Genetic diversity and interspecies hybridization in *Cossypha* robin-chats

---

Naadhirah Munshi

Student number: 540305

Supervisor: Dr Jean Mollett

Co-supervisor: Prof. Craig Symes

A Dissertation submitted to the Faculty of Science
University of the Witwatersrand
Johannesburg, South Africa
in fulfilment of the requirements for the degree of
Master of Science.

June 2017

DECLARATION

I, Naadhirah Munshi (Student number: 540305), am a student registered for Master of Science (Dissertation) in the academic year 2015 and 2016.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise.
- I have not submitted this work before for any other degree or examination at any other University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: Munshi **Date:** 31 May 2017

STRUCTURE OF THE DISSERTATION

This research dissertation is made up of four chapters. Chapter one presents an introduction to the study, together with a review of the literature that served to inform the study. In chapter two the methods used, i.e. *CO1* barcoding and microsatellite genotyping are discussed and explained in detail. The samples included all five species collected from different study sites where all five *Cossypha* species co-occur, as well as sites where they do not. Chapter three would conventionally be the results chapter; however, it is presented in the format of a research paper so that it can be read independently, consisting of a brief introduction, methods and materials, results and discussion. Chapter three interprets and discusses the outcome of the DNA barcoding of the five *Cossypha* species in conjunction with the microsatellite genotyping in determining the degree of relatedness between the species. The final fourth chapter discusses the findings, the limitations of the study, makes recommendations, and highlights the conclusions of the study. A single reference list has been compiled at the end of the dissertation. Because the thesis is structured in this manner there is some repetition between chapters.

RESEARCH OUTPUTS

Conference outputs

Tinderholm C, **Munshi N,** Symes C and Mollett J. 2014. Investigating interspecific hybridization and phylogenetics in *Cossypha* robin-chats. South African Genetics Society (SAGS) and South African Society for Bioinformatics (SASBi) Joint Congress. 23-26 September 2014. Kwalata Game Ranch. Dinokeng Game Reserve. *Poster presentation*.

Munshi N, Symes C and Mollett J. 2014. Investigating interspecies hybridization in *Cossypha* robin-chats using microsatellites. University of the Witwatersrand Molecular Biosciences Research Trust (MBRT) Research Day. Wits Club. 4 December 2014. *Poster presentation.*

Munshi N, Symes C and Mollett J. 2015. Genetic diversity and interspecies hybridization in *Cossypha* robin-chats. University of the Witwatersrand Molecular Biosciences Research Trust (MBRT) Research Day. Wits Club. 3 December 2015. *Poster presentation.*

Munshi N, Symes C and Mollett J. 2016. Genetic diversity and interspecies hybridization in *Cossypha* robin-chats. SA Genetics Society (SAGS) and SA Society for Bioinformatics (SASBi) Joint Congress. 20-23 September 2016, Nelson Mandela School of Medicine, University of KwaZulu-Natal, Durban. *Poster presentation*.

Munshi N, Symes C and Mollett J. 2016. Genetic diversity and interspecies hybridization in *Cossypha* robin-chats. Oral presentation. 7th Annual Oppenheimer De Beers Group Research Conference. 18-19 October 2016. De Beers Corporate Headquarters, Johannesburg. *Oral presentation - Won Best Presentation Award*.

Munshi N, Symes C and Mollett J. 2016. Genetic diversity and interspecies hybridization in *Cossypha* robin-chats. University of the Witwatersrand Molecular Biosciences Research Trust (MBRT) Research Day. Wits Club. 8 December 2015. *Poster presentation – Won 3rd place Best Poster Award*

ABSTRACT

Southern Africa boasts a high avian diversity with five *Cossypha* (robin-chat) species (*C. heuglini, C. caffra, C. humeralis, C. natalensis, C. dichroa*) distributed at varying levels of sympatry and allopatry. Due to the effects of global anthropogenic change many species which were once ecologically separated may now overlap, leading to possible genetic introgression and hybridization. This project investigates the genetic diversity and degree of relatedness between the five *Cossypha* robin-chat species that occur in South Africa. Genomic DNA was extracted from blood of all five species (n=92 individuals) using the standard phenol:chloroform extraction method. Mitochondrial and nuclear markers were analyzed using Likelihood and Bayesian methods to establish phylogenetic relationships and to determine speciation patterns. MtDNA barcoding using the cytochrome c oxidase subunit 1 (*CO1*) gene was used to assign individuals to species. The construction of a neighbour-joining and a maximum likelihood tree provided graphic representations of the pattern of divergences between the five *Cossypha* species. Individuals from a species clustered together with strong bootstrap values. These procedures were accomplished using MEGA software. PopART was used to construct a minimum spanning network. This network illustrated similarity between the five species with regards to the *CO1* barcode. Only seven of thirteen novel microsatellite markers were able to cross amplify in all five species. The Bayesian clustering analysis using the statistical programme STRUCTURE identified three genetic clusters (K=3) with the three distinct species being *C. dichroa, C. natalensis,* and *C. caffra*. *Cossypha heuglini* cluster amongst *C. dichroa, C. natalensis,* and *C. caffra,* while *C. humeralis* clusters amongst *C. natalensis*. Despite the hybridization events recorded between *C. dichroa* and *C. natalensis* these two species do not appear to be each other's closest relatives according to microsatellite and mtDNA analysis. The hybridization events indicate their ability to overcome reproductive isolation mechanisms such as vocalisations.

DEDICATION

This dissertation is dedicated to Uzair, an irreplaceable person in my life, who supported me through this journey but sadly won't be there till the end. You'll always be in my heart and in my thoughts.

ACKNOWLEDGEMENTS

I would like to acknowledge and extend my heartfelt gratitude to the following individuals and institutions:

My supervisor Dr Jean Mollett for the continuous support and encouragement, and for guiding me throughout this journey. I sincerely thank you for having the patience and understanding you have extended to me, both inside the laboratory and beyond.

My co-supervisor Prof. Craig Symes from the School of Animal, Plant and Environmental Sciences at the University of the Witwatersrand for the collection of blood samples. I am truly grateful for all that you have taught me.

My colleagues and friends in the Gate House 700 lab for all the laughs and assistance throughout these two years.

Most importantly my parents and sister for their unfaltering love, support, and constant encouragement. I am truly grateful. You were there to share my joys when experiments worked and most importantly my tears when they didn't. I would not be where I am today if it weren't for you.

My friend Claire Tinderholm for being there with me every step of the way regardless of the tantrums I threw. This journey would not have been the same without you.

Thank you to Greg Davies of the Ornithology Section, Ditsong National Museum of Natural History, Pretoria, South Africa for assistance with blood sample collection.

Lastly, the University of the Witwatersrand and the National Research Foundation for financial support.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AMOVA	Analysis of Molecular Variance
Bp	Base pairs
CO1	Cytochrome c oxidase 1
ddNTPs	Dideoxynucleotide triphosphates
DNA	Deoxyribose Nucleic Acids
dNTPs	Deoxynucleotide triphosphates
Fst	Fixation index
K2P	Kimura-2-Parameter
ML	Maximum likelihood
MtDNA	Mitochondrial DNA
NJ	Neighbour-Joining
PCR	Polymerase Chain Reaction
rRNA	ribosomal Ribonucleic Acid
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeats
STR	Short tandem repeats

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1. Selection pressures

Evolutionary biologists and ecologists have long been faced with predicting the effects of environmental change on the dynamics of biodiversity (McKinney & Drake, 1998). Climate change has fast become one of the major contributing factors which drive alterations in species and their habitats worldwide (Houghton *et al*., 2001). The selection of traits within a species can be influenced by natural pressures such as those presented by climate, natural disasters, or human-induced pressures such as land use changes, urbanization or even hunting patterns (Jones *et al*., 2001). These pressures could be a potential source of directional selection on traits that are important for fitness (Davis *et al*., 2005). Species faced with any one of these pressures could respond in three basic ways. Firstly, they could avoid the changes by moving to suitable habitats elsewhere. Secondly, they could remain in their changing environment and adjust to the conditions by means of phenotypic plasticity without any changes to their genetic make-up. Lastly, they could adapt to the changed environment by means of genetic changes through the process of evolution (Davis *et al*., 2005).

The option of evading the environment will essentially mean the extinction of the species in that particular habitat and the persistence of it elsewhere, whereas adaptation to the environment could prevent local extinction (Holt, 1990). Climate change has a considerable impact on phylogeography and genetics within and among closely related species of animals and plants (Hewitt, 1993). It can lead to previously isolated species or populations being in contact which could lead to hybridization in some cases and to speciation in other cases. Selection and divergence is thought to be more common in smaller populations that receive occasional immigrants, showing that peripheral populations play an important role in speciation and divergence (Hewitt, 1993).

1.2. Speciation and hybridization of species

Speciation is the evolutionary divergence of two or more populations (Maynard, 1996; Sobel *et al*., 2010). The type of speciation depends on the ecology of these populations with the two most common being allopatric and sympatric speciation (Maynard, 1996). Allopatric

speciation occurs when populations are geographically separated. Sympatric speciation occurs when individuals of different populations co-occur within range of one another (Coyne & Orr, 1997; Grant & Grant, 1997).

Hybridization between genetically distinct populations may be driven in nature by events such as climate change or land conversion which leads to previously isolated species coming in contact (Rhymer & Simberloff, 1996). The process of hybridization contributes significantly to speciation and can therefore contribute to evolution (Gholamhosseini *et al*., 2013). The occurrence of hybridization is hypothesized to be higher in sympatric zones where different populations of closely related species overlap (Grant & Grant, 1997). Natural hybridization is defined as the secondary contact between populations which have evolved separately and leads to gene exchange between populations (Mallet, 2005; Genovart, 2008). Hybrid speciation was thought to be rare in animals; however, over the last decade with the help of improved molecular techniques and research into this phenomenon the occurrence of hybridization is better documented (Brelsford *et al.*, 2011). Hybridization is relatively common in birds with approximately 1 in 10 species recorded to hybridize in nature (Grant & Grant, 1992). Hybridization in avian conservation is important when endangered and rare species are concerned. As species become rare there is generally an increase in cross mating with members of closely related species (Ma & Lambert, 1997). Hybridization in many species has been characterized as having a dual effect on biodiversity (Arnold *et al.*, 2006; Seehausen *et al*., 2008). Firstly, it plays a role in increasing species numbers by promoting evolutionary potential or increasing the genetic variation within the species. Alternatively, hybridization can lead to a decrease in the species numbers due to a decline in the parental species and the restriction of speciation (Felsenstein, 1981; Arnold, 1997).

Interspecific hybridization occurs when two species within the same genus mate (Grant & Grant, 1992). Hybridization studies are important in avian research because they provide information about evolutionary processes and insight into the development of new species or loss of species (Väli *et al*., 2010). The result of repeated backcrossing between an interspecies hybrid and the parental species is termed ‘introgressive hybridization’ which results in the movement of genes from one species into the gene pool of another species (Allendorf *et al*., 2001). Introgressive hybridization can only occur when individuals of the F1 generation from a hybridization event are fertile. When hybridization occurs naturally it generally results in positive evolutionary consequences due to the promotion of introgression of novel and potentially adaptive genotypes (Arnold *et al*., 1999). Hybridization in some cases may not be

the result of natural events as mentioned; however, aviculture or forced breeding has become a common practice (Ottenburghs *et al.*, 2016a). Selective breeding for specific traits to enhance hybrid vigour is implemented for many different species (Ottenburghs *et al.*, 2016a). For example, waterfowl have a high incidence of hybridization in nature and have been identified as a suitable bird group to infer evolutionary relationships and hybrid vigour (Ottenburghs *et al.*, 2016b). The brown teal (*Anas auklandica chlorotis*) was one of the most threatened of New Zealand's endemic waterfowl (Hayes & Williams, 1982). These waterfowl were bred in captivity for the re-establishment of the species (Hayes & Williams, 1982). The brown teal were kept in environments where they could easily choose to mate with grey ducks (*Anas superciliosa superciliosa*) or mallards (*Anas platyrhynchos*) to enhance the egg laying capacity of the females (Hayes & Williams, 1982). Natural and anthropogenic hybridization have both negative and positive effects on the genetic integrity of many species and subspecies worldwide (Rhymer & Simberloff, 1996). Species integrity will depend on the gene flow between the hybridizing species, and natural selection in the species involved and the resulting hybrid offspring (Genovart, 2008). The example mentioned above with regards to the grey ducks and mallards serves as an example of anthropogenic hybridization with a positive effect on the species as the egg laying capacity is enhanced, promoting the survival of the species (Hayes & Williams, 1982). In contrast to this, the introduction of the ruddy duck (*Oxyura jamaicensis*) in Great Britain in the 20th century led to the spread of this species throughout Europe, resulting in the hybridization of ruddy duck with the endangered white-headed duck (*Oxyura leucocephala*) leading to a decline in the endangered species (Munoz-Fuentes *et al*., 2007). Ruddy ducks were then restricted to certain areas or eradicated from certain areas to preserve the endangered white-headed duck (Muńoz-Fuentes *et al*., 2007).

1.3. Degree of relatedness between species

Considering the positive and negative effects of hybridization and introgression, and that ecosystems and biological communities could be altered, hybridization studies and studies of population structure and relatedness play an important role in informing avian conservation (Moritz, 1994). The conservation of any species aims to preserve maximum variation at all levels of biodiversity which highlights the importance of taxonomy and phylogenetics (Peterson & Eernisse, 2015). Geographic variation, genetic differentiation, and adaptations to changes in the species environment, be it natural or human-mediated, are vital aspects when assessing the ecology or population structure of related species (Randler & Bogner, 2002).

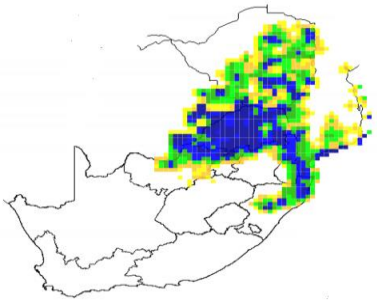
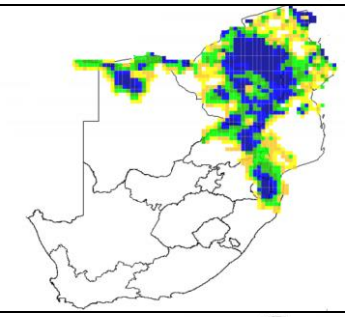
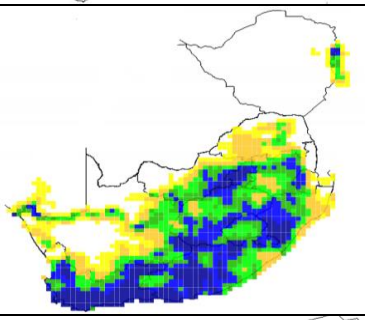
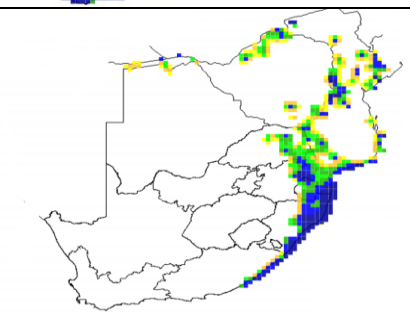
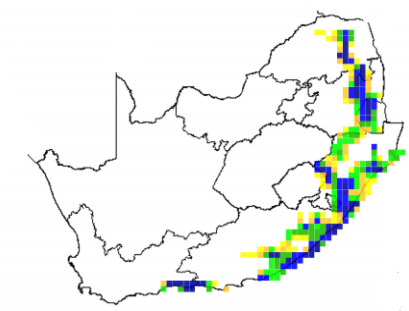
Subspecies have posed a challenge to conservation and accurate species delimitation (Zink, 2004). In some cases, the subspecies form separate phylogenetic clusters while in others their identification is misclassified (Gippoliti & Amori, 2007). The correct classification of subspecies or cryptic species is particularly relevant to birds. Phillimore and Owens (2006) illustrated that approximately thirty-six percent of defined avian subspecies from North America and Eurasia form distinct phylogenetic lineages.

Many evolutionary biologists assume sister taxa hybridize more easily than non-sister taxa because of the similar morphology as well as genetics enabling successful mating between the species and resulting in viable offspring (Randler & Bogner, 2002). Based on this assumption one would theoretically expect a hybrid to be the result of the two most closely related species (Randler & Bogner, 2002). *Cossypha* robin-chats support this very assumption with the emergence of a potential hybrid identified in the Eastern Cape region (Clancey, 1982; Davies *et al*., 2011).

1.4. *Cossypha* species and possible hybrids in South Africa

Southern Africa boasts a wide diversity of avian species (Hockey *et al*., 2005). Five *Cossypha* species are distributed across the region occurring at varying levels of sympatry and allopatry (Harrison *et al*., 1997; Hockey *et al*., 2005). They are relatively easily identified by their distinct colouring, unique vocalisations, the habitats in which they generally occur, and their distribution patterns across southern Africa (Table 1). To date there has been little debate regarding the identity of these taxa as distinct species.

Table 1: The distribution and habitat of the five *Cossypha* species found in southern Africa. Maps from Brookes (2015).

Species	Identifying characteristics	Habitat	Distribution Incidence of sighting (%) 0.25 – 0.73 0.73 – 1.73 1.73 – 3.45 3.45 – 6.2 6.2 - 11.39 11.39 +
C. humeralis (White-throated robin-chat)  Trevor Hardaker	White wing bar White chin and chest region White supercilium Rufous tail with black tips	Thickets, thorny scrub, edges of dune forest, also wooded suburbs and farm gardens.	
C. heuglini (White-browed robin-chat)  Stephen Davis	Conspicuous white supercilium	Riverine forest, shady trees and shrubs.	
C. caffra (Cape robin-chat)  Trevor Hardaker	White supercilium Pale grey lower breast	Forest edge, bushveld, scrubs, fynbos, gardens and parks.	
C. natalensis (Red-capped robin-chat)  Johan van Rensburg	Blue-grey back Orange face	Evergreen forest, riparian thickets in bushveld, dune forest, suburban gardens.	
C. dichroa (Chorister robin-chat)  Dennis Heidrich	Uniform black upper parts Lack of white supercilium	Afromontane evergreen forest.	

In 1909 Dr Jan Gunning described a new species of robin-chat collected in the Eastern Cape and named it *Cossypha haagneri* (Gunning, 1909). Many years later Phillip Clancey re-examined the inferred new species found in the Eastern Cape region and hypothesized that it was possibly a hybrid because it resembled an intermediate between *C. dichroa* and *C. natalensis* (Clancey, 1982). The hybrids were subsequently seen to have variable phenotypes (Figure 1).

Figure 1: Inferred *C. dichroa* x *C. natalensis* hybrids. Inferred hybrids have a variable phenotype; general characteristics include, blue-grey crown and unmottled blue-grey back similar to *C. dichroa*, but an all-orange face (or dusky cheeks) similar to *C. natalensis* (Davies *et al*., 2011). Photo credit: Craig Symes.

Recently the appearance of phenotypic hybrids between *C. dichroa* and *C. natalensis* (similar to those described above) in the Eastern Cape and KwaZulu-Natal province, suggests that hybridization between *C. natalensis* and *C. dichroa* may be more common than previously thought (Davies *et al.*, 2011). These inferred hybrids displayed a dark blue-grey crown, nape and forehead and their faces appeared orange with a lack of duskiness in their cheeks (Figure 1) (Davies *et al*., 2011). The weight of these inferred hybrids resembled that of *C. natalensis* although their tail and wing length more closely resembled *C. dichroa* (Clancey, 1982).

