests in South Africa were established less than 8000 years ago after the Last Glacial Maximum (Partridge *et al*., 1999). Lawes *et al.* (2000b) suggest that palaeoclimatic events had a greater impact on Afromontane forests because they are much older than the Last Glacial Maximum and have experienced more than one extinction filtering event, unlike the younger Indian Ocean Costal Belt forests. Afromontane forests are generally situated at higher altitude and therefore exposed to markedly seasonal climates (Lawes *et al.*, 2000a), whereas Indian Ocean Coastal Belt forests are close to the sea and have a more stable climate. Afromontane forests are generally smaller and more isolated than coastal forests, which occur in almost a continuous belt down the eastern coast of South Africa (Fig 2.1; Lawes *et al.*, 2007). Between the Indian Ocean Coastal Belt and Afromontane forests is a narrow band of Scarp forests. These forests consist of a mix of elements of both Afromontane and Coastal forests and some relictual tropical elements (Lawes, 1990; Lawes *et al*., 2000a). The current patterns of plant community composition of the two forest types in South Africa differ, suggesting that they may have been established before major forest fragmentation in the late Miocene, this period has been associated with increasing aridity. In this period, forest biota survived in areas which are now known as dispersal corridors (Mucina and Rutherford, 2006).

Of all of the forest types, Scarp forests contain the highest plant diversity. Along with the costal and Afrotemperate forests, Scarp forests are the most species-rich forests in the country (Eeley *et al.*, 1999). Mistbelt forests generally are not as species-rich as Scarp forests, but they share some floristic elements with Afrotemperate forests. Northern Mistbelt forests

are thought to have been more affected by the Last Glacial Maximum 8-6 kya (Deacon and Lancaster, 1988; Scott *et al*., 1997). Afrotemperate forests are thought to be generally less species-rich than Scarp and Mistbelt forests. Within the Coastal forests, the Northern Coastal forests contain more tropical species and are more species-rich than the Southern Coastal forests. This suggests that the Southern Coastal forests have established fairly recently (Mucina and Rutherford, 2006).

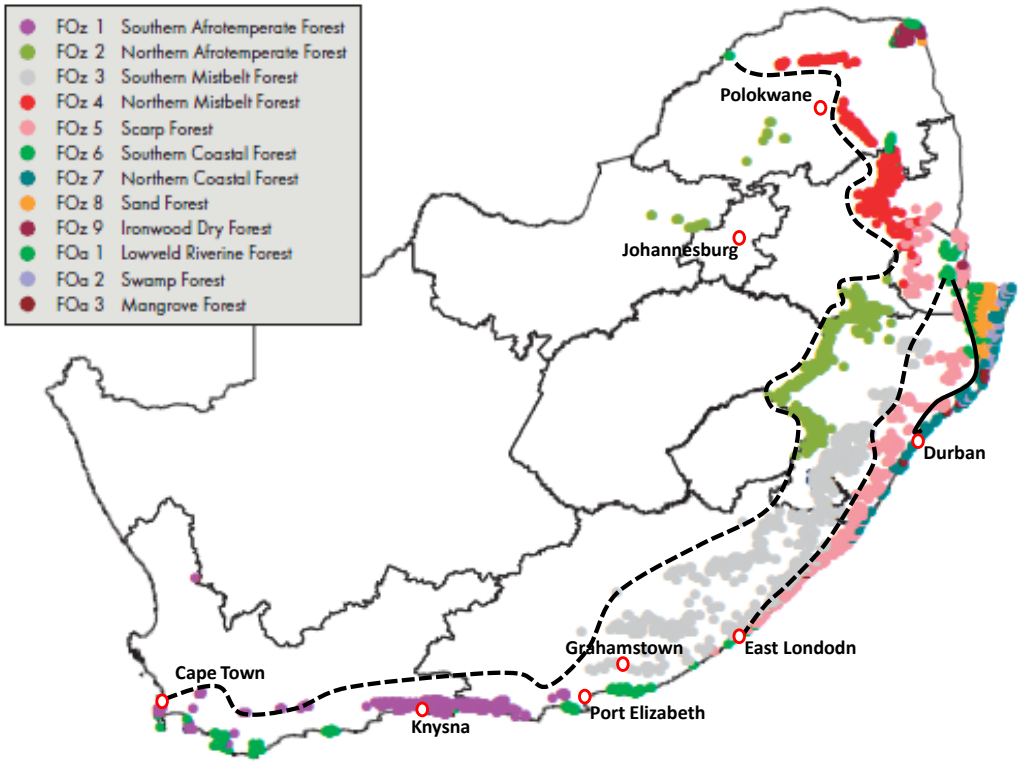

Figure 2.1: The distribution of indigenous forest in South Africa: Afrotemperate forest occurs between the two dashed lines; Scarp forest between the solid and dashed lines; and Indian Ocean Coastal Belt forest to the east of the solid line (modified from Lawes *et al.*, 2007). Major cities are included on the map, along with the distribution of the different forest types classified by Mucina and Rutherford (2006).

Mucina and Rutherford (2006) subdivide the forests in South Africa into 26 different types that are categorised into eight zonal groups within the two basic categories (IOCB and Afrotemperate). These different types occur only on the eastern and southern margins of the country in small and scattered fragments (Fig. 2.1). Most of the fragments are smaller than 100 ha, with only a few recognised larger forest complexes, the largest (60 560 ha) being around Knysna. The Amathole Forest complex (40 500 ha) is the second biggest patch in the country. The other larger patches are much smaller than these, all being below 8000 hectares.

Aims and Hypotheses

The aim of this study is to assess genetic diversity and phylogeographic patterns in two Afromontane, forest-dwelling butterflies, and to compare these to a forest-associated species and two species that are not associated with forests. Three hypotheses are tested using this approach:

Hypothesis 1: Forest-restricted (*P. ophidicephalus* and *P. echerioides*) and forest-associated (*P. dardanus*) species are expected to have more genetic structure than widespread species (*P. demodocus* and *P. nireus*) due to isolation in forest patches and lack of gene flow.

Hypothesis 2: It is expected that isolation by distance (IBD) is stronger in forest-restricted species than in non-restricted species because there is more restricted gene flow between different forest patches.

Hypothesis 3: Habitat fragmentation decreases genetic diversity in forests-restricted species due to restriction of suitable habitat along with a decrease in population size, genetic drift and inbreeding depression.

Species selected to test these hypotheses

To test the hypotheses above, five species of the genus *Papilio* from South Africa with different degrees of specificity to forests were selected for this study; two forest-specific, one forest-associated and two generalist species.

***Papilio ophidicephalus* Oberthür** (the Emperor Swallowtail) is a forest-restricted species that belongs to the *Papilio menestheus* group. It is the largest butterfly in South Africa with a wingspan of up to 12 cm; females are generally larger than males. There was some confusion among earlier writers about the taxonomy of this species, but Van Son (1939) showed that the African material falls into three distinct species and the forms found in South Africa are subspecies of the East African *P. ophidicephalus*. From that he described five new subspecies. The subspecies *P. o. entabeni* is found in the northern parts of the country (Limpopo province) along the Soutpansberg; *P. o. tranvaalensis* generally occurs a bit further south but north of the Olifants River along the Drakensberg and Wolkberg in Mpumalanga and Limpopo; *P. o. ayresi* can be found in the forests of KwaZulu-Natal northwards into south western Mpumalanga; *P. o. zuluensis* is mostly found around the Eshowe, Dlinza and Ngoye forests; and *P. o. phalusco* is the southernmost subspecies

occurring from the Tugela River southwards to the forests along the Amathole mountains (Dickson and Kroon, 1978; Mecenero *et al*., 2013). This species is bivoltine, where the first brood is from September to November and the second brood starts in December and goes to March, but they are most abundant in mid-summer (Mecenero *et al.*, 2013). The larvae feed on Rutaceae *Clausena antisata* (horsewood), *Zanthoxylum capense* (small knobwood), *Calodendrum capense* (Cape chestnut) and cultivated *Citrus* species (Kroon, 1999; Mecenero *et al.*, 2013; Woodhall, 2005).

Papilio echerioides **Trimen** (the White Banded Swallowtail) is a low-flying, forest-restricted species that occurs in more inland forests. It is distributed from the Amathole mountains through KwaZulu-Natal to the Soutpansberg, and is a very weak flyer that is often observed resting on plants. Individuals have been seen on forest edges briefly, after which they return into the forest. This swallowtail is bivoltine; the first brood flies from January to March and the second from September to December (Dickson and Kroon, 1978; Mecenero *et al*., 2013; Woodhall, 2005). Larval host plants include *Clausena anista*, *Vepris lanceolata* and *Zanthoxylum capense* (Dickson and Kroon, 1978; Henning *et al*., 1997; Kroon, 1999; Woodhall, 2005).

Papilio dardanus **Stoll (**the Mocker Swallowtail) is a forest-associated butterfly and is not found in the drier and more open parts of South Africa. It occurs from Knysna eastwards along the coastal forests to Port St. Johns and then northwards to Swaziland and Riverine and Montane forests in Mpumalanga and Limpopo. It has a much wider distribution than just South Africa. Different subspecies occur in the wetter parts of East and West Africa and on Madagascar. *Papilio d. dardanus* is the most widespread subspecies, occurring in most of western tropical Africa. However there are some subspecies that have a very narrow distribution range; some are even restricted to only one mountain top in northern Kenya (*P. d. ochracea* and *P. d. flavicornis*). The subspecies occurring in South Africa is *P. d. cenae*, which was sampled for this study (Clark and Vogler, 2009). Males generally fly along the edges of forests whereas females normally stay closer to the centre of the forest where the host plants for the larva are more abundant. This species flies year-round but is not as active in the cooler months as in summer (Dickson and Kroon, 1978; Henning *et al*., 1997; Mecenero *et al*., 2013; Woodhall, 2005). The main larval host plants belong to the Rutaceae (Kroon, 1999).

Papilio demodocus **Esper** (the Citrus Swallowtail) is a non-forest-associated species that is common and widespread throughout South Africa. It is more common in woodlands and gardens and absent from extremely arid areas (Mecenero *et al.*, 2013; Woodhall, 2005).

Their flight period is year-round in the subtropical areas, but restricted to summer months (September to April) in cooler areas such as the Cape Peninsula. When conditions are favourable, some individuals will emerge (Mecenero *et al.*, 2013). Main larval host plants include species of Rutaceae including cultivated *Citrus* species (Kroon, 1999).

***Papilio nireus* Doubleday** (the Blue-banded Swallowtail) is a non-forest-associated species that is distributed more commonly in the northern provinces of South Africa through to Mozambique and Zimbabwe. Only on rare occasions they have been observed in the Western Cape. They fly year-round in warmer areas and only in summer in cooler areas (Mecenero *et al.*, 2013). The main larval host plants belong to the Rutaceae and include cultivated *Citrus* species (Dickson and Kroon, 1978).

Materials and Methods

Sampling

Butterflies were sampled in a variety of different forest patches and types throughout South Africa on a number of field trips. The main field trip (January and February 2013) was five weeks long and covered the areas from Louis Trichardt in Limpopo province and ended in Grahamstown in Eastern Cape. Several shorter fieldtrips were done to forests closer to Grahamstown, such as Hogsback, Stutterheim, Pirrie forest and the forests around Mthata. Where possible, ten specimens from each region were collected and preserved. However, that was not always possible as we were time restricted and the weather in some places was not conducive for collection (rainy and cold). In KwaZulu-Natal, collection permits restricted us to five samples per forests. Another problem was that access to some forests was limited; open roads through forests are ideal for catching these fast flying butterflies, as they use them as patrol routes and one can chase them without difficulty. If there are no open roads, it is difficult to catch *Papilio*, since they cannot be trapped. Collected butterflies were pinned and three legs from each sample stored in 96% ethanol for subsequent DNA analysis. Voucher specimens are housed in the Albany Museum, Grahamstown.

Molecular techniques

DNA was extracted from the tissue of one of the ethanol preserved legs using the Invisorb ® Spin Tissue Mini Kit (Invitek ©). Part of the COI (cytochrome oxidase I) gene was amplified using a variety of primers (Table 2.1). The primer pair K807 and 1718 were first used; however, better primers pairs (Jerry and Pat2 and Lyn and K525) were found later

(Simon *et al*., 1994; Sperling *et al*., 1996; Wahlberg and Wheat, 2008) which made that pair (K807 and 1718) redundant, as it overlaps with the other two primer pairs. All PCR reactions were done in a 50 µl reaction volume. Reaction volumes for the different reagents in the PCR reaction are indicated in table 2.2. The thermal cycling profile for the COI primer pairs were 95°C for 5 minutes, 30 cycles of 95°C for 45 seconds, 45°C for 45 seconds and 72°C for 1.5 minutes, followed by a final extension period of 72°C for 5 minutes.

Table 2.1: Primer pairs used for amplification of the COI gene.

Primer Name	Direction	Sequence	Reference
Jerry	Forward	5'CAACATTTATTTTGATTTTTTGG3'	(Simon *et al.*, 1994)
Pat2	Reverse	5'TCCATTACATATAATCTGCCATATTAG3'	(Sperling *et al.*, 1996)
K807	Reverse	5'TGAAAATGAGCTACAACATAATA3'	(Caterino and Sperling, 1999)
1718	Forward	5'GGAGGATTTGGAAATTGATTAGTTCC3'	(Simon *et al.*, 1994)
Lyn	Reverse	5'ACAAATCATAAGGATATTGGAAC3'	(Wahlberg and Wheat, 2008)
K525	Forward	5'ACTGTAAATATATGATGAGCTCA3'	(Simon *et al.*, 1994)

PCR amplifications were checked for the presence of amplified PCR products by gel electrophoresis (0.5% agarose gel stained with SYBR green) and viewed with a UV-transluminator. Then successful PCR products were cleaned up using the Invisorb PCRapace ® Quick purification kit (Invitek©). Cleaned PCR products were then cycle sequenced using the ABI Big Dye Sequencing kit v.3.1, according to manufacturer's instructions. PCR products were sequenced in both directions using the same primers as for the PCR.

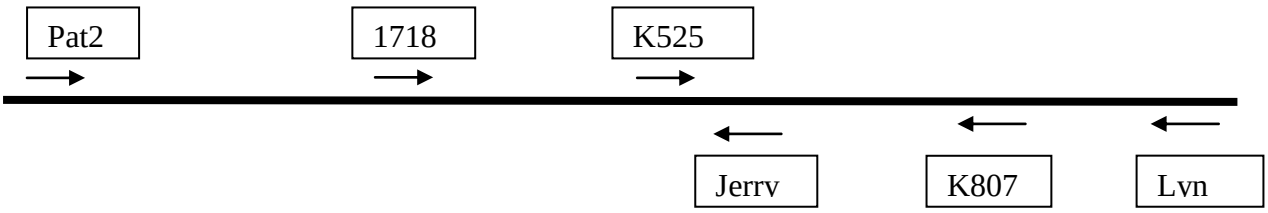

Figure 2.2: Positions of the different primer pairs along the amplified COI gene.

Phylogenetic analyses

The sequence trace files were viewed and checked using GeneStudio v.2.1.1.5 (GeneStudio. Inc). Sequence files were then aligned using the ClustalW algorithm, using default settings, which is included in the software package MEGA v.5 (Tamura *et al.*, 2011) (alignment on CD as Appendix 2). The positions of the different primer pairs in the COI gene region are indicated in Figure 2.2. The messy ends of each sequence have been trimmed and short sequences were filled with question marks to the end to indicate missing data. Each mutation was cross-checked in the original trace files.

Other African samples of *Papilio* from Genbank (Benson *et al*., 2008) were included in the analysis, this includes sequences of *P. demodocus* (Zakharov *et al*., 2004a, 2004b) and *P. dardanus* (Caterino and Sperling 1999). Sister species to the species selected for this study were also obtained from Genbank (*P. lormieri* gb AY569088.1, *P. constaninus* gb AF044002.1, *P. constantinus* gb EU153627.1, *P. demoleus* gb AF044000.1 Malaysia, *P. demoleus* gb AY569092.1 Australia, *P. oribazus* gb AY457591.1 Madagascar and *P. epiphorbas* gb AY457589.1 Madagascar). (Caterino and Sperling, 1999; Zakharov *et al*., 2004a, 2004b). These sequences were added to ascertain species-level differences as indicated by branch lengths.

A Bayesian Inference analysis was conducted using MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001) using the GTR model. Two runs of 20 million generations, sampling every 200 generations with four chains (3 heated and 1 cold), were plotted against the likelihood scores and tree length, to ascertain when the analysis reached stationarity, and the first 10% of the trees were discarded as burnt-in, which represents 20 000 of the 200 000 trees sampled.

Maximum likelihood analysis was conducted using the GARLI (Genetic Algorithm for Rapid Likelihood Inference) web service (Bazinet *et al*., 2014) running the GTR model. This model was selected as it is the most sophisticated model the program can run. The model selected by jModleTest (TIM2+I+G) was not available in GARLI, for that reason the next most sophisticated model was selected. The analysis was run with 2000 bootstrap replicates.

Molecular Dating

The molecular dating analysis was run in BEAST (Bouckaert *et al*., 2014). All identical sequences in the dataset were merged, so that the analysis was run on unique sequences only. This means that some of the final terminal names are concatenations of all

the individual sequences that are represented by a particular terminal. For the ucld mean a diffuse gamma distribution with and initial value of 1.0, shape value of 0.001 and a scale of 1000 were used. The calibration was set at 0.5 myr as this is the split in *P. dardanus* for the Gran Comores (Clark and Vogler, 2009). The tmrca prior was given a longitudinal distribution mean of 0.5 and log(SD) was set at 0.2, which gave a 5% to 95% age range to 0.353 to 0.681 for the calibration prior.

