Impact of hybridisation in two *Cossypha* robin-chat species in southern Africa

Jean Mollett [Corresp., Equal first author, 1], **Naadhirah Munshi** [Equal first author, 1], **Craig Symes** [2]

[1] Science Faculty, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, South Africa

[2] Science Faculty, School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa

Corresponding Author: Jean Mollett
Email address: Jean.Mollett@wits.ac.za

Chorister Robin-Chat *Cossypha dichroa*, a South African forest endemic, and Red-capped Robin-Chat *C. natalensis*, a widely distributed species in African forest and woodland, are inferred to hybridise in areas of sympatry. DNA was extracted from blood samples of *C. dichroa* (n = 18), *C. natalensis* (n = 47), and two phenotypic hybrids. The mitochondrial cytochrome c oxidase I (COI) gene was amplified by PCR and sequenced. Phylogenetic analysis was performed on the sequence data to investigate taxonomic status and putative interspecific hybridisation. Phenotypic hybrids grouped with *C. natalensis*, suggesting maternal parentage from that species. Intra- and interspecific genetic and geographic distances were compared between *C. dichroa* and *C. natalensis* to assess genetic introgression. Seven of the thirteen microsatellite primer pairs developed for *C. natalensis* cross amplified in *C. dichroa*. These seven markers were then used for further analysis. STRUCTURE v2.3.4 was used to assign individuals to a particular genetic cluster and determine any admixture. NEWHYBRIDS v1.1 was used to assign hybrid status to samples beyond the F1 generation. Despite the hybridisation events recorded between *C. dichroa* and *C. natalensis* they still form two separate clusters as expected, and two genetic clusters (K=2) were identified using STRUCTURE. These two species are proficient vocal mimics and it is likely that reproductive isolation mechanisms are overcome through vocalisations. Genotypic hybrids are evident in the sampled population and hybridisation and backcrossing across a zone of sympatry is occurring. However, hybridisation is expected to have very little evolutionary influence on the integrity of recently diverged species which retain reproductive isolation across a wide region of sympatry through call distinctness.

Impact of hybridization in two *Cossypha* robin-chat species in southern Africa

Jean Mollett[1], Naadhirah Munshi[1] and Craig Symes[2]

[1]School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, South Africa
[2]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa

Corresponding author:
Jean Mollett[1]
ORCID: 0000-0003-1057-0114
1 Jan Smuts Avenue, Braamfontein, 2000, Johannesburg, South Africa.
Email address: jean.mollett@wits.ac.za

Abstract

Chorister Robin-Chat *Cossypha dichroa*, a South African forest endemic, and Red-capped Robin-Chat *C. natalensis*, a widely distributed species in African forest and woodland, are inferred to hybridise in areas of sympatry. DNA was extracted from blood samples of *C. dichroa* (n = 18), *C. natalensis* (n = 47), and two phenotypic hybrids. The mitochondrial cytochrome c oxidase I (COI) gene was amplified by PCR and sequenced. Phylogenetic analysis was performed on the sequence data to investigate taxonomic status and putative interspecific hybridisation. Phenotypic hybrids grouped with *C. natalensis*, suggesting maternal parentage from that species. Intra- and interspecific genetic and geographic distances were compared between *C. dichroa* and *C. natalensis* to assess genetic introgression. Seven of the thirteen microsatellite primer pairs developed for *C. natalensis* cross amplified in *C. dichroa*. These seven markers were then used for further analysis. STRUCTURE v2.3.4 was used to assign individuals to a particular genetic cluster and determine any admixture. NEWHYBRIDS v1.1 was used to assign hybrid status to samples beyond the F1 generation. Despite the hybridisation events recorded between *C. dichroa* and *C. natalensis* they still form two separate clusters as expected, and two genetic clusters (K=2) were identified using STRUCTURE. These two species are proficient vocal mimics and it is likely that reproductive isolation mechanisms are overcome through vocalisations. Genotypic hybrids are evident in the sampled population and hybridisation and backcrossing across a zone of sympatry is occurring. However, hybridisation is expected to have very little evolutionary influence on the integrity of recently diverged species which retain reproductive isolation across a wide region of sympatry through call distinctness.

Introduction

Fifteen species of *Cossypha* robin-chat (Aves: Muscicapidae), distributed throughout sub-Saharan Africa, are currently recognised (BirdLife, 2016), five of which occur in southern Africa (Hockey *et al*., 2005; 2011). In 1909, *Cossypha haagneri* was described from the Eastern Cape (Gunning, 1909). Years later this individual, together with a number of other aberrant individuals in museum collections, was phenotypically identified as a Chorister Robin-Chat *Cossypha dichroa* x Red-capped Robin-Chat *Cossypha natalensis* hybrid (Clancey, 1981; 1982; 1991). More inferred hybrids have since been phenotypically identified and it has been suggested that this phenomenon may be more common than is currently recognised (Clancey, 1986; Davies *et al*., 2011; Symes, 2011; Figure 1). *Cossypha dichroa* and *C. natalensis* have long been recognised as closely related species that have overlapping ranges but prefer different vegetation types (Oatley and Arnott, 1998; Hockey *et al.,* 2005). *Cossypha dichroa*, endemic to South Africa, inhabits cool montane forests, whilst *C. natalensis*, inhabiting dense thickets and coastal forest, has a broader range into East and Central Africa (Oatley and Arnott, 1998; Hockey *et al.*, 2005). Inferred hybrids were found to have a variable phenotype intermediate between the two parental species (Davies *et al*., 2011; Figure 1E) and the probability of phenotypically identifying backcrosses, which may resemble F1 generation hybrids or the parental species, may be low or impossible (Avise, 2004). In each of *C. dichroa* and *C. natalensis*, sub-species status has been proposed for a number of taxa in the sub-region. While this may have little relevance for taxonomic introgression between the two species, the different distributions and proposed (or inferred) movement patterns of different sub-species may have bearing on intra-species genetic flow, and the likelihood of inter-species hybridisation.

Hybridisation in the wild is considered to have many important taxonomic and conservation implications (Jetz *et al.*, 2007; Price, 2008). In cases where hybridisation and backcrossing of interspecific hybrids into the parent populations is prevalent it may have an impact on species diversity (Rhymer and Simberloff, 1996; Allendorf *et al*., 2001). For example, the introduction of Mallard (*Anas platyrhynchos*) in New Zealand has led to the decline of the endemic Grey Duck (*A. superciliosa*) (Taysom *et al.*, 2014). Mitochondrial DNA (mtDNA) sequences, used to identify the maternal parent species, i.e. female partner, in interspecific pairing, are particularly useful in hybridization studies. The mitochondrial cytochrome c oxidase I (*COI*) gene is the standard DNA barcode for animals (Hebert *et al*., 2003; Jordaens *et al.*, 2015). 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), resulting in similar barcodes amongst different species due to introgression. Introgressive hybridisation may therefore result in specimens morphologically identified as one parental species having the *COI* haplotype of the other hybridising species (Hebert *et al*., 2004; Toews *et al.*, 2011). Introgression, and therefore sharing of DNA barcode haplotypes, generally only affects individuals in the vicinity of the hybrid zone where the species co-occur, while individuals of the parental species in allopatric populations remain “pure” and genetically unaffected (Rheindt and Edwards, 2011). While *COI* barcoding cannot provide indisputable evidence of evolutionary relationships it is a valuable tool to direct attention to species which require further taxonomic analysis or to support conclusions derived by conventional taxonomic methods (Hebert *et al*., 2004).

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 are similarly a popular tool for 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 and 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 and 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 and Toonen, 2006). 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).