The proposed hybridization between *C. dichroa*, endemic to South Africa and *C. natalensis*, with an extended range in southern Africa, is an example of interspecies hybridization due to previously ecologically segregated species being brought into contact based on phenotypic hybrids being identified in regions where these two species co-occur (Davies *et al*., 2011). In the case of the *Cossypha* robin-chats the inferred hybrids coexist with both the parental species and the hybrids are thought to be the result of the sympatric occurrence of the parental species (Roberts, 1914; Grant & Grant, 1997).

Cossypha caffra, *C. humeralis*, and *C. heuglini* differ in their colouration while sharing the similar stripe across their head area, whereas *C. heuglini* differ from *C. dichroa* by the white stripe across the head area. *Cossypha dichroa* has a blue-grey crown and a plain blue-grey back, while *C. natalensis* has an olive or reddish-brown crown and a heavily orange and blue mottled back (Davies *et al*., 2011; Symes, 2011). The inferred hybrids vary phenotypically which suggests possible backcrossing. The degree of relatedness between these five species could be inferred based on their phenotype; however, this may not be accurate as species may appear to be similar in phenotype solely based on convergent evolutionary processes and not genetics (Emery & Clayton, 2004). The potential hybridization event proposed by Davies *et al.* (2011) may suggest that *C. dichroa* and *C. natalensis* are each other's closest relatives although, this has not been confirmed. *Cossypha natalensis* and *C. dichroa* may be the two species which are able to overcome reproductive isolating mechanisms such as vocalisation while the other *Cossypha* species are not able to do so regardless of their degree of relatedness. Each species has a unique call although three of the species namely, *C. dichroa*, *C. natalensis* and *C. humeralis*, have the ability to mimic other bird species (Ferguson *et al.*, 2002); this could play a major role in overcoming reproductive isolation created through species-specific vocalisations. Therefore, analysis of these five species using different molecular techniques could allow the degree of relatedness between these five species to be more accurately determined.

1.5. Genetic markers in population genetic studies

The identification of a species is generally based on morphological characteristics, which according to Mallet (2008) is a weak diagnostic tool especially when hybridization is suspected. Identifying hybrids relying solely on phenotypic characteristics has proven to be a challenge in many cases. The probability of identifying backcrosses which may resemble F1 generation hybrids or the parental species is low or impossible in some cases based only on phenotypes; therefore, molecular methods are applied together with the morphological identification of hybrids and closely related species (Avise, 2004). An example of this is the identification of hybrids resulting from the hybridization of glossy ibis (*Plegadis falcinellus*) and white-faced ibis (*Plegadis chihi*). Both species are found to co-exist with the hybrid offspring in Weld and Boulder counties in Colorado (Leukering, 2008). F1 generation hybrids are often misidentified as glossy ibis due to the blueish upper and lower facial borders which is characteristic of the glossy ibis (Arterburn and Grzybowski, 2003; Semo, 2007; Leukering, 2008). On closer examination, reddish irides are characteristic of white-faced ibis as well as

grey bill colouration, which is often overlooked as many hybrids have two-toned bills with brown (characteristic of glossy ibis) overshadowing the grey colouration. Further backcrossing between the F1 generation hybrid and either of the parental species is most likely to result in an appearance with greater similarity to the parent species making the identification impossible. Genetic evidence to support theories of hybridization and to investigate the degree of relatedness between species based on morphological characteristics can be obtained with the use of DNA barcoding using mitochondrial DNA (mtDNA), single nucleotide polymorphisms (SNPs), and microsatellite genotyping (Schlotterer, 2004).

The use of nucleotide sequence differences in a single gene to investigate evolutionary relationships were first utilised in 1977 by Carl Woese (Balch *et al.*, 1977). Sequence differences in a conserved gene such as ribosomal RNA (rRNA) were initially used to infer phylogenetic relationships (Woese, 2000). It was later discovered that closely related organisms could not be differentiated using rRNA due to its slow evolving rate (Woese, 2000). Genes that evolve faster were then explored to determine if these rapidly evolving genes, such as mtDNA, were able to differentiate between closely related species (Brown *et al.*, 1979). MtDNA barcoding has since become a widely employed marker in phylogenetic studies as it has a higher mutation rate than nuclear DNA, allowing for the accumulation of differences between closely related species (Brown *et al.*, 1979).

DNA barcoding proposes that individuals can be easily and rapidly assigned to a species using a standardized DNA sequence from the mitochondrial genome with regards to vertebrates (Hebert *et al.*, 2003). The high mutation rate of mtDNA makes it a popular marker for biodiversity studies. It has been extensively used to investigate intraspecific and interspecific evolutionary relationships and for disentangling rapid, recent speciation events in vertebrate phylogenetic studies (Saetre *et al.*, 2001). DNA barcodes may be used for identification of morphologically similar species due to this method being based on the concept that every species has a unique genetic identity (Hebert *et al.*, 2004). The method is also useful in hybridization studies because of maternal inheritance patterns which help to identify which species was the maternal parent. DNA barcoding is based on the effective amplification and sequencing of a 648 base pair (bp) region of the mitochondrial cytochrome *c* oxidase 1 (*CO1*) gene (Hebert *et al.*, 2004). Genetic introgression and sharing of DNA barcode haplotypes often occurs in hybridization (Toews *et al.*, 2011). This is a result of backcrossing of the fertile first generation hybrids which have the mtDNA of their maternal parent, leading to the replacement of haplotypes in the different species or the transfer of alleles from one species to the other

(Rheindt & Edwards, 2011). This leads to similar barcodes amongst different species due to the introgression. An investigation performed on 643 previously recognised species of birds of North America demonstrated the effectiveness of DNA barcoding (Hebert *et al*., 2004; Kerr *et al*., 2007). In the study, 94% of the bird species possessed a unique monophyletic *CO1* cluster (Hebert *et al*., 2004; Kerr *et al*., 2007). The 6% which did not have a unique barcode were classified either as being mistakenly classified as a separate species, the species being closely related and potentially hybridizing, or the species has lost its identity by secondary contact (Hebert *et al*., 2004; Kerr *et al*., 2007). It is important to note that the appropriate set of markers is very important to help identify hybrids and genetic introgression in avian populations, and gene sequence results can be verified using microsatellite genotyping (Rheindt & Edwards, 2011).

Hybridization and parentage or relatedness is commonly identified by microsatellite genotyping and has become a popular choice for hybrid identification in the past two decades (Lu *et al*., 2001; Crochet *et al*., 2003; Toews *et al*., 2011; Coetzer *et al*., 2015; Germain-Aubrey *et al*., 2016; Samani *et al*., 2016). The efficiency of this identification with respect to ecology is based on the potential of microsatellites to determine relatedness of individuals (Selkoe & Toonen, 2006). Microsatellites also known as simple sequence repeats (SSR), or short tandem repeats (STR), are tandem repeats of one to six nucleotides in length (Li *et al*., 2002). Microsatellite loci can vary in repeat length with the most common choices for molecular genetic studies being di-, tri- or tetranucleotides (Jarne & Logoda, 1996).

The mutation rate of microsatellites differs among loci, alleles, and species (Ellegran, 2000). It depends on its intrinsic features such as the number of repeated units, the length in base pairs, and the repeated motif (Webster *et al*., 2002). Microsatellite markers with a higher number of repeats have a higher mutation rate due to the increased probability of slippage during replication (Webster *et al*., 2002). The mechanism of slippage assumes that during replication the nascent and template strand realign out of register (Schlotterer, 2000). When the DNA synthesis continues unabated the repeat number of the microsatellite is altered (Schlotterer, 2000).

The mutability of a microsatellite is also influenced by the flanking region within the primer region which is described as a single copy DNA sequence immediately upstream or downstream of the microsatellite loci (Buschiazzo & Gemmell, 2006). The microsatellite flanking regions are generally conserved across individuals of the same species and

occasionally of different species allowing microsatellite loci to be identified by the sequence of the flanking regions (Selkoe & Toonen, 2006). Differences in alleles are not only determined by the variation in repeat number, the substitutions, insertions and deletions in the flanking regions are also relevant (Viard *et al*., 1998). The identification of SNPs in the flanking regions may also be useful in identifying hybrids (Selkoe & Toonen, 2006). The use of SNPs in hybridization studies has increased in recent years (Väli *et al*., 2010). In some cases, a single SNP with two fixed alleles is sufficient to make a conclusive decision in assigning individuals to a particular species (Avise, 2004).

A combined analysis of mtDNA and microsatellite loci among lake whitefish (*Coregonus clupeaformis*) populations was used to assess the congruence between the two different types of markers in defining patterns of genetic structuring, introgressive hybridization, and for inferring population origins (Lu *et al*., 2001). The microsatellites allowed the identification of lineage specific allelic size groups which allowed better insight into the introgressive hybridization between the species (Lu *et al*., 2001). In another example, Townsend's warblers (*Dendroica townsendi*) and the black-throated green warblers (*Dendroica virens*) are phenotypically and genetically distinct species which occur in western and eastern North America, respectively (Toews *et al*., 2011). These two species co-occur in the Rocky Mountains of British Columbia and putative phenotypic hybrids have been identified (Toews *et al*., 2011). Evidence of the interbreeding between these species and determining whether the hybridization was a regular occurrence was based on morphology, plumage, mtDNA barcoding using the *CO1* gene, as well as nuclear microsatellite markers (Toews *et al*., 2011). The study showed extensive hybridization between these species (Toews *et al*., 2011). The confirmation of hybridization in the area of co-occurrence provides a system to study patterns of speciation and reproductive isolation (Toews *et al*., 2011). On the Chatham Islands, off the coast of New Zealand, the use of microsatellites has been implemented to reveal the hybridization between the endangered black robin (*Petroica traversi*) and the tomtit (*Petroica macrocephala*) (Ma and Lambert, 1997). Evidence of hybridization has had important implications for cross-fostering programs, which involves the separation of offspring from their biological parents at birth and allowing them to be raised by surrogates. Cross-fostering programs are being implemented in conservation programs worldwide (Ma and Lambert, 1997; Vali *et al.*, 2010).

According to Beresford (2003) many genera of African passerines are weakly diagnosed and remain untested both at species level and in cases of hybridization within the species. DNA

barcoding and microsatellite genotyping have been successful in identifying hybrids as well as determining the degree of relatedness between species (Beresford, 2003). Using molecular techniques such as DNA barcoding and microsatellite genotyping will help determine the population structure of *Cossypha* robin-chats and explore the genetic diversity in the five species found in South Africa (Beresford, 2003). The understanding of the population structure or the genetic structure of these species will allow further insight into the processes of adaptation, speciation, and interactions between these species in areas of allopatry and sympatry.

1.6. Hypothesis

The main objective of this study was to investigate genetic diversity and relatedness, including hybridization, in five *Cossypha* (robin-chat) species using genetic markers. Based on the hybridization event between *C. dichroa* and *C. natalensis*, as proposed by Davies *et al.* (2011), I hypothesize that, 1) *C. dichroa* and *C. natalensis* are the two most closely related species amongst the five *Cossypha* robin-chat species found in southern Africa and, 2) the five *Cossypha* species are genetically distinct from each other. DNA barcoding and microsatellite genotyping were used to test the two hypotheses and provide some insight into the proposed idea that identified hybrids are the result of matings between *C. dichroa* and *C. natalensis.* Depending on the level of relatedness between these species a greater understanding of the likelihood of hybridization between co-occurring species in the genus *Cossypha*, will be obtained.

Chapter 2

Methods and materials for DNA barcoding and microsatellite genotyping

---

2.1. Introduction

The inclusion of molecular data has enhanced the understanding of evolutionary history of populations or species (Zink & Barrowclough, 2008). The use of mitochondrial DNA (mtDNA) has been a popular choice of marker for the last two decades in phylogeographic studies and species identification (Bilgin *et al*., 2016). MtDNA has a rapid mutation rate of approximately 6.2×10^{-8} per site allowing an increased chance of accumulating variations (Avise, 1994; Haag-Liautard *et al*., 2008; Sturge *et al*., 2016). The realization that mtDNA evolved rapidly coupled with its ability to be potentially informative for taxonomic purposes rendered this marker an important tool for evolutionary research (Zink & Barrowclough, 2008). DNA barcoding using a standardized segment of the mitochondrial genome is proposed to allow rapid assignment of an individual to a species (Hebert *et al*., 2004). This technique is based on the effective amplification and sequencing of a 648 base pair (bp) region of the mitochondrial cytochrome *c* oxidase subunit 1 (*CO1*) gene (Hebert *et al*., 2004). DNA barcoding can be used in hybridization studies due to the maternal inheritance patterns which allow the identification of the maternal parent. Paternal parentage cannot be determined using mtDNA and therefore nuclear DNA analysis has become commonly used alongside DNA barcoding studies (Zink & Barrowclough, 2008).

The advantage of coupling nuclear DNA and DNA barcoding analysis is the potential to generate multiple gene estimates of evolutionary patterns (Zink & Barrowclough, 2008). The efficiency of microsatellites for identification with respect to ecology is based on the potential of microsatellites to provide estimates of migration rates and to determine relatedness of individuals (Selkoe & Toonen, 2006). The mutability of microsatellites differs among loci and species (Ellegran, 2000), and is affected by the flanking regions which are conserved amongst individuals of the same species (Buschiazzo & Gemmell, 2006). Therefore, the identification of single nucleotide polymorphisms (SNPs), insertions, deletions or substitutions found in the

flanking regions are as relevant as the allele frequencies, although this would require sequencing of the amplified microsatellite regions (Viard *et al*., 1998).

Based on the success of DNA barcoding and microsatellites in identifying hybrids and determining the degree of relatedness between species as illustrated by Lu *et al.* (2001) and Toews *et al.* (2011) amongst others, these techniques will be used to determine the population structure of the *Cossypha* robin-chats in this study. In light of the proposed phenotypic hybridization observed between *C. dichroa* and *C. natalensis* by Davies *et al.* (2011) we hypothesize that these two species are more closely related to each other than to *C. caffra*, *C. heuglini* and *C. humeralis* and that the five species are genetically distinct. This chapter will deal solely with the molecular techniques used in this study.

2.2. Methods and materials

2.2.1. Collection of blood samples

Blood samples from areas in South Africa where the five species co-occur sympatrically and allopatrically were collected (Table 2). A total of 92 blood samples were retrieved and analyzed in this study. The weight and length of the birds were recorded as well as the co-ordinates at which the birds were captured (Davies *et al*., 2011). The inferred hybrids (Safring no. BE37939 and BE37965 used in this study were caught in October 2010 (Davies *et al.*, 2011; Symes, 2011), the breeding season, at Vernon Crookes Nature Reserve (southern KwaZulu-Natal) where both species occur. The blood samples were taken from the brachial vein with the use of a 30-gauge needle to obtain a blood flow (Davies *et al*., 2011). Each blood sample was collected in a 75µl capillary tube and the samples were stored in ethanol (see Davies *et al*., 2011, where the collection procedure is explained in further detail). Ethics clearance was obtained from the University of the Witwatersrand Animal Ethics Screening Committee (permit number 2009.42/ZA).

Table 2: Gazetteer of sampling sites and number of samples collected for each species.

Sites	Co-ordinates	Altitude (m a.s.l.)	*C. dichroa*	*C. natalensis*	*C. caffra*	*C. heuglini*	*C. humeralis*
Vernon Crookes Nature Reserve, KwaZulu-Natal	30°16'28"S, 30°36'36"E	420	6	11			
New Forest, Nottingham Road, KwaZulu-Natal	29°27'50"S, 29°52'43"E	1610	5		4		
Twinstreams Education Centre, Mtunzini, KwaZulu-Natal	28°58'51"S, 31°44'09"E	20		4			
Pullen Farm, Nelspruit, Mpumalanga	25°34'22"S, 31°10'53"E	910		2		1	
Wits Rural Facility, Limpopo	24°33'07"S, 31°05'48"E	570		3		1	
Schoemansdal Environmental Education Centre, Limpopo	23°01'04"S, 29°43'32"E	980	7	18	9	6	4
Inhamitanga Forest, central Mozambique	18°09'17"S, 35°07'29"E	180		9			
Total number of samples			**18**	**47**	**13**	**8**	**4**

2.2.2. Genomic DNA extraction

Total genomic DNA extractions were performed according to the method designed by Blin and Stafford (1976). The blood samples were stored in ethanol at 4˚C. The ethanol was allowed to evaporate for the extraction. The samples were then resuspended in Queen's (Tris-EDTA, Sodium Chloride) buffer which contains n-lauroylsarcosine an anionic detergent to solubilise membranes and denature proteins as well as EDTA to protect the DNA from degradation by sequestration of divalent cations essential for DNase activity (Loparev *et al.*, 1991). Queen's buffer preserves the DNA and removes the proteins (Abd El-Aal *et al.*, 2010). Samples were then vortexed and centrifuged in order to pellet the DNA and proteins. The supernatant was discarded and replaced with STE (Sodium Chloride, Tris-Cl, EDTA) buffer. STE buffer was added to further lyse the cells along with sodium dodecyl sulfate (SDS), RNase and proteinase K (Wiegers and Hilz, 1971; Loparev *et al.*, 1991). SDS denatures proteins, proteinase K digests contaminating proteins, and RNase removes RNA. The lysate was then treated with phenol:chloroform:isoamylalcohol (25:24:1, v/v) which separates the proteins from the DNA (Loparev *et al.*, 1991). The solution was briefly vortexed to form an emulsion before

centrifugation (13 000 rpm for 8 minutes) which separates the mixture into a lower organic phase separated from the upper aqueous phase by a band of denatured protein (Loparev *et al.*, 1991). The aqueous phase containing the nucleic acids can then be recovered and the phenol:chloroform:isoamylalcohol (25:24:1, v/v) step was repeated on the remaining solution to maximise the DNA yield (Cler *et al.*, 2006). DNA was then precipitated in 95% ethanol and sodium acetate for 30 minutes at -20°C (Cler *et al.*, 2006). The precipitated DNA was collected by centrifugation (13 000 rpm for 10 minutes) and subsequently washed in 70% ethanol before resuspension in TE buffer (Loparev *et al.*, 1991). TE buffer contains Tris to buffer the pH of the solution and EDTA to sequester divalent cations thereby preventing DNA degradation (Loparev *et al.*, 1991). Samples were stored at 4°C (Loparev *et al*., 1991).

2.2.3. Quantitation

The extracted DNA was quantified with the use of a Nanodrop® ND-1000 spectrophotometer at a wavelength of 260 nm. DNA absorbs light at a wavelength of 260 nm and proteins absorb light at a wavelength of 280 nm. The NanoDrop-1000 measures absorbance at both these wavelengths to establish whether proteins are present in the sample (Cler *et al*., 2006). A purity ratio is determined by a ratio between the two wavelengths and this allows the presence of proteins as contaminants to be quantified (Cler *et al*., 2006). The NanoDrop-1000 provides the concentration as well as a purity ratio represented as $A_{260/280}$ (Cler *et al*., 2006). The purity ratio indicates a pure sample if the value obtained is between 1.8 and 2 (Cler *et al*., 2006). All samples used had a purity ratio well within the expected purity range with concentrations of 40 ng/µl or higher.

2.2.4. Gel electrophoresis of genomic DNA samples

The genomic DNA obtained from the extraction procedure was visualized using gel electrophoresis. The extracted DNA is mixed with a loading dye which allows the tracking of the sample and the visualization of the sample on the gel due to the tracking dye incorporated in the buffer (Smith, 1993). The buffer also contains glycerol or sucrose to allow the sample to gravitate to the bottom of the well of the gel (Smith, 1993). The gel was prepared with a corresponding percent agarose for the sample size. For the visualization of total genomic DNA, a 1% agarose gel was used with a TBE (Tris, Boric acid, EDTA) buffer for the purpose of this experiment (Brody & Kern, 2004). The fragments were visualized by fluorescent light which is facilitated by the GR green added to the gel; a fluorescent dye that interchelates between the DNA bases (Smith, 1993). The light emitted during the fluorescence is red-orange and this allows easy visualization on the gel (Smith, 1993).