The analysis was run with 50 million generations, sampling every 5000 generations, to give a total of 10 000 logged samples. To produce the final tree a burn-in of 10% was used.

Network-building Methods

Tree-building methods summarise phylogenetic ambiguity on dichotomising consensus trees using an optimality criterion that effectively hides the alternative solutions. An alternative approach is to show all of the ambiguity in a network diagram. Haplotyope networks were created using the software TCS ver. 1.21 (Clement *et al*., 2000). Non-interleaved nexus files were created in PAUP v.4.0b10 (Swofford, 2002) and imported into TCS, in which separate haplotype networks were created for each of the species.

Isolation By Distance

Isolation-by-Distance (IBD) analysis was conducted by using the Isolation By Distance Web Service version 3.23 (Jensen *et al*., 2005). This entailed constructing a data matrix of the genetic distances and the geographical distances. The genetic distances were calculated in MEGA 5 (Tamura *et al*, 2011) by comparing the number of pair-wise differences according to the Maximum Composite Likelihood model. The geographical distances were measured in Google Earth by measuring the direct distance between the different sample sites. The genetic and geographical data were log-transformed before analysis, which was run with 10 000 randomizations as part of the Mantel test.

Genetic Diversity

Numbers of haplotypes (H), haplotype diversity (H_d) and within-population nucleotide diversity (π), and numbers of polymorphic sites (S) were calculated for each species using DNASP ver. 5.10.01 (Rozas *et al.*, 2003). Only samples collected in South Africa were included in this analysis. Other studies of *Papilio* butterflies have used similar indicators to determine genetic diversity (Sperling and Harrison, 1994).

Results

Data characteristics

The final molecular dataset for the COI gene region consisted of 1130 characters, of which 798 were invariant, 46 variable but parsimony-informative and 286 parsimony informative, from 214 specimens including samples obtained from GenBank (Caterino and Sperling, 1999; Clark and Vogler, 2009; Zakharov *et al*., 2004b).

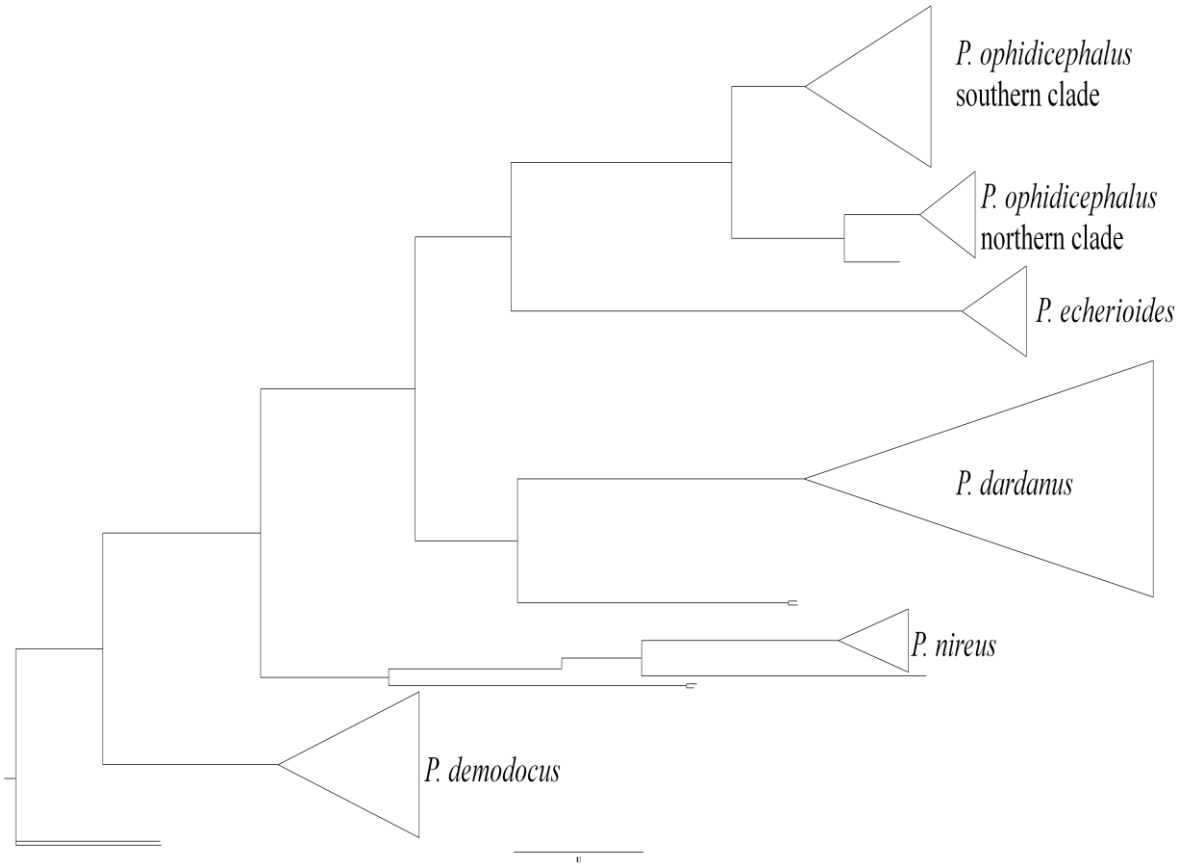

Figure 2.3: Summarised Bayesian Inference tree of all collected samples including sequences from GenBank (Benson *et al*., 2008), (Caterino and Sperling, 1999; Clark and Vogler, 2009; Zakharov *et al*., 2004b) .

Genetic structure of forest-restricted vs widespread species

Papilio ophidicephalus (forest-restricted)

The analysis of COI data shows that *P. ophidicephalus* is split into two clades, where one is well supported but the other group has no support. One of the groups (Fig. 2.4: black bar) consists of samples collected in Northern Mistbelt Forests. This clade covers three

different subspecies from the distribution (Mecenero *et al*., 2013), but there is no genetic evidence for clades that might represent each of the subspecies.

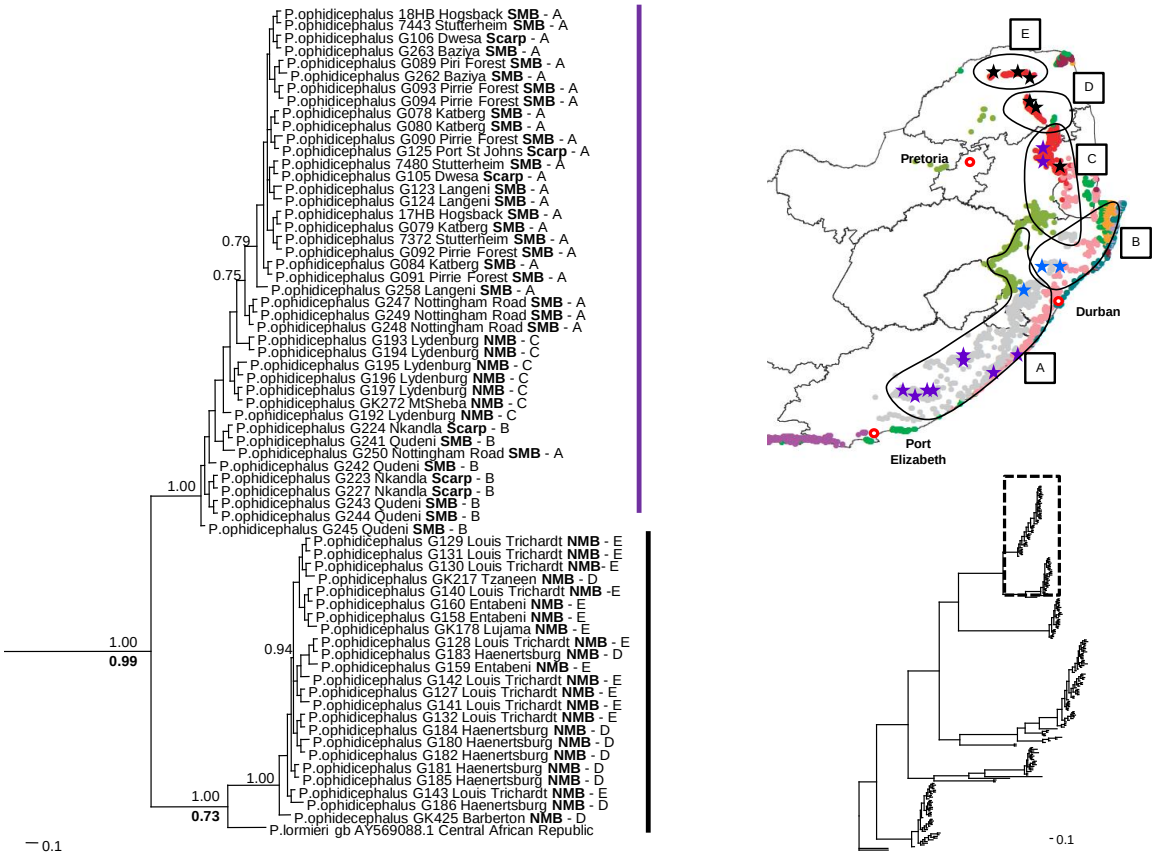

Figure 2.4: Bayesian Inference phylogram of *P. ophidicephalus* obtained from the COI data set. Posterior probabilities and Maximum likelihood support (below branch) > 0.7 are indicated at each node. Sample localities are mapped on the corresponding forest types. Colours of the stars correspond to the bars. The distribution ranges of the subspecies as derived from (Mecenero *et al*., 2013) are indicated. A: *P. o. phalusco*; B: *P. o. zuluensis*; C: *P. o. ayresi*; D: *P. o. transvaalensis*; E: *P. o. entabeni*.

Papilio echerioides (forest-restricted)

P. echerioides is split into three subgroups that are not well supported (Fig. 2.5). There is a basal paraphyletic group of two clades comprised of samples collected in SMB and Scarp forests. The NMB samples are sister to this, but there is a lack of support for this. This is similar to *P. ophidicephalus* where the group is split into northern and southern clade.

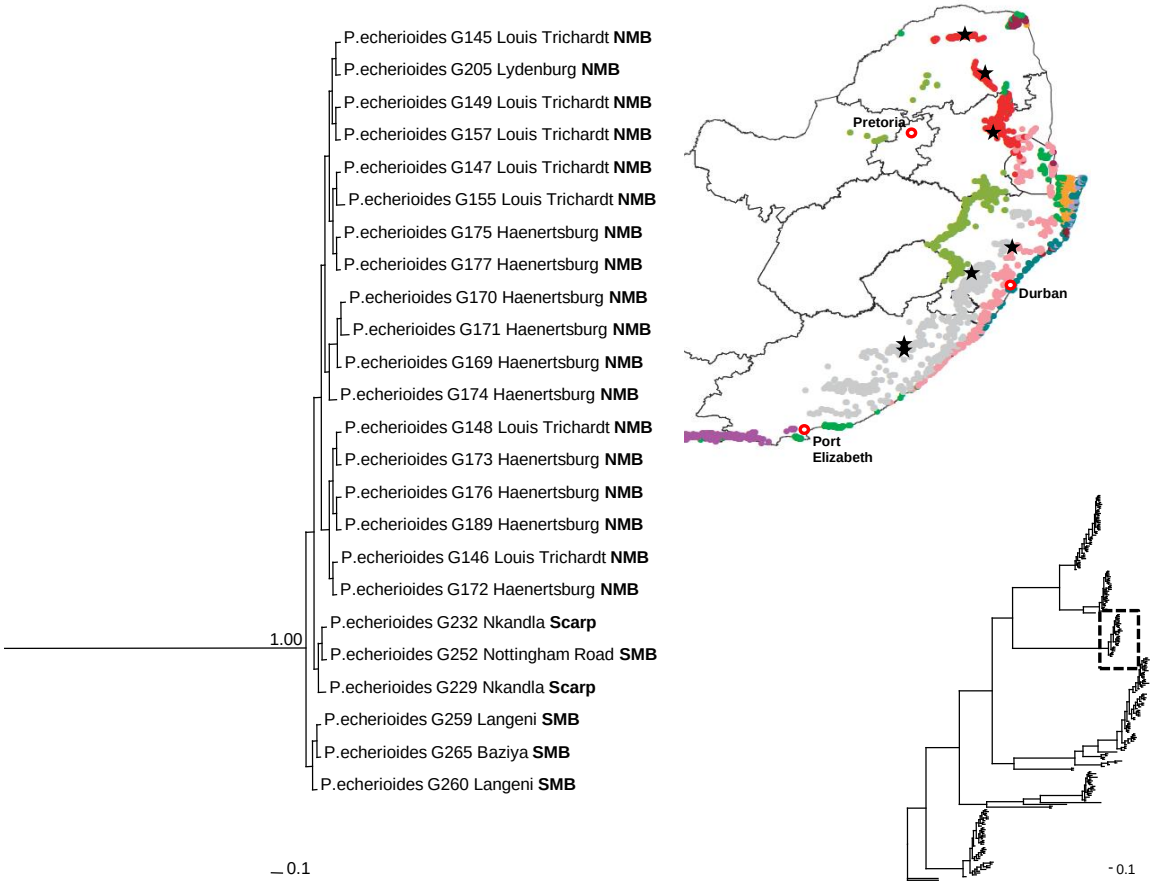

Figure 2.5: Bayesian Inference phylogram of *P. echerioides* obtained from the COI data set. Posterior probabilities and Maximum likelihood support (below branch) > 0.7 are indicated at each node. Sample localities are mapped on the corresponding forest types.

Papilio dardanus (forest-associated)

Papilio dardanus is split into four groups that generally lack support. There is no obvious structure within these subgroups because samples collected from the same area and forest type did not group together. However, it is evident that samples from East Africa and Madagascar form a basal grade. A single Ugandan sample is sister to the main clade consisting of South African and East African samples (PP 1.00; Fig. 2.6).

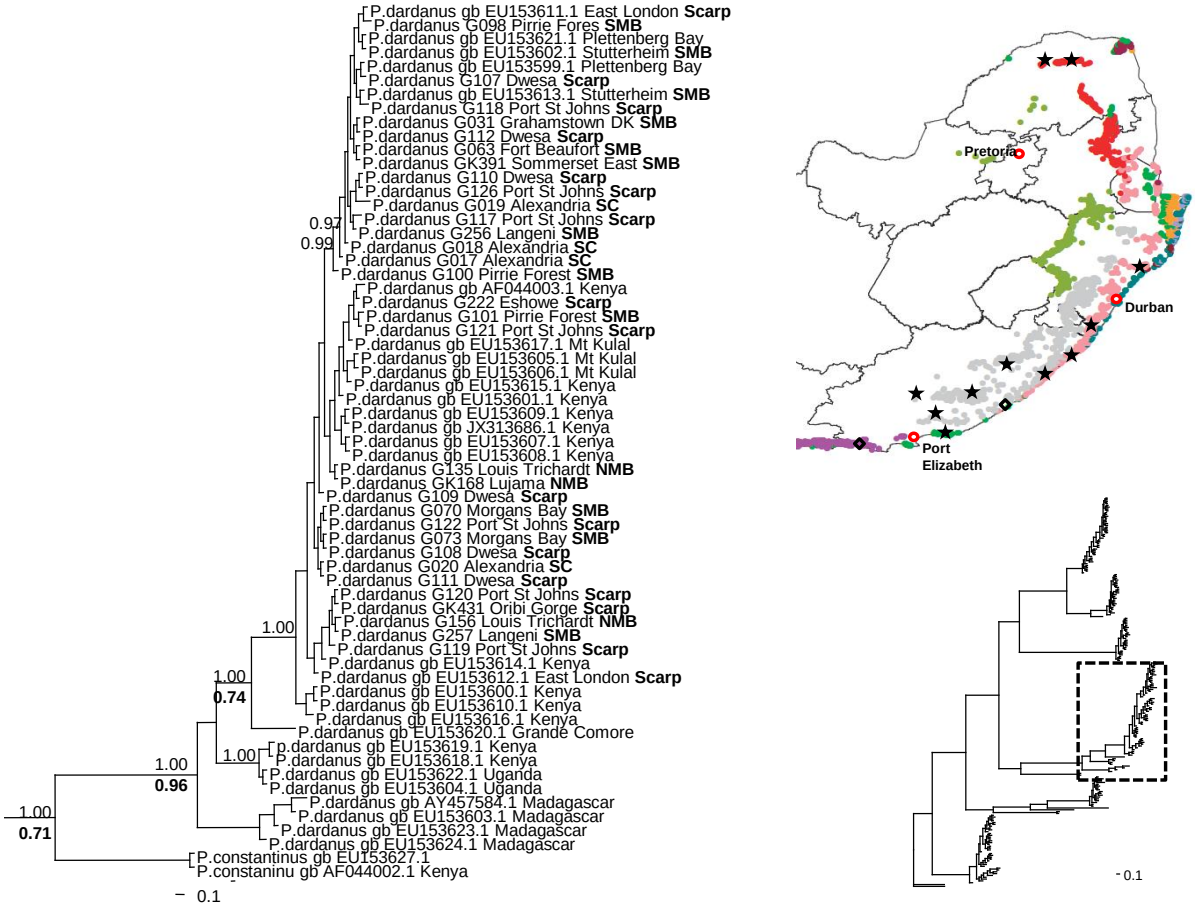

Figure 2.6: Bayesian Inference phylogram of *P. dardanus* obtained from the COI data set. Posterior probabilities and Maximum likelihood support (below branch) > 0.7 are indicated at each node. Sample localities are mapped on the corresponding forest types.