Given the morphological, ecological and behavioural differentiation between *C. dichroa* and *C. natalensis* in South Africa (Farkas, 1969; Oatley, 2005a, b; Davies et al., 2011), and the identification of hybridisation between taxa, we aimed to assess whether genetic data substantiate or invalidate hybridisation between these two species. Overall, the objectives were fourfold; i) to construct a phylogenetic tree based on *COI* barcode divergence in order to determine the evolutionary relationships between *C. dichroa*, *C. natalensis* and their inferred hybrids, ii) to compare intra- and interspecific genetic and geographic distances in order to assess whether hybridisation between *C. dichroa* and *C. natalensis* is resulting in genetic introgression, iii) to test Clancey's (1981 and 1991) hypotheses on geographical variation in *C. dichroa* and geographical structuring (subspeciation) within *C. natalensis* in the southern African region, and, iv) to investigate whether previously developed microsatellite markers specific for *C. natalensis* can be used to determine population structure in these species.

Insert Figure 1

Materials & Methods

Collection of robin-chat blood samples

Robin-chats were captured using mist-nets during March to December 2010 and blood samples collected (Davies *et al*., 2011). A blood sample (c. 60 µl) was withdrawn from each captured robin-chat specimen via venepuncture of the brachial vein and stored in 70-80% ethanol (Davies *et al.*, 2011). All sampled birds were weighed, measured, ringed, and released (Davies *et al.*, 2011). Bird capture was conducted under permit from Ezemvelo KZN Wildlife (OP1785/2010), Limpopo Province (001-CPM401-00005/2012-2013) and ethics clearance from the University of the Witwatersrand, Animal Ethics Screening Committee: blood and feather collection (2009.42/2A). This study included blood samples collected from, i) an allopatric *C. dichroa* population at New Forest, KwaZulu-Natal (nfa) and ii) allopatric *C. natalensis* populations at Twin Streams, KwaZulu-Natal (TW and Arabic numerals between 487 to 548); Wits Rural Facility, Mpumalanga (WR); Pullen farm, Mpumalanga (Pu), and Inhamitanga Forest, Mozambique (M). Also, blood samples of *C. dichroa, C. natalensis* and their inferred hybrids

were collected from sympatric populations occurring in the Soutpansberg, Limpopo (S), and Vernon Crookes Nature Reserve, KwaZulu-Natal (A and B). The geographical reference to all these sites together with sample sizes can be seen in Table 1.

Insert Table 1

Phenol-chloroform based extraction of total DNA from blood samples

Total DNA was extracted from blood samples via the phenol-chloroform based DNA extraction method originally designed by Blin and Stafford (1976). The ethanol in which the blood samples were stored was evaporated. The blood was then resuspended in Queen's lysis buffer (Seutin *et al.*, 1991). The samples were vortexed and centrifuged to collect the lysed cells (Loparev *et al.*, 1991). STE buffer was added to further lyse the cells along with SDS, RNase and proteinase K (Wiegers and Hilz, 1971; Loparev *et al.*, 1991). The solution was incubated for 2 hours at 50 °C for optimal activation of the proteinase K and RNase enzymes (Cler *et al.*, 2006). Phenol-chloroform was added. The solution was vortexed and centrifuged (13,000 rpm for 8 minutes) which separated the mixture into a lower organic and an upper aqueous phase (Loparev *et al.,* 1991). The aqueous phase containing the nucleic acids was recovered and the phenol-chloroform step was repeated on the remaining solution to maximise DNA yield (Cler *et al.*, 2006). DNA was 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). The resultant DNA solution was allowed to solvate overnight at 4 °C. The concentration and purity of DNA extracted was determined with the use of a Nanodrop 1000 and the solution was subsequently stored at 4 °C (Cler *et al.*, 2006).

Polymerase chain reaction to amplify the *COI* DNA barcode region

The polymerase chain reaction (PCR) was carried out according to the protocol presented in Hebert *et al.* (2004). The PCR primers used were developed by Hebert *et al.* (2004) to amplify a fragment of the 5′ terminus of the *COI* gene. The BirdF1-TTCTCCAACCACAAAGACATTGGCAC and BirdR1-ACGTGGGAGATAATTCCAAATCCTG primers amplify an approximately 750 bp region of the *COI* gene in most bird species. Each 50 µl PCR reaction included 2X KAPATaq Ready Mix (25 µl) (Lasec), the forward and reverse primers (0.1 µM each), template DNA (5 ng/µl) and nuclease free water. The PCR cycle consisted 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 synthesised fragments (Hebert *et al.*, 2004). A positive control PCR contained primers to amplify the chromo-helicase-DNA-binding (*CHD*) gene used routinely for molecular sexing of birds (Fridolfsson and Ellegren, 1999). A negative control PCR contained all reaction components except template DNA. The specificity of the PCR reaction was analysed by gel electrophoresis (Erlich, 1989).

***COI* barcode sequencing**

The *COI* barcodes amplified by PCR were sequenced by Inqaba Biotec™ (Johannesburg, South Africa) using computer-automated high-throughput DNA sequencing to sequence the *COI*

barcodes. Computer software, FinchTV, was then used to interpret the data into a DNA sequence (Geospiza Inc.). Sequences generated by this study were deposited on GenBank: accession numbers for *Cossypha dichroa* samples MK475629-MK475646 and *Cossypha natalensis* samples MK475579-MK475626 and for the two inferred hybrids MK475627-MK475628.

Analysis of *COI* barcode divergence

The *COI* 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) 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). These procedures were accomplished using Molecular Evolutionary Genetics Analysis (MEGAv7) software (Kumar *et al*., 2016; Tamura *et al.*, 2013).

Phylogenetic Analysis

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 two 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 v7) software (Kumar *et al*., 2016). *Cercotrichas quadrivirgata* (Bearded Scrub-Robin) 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). Bootstrapping statistics (1000 replicates) were calculated to evaluate the reliability of the inferred clades, where the higher the bootstrap value the more confidence there is that the branch point is correct (Felsenstein, 1985). Phylogenetic analyses were conducted in MEGAv7 (Kumar *et al*., 2016).

Introgression Analysis

In order to assess whether the inferred hybridisation is leading to genetic introgression the mean K2P distances between *C. dichroa* and *C. natalensis* from sympatric and allopatric populations were compared. Additionally, a Mantel test based on 10 000 permutations and a Pearson correlation coefficient was performed in XLSTAT version 2015.5.01.23251 (©Addinsoft 1995-2016). The Mantel test is the most commonly used method to evaluate the relationship between geographic distance and genetic divergence (Diniz-Filho *et al.*, 2013).

Microsatellite analysis

Amplification of microsatellite regions

PCR was carried out for the amplification of the microsatellite markers in the DNA. Fewer samples (n=45) 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 two *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 2) 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 were designed to correspond with the flanking

regions of the microsatellite markers. Seven of the thirteen microsatellite primer pairs cross amplified in both species. These seven markers were then used (bold in the Table 2 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.

Insert Table 2

Successful amplification was determined by gel electrophoresis. In order to obtain better resolution and determine heterozygosity of the samples, they were visualised using 3% agarose gels and viewed on the geldoc system.

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). Sample fragments were labelled with fluorescent labels on primers (Life Technologies, Inc, Johannesburg) to allow PCR reactions to be multiplexed. The seven microsatellite markers selected were fluorescently labelled. 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.

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™ 500 Liz™ (Applied Biosystem Inc.) internal size standard on an ABI 3130 Genetic Analyzer and Peak Scanner Software™ v1.0 was used. Peak Scanner Software™ 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.).