The use of electrophoresis has proved to be beneficial with regards to identifying, purifying or separating DNA fragments (Smith, 1993). The movement of the molecules along the gel is facilitated by the electric field applied to the gel (Smith, 1993). Molecules of different sizes migrate at different rates and therefore a smaller fragment migrates faster than larger fragments (Smith, 1993). The results of the agarose gel were captured using a Geldoc system which allows the quality of the sample to be analyzed. The samples were then subjected to polymerase chain reaction (PCR).

2.2.5 Mitochondrial DNA analysis

2.2.5.1. Amplification of the *CO1* gene

The polymerase chain reaction (PCR) allows for the amplification of the *CO1* region due to the repetitive cycles of the denaturing, annealing and elongation processes (Erlich,1989). This technique allows large quantities of the target DNA to be amplified (Erlich, 1989).

PCR requires template DNA, specific primers which will amplify the region of interest, a supply of deoxyribonucleic acids (dNTPs) for the elongation of the newly synthesized DNA, *Taq* polymerase, a heat stable enzyme, water and magnesium chloride, a cofactor for the *Taq* polymerase, in a single PCR tube. The use of the *Taq* polymerase, isolated from *Thermophilus aquaticus* and a thermostable enzyme (DNA polymerase), which yields a 3′ end with an additional adenine due to its terminal transferase activity (Cha & Thilly, 1993).

The PCR was carried out according to the protocol of Hebert *et al*. (2004). The PCR primers were developed by Hebert *et al*. (2004) to amplify a region near the 5′ terminus of the *CO1* gene. The BirdF1-TTCTCCAACCACAAAGACATTGGCAC and BirdR1-ACGTGGGAGATAATTCCAAATCCTG primers amplify an approximately 750 bp region of the *CO1* gene in most bird species. An alternative reverse primer (BirdR2-CTACATGTGAGATGATTCCGAATCCAG or BirdR3-AGGAGTTTGCTAGTACGATGCC) combined with the BirdF1 forward primer was developed to amplify this region in the event that the regular primer pair failed (Hebert *et al*., 2004). BirdR1, BirdR2, and BirdR3 were tested; the best amplification was obtained with the use of BirdR1. The PCR cycle consists of 1 minute at 94°C to denature the template DNA followed by 5 cycles of 1 minute at 94°C, l.5 minutes at 45°C and 1.5 minutes at 72°C and then 30 cycles of 1 minute at 94°C, l.5 minutes at 51°C and 1.5 minutes at 72°C to amplify the desired region and finally 5 minutes at 72°C for the final extension of any partially synthesized fragments (Hebert *et al.*, 2004).

2.2.5.2. Electrophoresis of the amplified *CO1* barcode

The specificity of the PCR reaction was analyzed by gel electrophoresis (Erlich, 1989). The PCR products were run on a 1% agarose gel (100 Volts for 45 minutes) following the same procedure as described in Section 2.2.4.

2.2.5.3. *CO1* barcode sequencing

The *CO1* barcodes amplified by PCR were sent to Inqaba Biotech for forward sequencing, this entails computer automated high throughput DNA sequencing (Sanger sequencing) which is a dideoxynucleotide chain-termination sequencing method. Computer-automated high-throughput DNA sequencing involves the synthesis of numerous DNA strands complementary to the template in the presence of each of the four deoxynucleotide triphosphates (dNTPs) and dideoxynucleotide triphosphates (ddNTPs) labelled with a different fluorescence dye (Karger *et al*., 1991). DNA polymerase incorporates a ddNTP instead of dNTP terminating DNA strand elongation, eventually the reaction solution contains strands of all possible lengths ending with a labelled ddNTP (Sanger *et al*., 1977). The resultant newly synthesised strands are separated according to size by capillary electrophoresis (Karger *et al.*, 1991). The fragments are detected as they pass by a detector apparatus which uses a laser to excite the fluorophore and measures the wavelength of light emitted (Karger *et al.*, 1991). Computer software is then used to interpret the data into a DNA sequence (Karger *et al*., 1991).

2.2.5.4. Analysis of *CO1* barcode divergence

The sequenced CO1 barcode amplicons (n=92) produced clean chromatograms, peaks were evenly spaced each with one colour present for each peak and there was minimal baseline noise. These *CO1* barcode sequences were aligned with the use of ClustalW (Larkin *et al*., 2007). The sequence divergence was calculated using the Kimura-2-Parameter (K2P) distance model, a statistical method for the estimation of evolutionary distances between homologous sequences based on the number of transition and transversion substitutions (Kimura, 1980). A neighbour-joining (NJ) tree of the K2P distances was constructed (Saitou & Nei, 1987). A maximum likelihood phylogeny was then inferred using the Kimura-2-model which was selected based on a model test (Olsen *et al*., 1994). The best model (K2+G) was selected based on the Bayesian information criterion (BIC). The model with the lowest BIC is the preferred model for the inference of an ML phylogeny. The phylogenetic trees were constructed to provide a hypothesis of the evolutionary relationships between the different *Cossypha* species and any inferred hybrids. Bootstrapping statistics were calculated for the evaluation of the reliability of the inferred clades, where a high bootstrap value (>50) provides more confidence that the

branch point is correct (Felsenstein, 1985). Bearded scrub robin (*Cercotrichas quadrivirgata*) was used as the outgroup, which also belongs to Muscicapidae family but is of the Muscicapinae subfamily distinct from the African forest robin group (Sangster *et al*., 2010). These procedures were accomplished using Molecular Evolutionary Genetics Analysis (MEGA v6) software (Tamura *et al*., 2013). PopART was then used to construct a minimum spanning haplotype network using the *CO1* barcodes (Bandelt *et al*., 1999).

2.2.6. Microsatellite analysis

2.2.6.1. Amplification of microsatellite regions

PCR was carried out for the amplification of the microsatellite markers in the DNA. A total of 78 samples were used for the microsatellite analysis. Fewer samples were used for the microsatellite analysis in comparison to the DNA barcoding due to depleted blood samples of the inferred hybrids and a number of blood samples from the five *Cossypha* species. The following thermocycling profile was followed: two minutes at 95°C followed by 40 cycles of thirty seconds at 95°C, annealing temperatures according to the primers (Table 3) and one minute at 72°C, and a final extension at 72°C for two minutes.

Thirteen microsatellite loci isolated from a genomic library of *Cossypha natalensis* according to Wogan *et al*. (2015) were tested. The primers have been designed to correspond with the flanking regions of the microsatellite markers. According to the authors these microsatellite markers cross amplify in *C. caffra*. Seven of the thirteen microsatellite primer pairs cross amplified in all five species. These seven markers were then used for the study (bold in the Table 3 below). The microsatellite loci are all tetranucleotide repeats which have been found to be best when scoring alleles as these give a high degree of error free data while remaining robust enough to survive degradation (Amos *et al.*, 2006). The temperatures for the denaturing, annealing and elongation were based on each primer pair used.

Table 3: Markers and primers tested for the amplification of microsatellites in robin-chats (Wogan *et al*., 2015). Bolded markers were selected for the microsatellite analysis.

Marker	**Annealing temperature (°C)**	**Allele range**	**Forward Primer 5′-3′**	**Reverse Primer5′-3′**	**Motief**
CNA111	54-56	143-230	CTAGCTAGCAGGCTCATTCG	ATATGAGGCATGCAAGCCTG	(TCCA)10
CNA130	**52-54**	**148-180**	**GTGATTAGCAGAGTTAGCTTC**	**TCCACAGAAATCTCGAACAG**	**(TGGA)10**
CNA139	**54-56**	**317-337**	**CCTAAGTAGCTGAACATCTC**	**GACTCTAATCAAGATGAGAC**	**(TCCA)13**
CNA142	50-54	181-213	AAGCAAGGCAGGATGCTCAC	TTGTCTATGATTCTTAGCAC	(TGGA)13
CNA69	**54**	**152-198**	**CCACCTTTAATACATTTCTAGTCAGTC**	**TTGTCCTTCCAAAACCAACC**	**(TGGA)13**
CNA99	**54**	**106-137**	**GGGTTCCTGTTCCCTTCTCT**	**CCATGTCCTGTGCATCTCAA**	**(TGGA)11**
CNA109	**52**	**170-214**	**GCACATATTGCCTTACAGTG**	**AATTGCACAGGCTAATATG**	**(GATG)14**
CNA113	**56**	**108-152**	**CAGCACTCAGGCAAATGAAA**	**AGCAGCTCAGAAGGCAAAAC**	**(TGGA)14**
CNA137	56	154-182	GGGATTGTCTTCTGCACTCAG	CCTCAGTTTGATCCGTCCAC	(TGGA)8
CNA162	56	240-260	TGAAACTAAAAACACCAAGGAAA	GCAATTTGTGAGCGCAACTA	(ATGG)10
CNA180	**56**	**101-125**	**ACATCTGCAGAGCACCATTG**	**GAGCCAGGGAAGGAAGGAT**	**(ATAC)9**
CNA233	56	84-136	TTGCCATTGAATTGGGAGTT	GAGAGTCACCTGGGATGGAG	(GATG)18
CNA214	56	227-259	TATGCAGGACGTGCTTCCTAC	TCTCTGAACACCAGTAGTAG	(TCCA)11

Successful amplification was determined by gel electrophoresis as mentioned in Section 2.2.4. In order to obtain better resolution and determine any heterozygosity of the samples 3% agarose gels were used. The Geldoc system was once again used for visualisation.

2.2.6.2. Multiplex PCR

Multiplex PCR is a variant of PCR which allows two or more target sequences to be amplified including multiple pairs of primers in a single reaction (Markoulatos *et al*., 2002). This technique was first described in 1988 and has been used widely in the identification of viruses, bacteria and parasites (Chamberlain *et al*., 1988). Sample fragments were labelled with fluorescent labels on primers (Life Technologies, Inc, Johannesburg) to allow PCR reactions to be multiplexed. The seven microsatellite markers were selected due to their ability to cross amplify in all five species were fluorescently labelled (refer to Chapter 3, Table 4). This allowed the identification of PCR products from different loci with overlapping sizes. The primers were pooled in two channels according to fluorescent labels and annealing temperatures. Every sample used two channels with the first containing PCR products using primers CNA69, CNA142, and CNA130 while the second channel contained PCR products using primers CNA99, CNA109, CNA113, and CNA180.

2.2.6.3. Microsatellite allele sizing (genotyping)

Microsatellite PCR product sizes were detected in an automated DNA analyzer instrument (ABI3100) run at the Stellenbosch University, Central Analytical Facility. For the purpose of this study Genescan^{TM} 500 Liz^{TM} (Applied Biosystem Inc.) internal size standard on an ABI 3130 Genetic Analyzer and Peak Scanner Software^{TM} v1.0 was used. Peak Scanner Software^{TM} v1.0 allows the visualisation of peaks to determine allele sizes. The algorithms integrated in this program have shown accurate results for fragment analysis applications in linkage analysis, paternity testing, animal parentage and animal genotyping (Applied Biosystem Inc.).

2.2.6.4. Data analysis and statistics

The results of the multiplex PCR provided the fragment lengths i.e. alleles. The mean number of alleles per locus, observed heterozygosities, expected heterozygosities and deviations from Hardy-Weinberg proportions were calculated using Arlequin 3.1 (Excoffier *et al*., 2005) which determines the level of genetic diversity. Linkage disequilibrium was tested using Arlequin 3.1 (Excoffier *et al*., 2005). The fixation index (Fst) and AMOVA tests were used to measure population differentiation due to genetic structure (Holsinger & Bruce, 2009). Fst and AMOVA analyses are estimated from genetic polymorphisms such as SNPs and microsatellites

(Holsinger & Bruce, 2009). The Fst test is based on the variance of allele frequencies between populations (Holsinger & Bruce, 2009). Interpretation of the results of the Fst test is a comparison of the genetic variability within and between populations (Holsinger & Bruce, 2009). The values range between zero and one, where zero implies that the two populations interbreed freely and a value of one implies that the two populations do not share alleles (Holsinger & Bruce, 2009). Both the Fst and AMOVA tests were carried out using Arlequin 3.1 (Excoffier *et al.*, 2005). MICRO-CHECKER (Van Oosterhout *et al*., 2004) was used for detecting null alleles and genotyping errors. The genetic relationships between the populations were then inferred using a Bayesian clustering analysis via a statistical programme called STRUCTURE v2.3.4 (Pritchard *et al*., 2000). Assessments were conducted with the USEPOPINFO = POPFLAG 0 option active. STRUCTURE was run for 5 replicates from K = 1-12, with a run-length of 500,000 repetitions of Markov chain Monte Carlo, following the burn-in period of 20,000 iterations. The five values for the estimated ln(*Pr*(X\K)) were averaged, from which the delta K was calculated. The K value with the highest delta K was used as the best K value for the dataset.

2.3. Conclusion

The molecular methods described above were used to obtain results which are depicted and discussed in chapter three and four. Images of the gel electrophoresis carried out after DNA extractions and PCR of the DNA barcodes and microsatellite loci, indicated clear discrete bands within the ranges expected in each instance (not shown in this dissertation). The results obtained from these techniques were used to test the hypotheses proposed in Section 1.6, Chapter 1.

Chapter 3

A population genetics study of *Cossypha* robin-chats in southern Africa

Abstract

Southern Africa boasts a wide diversity of avian species with five *Cossypha* (robin-chat) species (*C. heuglini*, *C. caffra*, *C. humeralis*, *C. natalensis* and *C. dichroa*) distributed at varying levels of sympatry and allopatry. Due to the effects of climate change and land use patterns that may have affected distributions, many species which were once ecologically segregated may now overlap in regions leading to possible genetic introgression and hybridization. This study investigates the genetic diversity and relatedness between the five *Cossypha* robin-chat species. Genomic DNA was extracted from blood samples of 92 individual birds across the range of all species using the standard phenol:chloroform extraction method. Mitochondrial and nuclear markers were analyzed using Bayesian and Likelihood methods to determine speciation patterns and phylogenetic relationships of the five *Cossypha* species. Barcoding using the cytochrome *c* oxidase subunit 1 (*CO1*) gene was used to confirm individual species identification. The construction of neighbour-joining and maximum likelihood trees provides a graphic representation of the pattern of divergence between the five *Cossypha* species; species clustered together with strong bootstrap values according to the *CO1* barcodes and a haplotype network constructed using PopART confirmed this. However, a Bayesian clustering analysis of microsatellite data using STRUCTURE indicated three distinct clusters (K=3). The data suggest that the five species have recently speciated; however, only three of the seven markers were polymorphic suggesting that they may not have provided the most accurate representation of the genetic relationships of these species. *Cossypha dichroa* and *C. natalensis* appear not to be each other's closest relatives despite the recorded hybridization events.

Keywords: Cossypha, robin-chats, DNA barcoding, microsatellite, population genetics

3.1. Introduction

Until the 1970's, evolutionary trees have been constructed mainly relying on morphological characteristics; however, the emergence of implementing molecular tools in phylogenetics has led to a genetic revolution (Avise, 2004; Ottenburgh *et al.*, 2016a). According to Joseph and Buchanan, (2015), this has led to a quantum leap in avian biology resulting in 27 publications in eight journals based on a genomic dataset of 48 bird species (Jarvis *et al.*, 2014). Contrasting results of introducing molecular tools in phylogenetic studies were reported which lead to further focus on biological processes, such as incomplete lineage sorting, gene duplication, and most importantly hybridization (Jarvis *et al*., 2014). Given the widespread occurrence of hybridization in birds, almost one in ten species, phylogenetic networks constructed using molecular tools are believed to be a powerful technique to display and analyze the evolutionary history and relatedness of species (Ottenburgh *et al*., 2016a).

Southern Africa boasts a wide diversity of avian species with five *Cossypha* robin-chat species (*C. dichroa, C. natalensis, C. caffra, C. heuglini, C. humeralis)* distributed at varying levels of sympatry and allopatry across the region (Hockey *et al*., 2005). The five species of *Cossypha* robin-chats have distinguishing characteristics including unique species specific calls (Hockey *et al*., 2005; Sinclair *et al*., 2005). *Cossypha natalensis*, *C. dichroa*, and *C. humeralis* are all capable of song mimicry while still maintaining their unique calls (Sinclair *et al*., 2005; Ferguson *et al.*, 2002).

Due to the effects of climate change and land use patterns that may have affected distributions, many species which were once ecologically segregated may now overlap in regions leading to possible genetic introgression and hybridization (Allendorf *et al.*, 2001). Possible hybridization in *Cossypha* robin-chats has been identified based on morphological characteristics (Clancey, 1982; Davies *et al*., 2011). *Cossypha haagneri* found in the Eastern Cape was a proposed new species (Gunning, 1909). This newly discovered *Cossypha* species was then hypothesized to be a possible hybrid as it resembled an intermediate between *C. dichroa* and *C. natalensis* (Clancey, 1982). The appearance of similar phenotypic hybrids (Figure 1) in the Eastern Cape and KwaZulu-Natal province, suggested that hybridization between *C. dichroa* and *C. natalensis* may be more common than previously thought (Davies *et al.*, 2011). In addition, inferred hybrids display variable phenotypes suggesting that back-crossing may be present (Davies *et al*., 2011).

Hybrid identification is generally based on phenotypic assessment as in the case of the five *Cossypha* species found in southern Africa (Mallet, 2008). The identification of hybrids has proven to be a challenge in many cases, relying exclusively on morphological characteristics (Avise, 2004). Molecular methods and morphological characteristics are now used in conjunction to identify hybrids and closely related species (Avise, 2004). Molecular techniques such as DNA barcoding and microsatellite genotyping can elucidate theories of hybridization and the degree of relatedness between species (Schlotterer, 2004).

DNA barcoding has become an important tool to catalogue the diversity of species and improve taxonomic classification (Bilgin *et al*., 2016; Purty & Chatterjee, 2016). The mitochondrial cytochrome c oxidase 1 (*CO1*) gene is the standard DNA barcode for animals (Hebert *et al*., 2003; Jordaens *et al*., 2015). DNA barcoding is based on the idea that a short standardized sequence can differentiate an individual of a species from those of another species because the sequence variation between species is assumed to be more than that within species (Bilgin *et al*., 2016). Genetic introgression and sharing of DNA barcode haplotypes often occurs in species after hybridization (Toews *et al*., 2011). This is a result of backcrossing of the fertile first generation hybrids which have the mtDNA of their maternal parent, leading to the replacement of haplotypes in different species or the transfer of alleles from one species to the other (Rheindt and Edwards, 2011).

Hebert *et al.* (2004) and Hajibabaei *et al.* (2005) have shown that more than 95% of vertebrate species possess unique *CO1* barcodes, allowing species level identification to be successful through DNA barcoding. Avian diversity using DNA barcodes have been implemented in North America, Korea, Argentina and Scandinavia (Hebert *et al*., 2004, Yoo *et al*., 2006, Kerr *et al*., 2009, Johnsen *et al*., 2010). However, in many cases DNA barcoding is coupled with microsatellite markers to encompass a large pool of genetic variation (Lu *et al*., 2001; Toews *et al*., 2011; Coetzer *et al*., 2015; Germain-Aubrey *et al.*, 2016; Samani *et al*., 2016)

Microsatellites have become a popular choice of marker for studies investigating parentage or species relatedness (Germain-Aubrey *et al*., 2016). They are found in abundance throughout the genome, are highly polymorphic and are inherited according to Mendelian inheritance (Morgante & Olivieri, 1993; Germain-Aubrey *et al*., 2016). The mutability of a microsatellite is also influenced by the flanking region which is described as a single copy DNA sequence immediately upstream or downstream of the microsatellite loci (Buschiazzo & Gemmell, 2006). The microsatellite flanking regions are generally conserved across individuals of the

same species and occasionally of different species allowing microsatellite loci to be identified by the sequence of the flanking regions (Selkoe & Toonen, 2006). Differences in alleles are not only determined by the variation in repeat number, the substitutions, insertions and deletions in the flanking regions are also relevant (Viard *et al*., 1998). With the variety of data provided by the use of microsatellite markers in conjunction with the maternal inheritance data from DNA barcoding it is widely acknowledged and advocated that both markers be used to gain a holistic understanding of species relatedness and identification (Yang *et al*., 2016).