Genetic structure of P. demodocus and P. nireus (non-forest species)

In both *P. demodocus* and *P. nireus*, there is no clear genetic structure (Fig 2.7). *Papilio demodocus* is split into four groups where the Central African Republic sample is the first to differentiate, then samples from South Africa and Africa follow with support of 0.95. With one exception, samples from Madagascar form a separate clade which is placed within a group of south African samples (PP = 0.93) *P. nireus* shows no genetic structure with (Fig 2.7) and samples from the same forest types do not form separate clades.

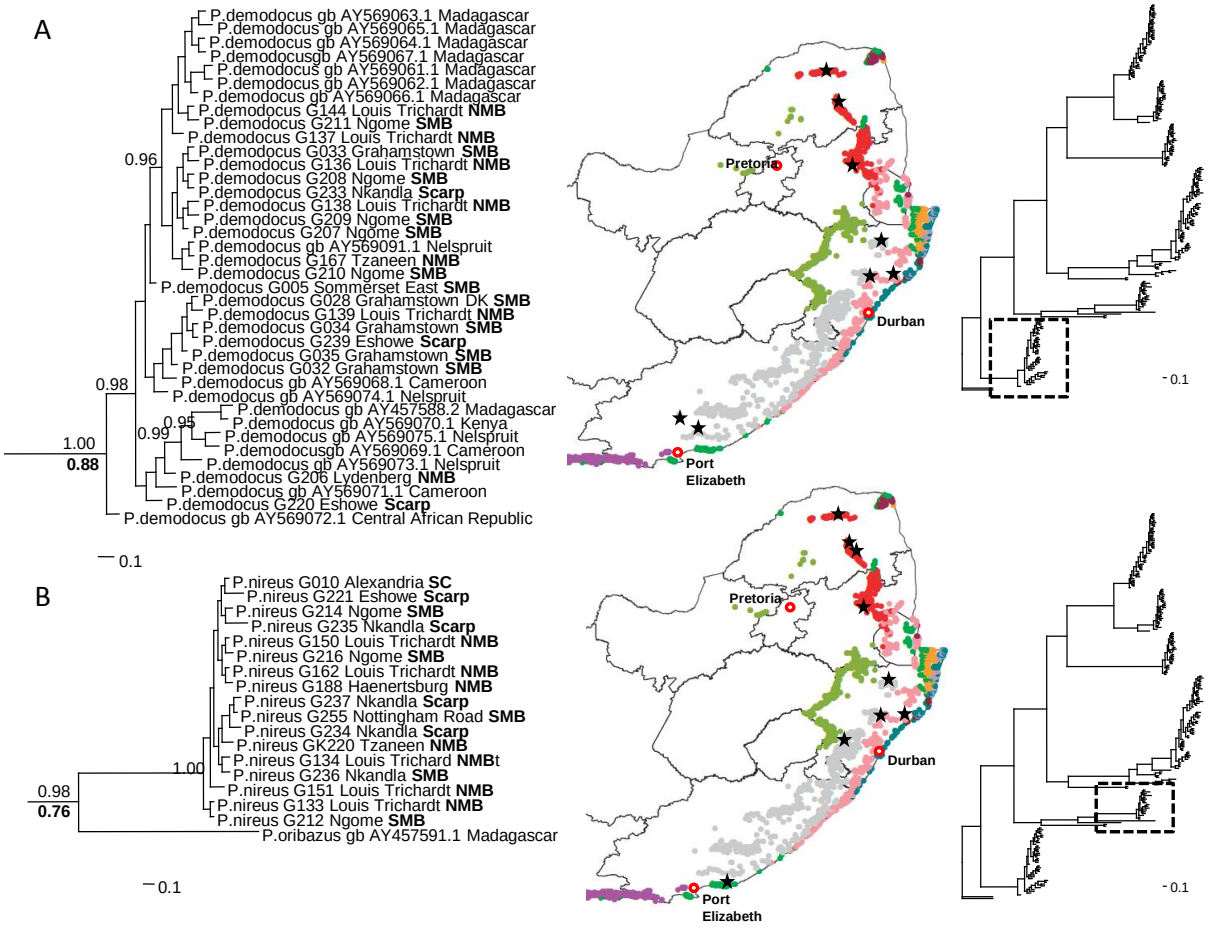

Figure 2.7: Bayesian Inference phylogram of *P. demodocus* and *p. nireus* obtained from the COI data set. Posterior probabilities and Maximum likelihood support (below branch) > 0.7 are indicated at each node. Sample localities are mapped on the corresponding forest types.

Haplotype networks

There are 25 haplotypes of *P. ophidicephalus*, and there are two haplotype networks (Fig. 2.8). The two networks compare to the northern and southern clade as retrieved by the Bayesian analysis. Samples from Lydenberg and parts of KZN form a group; this corresponds with the phylogenetic trees where those samples are embedded in the southern group even though they were collected in Northern Mistbelt forests.

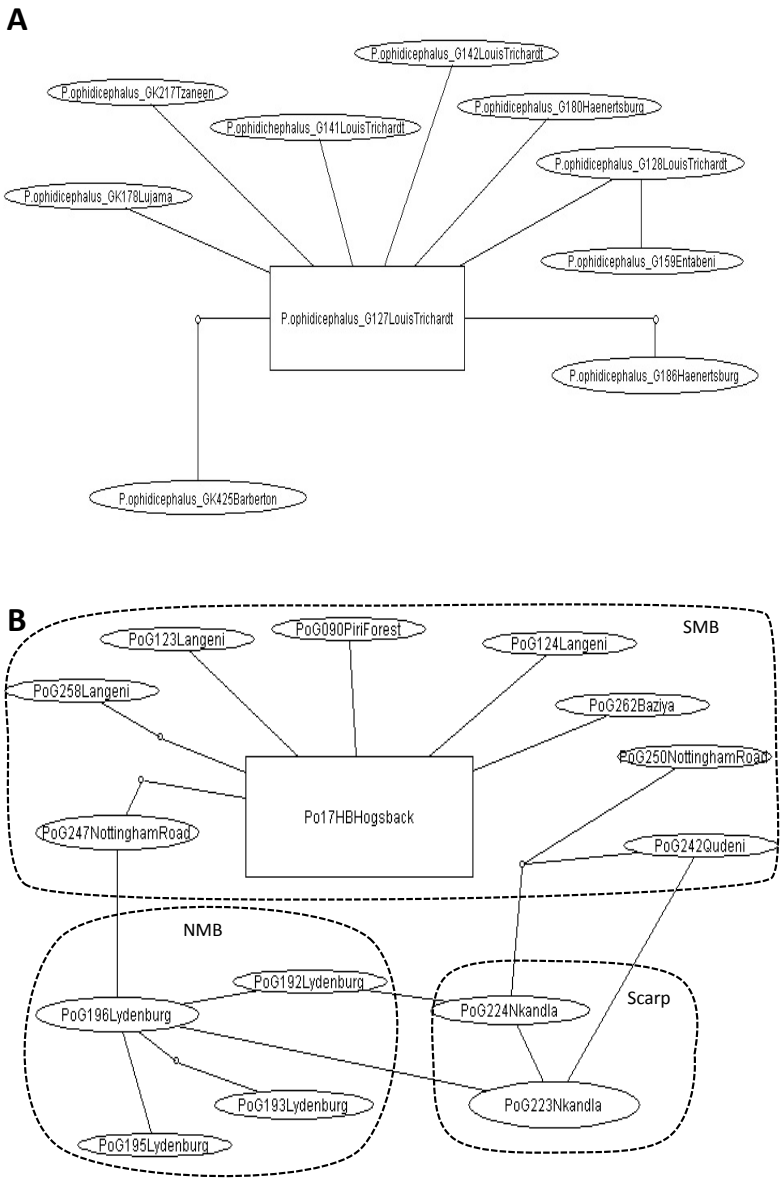

Figure 2.8: Haplotype network created in TCS for *P. ophidicephalus*. A = northern clade; B = southern clade.

P. echerioides has 9 haplotypes. As in *P. ophidicephalus*, there is some evidence in *P. echerioides* for a split into northern (NMB) and southern (SMB/Scarp) groups (Fig. 2.9), reflecting the results obtained in the Bayesian analysis.

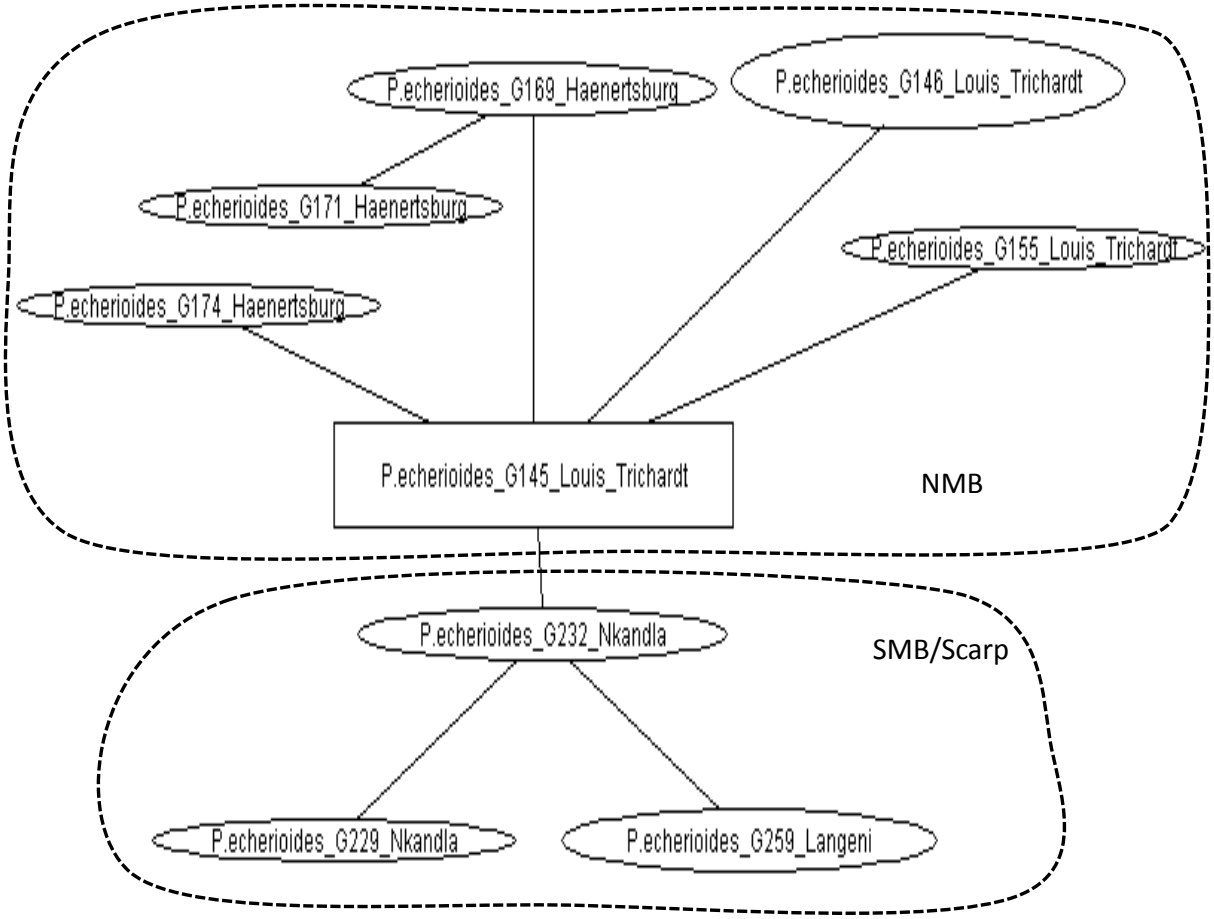

Figure 2.9*:* *P. echerioides* haplotype network along with indication of forest types.

P. dardanus has 31 haplotypes with six disconnected haplotypes and networks, of which most are from East Africa and Madagascar (Fig. 2.10). One South African sample from Plettenberg Bay is connected to Kenyan samples in one of these. The main network consists of most of the South African samples as well as one sample from Kenya.

Figure 2.10: TCS haplotype network of *P. dardanus* (including samples from GenBank (Benson *et al.*, 2008)).

Papilio demodocus has 27 haplotypes. There are two networks which are not connected (Fig. 2.11). The analysis infers that Madagascan Haplotype is ancestral to all other samples from East Africa and samples from Nelspruit in South Africa. The second network includes South African samples with one Madagascan sample.

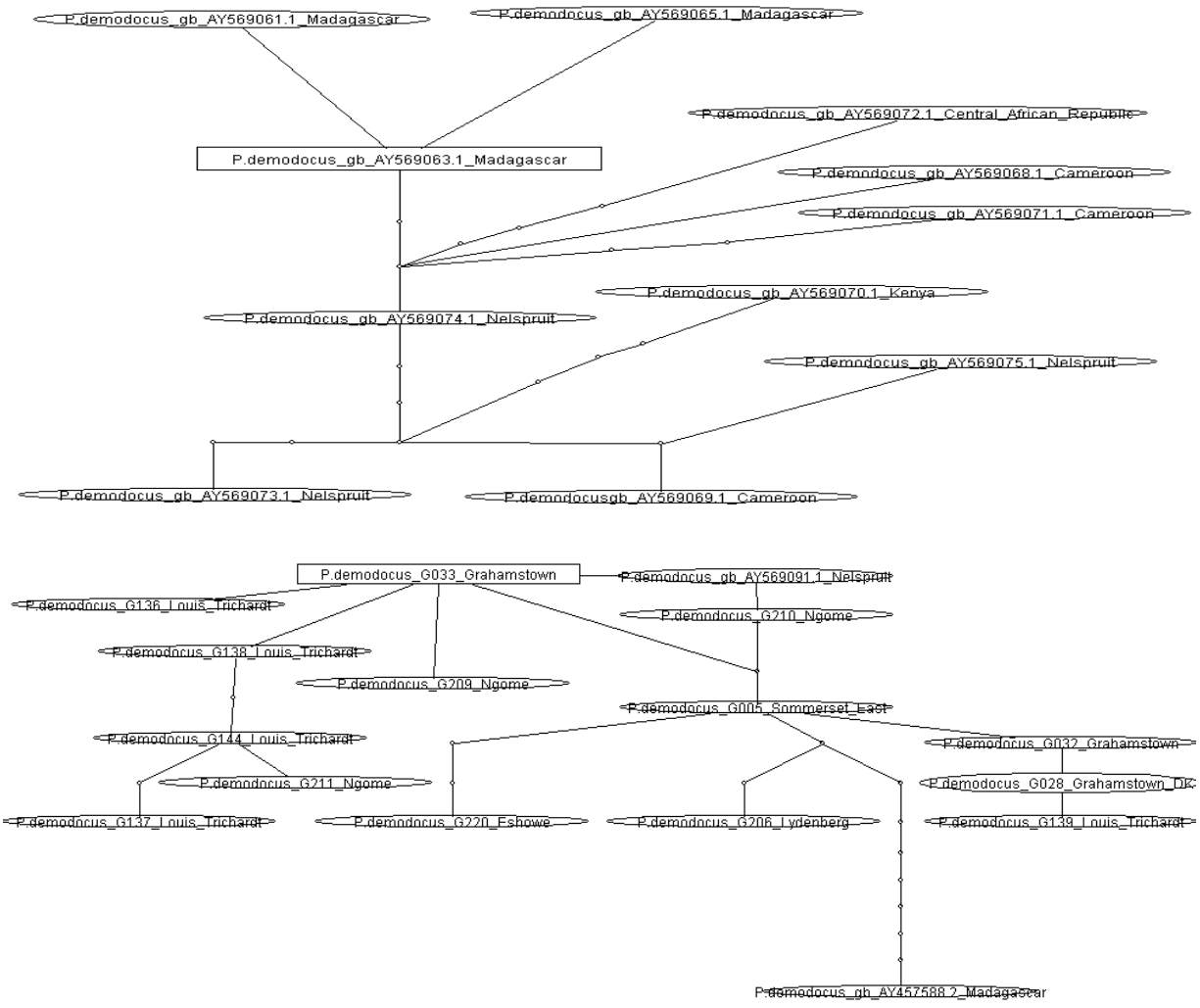

Figure 2.11: TCS haplotype network of *P. demodocus* (including samples from GenBank (Benson *et al*., 2008)).

The *P. nireus* samples from 14 haplotypes with the ancestral haplotype from NMB forests. Similar to *P. demodocus*, *P. nireus* shows no obvious genetic structure. Samples from Louis Trichardt are connected to samples from the Eastern Cape and again there is no consistent connectivity within forest types (Fig. 2.12).