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 and Bruce, 2009). The Fst test is based on the variance of allele frequencies between populations (Holsinger and Bruce, 2009). Interpretation of the results of the Fst test is a comparison of the genetic variability within and between populations (Holsinger and 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 and 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 (MCMC), 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.

NEWHYBRIDS v1.1 (Anderson and Thompson 2002) was used to assign hybrid status to samples beyond the F1 generation. The programme estimates the posterior probability (q) that an individual belongs to one of six classes which include the two parental species, F1 and F2 hybrids, and backcrosses to each of the parental species (Anderson and Thompson, 2002). Individuals were assigned to one of the six classes based on their q-value with a threshold of 0.5. There are two types of priors, "Jeffrey's-like" prior and the Uniform prior, both for mixing proportions and allele frequencies. Jeffrey's is the default setting as it minimizes a certain measure of the probable "discrepancy" between the model and reality (Jeffreys, 1946). If there is enough data and the k and n values are big enough the Jeffreys and Uniform will be similar. Uniform does not make any assumptions and has the smallest possible impact on the final answer and data (Anderson and Thompson, 2002; Jeffreys, 1946). The Jeffrey's-like and Uniform prior was used, with ten repetitions and a run-length of 100,000 repetitions of MCMC, following a burn-in period of 500,000 iterations. CLUMPP v1.2.2 (Jakobsson and Rosenberg, 2007) was used to find the optimal posterior probabilities from the ten repetitions using the Greedy algorithm with 1000 random input orders. The final plot was composed using the ggplot2 package in R v3.5.1 (R Core Team 2018).

For both STRUCTURE and NEWHYBRIDS inference of an individual's ancestry or posterior probability of belonging to a certain genotype frequency class is conducted using a MCMC simulation approach. This Bayesian approach helps obtain estimates from the posterior distribution reflecting the appropriate q-value of an individual in STRUCTURE or the level of certainty that an individual belongs to one of six genotype classes in NEWHYBRIDS (Vähä and Primmer, 2006)

Results

Total DNA extracted from robin-chat blood samples

Total DNA was extracted from 67 robin-chat blood samples; *C. dichroa* (n = 18), *C. natalensis* (n = 47), inferred hybrids (n = 2), and from *Cercotrichas quadrivirgata* as the outgroup. The concentration and purity of the DNA was measured using a NanoDrop 1000. DNA extraction solutions which had a concentration of higher than 40 ng/µl and A_{260}:A_{280} ratio between 1.8 and 2.0 were considered to be of sufficient concentration and purity respectively. The DNA integrity of these samples was then assessed by agarose gel electrophoresis (not shown here).

PCR amplified *COI* barcodes

The sequenced *COI* barcode amplicons (n = 68) produced clean chromatograms with phred quality scores Q≥20. Additionally, *COI* barcode sequences of three *C. natalensis* samples were downloaded from GenBank (Accession numbers JQ174552.1, JQ174553.1 and JQ174554.1;

Accessed 08 October 2015). The *COI* barcode sequences were aligned by ClustalW and the sequence divergence was calculated using the K2P distance model (Kimura, 1980; Larkin *et al.*, 2007).

Analysis of *COI* barcode divergence

Analysis of the K2P distances determined that the mean interspecific distance for the *COI* barcodes analysed in this study was 9.05% (n = 900) while the mean intraspecific distance was 0.59%. The threshold divergence of 10x the mean intraspecific distance (Hebert *et al.*, 2004) was therefore 5.9%. Both *C. dichroa* and *C. natalensis* had intraspecific divergences below the threshold divergence, 0.66% (n = 154) and 0.52% (n = 1225) respectively. *Cossypha dichroa* and *C. natalensis* combined had a within genus divergence of 9.0% (std. error 0.06%) for allopatric populations and 9.1% (std. error 0.02%) for sympatric populations, which was above the threshold divergence. *Cossypha natalensis* and the inferred hybrids combined had a within group divergence of 0.66% (n = 99), which was below the threshold divergence for a different species. *Cossypha dichroa* and the inferred hybrids had a within group divergence of 9.2% (n = 36) above the threshold divergence for a different species (Hebert *et al*., 2004). The divergence between the hybrids was 0.72%.

Phylogenetic analysis

The ML tree provided a graphic representation of the pattern of divergences between *C. dichroa* and *C. natalensis* and their inferred hybrids (Figure 2).

Insert Figure 2

Introgression analysis

The K2P distances between *C. dichroa* and *C. natalensis* from sympatric and allopatric populations were compared to determine if introgressive hybridisation was occurring. There was very little difference (0.1%) between the interspecific distances between the two species in sympatric versus allopatric regions. There was no correlation between the genetic distance and geographic distance matrices between *C. dichroa* and *C. natalensis* interspecies pairs (Mantel test, $r(AB) = -0.012$, $p > 0.05$).

Intraspecific divergence

There was a correlation between the genetic distance and geographic distance within the *C. dichroa* intraspecific pairs with a K2P genetic divergence of 0.66% (Mantel test $r(AB) = 0.388$, $n = 153$, $p < 0.05$). There was also a correlation between genetic distance and geographic distance within *C. natalensis* intraspecific pairs with a K2P genetic divergence within these sample pairs of 0.52%. (Mantel test ($r(AB) = 0.515$, $n = 1225$, $p < 0.05$).

Microsatellite analysis

Seven of the thirteen microsatellite primer pairs developed by Wogan *et al*. (2013) cross amplified and were polymorphic in both species. A pairwise linkage disequilibrium test was performed intraspecifically and interspecifically using Arlequin 3.1 (Excoffier *et al*., 2005). No

deviations from linkage disequilibrium were detected (Significance level of $p > 0.05$). 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. 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 species do not deviate from Hardy-Weinberg equilibrium across the seven loci as indicated in Table 3 below.

Insert Table 3

An AMOVA analysis was then conducted to determine the genetic variation between the species. The AMOVA analysis was run interspecifically and intraspecifically to allow a holistic understanding of the species. Fst values for *C. dichroa* and *C. natalensis* were negative and therefore considered as 0 which indicates complete panmixis. All seven loci yielded private alleles: 21 in *C. natalensis* and 12 in *C. dichroa*.

The Bayesian clustering analysis using the statistical programme STRUCTURE identified two genetic clusters (K=2) as shown in Figure 3. The red genetic cluster represents *C. natalensis* with four of these phenotypically *C. dichroa* (arrows 1 to 4). The green cluster represents *C. dichroa* with six of these phenotypically *C. natalensis* (arrows 5 and 10). The arrows represent the following individuals (1) *C. dichroa* A43, (2) *C. dichroa* A46, (3) *C. dichroa* A33, (4) *C. dichroa* A39, (5) *C. natalensis* S30, (6) *C. natalensis* S104 (7) *C. natalensis* 490 (8) *C. natalensis* S25, (9) *C. natalensis* A44 and (10) *C. natalensis* 488.

