Genetic and morphological analyses of historic and contemporary populations of western lowland gorilla: A multidisciplinary approach for the conservation of a critically endangered primate

by

Jaimie Morris

Canterbury Christ Church University

Thesis submitted for the Degree of Doctor of Philosophy of Biological Science

2020

Abstract

This study investigates the morphology and genetic diversity of the critically endangered sub-species, the western lowland gorilla (*Gorilla gorilla gorilla*). Regional variation of a historic wild population was assessed morphologically and genetically, and genetic comparisons between this and a contemporary captive population were made to assess the genetic fitness of the contemporary population with the aim of assisting future conservation planning.

Geometric morphometric analyses were applied to skulls and mandibles of both sexes in the historic population of gorillas to assess regional variation in relation to size and shape. No significant difference was found for regional size comparisons but shape variation between regions did find significant variation in skull morphology, particularly for males.

MtDNA and nuclear markers were employed to detect regional differentiation in the historic population of gorillas, and to compare genetic diversity between historic and contemporary populations. The mtDNA results were hindered by nuclear insertions (numts) yet 30 sequences of the mitochondrial Control Region Hypervariable Region I (HVI) were obtained and haplogroups identified, which revealed potential differences in the historic distribution of haplogroups than current literature reports.

Nuclear analysis based on microsatellites confirmed that all the gorillas used in this study were western lowland gorillas. Furthermore, the paternity of individuals in the contemporary population was confirmed. Comparisons between the historical population and the captive US population showed that genetic diversity of the contemporary population had been retained at similar levels to wild populations and the US captive population thus concluding that the contemporary population is genetically sustainable for the foreseeable future.

Contents

List of tables

List of figures

Figure 2.4. Placement and numbering of landmarks used for both sexes for (a) skulls and (b) mandibles of western lowland gorillas, and morphological definitions adapted from Cardini *et al*. (2007), Rommel *et al*. (2009), Marcus *et al*. (2000) and Holton *et al*. (2015).

Figure 2.5. Relative warps plots showing landmark placement for tests of (a) female skulls, (b) female mandibles, (c) male skulls and (d) male mandibles of western lowland gorillas. Although there is some variation among landmark placement, the multivariate analysis of variance (MANOVA) showed non-significant differences, thus confirming accurate placement of landmarks in regards to shape analysis.

Figure 2.6. The tpsSmall output for the female skull of western lowland gorillas confirming a positive correlation (R = 0.9999) between Procrustes distancw and orthogonal tangent space.

Figure 2.7. Scree plots based on eigenvalues for the four datasets of Western Lowland Gorillas, (a) female skull, (b) male skull, (c) female mandible and (d) male mandible, with the first five components showing >70% of the total variation in all cases.

Figure 2.8. Box plot for visualisation purposes of centroid sizes (transformed by natural logarithm) by region for (a) skulls and (b) mandibles of western lowland gorillas and standard error estimates for (c) skulls and (d) mandibles.

Figure 2.9. Deformation grids based on the relative warp data assist with the visualisation of shape change from the average configuration, (a) shows the shape change for the female skulls and (b) male skulls of western lowland gorillas, (c) female mandibles and (d) male mandibles. The specimens chosen to demonstrate the shape change were based on their position on the plot. The outer most specimens show the greatest variation from the average configuration; in addition, specimens were selected to represent the three regional groups A, B and C which are depicted by the capitalised letter accompanying each specimen.

Figure 2.10. Relative warps plots for (a) female skulls and (b) male skulls of western lowland gorillas. The data was renamed to correspond to the four demes as classified by Groves (1970). Group A, has been renamed as plateau and group B as coastal. Group C has remained as the Congo group which would be classed as belonging to the coastal deme orginally following Groves (1970). The redefined group inlcudes individuals which would be classfied as belonging to the plateau deme following Groves (1970) but were placed in the coastal deme (group B).

Figure 2.11. Principal Components Analysis (PCA) plots based on eigenvalues for the four datasets, (a) female skull, (b) male skull, (c) female mandible and (d) male mandible of western lowland gorillas. The first two factors were used to visualise the results. None of the four datasets showed any obvious clustering of regional groups A, B, and C.

Figure 2.12. Discriminant Function Analysis (DFA) plots for the four datasets, (a) female skull, (b) male skull, (c) female mandible and (d) male mandible of western lowland gorillas. The first two factors were used to visualise the results.

Figure 3.1. Map of the Powell-Cotton Museum (PCM) gorilla specimens used for mtDNA analysis showing the three defined subgroups A, B, C based on geographical origin.

Figure 3.2. Family trees for the six main silverback males from the Aspinall Foundation used in this study, (a) Djala, (b) Kifu, (c) Kijo, (d) Kouillou, (e)Bitam, (f) Djanghou (Djanghou is Djalas' offspring).

Figure 3.3. Agarose gel electrophoresis of Western Lowland Gorilla DNA (1% agarose in 1X TAE buffer, stained with SYBR®safe, run for 30 minutes at 100 V, and visualised under UV-light). From left to right, wells 1-18 contain Powell-Cotton Museum (PCM) samples: FC.122, FC.124, CamI.44, CamI.45, CamI.46, CamI.14, ZVI.32, M342, MI.30, CamI.97, CamI.106, CamI.109, CamI.110, MII.6, CamI.95, M135, M879 and MII.25. Wells 20-25 contain repeated DNA extractions for the Royal College of Surgeons (RCS) samples: PA61-A, PA61-B, PA62-A, PA62-B, PA63-A and PA63-B. Wells 27-33 contain the extracted DNA from the Aspinall Foundation tissue samples, including 58-Masindi infant, 59-Kwimba 2nd infant, 60-Kouilla, 61-Babydoll, 62-Louna, 63-Virginika infant and 64-Mouilla. Wells 19, 26 and 34 contain the molecular weight marker/ladder (100 bp DNA ladder). The gel shows high molecular weight and intact DNA (e.g. well 27), smeared DNA including high molecular weight fragments (e.g. well 33), and low molecular weight fragments (e.g. 14), and of different intensities due to varying DNA quantities (relative brightness).

Figure 3.4. Chromatograph showing DNA sequence data for Aspinall Foundation gorilla named Jah; (top) forward sequence using second-round primer L91-115 and (bottom) reverse sequence using second-round primer H402-27.

Figure 3.5. Mismatch distribution plots showing expected and observed frequencies of pairwise differences of (a) Aspinall Foundation (ASP), (b) Powell-Cotton Museum (PCM), (c) Central African Republic (CAR) and (d) Nigeria (NIG). ASP and PCM show multimodal distributions whereas CAR and NIG reveal unimodal distribution. The expected frequencies represent the expectations based on a neutral model of evolution for natural selection-genetic drift in a randomly reproducing theoretical population.

Figure 3.6. Neighbour Joining tree of the 11 groups of gorillas rooted with the eastern gorilla species; bootstrap values are shown on branches.

Figure 3.7. Phylogenetic trees for all sequences including nuclear inserts of mitochondrial DNA (numts), where (a) is the maximum likelihood tree and (b) is the Bayesian tree, both rooted with the

outgroup. Coloured branches depict the issues of numt sequences which could lead to misinterpretation of phylogeny if undetected. The red branches depict the numt sequence CAP L76760 and includes eight of the ASP gorillas (Matibi, Bitanu, Jubi, Ujiji, Thirza, FouFou, Louna and Yene), and two of the PCM gorillas (Mer.387, FC.130). The green branches include two known eastern lowland numt sequences EL AF240456 and ELAF240448 that did not cluster with known eastern gorillas (*G. g. beringei* (blue branches) and *G. g. graueri* (purple branches). Both trees showed PCM Mer.138 clustered with EL240448. Known numt sequences were included in this data to assist with numt identification of data generated from this study.

Figure 3.8. Phylogenetic networks of mitochondrial Hypervariable Region I (HVI) haplotypes of gorillas showing the haplotype distribution for: (a) the 11 groups and (b) captive versus wild populations.

Figure 3.9. Phylogenetic haplotype trees of true mtDNA data showing their haplogroup distribution for (a) Maximum Likelihood (MLH) and (b) Bayesian methods. Both trees included 30 mtDNA sequences generated from this study as well as data from Clifford *et al*. (2004) and Soto-Calderón *et al*. (2015), and other additional sequences from GenBank. The trees were rooted at the outgroup. ***** indicates more than one individual shared the same haplotype; the list of haplotypes is given in Appendix 4.

Figure 3.10. Network analysis showing the 11 groups of gorillas based on mitochondrial Hypervariable Region I (HVI) sequences with haplogroups A, B, C, D (and their respective subgroups C1, C2, C3, D1, D2, D3) identified based on Clifford *et al*. (2004).

Figure 3.11. Map of the Powell-Cotton Museum (PCM) gorillas haplogroup locations. Circle size refers to the number of individuals at the same location (range 1-7). Haplogroup C2 was the only haplogroup present in region A, whereas haplogroups C1 and D2 were both present in regions B and C.

Figure 3.12. Mismatch distribution plots showing multimodal distribution for (a) the total dataset, (b) captive population, (c) contemporary wild population and (d) the Powell-Cotton Museum (PCM) groups of gorillas.

Figure 3.13. Mismatch distribution plots showing unimodal and multimodal distributions for the four haplogroups (A, B, C and D) of gorillas based on mitochondrial Hypervariable Region I (HVI) DNA sequences. (a) Haplogroup A is the eastern mountain gorillas, (b) haplogroup B is the eastern lowland gorillas, both showing unimodal distribution. (c) Haplogroup C with a multimodal distribution and (d) haplogroup D with a bimodal distribution, where both are the western species; haplogroup C contains the genetically differentiated Cross River gorillas.

Figure 4.1. The precise geographical locations of the 73 gorilla genotypes from the Powell-Cotton Museum (PCM; historic) population used in this study. Geographical subgroups/regions are depicted by the yellow ovals and their designated subgroup name is indicated by the uppercase letter (A, B and C). The circles indicated the locations of the gorillas (blue for male, red for female) and the number next to each location refers to the number of gorillas captured from that precise location.

Figure 4.2. Gel electrophoresis results for primer tests using DNA from three gorillas 1) FC.130 (PCM), 2) ZII.63 (PCM), and 3) Bitono (ASP) using four primers for loci: A) vWF, B) D4s1627, C) D8s1106 and D) D10s1432. Two thermocycler protocols were tested, the microsatellite protocol (left image) and the mitochondrial protocol from Chapter 4 (right image). The mtDNA protocol produced more favourable results indicating more PCR bands at the desired molecular weights.

Figure 4.3. Multiplexing result for Baloo: (a) for loci vWF, D4s1627, D2s1326, D1s550 and D16s2624 and (b) singleplex for loci D16s2624.

Figure 4.4. Comparisons of allele frequencies between the contemporary and historic populations of gorillas for all 10 loci: (a) vWF, (b) D4s1627, (c) D8s1106, (d) D10s1432, (e) D2s1326, (f) D5s1470, (g) D7s2204, (h) D7s817, (i) D1s550, (j) D16s2624.

Figure 4.5. Allelic patterns for the contemporary and historic populations of gorillas summarising the number of alleles (*Na*), the number of effective alleles (*Ae*) and the number of private alleles.

Figure 4.6. Comparison between the contemporary and historic populations of western lowland gorillas in private allele distributions (shown as percentages). The historic population contained 25 individuals with private alleles and the contemporary population contained three.

Figure 4.7. The geographic distribution of private alleles for the historic population of western lowland gorillas. A total of 25 gorillas in the historic population were identified as carrying a private allele, however, MII.25 is not shown on map as coordinates were not available. The individuals with private alleles are depicted by the red boxes, and green circles indicate their exact geographical location. The number in the green circles identifies each geographical location and the gorillas at that specified location (Appendix 5).

Figure 4.8. Analysis of molecular variance (AMOVA) results shown as percentages for comparisons between (a) the four populations of gorillas consisting of two Cross River gorilla populations (contemporary and historic) and two western lowland gorilla populations (contemporary and historic), and (b) comparison between the contemporary and historic western lowland gorilla subspecies only.

Figure 4.9. Results from Structure analysis of the Cross River gorilla and western lowland gorilla populations, each containing a contemporary and historic population. (a) Structure Harvester results

with the Evanno method (Δ*K*) detected 1 or 2 clusters. (b) Structure Selector using the Puechmaille method detected four or five clusters.

Figure 4.10. Individual probability plots from Structure for the Cross River and western lowland gorilla analysis showing (a) the Δ*K* method identifying two clusters corresponding to the western lowland and Cross River subspecies, and (b) the Puechmaille method identifying four or five clusters (four shown) corresponding to the two subspecies plus additional substructure clusters in contemporary Cross River and in Contemporary (ASP) gorillas.

Figure 4.11. Results from Structure analysis of the 140 western lowland gorilla genotypes generated from this study for (a) the Evanno method which indicated that two clusters were present in the data as did (b) the Puechmaille method for all four statistical tests (MedMedk, MedMeak, MaxMedk and MaxMeak).

Figure 4.12. Individual probability plots from Structure for the western lowland gorilla. (a) Original order (contemporary and historic populations) of individuals. (b) Individuals ordered by *Q*. Each individual bar represents an individual gorilla. The 140 individuals were divided into two clusters with a clear break in the data. Hybrid individuals are depicted by the yellow arrows (their *Q* proportion fell below 80%). Cluster 1 contained only individuals from the contemporary population whereas Cluster 2 contained all the historical gorillas plus 35 gorillas from the contemporary population.

Figure 4.13. Results from Structure analysis of the 76 historical western lowland gorilla genotypes generated from this study for (a) the Evanno method which indicated that two clusters were present in the data as did (b) the Puechmaille method for all four statistical tests (MedMedk, MedMeak, MaxMedk and MaxMeak).

Figure 4.14. Individual probability plot of the 76 historical western lowland gorillas ordered by their original defined geographical regions (A, B, and C). This visual representation is indicative of no substructure. There is a slight observable difference in *Q* of subgroup C, but it does not indicate regional substructure as there is not a clear break in the data as observed in previous analyses.

Figure 4.15. Principal Coordinates Analysis (PCoA) for (a) the historical subgroup analysis showing no apparent substructure within the population, with the five gorillas from Congo (group C) not clustering together, and (b) PCoA for the historic and contemporary gorillas in the same analysis. There is a genetically distant cluster depicted by the red oval which includes individuals from Cluster 1 as identified in the Structure analysis. Kouillou has been labelled as he is one of the most genetically distant individuals as well as Mer.95, a gorilla from the historic (PCM) population who clustered with these contemporary individuals.

Figure 4.16. Principal Coordinates Analysis (PCoA) of all 140 gorillas plus an additional 34 Cross River individuals (Thalmann *et al*. 2011; Arandjelovic *et al*. 2015) across four loci. The cluster groups identified were based on the Structure results where the ASP contemporary population was divided into two clusters. ASP Cluster 2 and the PCM cluster shown here formed one cluster in Structure analysis but were divided here for visual purposes. The CR (Cross River gorilla) contemporary population is notably the most genetically distant cluster whereas the CR historic cluster shows considerable genetic similarity with primarily ASP Cluster 1, but also ASP Cluster 2 and the PCM cluster.

Figure 4.17. Identified migrant gorillas depicted by the yellow circles from each geographic region. Subgroup A contains seven migrant gorillas whereas subgroups B and C both contain two migrant gorillas each.

List of appendices

Appendix 1.

List of gorillas from the Aspinall Foundation samples that were used for in this study and their relationships, plus the mitochondrial founder of each.

Appendix 2.

Revised protocol for DNA extraction from FTA cards.

Appendix 3.

List of all GenBank sequences used in this study, including their group ID assigned for analyses and indication of whether the sequence is a numt sequence.

Appendix 4.

List of the 61 haplotypes from the numt free sequences, showing their haplotype, frequency, sequence ID and haplogroup.

Appendix 5.

List of the PCM specimens used in the microsatellite analyses showing specimen ID number, sex, locality of capture, geographical coordinates and subgroup allocation. The exact location (loc) refers to Fig. 4.7 where each specimen can be identified by the number in the green circle.

“You cannot get through a single day without having an impact on the world around you. What you do makes a difference, and you have to decide what kind of difference you want to make.”

— Jane Goodall

For my children, Isabelle and Oliver… Just don’t quit, no matter how hard things get.

Acknowledgements

There are many people that I would like to thank and which this research would not have been possible without. Firstly, my supervisor Dr. Rodrigo Vega, his patience, kindness, knowledge and support pulled me through the toughest times when I had lost all faith. He pushed me when I needed it, reassured me it was not “PhD in the bin time” and was always prepared with tissues and a good sense of humour. I hope I have not broken him, and he goes on to have many more PhD students because he is the best supervisor and I consider myself very lucky he inherited me.

Special thanks also to Professor Simon Harvey who begun this journey with me, his support and guidance will never be forgotten. Professor Peter Vujakovic, the Chair of the panel and my line manager for many years, I thank you for your support and guidance throughout. John Hills, the GIS guru, one of the most patient people on the planet, thank you for all your map making skills and support.

This research would not have been possible without the support of the Powell-Cotton Museum and the Aspinall Foundation particularly for providing specimens and samples. Thank you to Dr. Inbal Livne of the PCM for permitting the research to take place, sometimes at peculiar hours and for answering my questions, also to all the staff at the museum for being so kind and helpful, I loved my time there. Special thanks to Tony King and Jane Hopper of the Aspinall Foundation, for access to the samples, and especially to Tony for reading manuscripts and arranging some wonderful memories I will always treasure. Thanks to Vic Mathieson, gorilla keeper at the Aspinall Foundation for answering all my questions, her knowledge and love for the gorillas in her care is outstanding. Special thanks to Professor John Cooper of the Gorilla Pathology Group, for taking me under his wing and being so supportive, he and his wife Margaret are two of the kindest people I have had the pleasure of meeting. Thanks also to the Royal College of Surgeons for allowing the inclusion of their three gorilla specimens and facilitating the sample taking.

So many other colleagues I would like to thank in the Section of Life Sciences at Canterbury Christ Church University, too many to mention but thank you all for your support over the years. Special thanks to the research technicians and my treasured friends, Hannah and Naomi. Not only did you keep me stocked up with pipette tips, buffers and cake, you kept me laughing despite several science disasters and picked me up when I needed it, you are the best people to work with, nowhere else will be quite the same. Dr. Tyler Howell-Bray, I am honoured to call you a fellow colleague and friend, you held my hand in my darkest hour and didn’t let go, you kept me motivated and reminded me why I was doing all this in the first place, not to mention facilitating my idol and hero, Jane Goodall

emailing me words of motivation and encouragement. Special thanks to Jane Goodall, for taking the time to message and to remind me I was not alone, it meant a lot.

My friends and family, I am truly blessed to have such supportive and understanding friends who have stuck with me through it all, despite my sporadic attempts at staying in contact. I promise to be a better friend now this is over. My family, always providing support, never letting me give up. My mum in particular who has given up so much to see me succeed, I hope you are proud. My husband Ben, I could not have done it without you, you have been patient and understanding, your faith in me has not waivered, thank you. Lastly, but not least, my children, Isabelle and Oliver. You have been so patient and understanding, I know it has been hard for you when mummy is always working, I hope in time you understand why and if nothing else, I have taught you to follow your dreams and never give up.

Author declaration

I declare that this is my own work, unless stated otherwise.

Chapter 1
General Introduction

1.1. Structure and aims of thesis

This thesis aims to research the conservation biology of the western lowland gorilla (*Gorilla gorilla gorilla*), in terms of the genetic variation of historical (past and wild) and contemporary (captive) populations, and to explore the morphological regional diversity of past populations. This study is multidisciplinary and uses a combination of approaches including biogeography, population genetics, phylogeography and geometric morphometrics for informing the management of captive populations and for the conservation of this critically endangered species.

Chapter 1 provides a general introduction to the western lowland gorilla in terms of conservation, and genetic and morphological variation, and identifies unique contributions to the field. This chapter also provides a review of the literature in relation to biodiversity mapping and geographical information systems (GIS), geometric morphometrics, molecular ecology, and the importance of natural history collections (NHCs), with specific reference to the Powell-Cotton museum. The utilisation of natural history collections with geographical information systems for biodiversity mapping is reviewed and mapping of the historic western lowland gorilla populations is investigated with its application to further study in this research, and in broader terms, additionally, the importance of wildlife parks/zoos for scientific studies and conservation is reviewed. Chapter 2 focuses on geometric morphometrics and reviews previous geometric morphometric research focused on primates and specifically gorillas. The Powell-Cotton gorilla collection is visualised to investigate regional geographic variation. Chapter 3 is the first of two chapters which investigates genetic diversity and population structure in western lowland gorillas. This chapter investigates mitochondrial variation, firstly, by reviewing previous research in further detail and analysing the historic and contemporary populations which are the focus of this research. Chapter 4 focuses on the population genetics of western lowland gorillas using microsatellites. Again, a review of the literature specific to the western lowland gorilla is given followed by the analyses and comparison of the historic and contemporary populations. Chapter 5 is the general discussion which evaluates and discusses this research into broader context, providing conclusions and directions for further research.

The broader aims of this research were:

- To investigate regional variation (morphological and genetic) of a historic population of western lowland gorillas.
- To study the genetic diversity and structure within and among historic and contemporary populations of the western lowland gorillas, as well as to study the phylogeographic patterns.
- To 'bridge the gap' in the literature between historic and contemporary western lowland gorillas by combining geographical, morphological and genetic data of historic and contemporary populations, and reiterate the importance of museum natural history collections for conservation purposes.

Additionally, specific aims of this research were to seek:

- Confirmation that all gorillas in the captive population are (or were) genetically western lowland gorillas and not hybrids with any other subspecies, particularly the Cross River subspecies.
- Confirmation of relatedness and parentage of individuals in captivity through paternity testing.
- Whether there is genetic differentiation between captive individuals obtained from geographically distinct regions (the Republic of the Congo, Cameroon and Gabon).

1.2. An introduction to gorillas

On the 6th of May 2019 the United Nations (UN) reported a biodiversity crisis, with unprecedented species declines and accelerated extinction rates (UN 2019). This report was based on landmark findings from the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) Global Assessment, which was the first intergovernmental report of its kind and the most comprehensive report to date (UN 2019; IPBES 2019). With one in four species threatened with extinction, and 150-200 species lost every 24 hours, not only are anthropogenic activities eroding the very foundations (ecosystems) on which we as a species rely upon, but there is a moral and ethical responsibility to ensure the survival of biodiversity for future generations (UN 2019; IPBES 2019). Nature is in crisis, and the gorilla species are no exception.

Due to their similarities with humans and their importance to the ecosystem (see below), gorillas in general have been the focus of many research studies (Scally *et al*. 2012; Petre *et al*. 2013).

Gorillas are enigmatic animals and the largest of all living primates (Cassalett & Rothman 2018; Wright *et al*. 2020), they are one of the closest relatives to humans with 98% of similarity in terms of DNA (deoxyribonucleic acid) sequence (Toder *et al*. 2001). There are four genera which form the taxonomic family Hominidae, known as the hominids or great apes, which includes gorillas (*Gorilla* spp.), bonobos and chimpanzees (*Pan* spp.), orangutans (*Pongo* spp.), and the genus *Homo* (Dorado *et al.* 2018).

Gorillas share many other physical and behavioural traits and similarities with humans (Yaxley & Foley 2019). For example, gorillas are *K*-selected species, which tend to be larger, have a longer lifespan, invest more time and energy in offspring production but produce fewer offspring. Parental investment is high, and gorillas take care of their young for long periods of time, often teaching them behaviours to survive (E Crews & Gerber 2003). Additionally, gorillas are considered keystone species (Petre *et al*. 2013), an organism that performs a unique and crucial role in the functioning of an ecosystem (Paine 1974; Bond 1994; Power *et al*. 1996; Hale & Koprowski 2018), and which without it the integrity of the ecosystem would be dramatically compromised or cease to exist entirely (Garibaldi & Turner 2004; Hale & Koprowski 2018).

The diet of gorillas consists mainly of fruit and plant matter; therefore, their faeces contain many of the seeds of the plants and fruits which they consume making them one of the primary seed dispersers in their habitat (Rogers *et al*. 1998; Haurez *et al.* 2018). By roaming the forest and defecating, they promote the distribution of the plant species on which they and other species rely upon (Petre *et al*. 2013). In addition, gorillas build "nests" and when they move to another nesting site or die, other animals of the forest utilise the nests for themselves and populate the area (Tutin & Vedder 2001).