Based on the suspected hybridization between *C. dichroa* and *C. natalensis* by Davies *et al.* (2011), the degree of relatedness in five co-occurring *Cossypha* species was investigated. The observation of this possible hybridization event provides the basis for the following hypotheses to be tested:

1) *Cossypha dichroa* and *C. natalensis* are the two most closely related species amongst the five *Cossypha* robin-chat species found in southern Africa and

2) The five *Cossypha* robin-chat species are genetically distinct from each other.

3.2. Methods and materials[1]

3.2.1. Genomic DNA extraction

Blood samples were collected from *Cossypha* populations occurring in New Forest, KwaZulu-Natal (nfa); Twin Streams, KwaZulu-Natal (TW and Arabic numerals between 487 to 548); Wits Rural Facility, Mpumalanga (WR); Pullen farm, Mpumalanga (Pu); Inhamitanga Forest, Mozambique (M); Vernon Crookes Nature Reserve, KwaZulu-Natal, (A and B) and Soutpansberg region, Limpopo (S), where all five species co-occur. The geographical reference to all these sites can be seen in Table 2, Chapter 2. Total genomic DNA extractions were performed from ninety-two blood samples via the phenol-chloroform based DNA extraction method originally designed by Blin and Stafford (1976). Samples were stored at 4˚C (Loparev *et al*., 1991).

[1] A brief outline of the methods and materials used are found in this chapter. Chapter two describes the methods and materials in detail.

3.2.2. Mitochondrial DNA analysis

3.2.2.1. Amplification of the *CO1* gene

PCR was carried out according to the protocol of Hebert *et al*. (2004) using the BirdF1-TTCTCCAACCACAAAGACATTGGCAC and BirdR1-ACGTGGGAGATAATTCCAAATCCTG primers The PCR cycle consists of 1 minute at 94°C to denature the template DNA followed by 5 cycles of 1 minute at 94°C, l.5 minutes at 45°C and 1.5 minutes at 72°C and then 30 cycles of 1 minute at 94°C, l.5 minutes at 51°C and 1.5 minutes at 72°C to amplify the desired region and finally 5 minutes at 72°C for the final extension of any partially synthesized fragments (Hebert *et al.*, 2004). The *CO1* barcodes amplified by PCR were sent to Inqaba Biotech in Pretoria for sequencing.

3.2.2.2. Analysis of *CO1* barcode divergence

The *CO1* barcode sequences obtained were aligned with the use of ClustalW (Larkin *et al*., 2007). The sequence divergence was calculated using the Kimura-2-Parameter (K2P) (Kimura, 1980). A neighbour-joining (NJ) (Saitou & Nei, 1987) tree of the K2P distances and a maximum likelihood (ML) phylogeny using the Kimura-2-model was constructed based on a model test resulting in K2+G having the lowest BIC value (Olsen *et al.*, 1994), to provide a graphic representation of the evolutionary relationships between the different *Cossypha* species and any inferred hybrids. Bootstrapping statistics were calculated for the evaluation of the reliability of the inferred clades (Felsenstein, 1985). These procedures were accomplished using Molecular Evolutionary Genetics Analysis (MEGA v6) software (Tamura *et al*., 2013). A minimum spanning haplotype network was then constructed using PopART (Bandelt *et al*., 1999).

3.2.3. Microsatellite analysis

3.2.3.1. Amplification of microsatellite regions

PCR was carried out for the amplification of the DNA microsatellite markers. A total of seventy-eight samples were used for microsatellite analysis. Fourteen fewer samples were used for the microsatellite genotyping as the blood samples of the inferred hybrids and some of the species samples were depleted. The following thermocycling profile was followed: two minutes at 95°C followed by 40 cycles of thirty seconds at 95°C, annealing temperatures according to the primers (Table 4) and one minute at 72°C, and a final extension at 72°C for two minutes.

Thirteen microsatellite loci isolated from a genomic library of *C. natalensis* according to Wogan *et al*. (2015) were tested. Seven of the thirteen microsatellite primer pairs cross amplified in all five species. These seven markers were then used for the study.

3.2.3.2. Multiplex PCR

Sample fragments were labelled with fluorescent labels on the forward primers (Life Technologies, Inc) to allow PCR reactions to be multiplexed. Table 4 shows the seven microsatellite markers which were selected due to their ability to cross amplify in all five species. The primers were pooled in two channels according to fluorescent labels and annealing temperatures.

Table 4: Microsatellite markers (Wogan *et al*., 2015) used for multiplex PCR.

Microsatellite marker	**Forward (labelled) and reverse primers**	**Allele Range (bp)**
CNA69	F-VIC-CCACCTTTAATACATTTCTAGTCAGTC R-TTGTCCTTCCAAAACCAACC	152-198
CNA99	F-VIC- GGGTTCCTGTTCCCTTCTCT R- CCATGTCCTGTGCATCTCAA	106-137
CNA109	F-NED-GCACATATTGCCTTACAGTG R-AATTGCACAGGCTAATATG	170-214
CNA113	F-FAM- CAGCACTCAGGCAAATGAAA R-AGCAGCTCAGAAGGCAAAAC	108-152
CNA130	F-HEX-GTGATTAGCAGAGTTAGCTTC R-TCCACAGAAATCTCGAACAG	147-184
CNA142	F-FAM-AAGCAAGGCAGGATGCTCAC R-TTGTCTATGATTCTTAGCAC	181-213
CNA180	F-HEX-ACATCTGCAGAGCACCATTG R-GAGCCAGGGAAGGAAGGAT	101-125

3.2.3.3. Microsatellite allele sizing (genotyping)

Microsatellite PCR product sizes were detected in an automated DNA analyzer instrument (ABI3100) run at the Stellenbosch University, Central Analytical Facility. For the purpose of this study Genescan^{TM} 500 Liz^{TM} (Applied Biosystem Inc., Stellenbosch) internal size standard on an ABI 3130 Genetic Analyzer and Peak Scanner Software^{TM} v1.0 was used.

3.2.3.4. Data analysis and statistics

The results of the multiplex PCR provided the fragment lengths, i.e. alleles. The mean number of alleles per locus, observed heterozygosities, expected heterozygosities, and deviations from Hardy-Weinberg proportions were calculated using Arlequin 3.1 (Excoffier *et al*., 2005). The Fst and AMOVA tests were used to measure population differentiation due to genetic structure also carried out using Arlequin 3.1 (Holsinger & Bruce, 2009). MICRO-CHECKER (Van Oosterhout *et al*., 2004) was used for detecting null alleles and genotyping errors. The genetic similarities between the populations and possible admixture were then inferred using a Bayesian clustering analysis via a statistical programme called STRUCTURE v2.3.4 (Pritchard *et al*., 2000).

3.3. Results

3.3.1. MtDNA phylogenetic analysis

3.3.1.1. Analysis of *CO1* barcode divergence

CO1 barcoding was used to construct a NJ and a ML tree (Figure 2 & 3) to provide a graphic representation of the pattern of divergences between the five *Cossypha* species. The sequence of a bearded scrub robin (*Cercotrichas quadrivirgata*) (Muscicapidae, subfamily Muscicapinae) was used as the outgroup (Sangster *et al*., 2010).

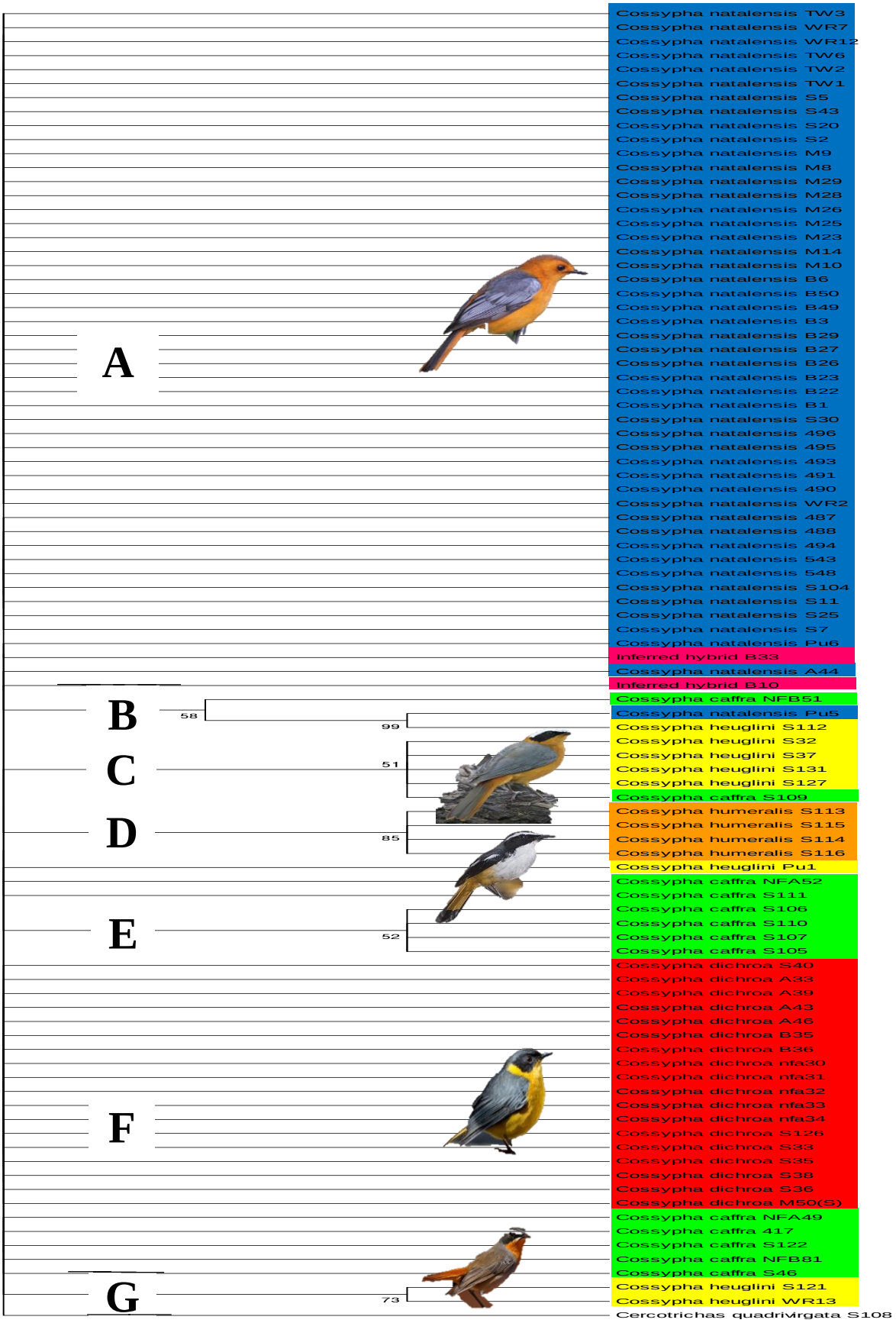


Figure 2*:* NJ Phylogenetic tree of *Cossypha* robin-chats.

In Figure 2, the evolutionary history was inferred using the NJ method (Figure 2, Saitou and Nei, 1987). The tree is drawn to scale where branch lengths represent evolutionary distances as computed using the K2P method (Kimura, 1980). Bootstrap (1000 replicates) values greater than 50% are shown next to the branches (Felsenstein, 1985). Evolutionary analyses were conducted in MEGA6 (Tamura *et al*., 2013). The bearded scrub robin (*Cercotrichas quadrivirgata*) was used as an outgroup and a polytomy is seen with the outgroup. The separation between the *Cossypha* species and the *Cercotrichas quadrivirgata* was supported with a bootstrap value of 1% (Appendix 1). Both the *Cossypha* species and the outgroup belong to the same order and family i.e. Passeriformes and Muscicapidae and both species are found in southern Africa.

The NJ phylogenetic tree grouped species as separate clades despite the polytomy. Individuals from a species cluster together and the boot strap values support separate lineages. *Cossypha caffra* did not separate completely and form a single clade (Clade E and F). Clade B was made up of *C. natalensis, C. caffra* and *C. heuglini*. *C. heuglini* shows a diverse clustering pattern with samples *C. heuglini* S112 and PU1 clustering amongst the *C. caffra* samples in clade B and E respectively, while majority of the samples are present in clade C with the exception of *C. heuglini* S121 and WR13 in clade G.

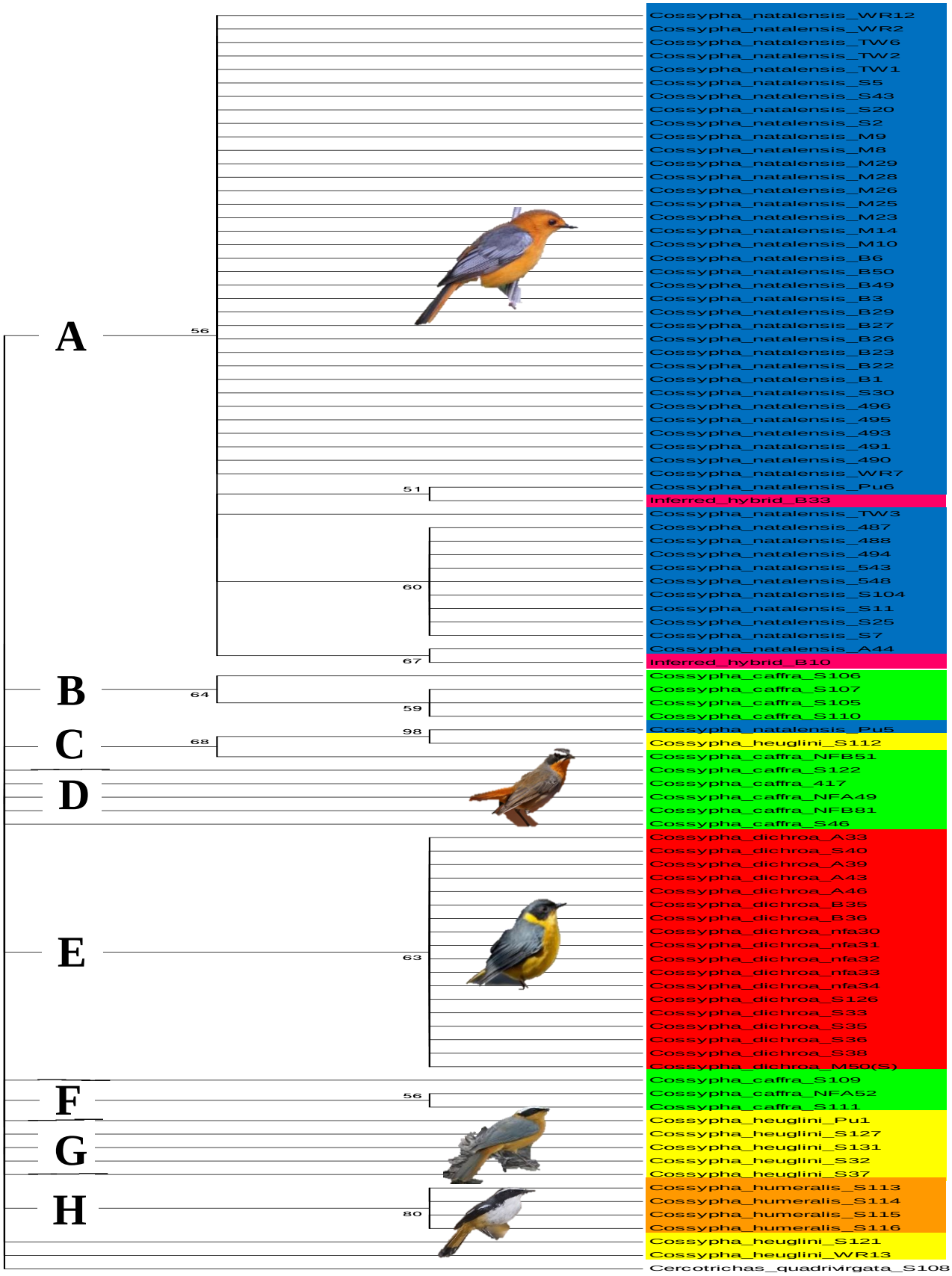


Figure 3: Maximum likelihood estimate of the phylogeny for *COI* barcode sequences from five different species of *Cossypha*.

The ML phylogenetic tree grouped the species as separate clades (Figure 3) although there was one clade (Clade C) which was made up of *C. natalensis*, *C. caffra* and *C. heuglini* as seen in the NJ tree as well (Figure 2, Clade B). *Cossypha caffra* splits into three separate clades (Clade B, D, and F). *Cossypha natalensis* and *C. dichroa* species cluster as separate monophyletic clades in a high percentage of the 1000 bootstrap replicates which is in agreement with the results of the NJ tree.

The DNA barcoding results showed admixture between the five species seen in Clade C. A separation can be seen between the five species despite the polytomy seen as a result of the absence of a clear root on the tree. The polytomy illustrates close genetic relationships between the species as well as the outgroup. The polytomy could be attributed to the use of an outgroup (*Cercotrichas quadrivirgata*) which appeared to be closely related genetically to the *Cossypha* species.

The NJ phylogenetic tree (Figure 2) differed slightly with *C. caffra* and *C. heuglini* clustering amongst the other clades. The branching patterns and similarity between the five *Cossypha* species was supported by the minimum spanning network haplotype (Figure 4) constructed using PopART (Bandelt *et al*., 1999).

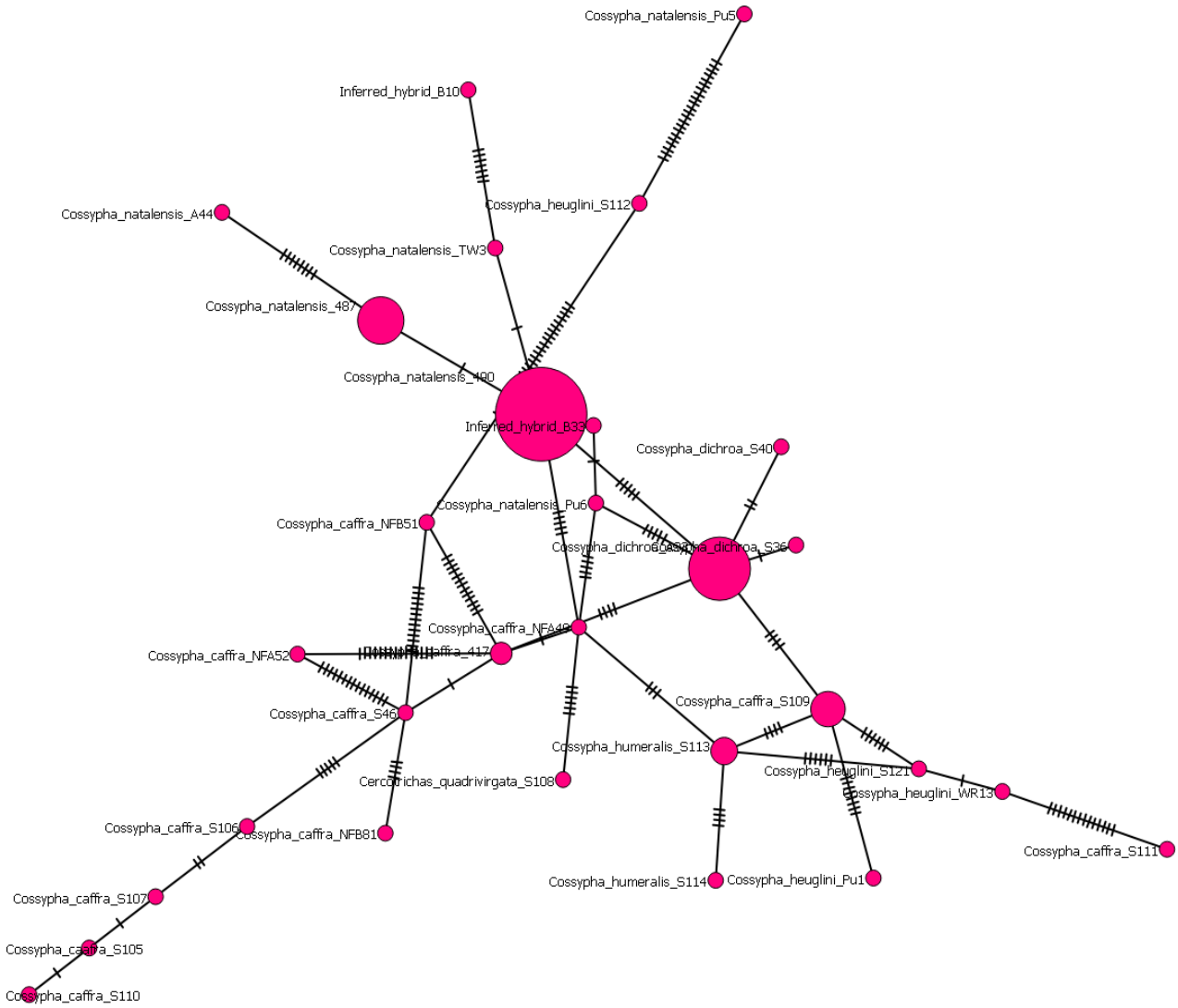


Figure 4: Minimum spanning haplotype network of five species of *Cossypha* robin-chats. The size of each circle is in proportion to the number of samples it represents.