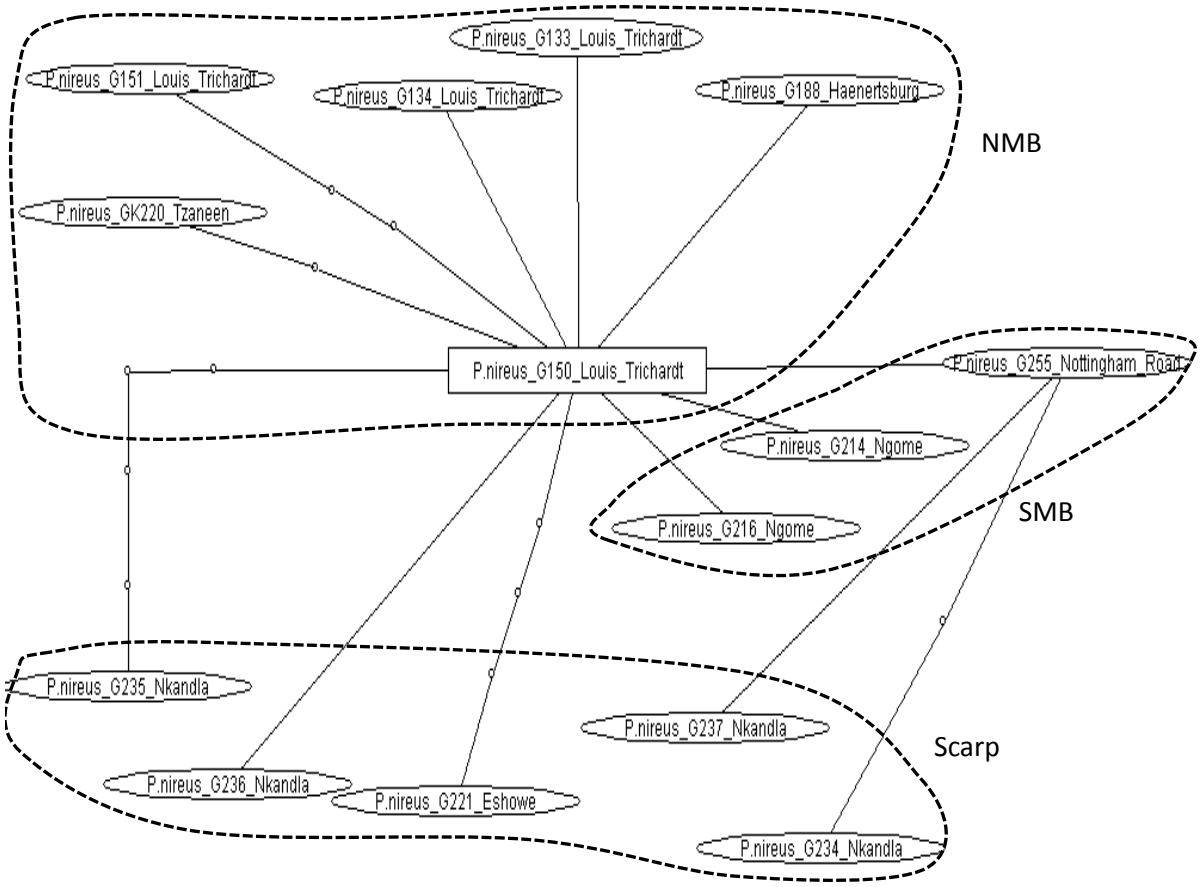

Figure 2.12: The haplotype network created for *P. nireus* in TCS. Different forest types are indicated on the figure.

Molecular dating

The crown age of the five study species is 1.4267 [0.6351; 2.3773] MYR. The split between the northern and southern clade of *P. ophidicephalus* occurred 0.5637 [0.2091; 0.9806] million years ago. The north/south split in *P. echerioides* is more recent, 0.2233 [0.0548; 0.4493] MYR. *P. ophidicephalus*, *P. echerioides* and *P. dardanus* have a crown age of 0.9591 [0.5977; 2.0048] million years and *P. nireus* and *P. demodocus* have a crown age of 1.0496 [0.3944; 1.779] million years. From the figure (2.13) it is clear that the different species diverged at around 0.75 – 1 million years ago and then further diversification within groups occurred more recent such as the north/south split in *P. ophidicephalus*.

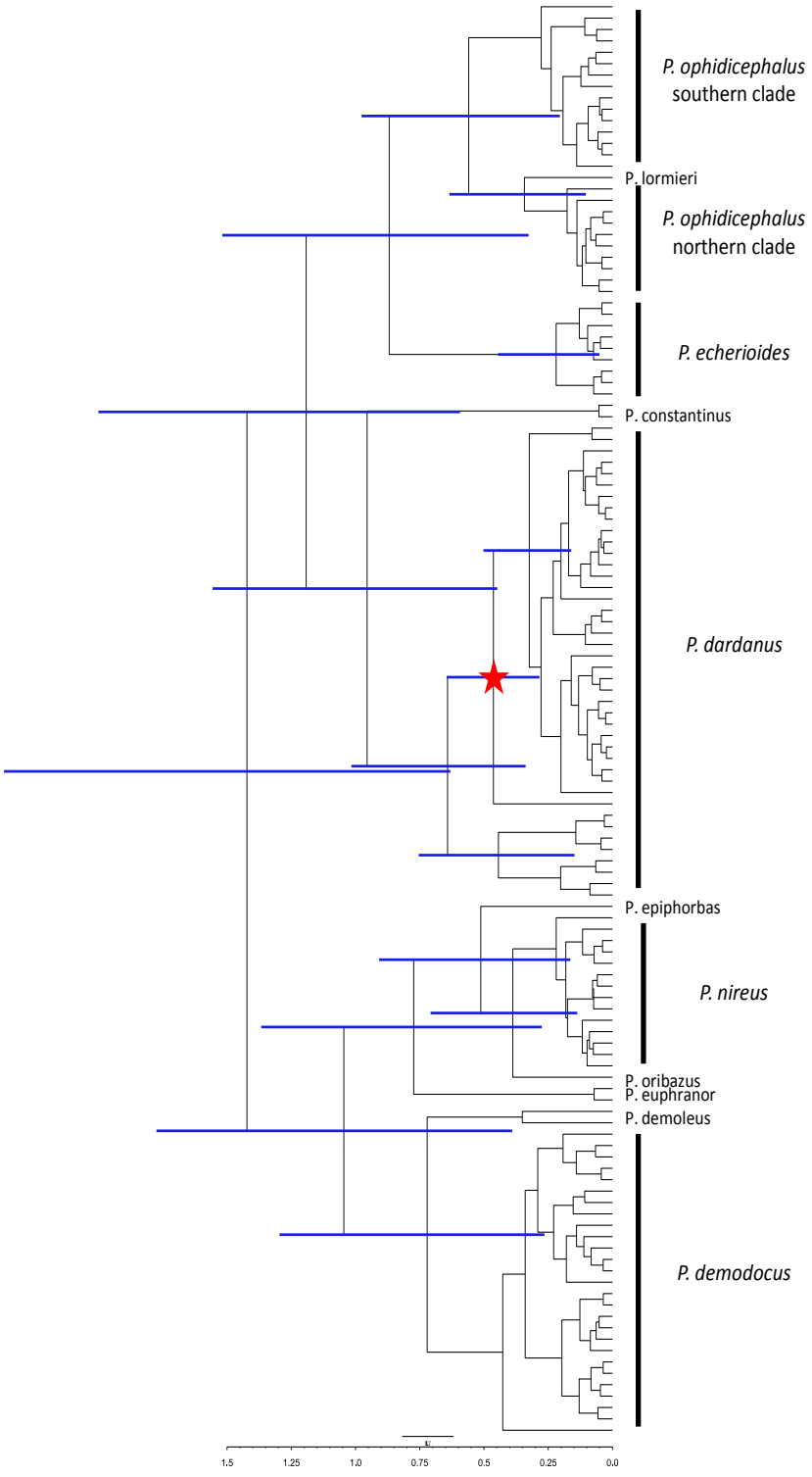

Figure 2.13: Showing a maximum clade credibility tree. The star indicates calibration point for the analysis. Node ages with confidence intervals are indicated on each node with blue bars. Axis below indicated ages in MYR.

Isolation By Distance

Forest-restricted species have a higher correlation between genetic distance and geographic distance than the non-forest-associated species. All *P. ophidicephalus* samples combined have a r value of 0.5553 (p = 1), but, when splitting *P. ophidicephalus* up into a northern and a southern clade, the correlation decreases, and the northern clade has a much lower correlation (r = 0.2085; p = 0.9604) than the southern clade (r = 0.4362; p =1). In *Papilio echerioides* the correlation between genetic and geographic distance is low (r = 0.2862; p = 0.9982), however it is still higher than the non-forest-associated species. The more widespread species *P. dardanus*, *P. nireus* and *P. demodocus*, had much lower values. (r = 0.0148; p = 0.6449, r = -0.0169; p = 0.4555 and r = 0.1455; p = 0.9660, respectively). The r-values are dependent on the sample size, so with higher sample sizes the values could change. However, more forest-specific species have higher correlations (Fig. 2.14).

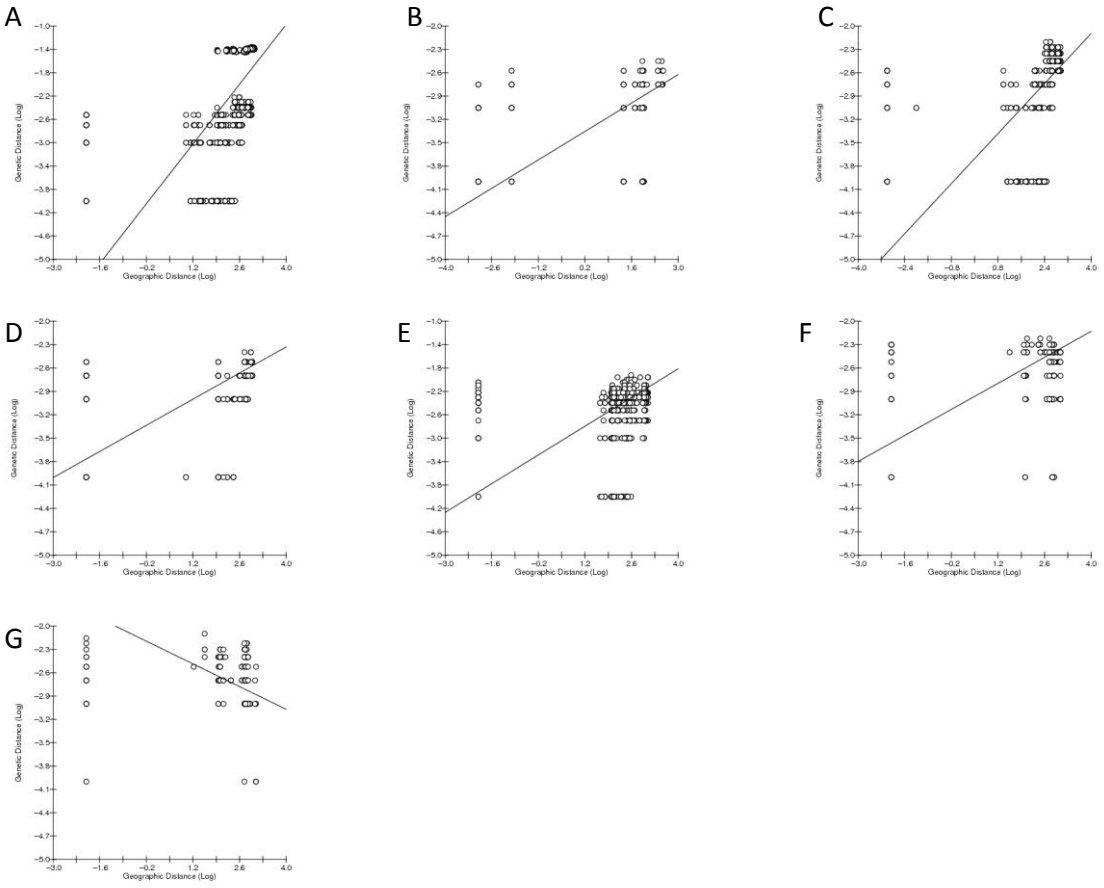

Figure 2.14: Showing the log-transformed graphs of the Isolation By Distance analysis of the five different *Papilio* species. A = *P. ophidicephalus* (all samples) B = *P. ophidicephalus* northern clade, C = *P. ophidicephalus* southern clade, D = *P. echerioides*, E = *P. dardanus*, F = *P. demodocus* and G = *P. nireus.*

Genetic diversity indices

Forest-restricted species (*P. ophidicephalus* and *P. echerioides*) have slightly lower haplotype diversity than the widespread species (Table 2.2). However, the sample size varies between the species collected, affecting estimates of haplotype diversity. *Papilio ophidicephalus* has the highest number of haplotypes (25) and *P. echerioides* (southern clade) the lowest (3). The nucleotide diversity for all of the species falls in the range between 0.0008 (*P. echerioides* southern clade) and 0.01 (*P. ophidicephalus*). The number of polymorphic sites is very variable between the different species. Again, *P. echerioides* (southern clade) has the lowest value (2) and *P. ophidicephalus* has the highest (57), but this is influenced by sample size. Upon splitting the two forest-restricted species into northern and southern clades, the sample sizes became more comparable, and the haplotype diversity decreases.

Table 2.2: Summary of genetic diversity for the five *Papilio* species sampled in South Africa.

Forest speceficity	Taxon	Sample size	Number of haplotypes	Haplotype diversity	Nucleotide diversity	Number of polymorphic sites
Forest-restricted	*P. ophidicephalus*	65	25	0.880	0.01063	57
Forest-restricted	*P. ophidicephalus northern clade*	23	10	0.668	0.00098	11
Forest-restricted	*P. ophidicephalus southern clade*	42	15	0.801	0.0023	12
Forest-restricted	*P. echerioides*	24	9	0.851	0.00139	8
Forest-restricted	*P. echerioides northern clade*	18	6	0.758	0.00097	5
Forest-restricted	*P. echerioides southern clade*	6	3	0.733	0.00083	2
Forest-associated	*P. dardanus*	32	18	0.929	0.00425	26
Wide spread	*P. demodocus*	21	15	0.943	0.00317	17
Wide spread	*P. nireus*	17	14	0.971	0.00275	22

When comparing genetic diversity of different forests, it is evident that the forest-restricted species (*P. ophidicephalus* and *P. echerioides*) always have a lower genetic diversity than the more generalist species (Table 2.3). This shows that the results from Table 2.2 are not skewed by data from only one or two forests. By comparing the genetic diversity of the different forests, the sample sizes become more similar and therefore the genetic diversity can be compared better. Nucleotide diversity in forest-restricted species is also always lower than that of the more generalist species.

Table 2.3: Summery of genetic diversity of the five different *Papilio* species from the different collection sites. In order to reach a sample size of at least 3 neighbouring forest patches were combined. Shaded rows indicate forest-restricted species.

Taxon	Forest	Sample size	#polymorphic sites	#Haplotypes	Haplotype diversity	Nucleotide diversity
dardanus	Alexandria	4	10	4	1	0.00457
dardanus	Dwesa	6	8	5	0.9333	0.00354
dardanus	Port St Johns	7	16	7	1	0.00559
dardanus	Pirrie/Ft Beaufort	4	10	4	1	0.00474
demodocus	Grahamstown	4	4	3	0.8333	0.00192
demodocus	Louis Trichardt	5	9	5	1	0.00373
demodocus	Ngome	5	7	4	0.9	0.00248
echerioides	Louis Trichardt	7	2	3	0.66667	0.00067
echerioides	Haenertsburg	10	4	5	0.82222	0.00124
echerioides	Nkandla	3	1	2	0.66667	0.00059
echerioides	Langeni/Baziya	3	0	1	0	0
nireus	Louis Trichardt	5	5	4	0.9	0.00177
nireus	Nkandla	4	10	4	1	0.00457
nireus	Ngome	3	3	3	1	0.00177
ophidicephalus	Louis Trichardt	10	3	4	0.3333	0.00053
ophidicephalus	Haenertsburg	7	4	4	0.71429	0.00101
ophidicephalus	Katberg	4	0	1	0	0
ophidicephalus	Pirrie	6	1	2	0.3333	0.00029
ophidicephalus	Lydenburg	6	4	4	0.86667	0.00154
ophidicephalus	Qudini	5	2	3	0.7	0.00071
ophidicephalus	Nottingham Road	4	3	2	0.5	0.00133
ophidicephalus	Hogsback/Stutterheim	5	0	1	0	0
ophidicephalus	Langeni/Baziya	4	2	3	0.83333	0.00089
ophidicephalus	Entabeni	3	2	2	0.6667	0.00118

Taxon	Forest	Sample size	#polymorphic sites	#Haplotypes	Haplotype diversity	Nucleotide diversity
ophidicephalus	Nkandla	3	1	2	0.66667	0.00059
Grouped by forest:						
demodocus	Louis Trichardt	5	9	5	1	0.00373
nireus	Louis Trichardt	5	5	4	0.9	0.00177
ophidicephalus	Louis Trichardt	10	3	4	0.3333	0.00053
echerioides	Louis Trichardt	7	2	3	0.66667	0.00067
echerioides	Haenertsburg	10	4	5	0.82222	0.00124
ophidicephalus	Haenertsburg	7	4	4	0.71429	0.00101
echerioides	Langeni/Baziya	3	0	1	0	0
ophidicephalus	Langeni/Baziya	4	2	3	0.83333	0.00089
echerioides	Nkandla	3	1	2	0.66667	0.00059
ophidicephalus	Nkandla	3	1	2	0.66667	0.00059
nireus	Nkandla	4	10	4	1	0.00457
dardanus	Pirrie + Ft Beaufort	4	10	4	1	0.00474
ophidicephalus	Pirrie	6	1	2	0.3333	0.00029

Discussion

Do forest-restricted butterfly species have more genetic structure than more widespread species?

In South Africa, *P. ophidicephalus* has previously been split into five subspecies based on the form of the discal spots on their wings (Van Son, 1939). Therefore, it is expected that *P. ophidicephalus* would have some genetic structure correlating to these subspecies. The results show that there is genetic structure within this group, in the form of a northern and a southern clade, but no evidence that mtDNA supports the recognition of five subspecies.