Insert Figure 3

For the NEWHYBRIDS analysis using the Jeffrey's prior for allele frequencies, even low frequencies, may provide astuteness to the inference about hybrid category of some individuals. In such a case Uniform prior may give different results because this prior asserts that at least one copy of each allele is found in both populations thus diminishing the effect of private alleles (Anderson, 2003). The Uniform then has a smaller impact on the final answer and data (Anderson and Thompson, 2002; Jeffreys, 1946). It is included here as it supports the clustering seen in STRUCTURE. The Jeffrey's-like prior, was used to infer further information on the 10 samples of interest from the STRUCTURE result. As seen in Figure 4(a), samples numbered 1-4 (mtDNA for *C. dichroa* and phenotypically *C. dichroa,* but clustered with *C. natalensis* in SRUCTURE) have an estimated membership for sub-structure classes including pure *C. natalensis*, and F2 hybrid; F1 and backcrosses to *C. natalensis* and *C. dichroa* (Samples 2-4). Sample numbers 5 to 10 (mtDNA for *C. natalensis* and phenotypically *C. natalensis*, but clustered with *C. dichroa* in SRUCTURE), all have estimated membership classes including F2 hybrid and backcrosses to *C. dichroa*; F1 and F2 hybrids, backcrosses to *C. natalensis* and *C. dichroa* (8 and 10); F2 hybrid and backcrosses to *C. dichroa* (6 and 7); and pure *C.dichroa*, F2 hybrid and backcrosses to *C. dichroa* (5 and 9). Plots for both Jeffery's and Uniform priors are shown in Figure 4 below.

Insert Figure 4

There is much admixture in the samples. According to the Jeffrey's prior the posterior probabilities of backcrossing to *C. dichroa* are higher than backcrossing to *C. natalensis* . However, according to the

Uniform prior there is general backcrossing leading to admixture and introgression with both parents (Table 4).

Insert Table 4

The following ML phylogenetic tree (Figure 5) is numbered as in the STRUCTURE and NEWHYBRID analyses (Figures 3 and 4) to highlight the corresponding samples in the phylogenetic tree.

Insert Figure 5

Discussion

Cossypha dichroa and *C. natalensis* are unquestionably distinct species (within a common genus), each clustering as separate monophyletic lineages. This is well supported at 100% and 96% (1000 bootstrap replicates) respectively (Figure 2). The *Cossypha* phylogenetic tree of the two species showed deep interspecific and shallow intraspecific divergence. Each species had a distinctly different *COI* barcode resulting in their separate grouping while individuals of the same species had similar *COI* barcodes and therefore grouped together. As reported in the results, the mean interspecific K2P distance was found to be 9.05% which is *c.* 15x the mean intraspecific distance of 0.59%, further supporting that they are different species (Hebert *et al.,* 2004). The standard screening threshold of sequence difference (10x average intraspecific difference) was thus 5.9% for these two *Cossypha* species. These mean inter- and intraspecific distances are similar to previously published values calculated on much larger sample sizes, for example a study of 260 North American bird species found a mean interspecific K2P distance of 7.93% and a mean intraspecific K2P distance of 0.27% for *COI* barcodes (Hebert *et al.,* 2004).

Both *C. dichroa* (K2P distance 0.66%) and *C. natalensis* (K2P distance 0.52%) have an intraspecific divergence well below the threshold and therefore are not deemed to contain cryptic species. There is however distinct separation between the southern African and Gabon *C. natalensis* indicating the geographically distant populations are divergent and possibly in the early stages of allopatric speciation. This is not surprising considering the distance between these locations. However, within the southern African birds there seems to be no genetic evidence for the support of different subspecies, viz. *C. n. natalensis*, *C. n. egregior*, *C. n. hylophona* (Clancey, 1982); thus our genetic data do not support Clancey's (1991) hypothesis of multiple distinct phenotype subspecies of *C. natalensis* within southern Africa. However, in *C. dichroa* the genetic divergence was correlated with geographic distance and three separate clades are apparent (Figure 2). Birds from KwaZulu-Natal cluster together (two separate clusters) and the birds in Limpopo province cluster together, which is consistent with Clancey's (1981) recognition of a distinct subspecies (*C. dichroa mimica*) in the eastern Limpopo province. There is an exception of one *C. dichroa* (nfa31) that clusters with this group. Given that, i) the species is involved in seasonal movements and ii) the distributional range is small compared to *C. natalensis*, this clustering is somewhat surprising, and may reinforce the species dependence and restriction to the naturally fragmented forest biome. Furthermore, it may lend support to the allocation of phenotypic subspecies within *C. dichroa* (Clancey, 1981; 1982).

The inferred hybrids of the *C. dichroa* and *C. natalensis* group phylogenetically with the *C. natalensis* cluster indicating that they are likely resulted from a hybridisation event involving a *C. natalensis* female (Figure 2). Both inferred phenotypic hybrids were male and had characteristics of both species (Davies *et al.*, 2011; Figure 1E), and mitochondrial DNA results indicated the female parent was a *C. natalensis*. *Cossypha dichroa* as the male parent of these hybrids makes intuitive sense on grounds that male *C. dichroa* (41-50 g) are significantly larger than male *C. natalensis* (31-36 g) (values from Davies *et al.*, 2011). 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 hybridisation as the blood of only two inferred hybrid specimens was obtained for this study. It would be interesting to expand this study to include more phenotypic hybrid specimens to determine whether the inferred hybridisation 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 analysed. Subsequent to the publication by Davies *et al*. (2011) more inferred hybrid phenotypes have been identified across a region of sympatry (Figure 1CandD), as well as an aberrant bird that we were unfortunately unable to sample (Figure 1F). Also, we were not able to include the two inferred hybrids in our STRUCTURE analysis, due to too little sample.

Backcrossing of interspecific hybrids into one or both of the parental species would result in introgression of *COI* barcodes from one species into the gene pool of the other, as suggested by findings of discordance between phenotypic identification and *COI* barcode haplotype (Toews *et al.,* 2011). For example, extensive hybridisation of Townsend's Green Warblers *Dendroica townsendi* and Black-throated Green Warblers *D. virens* resulted in frequent mismatch between mitochondrial haplotype and phenotypic identification, 10 out of 68 specimens phenotypically identified as *D. virens* had *D townsendi* mtDNA and four out of 35 specimens phenotypically identified as *D. townsendi* had *D. virens* mtDNA (Toews *et al.,* 2011). There were no findings of discordance between phenotypic identification and *COI* barcode haplotype for any of the individuals analysed in this study and therefore no evidence of genetic introgression between *C. dichroa* and *C. natalensis* for mtDNA. However it is important to note that the appropriate set of markers is very important to help identify hybrids and genetic introgression in avian populations. Väli *et al*. (2010) used a combination of microsatellites and single nucleotide polymorphisms (SNPs) to identify hybrids of Greater Spotted Eagle *Aquila clanga* and Lesser Spotted Eagle *A. pomarina*. To discuss the taxonomic status of two *Calliope* species, Alström *et al*. (2013) based their analyses on mitochondrial and nuclear DNA and songs. Further work was thus done on these *Cossypha* species using microsatellite markers that cross amplify in both species to help assess the taxonomic status within the genus and to confirm whether genetic introgression has occurred.

A Bayesian clustering analysis of the microsatellite data using STRUCTURE indicated that they separate into distinct clusters (Figure 3); however this time discordance between phenotypic identification and genotype was apparent.

The STRUCTURE analysis based on the microsatellite genotypes indicated two distinct clusters (K=2). Four out of 24 (17%) of the *C. natalensis* cluster were phenotypically classified as *C dichroa*. Theses samples were all from sympatric populations so could be the result of genetic introgression following hybridization and backcrossing, as supported in the NEWHYBRID plots. Six of the 21 samples (29%) in the *C. dichroa* cluster were phenotypically *C. natalensis* suggesting admixture and introgression, also supported in the NEWHYBRIDS analysis. Two of these were from an allopatric *natalensis* population on Mtunzini (KwaZulu-Natal) i.e. *C. natalensis* 490 and 488. The other samples were all from sympatric populations from Vernon Crookes Nature Reserve (KwaZulu-Natal) and Schoemansdal (Limpopo) (refer to Table 1 for co-ordinates). A pairwise Fst test was conducted in conjunction with the intraspecific and interspecific AMOVA tests. Fst test was negative and therefore considered as 0 which indicates complete panmixis i.e. the two populations are interbreeding.