1.2.1. Gorilla taxonomy

Groves (2002) notes that an English sailor, Andrew Battell, who was held prisoner in Angola by the Portuguese, provided an account of 'real gorillas' which entered European literature for the first time in the 16th century. Dr Jeffries Wyman, a Boston anatomist, published the first scientific description of the gorilla in 1847 (Groves 2002; Gott & Weir 2013) with its scientific name *Troglodytes gorilla*. However, this description was based on the work done by Dr Thomas S. Savage who described the external character and habits of a new species of *Troglodytes* (i.e. *T. gorilla*) discovered near the river Gabon, as stated by the Proceedings of the Boston Society of Natural History (1847) (Groves 2002; Cooper & Hull 2017). Paul du Chaillu, a French-American hunter, shot and collected gorilla specimens in 1862 in the currently named region of Gabon and was the first Westerner to give an eyewitness account of gorillas in the wild (Conniff 2009; Haikal 2020).

The research on gorillas (until relatively recently), has been unintentionally disproportionate within the field of primatology. Despite West Africa being the origin of initial first contact of gorillas by Western explorers, it is the East African gorillas from which the majority of comprehensive research studies and accounts have emerged from (Taylor & Goldsmith 2002). Eastern gorillas were discovered by western science as a new species in 1902 and named as the Mountain gorilla (*Gorilla beringei beringei*) (Tobias & Cooper 2003; Herzfield 2017). George Schaller and Dian Fossey (1967-1985) were field primatology pioneers and provided the first methodical and analytical accounts of gorillas which confirmed our perception of these large primates as quadrupedal, terrestrial knuckle-walking vegetarians (Schaller 1963; Fossey 1979; Doran 1997). Schaller (1963) acknowledged an attempt to document the behaviour of the western lowland gorilla, briefly noted as "*an interesting report on the little-known West African gorilla.*" A similar pattern of disproportionate research occurred regarding the knowledge about chimpanzees; Jane Goodall, who began her research in the late 1950s, provided most of the accounts from her work with the chimpanzees of Gombe, Tanzania. Similarly, for many years, the orangutan accounts from which scientists formed interpretations and analyses were based on research from sparse sites situated along the Bornean coast in Southeast Asia (Taylor & Goldsmith 2002). Although these pioneering works provided a vast amount of previously unreported observations and data, they focused only on a small sample set which was not representative of the species as a whole. For example, multiple local variations in behaviour patterns have been observed for chimpanzees and orangutans (Whiten *et al*. 1999; Van Schaik *et al*. 2003; Whiten *et al*. 2007).

Historically, defining species has been problematic for biologists and the issue is still prevalent (Isaac *et al*. 2004). There are numerous species concepts and extensive literature regarding them (Wiley 1978; Isaac *et al*. 2004). Many biologists and authors have reviewed the various species concepts throughout history including Mayr (1957, 1963, 1969), Simpson (1961), Dobzhansky (1970), Grant (1971), Sokal (1973), and Sneath & Sokal (1973) to mention just a few (Wiley 1978). However, the most prevalent and often debated concepts are the biological species concept and the phylogenetic or 'diagnostic' species concept and their variations (Wiley 1978; Isaac *et al*. 2004). Additionally, there are other concepts such as the genotypic cluster species concept, the recognition species concept, the cohesion species concept, the ecological species concept and the evolutionary species concept (Coyne & Orr 2004). It is no wonder that taxonomical implications for species are still under huge debate when biologists, or at least systematic biologists, are not able to agree on a single definition or concept of what a species is (Hey *et al*. 2003).

Primate taxonomy is no exception, wrought with numerous juxtaposing taxonomic hypotheses for multifarious 'species'. However, the consensus throughout the 20th century was that only one single species of gorilla existed with three subspecies: the western lowland gorilla (*Gorilla gorilla gorilla*), the eastern lowland gorilla (*G. g. graueri*) and the mountain gorilla (*G. g. beringei*). Not

only was this distinction made in terms of cranial and postcranial characteristics (Schaller 1963; Groves 1970; Groves & Stott 1979), but these subspecies have distinct geographical distributions (Clifford *et al*. 2003). Summarised hypotheses of the gorilla species based on morphological variation, adapted from Tuttle (2003), are shown in Table 1.1. The summary is not an exhaustive list as there are a multitude of studies relating to gorilla taxonomy, however, it serves to provide an overview of hypotheses based on morphological research.

Table 1.1 Summarised hypotheses of gorilla taxonomy based on morphological research.

Author	Date	Hypothesis summary of number of species and subspecies of *Gorilla*
Schwarz	1928	**One species** (*Gorilla gorilla*) **Seven subspecies** *(G. g. gorilla, G. g. matschiei, G. g. diehli, G. g. uellensis, G. g. rex-pygmaeorum, G. g. graueri, G. g. beringei)*
Coolidge	1929	**One species** (*G. gorilla*) **Two subspecies** (*Gorilla beringei* and *G. gorilla*)
Rzasnicki	1936	**Six "complexes" based on geographical distribution** **Three centres-of-origin (subspecies)** *(G. g. gorilla, G. g. beringei, G. g. diehli)*
Vogel	1961	**Two species** (*G. beringei* and *G. gorilla*) **Three subspecies** (*G. g. gorilla, G. g. beringei, G. g. graueri)*
Groves	1967, 1970	**Two species** (*G. beringei* and *G. gorilla*) Probably **four subspecies** (*Gorilla beringei berengei, G. b. graueri, G. gorilla gorilla, G. g. diehli)*
Albrecht & Miller	1993	**One species** (*G. gorilla*) **Three subspecies** (*G. g. gorilla, G. g. beringei, G. g. graueri)*
Inouye	1994	**One species** (*G. gorilla*), subspecies not defined
Stumpf *et al*.	1997	**Two species** (*G. beringei* and *G. gorilla*) Ambiguous regarding subspecies but do not challenge Groves' hypothesis
Albrecht *et al*.	2002	**One species** (*G. gorilla*) **Four subspecies** (*G. g. gorilla, G. g. diehli, G. g. beringei, G. g. graueri)*
Leigh *et al.*	2002	**One species** (*G. gorilla*) **Three subspecies** (*G. g. gorilla, G. g. beringei, G. g. graueri), G. g. gorilla* is regarded as highly diverse
Taylor	2002	**One species** (*G. gorilla*) **Three subspecies** (*G. g. gorilla, G. g. beringei, G. g. graueri)*

Hypotheses adapted from Tuttle (2003) plus authors additions.

The first extensive revision of gorilla taxonomy was that of Coolidge (1929) who classified all gorillas as one species, as did Schwarz (1928) in his taxonomic attempt (Groves 2002). Schwarz (1928) recognised seven subspecies whereas Coolidge (1929), recognised only two (Groves 2002). The work of Coolidge (1929) has been critically reviewed (Groves 2002), and specifically Haddow & Ross (1951)

investigated multiple inaccuracies relating to sex, age, locality, data arrangement and discrepancies amongst the data and sampling.

Groves' (1967, 1970) research on gorilla systematics and taxonomy reviewed the gorilla species and became the definitive revision following Coolidge's (1929) research and was the accepted standard for more than 30 years (Groves 2002). Groves had acquired a much larger dataset (which included the PCM specimens) consisting of 747 skulls and more than 100 skeletons since Coolidge's research (Stumpf *et al*. 2002). Additionally, methods had developed since Coolidge's attempts, particularly in regard to analytical methods, specifically multivariate analysis, which enabled a more robust and thorough investigation than was available previously (Groves 2002; Stumpf *et al*. 2002). Groves (1967, 1970) reviewed gorilla taxonomy in its entirety, concluding in agreement with Coolidge (1929) that there was one single gorilla species, but unlike Coolidge (1929), recognised three subspecies (Stumpf *et al*. 2002).

However, with the advances in molecular biology, DNA analysis of varying populations revealed the most significant differences were between the eastern and western populations rather than between the mountain and lowland populations (Scally *et al.* 2012). The eastern and western populations were so distinctive the classification was revised, and the gorilla species are currently represented as containing two species: the eastern gorilla (*G. beringei*) and the western gorilla (*G. gorilla*), with each species containing two subspecies: the mountain gorilla (*G. b. beringei*) and the eastern lowland or Grauer's gorilla (*G. b. graueri*), and the western lowland gorilla (*G. g. gorilla*) and the Cross River gorilla (*G. g. diehli*), with each species having a lowland and an upland subspecies (Gordon *et al.* 2013; Mittermeier *et al*. 2013). The Cross River gorilla contains fewer than 250 individuals in the wild dispersed into many subpopulations, and it is classified as the world's most endangered great ape (De Vere *et al.* 2011; IUCN 2019).

Tuttle (2003) noted that the authors in chapters 1-6 of Taylor & Goldsmith (2002) did not offer a proposal of common names for subgeneric taxa of gorillas and recommended that the scheme shown in Table 1.2. be accepted and enforced. The proposed classification by Tuttle (2003) was based on Groves (2001) and it is the taxonomic classification that will be employed in this research.

Table 1.2 Currently accepted gorilla taxonomy

Subgeneric taxa of *Gorilla*	Proposed common name
Gorilla gorilla gorilla	western gorilla
Gorilla gorilla diehli	Cross River gorilla
Gorilla gorilla beringei and *Gorilla beringei beringei*	mountain gorilla
Gorilla gorilla graueri and *Gorilla beringei graueri*	grauer gorilla

1.3. The western lowland gorilla

For the past several decades, research on gorillas has been accumulating in a dispersed and sporadic nature. More recently (the past 25 years), research and literature has attempted to bridge the gap regarding the imbalance of behavioural studies which centred predominantly on the eastern mountain gorilla and morphological studies which have focused primarily on the western lowland gorilla (Taylor & Goldsmith 2002). Of the four gorilla subspecies, the western lowland gorilla is the most numerous (Gordon *et al*. 2013) yet it is not the most studied (Doran & McNeilage 1998). Furthermore, most of our knowledge, particularly in relation to gorilla behaviour, is based on research of a small population of the mountain gorilla in Rwanda (Doran & McNeilage 1998). Therefore, the conservation efforts of western lowland gorilla subspecies would benefit from further studies specifically focusing on them, including behavioural, morphological and genetic. This research, via investigations into the genetic and morphological variation of the western lowland gorilla will therefore add to research specifically on this subspecies and help to bridge the gap in terms of scientific research on this critically endangered primate.

1.3.1. Gorilla distribution and population status

Despite the western lowland gorilla being the most numerous of the gorilla subspecies, estimates of population size have been difficult to obtain due to the dense forest habitat in which they occupy, habituation challenges, and continued observation difficulties (Arandjelovic *et al*. 2010; Hagemann *et al.* 2018). Previous wild population size estimates have varied from a few thousand to a few hundred thousand. The most recent and comprehensive survey to date was completed in 2018 and concluded that wild population size is considerably higher than previous estimates, with a predicted 360,000 wild individuals predicted to persist in extant populations (Strindberg *et al.* 2018; IUCN 2019). Their population is distributed amongst the rainforests of equatorial Africa and they are present in Nigeria, Gabon, Cameroon, Cabinda (Angola), Republic of Congo, Central African Republic, Equatorial Guinea and possibly in the Democratic Republic of Congo, covering over 7,000,000 km^2 (IUCN 2019), and all gorilla subspecies are separated geographically (Fig. 1.1). The western lowland gorilla has a large geographic range (IUCN 2019) found in varying altitudes in west Africa from 100-700 m above sea level (Yamagiwa *et al.* 2003). In comparison, the eastern lowland gorilla subspecies is estimated to have a population of 3800 individuals (Plumptre *et al*. 2016), the mountain gorilla with a minimum of 1004 (Granjon *et al.* 2020) and the Cross River gorilla with 250-300 individuals (Maisels *et al*. 2018).

Figure 1.1 Distribution of gorilla subspecies (based on Scally *et al*. 2012; IUCN 2019).

Estimating population sizes, demography, density and population dynamics is labour intensive and problematic particularly in forested habitats and the western lowland gorilla has proved inherently difficult due to their elusive behaviour and the remoteness of their habitat (Magliocca *et al*. 1999; Hagemann *et al.* 2018). Previous predictions of gorilla abundance focused on habitat availability and assumed that all suitable habitat in the Gabon region at the beginning of the 1980s, was populated by western lowland gorillas (IUCN 2019). This led to a commonly cited population size of 95,000 individuals (Harcourt 1996; Oates *et al.* 2008; Shutt *et al.* 2012). However, Redmond (2008a, b) noted that the population had decreased significantly to approximately half that (50,000 individuals) as recently as the mid-1990s, an estimate that was supported by more recent estimates until the comprehensive survey by Strindberg *et al.* (2018).

1.3.2. Physical traits and characteristics

Western lowland gorillas can be distinguished from the eastern gorilla species as they are on average smaller, lighter and have shorter hair (the eastern gorilla subspecies have notably longer hair). Extreme sexual dimorphism is observed with females being much smaller in comparison to males, approximately half the size (Caillaud *et al*. 2008).

Other physical distinctions include a pronounced brow ridge and wide skull, ears proportionally smaller in relation to the size of their heads. In addition to the females being physically half the size (approximately) of silverback males (i.e. an adult male typically over 12 years of age who has reached maturity (Yamagiwa *et al*. 2003), sexual dimorphism is also displayed in colouration. Coat colour is brownish grey with a crest of auburn or red. Male adults have white/light grey patches of hair present on their thighs which extend up into their back; where the term silverback is derived from distinguishing adult males (Redmond 2008a, b). Sexual maturity normally occurs at 8-10 years (typically males take longer to mature). Gestation is 8.5 to 9 months with one infant being born (twins are rare). Initially, the infant will grasp on to its mothers' stomach and will remain there for approximately the first five months of life. After this time, the infant is carried on the mother's back. Weaning is completed at three to four years of age, by this time the infant has learnt which food products are safe to consume and any additional preparation required for consumption. The primary mode of locomotion is quadrupedal knuckle-walking; however, they spend a proportion of their time climbing and limited amounts of time standing bipedally (Tutin *et al*. 1995; Rowe 1996). Gorillas have a life span of approximately 43-50 years in the wild (McLain & Faulk 2018).

1.3.3. Habitat and diet

Western lowland gorillas primarily inhabit lowland tropical forest (montane and swamp), mostly where herbaceous growth is dense at the ground level (IUCN 2019). This habitat has made estimating population size extremely challenging as mentioned previously. Gorillas are generally considered as folivores (a specialised leaf eating herbivore), with the staple of their diet primarily consisting of leaves, pith and shoots (IUCN 2019; van Casteren *et al*. 2019). Fruit, flowers, bark, seeds and roots are also consumed and omnivorous traits are displayed with the consumption of invertebrates such as weaver ants and termites which complement the diet (Lodwick & Salmi 2019). There have been multiple studies focusing on gorilla diet showing dietary variation according to the different habitats which the subspecies occupy (Table 1.3).

Table 1.3 Summary of gorilla subspecies diets

Subspecies	Diet
Western lowland gorilla	Sugary fruits, stems and seeds. Herbs and bark during the dry season and more fruit based during the wet season.
Cross River gorilla	Primarily tree bark and leaves, some fruit.
Mountain gorilla	Nettles, bamboo, celery and wild berries (when available). Consumes over 142 plant species. Ants and other insects compose 1% of their diet.
Eastern lowland gorilla	Primarily leaves and vegetation, some fruit depending on seasonality.

1.3.4. Behaviour and social structure

Western lowland gorillas predominantly live in one-male (the alpha male or silverback) groups with a polygynous mating system and long-term associations between both sexes (Hagemann *et al*. 2018). A group will usually consist of one silverback male, several females and their offspring, groups of mixed-sex non-breeding groups also exist (Forcina *et al.* 2019). Unusually for primates, maturing males and females both disperse from the natal group (Harcourt & Stewart 2007), with males becoming solitary or forming all-male groups before acquiring females and forming a mixed sex group, whereas females either form a new group by joining a solitary male or directly transfer from one group to another (Forcina *et al.* 2019). Gorilla groups most commonly occur when one or more females forms a long-term association with a silverback male and groups disband when the silverback dies (Harcourt & Stewart 2007; Forcina *et al*. 2019).

Lone silverbacks travel across the terrain, gaining confidence and displaying their strength and confidence to young females in other family groups, they will travel a greater distance than family groups in search for young females, eventually one or more young females will join him and start a new family group. Females leave their family groups at approximately eight years old and tend to give birth to their first offspring within two years (Robbins *et al*. 2004). Secondary transfer is common in female western lowland gorillas which is unusual for mammals, female gorillas may transfer between groups multiple times in their lifetime (Harcourt & Stewart 2007; Forcina *et al.* 2019).

1.3.5. Conservation status and population threats

Western lowland gorillas were listed as vulnerable in 1986 according to the IUCN red list of threatened species (IUCN 2019). Ten years later their status was ranked as endangered and in 2007 they became critically endangered, with their population having sustained a reduction in excess of 80% over three

generations. Habitat loss and fragmentation, infectious disease (such as Ebola virus) and illegal hunting are the main causes of decline in the population (Soto-Calderón *et al*. 2015). Nonhuman primates play an essential role in tropical biodiversity facilitating ecosystem health and contributing to forest regeneration, but anthropogenic pressures have resulted in 75% of primate species with decreasing population trends and 60% are now considered threatened with extinction (Estrada *et al*. 2017). Misconceptions about gorilla strength and alleged ferocity has led to conflicting opinions of the great apes amongst the tribal people sharing its habitat. Often, long respected and regarded as a neighbour or totem animal by tribes, however, this is not always the case. Other native people believe consuming the flesh of these mighty apes will bring them power and strength. For this reason, gorillas are hunted for meat and traditional African medicine, a practice which persists today despite their legal status and protection (Redmond 2008a, b).

1.4. The importance of natural history collections

Throughout the world, millions of specimens are housed in museums and other natural history collections (NHCs) and have emerged as a useful resource for a multitiude of biological baselines (Wandeler *et al*. 2007; Hedrick *et al.* 2020). The collections containing animal specimens yield great opportunities for a varied wealth of studies relating to morphology and analyses of genetic variation of past populations, they are known to have contributed to scientific research for decades (Hebert *et al*. 2004; Austin & Melville 2006; Beissinger & Peery 2007). Hofreiter *et al*. (2003) confirmed via mtDNA analysis of museum specimens that the "mystery ape" termed *Gorilla gorilla uellensis* was not a separate subspecies but was a western lowland gorilla population.

However, collections containing additional contextual data such as field notes, diaries and geographical coordinates are valuable due to their scarcity. This is particularly true for collections with contextual information allowing for geographic analysis (Burgman *et al*. 1995; Shaffer *et al*. 1998; Ponder *et al*. 2001; Gaubert *et al*. 2006). For example, Gaubert *et al*. (2006) analysed 667 specimens belonging to the genus *Genetta* held in fifteen natural history collections and applied ecological niche modelling (ENM) and geographic information systems (GIS) to predict the geographic distributions of three genet species (*G. servalina, G. cristata* and *G. victoriae*). The results indicated that the predicted species distributions far exceeded the protected area given within their distribution. Only 4-6% of their potential range was a protected area (Gaubert *et al*. 2006). This highlighted that the conservation needs of the species were not being met contrary to original belief and prompted their conservation management to be readdressed. Their research indicated that the use of natural history collections can play a significant role in the present and future conservation of species by providing additional

information on the geographic distribution and evolutionary history which is either absent or difficult to obtain from the species at present.

Museum collections provide a resource which is historically unique and can contribute significantly in a variety of ways to molecular studies (Austin & Melville 2006; Lister 2011; Spear *et al.* 2017). Retrieving DNA from specimens can be invaluable in conservation genetic studies where declining or extinct populations and species are the focus (Austin & Melville 2006). However, DNA from NHCs is typically degraded and/or sample size is small (Clifford *et al*. 2004; Wandeler *et al*. 2007; Sproul *et al.* 2017), and often specimens are of unknown origin or records are incomplete (Soberón & Peterson 2004; Pyke & Ehrlich 2010) which makes those specimens with additional accurate contextual information/data even more valuable.

Globally, habitat loss is regarded as the most significant threat to biodiversity (Haddad *et al.* 2015; Pardini *et al.* 2017), and it has long been known that museum collections can be used to map historical species distributions and/or identify important areas for conservation efforts (Ponder *et al*. 2001; Troudet *et al.* 2018). The expanding area of biodiversity informatics applies information technology to primary biodiversity data to assist with analysis and management most specifically at the species level (Soberón & Peterson 2004; Hortal *et al.* 2015; Troudet *et al*. 2018). However, only a limited fraction of the available information is available electronically, with some estimates suggesting less than 10% of museum specimens being accessible electronically (Krishtalka & Humphrey 2000; Canhos *et al*. 2004). Whilst this situation is improving, most museum data is still not computerised or accessible in any electronic format, leaving a considerable quantity of underutilised primary data (Newbold 2010; Hill *et al*. 2012). It is therefore critical that information about collections, and information held by collections, is both connected and accessible (Losos *et al*. 2013). This is particularly true for collections where associated contextual information would allow for wider analyses (Burgman *et al*. 1995; Shaffer *et al*. 1998; Ponder *et al*. 2001; Gaubert *et al*. 2006).

One such museum which is unique in its contextual information supporting its natural history collection is the Powell-Cotton museum in Birchington, Kent, UK. It is this museum that the data for this research regarding the historical population of the western lowland gorilla has been acquired.

1.5. The Powell-Cotton Museum natural history collection

The Powell-Cotton Museum (PCM) is one such museum whose natural history collection contains valuable contextual information rarely found in such detail. The PCM was established in 1896 in the grounds of Quex Park by the explorer Major Percy Powell-Cotton. Figure 1.2 shows the various trip routes and dates taken by him. The PCM contains over 6000 mammal specimens, collected between

1890 and 1940, primarily from equatorial Africa. Many of the specimens were collected by Major Powell-Cotton himself, although large numbers of the specimens were collected by Frederick Merfield, a close friend and collaborator of the Major. Primarily known for its extensive primate collection, the museum also holds important archaeological and anthropological collections created by other members of the Powell-Cotton family.

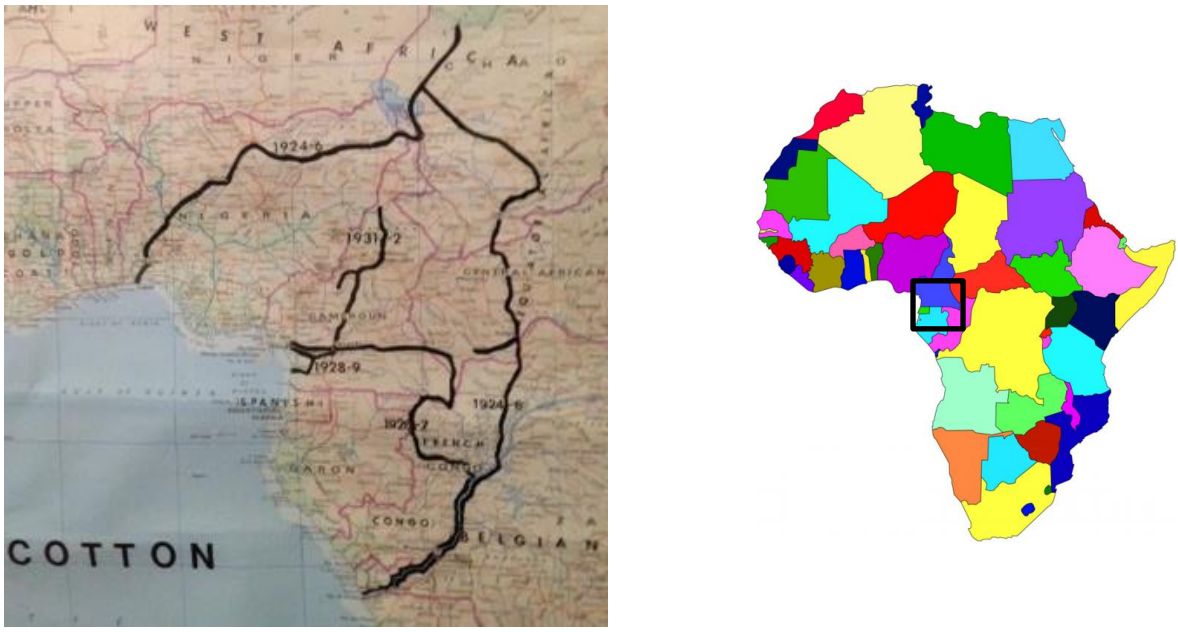


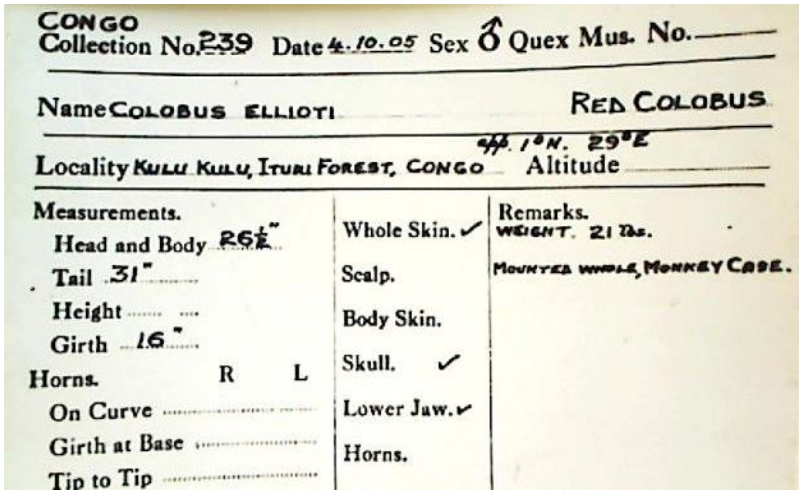
Figure 1.2 The routes and dates of the various exploration trips which contributed to the PCM collection; the map is on exhibition at the Powell-Cotton Museum, Quex Park, in Birchington, Kent, UK. The square on the Africa map (right) depicts the regions the trips explored (left), where the gorilla specimens used in this research originated from (equatorial Africa).