The Bayesian analysis splits *P. ophidicephalus* in two clades (Fig 2.4), corresponding to a northern and southern clade. The split to the northern clade is well supported (PP = 1.00). This clade includes samples collected in Northern Mistbelt (NMB) forests. The split to the southern clade does not have a good posterior probability indicating low support and includes samples from Southern Mistbelt (SMB) and Scarp forests. The split in the two groups seems to occur around the northern Swaziland border. Similar results have been obtained in a study on cicadas where samples from the Lydenberg area form a separate group (Price, 2010), this will be discussed further in the last chapter.

The other forest-associated species, *P. echerioides,* has the same genetic structure, while *P. dardanus* has no obvious genetic structure related to forests. The widespread species *P. demodocus* and *P. nireus* show no genetic structure at all, with the exception of Madagascan samples, which separate early from the other clades of *P. demodocus*.

As forest-restricted species have a more restricted habitat and are more prone to loss of genetic diversity within forest patches, it was expected that they would show more genetic structure between the different forest types.

The haplotype networks for *P. ophidicephalus* (Fig 2.8) shows the same split in clades as the Bayesian analysis, where there is a northern and a southern clade and the Lydenberg samples are included in the southern clade. The Lydenberg (NMB) samples are interlinked with samples from Nottingham Road (SMB) and Nkandla (Scarp).

The haplotype networks for *P. echerioides* show a general grouping where samples from NMB forests are not linked with samples from SMB forests.

The forest-associated species *Papilio dardanus* is split into four groups, but there is no evidence of spatial structure. This could be because this species is not strictly forest-

specific, but rather is restricted to woodlands and forest edges. Clark and Vogler (2009) reconstructed the lineage history of *P. dardanus*. This species has a very high phenotypic variation across its distribution range though Africa. Currently geographic forms have been divided into nine distinct subspecies across the wide distribution range of this species over sub-Saharan Africa and the Indian Ocean islands (Clark and Vogler, 2009).

For the haplotype networks, samples from other Central and East Africa and Madagascar were included, but the analysis shows that samples from all over Africa are interlinked (Figure 2.10).

Papilio demodocus shows no phylogeographic structure (Fig. 2.7). These butterflies are widespread and are not restricted to a specific vegetation type, but they are absent from extremely arid areas (Dickson and Kroon, 1978; Woodhall, 2005). *Papilio demodocus* is split into four unstructured groups but samples from the same forest types are not exclusively grouped together. Some Madagascan samples form a well-defined clade along with some samples from South Africa collected in NMB and SMB forests.

Papilio nireus show no phylogeographic structure. Like *P. demodocus*, this species is fairly widespread and has no real restrictions to any habitats, except that they do not occur in arid areas (Mecenero *et al.*, 2013).

Hypothesis 1 is thus not rejected by these results, as the forest-restricted butterflies collected for this study show evidence of genetic structure, which is correlated with geographic distribution and forest type. Forest-associated species such as *P. dardanus* show some structure, but this is not related to forest types. The widespread species show no structure at all.

Molecular dating and forest fragmentation

Over the last two million years southern Africa has gone through 20 climatic cycles each lasting about 100 000 years. This follows the trend in expansion and contraction periods of glacial ice sheets at higher latitudes (Deacon and Lancaster, 1988). These climate oscillations lead to wetter and dryer periods throughout southern Africa which has shaped the distribution of the forests (Partridge *et al*. 1999). As mentioned earlier Afromontane forests are more ancient than the IOCB forests, which have established only after the LGM about 8000 years ago. Therefore it is believed that the Afromontane forests have been through extinction filtering events, which shaped the species composition of these forests (Lawes *et al*., 2000a). The crown age of the five species studied here is about 1.4 million years, this is in the mid Pleistocene. Most of the within-species diversification occurred in the late

Pleistocene approximately 0.2 MYR ago. This correlates to the climate oscillation, which shaped today's forest distribution during the last glacial maximum about 18 000 years ago (Lawes, 1990). The split between the northern and southern clade in *P. ophidicephalus* occurred about 0.5 MYR ago. Therefore it is evident that the climatic changes throughout the Pleistocene have shaped the diversification of the different *Papilio* species in South Africa.

Is isolation by distance stronger in forest-restricted species?

Genetic isolation by distance is common in phytophagous insects, where there is significantly less gene flow with distance (Peterson and Denno, 1998). IBD relationships generally do not vary over different groups of insects. However, there is variation in IBD within some insect groups where gene flow decreases as geographic distance increases. Some species are characterised by steep declines in gene flow, such as the checkerspot butterfly and other species are not affected by IBD at all (Peterson and Denno, 1998). Mostly, where there is a decrease in gene flow with distance, it is represented by true IBD and not vicariance (Peterson and Denno, 1998). IBD slopes vary with the dispersal ability of a specific species: the more mobile the species, the greater the distance needed between populations to decrease gene flow, an aspect to be considered here as these butterflies are generally strong fliers and thus potentially highly vagile.

The IBD analysis (Fig 2.13) shows that *P. ophidicephalus* has the highest correlation between log genetic distance and log geographic distance ($r = 0.5553$), indicating that there is some restriction in gene flow between different populations across the country. Splitting *P. ophidicephalus* up into a northern and a southern clade shows that the northern clade has a lower correlation between genetic and geographic distance than the southern clade. The other forest-specific species, *P. echerioides*, has a fairly low r-value; but it is still higher than the more generalist species (Fig 2.14). This again indicates that there is some restriction in gene flow compared to the more generalist species (*P. dardanus, P. demodocus* and *P. nireus*) which have a very low correlation between log genetic and log geographic distance.

These results indicate that Hypothesis 2 is not rejected, as forest-restricted species have a higher IBD that indicates restriction in gene flow over distance. This is assumed to be a consequence of the considerable length of time that forest fragments have been isolated.

Does fragmentation decrease genetic diversity in forests?

P. ophidicephalus has the most haplotypes (25; Table 2.2). This was expected as they are forest-restricted and are thought to not travel long distances between forests. However,

the haplotype diversity for this species is lower than in the more generalist species such as *P. demodocus* and *P. nireus*. When comparing the haplotype diversity of the two different clades it is evident that the southern clade has a much higher diversity than the northern clade (Table 2.2). Since it was thought that the LGM had a greater effect on NMB forests (Deacon and Lancaster, 1988; Scott *et al*., 1997), the difference in haplotype diversity may be a consequence of historical genetic bottlenecks caused by fragmentation or founder effects in these forests. Comparing the genetic diversity of the northern and southern clade of *P. echerioides*, there is no major difference in haplotype diversity. Perhaps this is because the sample size of the southern clade is much lower than that of the northern clade.

The genetic diversity results of the forest-associated species *P. dardanus* (Table 2.2) show that there is less restriction in gene flow than in the more specific species, as *P. dardanus* has a higher haplotype diversity. However, it is still lower than that of the generalist species.

P. demodocus has 21 haplotypes with a haplotype diversity of 0.943. This was expected as this species is wide spread and a strong flyer, which allows it to inhabit a wide range of habitats (Dickson and Kroon, 1978). Similar results were obtained for *P. nireus*.

Comparing the genetic diversity of the different forest patches throughout the country it becomes clear that the forest-restricted species always have a lower genetic diversity. Forest-restricted species inhabit smaller areas, which can lead to inbreeding. As the area decreases, population size inevitably decreases, inbreeding increases and this could result in lower genetic diversity.

Forest fragmentation due to climate change, might decrease the diversity of plants within the forest (Danielsen, 1996). Butterfly diversity might not be directly affected as such, as they have a range of host plants and as long as these are present a forest patch should sustain a healthy population and avoid inbreeding depression and bottlenecks.

These results indicate that there is not enough evidence for Hypothesis 3 to be rejected; the forest-restricted species do have lower haplotype diversity than the more generalist species.

Conclusions

Forest-restricted butterflies have more phylogeographic structure than non-forest-specific species, and forest-associated species such as *P. dardanus* have intermediate levels of genetic structure. At the same time, they have lower haplotype diversity. These findings

suggest that historical fragmentation and forest community evolution has had considerable influence on forest biota. Results indicate a major historical split between NMB on the one hand and SMB and scarp forests on the other. However, the somewhat surprising result of *P. ophidicephalus* samples from Lydenberg not falling within the northern clade as expected based on geography requires further investigation. This pattern has been documented before in cicadas (Price, 2010). The reason for this is not entirely clear. It could be a distinct colonisation event from SMB to the forests near Lydenburg, or that the Lydenburg population could represent a refugium of a once more widespread SMB biota. Furthermore, the classification of forests in Mpumalanga is somewhat complicated. Lötter *et al*. (2013) recently reassessed the classification of forests in that province, in which Lydenburg falls. According to his classification, samples used here from Lydenbuerg were collected from Eastern Dry Afrotemperate forests. These forests are generally dryer than Mistbelt forests and fall outside the Mistbelt area (Lötter *et al*., 2013). However the Afrotemperate forests are similar to Mistbelt in species composition and evolutionary history. The forests in Mpumalanga are complicated, as the more recent forests (Afrotemperate and Mistbelt) are in close vicinity of the more ancient Scarp forests. This could play a role in species distribution of plants and animals as migration events between the different forests could easily happen (Lötter *et al*., 2013). There is a forest patch in northern KZN which is classified as Dry Afrotemperate forest. This could imply that the forests around Lydenberg are more similar to SMB forests than to NMB forests or even an intermediate between the two. This could be a reason for the Lydenburg samples turning out in the southern clade.

Another surprising result is that analyses of mtDNA of *P. ophidicephalus* only revealed evidence for two subspecies and not five. However, the variability of the COI gene (despite its recommended application to DNA barcoding (Elias *et al*., 2007; Hajibabaei *et al*., 2007; Hebert *et al*., 2003; Linares *et al*., 2009; Wiemers and Fiedler, 2007) may not be sufficient to retrieve the signal of recent evolutionary events leading to subspecies formation. The intraspecific taxonomy of this group will be explored further in the next chapter including analysis of a nuclear gene region and morphometric analyses.

3 A molecular and morphological investigation of the intraspecific taxonomy of *P. ophidicephalus*

Introduction

During his study of the genus *Papilio*, van Son (1939) noted several geographic variants of *P. ophidicephalus* in South Africa. He found that none of these 'races' could be referred to the East African *P. ophidicephalus*, even though some resembled them. By examining the genitalia of the different forms, he tried to establish a degree of relationship between them. He found differences in shape and structure of the valve which were specific for different forms. He also found distinct wing patterns, mainly discal spots, changing throughout the different 'races'. At that time, *P. ophidicephalus* was classified as a subspecies of *P. menestheus* and a similar rank was given to *P. lormieri*. Through the analysis of the genitalia, van Son established them as separate species and split *P. ophidicephalus* into ten subspecies, five of which occur in South Africa and one in Zimbabwe..

The key to identifying the three different species includes lengths of the tail and the arrangements of the discal spots on the forewing. Similarly, the key to identify the six southern African subspecies of *P. ophidicephalus* is based on the arrangement of the discal spots on the forewing. The main characters used for classification are the R5 spot (indicated in Fig. 3.1) and the arrangement of the discal spots in a line below it (van Son, 1939).

The Subspecies Concept

What is a subspecies? In 1942 Mayr defined subspecies as genetically distinct, geographically separated populations belonging to the same species and therefore interbreeding freely at the zones of contact (Wilson and Brown, 1953). However, there are problems in operationalising this popular definition, particularly regarding the tension between geographical separation and contact. The subspecies concept is also very subjective; where does one draw the line in defining a population of a species as a subspecies, particularly if evolutionary differentiation (e.g. speciation) is a continuous process? Even if the discrepancies from conflicting geographical variations are eliminated by the use of one or few characters, there is still is no real lower limit to the subspecies category. Wilson and

Brown (1953) state that "no arbitrary lower limit will ever be satisfactory, for even if only one character is used, there will always be borderline cases of an extremely vexing nature."

Aims

On the basis of mtDNA data, which suggest the presence of only two entities, the validity of *P. ophidicephalus* subspecies is investigated using mtDNA, nDNA and morphometrics.

Materials and Methods

Specimens were obtained from forest patches throughout South Africa, as described in Chapter 2. Samples were collected in distribution ranges of five of the ten subspecies of *P. ophidicephalus*. To include a sixth subspecies, pictures of two *P. o. chirinda* specimens from Dickson and Kroon (1978) were used as no samples were collected in the area of the distribution range of this subspecies.

Phylogenetic analysis

The ITS gene region was amplified and used as a separate dataset to test the mtDNA results. The primers ITS1-F (5'GCGTTCGAARTGCGATGATCAA3') and ITS1-R (5'GTAGGTGAACCTGCAGAAGG3'; Vogler and DeSalle, 1994) were used to amplify the region. The PCR reactions were done in a 50 µl reaction. The thermal cycling profile for the ITS region had the same temperatures and durations as in the COI profile (Chapter 2) but involved 35 cycles instead of 30 cycles. For the reaction volumes, the magnesium concentration was changed to 3 mM $MgCl_2$ and the volume of dH_20 was adjusted accordingly to make up the 50 µl reaction volume.

The sequence trace files were viewed and checked using GeneStudio v.2.1.1.5. (GeneStudio. Inc). Sequence files were then aligned using the ClustalW algorithm, which is included in the software package MEGA v.5 (Tamura *et al.*, 2011), and the alignment was then checked visually. The messy ends of each sequence have been trimmed and sequences lacking data at the 5' or 3' ends were filled with question marks to the end to indicate missing data (alignment in CD Appendix 2).

The Bayesian Inference analysis was conducted using MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001) using the GTR model. Two runs of 2 million generations, sampling every 200 generations with four chains (3 heated and 1 cold), were plotted against

the likelihood scores and tree length, to ascertain when the analysis reached stationarity, and the first 10% of the trees were discarded as burn-in, which represents 20 000 of the 200 000 trees sampled.

Morphometric analysis of wing spots

Three discal markings on the left wing (the R5 spot and another spot on the forewing, and a combination of spots on the hindwing: Fig. 3.1) were measured for all of the *P. ophidicephalus* samples collected. To take into account the shape of the spots, we used a network of box-truss dimensions that expresses shape as well as size, as traditional linear morphometric methods do not take into account the shape of a particular character (Cadrin, 2000).

Figure 3.1: A picture of *P. ophidicephalus* from (Dickson and Kroon, 1978) with the landmarks for the discal spot measurements superimposed on it. The R5 discal spot is indicated.

Pictures were taken of all of the specimens, including their sample numbers and a scale bar. ImageJ v.1.46r (Abràmoff *et al*., 2004) was used to measure the length and shape

of the landmarks. A scale bar was included in the image to calibrate ImageJ as this program works in pixels. The scale bar was reset for each picture to account for difference in focus between different pictures. In total, 23 measurements were taken on each sample (A-B, A-C, A-D, etc.; Fig 3.1). For each sample, all of the measurements were noted in a table along with their corresponding localities and sex. Images of specimens from Zimbabwe were obtained from Dickson and Kroon (1978) and measured, adjusting to the stated magnification of the images. Samples from all six of the southern African subspecies were therefore included in the analysis.

Statistical Analyses

Principal Component Analysis

A Principal Component Analysis (PCA) uses variables to determine structure within a group (StatSoft, 2013). We used PCA to examine structure in *P. ophidicephalus*. The analysis was done in PAST (PAleontological STatistics) (Hammer *et al*., 2001). Prior to analysis each variable was standardised (x-mean/stdev).

Discriminant Function Analysis

A DFA was done in STATISTICA 10 to determine which variables discriminate between different groups (StatSoft, 2013). In this case where, *P. ophidicephalus* is split into six subspecies, it was used to see what discal spot is the most variable and therefore determines the outcome of the grouping.

Results

ITS

The data set for the ITS gene region includes only samples of *P. ophidicephalus*. The data were generated to test the results obtained using mitochondrial COI sequence data (Chapter 2). This dataset consisted of 39 samples and the alignment was 447 base pairs in length (Provided in Appendix 2 on CD-ROM).

The Bayesian inference tree (Fig 3.2) is shown. The result is very similar to the COI tree where *P. ophidicephalus* is split into a northern and a southern clade with samples from Lydenberg embedded in the southern clade. No outgroups were included in the nDNA analysis, so the tree was midpoint rooted.