The minimum spanning network illustrates great similarity between the five species of *Cossypha* robin-chats with regards to the *CO1* barcode (Figure 4). There is one very common haplotype (large circle) with most variants spanning from it, alluding to limited genetic diversity. This representation of the *CO1* data supports those of the phylogenetic trees (Figure 2 and 3). The outgroup used (*Cercotrichas quadrivirgata*) does not diverge totally in the minimum spanning network supporting the polytomy seen in the phylogenetic trees. Appendix 2 generated by PopART illustrates the identical sequences found within the dataset.

3.3.1.2. Intraspecific and interspecific K2P distances

The intraspecific and interspecific genetic distances are shown in Table 5 and Table 6 respectively. The genetic distances are calculated using the K2P method (Kimura, 1980).

Table 5*: Cossypha* robin-chat intraspecific K2P distances.

Species	**Mean intraspecific distances (%)**	**Number of samples**
C. dichroa	1.4	18
C. natalensis	5.4	47
C. caffra	22.8	13
C. heuglini	39.3	8
C. humeralis	2.6	4
Inferred hybrids	0.0067	2
Mean	14.3	92

Table 6: Mean interspecific pairwise K2P distances (%).

	C. dichroa	***C. natalensis***	***C. caffra***	***C. heuglini***	***C. humeralis***
C. dichroa	*				
C. natalensis	16.0	*			
C. caffra	17.8	23.8	*		
C. heuglini	25.6	31.7	33.6	*	
C. humeralis	12.9	18.7	20.4	26.2	*

Analysis of the K2P distances determined that the average intraspecific distance for the *CO1* barcodes analyzed in this study was 14.3%. The standard screening threshold of sequence differences or the cut off to determine separate species requires the interspecific distance to be 10x the mean intraspecific distance (Hebert *et al*., 2004). All interspecific distances were less than 3x the mean intraspecific distance indicating that these species may be more closely related.

3.3.2. Microsatellite profile analysis

The seven cross amplifying microsatellite markers were fluorescently labelled for multiplex PCR and analyzed. The identification of species specific markers or private alleles as well as shared alleles across the five species was then documented. Loci CNA109 and CNA142 show a lack of private alleles while loci CNA180 and CNA69 show a lack of shared alleles across the five species as seen in Table 7 below. The summary of allele frequencies can be seen in Table 8 while a comprehensive table of the allele frequencies for each species can be seen in Appendix 3.

Table 7: Number of shared alleles and private alleles per microsatellite marker.

Locus	**Number of private alleles**	**Number of shared alleles across 5 species**	**Total number of alleles**
CNA113	6	1	14
CNA99	2	2	14
CNA109	0	2	10
CNA180	4	0	12
CNA142	0	1	11
CNA69	3	0	14
CNA130	3	1	14

Table 8*:* Summary of allele frequencies across seven loci for each *Cossypha* species.

Species	**Number of samples**	**Mean number of alleles across 7 loci**
C. dichroa	18	7
C. natalensis	26	8.71
C. caffra	18	7.43
C. heuglini	12	7.29
C. humeralis	4	3.29

MICRO-CHECKER (Van Oosterhout *et al*., 2004) was used to detect possible genotyping errors and null alleles. Null alleles were observed in four of the loci (i.e. CNA99, CNA180, CNA142, CNA130) where all five species were analysed as a single dataset due to sample sizes. The dataset was not adjusted as re-sequencing of the data was not possible.

A pairwise linkage disequilibrium test was performed intraspecifically for each species and interspecifically using Arlequin 3.1 (Excoffier *et al*., 2005). No deviations from linkage disequilibrium were detected (Appendix 4). Arlequin 3.1 (Excoffier *et al*., 2005) provides a chi-squared test and P-value for the indication of linkage disequilibrium as opposed to the conventional R^2 and D values (Appendix 4). This form of representation of linkage disequilibrium has been used in the literature (Woolaver *et al*., 2013).

Hardy-Weinberg statistics were determined using Arlequin 3.1 (Excoffier *et al*., 2005) for each of the species. The P-values show that the five species do not deviate from Hardy-Weinberg equilibrium across the seven loci as indicated in Table 9.

Table 9*:* Hardy-Weinberg statistics in the *Cossypha* robin-chats.

	Mean observed heterozygosity	**Mean expected heterozygosity**	**Mean P-values**
C. dichroa	0.78708	0.76951	0.23033
C. natalensis	0.84668	0.80959	0.39041
C. caffra	0.69692	0.77189	0.19586
C. heuglini	0.72034	0.83862	0.36659
C. humeralis	0.75000	0.71360	0.70207

An AMOVA analysis was then conducted to determine the genetic variation between the five species. The AMOVA analysis was run interspecifically and intraspecifically to allow a holistic understanding of the species.

Table 10: Analysis of molecular variance (AMOVA) for *Cossypha* species.

Source of variation	**d.f.**	**Sum of Squares**	**Variance components**	**Percentage of variation**	***F*st**
Among Populations	4	7.65	0.04	4.89	
Among individuals within populations	75	55.21	-0.02	-2.45	
Within Individuals	80	62.00	0.78	97.56	
Total	159	124.86	0.79		0.05

Table 10 shows an Fst value of 0.049 when considering the five different species together. This value is closer to 0 than to a value of 1 indicating slight genetic differentiation. When compared with the Fst values seen in Appendix 5 this value is higher than those of *C. dichroa* and *C. natalensis* which are considered as 0 due to the negative Fst values. *Cossypha caffra* and *C. heuglini* indicate greater genetic diversity in comparison to *C. dichroa* and *C. natalensis* with values of 0.44041 and 0.10725 respectively. A pairwise Fst test was conducted in conjunction with the intraspecific and interspecific AMOVA tests. The pairwise Fst test showed similar results and the interspecific and intraspecific AMOVA test show Fst values which are an approximate average of the pairwise Fst test.

The genetic similarity between the five species was then inferred via a Bayesian clustering analysis statistical program which could determine the best K-value or number of distinct species/populations present in a data set. The analysis using STRUCTURE v2.3.4 (Pritchard *et al*., 2000) suggested the best K-value as K=3 (Figure 5) indicating the presence of three distinct species instead of five. The STRUCTURE analysis was then performed using K=3 (Figure 6) to get a clear comparison between the five species and determine which were closely related and shared genetic material. The proportion of membership was approximately 1:3:5:5:7 for *C. humeralis*, *C. heuglini*, *C. caffra*, *C. dichroa* and *C. natalensis* respectively which illustrates the unequal representation of the five *Cossypha* species used in this study.

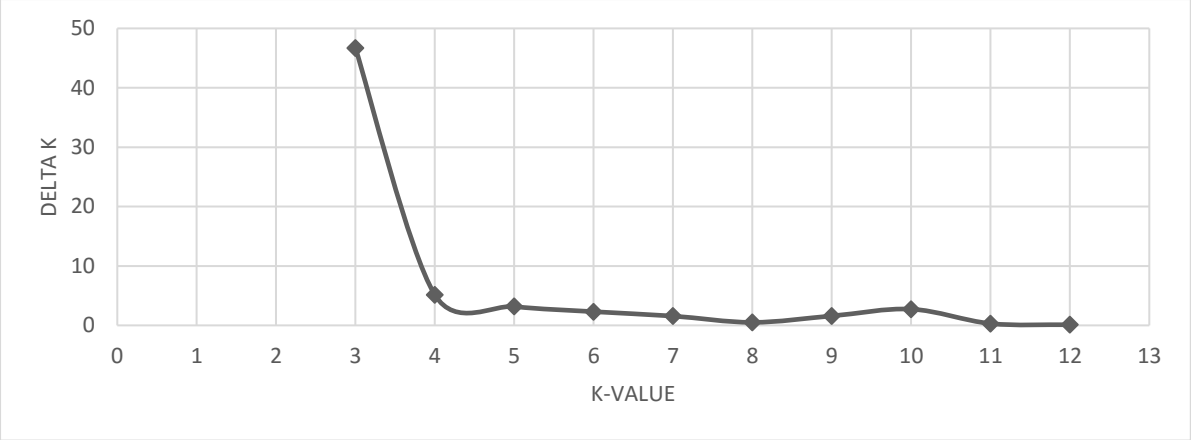


Figure 5*:* Delta K values for species structures of K=1-12.

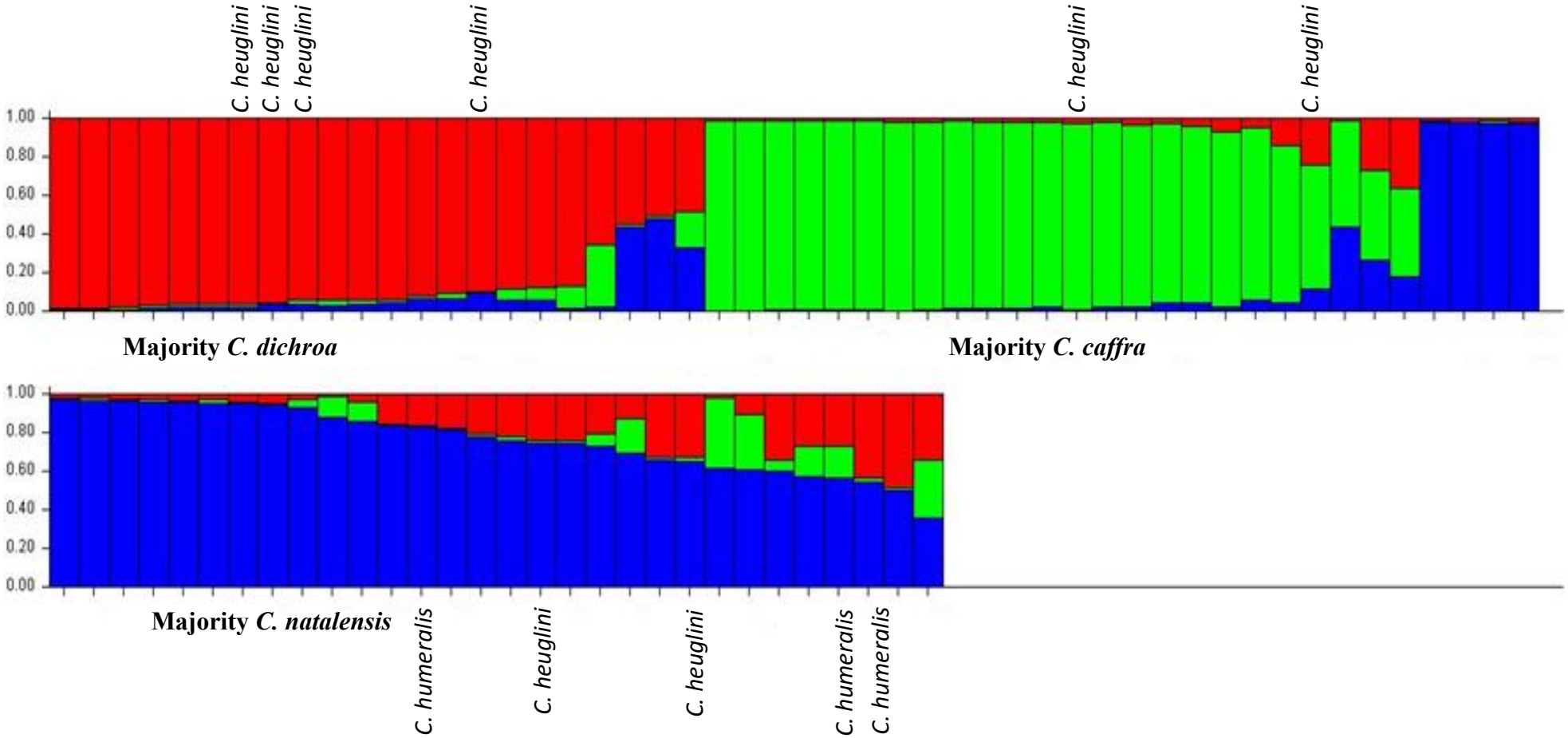


Figure 6: STRUCTURE analysis based on the microsatellite genotypes indicated three distinct clusters (K=3) of *C. dichroa*, *C. natalensis*, *C. caffra*, *C. heuglini*, and *C. humeralis*. Each individual is represented by a single vertical line, with lengths proportional to the estimated membership in each cluster. The red genetic cluster represents majority *C. dichroa* with five *C. heuglini* clustered amongst them. The green cluster represents majority of *C. caffra* with two *C. heuglini* clustered with *C. caffra*. The blue cluster represents majority of *C. natalensis* with four *C. humeralis* in this cluster.

It can be seen that individuals from *C. dichroa* and *C. heuglini* share similar microsatellite loci profiles while *C. caffra* and *C. heuglini* also share microsatellite loci profiles. The blue represents majority of *C. natalensis* with *C. humeralis* sharing genetic makeup based on the seven microsatellite markers used.

3.4. Discussion

The five *Cossypha* species are unquestionably separate species based on phenotypic, morphological, ecological and behavioural differentiation, however, the markers used in this study do show admixture between the sympatric species. The STRUCTURE analysis resulted in three distinct clusters for the *Cossypha* individuals used in this study. Analysis of both nuclear and mtDNA support that *C. dichroa* and *C. natalensis* are not each other's closest relatives despite recorded hybridization events. The ML phylogenetic tree (Figure 3) indicates the clustering patterns of the five species. Individuals of a species form a cluster and the bootstrap values support separate lineages with a percentage higher than 50% although both trees illustrate polytomies or multifurcating branches which are thought to be more accurate in reflecting evolutionary relationships (Lin *et al*., 2010). The outgroup (*Cercotrichas quadrivirgata*) diverges with a bootstrap value of 1% (Appendix 1) which indicates that *Cercotrichas quadrivirgata* is not as genetically distinct in comparison to the *Cossypha* robin-chats. Polytomies are thought to be common when constructing phylogenetic trees of species which have recently speciated due to multiple vicariant events (Hoelzer & Meinick, 1994). An example of polytomous phylogenetic trees have been studied in Macaque monkeys, where three species namely, *Macaca mulatta*, *M. fuscata* and *M. cyclopis* are geographically segregated but form a monophyletic group or polytomy (de Queiroz & Donoghue, 1990). A polytomous relationship is also seen in the sitellas (*Daphoenositta*) of Australia (Cracraft, 1989). The five species of sitellas are isolated species in humid conditions however, once the climate changes, the differentiated species come into contact and are able to hybridize in areas of overlap, while maintaining distinctiveness (Cracraft, 1989). Based on the polytomous phylogenetics of these birds the five species are now regarded as megasubspecies (Cracraft, 1989). Similarly, the phylogenetic trees constructed in this study of the *Cossypha* species could indicate recent speciation or hybridization based on the polytomy. The clades predominantly reflect a single species with the exception of Clade B, C, D, and F. Clade C contains 3 different species ie. *C. natalensis*, *C. caffra*, and *C. heuglini*, while *C. caffra* forms clades B, D, and F. The minimum spanning network (Figure 4) supports the results obtained from the phylogenetic trees. The genetic diversity seen in *C. caffra* is mirrored in all three analyses. The identical sequence list (Appendix 2) supports the branching patterns seen in the haplotype network and the ML and NJ phylogenetic trees as well.

Inferred hybrids were phenotypically identified as hybrids as they have characteristics of both species *C. dichroa* and *C. natalensis* (see Figure 1 in Davies *et al.*, 2011). The mitochondrial

DNA result indicates the female parent was a *C. natalensis* and the male parent a *C. dichroa*. This is supported based on male *C. dichroa* being larger than male *C. natalensis* by at least 10 g. This size difference might allow male *C. dichroa* to dominate *C. natalensis* males in territory and mate acquisition. Although both inferred hybrids are of *C. natalensis* maternal parentage no conclusion could be drawn with regards to matings being unidirectional and thus sex-bias hybridization as the blood of only two inferred hybrid specimens was obtained in this study. It would be interesting to expand this study to include more hybrid specimens to determine whether the inferred hybridization is significantly sex-biased and occurring mainly between *C. dichroa* males and *C. natalensis* females. Also, while we only identified two phenotypic hybrids we cannot be sure that, given the phenotypic variability of inferred hybrids, more genotypic hybrids do not exist in the samples we analyzed. Subsequent to the publication by Davies *et al.* (2011) more inferred hybrids have been identified across a region of sympatry.

The DNA barcoding results showed a clear separation can be seen between *C. dichroa* and *C. natalensis*. *Cossypha caffra* clusters as two separate clades in the NJ tree (Figure 2, Clade E and F) while it formed three separate clades in the ML tree (Figure 3, Clade B, D, and F). A divergence is seen between two *C. caffra* samples from different regions (Soutspansberg and New Forest). However, divergence is also seen amongst *C. caffra* samples originating from the same region (Soutspansberg). Although the five *Cossypha* species do separate into different clades a high degree of relatedness is observed based on the data obtained from the interspecific and intraspecific distances. The unpredictable clustering of *C. caffra* indicates a large amount of genetic diversity within the species as compared to the other four species and as supported by the Fst values. The high percentage of the 1000 bootstrap replicates (98%) divergence between *C. natalensis* PU5 with *C. heuglini* S112 preceded by a divergence splitting with a *C. caffra* NFB51 sample is of interest as it may indicate genetic introgression between these species or it could be that this technique is not sensitive to recent divergence events. The NJ tree mirrors the particular clade patterning of the ML tree however it differs with regards to *C. caffra* lineages (Figure 2 & 3). *Cossypha caffra* clusters as 3 separate clades in the ML tree (Clade B, D, and F) as well as with *C. heuglini* and *C. natalensis* in clade C.