The PCM Natural History Collection has contributed to mammal research, particularly of primates, for many years (e.g. Ashton & Zuckerman 1951; Wood 1976; McHenry 1982; Jurmain 1997; d'Huart & Grubb 2003; Pitra *et al*. 2006; Yamaguchi *et al*. 2009). This is primarily because the collection is vast (over 2000 primate specimens) and, in comparison to many natural history collections of this period, well documented (Jenkins 1990; B. Wood pers. com.), with geographical locations and both morphological and contextual information recorded for most of the specimens in the collection. There remains, however, an abundance of unutilised or generally inaccessible information within the PCM, particularly relating to the geographical location at which specimens were collected.

Most of the research with has utilised the PCM primate collection has focused on morphological studies, particularly of the chimpanzee and gorilla specimens and spans many decades. For example, Zipfel *et al*. (2009) researched the lateral column of the foot for hominin evolution

purposes. Taylor (2006) investigated size and shape dimorphism in great ape mandibles. Hager (1996) studied sex differences in the sciatic notch of great apes and modern humans, and Wood (1979) researched the relationship between body size and long bone lengths in *Pan* and *Gorilla*. One of the earliest publications by Ashton & Zuckerman (1952) investigated age changes in the position of the occipital condyles in the chimpanzee and gorilla.

Major Powell-Cotton was meticulous in his record-keeping. Consequently, and unusually for a collection of this period, geographical coordinates, which would have been recorded with a sextant, and both morphological and contextual information were recorded for most of the mammal specimens in the PCM collection. Information about specimens is stored on specimen cards, with much of this information replicated in an electronic database. In total, the PCM holds over 6000 specimen cards for the mammal collection, with cards normally including the date and location of collection, morphological measurements, sex, and the nature of the museum holdings. Most cards record only a single individual, but some contain information for multiple individuals, e.g. one card records seventeen black colobus monkeys. The data recorded on the specimen cards varies in detail, the date and sex are almost always present, and in most cases the geographical coordinates are given and, depending on which mammal is being investigated, anatomical and morphological information is often detailed. Figure 1.3 is an example of one of the specimen cards for a male red colobus monkey captured on the 4th of October 1905 in the Congo.


CONGO
Collection No. 239 Date 4.10.05 Sex ♂ Quex Mus. No.

Name COLOBUS ELLIOTI RED COLOBUS

Locality KULU KULU, ITURI FOREST, CONGO 1°N. 29°E Altitude

Measurements.
Head and Body 26½"
Tail 31"
Height
Girth 16"

Horns. R L
On Curve
Girth at Base
Tip to Tip

Whole Skin. ✓
Scalp.
Body Skin.
Skull. ✓
Lower Jaw. ✓
Horns.

Remarks.
WEIGHT. 21 lbs.
MOUNTED WHOLE, MONKEY CASE.

Figure 1.3 Specimen card from the Powell-Cotton Museum primate collection for a red colobus monkey detailing sex, capture date, measurements including weight and geographical coordinates.

In addition to the specimen cards, further information relating to the PCM mammal collection is kept in a variety of forms such as notes, diaries and letters. Some of this material is archived in

specific 'trip boxes' that contain notes, letters, receipts, pictures and financial records to specific trips. Whilst Major Powell-Cotton's diaries have been transcribed from the originals, the diaries are not yet freely available electronically and the other contextual material remains in its original form with no electronic duplication. To provide a preliminary assessment of the potential utility of the contextual information, material associated with specific trips and specimens was manually searched for relevant information. The following Table 1.4 gives five examples taken from the gorilla field notes of the Merfield collection. They demonstrate the detail and observations that were recorded in the field which as previously mentioned, is unusual, particularly for this period.

Table 1.4 Examples from the gorilla field notes recorded by Fred Merfield from the Powell-Cotton Museum collection

Example	Collection no.	Description and additional information
1.	**902. Female adult**	Head to fork 36¾". Height 60". Span 78". Girth 44½". Belly 49½". Head round chin 29½". Neck 19¼". Biceps left arm 12¼". Biceps right arm 9½". Forearm left 12½". Forearm right 10½". Thigh 20¾". Calf 10¾". Hand 8½" x 6¼". Foot 11" x 7". Ears 2" x 1¾". **Note.** Left eye looks blind and right one going ? Right hand deformed and very bad sores between fingers, big sore on top of hand. Old scars on wrist, hand or arm not been used for sometime. Forearm skin and bone, first finger looks as if occasionally put on ground. Sexual organs diseased also surrounding parts. The running sores known to natives as "Marjal" they suffer considerably themselves with this disease, they say sores on hands is the same thing.
2.	**932. Female adult**	Head to fork 37". Height 56". Span 78¼". No other measurements taken owing to beast being diseased. Upper lip, chin, right cheek, under side right elbow, right ankle white. Chest and arms big white blotches. Right eye bunged up. All the right side of neck badly diseased and large open sores eaten right through skin, other places epidermis eaten away. **Note.** Disease known as "Marjal" (Yaunde "Mabada) see No 902. Disease of various Gorilla should be interesting to medical men if on the spot.
3.	**340. Male adult**	Very old beast, 2 upper incisors and 1 lower canine missing. All teeth very much worn, canine teeth in upper jaw worn level with other teeth. All Gorilla coming from the forest NW and W. from here (Arteck 3¾ N. 14¼ E.) seem to have smaller crests in comparison with those coming from the E & S districts.
4.	**720. Male adult**	Head to fork 45½". Height 69¼". Span 106". Girth 65½". Belly 66". Head round chin 38¾". Neck 32". Biceps 19". Forearm 19". Hand 11" x 8". Thumb 2½" x 3¾". Middle finger 4¼" x 4". Measurements broken off owing to accident, see letter 14.

		Note. This beast was located in plantation and followed into the forest, where after a time beast turned on chief Key – ar – Bar of village Bey – cum – a – Dee, who threw spear at it and killed it.
5.	**868. Female yg**	Head to fork 25¾". Height 40¼". Span 58¼". Girth 33¼". Belly 36". Ear 2" x 1½". Head round chin 223/8". Neck 12¾". Biceps 8½". Forearm 9". Wrist 6¼". Thigh 14¼". Calf 8". Ankle 7¼". Hand 71/8" x 5". Foot 85/8" x 5". Fingertip to armpit 23¼". Weight 26 kilos. Red crest very noticeable and large, going well down into nape of neck and whole face particularly across eyebrow ridge. Whole beast grey with slightly brownish tinge in places, rather darker on arms and legs, under hair very grey and skin also greyish. Hair on buttocks and bottom of stomach reddish brown, first specimen of this colouration procured.

The PCM has contributed to an abundance of scientific research by providing access to its collections (e.g. d'Huart & Grubb 2003; Pitra *et al*. 2006; Yamaguchi *et al*. 2009; Brimacombe *et al.* 2015; Coutu 2015; Towle *et al.* 2018; Dunmore *et al.* 2019), however, there is still a large amount of information left untouched particularly relating to the geographical records, and in some instances where research has been conducted, these records are incomplete, and work has not been published.

1.6. Geographical Information Systems (GIS) and biodiversity mapping

The Navstar Global Positioning System (GPS) was completed in 1995, and since then, the integration of GPS and Geographical Information Systems (GIS) technology has developed and increased to encompasses a range of ecological and conservation applications (Dominy & Duncan 2002; Nowak *et al.* 2020), with the last decade or so seeing a rapid rise in its application to ecological studies (Wegmann *et al.* 2016). Scholten & de Lepper (1991) described the function of GIS as being more sophisticated in that it can store, manage, and integrate spatially referenced data relating to points, lines and polygons. It can perform spatial queries, conduct geographic analyses and display the data in the form of high-quality maps.

The use of GIS has many applications in the real world, including but not limited to traffic and transport planning, marketing, public and environmental health, land use, agricultural planning, service planning (police, health etc.) and environment and natural resource planning (Scholten & de Lepper 1991). Since the turn of this century, GIS has played an increasing role in biogeography and its

functions have been increasingly applied in ecology and conservation (Nowak *et al.* 2020). Although biodiversity is known to be under immense threat, knowledge of biodiversity is often uncertain (Foody 2008). GIS is increasingly being applied to assist with research and planning for biodiversity related issues (Nowak *et al.* 2020).

The increase in the number of publications using GIS reflects the growth and application of this useful technological tool. Andersen (2008) reported that the number of published papers in the journal Landscape Ecology in the previous ten years, had approximately doubled. In July 2019, there were approximately 4,940,000 publications on Google Scholar using the simple search criteria of 'geographical information system'. The application of GIS tools has rapidly grown, its use has diversified in terms of applications in ecology and conservation, and a multitude of species have found themselves subjected to GIS studies. Primates are no exception, and numerous studies have employed GIS for a variety of conservation purposes (Gregory *et al*. 2014). For example, Kamilar *et al*. (2012) used GIS-based data to test Bergmann's rule (i.e. an ecogeographical rule whereby animals from colder climates are larger than those from warmer climates) and the resource seasonality hypothesis in Malagasy primates.

GIS has been used to create Habitat Suitability Models (HSM) to predict the distribution of a species based on environmental data and occurrence records (Gregory *et al*. 2014), including many species of primates. For example: Buckingham & Shanee (2009) investigated this in the critically endangered yellow-tailed woolly monkey (*Oreonax flavicauda*) (IUCN 2019) which has been on the list of the worlds twenty-five most endangered primates since 2006 (DeLuycker *et al*. 2007). They used GIS to create an HSM for assessing the current distribution of the species, thus aiding current conservation initiatives and to determine conservation priorities for the future management of the species. Shanee *et al*. (2013) used the same GIS application to determine an HSM for the Andean titi monkey (*Callicebus oenanthe*), another critically endangered primate (IUCN 2019). They used GIS to evaluate threat levels to habitat areas highlighted by the HSM, again for current of future conservation planning of the species. Boubli & De Lima (2009) modelled the geographic distribution of the brown-backed bearded sakis (*Chiropotes israelita*) and three black uakaris (*Cacajao melanocephalus, C. hosomi,* and *C. ayresi*). Their analyses identified areas of high probability for which the species may inhabit, additionally it identified potential areas for study and field/survey expeditions. Waters & Ulloa (2007) executed a preliminary survey on the current distribution of primates in Belize. Their study focused on the only two non-human primates in Belize, the black howler monkey (*Alouatta pigra*) and the Yucatan spider monkey (*Ateles geoffroyi yucatanensis*), listed as Endangered and Vulnerable, respectively, on the IUCN Red List (IUCN 2019). They conducted a countrywide assessment of wildlife/human conflict among subsistence farmers in Belize via qualitative methods involving questionnaires and recorded the GPS locations for each species reported by the respondents. A GIS

map was created showing the distribution of the two species generated from the acquired data. The report was an initial first step in identifying populations of the two species residing outside of the protected areas that require active conservation management.

GIS can also be used to identify disease transmission patterns (Gregory *et al*. 2014) as in the research conducted by Pigott *et al*. (2014) who investigated and mapped the zoonotic Ebola virus. Using GIS and species distribution models, Pigott *et al*. (2014) collated all recorded zoonotic transmissions of the Ebola virus infection to humans from primates and bats to predict a zoonotic transmission niche spanning twenty-two countries.

Replication of potential routes of gene flow can also be achieved with the application of GIS methods (Gregory *et al*. 2014). Quéméré *et al*. (2010) investigated the impact of forest fragmentation in relation to patterns of genetic differentiation in the endangered primate, the golden-crowned sifaka (*Propithecus tattersalli*). Additionally, GIS can allow for the study of navigation routes and strategies amongst primates (Phillips *et al*. 1998; Gregory *et al*. 2014; Siljander *et al*. 2020). Examples of this application include Porter & Garber (2014) who investigated daily movement patterns of Weddell's saddleback tamarins (*Saguinus fuscicollis weddelli*) using GIS, and Hopkins (2016) who researched mantled howler monkey (*Alouatta palliata*) groups and their arboreal pathway networks. Siljander *et al*. (2020) used GIS to analyse the vulnerability of households to crop-raiding by primates, which is of concern to people living near protected areas and increases human-wildlife conflict.

GIS and remote sensing applications are continually being developed and utilised, more recent studies employing these techniques for primate conservation and ecology include Mekonnen *et al.* (2020), Moraes *et al.* (2020) and Siegel *et al.* (2020). These mapping applications are globally utilised for an abundance of research and monitoring purposes for diverse scenarios, currently, two of the most prevalent crises facing humans and animals are the climate change induced bushfires in NSW, Australia and the global pandemic, COVID-19. GIS and remote sensing applications are being employed in a multitude of studies relating to these life-threatening crises, e.g. Mutai & Chang (2020), Todd & Maurer (2020), Martellucci *et al.* (2020) and Zhou *et al.* (2020).

This thesis aims to utilise the geographical information from the records available in the Powell-Cotton Museum (PCM) Natural History Collection and make it more accessible for future research. The geographic information for the western lowland gorilla specimens is investigated for its potential use in following chapters, with the aim of utilising the geographical data with morphometric and genetic data for regional comparisons. Furthermore, this research utilises the contextual information contained in the 'trip boxes' and Major Powell-Cotton's diaries, which have not been previously investigated to this scale. Although this thesis focuses on the western lowland gorillas, the same approach can be applied to other species, presented here was an early attempt to digitise and map the entire collection to demonstrate the utility of the collection to multiple species.

1.7. Biodiversity mapping of the Powell-Cotton Museum data

To facilitate the following chapters of this thesis and for future research/accesibilty purposes, all the specimen cards relating to the PCM mammal collection were digitised using a Logitech C905WEBCAM. The images were collated and renamed to correspond with the mammal database of the museum catalogue. The geographical coordinates recorded on all the mammal specimen cards were added into the mammal database. In total, 6164 specimen cards were digitised, which provide information for all 6429 specimens in the PCM mammal collection. Over a third of the specimen cards represent primates, however, the remainder of the mammal collection represents a variety of fauna including (but not exclusively): lions (20 specimens), leopards (35 specimens), bears (25 specimens), duikers (804 specimens), antelope (101 specimens), hartebeest (131 specimens), elephant (49 specimens) and buffalo (160 specimens).

Geographical coordinates are provided on the specimen cards for 5449 of the 6429 mammal specimens. This primary biodiversity data was used to produce maps for each species in the collection, the geographical data from the updated mammal database was visualised using ArcGIS version 10 (Redlands, 17 California). The maps for all species are available from the collection's manager at the PCM (http://www.powell-cottonmuseum.org/). Of the total 6429 specimens in the collection, 181 are of Asian origin. Of those 181 Asian origin specimens, only 29 have geographical coordinates recorded. Specimens of African origin (of which there are 6248 in total and 5420 with geographical coordinates available) are predominantly from sub-Saharan Africa and represent distributions for great apes, monkeys, carnivores and other fauna.

The primary biodiversity data from the PCM for the 5449 specimens in the mammal catalogue which have known geographical coordinates is displayed in Figure 1.4 and includes information for the great apes (a), monkeys (b), carnivores (c) and other fauna (d) in the collection.

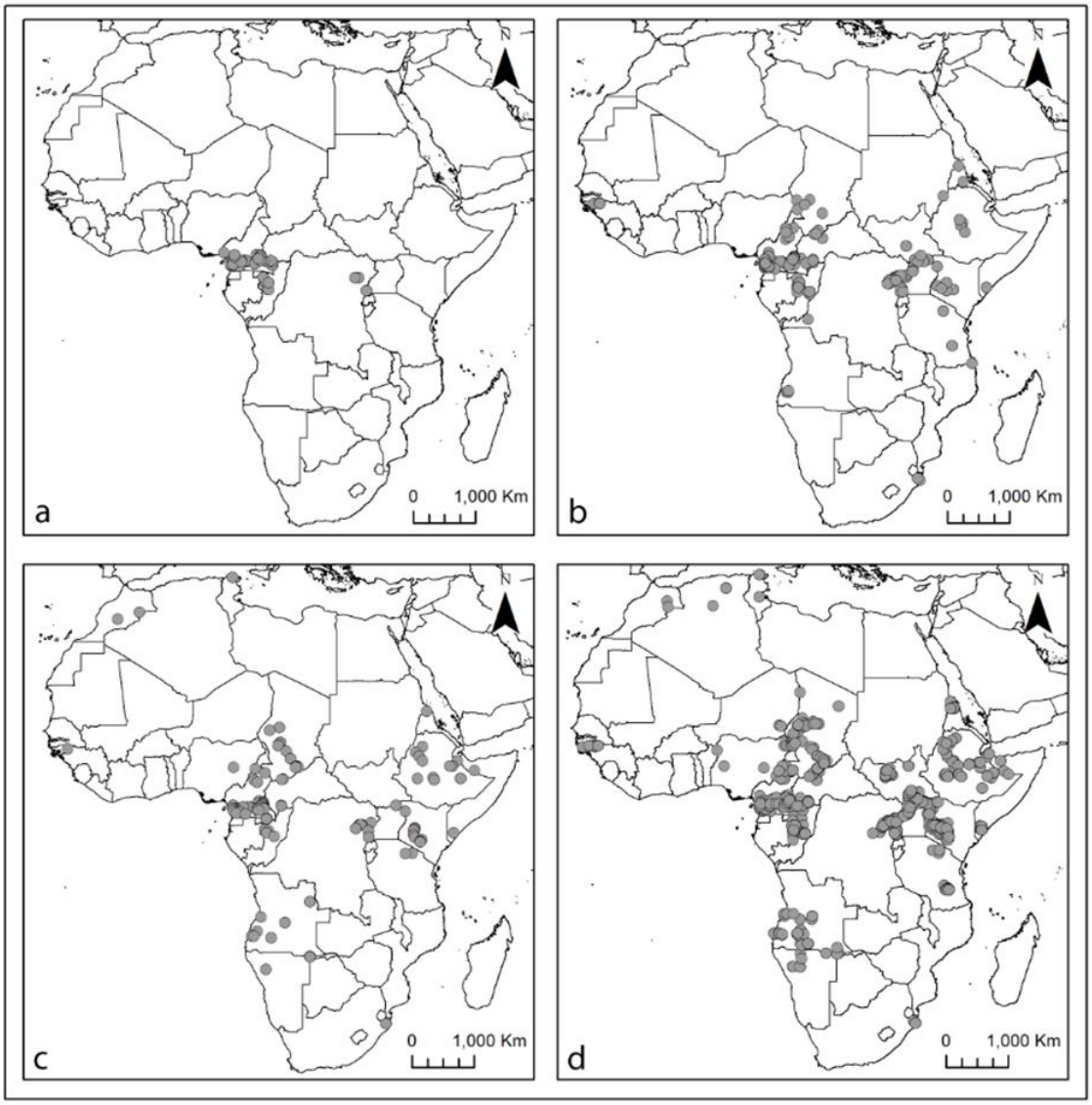


Figure 1.4 Primary biodiversity data from the Powell-Cotton Museum for the 5449 specimens in the mammal catalogue which have known geographical coordinates. Maps represent (**a**) the great apes, (**b**) monkeys, (**c**) carnivores, and (**d**) other fauna in the collection.

1.7.1. Western lowland gorilla mapping

GIS methods have also been applied to studies involving the great apes and are varied in their application as with previously mentioned studies. GIS has been used to investigate behaviour, e.g. Smith (2014) who mapped spatial movements, behaviours and interactions of captive orangutans with the application of GIS methods for furthering our understanding of their behaviour with each other, and their interaction with their surrounding environment.

Habitat use by humans and threats to great ape habitat has been studied by Morgan (2007) who produced best practice guidelines using GIS methods, in relation to the impact of commercial

logging on great apes in Western Equatorial Africa (WEA). Over half of the gorillas and chimpanzees' range in WEA (at the time of the publication) was allocated to logging concessions and 36% of the total area, which was considered an exceptional priority area for ape conservation, was subject to logging concessions. Additionally, Bender & Ziegler (2011) applied GIS methods to investigate threats to gorilla habitat in the Congo basin which included forest loss, mineral exploitation, armed conflicts, human footprint and the Ebola virus.

The Global Ape Populations, Environments and Surveys (A.P.E.S) status report (Campbell *et al*. 2012; http://apesportal.eva.mpg.de/) used GIS methods to summarise ape population status for all four African ape species and the two Asian ape species, as well as information for all subspecies belonging to the African ape species. It was the first report based on the APES database and provides information on ape populations over a large scale. The report found that most ape species were experiencing a drastic reduction in suitable habitat, with gorillas in Africa experiencing the worst of the decline.

This research employs biodiversity mapping to investigate the regional distribution of western lowland gorilla specimens within the PCM. The data is used to assess regional morphological and genetic variation of the subspecies. In total, the PCM holds 242 specimen records for *Gorilla gorilla gorilla*, of which 239 have the geographical data recorded. Employing the same methods previously described, a map was produced showing the distribution of the gorilla specimens from the PCM (Fig. 1.5).

Visualising this data assists with other methods used in this research, as regional variation amongst the western lowland gorilla subspecies is under investigation. Figure 1.5 enables the specimens to be broadly defined into three regional clusters in terms of their distance from one another. However, it is important to acknowledge that the figures represent the distribution of the specimens/samples in the PCM collection and generally follows previous literature relating to identification of demes by Groves (1970), not of the distribution of the subspecies in its entirety. Due to the widespread distribution of the western lowland gorilla, it is possible, that the defining of regional groups would change if there were more specimens present. However, using the data available and as an initial starting point, three regional subgroups were defined as the following regions A, B and C.

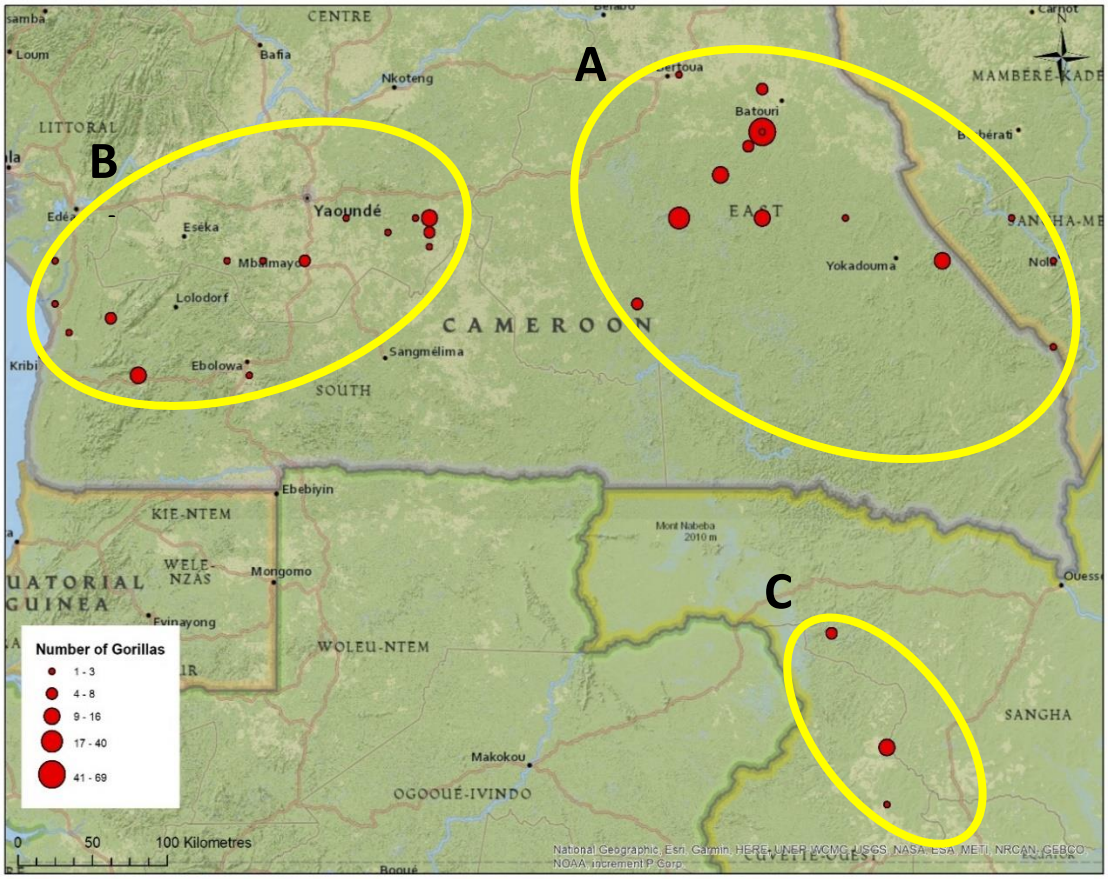


Figure 1.5 Distribution of the 239 western lowland gorilla specimens from the Powell-Cotton Museum for which geographical coordinates were recorded. The gorilla specimens have been defined into three regional subgroups: A, B and C, based on furthest geographical distance from one another and broadly following Groves (1970).