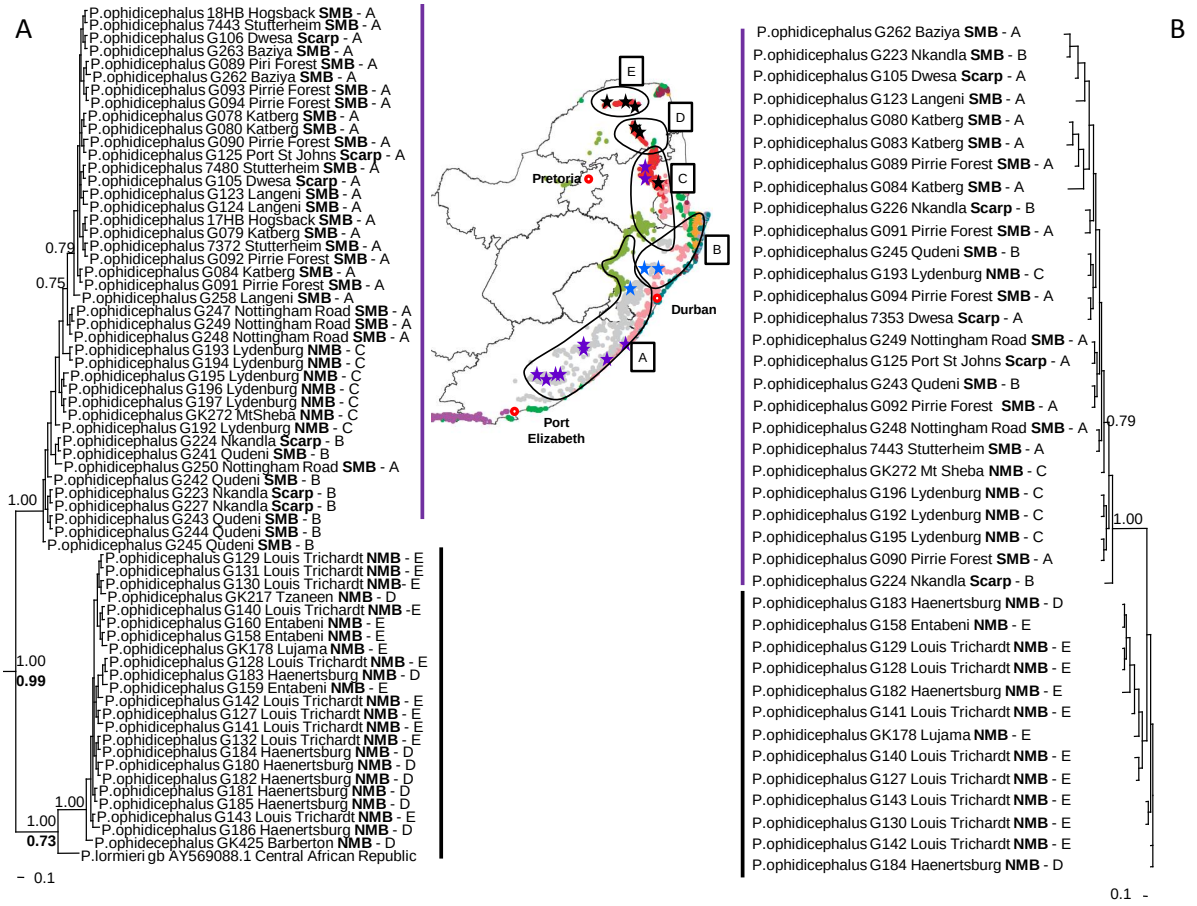

Figure 3.2: Bayesian Inference phylogram of the *P. ophidicephalus* samples showing the results of COI data (A) and ITS data only (B). Posterior probabilities > 0.70 are indicated at each node. Sample localities are mapped on the corresponding forest types. Colours of the stars correspond to the bars. The distribution ranges of the five subspecies are mapped (Mecenero *et al*., 2013). A: *P. o. phalusco*; B: *P. o. zuluensis*; C: *P. o. ayresi*; D: *P. o. transvaalensis*; E: *P. o. entabeni.*

Both the separate data sets of COI and ITS show a very similar result, and for that reason the data sets were combined to see if increased the resolution of the analysis (McDonald and Daniels, 2012). The combined dataset consisted of 1579 characters of 36 *P. ophidicephalus* samples. The Bayesian Inference tree is shown in Figure 3.3.

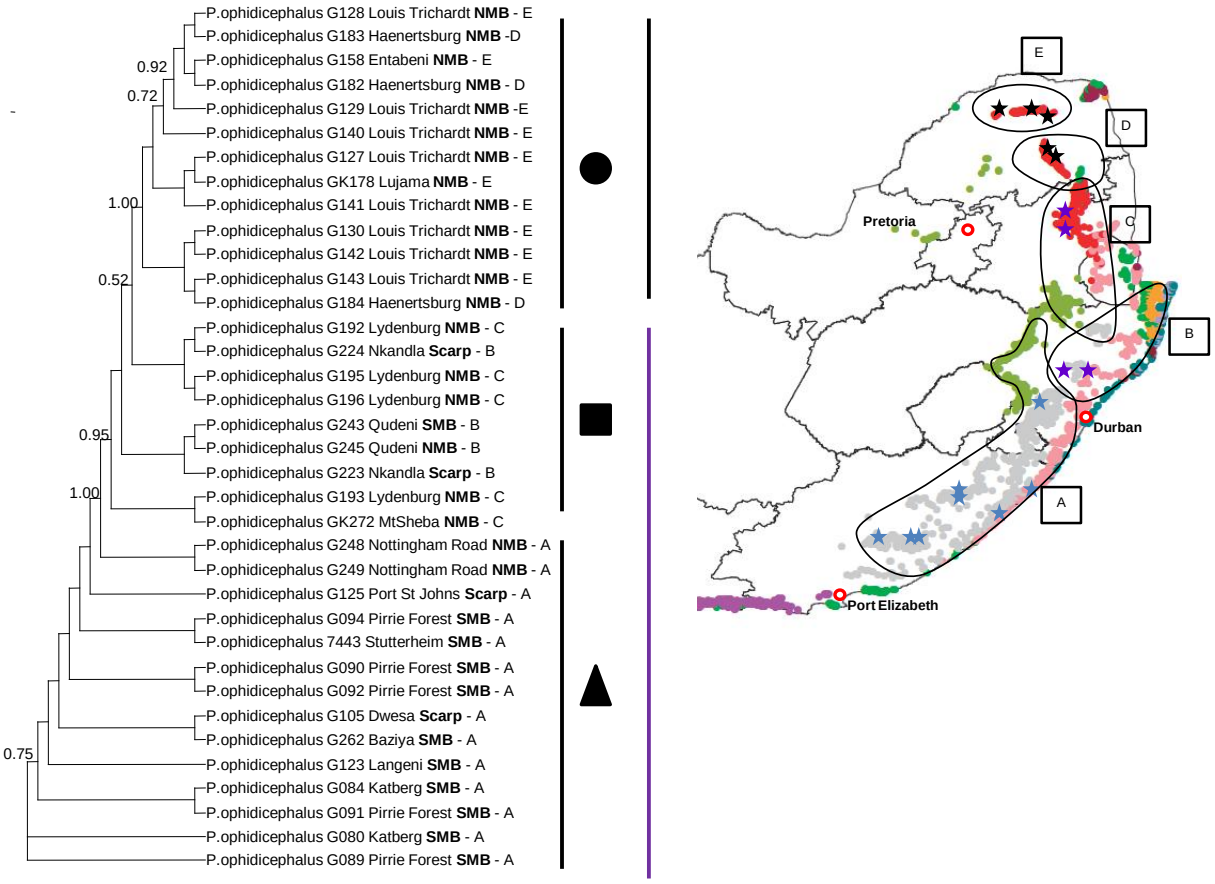

Figure 3.3: Bayesian Inference cladogram of the combined mtDNA and ITS dataset for *P. ophidicephalus* samples showing the results of COI and ITS data combined. Posterior probabilities >0.70 are indicated at each node. Sample localities are mapped on the corresponding forest types. Colours of the stars correspond to the bars. The distribution ranges of the five subspecies are mapped. A: *P. o. phalusco*; B: *P. o. zuluensis*; C: *P. o. ayresi*; D: *P. o. transvaalensis*; E: *P. o. entabeni* (Mecenero *et al*., 2013). Triangle = group which includes *P. o. phalusco* indicated by light blue stars; Square = group which includes *P. o. zuluensis P. o. ayresi* indicated by the purple stars; Circle = group which includes *P. o. transvaalensis* and *P. o. entabeni* indicated by black stars.

The combined dataset for *P. ophidicephalus* shows a very similar result (Fig. 3.3) to the single-gene analyses of COI and ITS. The group is split into two well-supported subgroups. One consists of samples collected in the northern parts of the country in Northern Mistbelt Forests (indicated by the black bar in Fig. 3.3) and the other one of samples collected further south in KwaZulu-Natal and Eastern Cape in Southern Mistbelt Forests and Scarp Forests with the exception of samples from Mt. Sheba and Lydenburg that fall within the Northern Mistbelt forests distribution. However there is evidence for more structure. There is good support (PP = 0.95) for *P. o. phalusco* forming a separate clade (triangle in Fig. 3.3). There is also some evidence for *P. o. zuluensis* and *P. o. ayresi* forming a separate

paraphyletic group separated from the northern and southern clade (square on Fig 3.3). The northern clade is made up of a mixture of *P. o. transvaaliensis* and *P. o. entabeni* (circle in Fig 3.3).

Principal Component Analysis

The results of the PCA show that there are two distinct phenotypes of *P. ophidicephalus*, a northern form that includes samples from Louis Trichardt and Haenertsburg and a southern form that consists of samples from KZN and EC. Samples from Lydenberg fall within the southern group and not the northern group, as found in mtDNA and nDNA analyses. Samples from Zimbabwe are similar to the northern group, but they do not fall within the group. Table 3.1 shows the Eigenvalues and the percentage of variance of the first four principal components.

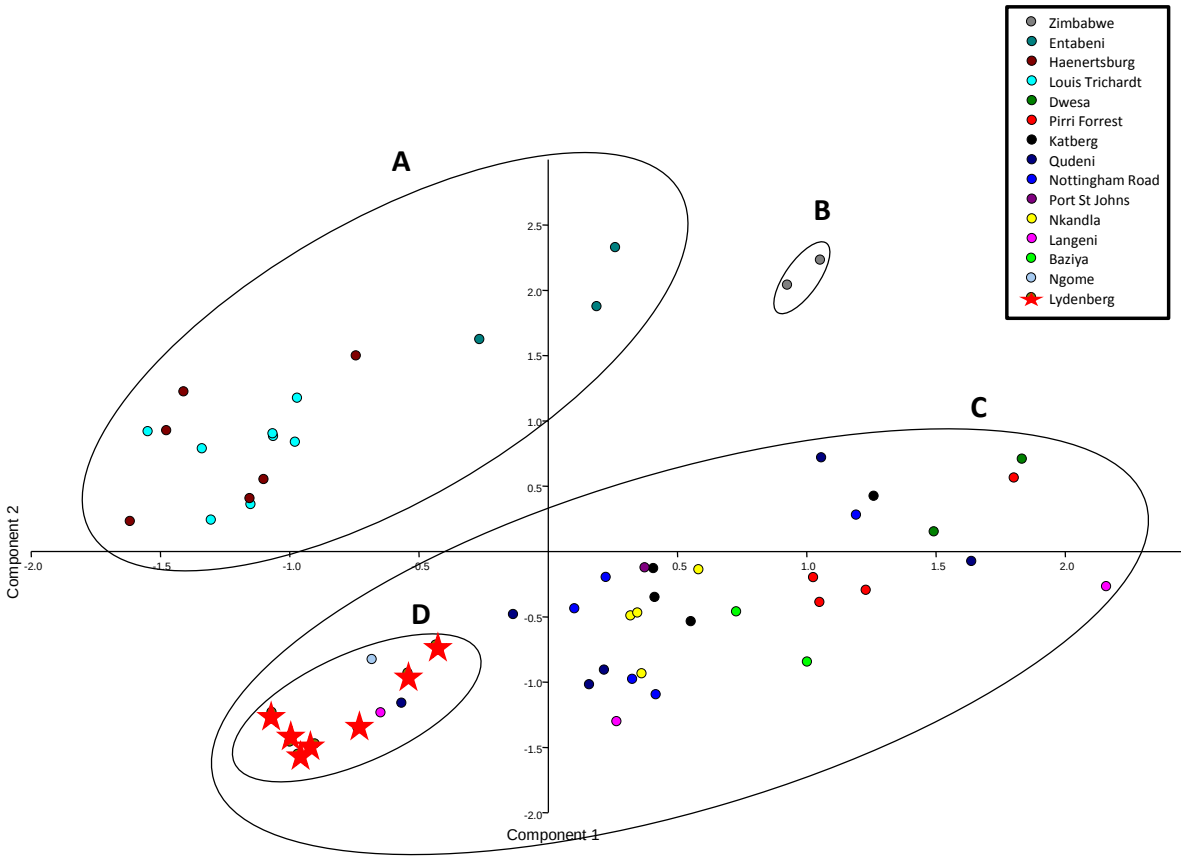

Figure 3.4: PCA ordination plot of *P. ophidicephalus* samples. A represents the northern clade. B are samples from Zimbabwe obtained from Dickson and Kroon (1978). C is the southern clade including samples from northern Mistbelt forests around Lydenberg (D).

Table 3.1: The Eigenvectors of the PCA along with the percentage contribution to total variance.

PC	Eigenvalues	% variance
1	10.7454	46.719
2	3.6317	15.79
3	2.87066	12.481
4	1.26707	5.509

Discriminant Function Analysis

The Discriminant Function Analysis ordination plot (Fig. 3.3) clearly shows five clusters, corresponding to subspecies. The two northern most subspecies in South Africa *P. o. entabeni* and *P. o. transvaalensis* do not form a separate cluster but seem to be grouped together. Again, there is evidence of a northern and southern group. The subspecies are ordered south to north in the southern clade. The northern group is formed of *P. o. entabeni* and *P. o. transvaalensis*.

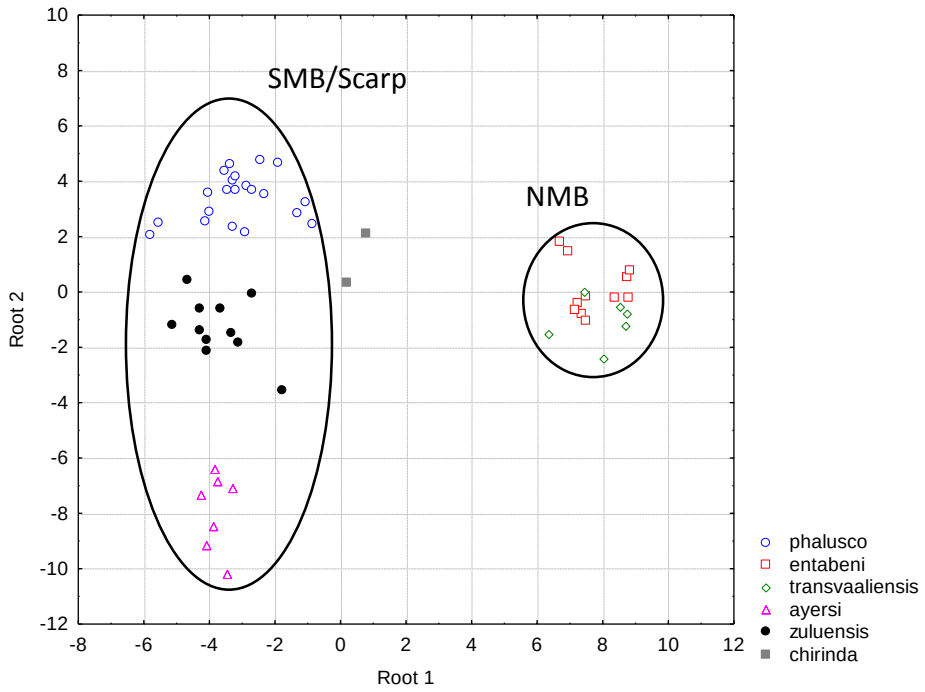

Figure 3.5: Graph showing the results for the different subspecies of the DFA.

The classification table (Table 3.2) shows that only *P. o. transvaalensis* was ambiguously identified because one of the six samples is more similar to *P. o. entabeni*; all of the other specimens were correctly identified. The structure matrix showing the Canonical

Roots show which variables are the most diagnostic. It can clearly be seen that the R5 discal spot is important, as it has the highest or lowest coefficients (Table 3.3).

Table 3.2: Classification table of the DFA.

Subspecies	**% correct**	***phalusco***	***entabeni***	***transv***	***ayresi***	***zuluensis***	***chirinda***
phalusco	100.0000	21	0	0	0	0	0
entabeni	100.0000	0	11	0	0	0	0
transv	83.3333	0	1	5	0	0	0
ayersi	100.0000	0	0	0	7	0	0
zuluensis	100.0000	0	0	0	0	11	0
chirinda	100.0000	0	0	0	0	0	2

Table 3.3 Factor Structure Matrix including Correlations Variables - Canonical Roots (Pooled-within-groups correlations). Most important variables are highlighted.

	Root 1	**Root 2**	**Root 3**	**Root 4**	**Root 5**
A-B	**-0.394697**	**0.419251**	-0.001960	-0.018450	0.029164
A-C	-0.168411	0.162252	-0.006964	-0.228206	-0.031096
A-D	**-0.382105**	**0.412528**	0.007400	0.046744	-0.001620
B-D	-0.138202	-0.255869	-0.051011	-0.141130	0.175520
B-C	**-0.214911**	0.324649	0.069118	0.160114	0.017963
R5 C-D	**-0.206937**	0.353226	0.073694	0.211690	-0.022980
C-E	-0.177656	0.069865	0.092255	-0.205132	-0.093505
C-F	-0.107300	0.288666	0.039037	0.100586	-0.291446
D-F	-0.138577	-0.275105	-0.051778	-0.259566	0.259544
D-E	**-0.237523**	0.398179	-0.020294	0.140858	0.124509
E-F	**-0.219514**	**0.444090**	-0.014969	0.195489	0.069945
G-H	0.136141	0.145219	0.154725	0.089827	0.082552
G-I	-0.172856	0.217925	0.064865	0.067835	0.326511
G-J	-0.015223	0.035636	-0.012825	-0.074993	0.026750
H-I	0.120992	0.182742	0.217528	0.053967	0.106632
I-J	0.188670	0.177946	0.361163	0.026037	0.147096
H-J	-0.186215	0.075494	0.160077	0.277104	-0.039775
K-L	-0.072624	0.093370	0.096934	0.142007	0.159099
K-M	-0.063051	0.028359	0.138908	0.103920	0.204699
K-N	-0.066047	0.071571	0.133077	0.044890	0.204952
L-M	-0.076116	0.007128	0.083978	0.133354	0.137483
L-N	-0.071288	0.028285	0.071207	0.086250	0.078435
M-N	0.004990	-0.072482	0.047325	0.260948	0.075442

Discussion

The Bayesian analysis of the ITS gene region alone and combined with mtDNA shows *P. ophidicephalus* to be split into two main clades (Figs 3.2, 3.5 and 3.6). Furthermore, the combined dataset suggests there is some evidence to split the South African populations of *P. ophidicephalus* into separate subspecies. For a concrete conclusion about the subspecies classification, more work needs to be done. For example, it is possible that van Son's subspecies classification could be genuine, and morphological mutation has outpaced mtDNA and nDNA mutation rates. A more sensitive DNA approach, such as microsatellites analysis, may detect a signal that correlates with morphology.