Genetic introgression between *C. dichroa* and *C. natalensis* could result in loss of genetic integrity and therefore lower interspecific *COI* barcode divergence in areas where they hybridise compared to geographically isolated areas where parental species remain pure and genetically distinct (Rheindt and Edwards, 2011). However, the mean interspecific distance between *C. dichroa* and *C. natalensis* from sympatric (9.1%) and allopatric populations (9.0%) was found to be almost equal. Furthermore, interspecific geographic and genetic distance was found to be unrelated when comparing *C. dichroa* and *C. natalensis* interspecies pairs whereas introgression would have caused a positive correlation such that interspecific genetic distance would increase with increasing geographic distance. It appears, therefore, that hybridisation between *C. dichroa* and *C. natalensis* is not resulting in genetic introgression according to mtDNA analysis, and that any backcrossing of interspecific hybrids is of little genetic influence on either parental species. Given the variable phenotypic expression of inferred hybrids (Davies et al., 2011; Figure 1C-F) we hypothesized that introgression and back crossing is likely. Based on NEWHYBRIDS, as seen in Figure 4, the *C. dichroa* samples appear not to be pure *C. dichroa* robin-chats but instead the result of F2 hybrids and backcrosses with pure *C. dichroa*. *Cossypha natalensis* S30 (5), *C. natalensis* S104(6), -*C. natalensis* 490-(7), *C. natalensis* S25 (8), *C. natalensis* A44 (9) and *C. natalensis* 488 (10) show q values in support of backcrossing with *C. dichroa* indicating genetic introgression following hybridization. F2 hybrids seem to be more common than F1 hybrids based on the Jeffrey's prior q-values. Also the backcrossing appears to be predominantly with *C. dichroa.*

Conclusions

We acknowledge that the markers used may not be informative enough and suggest further field and genetic based studies will inform these interpretations. Vähä and Primmer (2006) found that the method of NEWHYBRIDS performed slightly better than STRUCTURE when individuals from both backcross and F1 hybrid classes were present in the sample and they recommend that to separate backcrosses from purebred parental individuals requires at least 48 loci, even with high divergence between parental populations. Although our study only used seven microsatellite loci, the NEWHYBRIDS analyses did show hybridisation, admixture, and backcrossing amongst individuals of both *C. dichroa* and *C. natalensis*.

Many genera of African passerines are weakly diagnosed and remain untested (Beresford 2003). This paper adds to information on the phylogenetic relationships of two species of *Cossypha* robin-chat and further work will include adding more species of *Cossypha* robin-chats to the phylogenetic tree. Our analyses suggest that hybridisation occurs across the genus because of recent divergence in the taxon, with call distinctness being a strong driver and definer of reproductive isolation. Under such a scenario species isolating mechanisms ensure that hybridisation between other *Cossypha* species is rare (or absent), with little introgression. These two species are strong mimics and it is hypothesized that in regions of sympatry where one species is rarer the likelihood of pairings may be more common. These important questions regarding the evolutionary processes within the genus remain to be resolved.

Acknowledgements

Thank you to Claire Tinderholm for initial DNA barcoding experiments performed, to Stacey de Souza for help with NEWHYBRID analyses and to Gregory Davies for his insightful comments and guiding critique and for the bird blood samples.

References

Allendorf, F.W., Leary, R.F., Spruell, P. and Wenburg, J.K. 2001. The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution* 16: 613-622.

Alström, P., Song, G., Zhang R., Gao X., Holt P.I., Olsson, U. and Lei F. 2013. Taxonomic status of Blackthroat *Calliope obscura* and Firethroat *C. pectardens*. *Forktail* 29: 94-99.

Anderson EC. 2003. User's guide to the program NEWHYBRIDS version 1.1 beta. Available at http://ib.berkeley.edu/labs/slatkin/eriq/software/new_hybs_doc1_1Beta3.pdf [accessed 8 July 2019].

Anderson E.C. and Thompson E.A. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160: 1217–1229.

Anderson (2003)

Amos, W., Hoffman, J.I., Frodsham, A., Zhang, L., Best, S. and Hill, A.V.S. 2006. Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology* 7: 10-14.

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

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

BirdLife International. 2016. IUCN Red List for birds. Downloaded from http://www.birdlife.org on 26/01/2016.

Blin, N. and Stafford, D.W. 1976. A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Research* 3: 2303-2308.

Buschiazzo, E. and Gemmell, N.J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays*, 28(10): 1040-1050.

Clancey P.A. 1981. Variation in the Chorister Robin *Cossypha dichroa* (Gmelin), 1789. *Durban Museum Novitates* 13(1): 6-10.

569 Clancey P.A. 1982. On the robins *Cossypha dichroa* and *C. natalensis* (Aves: Turdidae) in
570 southern Africa. *Bonnerzoologische Beiträge* 33: 293–302.
571 Clancey P.A. 1986. Endemicity in the southern African avifauna. *Durban Museum Novitates*
572 13(20): 245-284.
573 Clancey P.A. 1991. Subspeciation in the Natal Robin *Cossypha natalensis* Smith, 1840. *Durban*
574 *Museum Novitates* 16: 25-30.
575 Cler, L., Bu, D., Lewis, C. and Euhus, D. 2006. A comparison of five methods for extracting
576 DNA from paucicellular clinical samples. *Mol Cell Probes* 20: 191-196.
577 Coetzer, W.G., Downs, C.T., Perrin, M.R. and Willows-Munro, S. (2015). Molecular systematics
578 of the Cape Parrot (Poicephalus robustus): Implications for taxonomy and conservation.
579 *Plos ONE*, DOI 10.1371/journal.pone.0133376.
580 Davies, G.B.P., Symes, C.T., Boon, R.G.C. and Campbell, H.A. 2011. Inferred hybridisation,
581 sympatry and movements of Chorister Robin-Chat *Cossypha dichroa* and Red-capped
582 Robin-Chat *C. natalensis*. *Ostrich* 82: 231-241.
583 Diniz-Filho, J.A.F., Soares, T.N., Lima, J.S., Dobrovolski R., Landeiro, V.L., de Campos Telles,
584 M.P., Rangel, T.F. and Bini, L.M. 2013. Mantel test in population genetics. *Genetics and*
585 *Molecular Biology* 36(4): 475-485.
586 Erlich, H.A. 1989. Polymerase chain reaction. *Journal of Clinical Immunology* 9: 437-447.
587 Excoffier, L.G., Laval, G. and Schneider, S. 2005. Arlequin Ver. 3.0 : An integrated software
588 package for population genetics data analysis. Evolutionary Bioinformatics Online, 1, pp.
589 47-50.
590 Farkas T. 1969. Notes on the biology and ethology of the Natal Robin *Cossypha natalensis*. *Ibis*
591 111: 281-92.
592 Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap.
593 *Evolution* 39(4): 783-791.
594 Fridolfsson, A.-K. and Ellegren, H. 1999. A simple and universal method for molecular sexing of
595 non-ratite birds. *Journal of Avian Biology* 30: 116-121.
596 Germain-Aubrey, C.C., Nelson, C., Soltis, D.E., Soltis, P.S. and Gitzendanner, M.A. 2016. Are
597 microsatellite fragment lengths useful for population-level studies? The case of *Polygala*
598 *lewtonii* (Polygalaceae). *Applications in Plant Sciences*, 4(2), DOI 10.3732/apps.1500115.
599 Gunning, J.W.B. 1909. A new species of *Cossypha* from West Pondoland. *Annals of the*
600 *Transvaal Museum* 1: 174.
601 Hebert, P.D., Cywinska, A. and Ball, S.L. 2003. Biological Identifications through DNA
602 Barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270:
603 313-321.
604 Hebert, P.D., Stoeckle, M.Y., Zemlak, T.S. and Francis, C.M. 2004. Identification of birds
605 through DNA barcodes. *PLoS Biology* 2: e312.
606 Hockey, P.A., Sirami, C., Ridley, A.R., Midgley, G.F. and Babiker, H.A. 2011. Interrogating
607 recent range changes in South African birds: confounding signals from land use and
608 climate change present a challenge for attribution. *Diversity and Distributions* 17: 254-261.
609 Hockey, P.A.R., Dean, W.R.J. and Ryan, P.G. 2005. Roberts - Birds of southern Africa, VIIth
610 ed. The Trustees of the John Voelcker Bird Book Fund, Cape Town.
611 Holsinger, K.E. and Bruce, S.W. 2009. Genetics in geographically structured populations:
612 defining, estimating and interpreting Fst. *Nature Reviews Genetics*, 10(9): 639-650.