From this initial regional clustering, specimens could then be grouped and investigated further in terms of regional morphological variation (Chapter 2) and regional genetic variation (Chapters 3 and 4). There are, however, many other investigations where the visualisation of this data could be of use but is outside the scope of this research. For example, more detailed investigations into home ranges, age related demographics, sex distribution and habitat occupancy are all important areas of research which contribute to the conservation of this critically endangered primate.

This preliminary biodiversity mapping is useful for researchers interested in a specific geographical region and/or specific species in a certain location. The majority of the PCM primate collection is from Cameroon and the Republic of Congo, followed by the area of Kenya, Uganda and eastern Democratic Republic of Congo, with a clear absence of specimens centrally (e.g. Fig. 1.4.b and d). The pattern of specimen data correlates with the documentation regarding the locations of Major Powell-Cotton's expeditions, adding credibility to the coordinates of the specimens.

In relation to the western lowland gorilla, the visualisation of the specimen data is a useful tool for a multitude of scientific investigations, its primary purpose for this study is to assist with the identification of regional areas to further morphological and genetic analysis. The mapped results have enabled regional areas to be defined for subsequent analysis, and in this respect the investigations of this chapter have served to provide the necessary data for continuous investigations of the western lowland gorilla. In addition, the results of the mapping in general, have confirmed and reiterated the importance of NHCs for species conservation purposes, and supports other publications that have highlighted the importance of NHCs, e.g. Kitchener (1997), Shaffer *et al*. (1998), Wandeler *et al*. (2007), Lister (2011), Holmes *et al*. (2016), and Kharouba *et al*. (2019).

Additionally, it has highlighted some important caveats relating to the use of museum specimens which need to be acknowledged in studies including them. Firstly, the PCM is regarded as exceptional in terms of its additional contextual information, this is not the case for all NHCs, therefore caution must be exercised particularly for studies using museum specimens for regional comparisons where the data may not be as precise or reliable. Secondly, the species under investigation will be limited by the specimens available. This will be true of all NHCs which means there is certainly the potential for bias to exist when analysing data, however, this does not mean that NHCs are not useful for species conservation purposes, it simply highlights that caution must always be given to the data and any potential bias addressed and where possible minimised. This may include increasing sample size in some instances, if possible, within the collection or by collaboration with multiple NHCs, if a sufficient number of specimens does not exist in the NHC being investigated. In terms of the geographical data, if an NHC does not hold the information, then simply, there is nothing that can be done to rectify that, the records are what they are. This is the very reason why the collection at the PCM is remarkable, as it allows for an accurate regional analysis of the western lowland gorilla (and other species) which is absent from most of the previous literature at this scale.

1.8. An introduction to geometric morphometrics

The study of shape variation and its covariation with other variables is termed morphometrics (Bookstein 1991; Dryden & Mardia 1998; Adams *et al*. 2004). It is the measurement (metron) of shape (morphe) and is a section of statistics that has a long history dating back to the origins of statistics as a discipline (Mitteroecker & Gunz 2009). In morphometric studies, it is essential to distinguish between shape and form. The shape of an object refers to the geometric properties that are constant, or invariable irrespective to rotation, scaling and translation (Mitteroecker *et al*. 2013). Shape is distinctly different from form in that relates to the geometric properties that are constant only in

relation to translation and rotation, not scaling. Form can therefore be defined as size and shape (Dryden & Mardia 1998; Mitteroecker *et al*. 2013). The size of an object can be defined as a length, volume, area, weight or centroid and is commonly represented as a single measure i.e. total weight (Richtsmeier *et al*. 2002).

In traditional linear morphometric studies, size is commonly the only estimator used (Outomuro & Johansson 2017). However, because the analysis of shape and shape change is a fundamental component of much biological research (Richtsmeier *et al*. 2002; Slice 2007), during the 1960s and 1970s multivariate statistical tools were beginning to be utilised by biometricians to analyse shape variation among and within groups (Adams *et al*. 2004). This approach was termed multivariate morphometrics (Blackith & Reyment 1971) or more commonly, traditional morphometrics (Marcus 1990; Reyment 1991; Adams *et al*. 2004, 2013; Slice 2007). Traditional morphometrics encompassed the application of multivariate statistical analyses to an assembly of morphological variables (Adams *et al*. 2013). Commonly, measurements of the linear variety were used, however, ratios, counts and angles were occasionally incorporated (Slice 2007).

Multivariate morphometrics, while combining quantitative morphology and multivariate statistics, has its limitations (Adams *et al*. 2004, 2013). Measurements of the linear variety are frequently correlated with size (Bookstein *et al*. 1985; Cooke & Terhune 2015) which led to considerable time creating methods for size correction (Sundberg 1989; Jungers *et al*. 1995; Adams *et al*. 2004). An abundance of methods was recommended but opinions on which method was the best, and should therefore be applied, differed immensely, due to each method yielding marginally different results, methodological disagreements were a fundamental problem in the field of multivariate morphometrics (Adams *et al*. 2004). Moreover, due to many distances (such as maximum width) not being characterised by homologous points, the assessment of linear distances in homologous terms, was arduous (Adams *et al*. 2004). Additionally, two disparate shapes could produce duplicate results (Fig. 1.6) because the data relating to the locations of where the distances were made in relation to one another was absent (Adams *et al*. 2004). The final limitation presents itself with the lack of geometry captured from a data set of linear distances, resulting in the loss of some of the elements of the shape, therefore, different methods for quantifying and analysing morphological shape were required (Adams *et al*. 2004).

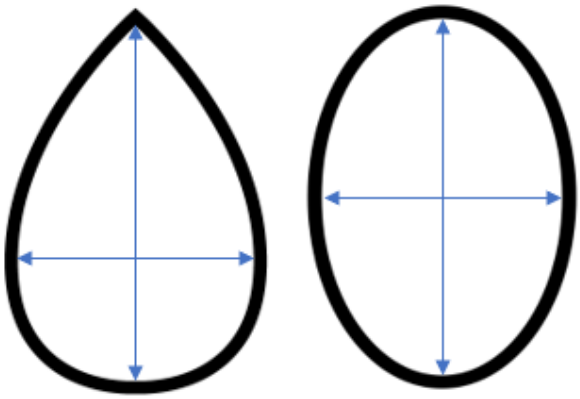


Figure 1.6. An example of two disparate shapes that can produce the same length and width measurements.

The issue which captured a great deal of attention was the loss of the geometry regarding the morphological structure. Synchronously, statisticians composed and refined a robust statistical theory for shape analysis (Adams *et al*. 2004). This simultaneous development of methods combined the methods of visualisation in biological form with multivariate statistical methods, known as the "morphometric synthesis" (Bookstein 1996; Adams *et al*. 2004, MacLeod 2017). Consequently, a transition occurred in the late 1980s/early 1990s which observed a change in the way morphological structures were quantified and the methods of data analyses (Adams *et al*. 2004). In 1993, the emergence of the term 'geometric morphometrics' was coined. This was a new approach born from a review by Rohlf & Marcus (1993) in the field of morphometrics and was considered a "revolution in morphometrics" (Rohlf & Marcus 1993; Adams *et al*. 2004, 2013). The term 'geometric morphometrics' refers to the analysis of morphological structures using Cartesian geometric coordinates as opposed to linear, provincial or volumetric variables (Lawing & Polly 2010). The first geometric morphometrics methods to be applied were outline methods followed by landmark-based methods (Zelditch *et al*. 2004; Adams *et al.* 2013; MacLeod 2017). In morphometrics, landmarks are defined as discrete anatomical loci (also known as homologous points) that can be recognised in all specimens of the study as the same point (Zelditch *et al*. 2004; MacLeod 2017).

Instead of using linear measurements as a size variable, in geometric morphometrics, centroid size is the most commonly applied size estimator (Richtsmeier *et al*. 2002; MacLeod 2017). Centroid size is defined as the square root of the summed squared Euclidean distances from each landmark to the centroid of the configuration of landmarks, and it describes the object's overall size reliably (Zelditch *et al*. 2004; Niewoehner 2005; Lycett *et al*. 2010; Adams *et al.* 2013).

Registration-based morphometry refers to systems that insert organisms into a common frame of reference, preventing inconsistent differences in orientation, from being inappropriately

interpreted as biologically relevant results in variation in form (Cole 1996; Cooke & Terhune 2015). However, registration methods often lead to inaccuracies (Richtsmeier & Cheverud 1986). Cole (1996) proclaims that Sneath (1967) is commonly cited as the earliest use of "registration-free" morphometric methods. Sneath (1967) applied the superimposition (Procrustes) technique in a study comparing living and fossilised cranial shapes of human and great apes. Procrustes is an analysis theory/technique that relates to statistical shape analysis. The name originates from Greek mythology; Procrustes was a giant who tortured his victims. He would make his victims fit his iron bed either by stretching their bodies or cutting off body parts that were longer (Crosilla *et al*. 2019) – thus 'fitting' his victims to his bed, albeit the resulting shape change which is not the intention of Sneaths' (1967) superimposition method. In addition to the Procrustes technique, other registration free methods to compare form have been developed, these include thin-plate splines, relative warp analysis, finite element scaling and Euclidean distance matrix analysis (Cole 1996). However, Cole (1996) argues that an earlier paper by Boas (1905) on the "method of least differences" illustrated a geometric morphometric method and acknowledged the biological inaccuracies of registration systems. Boas is well regarded as the founder of the modern anthropology science, but Cole (1996) notes Boas' contribution to morphometrics and his lack of recognition for it in the literature of morphometrics. Cole (1996) also acknowledges Phelps (1932) who continued and expanded Boas' (1905) method but also, like Boas, is rarely cited despite publications in prominent journals. The early 1990s marked the beginning of a notable increase in geometric morphometric publications and citations, as biologists acknowledged the superiority of geometric morphometric methods and applied them more avidly to a variety of hypotheses (Adams *et al*. 2004).

Biological processes such as mutation, disease, injury, local geographic adaptation, ontogenetic development or long-term evolutionary diversification produce variations in shape between individuals or components of them, thus the analysis of shape and size performs an important role in an abundance of biological studies (Zelditch *et al*. 2004). Geometric morphometrics measures phenotypic shape and size changes, helping the development of hypotheses in research that require a sophisticated quantitative representation of the phenotype in terms of genetic, functional or developmental attributes (Lawing & Polly 2010).

The application of geometric morphometric analysis to biological questions is vast and there have been many studies since the geometric morphometrics 'revolution'. Anthropological research and geometric morphometrics methods follow suit in their application across species of primates, and research into dentition, identification and characterisation of species and their relationships (Clark 1950; Irish 1998; Bailey 2002; Gómez-Robles *et al*. 2007) and ecological studies (Gómez-Robles *et al*. 2007) reiterate the capability of geometric morphometric techniques to accurately determine morphological variation among species. Geometric morphometrics can also be used to ascertain

sexual dimorphism, for example from lizards (Kaliontzopoulou *et al*. 2007) to fur seals (Oliveira *et al*. 2005).

Previous studies show the realm of literature in the growing field of geometric morphometrics, demonstrating the diversity of its applications across and within species, from plants (Viscosi & Cardini 2011) and insects (Tofilski 2008) to small/medium mammals (Cardini *et al*. 2010) and larger bodied animals (Bignon *et al*. 2005) including *Homo sapiens* and other hominins (Hennessy & Stinger 2002; Perez *et al*. 2006; Gayzik *et al*. 2008). Regardless of the questions being asked, and to what species they concern, one theme which is prevalent in the majority of the literature is that geometric morphometrics supersedes traditional and standard techniques and adds scientific rigour to the research in question. Aside from anthropological geometric morphometric research (Franklin *et al*. 2007; Gómez-Robles *et al*. 2007) other primate species have also been the centre of geometric morphometric studies, including chimpanzees and gorillas, detailed in Chapter 2.

Geometric morphometrics continues to rapidly evolve and is considered a valuable tool for biological studies (Adams *et al*. 2013; MacLeod 2017). One of the more recent and most significant changes in geometric morphomterics is the use of three-dimensional data (Adams *et al.* 2013), which has seen a rapid growth in its applications across a variety of fields including ecology and primatology. For example, Garrod (2017) investigated *Chlorocebus* monkeys and patterns of island evolution via mitochondrial and cranial three-dimensional geometric morphometric analyses. Fiorenza & Bruner (2018) investigated cranial shape variation in adult howler monkeys and Ito & Koyabu (2018) focused on biogeographic skull morphology of dusky leaf monkeys.

1.9. The role of wildlife parks/zoos in species conservation

Wildlife parks and zoos have held a prominent place in our society for over a century, yet the last couple of decades have seen them undergo significant changes in their structure and function (Tribe & Booth 2003). It is estimated that there are 10,000 zoos worldwide, defined as collections of captive wild animals that are displayed to the public so that they are easier to observe than in nature (Tribe & Booth 2003), and for many people, zoos may be the only place they are likely to observe the species (Consorte-McCrea *et al.* 2016), albeit in captivity, rather than their native habitat. Zoos have an important role in species conservation via a multitude of projects, ventures and activities including field conservation projects and increasing public awareness (Breuer *et al.* 2018). Despite their place in society and the fact they attract approximately 700 million visitors annually (Gilbert & Soorae 2017), there has been much controversy and lengthy debates for decades regarding the ethics and morality of keeping animals in zoos, with members of the public, professionals, educators and conservationists having diverse and varied opinions, and often highly emotive. There are arguments for keeping, or

not, animals in captivity. One of the objections to zoos often emerging is “Some zoos are bad” (Safina 2018). There are some zoos worldwide which clearly do not have animal welfare as a priority. However, the introduction of the Zoo Licensing Act (1981) aims to ensure that all animals that are kept in enclosures in Great Britain are provided with a suitable environment. In addition, the environment is to provide opportunities for the animals to express normal behaviour, and the Act also requires licensing and inspection of all zoos. Section 1A of the Act specifies conservation measures for zoos which they are required to abide and specifies the breeding of wild animals in captivity, and/or the repopulation or reintroduction of wild organisms into an area (Zoo Licensing Act 1981, section 3.1iv and 3.1v).

Arguments against zoos include valid points in terms of animal health and welfare, animal rights and ethics (Regan 1996; Hogan & Tribe 2007; Morgan & Tromborg 2007). Some of the more poignant arguments against zoos which are often stated have been discussed in many publications. Jamieson (1985), Bertram (2004) and Lin (2014) have highlighted the following points as arguments against zoos:

- Keeping animals in captivity for entertainment is degrading and wrong.
- Zoos cull surplus animals.
- Restricting an animal’s freedom is wrong.
- Captive animals suffer from more stress due to confinement.
- Removing individuals from the wild will further endanger the wild population.
- An individual’s rights should not be superseded for the sake of the species.
- Baby animals are bred to attract visitors and leads to overpopulation.
- Animals do not live as long in captivity as they would in the wild.
- Animals on occasion, escape their enclosures endangering themselves, other animals and people.
- Conservation efforts are hindered because zoos draw attention away from the wild animals.
- Funds should be spent on conserving wild populations, not captive individuals.
- Zoos are not educational because they send the wrong messages.
- Reintroduction of animals back to the wild does not work.
- Television and film productions make zoos unnecessary, people can go and see them in the wild.
- Zoos conserve single species whereas we should be conserving the habitat.
- Zoos only save a small proportion of all the species under threat.

Whilst many of these arguments are valid there are also very good reasons for zoos to exist. These include the following which have been discussed in various publications, e.g. Chiszar *et al*. (1990), Hutchins *et al*. (2003), and Borrell (2016), and are summarised here:

- Currently 39 species listed by the IUCN which only exist in captivity, they are extinct in the wild, without zoos, these species would be entirely extinct.
- Zoos promote a connection with the wild, with 700 million visitors on average per year, education and awareness is improved and raised.
- Zoos raise funds for conservation purposes.
- Captive animals teach us about animals in the wild, we can learn about their behaviour, reproduction and diet thus aiding scientific research.
- Captive populations are 'insurance' for wild populations should the wild population become extinct; they are a 'store' for individuals as well as a genetic bank.
- Captive animals are used for reintroduction programmes.

Regardless of personal opinions surrounding the keeping of animals in captivity. There have been numerous success stories for reintroductions programs, that without a captive breeding program, the species would likely have become extinct. For example, the Arabian oryx *(Oryx leucoryx)* became extinct in the wild in 1972 due to overhunting, but captive breeding and reintroduction programs saw the species surpass a population of 400 individuals in the Arabian Peninsula during the 1990s (Sodhi *et al*. 2011). Habitat loss and fragmentation reduced the golden lion tamarin *(Leontopithecus rosalia)* to 562 individuals in the early 1990s, but successful reintroductions from captive populations saw the population boost to over 1500 individuals (Sodhi *et al*. 2011). One conservation foundation that is involved in captive breeding and reintroduction programs is the Aspinall Foundation located in Kent, UK. The Foundation has permitted this research to be performed and has provided the necessary samples to analyse a contemporary, captive population of western lowland gorillas at a molecular level. The Aspinall Foundation is described in more detail in Chapter 3.

The US Endangered Species Act of 1973 was introduced with the purpose of protecting critically imperilled species from extinction (Hemley 1994; Huxley 2013). The Act came into force on 28th December 1973 and is the enacting legislation for the implementation of the provisions set out in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (Hemley 1994; Huxley 2013). CITES is a multilateral treaty for the protection of endangered species, and it was created as a result of a meeting in 1963 by members of the IUCN. CITES came into force 1st July 1975 with the purpose of ensuring that international trade of wild plants and animals does not threaten the species' survival in the wild (Huxley 2013). The Endangered Species (Import and Export)

Act of 1976, is the UK legislation that supports CITES, and it came into force on 31st October 1976 (Endangered Species Act 1976).

Gorillas are listed as critically endangered by the IUCN (2019) and are listed on Appendix I of CITES, which means they receive the highest level of legislative protection and trade is prohibited (Hemley 1994). The introduction of the aforementioned legislation meant that taking gorillas from the wild for any purpose, including to house them in captivity, ceased and remains enforce. Therefore, there have not been any introductions to the captive populations of gorillas since the mid-seventies (Nsubuga *et al*. 2010). Mace (1988) reported the International Studbook of the gorilla (Kirchshofer 1985), recorded a total of 563 wild gorillas introduced to captivity from 1935 until the legislation came into force. The oldest known gorilla in captivity who was taken from the wild, Trudy, recently died in 2019 aged 63. The average captive lifespan ranges for gorillas are 40-55 years (Perez *et al*. 2013). In 1956, the first gorilla was born in captivity with a further 379 recorded from 1956 – 1988 (Mace 1988).

1.10. The importance of genetic diversity

Until the 1960s, the evolutionary establishment of relationships among species, was primarily based on morphological data (Moritz 1995; Clifford *et al*. 2003). Although genetic studies focusing on gorillas are comparatively few, their significance is becoming increasingly relevant. Genetic studies often complement ecological and morphological data but can also contradict it (Garner & Ryder 1996; Hillis *et al*. 1996; Harris & Disotell 1998; Gagneux *et al*. 1999; Clifford *et al*. 2003).

Genetic diversity, which is the variation of alleles and genotypes within a population, is the foundation on which the potential for adaptation and evolution depends (Frankham *et al*. 2002; Garner *et al*. 2005; Leigh *et al.* 2019). Genetic variation is a trait of populations and of the individuals within the populations (Lacy 1997; Leigh *et al.* 2019). The characterisation of genetic variation is commonly represented by the percentage of loci at which an individual is heterozygous (Lacy 1997; Leigh *et al.* 2019). A locus (plural loci) is essentially a place/position on the chromosome, and in diploid organisms (e.g. primates) there are two alleles at one locus, one allele inherited from each parent (Hamilton 2011). Heterozygosity can be simply described as when the two alleles at a given locus are chemically different, and homozygosity refers to when they are chemically the same (Hartl & Clark 1997). Thus, heterozygosity promotes evolutionary potential for adaptation as it provides a diversity of alleles for natural selection to occur (Hartl & Clark 1997). Heterozygosity is reduced by inbreeding (mating between closely related individuals) and genetic drift which leads to a greater chance of homozygosity in populations, as the probability of two identical alleles at a locus being inherited by future generations is increased (Lacy 1997). Inbreeding depression is the collective term for the array of impacts which can be observed from inbreeding. The impacts of inbreeding reduce an individual's

fitness in a multitude of ways which include slower growth, higher mortality, reduced mating ability, increased susceptibility to disease, lowered ability to tolerate stress, increased developmental defects, developmental instability, reduced intra- and inter-specific competitive ability (Lacy 1997). Inbreeding not only affects an individual but can have severe consequences for the population's fitness as a whole. Inbred populations are at higher risk of becoming extinct, due to the lower fecundity and survival rate of individuals that are inbred which ultimately slows the growth rate of the population. This is particularly true for populations that are also under other types of pressure in stressful environments, including habitat loss and fragmentation (Lande 1988; Avise 1994; Frankham 1995; Lacy 1997). Prolonged bottlenecks in terms of population size increase the loss of genetic variation further via genetic drift (Lacy, 1997). Therefore, a population that has low heterozygosity, low heritability and few polymorphic loci will be slower to adapt to the pressures of selection than a more diverse population (Avise 1994; Frankham 1995; Lacy 1997). Thus, the importance of conserving genetic variation in populations is paramount for their long-term survival (Garner *et al*. 2005). The IUCN recognises the preservation of genetic diversity as one of three conservation priorities (Garner *et al*. 2005). The evidence now suggests that the genetic consequences of a small population size may be a more significant threat to survival than previously identified (Bergl *et al*. 2008; Frankham *et al.* 2017). In addition, fragmented populations often contain lower levels of genetic diversity versus populations in continuous landscapes, due to the increase in genetic drift and restricted gene flow (Fünfstück & Vigilant 2015).

The "genetic health" of a species or a population is a term often used in molecular studies, e.g. Citek *et al*. (2006), Lippe *et al*. (2006), Kendall *et al*. (2009), Nsubuga *et al*. (2010), Zeisset & Beebee (2010) and Flanagan *et al.* (2018), and it refers to a species/populations genetic diversity and inbreeding level (Muniz *et al*. 2019) with endangered species generally having low gene flow and reduced levels of genetic diversity because their populations tend to be smaller and/or fragmented (Honnay & Jacquemyn 2007; Muniz *et al*. 2019). In this thesis, where the term genetic health is used, it is referring to levels of genetic diversity and inbreeding, with populations/groups having higher levels of genetic diversity and lower levels of inbreeding being considered as showing greater genetic health than others.