The results of the PCA do not show six groups as expected, but only two groups. As in the previous chapter for the genetic analysis, there is a northern and a southern group. Samples from Lydenberg again fall within the southern group (Fig 3.2).

The difference in the results obtained between van Son and the data presented here could be due to difference in interpretation. Van Son classified the characters by eye and also assessed them by how much one discal spot touched adjacent spots. This form of classification can be subjective. By using a fixed set of landmarks and measuring the shape of the discal spots, subjectivity can be reduced. The spot that changes the most is R5 (Table 3.3). This spot changes from a teardrop shape, with a pointed side, in the northern populations, to a more rectangular shape in the southern populations. The other discal spots do not change as significantly as the R5 spot.

However, the DFA shows that *P. ophidicephalus* can be grouped into some of the previously identified subspecies (Fig 3.3). The south-north split can still be noticed in the DFA. The southern subspecies *P. o. phalusco*, *P. o. ayresi* and *P. o. zuluensis* could be grouped together and then the northern subspecies *P. o. entabeni* and *P. o. transvaaliensis* could be grouped together, but there is not enough evidence from the DFA that these two can be identified as separate subspecies. Interestingly the samples from Zimbabwe (*P. o. chirinda*) seem to fit in better with the southern group in the Discriminant Function Analysis, but only two samples were measured from photos, and an increase in sample size could change the result.

Conclusions

There is not enough evidence to divide South African populations of *P. ophidicephalus* into five different subspecies. This study suggests two subspecies or probably

even two separate species. This conclusion is stronger when comparing the genetic and morphometric results. Both show a similar pattern of northern and southern populations with samples from Lydenburg related to the southern population. However before final conclusions can be drawn regarding the subspecies classification, more work needs to be done as there is some evidence for more divisions. A more sensitive DNA approach, such as microsatellites, may resolve the samples into the corresponding subspecies. In addition, van Son (1939) analysed the genitalia of different subspecies in combination with the discal spots from which he classified the different subspecies. The DNA data and the wing morphological data should be examined together with data from genitalia to make a final decision on the intraspecific taxonomy of *P. ophidicephalus*. However, in light of current available data, it is suggested that *P. ophidicephalus* be split into two separate species rather than five subspecies. However before that conclusion is drawn other recognised subspecies of *P. ophidicephalus* not included here should also be evaluated to see where they would fit into molecular and morphological analyses. Nevertheless, as mentioned in Chapter 1, the issue around the subspecies concept is subjective and no lower limit is available to discriminate between subspecies or species. Therefore the state of this species is open to discussion.

4 General Discussion

Summary of findings

Forest-restricted butterflies have more genetic structure than forest-associated or non-forest-specific butterflies. Furthermore, we found a stronger correlation between genetic and geographic distance in forest-restricted butterflies (*Papilio ophidicephalus* and *P. echerioides*) than in less habitat-restricted species. This is attributed to the effects of fragmentation, which decreases genetic diversity in forest butterflies due to restriction of suitable habitat, which affects the population size, resulting in genetic bottlenecks.

Correlation with Forest types

Forest types do not adequately describe the population structure of the forest-restricted species *P. ophidicephalus* and *P. echerioides*. There is some evidence that the diversification of *Papilio* populations is related to the main forest types. It is suggested here that this diversification occurred in the Pliocene or Pleistocene, as this has been documented in a number of studies on a number of different groups of forest-dwelling organisms. These studies show species radiation follows the expansion and contraction cycles of forests in Africa (Hughes *et al*., 2005; Lawes, 1990; Lawes *et al*., 2000a; McDonald and Daniels, 2012; Roy *et al*., 2001; Tolley *et al*., 2006). The more generalist species (*P. dardanus*, *P. demodocus* and *P. nireus*) show no real genetic structure as they have a wide distribution range.

Genetic and morphological investigation of *P. ophidicephalus*

In South Africa, *P. ophidicephalus* has previously been split into five subspecies according to the form of the discal spots on the wings (Van Son, 1939). Samples collected in Louis Trichardt, Entabeni and Lajama (Fig 2.4) fall within the distribution range of *P. o. entabeni*. Samples from Haenertsburg and Tzaneen fall within the geographical range of *P. o. transvaalensis* (Mecenero *et al.*, 2013). Furthermore samples collected in Lydenburg, Mt. Sheba and Barberton were collected in the distribution range of *P. o. ayresi*. Samples from

Qudeni and Nkandla fall within the distribution range of *P. o. zuluensis*. *Papilio o. phalusco* has the biggest distribution range, ranging through the Eastern Cape and south western parts of KZN. All of the samples collected in the Eastern Cape and samples from Nottingham Road fall within this distribution range (Mecenero *et al.*, 2013).

The mtDNA data suggests that there is not enough evidence to divide southern African *P. ophidicephalus* up into six subspecies. Comparing these results to nDNA data and morphological data, it becomes evident that there is more genetic structure than initially thought, based on CO1 data alone (Chapter 2). Upon combining mtDNA and nDNA data more segregation can be detected. In the combined data set there is evidence for three different subspecies (Fig 3.3). The morphological data also show a similar result to the mtDNA data where there are two clades. However the DFA discriminates the most, if not all, different subspecies. Since there is evidence for more structure, van Son's (1939) classifications might hold up. However this study provides enough evidence, both molecular and morphological, to suggests that *P. ophidicephalus* in South Africa should be split into two separate species rather than five subspecies.

The Lydenburg anomaly

This study revealed that samples from Lydenburg do not fall within the northern clade, but in the southern clade. mtDNA, nDNA and morphological data showed the same trend. The reason for this is not entirely clear. The one reason could be that the forests around Lydenburg have been misclassified and are actually SMB forests and not NMB forests. A study by Lötter *et al*. (2013) revealed that the classification of forests in Mpumalanga is somewhat complicated. By comparing the collection sites of the samples to their forest classification it becomes clear that the forests are not Mistbelt forests but Afrotemperate forests. It is suggested that that region is complicated when it comes to species composition as the more recent forests (Mistbelt and Afrotemperate) occur in close vicinity of the more ancient Scarp forests (Lötter *et al.*, 2013).

Similar results have been found in a study on forest cicadas (Price, 2010). The results show a split into a northern and a southern clade with well supported clades throughout South Africa. The samples from the Lydenberg area fall within the southern clade; however, samples from Mpumalanga form a separate clade. To resolve this issue, more genetic data of different taxa need to be collected from across the study area, which also covers the different forest types in Mpumalanga identified by Lötter *et al*. (2013).

The work presented here is the first comparative phylogeographic study of genetic diversity of lepidopteran species in southern Africa. While perhaps superficial in its sampling range and coverage, it serves to highlight the value of phylogeographic studies of invertebrates in elucidating the historical complexities of the vegetation with which they are associated, and forms a point of departure for more nuanced and informed future studies of this type.

References

Abràmoff, M.D., Hospitals, I., Magalhães, P.J., Abràmoff, M., 2004. Image Processing with ImageJ. Biophotonics Int. 11, 36–44.

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge.

Balmford, A., 1996. Extinction Filters and current resilience: the significance of past pressures for conservation biology. TREE 11, 193–196.

Bazinet, A.L., Zwickl, D.J., Cummings, M.P., 2014. A Gateway for Phylogenetic Analysis Powered by Grid Computing Featuring GARLI 2.0. Syst. Biol.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L., 2008. GenBank. Nucleic Acids Res. 36, D25–30.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. a, Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.

Cadrin, S.X., 2000. Advances in morphometric identification of fishery stocks. Rev. Fish Biol. Fish. 10, 91–112.

Caterino, M.S., Sperling, F.A.H., 1999. Papilio phylogeny based on mitochondrial cytochrome oxidase I and II genes. Mol. Phylogenet. Evol. 11, 122–37.

Clark, R., Vogler, A.P., 2009. A phylogenetic framework for wing pattern evolution in the mimetic Mocker Swallowtail Papilio dardanus. Mol. Ecol. 18, 3872–84.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–9.

Collins, N.M., Morris, M.G., 1985. Threatened Swallowtail Butterflies of the World. The IUCN Red Data Book. IUCN.

Condamine, F.L., Toussaint, E.F.A., Cotton, A.M., Genson, G.S., Sperling, F.A.H., Kergoat, G.J., 2013. Fine-scale biogeographical and temporal diversification processes of peacock swallowtails (Papilio subgenus Achillides) in the Indo-Australian Archipelago. Cladistics 29, 88–111.

Cornell, H. V., Lawton, J.H., 1992. Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. J. Anim. Ecol. 61, 1–12.

Danielsen, F., 1996. Stable environments and fragile communities: does history determine the resilience of avian rain-forest communities to habitat degradation? Biodivers. Conserv. 6, 423–434.

Deacon, J., Lancaster, N., 1988. Late Quaternary palaeoenvironments of southern Africa. Clarendon Press, Oxford.

Dickson, C.G.C., Kroon, D.M. (Eds.), 1978. Penningtons Butterflies. AD. Donker/Publishers, Johannesburg.

Eeley, H.A.C., Lawes, M.J., Piper, S.E., 1999. The influence of climate change on the distribution of indigenous forest in. J. Biogeogr. 26, 595–617.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J., Jiggins, C.D., 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc. R. Biol. Sci. / Biol. Sci. / 274, 2881–9.

Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N., Hickey, D.A., 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 23, 167–72.

Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST : PALEONTOLOGICAL STATISTICS SOFTWARE PACKAGE FOR EDUCATION AND DATA ANALYSIS 4, 1–9.

Hebert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. Biol. Sci. 270 Suppl , S96–9.

Henning, G.A., Henning, S.F., Joannou, J.G., Woodhall, S.E., 1997. Living Butterflies of Southern Africa: Biology, Ecology, Conservation Volume 1: Hesperiidae, Papilionidae and Pieridae of South Africa. Umdaus Press, Johannesburg.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Hughes, M., Möller, M., Bellstedt, D.U., Edwards, T.J., de Villiers, M., 2005. Refugia, dispersal and divergence in a forest archipelago: a study of Streptocarpus in eastern South Africa. Mol. Ecol. 14, 4415–26.

Jensen, J.L., Bohonak, A.J., Kelley, S.T., 2005. Isolation by distance, web service. BMC Genet. 6, 13.

Kroon, D.M., 1999. Lepidoptera of Southern Africa: Host-Plant & Other Associations. Lepidoperists' Society of Africa and D.M. Kroon, Sasolburg.

Lawes, M.J., 1990. The distribution of the samango monkey (Cercopithecus mitis erythrarchus Peters, 1852 and Cercopithecus mitis labiatus l. Geoffroy, 1843) and forest history in south Africa. J. Phylogeography 17, 669–680.

Lawes, M.J., Eeley, H.A.C., Findlay, N.J., Forbes, D., 2007. Resilient forest faunal communities in South Africa: a legacy of palaeoclimatic change and extinction filtering? J. Biogeogr. 34, 1246–1264.

Lawes, M.J., Eeley, H.A.C., Piper, S.E., 2000a. The relationship between local and regional diversity of indigenous forest fauna in KwaZulu-Natal Province, South Africa. Biodivers. Conserv. 9, 683–705.

Lawes, M.J., Mealin, P.E., Piper, S.E., 2000b. Occupancy of and Forest Metapopulation in Dynamics Three Forest South Africa Fragmented Afromontane in. Conserv. Biol. 14, 1088–1098.

Linares, M.C., Soto-Calderón, I.D., Lees, D.C., Anthony, N.M., 2009. High mitochondrial diversity in geographically widespread butterflies of Madagascar: a test of the DNA barcoding approach. Mol. Phylogenet. Evol. 50, 485–95.

Lötter, M.C., Mucina, L., Witkowski, E.T.F., 2013. Classification of the indigenous forests of Mpumalanga Province , South Africa. South African J. Bot. 90, 37–51.

Mayr, E., 1942. Systematics and the origin of species. Columbia Univ. Press, Columbia.

McDonald, D.E., Daniels, S.R., 2012. Phylogeography of the Cape velvet worm (Onychophora: Peripatopsis capensis) reveals the impact of Pliocene/Pleistocene climatic oscillations on Afromontane forest in the Western Cape, South Africa. J. Evol. Biol. 25, 824–35.

Mecenero, S., Ball, J.B., Edge, D.A., Hamer, M.L., Henning, G.A., Kruger, M., Pringle, E.L., Terblanche, R.F., Williams, M.C. (Eds.), 2013. Conservation Assessment of Butterflies of South Africa, Lesotho andSwaziland: Red List and Atlas. Saftronics (Pty) Ltd and Animal Demography Unit, Johannesburg.

Mucina, L., Geldenhuys, C.J., Rutherford, M., Powrie, L.W., Lötter, M.C., von Maltitz, G., Euston-Brown, D.I.W., Matthews, W.S., Dobson, L., Mckenzie, B., 2006. Afrotemperate, subtropical and azonla forests, in: Mucina, L., Rutherford, M. (Eds.), The Vegetation of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute, Pretoria, pp. 584–615.

Mucina, L., Rutherford, M., 2006. The Vegetation of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute, Pretoria.

Partridge, T.C., Scott, L., Hamilton, J.E., 1999. Synthetic reconstructions of southern African environments during the Last Glacial Maximum (21-18 kyr) and the Holocene Altithermal (8-6 kyr) 58, 207–214.

Peterson, M.A., Denno, R.F., 1998. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. Am. Nat. 152, 428–46.

Price, B.W., 2010. Historical biogeography of the tribe Platypleurini Schmidt, 1918 (Hemiptera: Cicadidae) with a focus on southern Africa. Rhodes University.

Roy, M.S., Sponer, R., Fjeldså, J., 2001. Molecular systematics and evolutionary history of Akalats (Genus Sheppardia): A Pre-Pleistocene Radiation in a Group of African forest birds. Mol. Phylogenet. Evol. 18, 74–83.

Sanderson, M.J., 1997. A Nonparametric Approach of Rate Constancy to Estimating Divergence Times in the Absence 1218–1231.

Scott, L., Anderson, H., Anderson, J.M., 1997. Vegetation history. In: Vegetation in Southern Africa. Cambridge University Press, Cambridge.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. Ann. Entomol. Soc. Am. 87, 651–701.

Sperling, F.A.H., Byers, R., Hickey, D., 1996. Mitochondrial DNA sequence variation among pheromotypes of the dingy cutworm, Feltia jaculifera (Gn.) (Lepidoptera: Noctuidae). Can. J. Zool. 74, 2109–2117.

Sperling, F.A.H., Harrison, R.G., 1994. Mitochondrial DNA Variation Within and Between Species of the Papilio machaon Group of Swallowtail Butterflies. Evolution (N. Y). 48, 408–422.

StatSoft, I., 2013. Electronic Statistics Textbook [WWW Document]. URL http://www.statsoft.com/textbook/ (accessed 2.26.14).

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0 Beta. Folrida State University.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28, 2731–2739.

Tolley, K.A., Burger, M., Turner, A.A., Matthee, C.A., 2006. Biogeographic patterns and phylogeography of dwarf chameleons (Bradypodion) in an African biodiversity hotspot. Mol. Ecol. 15, 781–93.

Van Son, G., 1939. Notes on the Menestheus group of Papilio L. and the South African Races of Papilio ophidicephalus Oberthur, containing descriptions of five new Subspecies and a new Variety. Ann. Transvaal Museum 19, 53–64.

Vogler, A., DeSalle, R., 1994. Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle Cicindela dorsalis. Mol. Biol. Evol. 11, 393–405.

Von Maltitz, G., Mucina, L., Geldenhuys, C.J., Lawes, M.J., Eeley, H.A.C., Aidie, H., Vink, D., Fleming, G., Bailey, C., 2003. Classification system for South African indigenous forests. An objective classification for the Department of Water Affairs and Forestry. Pretoria.

Wahlberg, N., Wheat, C.W., 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of lepidoptera. Syst. Biol. 57, 231–42.

Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae). Front. Zool. 4, 8.

Wilson, E.O., Brown, W.L., 1953. The Subspecies concept and Its Taxonomic Application. Syst. Zool. 2, 97–111.

Woodhall, S.E., 2005. Field Guide to Butterflies of South Africa. Struik Publishers, Cape Town.

Zakharov, E. V, Caterino, M.S., Sperling, F.A.H., 2004a. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus Papilio (Lepidoptera: Papilionidae). Syst. Biol. 53, 193–215.

Zakharov, E. V, Smith, C.R., Lees, D.C., Cameron, A., Vane-Wright, R.I., Sperling, F.A.H., 2004b. Independent gene phylogenies and morphology demonstrate a malagasy origin for a wide-ranging group of swallowtail butterflies. Evolution 58, 2763–82.