613 Jakobsson M. and Rosenberg N.A. 2007. CLUMPP: a cluster matching and permutation program
614 for dealing with label switching and multimodality in analysis of population structure.
615 *Bioinformatics* 23: 1801–1806.
616 Jeffreys, H. 1946. An Invariant Form for the Prior Probability in Estimation Problems.
617 Proceedings of the Royal Society of London. Series A, Mathematical and Physical
618 Sciences. 186 (1007): 453–461. doi:10.1098/rspa.1946.0056. JSTOR 97883.
619 Jetz, W., Wilcove, D.S. and Dobson, A.P. 2007. Projected impacts of climate and land-use
620 change on the global diversity of birds. *PLoS Biology* 5: e157.
621 Jordaens, K., Goergen, G., Virgilio, M., Backeljau, T., Vokaer, A. and De Mayer, M. 2015. DNA
622 barcoding to improve the taxonomy of the Afrotropical Hoverflies (Insecta: Diptera:
623 Syrphidae). *PLoS ONE*: 10(10): e0140264, DOI 10.1371/journal.pone.0140264.
624 Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions
625 through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:
626 111-120.
627 Kumar S., Stecher G., and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
628 version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874.
629 Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., Mcwilliam, H.,
630 Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and
631 Higgins, D.G. 2007. Clustal W and Clustal X Version 2.0. *Bioinformatics* 23: 2947-2948.
632 Loparev, V., Cartas, M., Monken, C., Velpandi, A. and Srinivasan, A. 1991. An efficient and
633 simple method of DNA extraction from whole blood and cell lines to identify infectious
634 agents. *Journal of Virological Methods* 34: 105-112.
635 Lu, G., Basley, D.J. and Bernatchez, L. 2001. Contrasting patterns of mitochondrial DNA and
636 microsatellite introgressive hybridization between lineages of lake whitefish (Coregonus
637 clupeaformis); relevance for speciation. *Molecular Ecology* 10: 965-985.
638 Markoulatos. P., Siafakas, N. and Moncany, M. (2002). Multiplex polymerase chain reaction: A
639 practicla approach. *Journal of Clinical Laboratory Analysis* 16: 47-51.
640 Morgante, M. and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant
641 genetics. *Plant Journal*, 3: 175-182.
642 Oatley, T. and Arnott, G. 1998. *Robins of Africa.* Acorn Books. ISBN: 9-781874-802105
643 Oatley, T.B. 2005a. Chorister Robin-Chat. *In:* Hockey, P.A.R., Dean, W.R.J. and Ryan,P. (eds.)
644 *Roberts' birds of southern Africa.* Cape Town: Trustees of the John Voelcker Bird Book
645 Fund. Pp. 934-935.
646 Oatley, T.B. 2005b. Red-Capped Robin-Chat. *In:* Hockey, P.A.R., Dean, W.R.J. and Ryan, P.
647 (eds.) *Roberts' Birds of Southern Africa.* Cape Town: Trustees of the John Voelcker Bird
648 Book Fund. Pp. 932-934.
649 Olsen, G.J., Matsuda, H., Hagstrom, R. and Overbeek, R. (1994). FastDNml: a tool for
650 construction of phylogenetic trees of DNA sequences using Maximum likelihood.
651 *Computational Applied Biosciences* 10(1): 41-48.
652 Price, T. 2008. *Speciation in birds,* Greenwood Village, Colorado: Roberts and Company
653 Publishers.
654 Pritchard, J.K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using
655 multilocus genotype data. *Genetics* 155: 945-959.
656 R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for
657 Statistical Computing. Vienna, Austria. Available at https://www.r-project.org.

Rheindt, F.E. and Edwards, S.V. 2011. Genetic introgression: an integral but neglected component of speciation in birds. *The Auk* 128: 620-62.

Rhymer, J.M. and Simberloff, D. 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27: 83-109.

Samani, N.K., Esa, Y., Nurul Amin, S.M. and 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.

Sangster, G., Alström, P., Forsmark, E. and 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 Phylogenetics and Evolution* 57: 380-392.

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

Seutin, G., White, B.N. and Boag P.T. 1991. Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* 69: 82-90.

Symes, C. 2011. Cross-hatched – The search for Haagner's Robin-Chat. *Africa – Birds and Birding* 16(2): 30-32.

Tamura, K. and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10:512-526.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.

Taysom, A., Johnson, J. and Guay, P.J. 2014. Establishing a genetic system to distinguish between domestic Mallards, Pacific black ducks and their hybrids. *Conservation Genetic Resource* 6: 197-199.

Toews, D.P.L., Brelsford, A. and 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: 434-446.

Vähä J.P. and Primmer C.R. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology* 15: 63–72.

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

Wiegers, U. and Hilz, H. 1971. A new method using 'proteinase K' to prevent mRNAdegradation during isolation from Hela cells. *Biochemical and Biophysical Research Communications* 44: 513-519.

Wogan, G.O.U., Feldheim, K., Voelker, G., and Bowie, R.C.K. 2014. 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 Resources* 7: 151-154.

Woolaver, L.G., Nicols, R.K., Mortan, E.S. and Stutchbury, B.J.M. (2013). Population genetics and relatedness in a critically endangered island, Ridgeway's Hawk Buteo ridgwayi. *Conservation Genetics* 14: 559-571.

Yang, J., Hsu, K., Liu, Z., Su, L., Kuo, P., Tang, W., Zhou, Z., Liu, D., Bao, B. and 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.