1.10.1. Mitochondrial DNA

Mitochondrial DNA (mtDNA) is inherited via the maternal lineage in a single copy and studies have proven that on average it evolves significantly faster (five to ten times) than nuclear DNA (DeSalle *et al.* 2017). MtDNA studies are particularly useful for population genetics research in mammals and there have been numerous publications supporting this, e.g. Taberlet & Bouvet (1994) on brown

bears, Matthee & Robinson (1999) on roan and sable antelopes, Effenberger & Suchentrunk (1999) on otters and Vega (2010) and Vega et al. (2016) on pygmy shrews. There is also an abundance of mitochondrial research on primates published over the years, e.g. Brown *et al*. (1982), Hayasaka *et al*. (1988), Hasegawa *et al*. (1990), Horovitz & Meyer (1995) and Fredsted *et al*. (2004) and continues to more recent years, e.g. Debray *et al*. (2018), Zinner *et al*. (2018) and Zahidin *et al.* (2019).

MtDNA research has often been described as the workhorse of molecular studies particularly in relation to phylogeographic studies (Zink & Barrowclough 2008; DeSalle *et al*. 2017; Burgos *et al.* 2019) and is considered a powerful tool in evolutionary biology (Moritz 1994; Burgos *et al.* 2019). MtDNA sequence data has the advantage that it represents one single, maternally inherited, nonrecombining locus, that allows relationships to be analysed by the production of genealogical trees, which trace ancestral lineages (Thalmann *et al*. 2004). However, mtDNA analysis is not without its limitations. Thalmann *et al*. (2004) discussed the unreliability of mtDNA due to the existence of numts (nuclear mitochdrial DNA). These translocated copies can pose various problems in analysis, and despite the lack of substantial investigation into numts, it has been implied that numts are prevalent in gorilla studies (Thalmann *et al*. 2004). The implications of mtDNA analysis and numts are discussed in further detail in Chapter 3.

Soto-Calderón *et al.* (2015) highlighted the limited amount of previous studies (of which there were only two: Nsuguba *et al*. (2010) and Simons *et al*. (2013), that investigated genetic variation in captive gorillas, noting that those studies were restricted to nuclear data of unknown or country-wide geographic origin for a proportion of the gorillas involved in the studies. Therefore, research on captive gorillas that compare mitochondrial variability with founder and wild populations is essential for ascertaining the retention of genetic variability in captive gorilla populations (Soto-Calderón *et al*. 2015).

1.10.2. Microsatellites

Microsatellites are regions of short tandem repeats (He *et al*. 2003; Garner *et al*. 2005; Šarhanová *et al.* 2018). Unlike mtDNA markers, microsatellites provide results for both parental lineages and not just the maternal line, thus making them more informative in this manner. There are studies which use a combination of mtDNA and microsatellite analyses for primates, e.g. orangutans (Kanthaswamy *et al*. 2006), eastern Assamese macaques (Sukmak *et al*. 2014) and the Grey Mouse Lemur (Wimmer *et al*. 2002), and results of the two methods occasionally produce contradictory results as is the case for Guizhou snub-nosed monkeys (*Rhinopithecus brelichi*) (Kolleck *et al*. 2013). This combination of genetic markers forms a comprehensive set of data, with the mtDNA markers often used for phylogeographic analysis of evolutionary events that occurred further back in time than those studied

with microsatellites, and the microsatellite markers, which are highly polymorphic can detect small-scale changes in demography due to recent ecological processes (Feulner *et al*. 2004).

Microsatellite markers are widely applied in population genetic studies (Šarhanová *et al.* 2018) and they have been the most abundantly applied molecular markers in conservation breeding programs, as the use of genetic information is crucial to establish such programs to maintain genetic diversity and avoid inbreeding (Roques *et al.* 2019). Microsatellite genotyping is traditionally implemented via fragment length recording and is still applied in current research (Šarhanová *et al.* 2018), however, next-generation sequencing (NGS) approaches which use genome-wide markers such as single nucleotide polymorphisms (SNPs), have become increasingly applied in the last decade and are replacing microsatellite markers as the preferred method (Šarhanová *et al.* 2018; Roques *et al.* 2019). However, every method has advantages and disadvantages, and whilst methods to obtain genetic information data such as SNPs using Restriction-site-Associated DNA sequencing (RADseq), are now the more commonly preferred choice, they are the most resource demanding (Lemopoulos *et al.* 2019) and usually, require high quality DNA (Maigret 2019). Historical samples, such as those housed in NHCs are renowned for containing low quality, degraded and fragemented DNA (Sproul & Maddison 2017) which poses problems for these more technologically advanced methods when historical specimens are the focus of the research.

The acquisition of precise genetic information for population genetics should therefore seek to acquire data from a variety of methods. Studies comparing microsatellites and more recent methods such as SNPs are limited, therefore there is a great interest and requirement for studies which evaluate and compare data obtained via different microsatellite and SNP methods (Roques *et al.* 2019). In addition, to obtain the most comprehensive genetic information which is essential for the conservation and breeding programs of captive populations, using a variety of methods will only serve to aid conservation planning as comprehensive studbooks encompassing a multitude of methods can be established, this will provide conservation and breeding program managers, the tools to make the most informed decisions regarding the management of the species (Roques *et al.* 2019).

1.11. Regional variation of western lowland gorillas

The literature regarding regional variation of gorillas primarily consists of studies that examine differences between the gorilla species and subspecies rather than within the subspecies, i.e. research investigating morphological and/or genetic differentiation between the eastern and western species. There are significantly more studies focusing on morphological differentiation as opposed to genetic differentiation, which reflects morphological research existing before molecular investigations emerged. Substantial morphological and ecological variation has been reported amongst the gorilla

species (Groves 1970; Sarmiento *et al*. 1996; Sarmiento & Oates 2000; Clifford *et al.* 2004), as well as within (Groves 1970), and molecular studies have confirmed genetic variation between and within gorilla species (Garner & Ryder 1996; Saltonstall *et al*. 1998; Clifford *et al*. 2003), but those studies generally focused on the eastern gorillas and/or gorillas species as a whole with investigations into the western lowland subspecies, specifically, being less represented. However, Garner & Ryder (1996) and Clifford *et al*. (2003, 2004) identified variation within the western lowland gorilla subspecies. Clifford *et al*. (2003) concluded from their analysis that there were three distinct clusters based on genetic analysis of the western lowland gorilla. Groups were defined from Nigeria, the Central African Republic and Cameroon. The Cameroon group consisted of most of the gorillas in their dataset and did not identify any further subdivision. Likewise, mtDNA analysis by Garner & Ryder (1996) and Clifford *et al*. (2004) found that distinct haplogroups were identifiable in western lowland gorillas but gene flow was still occurring regardless of geographic boundaries such as major rivers (Clifford *et al*. 2004). Garner & Ryder (1996) reported that despite considerable genetic differences in sequences within western species and high genetic variability the western gorillas used in their study were from captive populations only and thus could not be used in terms of geographical correlation.

Clifford *et al*. (2003) noted that further studies may expose greater variation in surviving populations, and Clifford *et al*. (2004) reported that further genetic research is required for western lowland gorillas of haplogroup C (from Nigeria, Cameroon and Gabon) to establish any further genetic differentiation and population structure. In addition, it is clear from the literature and has been reported in several cases, that the lack of geographical data is of hinderance when examining genetic population structure. This study aims to contribute to the genetic characterisation of western lowland gorillas in a geographical, population genetics, phylogeographic, and geometric morphometrics context.

1.12. The gorillas in this study

The dataset for this study comprises of a historic wild population of western lowland gorillas from the Powell-Cotton Museum (PCM), plus three further museum specimens from the Royal College of Surgeons (RCS). The contemporary population is a captive population based in the Aspinall Foundation (ASP). These two populations are the ones which are the focus of this research. In terms of the PCM collection, despite the collection being used extensively for years with regards to morphological analysis, very little has been performed genetically. A few of the gorilla specimens have had DNA extracted previously, but this is the first time the gorilla collection has been studied to this extent in terms of genetic analysis, geometric morphometrics, and combined with the geographical data. To my

knowledge, the ASP gorillas have not had any genetic investigations performed before, thus, this is the first study to assess the genetic variation and diversity of this contemporary population.

1.13. Summary

The aims of this research are to identify regional variation in historic populations of *Gorilla gorilla gorilla* using geometric morphometrics and molecular techniques (mtDNA and microsatellites), to compare the genetic variation of historic and contemporary populations to ascertain the genetic health of the contemporary population under study, and to provide reliable and informative recommendations for the conservation planners and gorilla keepers and zoo managers to consider for implementation in future breeding programs.

There is a lack of datasets where geographical origins are known for past populations of western lowland gorillas; therefore, this research aims to add to existing limited literature regarding historic populations of this critically endangered gorilla subspecies.

Chapter 2

Regional morphological variation of a historical population of western lowland gorilla

2.1. Introduction

Chapter 1 introduced geometric morphometrics including its emergence and applications, as well as discussing the importance of natural history collections to the field of biological conservation. The importance of natural history collections and specifically the Powell-Cotton Museum (PCM), was reviewed and additionally, biodiversity mapping and geographical information systems (GIS) was investigated in relation to their applications in the field of biodiversity and conservation management. This chapter utilises the gorilla specimens and additional contextual information from the PCM natural history collection and combines it with biodiversity mapping and geometric morphometric methods to investigate regional morphological variation amongst historic populations of western lowland gorillas. From the contextual information held at the PCM, not only can this research utilise the specific geographic coordinates for applying biodiversity mapping methods, but also the additional information in the trip logs, diaries, and field notes. For example, the following two inserts from the gorilla field notes indicate morphological and phenotypic variation observed in the field at the time of collection:

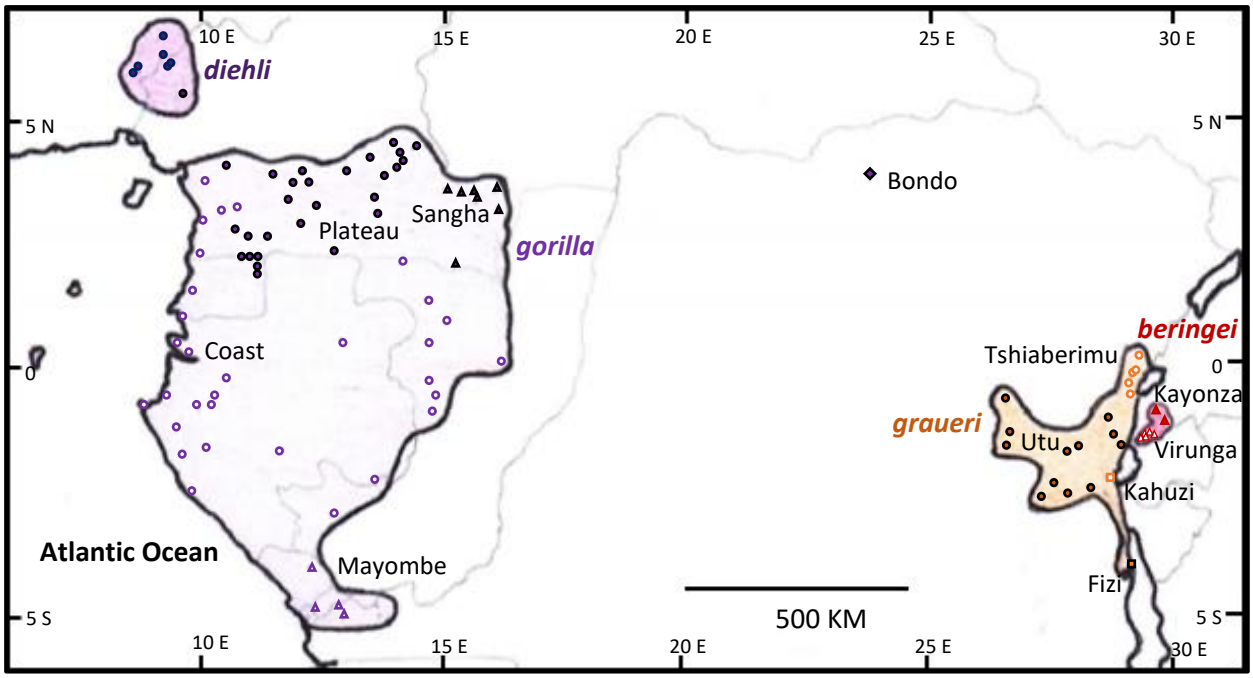
> ***340. Male adult*** *Very old beast, 2 upper incisors and 1 lower canine missing. All teeth very much worn, canine teeth in upper jaw worn level with other teeth. All Gorilla coming from the forest NW and W. from here (Arteck 3¾ N. 14¼ E.) seem to have smaller crests in comparison with those coming from the E & S districts.*
>
> ***Mer 138 and 139***
>
> ***138. Female adult*** *Skinned in bush, skin badly speared. Long narrow faced, no sign of red crest, beast all black.*
>
> ***139. Female adult*** *Skinned in bush, plump and round faced, red crest and hair tipped with grey. These two beasts 138 and 139, show the difference in shape of face, as so well shown in Carl Akerley's "Brightest Africa" page 222,* (Akeley 1923).

As reviewed in Chapter 1, geometric morphometrics has seen a rapid growth since the turn of the century and supersedes traditional morphometric approaches (Tofilski 2008). Studies in the

field of primatology have also employed geometric morphometric methods for morphological studies. Some of the more recent publications in this area include Rocatti *et al*. (2017), Fiorenza & Bruner (2018), Ito & Koyabu (2018), Püschel *et al*. (2018), Simons *et al*. (2018), Aristide *et al*. (2019) and Nishimura *et al*. (2019). Geometric morphometric studies, specifically involving the great apes, address a variety of scientific interests such as evolution, locomotion, and diet. Recent publications including great apes (non-human) and geometric morphometric analysis include: Knigge *et al*. (2015), Pearman & Jabbour (2015), Martinez-Maza *et al*. (2016) and Fatica *et al*. (2019). As mentioned in the previous chapter, the PCM has contributed to morphological research for many decades. For example, Terhune *et al*. (2007) analysed the temporal bone of gorillas and chimpanzees for variation; Gilbert (2011) sampled over 800 skulls (the basicranium) of extant African papionins (a tribe of several large old-world monkey species) from ten different natural history collections including the PCM; Coolidge (1929) and Groves (1967, 1970) incorporated the PCM gorilla specimens in their research, both of which provided the foundations for gorilla systematics which practically all contemporary studies of gorillas focus upon (Leigh *et al*. 2003).

Previous studies on gorillas using traditional morphological analysis were focused on taxonomy and systematics. Groves (1967, 1970) identified four 'demes' within the western species (Fig. 2.1.a) which were termed Nigerian, Coastal, Sangha and Plateau. The Nigerian cluster was observed to be significantly morphologically distinct from the other clusters, and in 1904 was classified as a subspecies of the western gorilla, *G. g. diehli* (Sarmiento & Oates 2000). The remaining western gorillas, Groves (1967, 1970) concluded, were one subspecies showing considerable morphological overlap but sufficient variation to separate them (based on skull morphology) into three demes. Figure 2.1.b demonstrates the distribution of the demes as identified by Groves (1970). Interestingly, western lowland gorillas from the Republic of the Congo (previously the French Congo) were classified in the coastal deme which has the largest deme distribution, although Groves (1970) noted that gorillas from the Mambili region (number 6 on Fig. 2.1.b) were closely associated with gorillas from the coastal groups (numbers 2-5 on Fig. 2.1.b) thus grouped together. The plateau deme lies in the central/North-east of Cameroon and the Sangha deme resides in the East of Cameroon.

(a)



(b)

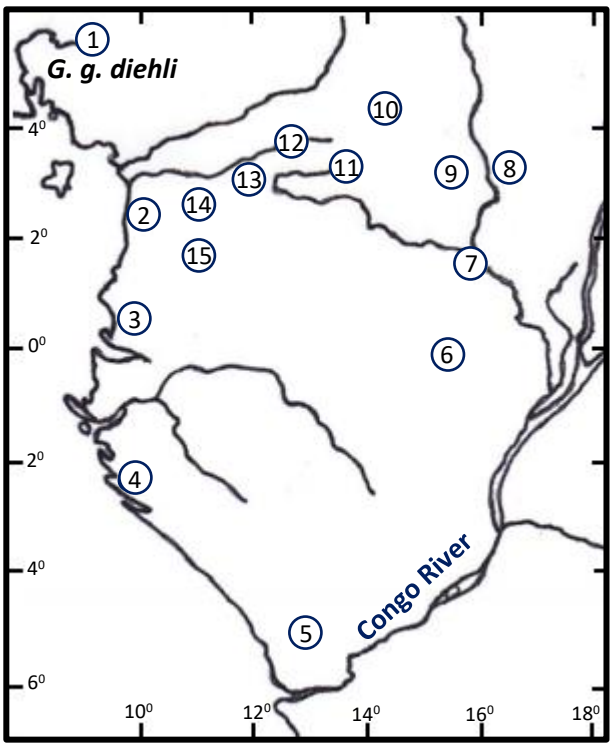


Figure 2.1 Map of central Africa showing (a) gorilla subspecies, localities and demes based on results from Albrecht *et al*. (2003, page 68) and (b) the geographical locations of the 15 groups of western gorillas studied by Groves (1970, page 289). Groups 16-19 belong to the eastern species and are not shown in this figure.

Using the PCM collection, the regional morphological variation of western lowland gorillas was studied to compare it with previous assessments using traditional morphological tools, and to further explore the division of western lowland gorillas into demes.

2.2. Aims

The aims of this study were:

- To investigate regional morphological variation of a historic wild population of the western lowland gorilla subspecies, by analysing skulls and mandibles of specimens held in the natural history collection at the Powell-Cotton Museum (PCM).
- To compare results found here that use geometric morphometric methods to previous investigations that used traditional morphometrics.
- To draw conclusions in support or objection to the existence of regional demes among western lowland gorillas and thus, add to existing literature.

2.3. Hypotheses and predictions

Previous literature using traditional morphometrics has indicated regional morphological variation and classified the western lowland gorilla population into three demes (the fourth deme being the Cross River gorilla), therefore the hypotheses and predictions for this study are:

- Regional morphological variation in size and shape will be observed between groups A and B which reflect the plateau and coastal demes defined by Groves (1970).
- No significant variation in size and shape will be found between groups B and C as they are both classed as belonging to the coastal deme as defined by Groves (1970).
- Regional morphologic variation will be more significant in the males than females, as females have been shown to be more homogenous morphologically in previous studies (Albrecht *et al.* 2003).

2.4. Methods

2.4.1. Historical western lowland gorilla specimens

The skulls and mandibles used in this research are from the western lowland gorilla PCM collection. A total of 138 skulls (66 males and 72 females, ventral position), and 130 mandibles (67 males and 63 females, lateral view) were included in this study. Although the PCM has considerably more western lowland gorilla specimens than used in this research, some of the skulls and mandibles were unable to be included in the sample set because they were either too badly damaged or were juveniles. Juveniles were excluded from analyses because they are not fully developed individuals, potentially biasing the morphological analysis, particularly in terms of size where a downward bias would be observed (Coolidge 1929; Haddow & Ross 1951). The juvenile and adolescent stage of gorilla development ranges from 3-8 years of age (Palagi *et al.* 2007; Pafčo *et al.* 2019), therefore, all specimens aged 8 years and younger were removed from analyses. This was achievable due to the meticulous record keeping where each specimen card had an (approximate) age recorded or juvenile was stated.

Extreme sexual size dimorphism is known to exist amongst adult gorillas (Stumpf *et al*. 2002; Albrecht *et al*. 2003). Younger (juvenile and adolescent) gorillas do not demonstrate sexual dimorphism, only in the adult life stage does sexual dimorphism become apparent (Berge & Penin 2004) therefore, previous literature supports the methods applied here to remove juveniles entirely from analyses and divide the remaining adult specimens into independent male and female datasets. Thus, the adult skulls and mandibles were divided into four separate datasets: male skulls and male mandibles, and female skulls and female mandibles. Figure 2.2 demonstrates the observable sexual dimorphism shown between male and female gorilla skull morphology. Specimens FC.123 and M.283 (Fig. 2.2.a and b respectively) are males, notably larger than the female specimens FC.154 and M.58 (Fig. 2.2.c and d respectively) with the sagittal crest significantly prominent in the males.

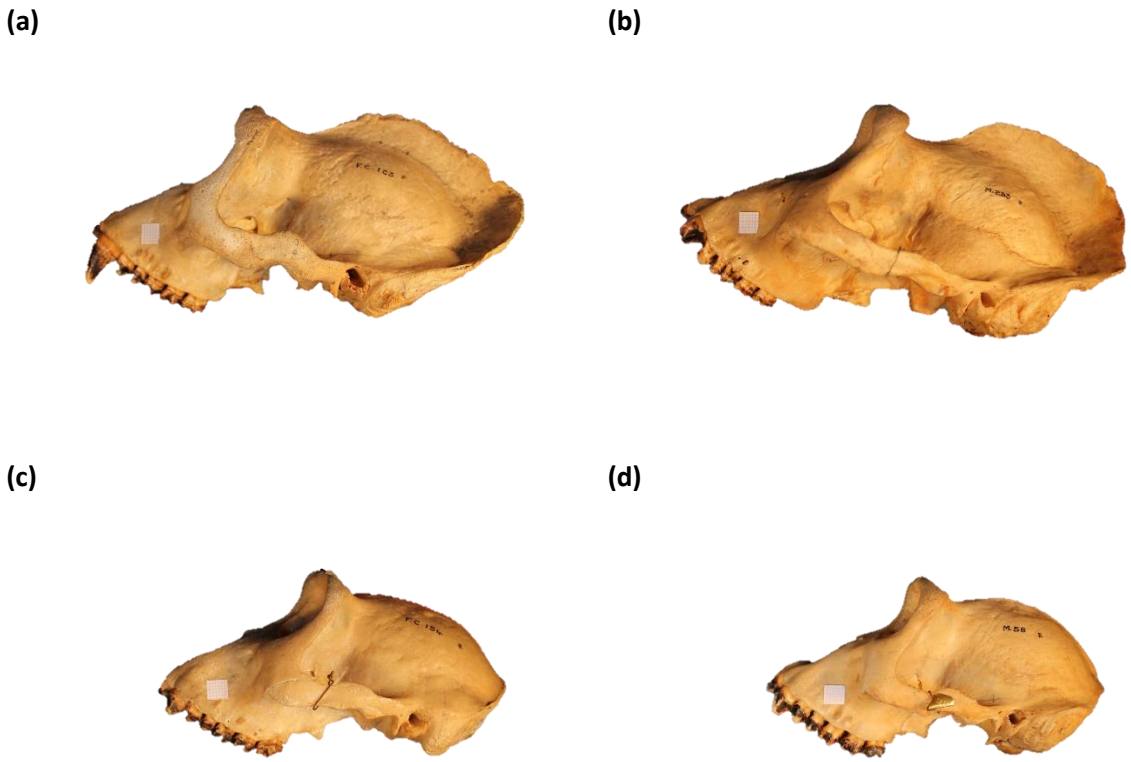


Figure 2.2 Pictures demonstating the observable sexual dimorphism between (a, b) males and (c, d) females. Notable differences can be observed between size, with males being considerably larger, and the prominence of the sagital crest in the males.

2.4.2. Geographical location of historical specimens

The biodiversity mapping results from Chapter 1 (Fig. 1.5) were used to visualise the geographic distribution of the samples used in this chapter. The data was separated for analyses into three regional clusters, as well as male and female skulls and mandibles (Fig. 2.3.a, b). The reason for this initial clustering, as described previously, was based on several factors. Firstly, the visualisation of the exact location of the specimens in this dataset provided an observable clustering incorporating extremes of the geographic ranges. Secondly, to provide consistency throughout this research including the genetic analyses, groups needed to remain the same. Thirdly, despite the Republic of Congo specimens falling into the Coastal deme as classified by Groves (1970), there were differences observed. Due to these specimens in this dataset being more geographically distinct from the other specimens, and Groves (1970) noting dissimilarities, albeit marginal, the decision was made to treat them as a separate group in this research to allow for comparison. This clustering mainly followed the demes as distinguished by Groves (1970), however, some distinctions were made: this dataset did not contain many individuals from the Sangha deme, so they were incorporated into the neighbouring

cluster (group A); the Republic of Congo specimens were treated as a separate cluster, and a few individuals of each sex were designated into group B, although they would have been placed in group A if strictly following the demes classification.