The branching patterns seen in the analysis of the DNA barcodes show the divergence of the five species and the five different species are seen to separate with the exception of a few samples. These exceptions may be due to interbreeding of the samples in regions where they co-occur such as Soutspansberg. The sample size used may also affect the results of the tree due to the small number of samples for certain species such as *C. humeralis* (n=4) and *C. heuglini* (n=8) in

comparison to *C. natalensis* (n=47). Hebert *et al*. (2004) proposed a threshold when determining the separation of species using DNA barcoding. This threshold is regarded as high enough to separate specimens belonging to different species and low enough to recognise recently diverged species (Hebert *et al.*, 2004; Bilgin *et al*., 2016). This threshold or barcoding gap is defined as ten times the mean intraspecific variation for the samples studied (Hebert *et al*., 2004). In this study, the mean intraspecific divergence (Table 5) was 14% for the samples analyzed and the mean interspecific pairwise K2P distances (Table 6) were all no more than three times the mean intraspecific variation. This result implies that these five species may not actually be different species and may be closely related sister species or sub-species (Hebert *et al*., 2004). The polytomy of the phylogenetic trees in Figure 2 and Figure 3 both support and contribute to the barcoding gap being below the threshold. Taylor & Harris (2012) do argue that DNA barcoding may not be as effective in determining the identification of species that have recently diverged. By 2012, only a relatively small number of studies were conducted using DNA barcoding in birds, and as these studies represent aggregations of large numbers of bird species barcodes they may be misleading (Taylor & Harris, 2012; Hebert *et al*., 2004; Yoo *et al*., 2006; Kerr *et al*., 2007, 2009). DNA barcoding may therefore not be the most effective for smaller sample sizes, recently diverged species or species which have recently come into close contact with each other (Taylor & Harris, 2012). Hebert *et al.* (2003) argues that DNA barcoding of the *CO1* gene is the best candidate for a universal barcode. Will and Rubinoff (2004) proposed that the DNA barcoding of the *CO1* gene could not replace the morphological identification and classification of species. MtDNA sequences are known to vary and the rate of evolution is inconsistent within as well as between species (Hebert *et al*., 2004). Based on this statement, the level of divergence is not standard and the use of standard pseudogroups by Hebert *et al.* (2004) which still relies largely on the cladistics method using phenotype as an identification tool, renders DNA barcoding a supporting tool for identification instead of a replacement. A trend noticed in the DNA barcoding studies which have been conducted by Hebert *et al.* (2004) is a single representative of data is used in the distance matrix (Will & Rubinoff, 2004). This approach is thought to prevent the observation of specified variation between individuals and at the same time it does not represent the degree of variability seen within species (Will & Rubinoff, 2004). Hence the results obtained from a DNA barcoding study using the *CO1* gene needs to be supported by phenotypic classification, which is still the basis of the identification for species (Will & Rubinoff, 2004).

This study supports the arguments put forth by Will and Rubinoff (2004) where the barcoding gap indicates that the five *Cossypha* robin-chat species may have recently speciated or may not be separate species. However, the classification of these five separate species have been well documented and supported based on phenotypic identification as well as unique vocalisations of each species (Ferguson *et al.*, 2002; Sinclair *et al*., 2005). The use of DNA barcoding may be more effective in cases where the background of the taxa in question is well documented (Will & Rubinoff, 2004). MtDNA reflects matrilineal history which could be seen as a biased portrayal of the overall lineage of a species (Zink & Barrowclough, 2008). Another concern of using mtDNA exclusively is the evolution of the mitochondrial genome as a single linkage unit (Zink & Barrowclough, 2008). Multiple population divergences or speciation events which occur in a short space of time may not be reflected in a single gene tree. Due to these concerns, many studies advocate the use of multiple, unlinked nuclear loci along with DNA barcoding in determining the genetic structure and relatedness of species (Zink & Barrowclough, 2008; Rubinoff & Holland, 2005).

Therefore, the results from microsatellite analysis were used in conjunction with DNA barcoding to address the hypotheses proposed in this study. The seven microsatellite markers isolated from a *C. natalensis* genomic library (Wogan *et al.*, 2015) cross amplified in all five *Cossypha* species (*C. dichroa*, *C. natalensis*, *C. caffra*, *C. heuglini* and *C. humeralis*). Analysis of the allele frequencies (Table 7, Appendix 3) in each species revealed eighteen private alleles amongst the total eighty-nine identified. From these eighteen private alleles, five alleles were exclusive in *C. natalensis*, two in *C. dichroa*, six in *C. caffra* and five in *C. heuglini* (Appendix 3). The identification of these private alleles can lead to the identification of a species specific marker (Van Wyk *et al*., 2013). Private alleles reveal a clear genetic distinction between the five species. The analysis also indicates a higher diversity with respect to the mean number of alleles per locus (A) and observed heterozygosity (Ho) in *C. natalensis* when compared to the other four *Cossypha* species (Table 9). The number of alleles was lowest in *C. humeralis* which was expected given the sample size of four. Low allelic richness has been seen in many recent studies in fish (Palti *et al*., 2012; McCracken *et al*., 2014; Priest *et al.*, 2014). A possible reason for the low allelic richness may be due to small sample sizes for some of the species (Hale *et al.*, 2012). A study by Hale *et al.* (2012) illustrates the positive effects of having a sample size between 25-30 individuals per species. However, smaller representations of the species studied would provide a basic understanding of the genetics of the population being studied (Hale *et al.*, 2012). The results from MICRO-CHECKER (Van Oosterhout *et al*., 2004) displayed the presence of

null alleles in four of the seven loci used (CNA99, CNA180, CNA142 and CNA130). Null alleles are alleles which fail to amplify in a PCR because the conditions were not ideal or the primer binding region contains mutations which inhibit binding (Van Oosterhout *et al*., 2004). The result being the appearance of heterozygotes as homozygotes and some may not even amplify any alleles (Van Oosterhout *et al*., 2004). This result indicates that there may be more heterozygotes in the data set used in this study than actually presented. This may have skewed the results of the study with regards to the statistical analysis performed based on the heterozygosity and homozygosity of the samples as the data was not adjusted.

Analysis using Arlequin 3.1 (Excoffier *et al*., 2005) provided data regarding heterozygosities, homozygosities, allele frequencies and Fst values which showed that only three of the seven markers were polymorphic when analyzed which was a disadvantage for the study. Balloux *et al.* (2004) argue that more than five polymorphic markers should be used to accurately determine inbreeding and heterozygosities. According to Wimmers *et al.* (2000) the average number of alleles in a population or in this case a species, needs to be at least four in order for microsatellites to be used in the estimation of genetic diversity and genetic distances (Olowofeso *et al.*, 2005). The average number of alleles in each population per loci were all above four with *C. humeralis* being the exception due to the small sample size (n=4). Each species had an average sample size of sixteen whereas *C. humeralis* fell short in this regard with only four samples. These four samples could not be regarded as a correct representation of the *C. humeralis* population in South Africa and the data obtained was not sufficient for tests such as an AMOVA analysis.

There were no departures from Hardy-Weinberg equilibrium in all five species across the seven loci (Table 9). The AMOVA analysis in Table 10 shows an Fst value of 0.04891 when considering the five different species together. This value, closer to 0 than to 1, indicates slight genetic differentiation and much admixture in the population (Holsinger & Bruce, 2009). When compared with the Fst values seen in Appendix 5 this value is higher than those of *C. dichroa* and *C. natalensis* which are considered as 0 due to the negative Fst values. *Cossypha caffra* and *C. heuglini* indicate greater genetic diversity in comparison to *C. dichroa* and *C. natalensis* with values of 0.44041 and 0.10725 respectively. The high Fst value within *C. caffra* supports the branching seen in the phylogenetic analysis where sub-branching is seen in the NJ tree and the multiple clades seen in both trees (Figure 2 and 3).

A Bayesian clustering analysis using STRUCTURE v2.3.4 (Pritchard *et al*., 2000) revealed the best K value is three (Figure 5 & 6). Shared alleles between *C. dichroa* and *C. heuglini* as well

as shared alleles between *C. natalensis* and *C. humeralis* could mean there is interbreeding between these populations. An alternative explanation could be the divergence of *C. humeralis* and *C. heuglini* from the three genetically distinct species, i.e. *C. dichroa, C. natalensis* and *C. caffra*. According to a study by Horreo *et al.* (2016) on the threatened West-Pannonian population of great bustard (*Otis tarda*), the results obtained from STRUCTURE need to be correlated with knowledge of the species or populations in question in the field. Background knowledge of the species is required to establish if the estimation of the number of genetic units present is in fact a correct representation of the species (Horreo *et al.*, 2016). In this case, the expected result would be K=5 indicating five distinct species instead of three.

Based on the results obtained from the DNA barcoding of the *CO1* gene, the haplotype network and the use of seven microsatellites it can be seen that these five *Cossypha* species are closely related, as previously recognised (Ferguson *et al.*, 2002; Sinclair *et al.*, 2005). The DNA barcoding gap however would not classify the five species as different species based on the K2P distances and the screening threshold described by Hebert *et al.* (2004) which is not in agreement with the accepted classification of these species as well as the microsatellite analysis which according to the STRUCTURE analysis recognises three distinct genetic units. Discrepancies between DNA barcoding results and microsatellite analysis have been documented in other studies (Horreo *et al*., 2016; Samani *et al*., 2016; Yang *et al*., 2016). For example, the study by Samani *et al.* (2016) on the phylogenetics and population genetics of catfish (*Plotosus canius*) showed a relatively counter-outcome in comparison with the mitochondrial results where Fst estimations of former populations were lowest amongst *P. canius* samples but differed when comparing the results of the two techniques.

These five *Cossypha* species have been recognised as separate species without question based on their phenotypes and their unique vocalisations (Ferguson *et al.*, 2002; Sinclair *et al.*, 2005). *Cossypha dichroa* has a call which is plaintive *toy-toy, toy-toy* however, its song is loud and most importantly it has the ability to mimic other birds (Sinclair *et al*., 2005). *Cossypha natalensis* and *C. humeralis* also possess this ability of vocal mimicry but each have a distinct soft call (*see-saw, see-saw* and *seet-cher, seet-cher,* respectively) (Sinclair *et al*., 2005). The ability to mimic the songs of different bird species could allow some species to overcome the reproductive isolation created through species-specific vocalisations (Kelley *et al*., 2008). A study by Ferguson *et al.* (2002) explored the vocal mimicry in African *Cossypha* robin-chats with a focus on *C. natalensis* and *C. dichroa*. The study showed that *C. natalensis* and *C. dichroa* are able to imitate the acoustic environment within their habitats (Ferguson *et al*., 2002). Vocal imitation

has been hypothesised to play a role in many different spheres of avian interactions. The ability to imitate different species vocalisations in an environment could lead to a competitive advantage where the imitators' access to a food source shared with the species being imitated could become accessible (Catchpole & Baptista, 1988). Imitation of a predator could allow the imitator to appear as a threat to other bird species in the environment (Robinson, 1975; Igic *et al*., 2015) and it could play a role in sexual selection (Hartshorne, 1956). Vocal mimicry in this case can be used by males to expand their vocal repertoire to appear more attractive to females of the same species or in some cases it could allow for interspecies hybridization to occur based on the imitation of a different bird species leading to females of the imitated species to be attracted based on the vocalisation (Dobkin, 1979; Catchpole, 1980; Baylis, 1982; Searcy, 1984; Catchpole & Slater, 2008; Kelley *et al*., 2008).

Based on the possibilities of overcoming reproductive isolation mechanisms such as vocalisations, we propose that once *C. dichroa* and *C. natalensis* began to coexist in a region like the Eastern Cape or KwaZulu-Natal where the inferred hybrids were documented (Clancey, 1982; Davies *et al*., 2011) they may have begun to mimic each other's songs leading to them overcoming this isolation mechanism and recognising each other as mates. The other *Cossypha* species may not have achieved this and reproductive isolation is maintained, rendering hybridization less likely.

We were able to partially test the parentage of the inferred hybrid by showing that the maternal parent of the inferred hybrid was *C. natalensis* using DNA barcoding. The paternal parentage was not genetically determined due to the small sample size and the lack of blood samples of the inferred hybrid to continue further analysis with microsatellite markers. However, the morphology of these inferred hybrids (Davies *et al.*, 2011), display characteristics seen in *C. dichroa* such as wing length, size and even plumage in some cases, as the phenotypes are variable. A suspected hybrid documented by McKenzie (1998) at Vryheid Hill Nature Reserve, KwaZulu-Natal in November 1998 was seen in the presence of a *C. dichroa*. These phenotypic characteristics provide evidence to suggest that the paternal parent was *C. dichroa* although the collection of more blood samples of inferred hybrids could allow genetic evidence to elucidate the paternal parentage.

The hypotheses proposed that *C. dichroa* and *C. natalensis* are each other's closest relatives and that the five *Cossypha* species are genetically distinct. DNA barcoding and microsatellite genotyping suggested that these two species are not each other's closest relatives despite their

proposed hybridization by Davies *et al.* (2011). DNA barcoding of the *CO1* gene showed separate lineages supported by bootstrap value of 56% between *C. dichroa* and *C. natalensis.* Literature supports the use of a 50% cut off in bootstrap values when constructing both a NJ and ML phylogenetic tree (Johnsen *et al.*, 2010; Sangster *et al.*, 2010; Samani *et al.*, 2016; Zarza *et al.*, 2016). A study by Johnsen *et al.* (2010) focusing on the DNA barcoding of Scandanavian birds showed bootstrap support as low as 60% for the divergence of common scoter (*Melanita nigra*). According to their study 56% bootstrap support is seen for the divergence between sandwich tern (*Thalasseus sandvicensis*) found in Sweden as opposed to sandwich tern (*Thalasseus sandvicensis*) sampled from the United States (Johnsen *et al*., 2010). In a separate experiment, using only *C. dichroa* and *C. natalensis* samples, a NJ tree of the K2P distances (Saitou & Nei, 1987) was constructed to provide a hypothesis of the evolutionary relationships between *C. dichroa* and *C. natalensis* and their inferred hybrids. In the tree, *C. dichroa* and *C. natalensis* each clustered as separate monophyletic lineages that were very well supported at 100% and 99% (1000 bootstrap replicates) respectively. Results not shown here.

This study illustrated that DNA barcoding may be a useful tool in assisting with the identification of many species; however, it may not be applicable to all species. This study also showed the reliance on phenotypic evidence and support when inferring phylogenetic relationships between species. The degree of relatedness between species based on the microsatellite analysis infers that *C. humeralis* may be more closely related to *C. natalensis* and *C. heuglini* could possibly be more closely related to *C. dichroa*. The outcome of the microsatellite analysis proved to be informative in determining the genetic structure of the species although a clearer and more conclusive decision will require the addition of more polymorphic markers as well as an increased sample size particularly for *C. heuglini* (n=8) and *C. humeralis* (n=4).

Acknowledgements

Thank you to Professor Craig Symes from the School of Animal, Plant and Environmental Sciences at the University of the Witwatersrand for the collection of blood samples and Greg Davies of the Ornithology Section, Ditsong National Museum of Natural History, Pretoria, South Africa for his assistance in this process. The University of the Witwatersrand and the National Research Foundation for financial support.

Appendix 1. A neighbour-joining tree with a 0% bootstrap cut off.

Appendix 2. Table illustrating *CO1* sequences which are identical in correlation with the haplotype network in Figure 4.

Identical sequences					
C. natalensis_490	***C. natalensis_487***	***C. dichroa_A33***	***C. caffra_S109***	***C. caffra_417***	***C. humeralis_S113***
C. natalensis_491	*C. natalensis_488*	*C. dichroa_A39*	*C. heuglini_S127*	*C. caffra_S122*	*C. humeralis_S115*
C. natalensis_493	*C. natalensis_494*	*C. dichroa_A43*	*C. heuglini_S131*		*C. humeralis_S116*
C. natalensis_495	*C. natalensis_543*	*C. dichroa_A46*	*C. heuglini S32*		
C. natalensis_496	*C. natalensis_548*	*C. dichroa_B35*	*C. heuglini_S37*		
C. natalensis_S30	*C. natalensis_S104*	*C. dichroa_B36*			
C. natalensis_B1	*C. natalensis_S11*	*C. dichroa_nfa30*			
C. natalensis_B22	*C. natalensis_S25*	*C. dichroa_nfa31*			
C. natalensis_B23	*C. natalensis_S7*	*C. dichroa_nfa32*			
C. natalensis_B26		*C. dichroa_nfa33*			
C. natalensis_B27		*C. dichroa_nfa34*			
C. natalensis_B29		*C. dichroa_S126*			
C. natalensis_B3		*C. dichroa_S33*			
C. natalensis_B49		*C. dichroa_S35*			
C. natalensis_B50		*C. dichroa_S38*			
C. natalensis_B6		*C. dichroa_M50(S)*			
C. natalensis_M10					
C. natalensis_M14					
C. natalensis_M23					
C. natalensis_M25					
C. natalensis_M26					
C. natalensis_M28					
C. natalensis_M29					
C. natalensis_M8					
C. natalensis_M9					
C. natalensis_S2					
C. natalensis_S20					
C. natalensis_S43					
C. natalensis_S5					
C. natalensis_TW1					
C. natalensis_TW2					
C. natalensis_TW6					
C. natalensis_WR12					
C. natalensis_WR2					
C. natalensis_WR7					

Appendix 3. Frequency of private alleles and shared alleles in each of the five *Cossypha* robin-chats. Yellow represents private alleles and green represents alleles shared across all five species.

Locus	Allele	Frequency of alleles for each species				
		C. dichroa	*C. natalensis*	*C. caffra*	*C. heuglini*	*C. humeralis*
CNA113	102		2			
	105			4		
	109		17	7	2	
	113		3	9	1	
	117	1		9		
	120			1		
	124	1	2	2		
	128	2	5	1	3	
	132	8	10	2	4	
	137	1	8	1	3	3
	141	5				
	145				3	5
	154				2	
	94				2	
CNA99	105	1	2	10	1	3
	109			3		
	118	3	1	1		
	122	2	1			
	126	4	5		3	
	131	3	3	2		
	135	10	18	5	5	2
	139	9	14	5	3	
	144		1			1
	148		3		2	
	174			2	1	
	89	2	2	10		
	94				3	
	98	2	2		5	
CNA109	174	2	7	3	2	3
	178	1	3	10	7	
	182	22	23	10	5	2
	186	5			1	
	190		2	1		
	194	2	1	3	1	
	198	2	7			1
	202	4	8	8	1	
	207				3	
	211		1		4	
CNA180	101		9		6	
	109		14	3	6	3
	112	2		10	1	
	116			1		
	120	11	5	11	1	

Locus	Allele	Frequency of alleles for each species				
		C. dichroa	*C. natalensis*	*C. caffra*	*C. heuglini*	*C. humeralis*
	124	1	1		2	
	128		4			3
	132	2		1	4	
	135	2	7	1		
	139				2	
	89			1		
	98	2				
CNA142	171		1	3		
	175	2	1	10	2	
	179	2		3	1	
	185	1	2	6		
	189	1	5	4	4	
	193		9	7	6	
	197		9		5	1
	201	3	6	2	3	4
	205	3	4		1	2
	210	5	2	1		1
	214	7	1			
CNA69	143		3		1	
	153	2	14		4	
	162		1			
	166			10	4	
	170		3			
	174	3		19	6	
	178	1			4	
	182		4	1	4	
	186	5	6			1
	190	10		4	1	4
	194	6	8	2		2
	198		5			1
	202	1	1			
	210		1			
CNA130	149			2		
	154		9	2	1	2
	158				1	
	166		5	5	4	
	170		3			
	174	9	11	17	12	2
	178	4	12			
	182	5	1	1	2	
	186	2	1			1
	190	3		2		2
	194		1	1		
	198		1			1
	212	1		1		
	237			1	2	

Appendix 4. Pairwise linkage disequilibrium in the *Cossypha* robin-chats across 7 loci.