Figure 1

Phenotypic comparison of *Cossypha* robin-chats

a) *Cossypha dichroa*, Kurisa Moya, Mpumalanga province (Photograph credit: Warwick Tarboton); **b)** *Cossypha natalensis*, Eshowe, KwaZulu-Natal province (Photograph credit: Warwick Tarboton); **c)** *C. dichroa* x *C. natalensis*, Lorraine Estates, Umzumbe, KwaZulu-Natal (note grey crown, 4 July 2013, ring BE34709; Mass = 28g; Wing = 95mm; Tail = 81mm) (Photograph credit: Andrew Pickles); **d)** likely *C. dichroa* x *C. natalensis* hybrid, Mazeppa Bay, Eastern Cape (note dark cheeks - see inset, 9 April 2011, ring BE63158; Mass = 35.8g; Wing = 95mm; Tail = 83mm) (Photograph credit: Karin Nelson); **e)** *C. dichroa* x *C. natalensis* hybrid (note grey back; ring BE37965; Davies *et al.*, 2011); **f)** aberrant *C. natalensis*, Levubu, Soutpansberg (note four central tail rectrices, as opposed to two typical for the species; Photograph credit: Craig Symes).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27912v1 | CC BY 4.0 Open Access | rec: 21 Aug 2019, publ: 21 Aug 2019

Figure 2

Molecular Phylogenetic analysis by Maximum Likelihood method

The evolutionary history of *Cossypha* robin-chats (n=70) was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993). The tree with the highest log likelihood (-1530.0646) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 71 nucleotide sequences. There were a total of 561 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 Evolutionary analyses were conducted in MEGA7 (Kumar *et al*., 2016). *Cercotrichas quadrivirgata* was the outgroup. The species marked with an asterisk (*) had *COI* barcode sequences downloaded from GenBank. The species shaded in grey represent those found in the northern geographic distribution

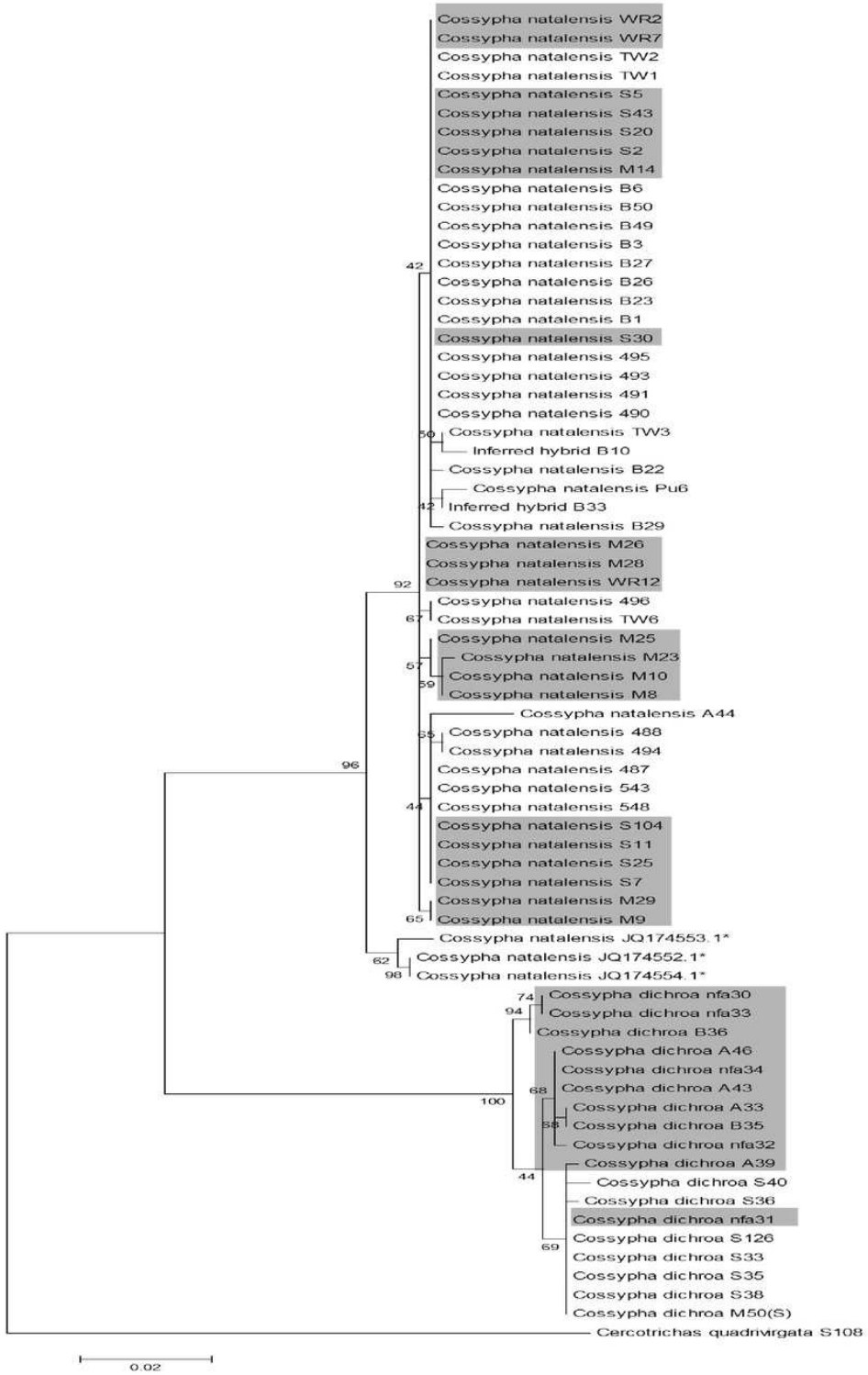


Figure 3

STRUCTURE analysisbased on microsatellite genotyping

STRUCTURE analysis based on the microsatellite genotypes indicated two distinct clusters (K=2) of *C. natalensis* (red), and *C. dichroa* (green). Each individual is represented by a single vertical line, with lengths proportional to the estimated membership in each cluster based on genetic makeup of the seven microsatellite markers. Individuals marked , are all phenotypically *C. dichroa*.

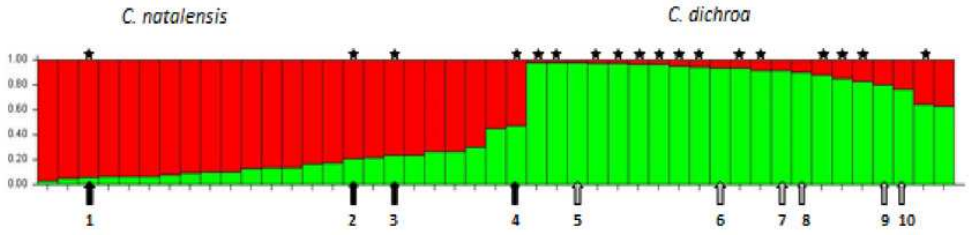


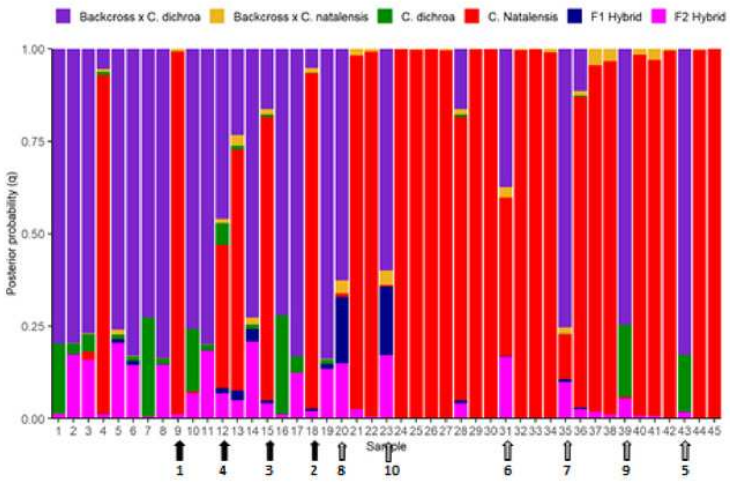
Figure 4

NEWHYBRIDS analyisis of microsatellite genotypes

NEWHYBRIDS analysis of microsatellite genotypes for *C. dichroa* (numbers 1 to 19) and *C. natalensis* (numbers 20 to 45). Each individual is represented by a single horizontal line, with lengths proportional to the estimated membership for each sub-structure class, i.e. purely parental, F1hybrids, F2 hybrids, and backcrosses to the two parental species. (a) Jeffrey's prior and (b) Uniform prior.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27912v1 | CC BY 4.0 Open Access | rec: 21 Aug 2019, publ: 21 Aug 2019