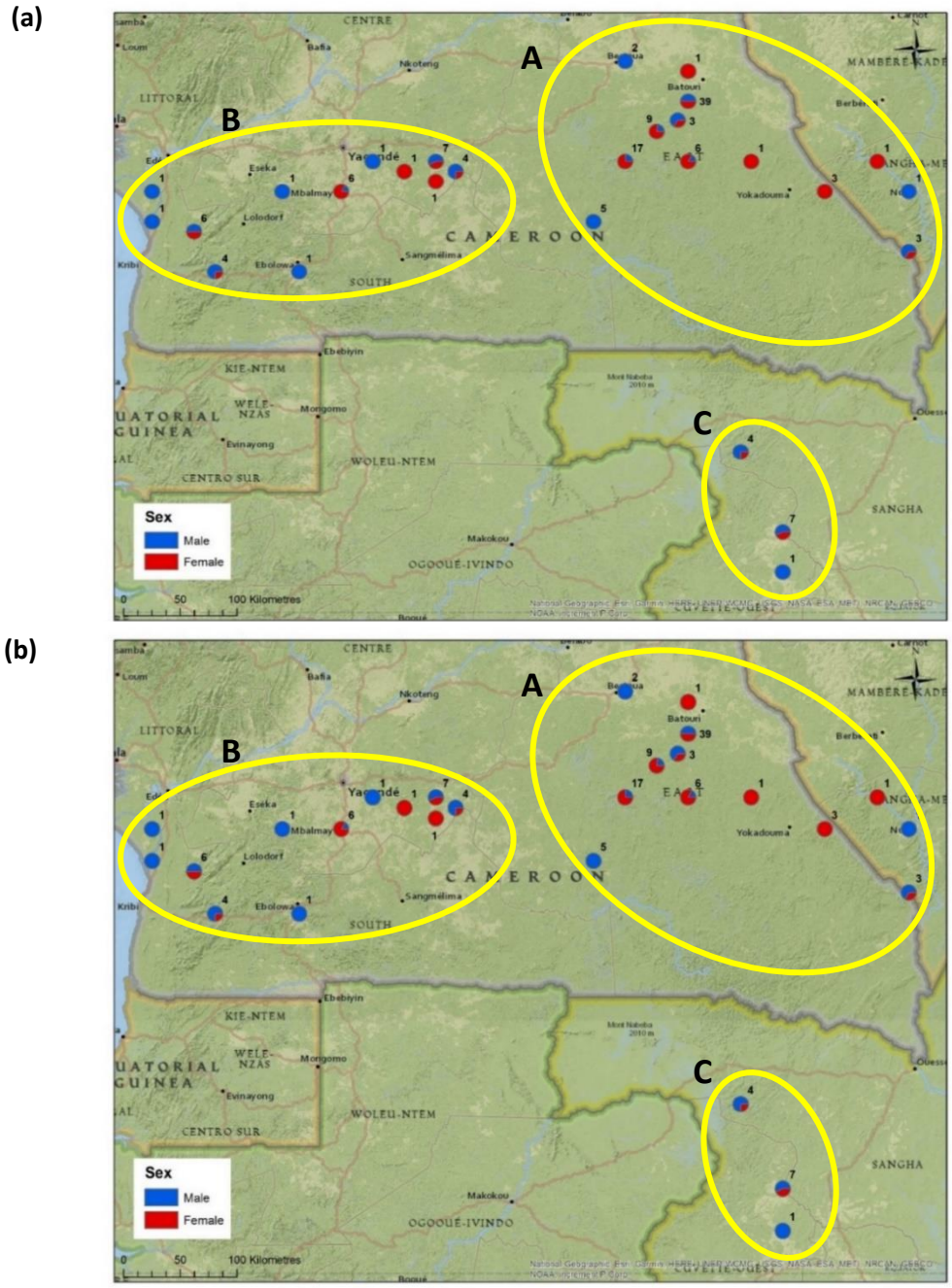


Figure 2.3 Geographical locations of the PCM gorilla specimens for (a) skulls and (b) mandibles (western lowland gorillas). The regional groups are defined by the yellow borders and represented by capital letters (A, B and C). The number next to each location represents the total number of gorillas at that location, the red and blue division relate to the number of gorillas of each sex at the given location.

2.4.3. Digitisation of samples

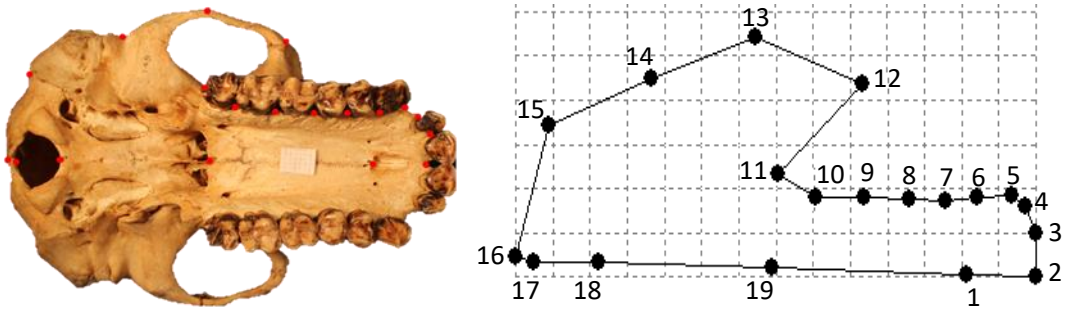
Photographic images of gorilla skulls and mandibles were taken using a Canon EOS 1100D mounted on a copy stand set at a distance of 48.26 cm (19 inches) for all samples, with camera settings: ISO Auto, aperture setting F32 and shutter speed 0”8. One of the largest male specimens was selected to set the distance of the camera to ensure that subsequent specimens did not exceed the frame of the photograph. To support the specimens and hold them in position, bean bags were used. Due to the apparent variation in size and shape of the specimens, it was necessary on occasion to slightly alter the position of the supporting bean bags. To ensure specimens maintained the correct orientation, a spirit level was placed on each specimen for every image taken. In additon, a 1 cm^2 piece of graph paper was positioned centrally on each specimen to allow for scaling during analyses.

Morphological analyses were executed using the freely available “TPS-Series” software developed by Rohlf (2015) which has been used extensively for geometric morphometric analyses. Prior to analyses, all photographs were renamed with their individual specimen identification as per the PCM catalogue.

Prior to the application of landmarks, the samples were randomly ordered in tpsUtil version 1.65 to avoid any landmark placing bias. A total of 17 landmarks were digitised on the lateral side of the left mandibles and 19 landmarks were digitised on the right side of the ventral view of the skulls. Landmark digitisation was performed in tpsDig2 version 2.22. Figure 2.4.a and b show the placement of landmarks and provides a morphological definition of landmark sites. Asymmetry was not the focus of this study, hence only the right side of the skulls received landmark placement; this is common practice and reduces data redundancy (Webster & Sheets 2010). Additionally, and again to avoid bias, landmarks were placed over a duration of time (weeks) to ensure that placement was consistent and repeatable on separate occasions. The landmarks were selected to provide the most inclusive scope of the morphology of the samples and to ensure that all landmarks could be placed in all specimens.

To further ensure consistent positioning of samples and placement of landmarks (i.e. landmark positioning error), a subset of five specimens from each of the four datasets was digitised separately and repeatedly. This provided a dataset of 25 pictures for each dataset to test for accuracy. A tps file was built from the landmark data using tpsUtil for each of the four datasets and a scale set within each dataset.

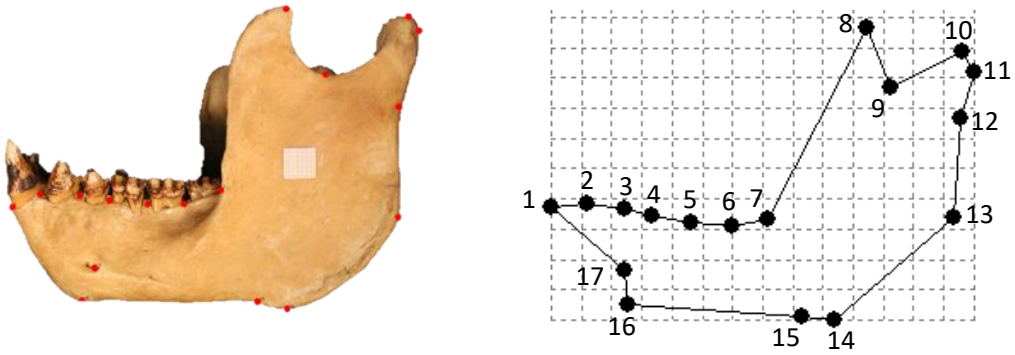
(a)



Landmark	Definition:
1	Incisive foramen: most posterior point
2	Prosthion: between central incisors
3	Prosthion: between central and lateral incisors
4	Incisor alveolus: most posterior point
5	Canine alveolus: most anterior point
6	Contact point between canine alveolus and first premolar
7-10	Contact points between adjacent pre-molars and molars
11	Posterior midpoint of 3rd molar alveolus

Landmark	Definition:
12	Most anterior point of zygomatic arch
13	Most lateral point of zygomatic arch
14	Meeting point of zygomatic arch and temporal bone
15	Meeting point of temporal bone and occipital bone
16	Most posterior point of cranium
17	Opisthion: most posterior point of the foramen magnum
18	Basion: most anterior point of the foramen magnum
19	Tip of posterior nasal spine

(b)



Landmark:	Definition:
1	Infradentale
2	Contact point between incisor and canine
3	Contact point between canine and 1st premolar
4-6	Contact points between pre-molars and molars
7	Alveolus on distal aspect of molars
8	Tip of coronoid process
9	Deepest point of mandibular notch
10	Tip of condylar surface

Landmark:	Definition:
11	Most lateral point of condylar surface
12	Posterior border of the condylar articular surface
13	Gonion
14	Inferior border of the gonial region
15	Antegonial notch
16	Menton
17	Mental foramen

Figure 2.4 Placement and numbering of landmarks used for both sexes for (a) skulls and (b) mandibles of western lowland gorillas, and morphological definitions adapted from Cardini *et al*. (2007), Rommel *et al*. (2009), Marcus *et al*. (2000) and Holton *et al*. (2015).

2.4.4. Procrustes superimposition

Partial Procrustes superimposition, generally referred to as Procrustes superimposition, is the most commonly used superimposition method in geometric morphometrics (Webster & Sheets 2010). Essentially, Procrustes superimposition is the removal of differences in location and is accomplished by centring configurations. This is achieved by the centroid of each configuration being calculated, the centroid is then the origin of a new coordinate system and the configurations are rescaled so that a common centroid is shared, thus removing the differences in size (Zelditch *et al*. 2004; Webster & Sheets 2010). The differences in orientation are removed by the rotation of one configuration around its centroid, relative to another configuration, until minimum offset in landmark location is achieved (Zelditch *et al*. 2004; Webster & Sheets 2010). This method of translation, rescaling and rotation is referred to as Generalised Procrustes Analysis (GPA), once this procedure has been performed, all the differences in landmark location, scale and orientation have been removed, which results in any observed remaining differences in landmark data being the result of shape differences between the configurations (Zelditch *et al*. 2004; Webster & Sheets 2010). Procrustes superimposition, or more simply, the alignment of all configurations, were performed in tpsRelw version 1.70 for each dataset.

Procrustes superimposition aligns the configurations in a non-Euclidean shape space, but to enable statistical analyses, Euclidean shape space is required (Zelditch *et al*. 2004). The Procrustes shape distance is a measure of the difference in shape between two landmark configurations and is given in Kendall's (non-Euclidean) shape space; therefore, Procrustes distances must be approximated to Euclidean space of the same dimension, known as a tangent space (Stegmann & Gomez 2002; Zelditch *et al*. 2004; Mitteroecker *et al*. 2013). Approximating Euclidean shape space has been shown to be acceptable for most biological datasets (Marcus *et al*. 2000; Mitteroecker *et al*. 2013). Euclidean shape space was approximated using tpsSmall version 1.34 with an orthogonal alignment projection method to validate whether the orthogonal tangent space was appropriate for the data. All four datasets were tested to confirm that the shape variation amongst the specimens was of acceptable discrete distance to allow an approximation of Euclidean tangent space.

2.4.5. Western lowland gorilla skull and mandible centroid size analyses

The data for the size of the skulls and mandibles was generated by calculating the centroid size for each specimen. The centroid of a configuration literally refers to the centre of the object (Webster & Sheets 2010). Centroid size is considered mathematically independent of shape as it is orthogonal to shape (Zelditch *et al*. 2004). Centroid size is calculated as the square root of the sum of the squared distances between each landmark and the centroid of the form (Bookstein 1991; Zelditch *et al*. 2004; Webster & Sheets 2010). Centroid size for each skull and mandible was calculated using tpsRelw and transformed with natural logarithms. To test for landmark positioning error in terms of size, an analysis of variance (ANOVA) was carried out in SPSS version 23.0 (IBM Corp., Released 2015, Armonk, NY) on the transformed centroid size measurement (LnCS) of the 25 test photographs for each dataset. The mean, median, standard deviation and standard error were calculated for the skulls and mandibles (sex separated), and by regional groups as previously defined (A, B and C). An ANOVA by region on LnCS was performed in SPSS for each dataset to test for significant differences among the regions. Tukey-Kramer post-hoc tests were also performed on LnCS as the test takes into account unequal group size (Ashcroft & Pereira 2002; Field 2013; Fowler *et al*. 2013). In addition, a box plot of CS was produced for visualisation purposes.

2.4.6. Western lowland gorilla skull and mandible shape analyses

The shape data for each dataset is represented as relative warps and partial warps scores and plots and was calculated in the TPS-series software. Eigenanalysis of the bending energy matrix produces the partial warps shape data (Zelditch *et al*. 2004). Relative warps can be obtained from Procrustes residuals or partial warps, and they are principal components of a distribution of shapes in a tangent space (Pavlinov 2001). To test for landmark positioning error in terms of shape, a multivariate analysis of variance (MANOVA) was carried out in PAST version 3.14 (Hammer *et al*. 2001) on shape variables (relative warps) obtained from the 25 test photographs for each dataset, and relative warps plots were produced in tpsRelw to visualise the accuracy of landmark placement. MANOVA uses more than one dependent variable, it tests whether two or more groups have the same multivariate mean, and it is viewed as an extension of an ANOVA (Qeadan 2015). The MANOVA requires additional assumptions to be met in comparison to the ANOVA. The absence of multivariate outliers and the equality of covariance matrices are two of the additional assumptions which require confirmation. The absence of multivariate outliers is confirmed by assessing Mahalanobis distances, and the equality of covariance matrices is assessed via a Box's *M* test (O'Brien & Kaiser 1985; Garson 2012).

Other multivariate methods frequently applied in morphometric analyses are principal components analysis (PCA), canonical correlation (CCA), the related canonical variates analysis (CVA) and discriminant functions analysis (DFA) (Kapoor & Khanna 2004). All these methods are related in that they use a linear combination of the number of variables multiplied by their respective coefficients, which maximises intragroup variance (Kapoor & Khanna 2004). The two ordination methods (also called gradient analyses) used in this chapter were PCA and DFA. Both methods are exploratory ordination methods that endeavour to order objects (here skulls and mandibles) based on the variables measured in the objects (Paliy & Shankar 2016). Exploratory methods are useful because they provide a visualisation of object similarities and dissimilarities which aids interpretation of the results (Paliy & Shankar 2016). PCA simplifies descriptions of variation among individuals (Zelditch *et al*. 2004); it uses Euclidean distance to measure dissimilarity among objects (Zelditch *et al*. 2004; Paliy & Shankar 2016). The purpose of PCA is to replace original variables with new composite variables (principal components, PCs) that are linear combinations of the original variables; additionally, they are independent from one another.

PCA of the partial warps was performed on all four datasets in SPSS. In the first instance, the Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy was performed in addition to the Bartlett's test of sphericity on each of the datasets; these tests indicate whether the data is suitable for PCA. A total of 10 specimens consisting of six females (CamI.149, CamI.150, CamI.139, CamI.97, CamI.109 and CamI.98) and four males (MI.28, ZVI.32, CamI.107 and CamI.134) were used in this analysis which were classified as belonging to the regional group B, whereas their deme classification would have placed them in the plateau deme or group A (this research). These specimens were classified as a 'redefined' group to visualise their placement and observe any group clustering. Scree plots were produced and PCA plots using the first two principal components were generated.

DFA, also referred to as linear discriminant analysis (LDA) or canonical discriminant analysis (CDA), consists of ordination techniques that find linear combinations that maximise the groupings of objects placing them into separate classes based on the observed variables (Paliy & Shankar 2016). DFA can be used to assign specimens, including unknown specimens to groups (Zelditch *et al*. 2004). DFA uses an eigenvector-based solution as does PCA, however, unlike PCA, DFA explicitly maximises the between-class group dispersion (Paliy & Shankar 2016). DFA was performed on the partial warps in SPSS for each dataset to ascertain differentiation among groups and predict group membership, and plots were produced using the first two functions for visualisation. Additionally, a cross-validation method was performed in SPSS; the cross-validation method repeatedly treats $n - 1$ of n samples as the verifying dataset and uses this to establish the discriminant rule which is then applied to the one observation left out and classifies it.

The relative warps data from the four datasets was subjected to MANOVA analysis using PAST, accompanied by Wilks' lambda (λ) test statistic, followed by pairwise comparisons using Hotelling's T^2 tests. The Bonferroni correction was applied to the pairwise *p*-values. The Wilks' lambda (λ) test statistic is the multivariate equivalent of the *F*-test statistic in a one-way ANOVA and it tests for differences between groups (O'Brien & Kaiser 1985; Garson 2012). The MANOVA also reports a *p*-value and an *F*-statistic which are related to the degrees of freedom reported. If the test produces a significant *p*-value then further pairwise and post-hoc testing can be performed (O'Brien & Kaiser 1985; Garson 2012). The absence of multivariate outliers was tested by performing a multiple linear regression in SPSS. To identify outliers, the Mahalanobis distances obtained from the linear regressions were investigated. Box's *M* tests were performed in SPSS to assess the equality of covariance matrices. The Box's *M* test is considered to be very stringent (Ashcroft & Pereira 2002; Field 2013; Fowler *et al*. 2013) thus a *p*-value above 0.001 is deemed appropriate to meet the assumption. Levene's tests of equality of variances were performed in SPSS for each shape variable among groups; a significant result from the Levene's test indicates the assumption of homogeneity of variances has been violated and non-parametric tests (for example, PERMANOVA) should be performed.

In addition, the relative warps data was used to produce deformation grids using the tpsRelw. The deformation grids enable visualisation of the shape variation between samples. An average configuration was produced for each dataset and deformations grids of variations from the average were produced for a few specimens in each dataset.

2.5. Results

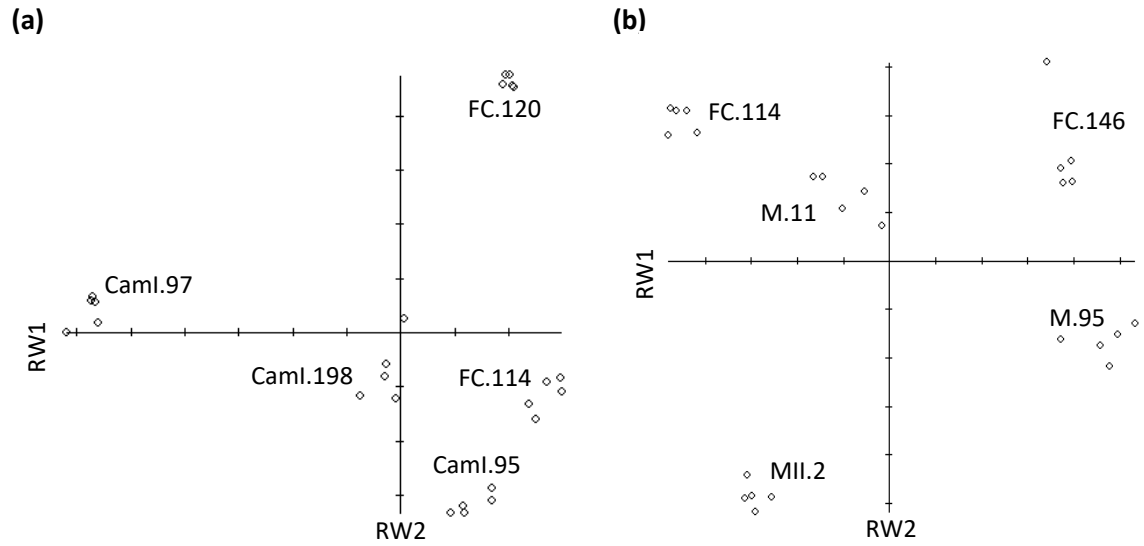
2.5.1. Tests for accuracy of landmark placement

For landmark placement repeatability, the ANOVA on LnCS (Table 2.1) showed that there were no significant differences amongst the five repeated photographs. This confirmed that the positioning of specimens for photographing and landmark placement on mandibles and skulls was accurate, and any significant results obtained in further analyses were not due to errors of specimen or landmark placement.

Table 2.1 Analysis of variance (ANOVA) results for landmark placement using centroid size (transformed with natural logarithm) of western lowland gorillas

		Sum of Squares	DF	Mean Square	F	Sig.
Female Skull	Between groups	0.000	4	0.000	0.03	1.000
	Within groups	0.105	20	0.005		
	Total	0.105	24			
Female Mandible	Between groups	0.000	4	0.000	0.001	1.000
	Within groups	0.291	20	0.015		
	Total	0.291	24			
Male Skull	Between groups	0.000	4	0.000	0.047	0.996
	Within groups	0.046	20	0.002		
	Total	0.047	24			
Male Mandible	Between groups	0.000	4	0.000	0.006	1.000
	Within groups	0.043	20	0.002		
	Total	0.043	24			

The relative warps plots complemented the ANOVA results for centroid size regarding placement of specimens and landmarks. Although there were some observable differences amongst the repeated photographs and landmarks, they were not significant. The relative warps plots for male and female skull and mandible tests are shown in Fig. 2.5.a-d. The MANOVA showed non-significant differences for all of the four datasets: female skulls ($F = 0.5531$; $p > 0.05$; Wilk's $\lambda = 0.5391$), female mandibles ($F = 1.334$; $p > 0.05$; Wilk's $\lambda = 0.264$), male skulls ($F = 0.6919$; $p > 0.05$; Wilk's $\lambda = 0.4693$), male mandibles ($F = 0.5911$; $p > 0.05$; Wilk's $\lambda = 0.5187$).



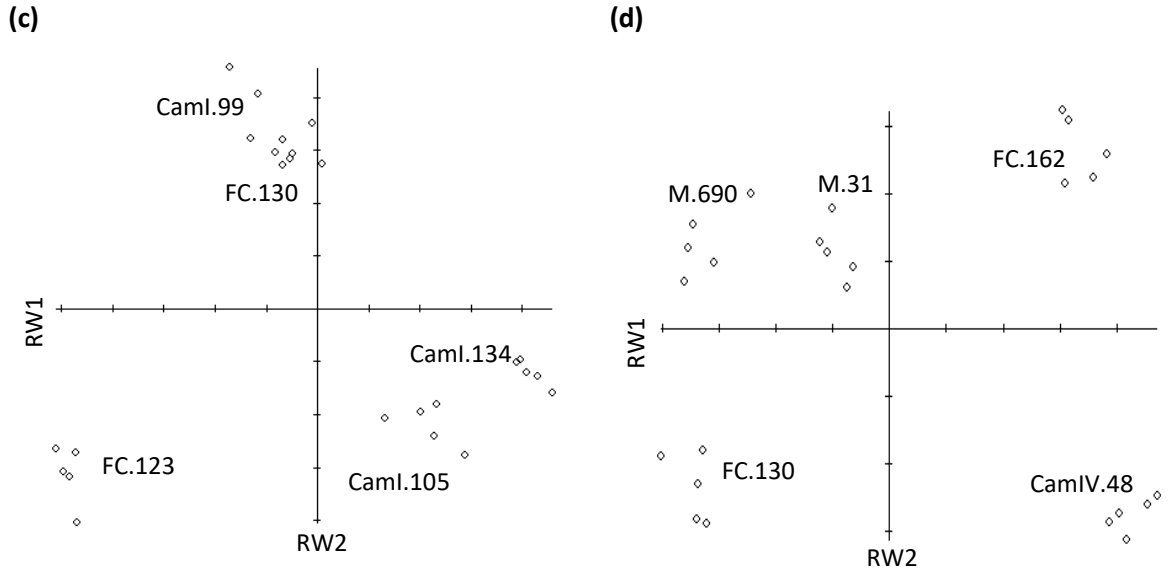


Figure 2.5 Relative warps plots showing landmark placement for tests of (a) female skulls, (b) female mandibles, (c) male skulls and (d) male mandibles of western lowland gorillas. Although there is some variation among landmark placement, the multivariate analysis of variance (MANOVA) showed non-significant differences, thus confirming accurate placement of landmarks in regards to shape analysis.