Locus	**CNA113**	**CNA99**	**CNA109**	**CNA180**	**CNA142**	**CNA69**	**CNA130**
CNA113	*	+	+	+	+	+	+
CNA99		*	+	+	+	+	+
CNA109			*	+	+	+	+
CNA180				*	+	+	+
CNA142					*	+	+
CNA69						*	+
CNA130							*

Significance level of $p > 0.05$ (+)

Appendix 5. Analysis of molecular variance (AMOVA) of each *Cossypha* species.

Analysis of molecular variance (AMOVA) of *C. dichroa.*

Source of variation	**d.f.**	**Sum of squares**	**Variance components**	**Percentage of variation**	***F*st**
Among individuals within populations	18	3.92	-0.10	-31.81	
Within individuals	19	8.00	0.42	131.81	
Total	37	11.92	0.32		-0.32

Analysis of molecular variance (AMOVA) of *C. natalensis*

Source of variation	**d.f.**	**Sum of Squares**	**Variance components**	**Percentage of variation**	***F*st**
Among individuals within populations	25	16.65	-0.11	-14.09	
Within individuals	26	23.00	0.88	114.09	
Total	51	39.65	0.77		-0.14

Analysis of molecular variance (AMOVA) of *C. caffra*

Source of variation	**d.f.**	**Sum of Squares**	**Variance components**	**Percentage of variation**	***F*st**
Among individuals within populations	18	10.97	0.19	44.04	
Within individuals	19	4.50	0.24	55.96	
Total	37	15.47	0.42		0.44

Analysis of molecular variance (AMOVA) of *C. heuglini*

Source of variation	**d.f.**	**Sum of Squares**	**Variance components**	**Percentage of variation**	***F*st**
Among individuals within populations	11	15.92	0.14	10.72	
Within individuals	12	14.00	1.17	89.28	
Total	23	29.92	1.31		0.11

Chapter 4

Discussion

Climate change and human-mediated pressures such as urbanisation pose a threat on the genetic diversity of many species and subspecies worldwide (Houghton *et al*., 2001). Species which were once ecologically segregated are now found in areas of sympatry and they have the opportunity to possibly hybridize (Rhymer & Simberloff, 1996; Gholamhosseini *et al*., 2013). DNA barcoding and microsatellite genotyping showed that there is a great degree of relatedness between the five *Cossypha* robin-chat species found in southern Africa, i.e. *C. dichroa*, *C. natalensis*, *C. caffra*, *C. heuglini* and *C. humeralis*. Inferred phenotypic hybridization between two of the *Cossypha* robin-chat species (*C. dichroa* and *C. natalensis*) proposed by Davies *et al.* (2011) was also explored.

4.1. Outcomes of this study

The results of the DNA barcoding analysis using the *CO1* gene illustrated separate lineages and clustering patterns of the five species however, *C. caffra* and *C. heuglini* appear to cluster amongst the other species. *Cossypha caffra* is also seen to have a divergence pattern or further branching which indicates a high degree of intraspecific variation. Both the ML and NJ phylogenetic trees showed a polytomy which formed with the outgroup (*Cercotrichas quadrivirgata*). The polytomy indicates a high degree of similarity between the *Cossypha* species and *Cercotrichas quadrivirgata* which is indicated by a 1% bootstrap value supporting a divergence between the two species (Appendix 1). This was supported by the AMOVA analysis using the seven microsatellite markers, which revealed an Fst value of 0.44 within this population. The Fst value of 0.44 within *C. caffra* was the closest to a value of 1 when compared to the Fst values of the other species (*C. dichroa* = 0.32, *C. natalensis* = 0.14, *C. heuglini* = 0.11) as well as the Fst value interspecifically (0.05). An analysis of the K2P distances revealed that the interspecific distances were no more than 3X the mean intraspecific distance which was 14% in this study. According to Hebert *et al*. (2004) in order to identify separate species based on the *CO1* barcode the interspecific distance should be at least 10X the mean intraspecific distance. This approach in delimiting species proved to be ineffective for this particular study as the results

indicated that there was no separation of the five *Cossypha* species. Implementing the species threshold proposed by Hebert *et al*. (2004) indicated a high degree of relatedness between the five species as well as a possibility of admixture between these populations (Hebert *et al*., 2004; Yang *et al.*, 2016). This study supports the arguments put forth in the literature that *CO1* barcodes cannot distinguish closely related sister species using the 10X rule of among to within species divergence (Moritz & Cicero, 2004; Hickerson *et al.*, 2006; Zink & Barrowclough, 2008).

Analysis of the microsatellite markers used displayed multiple results showing that there were no deviations from Hardy-Weinberg or linkage disequilibrium with P-values greater than 0.05 although null alleles were present in four of the loci. Using seven microsatellite markers STRUCTURE v2.3.4 analysis (Pritchard *et al*., 2000) revealed three distinct clusters within the data set instead of five as expected. *Cossypha heuglini* is seen to have similar alleles to *C. dichroa* and *C. caffra*, whereas *C. humeralis* is seen to have similar alleles to *C. natalensis.* A possible explanation could be that *C. humeralis* and *C. heuglini* may have recently diverged from the three genetically distinct species, however a more likely explanation could be that the markers used were not sufficient in detecting variation between these populations and could not accurately distinguish between the genetics of these five species (Germain-Aubrey *et al*., 2016). A disadvantage of the markers used was only three of the seven markers were polymorphic. According to literature a minimum of five polymorphic markers are used to elucidate conclusive genetic relationships between species (Olowofeso *et al*., 2005; van Wyk *et al.*, 2013; Coetzer *et al.*, 2015; Horreo *et al.*, 2016; Yang *et al.*, 2016).

The sample size used in this study was relatively small for the different species, i.e. *C. natalensis* (n=26), *C. caffra* (n=18), *C. dichroa* (n=18), *C. heuglini* (n=12) and *C. humeralis* (n=4). According to Hale *et al.* (2012) microsatellite based population genetic studies are most accurate with a sample size of 25 to 30 individuals per species. Their study looked at allele frequencies, expected heterozygosities, and pairwise Fst values by randomly subsampling 5-100 individuals from four microsatellite genotype sets. Variability in the results was minimal above sample sizes of 25-30 (Hale *et al.*, 2012). The authors thus concluded that there was no need for sample sizes greater than 25-30 individuals per population for microsatellite based genetic population studies. It can therefore be concluded that the sample sizes used in this study were insufficient and prevented conclusive decisions to be made regarding the genetic relationships of these five *Cossypha* robin-chat species. The results of the study do, however, allow some inferences to be made with regards to the degree of relatedness between the species.

4.2. Effectiveness of DNA barcoding and microsatellites in population genetics studies

Microsatellite analysis was coupled with mtDNA for a number of reasons. The analysis of mtDNA exclusively in determining evolutionary relationships and in phylogeographic studies has become a concern (Yang *et al*., 2016). The maternal inheritance of mtDNA provides a bias portrayal of the overall evolutionary history (Zink & Barrowclough, 2008). Another concern is the rapid evolution of the mitochondrial genes as a single linked unit. Hence, the sequencing of multiple mitochondrial genes would result in a single gene tree which may be incorrect if multiple divergence and speciation events occurred in a short space of time (Hickerson *et al*., 2006; Knowles & Carstens, 2007; Edwards *et al*., 2008; Zink & Barrowclough, 2008). As proposed by Rubinoff and Holland (2005) using multiple, unlinked nuclear loci in conjunction with DNA barcoding could provide a better understanding of the evolutionary history of species. Multiple independent nuclear loci would reduce the error associated with estimating parameters for algorithms used to generate phylogenetic trees (Dolman & Moritz, 2006). Due to the debate around the use of DNA barcoding, microsatellite genotyping has become coupled with this molecular technique (Lu *et al*., 2001; Toews *et al*., 2011; Coetzer *et al*., 2015; Germain-Aubrey *et al*., 2016; Samani *et al*., 2016). Nuclear DNA evolves much slower than mtDNA; however, it provides a larger pool of genetic variation and is not biased with regards to maternal or paternal parentage (Germain-Aubrey *et al.*, 2016). Microsatellites are a popular choice of marker as they are useful in studies of species which have low evolutionary history (Putman & Carbone, 2014).

A study by Yang *et al*. (2016) found discrepancies between mtDNA lineages and microsatellite data. Their study revealed discordance between the two molecular markers attributed to incomplete lineage sorting of ancestral polymorphisms, differences between male and female dispersal rates, and recent admixture. Previous studies conducted in vinous-throated parrotbills supported the Yang *et al*. (2016) study (Ballard & Whitlock, 2004; Qu *et al.*, 2012). Zhang and Hewitt (2003) proposed that without admixture, the mtDNA and microsatellite data would show similar divergence patterns.

DNA barcoding and microsatellite analysis in this study both illustrated separate lineages with regards to *C. natalensis* and *C. dichroa.* The results of the microsatellite genotyping inferred that *C. dichroa* may actually be more closely related to *C. heuglini,* and *C. natalensis* may be more closely related to *C. humeralis*. However, as previously mentioned the use of only three polymorphic markers and the small sample sizes may have skewed the results and not provided the most accurate representation of the genetic divergence of these species. The addition of at least two more polymorphic microsatellite markers would allow the results to be rendered more

accurate as a minimum of five markers are required to render accurate conclusive results with regards to population genetics studies (Olowofeso *et al*., 2005; van Wyk *et al.*, 2013; Woolaver *et al.*, 2013; Coetzer *et al.*, 2015; Horreo *et al.*, 2016; Yang *et al*., 2016).

4.3. Final conclusions drawn from this study

This study demonstrated a similarity between the data obtained from the two molecular techniques used as a high degree of relatedness is evident between the species with both techniques. Separate lineages are depicted for *C. dichroa*, *C. natalensis* and partially for *C. caffra* with both techniques although there are discrepancies with regards to *C. humeralis* and *C. heuglini*, but they were the smallest sample sizes. Of the five *Cossypha* species found in southern Africa *C. dichroa* and *C. natalensis* are not each other's closest relatives despite the fact that they hybridize, with DNA barcoding clearly indicating that the maternal parent of the two inferred hybrids was *C. natalensis*. Based on the assumption that *C. dichroa* and *C. natalensis* were able to hybridize, and that the five species are closely related (based on the specific techniques used), we could predict that all five *Cossypha* species could potentially hybridize. However, there are reproductive isolation mechanisms in place between these species with one of the most important isolation mechanisms between them likely being vocalisations (Kelley *et al*., 2008).

Each *Cossypha* robin-chat species in southern Africa has a distinct call with *C. dichroa*, *C. natalensis* and *C. humeralis* having the ability to mimic other birds. The ability to mimic the songs of different bird species could allow some species to overcome this reproductive isolation mechanism between species (Kelley *et al*., 2008). Based on the possibilities of overcoming reproductive isolation mechanisms such as vocalisations through mimicry, we could predict that in any area where *C. dichroa* and *C. natalensis* co-occur we would be likely to find hybridization (Clancey, 1982; Davies *et al*., 2011). Hybridization between the other *Cossypha* species may be less likely because reproductive isolation is maintained through species specific song and the lack of mimicry.

4.4. Genetic and ecological contribution of this study

This study has provided a deeper understanding of the genetic structure of the five *Cossypha* robin-chat species found in southern Africa and contributed to the effectiveness of using molecular techniques as a valid method of species identification. This study also provided comparative data obtained from nuclear DNA and mitochondrial DNA as well as illustrated that even though a technique such as DNA barcoding is being employed largely for identification

purposes (Hebert *et al.*, 2004; Kerr *et al.*, 2007; Tavares & Baker, 2008; Johnsen *et al*., 2010; Sangster *et al.*, 2010; Heller *et al.*, 2016; Purty & Chatterjee, 2016; Thaler & Stoeckle, 2016; Zarza *et al*., 2016), it may not be the most reliable and/or effective tool (Avise, 1994; Ballard & Whitlock, 2004; Moritz & Cicero, 2004; Rubinoff & Holland, 2005; Bazin *et al.*, 2006; Ruegg, 2008; Zink & Barrowclough, 2008; Edwards & Bensch, 2009).

References

Abd El-Aal, A.A., Abd Elghany, N.A., Mohamadin, A.M. & El-Badry, A. (2010). Comparative study of five methods for DNA extraction from whole blood samples. *International Journal of Health Science*, 3(1), pp. 285-7.

Allendorf, F.W., Leary, R.F., Spruell, P. & Wenburg, J.K. (2001). The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, 16, pp. 613-22.

Amos, W., Hoffman, J.I., Frodsham, A., Zhang, L., Best, S. & Hill, A.V.S. (2006). Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology*, 7, pp. 10-4.

Arnold, M.L. (1997). Natural hybridization and evolution. *Oxford UK: Oxford University Press.*

Arnold, M.L., Bulger, M.R., Burke, J.M., Hampel, A.L. & Williams, J.H. (1999). Natural hybridization: how low can you go and still be important. *Ecology*, 80, pp. 371-81.

Arnold, I., Wagner-Ecker, M., Ansorge, W. & Langer, T. (2006). Evidence for a novel mitochondria-to-nucleus signalling pathway in respiring cells lacking i-AAAprotease and the ABC-transporter Mdl1. *Gene*, 367, pp. 74-88.

Aterburn, J.W. & Grzybowski, J.A. (2003). Hybridization between glossy and white-faced ibises. *North American Birds*, 57, pp. 136-9.

Avise, J.C. (1994). Molecular markers, natural history and evolution. Chapman and Hall, London, U.K.

Avise, J.C. (2004). Molecular markers, natural history, and evolution. Sinauer Ass., Sunderland.

Balch, W.E., Magrum, L.J., Fox, G.E., Wolfe, R.S. & Woese, C.R. (1977). An ancient divergence among the bacteria. *Journal of Molecular Evolution*, 9(4), pp. 305-11.

Ballard, J.W. & Whitlock, M.C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, 13, pp. 729-44.

Balloux, F., Amos, W. & Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology*, 13(10), pp. 3021-31.

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

Baylis, J.R. (1982). Avian vocal mimicry: its function and evolution. *Acoustic Communication in birds*, pp. 51-80.

Bazin, E., Glemin, S. & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, 312, pp. 570-2.

Beresford, P. (2003). Molecular systematics of *Alethe*, *Sheppardia* and some other African robins (Muscicapoidea). *Ostrich*, 74, pp. 58-73.

Bilgin, R., Ebeoğlu, N., Inak, S., Kirpik, M.A., Horns, J.J. & Şekercioğlu, C.H. (2016). DNA barcoding of birds at a migratory hotspot in Eastern Turkey highlights continental phylogeographic relationships. *Plos ONE*, 11(16):e0154454.

Blin, N. & Stafford, D.W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Resources*, 3(9), pp. 2303-8.

Brelsford, A., Mila, B. & Irwin, D.E. (2011). Hybrid origin of Audubon's warbler. *Molecular Ecology*, 20(11):2380-9, DOI 10.1111/j.1365-294X.2011.05055.x.

Brody, J.R. & Kern, S.E. (2004). History and principles of conductive media for standard DNA electrophoresis. *Analytical Biochemistry*, 333(1), pp. 1-13.

Brooks, M. (2016). South African Bird Atlas Project 2. [Online] Available from: http://sabap2.adu.org.za [Accessed: 20 November 2016].

Brown, W.M., George, M. & Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, 76, pp. 1967-71.

Buschiazzo, E. & Gemmell, N.J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays*, 28(10), pp. 1040-50.

Catchpole, C.K. (1980). Sexual selection and the evolution of complex songs among European warblers of the genus *Acrocephalus*. *Behaviour*, 74, pp. 149-66.

Catchpole, C.K. & Baptista, L.E. (1988). A test of the competition hypothesis of vocal mimicry, using song sparrow imitation of white-crowned sparrow song. *Behaviour*, 106, pp. 119-28.

Catchpole, C.K. & Slater, P.J.B. (2008). Bird song: biological themes and variations. *Cambridge University Press, United Kingdom.*

Cha, R.S. & Thilly, W.G. (1993). Specificity, efficiency, and fidelity of PCR. *PCR Methods and Applications*, 3(3), pp. S18–29.

Chamberlain, J.S., Gibbs, R.A., Ranier, J.E., Nguyen, P.N. & Caskey, C.T. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Resources*, 16(23), pp. 11141-56.

Clancey, P.A. (1982). On the robins *Cossypha dichroa* and *Cossypha natalensis* (Aves: Turdidae) in southern Africa. *Bonner Zoologische Beitrage*, 33, pp. 293-302.

Cler, L., Bu, D., Lewis, C., & Euhus, D.(2006). A comparison of five methods for extracting DNA from paucicellular clinical samples. *Molecular and Cellular Probes*, 20(3-4), pp. 191-6.

Coetzer, W.G., Downs, C.T., Perrin, M.R. & Willows-Munro, S. (2015). Molecular systematics of the Cape Parrot (*Poicephalus robustus*): Implications for taxonomy and conservation. *Plos ONE*, DOI 10.1371/journal.pone.0133376.

Coyne, J.A. & Orr, H.A. (1997). Patterns of speciation in *Drosophila melanogaster* revisited. *Evolution*, 51, pp. 295-303.

Cracraft. J. (1989). Species, speciation, and phylogenetic systematics. *Cladistics*, 6(2), pp. 205-9.

Crochet, P.A., Chen, J.J. Z., Pons, J.M., Lebreton, J.D., Hebert, P.D.N. & Bonhomme, F. (2003). Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? *Evolution*, 57, pp. 2865-78.

Davies, G., Symes, C., Boon, R.G.C. & Campbell, H.A. (2011). Inferred hybridization, sympatry and movements of Chorister Robin-Chat *Cossypha dichroa* and Red-capped Robin-Chat *C. natalensis*. *Ostrich*, 82(3), pp. 231-41.

Davis, C.S., Thwaites, F.T., Gallager, S.M. & Hu, Q. (2005). A three-axis fast tow digital video plankton recorder for rapid surveys of plankton taxa and hydrography. *Limnology and Oceanography: Methods*, 3, pp. 59-74.

De Queiroz, K. & Donoghue, M.J. (1990). Phylogenetic systematics and species revisited. *Cladistics*, 6(1), pp. 61-75.

Dobkin, D.S. (1979). Functional and evolutionary relationships of vocal copying phenomena in birds. *Ethology*, 50(4), pp. 348-63.

Dolman, G. & Moritz, C. (2006). A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). *Evolution*, 60, pp. 573-82.

Edwards, D., Horn, A., Taylor, D., Savolainen, V. & Hawkins, J.A. (2008). DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). *Taxon*, 57, pp. 1317-27.

Edwards, S. & Bensch, S. (2009). Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*, 18, pp. 2930-3.

Ellegran, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, 16(12), pp. 551-8.

Emery, N.J. & Clayton, N.S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science*, 306(5703), pp. 1903-7.

Erlich, H. (1989). Polymerase chain reaction. *Journal of Clinical Immunology*, 9(6), pp. 437–47.

Excoffier, L.G., Laval, G. & Schneider, S. (2005). Arlequin Ver. 3.0 : An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, pp. 47-50.

Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, 35, pp. 124-38.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution***,** pp. 783-91.

Ferguson, J.W.H., van Zyl, A. & Delport, K. (2002). Vocal mimicry in African *Cossypha* robin-chats. *Journal of Ornithology*, 143, pp. 319-30.

Genovart, M. (2008). Natural hybridization and conservation. *Biodiversity and Conservation*, 18(6), pp. 1435-9, DOI 10.1007/s10531-008-9550-x.

Germain-Aubrey, C.C., Nelson, C., Soltis, D.E., Soltis, P.S. & Gitzendanner, M.A. (2016). Are microsatellite fragment lengths useful for population-level studies? The case of *Polygala lewtonii* (Polygalaceae). *Applications in Plant Sciences*, 4(2), DOI 10.3732/apps.1500115.