(a)



(b)

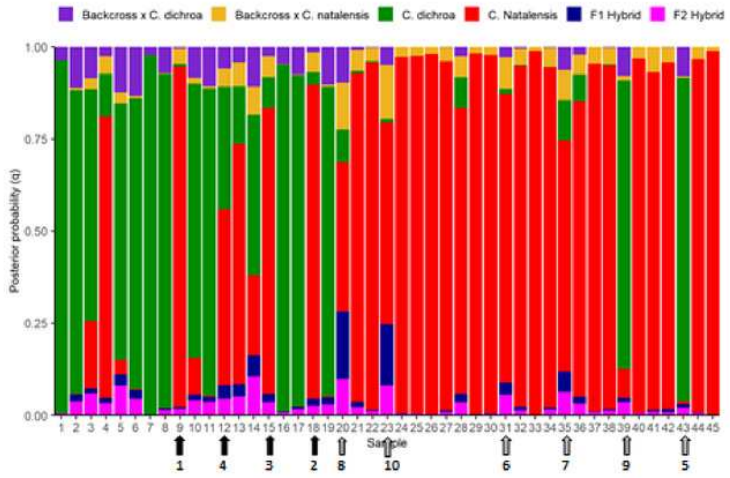


Figure 5

Phylogenetic tree including the samples used in STRUCTURE and NEWHYBRIDS

Molecular Phylogenetic analysis by ML method of the samples shown in the STRUCTURE analysis. The ML phylogeny tree was generated using 1000 bootstrap iterations.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27912v1 | CC BY 4.0 Open Access | rec: 21 Aug 2019, publ: 21 Aug 2019

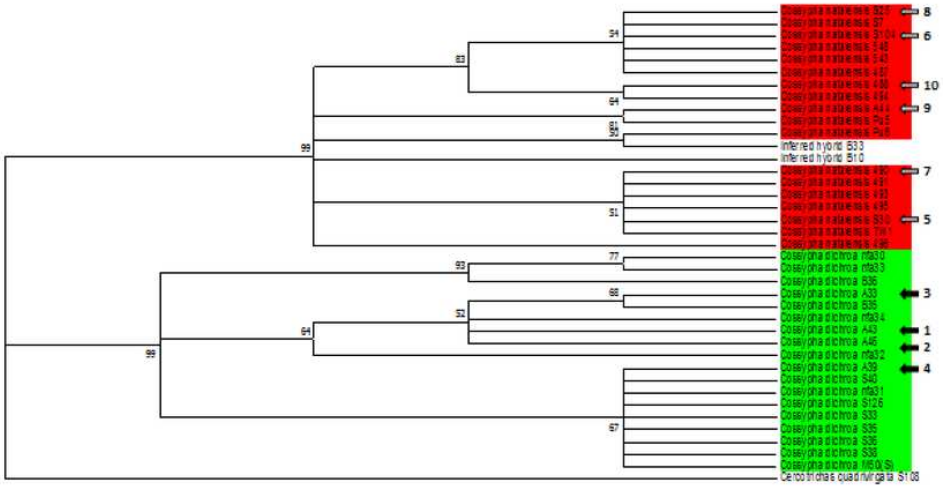


Table 1(on next page)

Sampling sites and samples collected

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

1

Location	Co-ordinates	Altitude (m.a.s.l.)	*C. dichroa* (n)	*C.natalensis* (n)	Inferred hybrids (n)
Vernon Crookes Nature Reserve, KwaZulu-Natal (A and B)	30°16'28"S, 30°36'36"E	420	6	11	2
New Forest, Nottingham Road, KwaZulu-Natal (nfa)	29°27'50"S, 29°52'43"E	1610	5		
Twinstreams Education Centre, Mtunzini, KwaZulu-Natal (TW or arabic numeral 487-548)	28°58'51"S, 31°44'09"E	20		14	
Pullen Farm, Nelspruit, Mpumalanga (Pu)	25°34'22"S, 31°10'53"E	910		1	
Wits Rural Facility, Limpopo (WR)	24°33'07"S, 31°05'48"E	570		3	
Schoemansdal Environmental Education Centre, Schoemansdal, Limpopo (S)	23°01'04"S, 29°43'32"E	980	7	9	
Inhamitanga Forest, central Mozambique (M)	18°09'17"S, 35°07'29"E	180		9	
Sample size			18	47	2

2

3

Table 2(on next page)

Microsatellite primers tested for cross amplification .

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

1

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27912v1 | CC BY 4.0 Open Access | rec: 21 Aug 2019, publ: 21 Aug 2019

2

3

4

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27912v1 | CC BY 4.0 Open Access | rec: 21 Aug 2019, publ: 21 Aug 2019

Table 3(on next page)

Allele frequencies and Hardy-Weinberg statistics

Summary of allele frequencies and Hardy-Weinberg statistics in the two *Cossypha* robin-chats.

1

	Number of samples	Mean number of alleles across 7 loci	Mean observed heterozygosity	Mean expected heterozygosity	Mean P-values
C. dichroa	18	7	0.78708	0.76951	0.23033
C. natalensis	26	8.71	0.84668	0.80959	0.39041

2
3

Table 4(on next page)

Posterior probabilities (q) assigned using NEWHYBRIDS

Table 4: Posterior probabilities (q) assigned to the six classes of individuals using NEWHYBRIDS.

1

		q – values											
Sample		***C. dichroa***		***C. natalensis***		**F1 Hybrid**		**F2 Hybrid**		**Backcross x *C. dichroa***		**Backcross x *C. natalensis***	
		Jeffreys	**Uniform**	**Jeffreys**	**Uniform**	**Jeffreys**	**Uniform**	**Jeffreys**	**Uniform**	**Jeffreys**	**Uniform**	**Jeffreys**	**Uniform**
1	A43	0	0.0071	0.9812	0.924	0	0.0045	0.0116	0.0182	0.0002	0.0040	0.007	0.0422
2	A46	0.0009	0.034	0.9075	0.8522	0.0058	0.0189	0.0205	0.0265	0.0515	0.0146	0.0138	0.0539
3	A33	0.0055	0.0837	0.7672	0.7752	0.0084	0.0234	0.0407	0.0355	0.1628	0.0250	0.0154	0.0571
4	A39	0.0585	0.3324	0.3868	0.4775	0.0159	0.0373	0.0672	0.0448	0.4596	0.0583	0.012	0.0497
5	S30	0.1532	0.8787	0	0.0072	0.0004	0.0101	0.0181	0.0206	0.8283	0.0788	0	0.0044
6	S104	0	0.0142	0.4293	0.7832	0.0023	0.0329	0.1662	0.0560	0.3746	0.0278	0.0276	0.0859
7	489	0.0003	0.1092	0.1228	0.6279	0.008	0.0559	0.0979	0.0630	0.7534	0.0621	0.0175	0.0819
8	S25	0.0001	0.0891	0.0083	0.4059	0.1797	0.1835	0.1496	0.0977	0.6268	0.0975	0.0355	0.1263
9	A44	0.1949	0.7819	0.0044	0.0794	0.0004	0.0123	0.054	0.0355	0.746	0.0781	0.0002	0.0128
10	488	0	0.0103	0.0048	0.099	0.186	0.1658	0.1706	0.0814	0.599	0.0.488	0.0395	0.1462