2.5.2. Preliminary analyses of data

To test the suitability of the data prior to analyses, initial tests are required. Figure 2.6 shows the tpsSmall output for the female skull (the male skull and both mandible datasets showed similar outputs but are not shown), showing a positive correlation (R = 0.9999) between the Procrustes distances and the orthogonal tangent space (R = 0.9999 was observed in all four datasets), thus tangent space approximation was an appropriate method for the data and further statistical analyses was permitted.

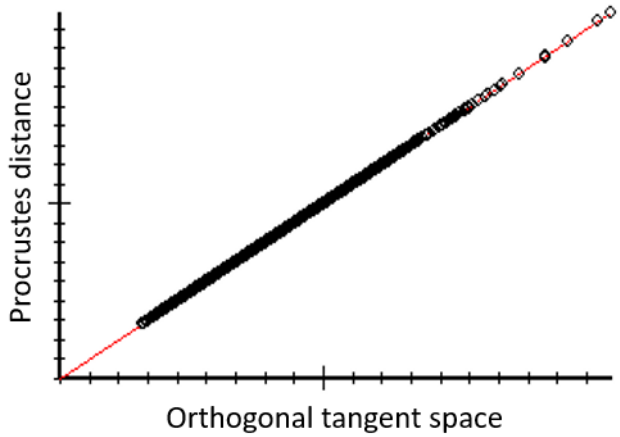


Figure 2.6 The tpsSmall output for the female skull of western lowland gorillas confirming a positive correlation (*R* = 0.9999) between Procrustes distancw and orthogonal tangent space.

To proceed with PCA analyses, the Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy was performed on all four datasets, in addition to the Bartlett's test of sphericity. Table 2.2. summarises the results. Generally, a KMO value of ≥0.500 indicates the data is suitable for PCA analysis. The male skull dataset was slightly lower than this recommendation with a value of 0.487. However, in all four datasets the Bartlett's significance value was <0.001, indicating that the data is suitable to perform PCA.

Table 2.2 Summary of results for sampling adequacy and tests of sphericity for four datasets of western lowland gorillas

	Female Skull	Female Mandible	Male Skull	Male Mandible
KMO	0.550	0.500	0.487	0.559
Bartlett's significance value	<0.001	<0.001	<0.001	<0.001

Scree plots were produced for each of the four datasets (Fig. 2.7.a-d). The results showed that the first three to five components based on eigenvalues retained the most variation, thus, five components were selected for the PCAs.

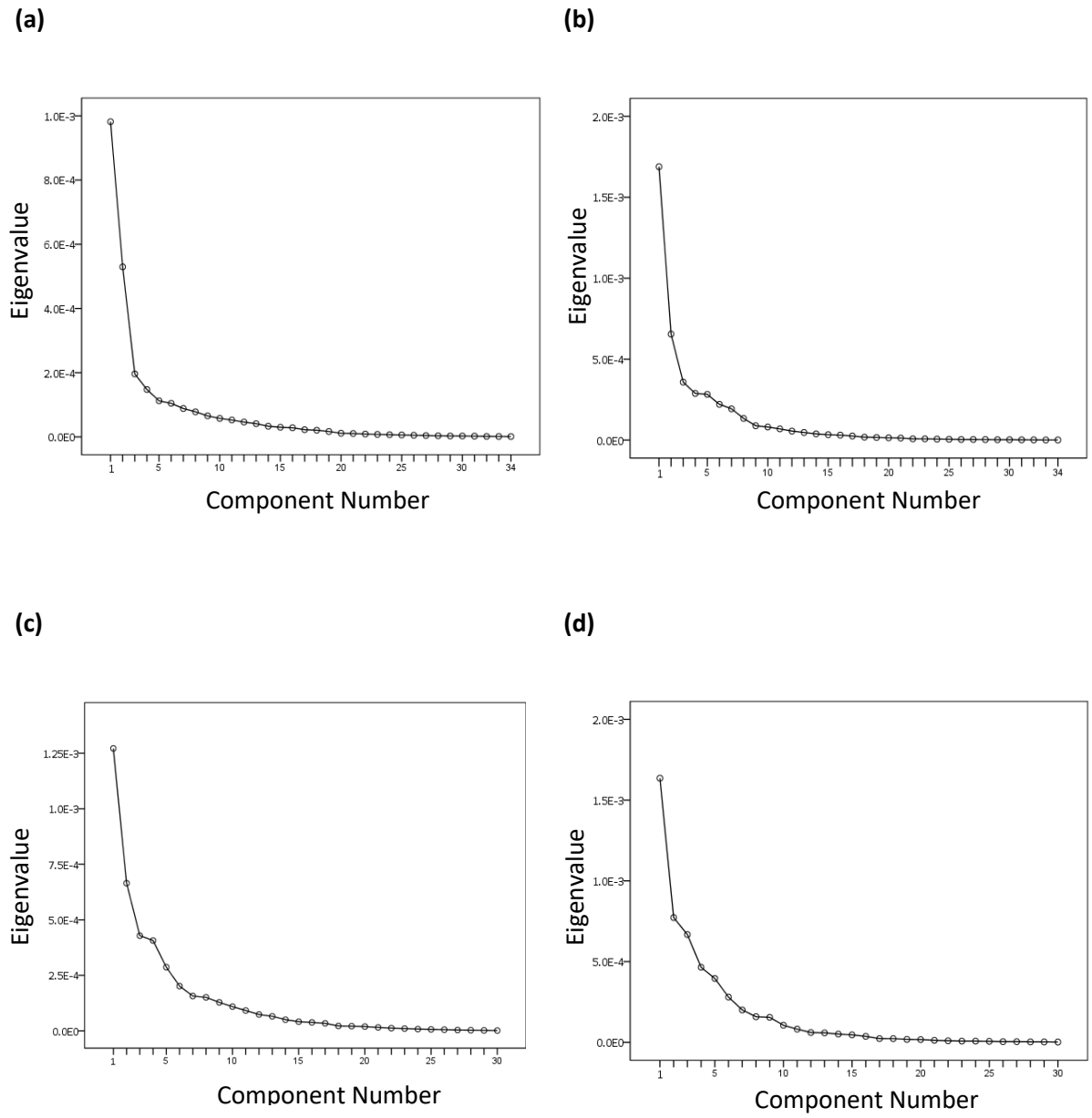


Figure 2.7 Scree plots based on eigenvalues for the four datasets of Western Lowland Gorillas, (a) female skull, (b) male skull, (c) female mandible and (d) male mandible, with the first five components showing >70% of the total variation in all cases.

The total variance of components for each of the four datasets is summarised in Table 2.3. The results for the first five raw components only are shown because they account for over 70% of the variation in each dataset.

Table 2.3 Total variance of principal components based on eigenvalues for all four datasets of western lowland gorillas

	Initial Eigenvalues			Initial Eigenvalues	
Raw component	% of variance	Cumulative %	Raw component	% of variance	Cumulative %
Female Skull			**Female Mandible**		
1	35.954	35.954	1	29.356	29.356
2	19.395	55.349	2	15.349	44.706
3	7.188	62.538	3	9.892	54.598
4	5.406	67.944	4	9.392	63.990
5	4.114	72.057	5	6.625	70.615
Male Skull			**Male Mandible**		
1	38.382	38.382	1	30.781	30.781
2	14.878	53.261	2	14.541	45.323
3	8.147	61.408	3	12.564	57.886
4	6.552	67.960	4	8.748	66.634
5	6.423	74.383	5	7.450	74.084

For visualisation purposes the first two principal components (which also show the greatest percentage of variation) were used to display the results.

2.5.3. Regional variation in skull and mandible size of historical western lowland gorillas

The mean centroid size for females was smaller than males in both the skull and mandible analyses; this result was expected as the subspecies shows extreme sexual size dimorphism and thus, the reason why the sexes were analysed separately. In terms of regional variation, the results revealed similar centroid size across the three regions, but the skulls and mandibles in both males and females from region C (Congo) were smaller in all cases (Fig. 2.8).

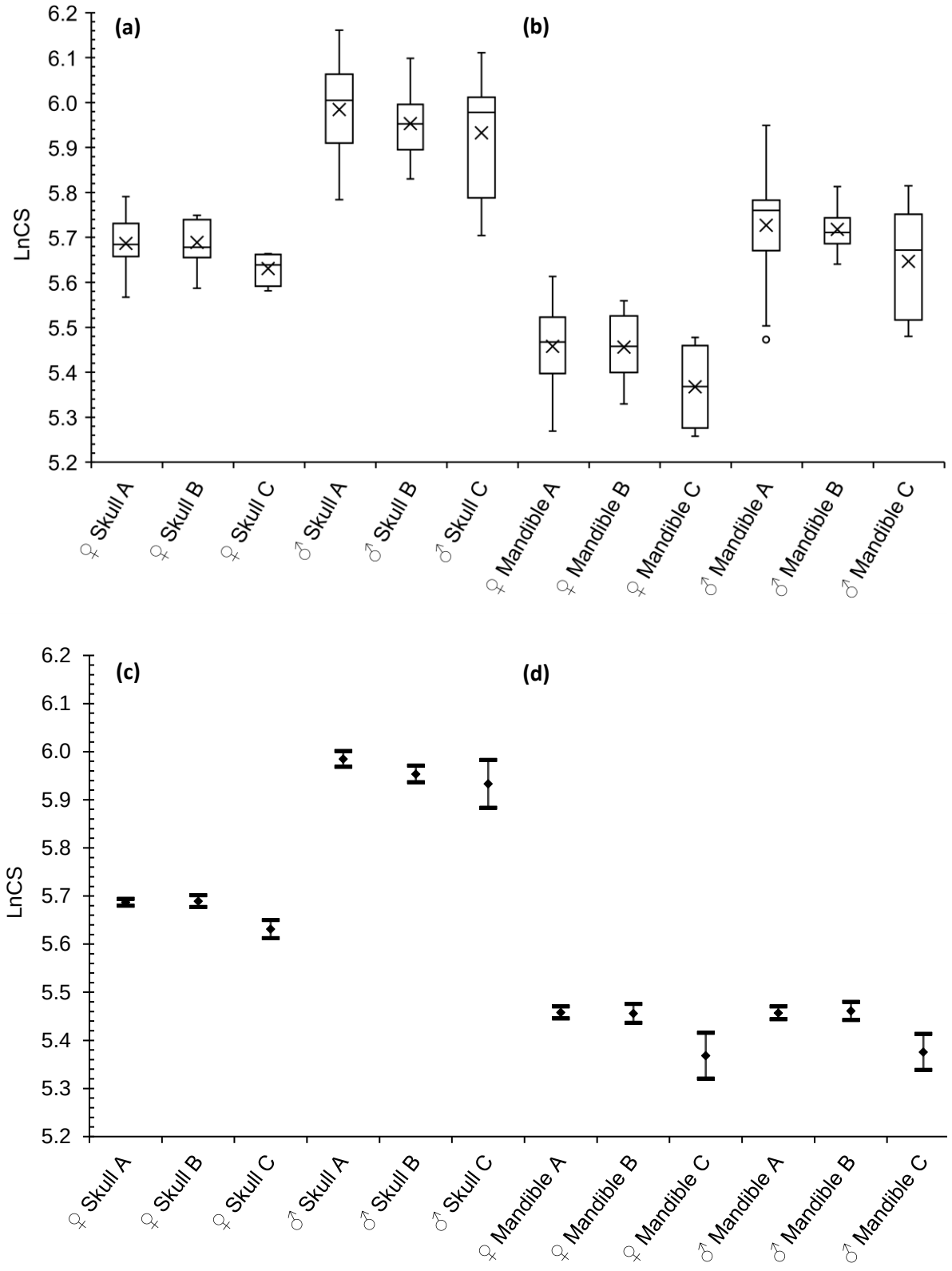


Figure 2.8 Box plot for visualisation purposes of centroid sizes (transformed by natural logarithm) by region for (a) skulls and (b) mandibles of western lowland gorillas and standard error estimates for (c) skulls and (d) mandibles.

Statistical analysis by means of ANOVA by region on LnCS revealed that the observed differences in size were not significant in any of the four datasets (Table 2.4).

Table 2.4 Analysis of variance (ANOVA) results for centroid size (transformed by natural logarithm) of western lowland gorillas

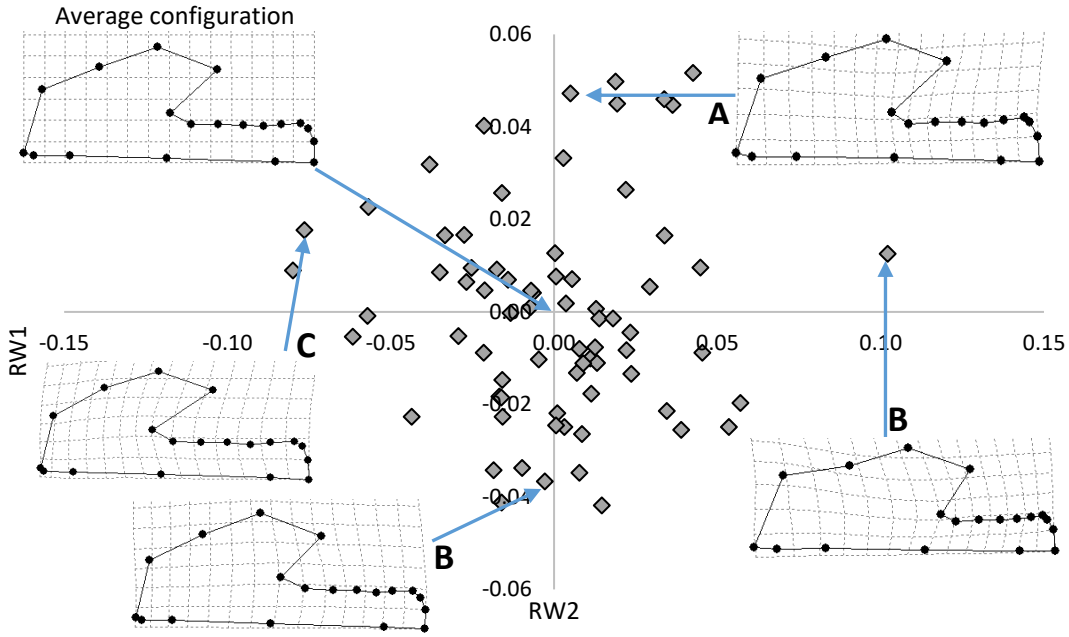
		Sum of Squares	*DF*	*Mean Square*	*F*	*Sig.*
Female Skull	Between groups	0.012	2	0.006	2.341	0.104
	Within groups	0.178	69	0.003		
	Total	0.190	71			
Female Mandible	Between groups	0.030	2	0.015	2.205	0.119
	Within groups	0.408	60	0.007		
	Total	0.439	62			
Male Skull	Between groups	0.025	2	0.012	1.239	0.297
	Within groups	0.626	63	0.010		
	Total	0.651	65			
Male Mandible	Between groups	0.048	2	0.024	2.738	0.072
	Within groups	0.559	64	0.009		
	Total	0.607	66			

The Tukey-Kramer post-hoc tests also revealed no significant results for any of the datasets, female skulls (p = 0.055), female mandibles (p = 0.070), male skulls (p = 0.345) and male mandibles (p = 0.088).

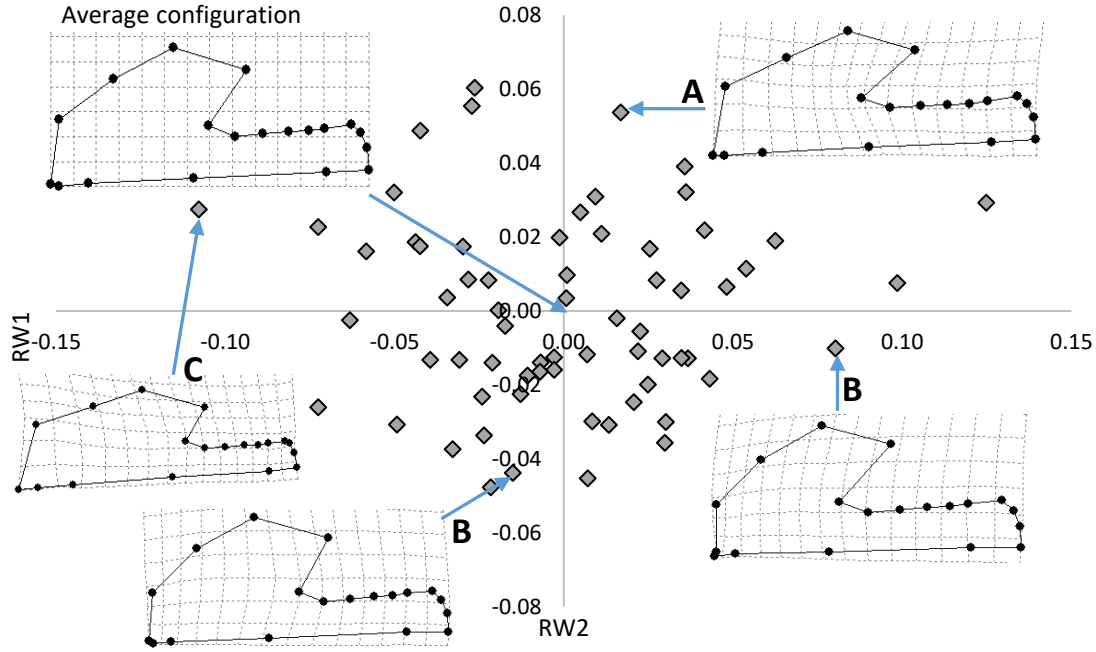
2.5.4. Regional variation in skull and mandible shape of historical western lowland gorillas

Deformation grids using the relative warps data for each of the four datasets help to visualise shape changes from the average configuration (Fig. 2.9.a-d). The outermost specimens demonstrate the most shape change from the average configuration, and examples of these have been chosen to demonstrate the extremes as well as selecting specimens from the geographical subgroups of regions A, B and C.

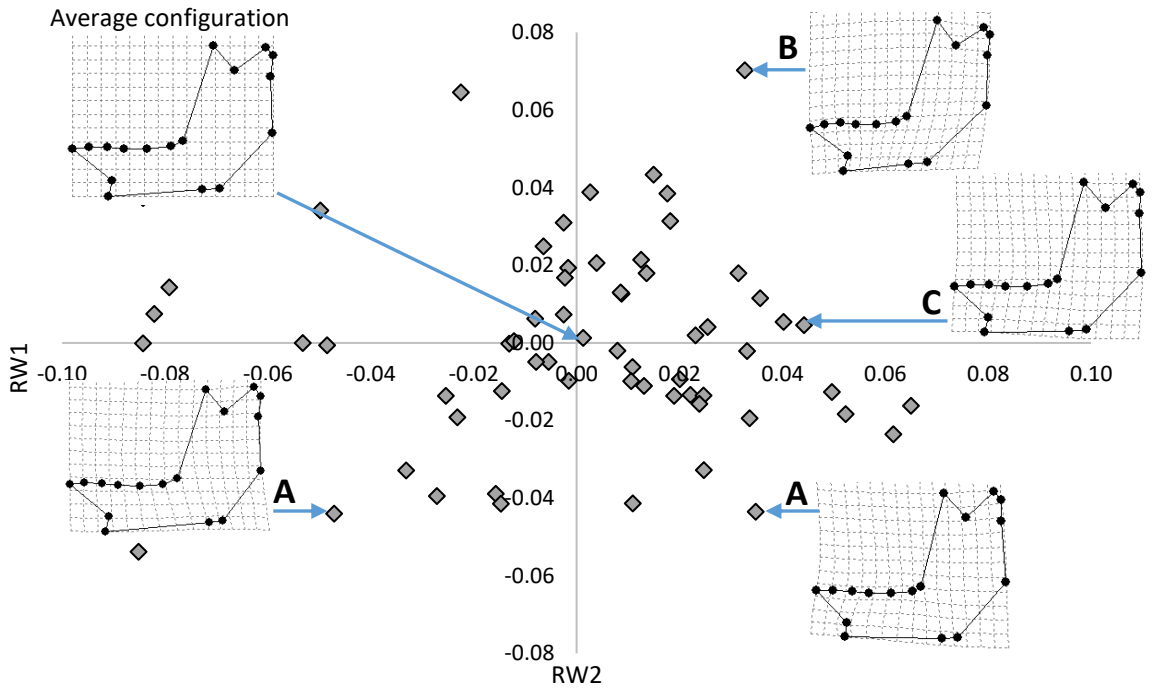
(a)


Average configuration
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
RW1
RW2
A
B
C
B

(b)


Average configuration
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
RW1
RW2
A
B
C
B

(c)



(d)

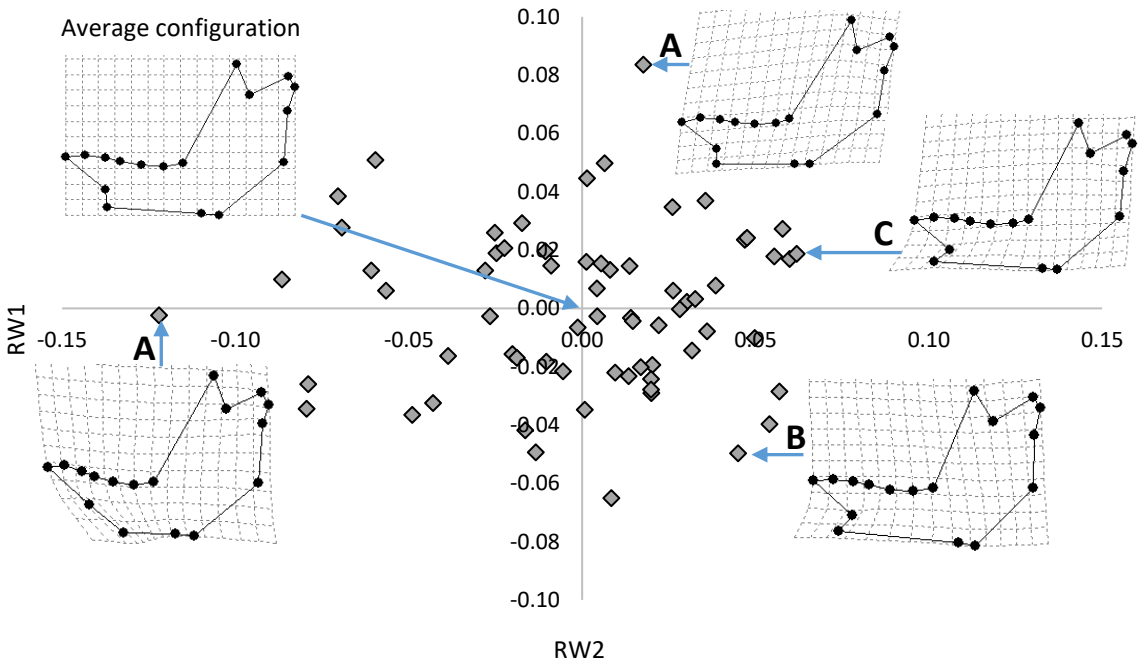


Figure 2.9 Deformation grids based on the relative warp data assist with the visualisation of shape change from the average configuration, (a) shows the shape change for the female skulls and (b) male skulls of western lowland gorillas, (c) female mandibles and (d) male mandibles. The specimens chosen to demonstrate the shape change were based on their position on the plot. The outer most specimens show the greatest variation from the average configuration; in addition, specimens were selected to represent the three regional groups A, B and C which are depicted by the capitalised letter accompanying each specimen.

Regional shape variation was assessed using MANOVA. Statistical tests were performed to assess the suitability of the parametric MANOVA testing: Mahalanobis distances for each dataset (male and female skulls and mandibles) confirmed the absence of any outliers in the data, therefore all specimens were retained for analyses. The Box's *M* tests for the female and male skulls and mandibles did not produce significant results with *p* values of 0.315 and 0.960 for female skulls and mandibles, respectively, and 0.216 and 0.89 for male skulls and mandibles, respectively; therefore, the assumptions of the equality of covariance matrices were met. Levene's tests performed on each shape variable among groups showed no significant deviations from the assumption of homogeneity of variances.

The MANOVA revealed significant regional differences for both the female (Wilks' λ = 0.7312, $F = 3.784$, $p < 0.0016$) and male (Wilks' λ = 0.6493, $F = 4.901$, $p < 0.0002$) skulls (Table 2.5). There was, however, no significant regional variation found for the mandibles in either sex. Upon further analyses, pairwise tests (Hotelling T^2), revealed that there was only a significant difference between the female skulls in regions A and B ($p < 0.0153$). Regarding the male skulls, significant regional variation was observed between regions A and B ($p < 0.0262$), and between regions A and C ($p < 0.0039$).