Gholamhosseini, A., Vardakis, M., Aliabadian, M., Nijman, V. & Vonk, R. (2013). Hybridization between sister taxa versus non-sister taxa: A case study in birds. *Bird Study*, 60(2), pp. 195-201, DOI 10.1080/00063657.2013.770815.

Gippoliti, S. & Amori, G. (2007). The problem of subspecies and biased taxonomy in conservation lists: the case of mammals. *Folia Zoologica*, 56, pp. 113-7.

Grant, P.R. & Grant, B.R. (1992). Hybridization of bird species. *Science*, 256, pp. 193-7.

Grant, P.R. & Grant, B.R. (1997). Genetics and the origin of bird species. *Proceedings of the National Academy of Sciences, USA*, 94, pp. 1168-7775.

Gunning, J.W.B. (1909). A new species of *Cossypha* from west Pondoland. *Annals of the Transvaal Museum*, 1, pp. 174.

Haag-Liautard, C., Coffrey, N., Houle, D., Lynch, M., Charlesworth, B. & Keightley, P.D. (2008). Direct estimation of the mitochondrial DNA mutation rate in *Drosphila melanogaster*. *PLoS Biology*, 6(8):e204.

Hajibabaei, M., deWaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T. & Kirk, S.L. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society*, 360, pp. 1959-67.

Hale, M.L., Burg, T.M. & Steeves, T.E. (2012). Sampling for microsatellite-based population genetic studies: 25-30 individuals per population is enough to accurately estimate allele frequencies. *PLoS ONE*, 7(9): e54170, DOI 10.1371/journal.pone.0045170.

Harrison, J.A., Allan, D.G., Underhill, L.G., Herremans, M., Tree, A.J., Parker, V. & Brown, C.J. (1997). The atlas of southern African birds. *Johannesburg BirdLife South Africa.*

Hartshorne, C. (1956). The monotony threshold in singing birds. *Auk*, 73, pp. 176-91.

Hayes, F.N. & Williams, M. (1982). The status, aviculture and re-establishment of Brown Teal in New Zealand. *Wildfowl*, 33, pp. 73-80.

Hebert, P.D., Cywinska, A. & Ball, S.L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270**,** pp. 313-21.

Hebert, P.D., Stoeckle, M.Y., Zemlak, T.S. & Francis, C.M. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, 2**,** pp. e312.

Heller, E.L., Kerr, K.C.R., Dahlan, N.F., Dove, C.J. & Walters, E.L. (2016). Overcoming challenges to morphological and molecular identification of *Empidonax* flycatchers: a case study with a Dusky Flycatcher. *Journal of Field Ornithology*, 87(1), pp. 96-103.

Hewitt, G.M. (1993). Postglacial distribution and species sub-structure: lessons from pollen, insects and hybrid zones. *Evolutionary Patterns and Processes*, pp. 97-123.

Hickerson, M.J., Meyer, C.P. & Moritz, C. (2006). DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, 55, pp. 729-39.

Hockey, P.A.R., Dean, W.R.J. & Ryan, P.G. (2005). Roberts - Birds of southern Africa, VIIth ed. *The Trustees of the John Voelcker Bird Book Fund*, Cape Town.

Hoelzer, G.A. & Melnick, D.J. (1994). Patterns of speciation and limits to phylogenetic resolution. *Tree*, 9, pp. 104-7.

Holsinger, K.E. & Bruce, S.W. (2009). Genetics in geographically structured populations: defining, estimating and interpreting Fst. *Nature Reviews Genetics*, 10(9), pp. 639-50.

Holt, R.D. (1990). The microevolutionary consequences of climate change. *Trends in Ecology and Evolution*, 5, pp. 311-5.

Horreo, J.L., Raab, R., Spakovszky, P. & Alonso, J.C. (2016). Genetic structure of the threatened West-Pannonian population of Great Bustard (*Otis tarda*). *PeerJ*, DOI 10.7717/peerj.1759.

Houghton, J.T., Ding, Y., Griggs, D,J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K. & Johnsen, C.A. (2001). Climate change 2001. *Cambridge University Press, United Kingdom* -ISBN 0521807670.

Igic, B., McLachlan, J., Lehtinen, I. & Magrath, R.D. (2015). Crying wolf to a predator: deceptive vocal mimicry by a bird protecting young. *Proceedings of the Royal Society B*, 282: 20150798.

Jarne, P. & Logoda, P.J.L.(1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, 11, pp. 424-9.

Jarvis, E.D.S., Mirarab, A.J., Aberer, B., Li, P., Houde, C., Li, S.Y.W., Ho, B.C., Faircloth, B., Nabholz, J.T. & Howard, A. (2014). Whole-genome analysis resolve early branches in the tree of life of modern birds. *Science*, 346, 1320-31.

Johnsen, A., Rindal, E., Ericson, P.G.P., Zuccon, D., Kerr, K.C.R. & Stoeckle, M.Y. (2010). DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. *Journal of Ornithology*, 151, pp. 565-78.

Jones, P.D., Osborn, T.J. & Briffa, K.R. (2001). The evolution of climate over the last millennium. *Science*, 292(5517), pp. 662-7.

Jordaens, K., Goergen, G., Virgilio, M., Backeljau, T., Vokaer, A. & De Mayer, M. (2015). DNA barcoding to improve the taxonomy of the Afrotropical Hoverflies (Insecta: Diptera: Syrphidae). *PLoS ONE*, 10(10): e0140264, DOI 10.1371/journal.pone.0140264.

Joseph, L. & Buchanan, K.L. (2015). A quantum leap in avian biology. *Emu*, 115, pp. 1-5.

Karger, A.E., Harris, J.M. & Gesteland, R.F. (1991). Multiwavelength fluorescence detection for DNA sequencing using capillary electrophoresis. *Nucleic Acids Research*, 19**,** pp. 4955-62.

Kelley, L.A., Coe, R.L., Madden, J.R. & Healy, S.D. (2008). Vocal mimicry in songbirds. *Animal Behaviour*, 76, pp. 521-8.

Kerr, K.C.R., Stoeckle, M.Y., Dove, C.J., Weigt, L.A., Franciss, C.M. & Hebert, P.D.N. (2007). Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, 7, pp. 535-43.

Kerr, K., Birks, S.M., Kalyakin, M.V., Redkin, Y.A., Koblik, E.A. & Hebert, P.D.N. (2009). Filling the gap - *CO1* barcode resolution in eastern Palearctic birds. *Frontiers in Zoology*, 6, pp. 10.

Kimura, M.(1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16**,** pp. 111-20.

Knowles, L.L. & Carstens, B.C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, 56(6), pp. 887-95.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., Mcwilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2007). Clustal W and Clustal X Version 2.0. *Bioinformatics*, 23**,** pp. 2947-8.

Leukering, T. (2008). Glossy ibis and the identification challenges of hybrid *Plegadis* in Colorado. *Colorado Birds*, 42(2), pp. 147-9.

Li, Y,C., Korol, A.B., Fahima, T., Beiles, A. & Nevo, E. (2002). Microsatellites: genomic distribution, putative functions, and mutational mechanisms: A review. *Molecular Ecology*, 11, pp. 2453-65.

Lin, G.N., Zhang, C. & Xu, D. (2010). Polytomy identification in microbial phylogenetic reconstruction. *BMC Systems Biology*, 5(Suppl. 3): S2.

Loparev, V.N., Cartas, M.A., Monken, C.E., Velpandi, A. & Srinivasan, A. (1991). An efficient and simple method of DNA extraction from whole blood and cell lines to identify infectious agents. *Journal of Virological Methods*, 34(1), pp. 105-12.

Lu, G., Basley, D.J. & Bernatchez, L. (2001). Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, 10, pp. 965-85.

Ma, W. & Lambert, D. (1997). Minisatellite DNA markers reveal hybridization between the endangered Black Robin and Tomtit. *Electrophoresis*, 18**,** pp. 1682-7.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, 20, pp. 229-37.

Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B*, 363, pp. 2971-86.

Markoulatos. P., Siafakas, N. & Moncany, M. (2002). Multiplex polymerase chain reaction: A practicla approach. *Journal of Clinical Laboratory Analysis*, 16, pp. 47-51.

Maynard, S.J. (1996). Sympatric speciation. *American Naturalist*, 100, pp. 637-50.

McCracken, G.R., Wilson, K.L., Brewis, H.T., Mcbride, M.C., Paterson, I., Perry, R., Keefe, D. & Ruzzante, D.E. (2014). Development of 26 novel microsatellite markers for the round whitefish (*Prosopium cylindraceum*) and successful polymorphic cross-specific amplification of seven previously developed salmonid markers. *Conservation Genetic Resources*, 6(4), pp. 1023-6.

McKenzie, D. (1998). Chorister-Natal Robin hybrid. *Blue Swallow*, 11(3), pp. 17.

McKinney, M.A. & Drake, J.A. (1998). Biodiversity Dynamics. *Turnover of Populations, Taxa, and Communities*, 137(3), pp. 341.

Morgante, M. & Olivieri, A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. *Plant Journal*, 3, pp. 175-82.

Moritz, C. (1994). Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, 3(4), pp. 401-11.

Moritz, C. & Cicero, C. (2004). DNA barcoding: Promise and pitfalls. *PLoS Biology*, 2(10), pp. 1529-31.

Muńoz-Fuentes, V., Vilá, C., Green, A.J., Negro, J.J. & Sorenson, M.D. (2007). Hybridization between white-headed ducks and introduced ruddy ducks in Spain. *Molecular Ecology*, 16, pp. 629-38.

Olowofeso, O., Wang, J.Y., Dai, G.J., Yang, Y., Mekki, D.M. & Musa, H.H. (2005). Measurement of genetic parameters within and between Haimen chicken populations using microsatellite markers. *International Journal of Poultry Science*, 4(3), pp. 143-8.

Olsen, G.J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994). FastDNml: a tool for construction of phylogenetic trees of DNA sequences using Maximum likelihood. *Computational Applied Biosciences*, 10(1), pp. 41-8.

Ottenburghs, J., van Hooft, P., van Wieren, S.E., Ydenberg, R.C. & Prins, H.H.T. (2016a). Birds in a bush: Toward an avian phylogenetic network. *The Auk*, 133, pp. 577-82.

Ottenburghs, J., van Hooft, P., van Wieren, S.E., Ydenberg, R.C. & Prins, H.H.T. (2016b). Hybridization in geese: a review. *Frontiers in Zoology*, 13:20, DOI 10.1186/s12983-016-0153-1.

Palti, Y., Genet, C., Gao, G., Hu, Y., You, F.M., Boussaha, M., Rexroad, C.E. & Luo, M.C. (2012). A second generation integrated map of the rainbow trout (*Oncorhynhus mykiss*) genome: analysis of conserved synteny with model fish genomes. *Marine Biotechnology*, 14(2), pp. 343-57.

Peterson, K.J. & Eernisse, D.J. (2015). The phylogeny, evolution developmental biology, and paleobiology of the Deuterostomia: 25 years of new techniques, new discoveries, and new ideas. *Organisms Diversity & Evolution*, DOI 10.1007/s13127-016-0270-x.

Phillimore, A.B. & Owens, P.F. (2006). Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society Biological Sciences*, 273(1590), pp. 1049-53, DOI 10.1098/rspb.2005.3425.

Priest, M.A., Almany, G.R., Braun, C.D., Hamilton, R.J., Lozano-Cortes, D.F., Saenz-agudelo, P. & Beruman, M.L. (2014). Isolation and characterisation of 29 microsatellite markers for the bumphead parrotfish, *Bolbometopon muricatum*, and cross amplification in 12 related species. *Marine Biodiversity*, 44(4), pp. 1-6.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, pp. 945-59.

Purty, R.S. & Chatterjee, S. (2016). DNA barcoding: An effective technique in molecular taxonomy. *Austin Journal of Biotechnology & Bioengineering*, 3(1), pp. 1059-66.

Putman, A.I. & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, 4, pp. 4399-428.

Qu, Y., Zhang, R., Quan, Q., Song, G., Li, S.H. & Lei, F. (2012). Incomplete lineage sorting or secondary admixture disentangling historical divergence from recent gene flow in the Vinous-throated parrotbill (*Paradoxornis webbianus*). *Molecular Ecology*, 21, pp. 6117-33.

Randler, C. & Bogner, F. (2002). Comparing methods of instruction using bird species identification skills as indicators. *Journal of Biological Education*, 36, pp. 181-8.

Rheindt, F.E. & Edwards, S.V. (2011). Genetic introgression: An integral but neglected component of speciation in birds. *The Auk*, 128**,** pp. 620-632.

Rhymer, J.M. & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology, Evolution and Systematics*, 27, pp. 83-109.

Roberts, A. (1914). Notes on birds in the collection of the Transvaal Museum with descriptions of several new sub-species. *Annals of the Transvaal Museum*, 4(4), pp. 169-79.

Rubinoff, D. & Holland, B.S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, 54, pp. 952-61.

Robinson, F.N. (1975). Vocal mimicry and the evolution of bird song. *Emu*, 75, pp. 23-7.

Ruegg, K. (2008). Genetic, morphological, and ecological characterization of a hybrid zone that spans a migratory divide. *Evolution*, 62, pp. 542-66.

Saetre, G.P., Borge, T. & Lindell, J. (2001). Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. *Molecular Ecology*, 10, pp. 737-49.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4**,** pp. 406-25.

Samani, N.K., Esa, Y., Nurul Amin, S.M. & Ikhsan, N.F.M. (2016). Phylogenetics and population genetics of *Plotosus canius* (Siluriformes: Plotosidae) from Malaysian coastal waters. *PeerJ*, 4:e1930, DOI 10.7717/peerj.1930.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74**,** pp. 5463-7.

Sangster, G., Alstrom, P., Forsmark, E. & Olsson, U. (2010). Multi-locus phylogenetic analysis of Old World chats and flycatchers reveals extensive paraphyly at family, subfamily and genus level (Aves: Muscicapidae). *Molecular Phylogentics and Evolution*, 57, pp. 380-92.

Schlotterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109, pp 365-71.

Schlotterer, C. (2004). The evolution of molecular markers – just a matter of fashion ? *Nature Reviews Genetics*, 5, pp. 63-9.

Searcy, W.A. (1984). Song repertoire and female preferences in song sparrows. *Behavioural Ecology Sociobiology*, 14, pp. 281-6.

Seehausen, O., Takimoto, G., Roy, D. & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, 17, pp. 30-44.

Selkoe, K.A. & Toonen, R.J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, 9, pp. 615-29.

Semo, L.S. (2007). The 44th report of the Colorado bird records committee. *Colorado Birds*, 41, pp. 241-58.

Sinclair, I., Hockey, P.A.R. & Arlott, N. (2005). The larger illustrated guide to birds of southern Africa. *Nature,* ISBN 1770072438.

Smith, L.M. (1993). DNA electrophoresis. *Current Opinion in Biotechnology*, 4(1), pp. 37-40.

Sobel, J.M., Chen, G.F., Watt, L.R., & Schemske, D.W. (2010). The biology of speciation. *Evolution*, 64(22), pp. 295-315.

Sturge, R.J., Cortés-Rodriguez, M.N., Rojas-Soto, O.R. & Omland, K.E. (2016). Nuclear locus divergence at the early stages of speciation in the Orchard Oriole complex. *Ecology and Evolution*, 6(13), pp. 4307-17.

Tamura, K., Stecher, G. & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: pp. 2725-9.

Tavares, E.S. & Baker, A.J. (2008). Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *BMC Evolutionary Biology*, 8, pp. 81-94.

Taylor, H.R. & Harris, W.E. (2012). An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, 12(3), pp. 377-88.

Thaler, D.S. & Stoeckle, M.Y. (2016). Bridging two scholarly islands enriches both: *CO1* DNA barcodes for species identification versus human mitochondrial variation for the study of migrations and pathologies. *Ecology and Evolution*, 6(19), pp. 6824-35, DOI 10.1002/ece3.2394.

Toews, D.P.L., Brelsford, A. & Irwin, D.E. (2011). Hybridization between Townsend's *Dendroica townsendi* and Black-Throated Green Warblers *D. virens* in an Avian Suture Zone. *Journal of Avian Biology*, 42**,** pp. 434-46.

Vali, U., Saag, P., Dombrovski, D., Meyburg, B.U., Maciorowski, G., Mizera, T., Treinys, R. & Fagerberg, S. (2010). Microsatellites and single nucleotide polymorphisms in avian hybrid identification: a comparative case study. *Journal of Avian Biology*, 41, pp. 34-49.

Van Oosterhout, C., Hutchinson, W.E., Wills, D.P.M. & Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3), pp. 535-8

Van Wyk, A.M., Kotze, A., Randi, E., & Dalton, D.L. (2013). A hybrid dilemma: a molecular investigation of South African bontebok (*Damaliscus pygargus*) and blesbok (*Damaliscus pygargus phillipsi*). *Conservation Genetics*, 14, pp. 589-99, DOI 10.1007/s10592-013-0448-0.

Viard, F., Franck, P., Dubois, M.P., Estoup, A. & Jarne, P. (1998). Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *Journal of Molecular Evolution*, 47, pp. 42-51.

Webster, M.D., Babiker, M., Mayer, M., Reilly, J.M., Harnisch, J., Sarofim, M.C. & Wang, C. (2002). Uncertainty in emissions projections for climate models. *Atmospheric Environment*, 36(22), pp. 3659-70.

Wiegers, U. & Hilz, H. (1971). A new method using 'proteinase K' to prevent mRNA degradation during isolation from HeLa cells. *Biochemical Biophysical Research Communications*, 16(44:2), pp. 513-9.

Will, K.W. & Rubinoff, D. (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, 20, pp. 47-55.

Wimmers, K., Ponsuksili, S., Hardge, T., Valle-Zarate, A., Mathur, P.K. & Horst, P. (2000). Genetic distinctness of African, Asian and South American local chickens. *Animal Genetics*, 31(3), pp. 159-65.

Woese, C.R. (2000). Interpreting the universal phylogenetic tree. *PNAS*, 97(15), pp. 8392-6.

Wogan, G.O.U., Feldheim, K., Voelker, G., & Bowie, R.C.K. (2015). Microsatellite markers for the Cape robin-chat (*Cossypha caffra*) and the Red-Capped robin-chat (*Cossypha natalensis*) for use in demographic and landscape genetics analyses. *Conservation Genetics Resource*, 7(1), pp. 151-54, DOI 10.1007/s12686-014-0315-4.

Woolaver, L.G., Nicols, R.K., Mortan, E.S. & Stutchbury, B.J.M. (2013). Population genetics and relatedness in a critically endangered island, Ridgeway's Hawk *Buteo ridgwayi*. *Conservation Genetics*, 14, pp. 559-71.

Yang, J., Hsu, K., Liu, Z., Su, L., Kuo, P., Tang, W., Zhou, Z., Liu, D., Bao, B. & Lin, H. (2016). The population history of *Garra orientalis* (Teleostei: Cyprinidae) using mitochondrial DNA and microsatellite data with approximate Bayesian computation. *BMC Evolutionary Biology*, 16(73), DOI 10.1186/s12862-016-0645-9.

Yoo, H.Y., Eah, J., Kim, J.S., Young, J.K., Min, M. & Peak, W.K. (2006). DNA barcoding Korean birds. *Molecules and Cells*, 22, pp. 323-7.

Zarza, E., Faircloth, B.C., Tsai, W.L.E., Bryson, R.W., Klicka, J. & McCormack, J.E. (2016). Hidden histories of gene flow in highland birds revealed with genomic markers. *Molecular Ecology*, 25, pp. 5144-57.

Zhang, D.X. & Hewitt, G.M. (2003). Nulear DNA analyses in genetic studies of populations practice, problems and prospects. *Molecular Ecology*, 12, pp. 563-84.

Zink, R.M. (2004). The role of subspecies in obscurring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 271, pp. 561-4.

Zink, R.M. & Barrowclough, G.F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, 17, pp. 2107-719.