Appendix 1

Table A1: Sample localities of all collected butterflies along with collection date and collector.

Sample #	Taxon	Locality	Collection Date	Collector
G263	*ophidicephalus*	Baziya	06-Mar-13	G.G. Neef
G090	*ophidicephalus*	Pirrie Forest	24-Feb-12	J. Nicolau
G089	*ophidicephalus*	Pirrie Forest	24-Feb-12	J. Nicolau
G251	*ophidicephalus*	Nottingham Road	12-Feb-13	G.G. Neef
G082	*ophidicephalus*	Katberg	31-Jan-12	G.G. Neef
G078	*ophidicephalus*	Katberg	31-Jan-12	G.G. Neef
G086	*ophidicephalus*	Katberg	31-Jan-12	G.G. Neef
G079	*ophidicephalus*	Katberg	31-Jan-12	G.G. Neef
G080	*ophidicephalus*	Katberg	31-Jan-12	G.G. Neef
G250	*ophidicephalus*	Nottingham Road	12-Feb-13	G.G. Neef
G262	*ophidicephalus*	Baziya	06-Mar-13	G.G. Neef
G091	*ophidicephalus*	Pirrie Forest	24-Feb-12	J. Nicolau
G106	*ophidicephalus*	Dwesa	04-Mar-12	G.G. Neef
G241	*ophidicephalus*	Qudeni	08-Feb-13	G.G. Neef
G242	*ophidicephalus*	Qudeni	08-Feb-13	G.G. Neef
G243	*ophidicephalus*	Qudeni	08-Feb-13	G.G. Neef
G249	*ophidicephalus*	Nottingham Road	12-Feb-13	G.G. Neef
G244	*ophidicephalus*	Qudeni	08-Feb-13	G.G. Neef
G085	*ophidicephalus*	Katberg	31-Jan-21	G.G. Neef
G245	*ophidicephalus*	Qudeni	08-Feb-13	G.G. Neef
G246	*ophidicephalus*	Qudeni	08-Feb-13	G.G. Neef
G247	*ophidicephalus*	Nottingham Road	12-Feb-13	G.G. Neef
G248	*ophidicephalus*	Nottingham Road	12-Feb-13	G.G. Neef
G083	*ophidicephalus*	Katberg	31-Jan-12	G.G. Neef
G084	*ophidicephalus*	Katberg	31-Jan-12	G.G. Neef
G258	*ophidicephalus*	Mthata	04-Mar-13	G.G. Neef
G185	*ophidicephalus*	Haenertsburg	26-Jan-13	G.G. Neef
G092	*ophidicephalus*	Pirrie Forest	24-Feb-12	J. Nicolau
G198	*ophidicephalus*	Lydenburg	29-Jan-13	G.G. Neef
G131	*ophidicephalus*	Louis Trichardt	22-Jan-13	G.G. Neef
G197	*ophidicephalus*	Lydenburg	29-Jan-13	G.G. Neef
G196	*ophidicephalus*	Lydenburg	29-Jan-13	G.G. Neef
G195	*ophidicephalus*	Lydenburg	29-Jan-13	G.G. Neef
G194	*ophidicephalus*	Lydenburg	29-Jan-13	G.G. Neef
G193	*ophidicephalus*	Lydenburg	29-Jan-13	G.G. Neef
G142	*ophidicephalus*	Louis Trichardt	23-Jan-13	G.G. Neef
G186	*ophidicephalus*	Haenertsburg	26-Jan-13	G.G. Neef
G143	*ophidicephalus*	Louis Trichardt	23-Jan-13	G.G. Neef
G184	*ophidicephalus*	Haenertsburg	26-Jan-13	G.G. Neef
G158	*ophidicephalus*	Entabeni	24-Jan-13	G.G. Neef

Sample #	Taxon	Locality	Collection Date	Collector
G159	*ophidicephalus*	Entabeni	24-Jan-13	G.G. Neef
G160	*ophidicephalus*	Entabeni	24-Jan-13	G.G. Neef
G183	*ophidicephalus*	Haenertsburg	26-Jan-13	G.G. Neef
G182	*ophidicephalus*	Haenertsburg	26-Jan-13	G.G. Neef
G181	*ophidicephalus*	Haenertsburg	26-Jan-13	G.G. Neef
G180	*ophidicephalus*	Haenertsburg	26-Jan-13	G.G. Neef
G192	*ophidicephalus*	Lydenburg	29-Jan-13	G.G. Neef
G127	*ophidicephalus*	Louis Trichardt	22-Jan-13	G.G. Neef
G093	*ophidicephalus*	Pirrie Forest	24-Feb-12	J. Nicolau
G094	*ophidicephalus*	Pirrie Forest	24-Feb-12	J. Nicolau
G105	*ophidicephalus*	Dwesa	04-Mar-12	G.G. Neef
G227	*ophidicephalus*	Nkandla	05-Feb-13	G.G. Neef
G226	*ophidicephalus*	Nkandla	05-Feb-13	G.G. Neef
G225	*ophidicephalus*	Nkandla	05-Feb-13	G.G. Neef
G224	*ophidicephalus*	Nkandla	05-Feb-13	G.G. Neef
G223	*ophidicephalus*	Nkandla	05-Feb-13	G.G. Neef
G217	*ophidicephalus*	Ngome	02-Feb-13	G.G. Neef
G123	*ophidicephalus*	Langeni	05-Dec-12	G.G. Neef
G125	*ophidicephalus*	Port St Johns	03-Dec-12	G.G. Neef
G141	*ophidicephalus*	Louis Trichardt	23-Jan-13	G.G. Neef
G128	*ophidicephalus*	Louis Trichardt	22-Jan-13	G.G. Neef
G129	*ophidicephalus*	Louis Trichardt	22-Jan-13	G.G. Neef
G130	*ophidicephalus*	Louis Trichardt	22-Jan-13	G.G. Neef
G132	*ophidicephalus*	Louis Trichardt	22-Jan-13	G.G. Neef
G140	*ophidicephalus*	Louis Trichardt	23-Jan-13	G.G. Neef
G124	*ophidicephalus*	Langeni	05-Dec-12	G.G. Neef
G103	*echerioides*	Pirrie Forest	24-Feb-12	J. Nicolau
G145	*echerioides*	Louis Trichardt	23-Jan-13	G.G. Neef
G265	*echerioides*	Mthata	06-Mar-13	G.G. Neef
G146	*echerioides*	Louis Trichardt	23-Jan-13	G.G. Neef
G147	*echerioides*	Louis Trichardt	23-Jan-13	G.G. Neef
G189	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G148	*echerioides*	Louis Trichardt	23-Jan-13	G.G. Neef
G266	*echerioides*	Mthata	06-Mar-13	G.G. Neef
G261	*echerioides*	Mthata	04-Mar-14	G.G. Neef
G260	*echerioides*	Mthata	04-Mar-13	G.G. Neef
G149	*echerioides*	Louis Trichardt	23-Jan-13	G.G. Neef
G176	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G205	*echerioides*	Lydenburg	29-Jan-13	G.G. Neef
G202	*echerioides*	Lydenburg	29-Jan-13	G.G. Neef
G203	*echerioides*	Lydenburg	29-Jan-13	G.G. Neef
G204	*echerioides*	Lydenburg	29-Jan-13	G.G. Neef
G240	*echerioides*	Entumeni	07-Feb-13	G.G. Neef
G259	*echerioides*	Mthata	04-Mar-13	G.G. Neef
G230	*echerioides*	Nkandla	05-Feb-13	G.G. Neef
G175	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G174	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G173	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G172	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G171	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G170	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G169	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G177	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef

Sample #	Taxon	Locality	Collection Date	Collector
G178	*echerioides*	Haenertsburg	26-Jan-13	G.G. Neef
G231	*echerioides*	Nkandla	05-Feb-13	G.G. Neef
G229	*echerioides*	Nkandla	05-Feb-13	G.G. Neef
G228	*echerioides*	Nkandla	05-Feb-13	G.G. Neef
G155	*echerioides*	Louis Trichardt	23-Jan-13	G.G. Neef
G232	*echerioides*	Nkandla	05-Feb-13	G.G. Neef
G252	*echerioides*	Nottingham Road	12-Feb-13	G.G. Neef
G157	*echerioides*	Louis Trichardt	23-Jan-13	G.G. Neef
G087	*echerioides*	Katberg	31-Jan-12	G.G. Neef
G088	*echerioides*	Katberg	31-Jan-12	G.G. Neef
G076	*echerioides*	Katberg	31-Jan-12	G.G. Neef
G104	*echerioides*	Pirrie Forest	24-Feb-12	J. Nicolau
G077	*echerioides*	Katberg	31-Jan-12	G.G. Neef
G102	*echerioides*	Pirrie Forest	24-Feb-12	J. Nicolau
G116	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G121	*dardanus*	Port St Johns	03-Dec-12	G.G. Neef
G126	*dardanus*	Port St Johns	03-Dec-12	G.G. Neef
G107	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G119	*dardanus*	Port St Johns	03-Dec-12	G.G. Neef
G117	*dardanus*	Port St Johns	03-Dec-12	G.G. Neef
G070	*dardanus*	Morgans Bay	06-Mar-11	G.G. Neef
G222	*dardanus*	Eshowe	05-Feb-13	G.G. Neef
G118	*dardanus*	Port St Johns	03-Dec-12	G.G. Neef
G109	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G110	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G135	*dardanus*	Louis Trichardt	22-Jan-13	G.G. Neef
G115	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G071	*dardanus*	Morgans Bay	06-Mar-11	G.G. Neef
G120	*dardanus*	Port St Johns	03-Dec-12	G.G. Neef
G063	*dardanus*	Fort Beaufort	21-Mar-11	G.G. Neef
G156	*dardanus*	Louis Trichardt	23-Jan-13	G.G. Neef
G038	*dardanus*	Grahamstown	20-Feb-11	G.G. Neef
G256	*dardanus*	Langeni	04-Mar-13	G.G. Neef
G074	*dardanus*	Morgans Bay	06-Mar-11	G.G. Neef
G017	*dardanus*	Alexandria	13-Feb-11	G.G. Neef
G018	*dardanus*	Alexandria	13-Feb-11	G.P. Keevey
G019	*dardanus*	Alexandria	13-Feb-11	G.G. Neef
G020	*dardanus*	Alexandria	13-Feb-11	G.P. Keevey
G021	*dardanus*	Alexandria	19-Feb-11	G.G. Neef
G022	*dardanus*	Alexandria	19-Feb-11	G.G. Neef
G073	*dardanus*	Morgans Bay	06-Mar-11	G.G. Neef
G072	*dardanus*	Morgans Bay	06-Mar-11	G.G. Neef
G122	*dardanus*	Port St Johns	03-Dec-12	G.G. Neef
G031	*dardanus*	Grahamstown	20-Feb-11	G.G. Neef
G114	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G257	*dardanus*	Langeni	04-Mar-13	G.G. Neef
G095	*dardanus*	Pirrie Forest	24-Feb-12	J. Nicolau
G099	*dardanus*	Pirrie Forest	24-Feb-12	J. Nicolau
G264	*dardanus*	Baziya	06-Mar-13	G.G. Neef
G100	*dardanus*	Pirrie Forest	24-Feb-12	J. Nicolau
G101	*dardanus*	Pirrie Forest	24-Feb-12	J. Nicolau
G096	*dardanus*	Pirrie Forest	24-Feb-12	J. Nicolau
G098	*dardanus*	Pirrie Forest	24-Feb-12	J. Nicolau

Sample #	Taxon	Locality	Collection Date	Collector
G111	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G112	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G113	*dardanus*	Dwesa	04-Mar-12	G.G. Neef
G097	*dardanus*	Pirrie Forest	24-Feb-12	J. Nicolau
G161	*demodocus*	Entabeni	24-Jan-13	G.G. Neef
G023	*demodocus*	Alexandria	19-Feb-11	G.G. Neef
G016	*demodocus*	Alexandria	13-Feb-11	G.P. Keevey
G015	*demodocus*	Alexandria	13-Feb-11	G.P. Keevey
G006	*demodocus*	Summerset East	12-Feb-11	G.G. Neef
G144	*demodocus*	Louis Trichardt	23-Jan-13	G.G. Neef
G167	*demodocus*	Tzaneen	25-Jan-13	G.G. Neef
G168	*demodocus*	Tzaneen	25-Jan-13	G.G. Neef
G032	*demodocus*	Grahamstown	20-Feb-11	G.G. Neef
G005	*demodocus*	Summerset East	12-Feb-11	G.G. Neef
G007	*demodocus*	Summerset East	12-Feb-11	G.P. Keevey
G012	*demodocus*	Alexandria	13-Feb-11	G.G. Neef
G013	*demodocus*	Alexandria	13-Feb-11	G.G. Neef
G028	*demodocus*	Grahamstown	20-Feb-11	G.G. Neef
G035	*demodocus*	Grahamstown	20-Feb-11	G.G. Neef
G034	*demodocus*	Grahamstown	20-Feb-11	G.G. Neef
G033	*demodocus*	Grahamstown	20-Feb-11	G.G. Neef
G139	*demodocus*	Louis Trichardt	22-Jan-13	G.G. Neef
G138	*demodocus*	Louis Trichardt	22-Jan-13	G.G. Neef
G137	*demodocus*	Louis Trichardt	22-Jan-13	G.G. Neef
G136	*demodocus*	Louis Trichardt	22-Jan-13	G.G. Neef
G026	*demodocus*	Alexandria	19-Feb-11	G.G. Neef
G207	*demodocus*	Ngome	01-Feb-13	G.G. Neef
G039	*demodocus*	Eshowe	07-Feb-13	G.G. Neef
G220	*demodocus*	Eshowe	03-Feb-13	G.G. Neef
G211	*demodocus*	Ngome	01-Feb-13	G.G. Neef
G210	*demodocus*	Ngome	01-Feb-13	G.G. Neef
G208	*demodocus*	Ngome	01-Feb-13	G.G. Neef
G233	*demodocus*	Nkandla	05-Feb-13	G.G. Neef
G206	*demodocus*	Lydenburg	30-Jan-13	G.G. Neef
G209	*demodocus*	Ngome	01-Feb-13	G.G. Neef
G219	*euphranor*	Ngome	02-Feb-13	G.G. Neef
G218	*euphranor*	Ngome	02-Feb-13	G.G. Neef
G237	*nireus*	Nkandla	05-Feb-13	G.G. Neef
G238	*nireus*	Nkandla	05-Feb-13	G.G. Neef
G200	*nireus*	Lydenburg	29-Jan-13	G.G. Neef
G133	*nireus*	Louis Trichardt	22-Jan-13	G.G. Neef
G234	*nireus*	Nkandla	05-Feb-13	G.G. Neef
G212	*nireus*	Ngome	01-Feb-13	G.G. Neef
G213	*nireus*	Ngome	01-Feb-13	G.G. Neef
G235	*nireus*	Nkandla	05-Feb-13	G.G. Neef
G221	*nireus*	Eshowe	04-Feb-13	G.G. Neef
G216	*nireus*	Ngome	01-Feb-13	G.G. Neef
G214	*nireus*	Ngome	01-Feb-13	G.G. Neef
G236	*nireus*	Nkandla	05-Feb-13	G.G. Neef
G215	*nireus*	Ngome	01-Feb-13	G.G. Neef
G134	*nireus*	Louis Trichardt	22-Jan-13	G.G. Neef
G011	*nireus*	Alexandria	13-Feb-11	G.G. Neef
G151	*nireus*	Louis Trichardt	23-Jan-13	G.G. Neef

Sample #	Taxon	Locality	Collection Date	Collector
G152	*nireus*	Louis Trichardt	23-Jan-13	G.G. Neef
G150	*nireus*	Louis Trichardt	23-Jan-13	G.G. Neef
G188	*nireus*	Haenertsburg	26-Jan-13	G.G. Neef
G002	*nireus*	Summerset East	12-Feb-11	G.G. Neef
G003	*nireus*	Summerset East	12-Feb-11	G.G. Neef
G201	*nireus*	Lydenburg	29-Jan-13	G.G. Neef
G036	*nireus*	Grahamstown	20-Feb-11	G.G. Neef
G253	*nireus*	Nottingham Road	12-Feb-13	G.G. Neef
G024	*nireus*	Alexandria	19-Feb-11	G.G. Neef
G010	*nireus*	Alexandria	13-Feb-11	G.G. Neef
G004	*nireus*	Summerset East	12-Feb-11	G.G. Neef
G162	*nireus*	Louis Trichardt	24-Jan-13	G.G. Neef
G255	*nireus*	Nottingham Road	12-Feb-13	G.G. Neef
G037	*nireus*	Grahamstown	20-Feb-11	G.G. Neef
G014	*nireus*	Alexandria	13-Feb-11	G.G. Neef
G039	*nireus*	Grahamstown	20-Feb-11	G.G. Neef
G199	*nireus*	Lydenburg	29-Jan-13	G.G. Neef
G029	*nireus*	Grahamstown	20-Feb-11	G.G. Neef
G163	*nireus*	Louis Trichardt	24-Jan-13	G.G. Neef
G001	*nireus*	Summerset East	12-Feb-11	G.G.Neef
G254	*nireus*	Nottingham Road	12-Feb-13	G.G. Neef
G030	*nireus*	Grahamstown	20-Feb-11	G.G. Neef

Appendix 2

Papilio alignments of COI, ITS and the combined data in nexus format on enclosed CD.