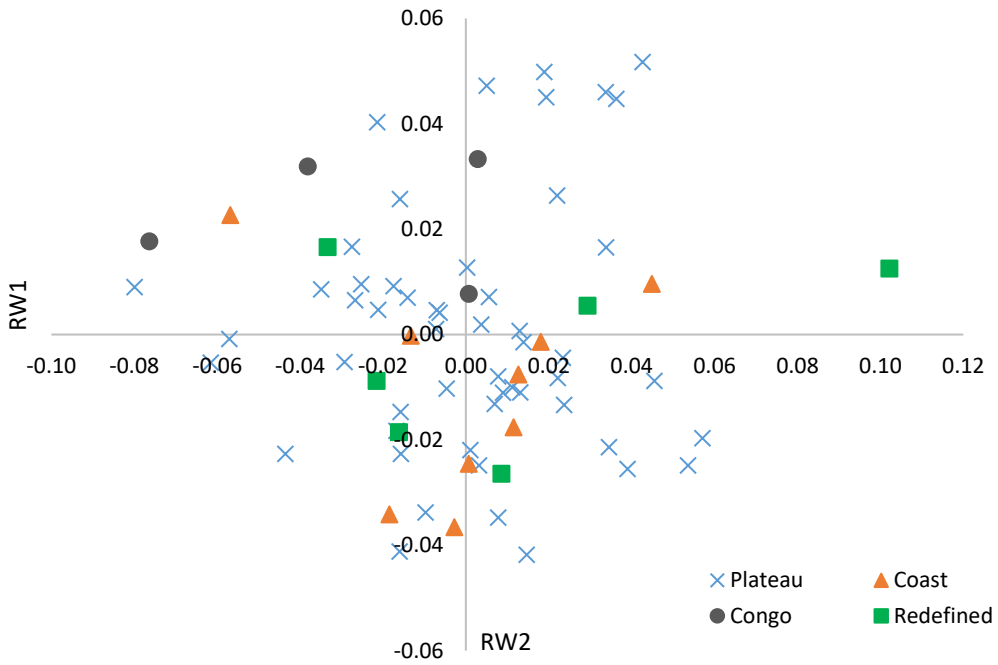
Table 2.5 Multivariate analysis of variance (MANOVA) and pairwise tests for regional variation in shape for western lowland gorillas.

	Female Skull	**Female Mandible**	**Male Skull**	**Male Mandible**
Wilks' lambda (λ)	0.7312	0.9698	0.6493	0.8202
df1	6	6	6	6
df2	134	116	122	124
F	3.784	0.2989	4.901	2.153
P	0.0016	0.9363	0.0002	0.0520
Pairwise tests p-values (uncorrected)				
Region A and B	0.0051	0.7230	0.0087	0.2244
Region A and C	0.0431	0.9198	0.0013	0.0285
Region B and C	0.0671	0.9140	0.0216	0.2297
Pairwise tests p-values (*corrected)				
Region A and B	**0.0153**	1.0000	**0.0262**	0.6731
Region A and C	0.1292	1.0000	**0.0039**	0.0854
Region B and C	0.2012	1.0000	0.0647	0.6891

** significant results are shown in bold*

Significant differences were found regionally for both the male and female skulls. However, no obvious regional clustering was observed from the relative warps data for females (Fig. 2.10.a) or males (Fig. 2.10.b). Both the redefined females and males did not show any obvious clustering and were widespread amongst the other specimens. In both sexes for the skulls and mandibles, the Congo gorillas (group C) were not significantly different from those in region B. The only significant difference observed for group C specimens was with those in group A for the male skulls, this was however, the most significant result.

(a)



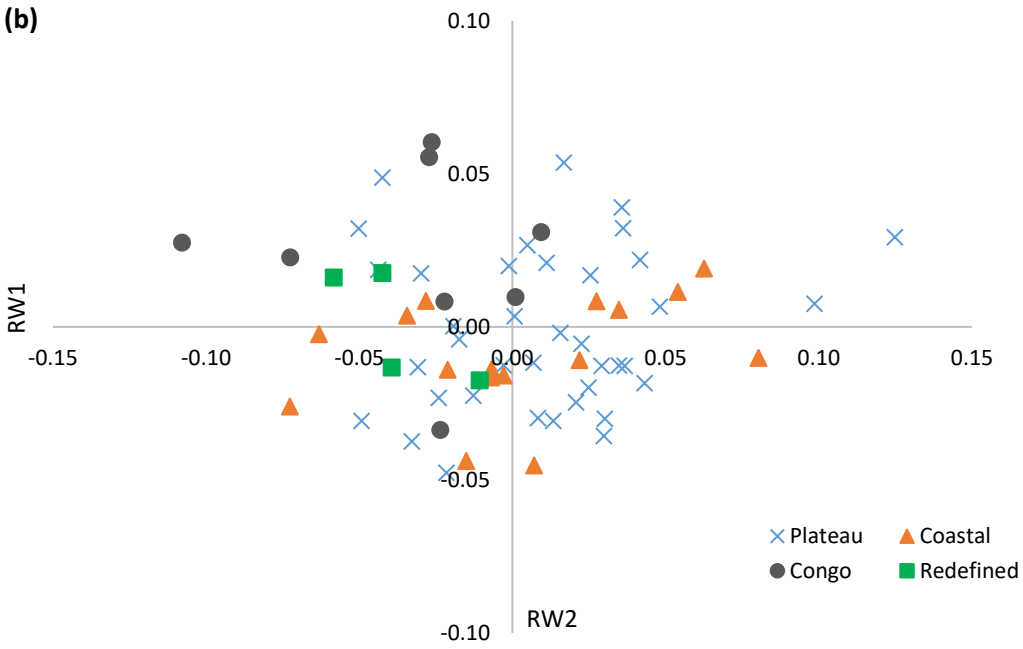


Figure 2.10 Relative warps plots for (a) female skulls and (b) male skulls of western lowland gorillas. The data was renamed to correspond to the four demes as classified by Groves (1970). Group A, has been renamed as plateau and group B as coastal. Group C has remained as the Congo group which would be classed as belonging to the coastal deme orginally following Groves (1970). The redefined group inlcudes individuals which would be classfied as belonging to the plateau deme following Groves (1970) but were placed in the coastal deme (group B).

The PCA for the female and male skulls and mandibles (Fig. 2.11.a-d) did not reveal any regional differentiation. The individuals from region C (Congo) did not cluster together and although there were more specimens in regions A and B, there does not appear to be any obvious clustering.

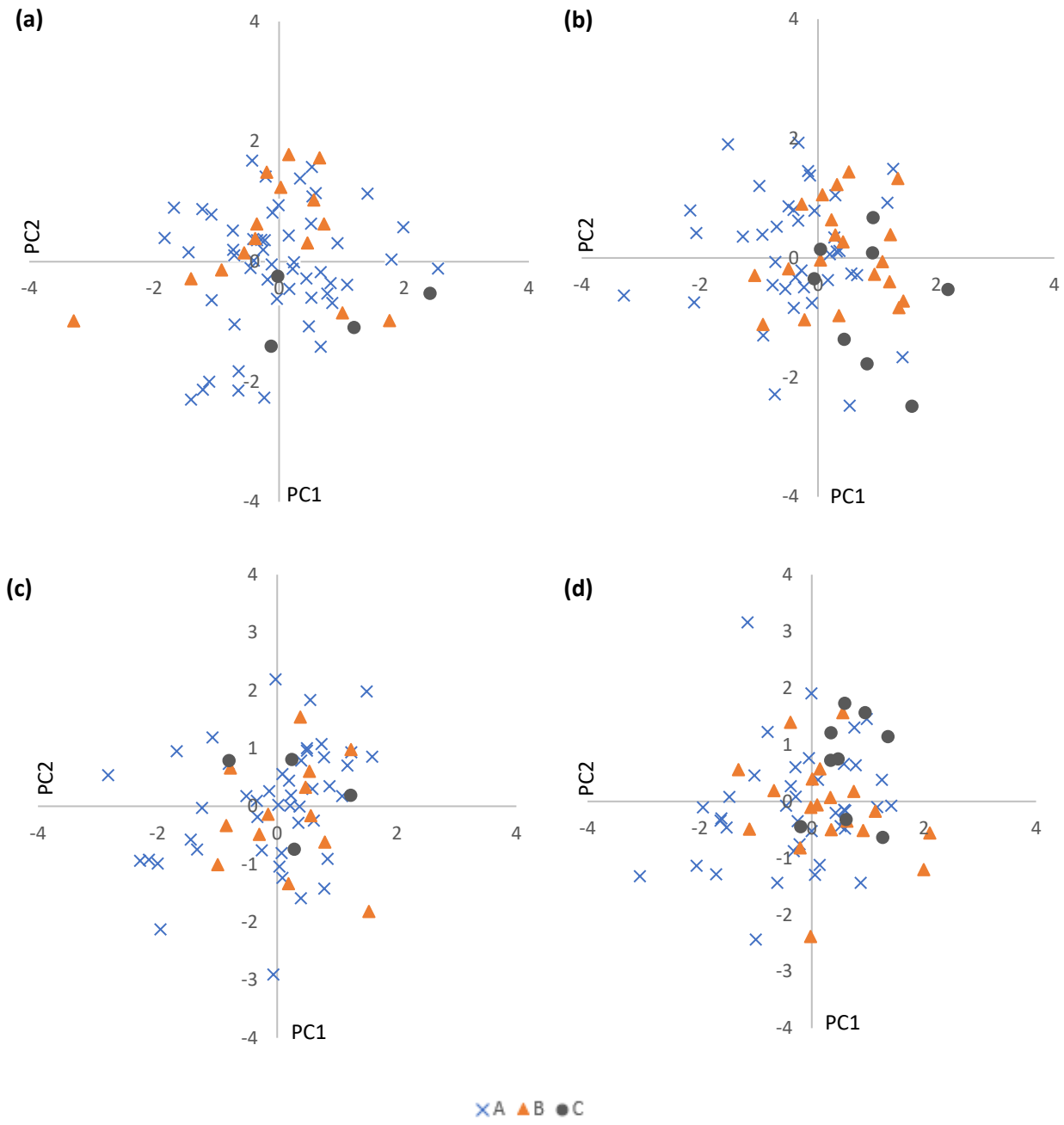


Figure 2.11 Principal Components Analysis (PCA) plots based on eigenvalues for the four datasets, (a) female skull, (b) male skull, (c) female mandible and (d) male mandible of western lowland gorillas. The first two factors were used to visualise the results. None of the four datasets showed any obvious clustering of regional groups A, B, and C.

The DFA correctly assigned 97.2% of individuals to their predefined regional group for the female skulls, however the cross-validation method correctly assigned 73.6%. For the female mandibles, 85.7% were correctly classified which decreased to 52.4% in the cross-validation. A total of 95.5% for the male skulls were correctly classified decreasing to 42.4% in the cross-validation and

male mandibles had 92.5% correctly classified and again, decreased to 59.7% in the cross-validation. The DFA plots using the first two factors (Fig. 2.12.a-d), however, showed certain level of shape differentiation among the groups. The cross-validation indicates that regional groupings could occur simply by chance as opposed to strictly defined regional variation.

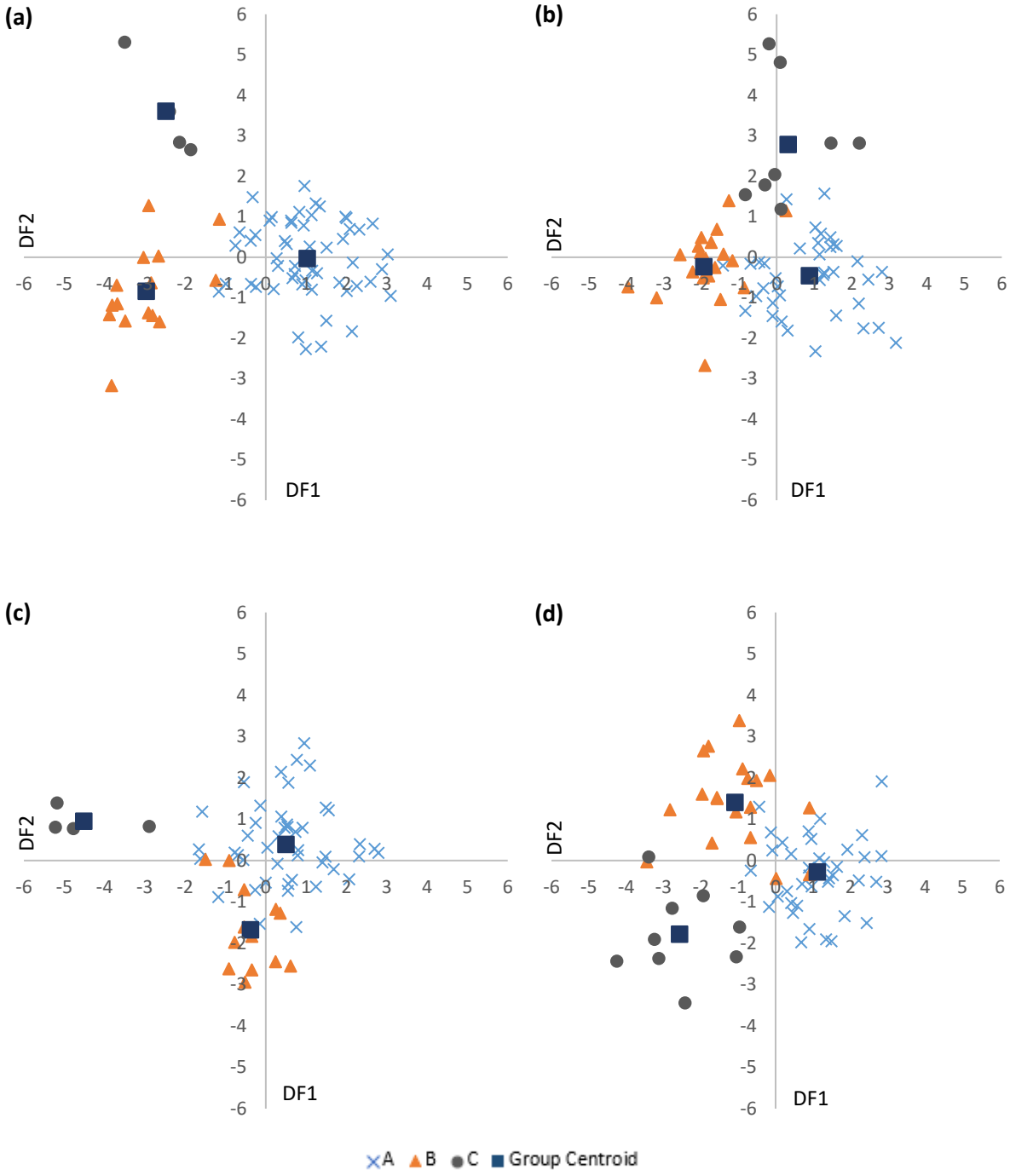


Figure 2.12 Discriminant Function Analysis (DFA) plots for the four datasets, (a) female skull, (b) male skull, (c) female mandible and (d) male mandible of western lowland gorillas. The first two factors were used to visualise the results.

2.6. Discussion

Geographic variation is common among primate species and knowledge of geographic variation within species is fundamental for taxonomic classification and conservation management (Albrecht & Miller 1993). Primatologists investigating geographic variation rely on the significance of geographic differences in external (colour coat and pattern) and internal (skeletal) morphology to confirm or contradict results (Albrecht & Miller 1993). Traditional morphometric methods of the gorilla species in previous literature has revealed significant variation between the four sub-species with most of the variation observed between eastern and western populations rather than between populations residing in the north and south (Groves 1970; Stumpf *et al*. 2002; Albrecht *et al*. 2003).

The founding research of gorilla systematics and taxonomy is based on the works of Coolidge (1929) and Groves (1967, 1970). However, as reviewed by Haddow & Ross (1951) the research by Coolidge (1929) is not without inconsistencies and errors. For example, one of Haddow and Ross' (1951) critiques of Coolidge (1929) was that despite noting that young/juveniles should be excluded from morphometric analyses to avoid biasing the data, and that female skulls are frequently misclassified as young males (and therefore should also be excluded), Coolidge (1929) went on to include juveniles and females in the analyses despite justifying his research methods for basing classification on male adult male skulls only.

This study used the same specimens from the Powell-Cotton Museum that were used by Coolidge (1929) and subsequently Groves (1967, 1970), and also found discrepancies in Coolidge's methods. For example, Coolidge (1929) has the locality of the PCM specimens FC.195, FC.163, FC.216, FC.225, FC207, FC.196 and FC.207 as Gabon. Visualising the geographic data using GIS and investigating the contextual information further, revealed none of the PCM specimens are from Gabon. They were all collected in what is now the Republic of Congo which belonged at the time of collection to the French Congo. Further to this, those specimens are also geographically widespread. Specimens FC.195, FC.196 and FC.207 were captured in region A of this research, unlike the remaining FC specimens which were grouped in region C of this research. Investigating the trip logs and specimen cards available in the PCM confirms that this difference in geographical location is valid as the collection dates of those specimens placed in region A, were collected one month after the region C specimens, and follow the trajectory of the field notes and diaries of Major Powell-Cotton. Identifying the correct origin of samples is important while studying geographical patterns of morphological variation, therefore highlighting the usefulness of contextual data for this study.

Groves' (1970) revision, which contains hundreds of specimens (including the PCMs) and includes a more thorough and detailed analyses using multivariate methods, provides a more robust and solid foundation than that of Coolidge (1929) and is supported by analogous studies such as Uchida (1998) who concluded that the dental morphology of *G. g. gorilla* is highly variable (Leigh *et*

al. 2003). In summary of Groves' work and others such as Uchida (1998), Stumpf *et al*. (2002), Albrecht *et al*. (2003) and Leigh *et al*. (2003), the sub-species *G. g. gorilla* studied here showed high morphological variability throughout its geographic distribution with considerable 'overlap' of variation. Despite the high variability and overlap, significant regional morphological variation was observed, which is consistent with the contemporary designation of populations into demes (Groves 1970; Leigh *et al*. 2003).

2.6.1. Skull and mandibular regional variation in western lowland gorillas

The results of this research were generally consistent with morphological variation findings of previous research, e.g. Groves (1970), Stumpf *et al*. (2002), Leigh *et al*. (2003); however, some dissimilarities do occur. In terms of size (LnCS), the mean size for females was considerably smaller than males reflecting the extreme sexual dimorphism within the species in general. In all four datasets, individuals in region C showed smaller mean sizes compared to individuals in regions A and B, which was a surprising result given that the gorillas in region C were placed in the same deme (coastal) by previous research by Groves (1970), however, Groves (1970) did note observable (minor) morphological differences of individuals from this region.

Female skulls in region A were on average, slightly smaller than those in region B. The reverse was found for male skulls with the average size in region B being smaller than those in region A. The observed morphological differences may be a result of regional diet/habitat, but previous studies have found that the diet of western lowland gorillas is comparable among regional sites and with considerable overlap (Doran & McNeilage 1998; Rogers *et al*. 2004), however behavioural studies and those involving diet are considerably difficult to carry out due to the inaccessibility of the habitat, therefore research in this area is somewhat lacking (Cipolletta 2003; Genton *et al*. 2012). Additionally, the same result was not found across both sexes consistently by region, in fact, the reverse was found, indicating that diet is not likely to be the factor influencing size differences as the diet for both sexes would be regionally the same.

Slightly more variation in mean size was seen among the male skulls than female skulls. Females have been reported to be more homogeneous in their skull variation compared with males (Albrecht *et al*. 2003). Regarding the mandibles, the same pattern was observed in the males with individuals from region A being larger than those in region B, the same pattern was found in the female skulls with individuals from region A being larger than those in region B (Fig. 2.7). Despite observed differences, statistical analysis by means of ANOVA showed that none of these differences were

significant. The results for centroid size were interesting, as they confirmed observed regional differences but not at significant levels.

If this study had only investigated skull and mandibular size, then it would have concluded that there was no regional variation amongst the western lowland gorillas. The shape analyses, however, did produce some significant findings. This indicates that shape, rather than size, was the most notable difference within this subspecies of gorilla. The MANOVA for mandibles in both sexes revealed no significant regional variation, with the female mandibles showing the least variation of the two sexes. Again, this followed the observations from previous studies that the females of the gorilla species are more homogenous than the males in terms of morphology (Albrecht *et al*. 2003). However, in terms of skull morphology, pair-wise tests (Hotelling T^2 and Bonferroni corrected *p*-values) did reveal regional variation in both sexes. For females, this significant difference in skull morphological variation was found between individuals in region A and those in region B, however, no significant difference was observed in relation to region C. This was interesting, firstly, as it contradicted the results found when looking at just size, where region C was observed to be the 'most' different, although not significantly. Secondly, when comparing these results to those of Groves (1970) some similarities and differences were noted: this research supports Groves' (1970) findings in relation to deme morphological variation, where regions A and B (the plateau and coastal demes) were significantly different; and that no regional variation was found between individuals in region C compared with region A and B. Region C individuals were most similar to region B (the coastal deme) which is the deme Groves (1970) had originally assigned them to.

The pattern of deme morphological variation was also found when investigating the male skulls. Here, significant differences were observed between regions A and B (plateau and coastal demes) and between regions A and C, the most significant difference was observed in the latter (p = 0.0039). Again, this fitted with Groves' (1970) demes because those individuals in region C should be classed in the coastal deme (region B), and because there was no significant difference between male skull morphology between those individuals in region B and C, thus confirming that individuals in region C were more similar to those in region B (coastal deme) as described by Groves (1970).

The relative warps plots for the female and male skulls (Fig. 2.10. a, b) however, did not show any clear regional clustering. In fact, individuals from all three regions were widely distributed and overlapping, confirming as did Groves (1970) and other studies (e.g. Stumpf *et al*. 2002) that there is high morphological variability within the western lowland gorilla subspecies. Even with the re-classification of the ten specimens originally placed in region B (coastal deme) which should have been placed in region A (plateau deme) (Fig. 2.10.a, b), including six females (CamI.149, CamI.150, CamI.139, CamI.97, CamI.109 and CamI.98) and four males (MI.28, ZVI.32, CamI.107 and CamI.134), it was not clear which cluster/region/deme they belonged to as they were widely distributed and did

not form any observable clustering. Had they been classified as plateau or coastal gorillas' group (group A or B), this would not have influenced the general outcome of the results, and the specimens would have remained regionally widespread in their skull shape variation with no obvious clustering.

Likewise, the PCA did not reveal any obvious regional clustering in any of the four datasets (Fig. 2.11). The PCA results obtained here showed that generally, across the four datasets, gorillas in region A (plateau deme) contained the most variation as they were the most widely distributed in the plots. The DFA, which is a multivariate technique that maximises the differences of predefined groups, did reflect regional groups (Fig. 2.12). Both the male and female plots for the skull data had minimal overlap, slightly more overlap could be seen for the mandible data plots. They all yielded high percentages for correct assignment to groups ranging from 85.7 to 97.2%. However, the cross-validation results did see these percentages fall to a range of 42.4 to 73.6% indicating that in some instances at least, chance was very much a factor. However, these results do indicate that some regional morphological variation in shape (not size), does exist in the western lowland gorilla. These results appear to be plausible when viewing the results in their entirety. The relative warps plots and PCAs did not show any clear regional clustering but the MANOVA did identify significant regional variation. These findings were similar to Groves (1970) in that morphological variation among the western lowland gorilla is highly variable throughout its distribution, but there does appear to be significant regional variation at least when examining skull morphology, not mandibular. Revisiting the gorilla field notes from Major Percy Powell-Cotton confirmed that morphological and phenotypic variation was extremely evident even in instances where individuals were captured at the same location and that this had also been recorded previously by Akerley (1923). This was confirmed by inserts such as the one earlier regarding specimens Mer.138 and 139, where the Major noted *"these two beasts 138 and 139, show the difference in shape of face, as so well shown in Carl Akerley's "Brightest Africa" page 222."*

The aims of this chapter were to investigate regional morphological variation of a historic wild population of western lowland gorillas by application of geometric morphometric methods and compare those findings to previous studies that used traditional morphometric methods e.g. Groves (1970), and to draw conclusions upon the existence of regional demes as classified by Groves (1970). The findings of this study produced mixed results with no significant levels of regional morphological variation found in either sex when investigating skull and mandibular size. However, some level of regional shape variation was detected in both sexes for skulls but not for mandibles, and regional shape variation was more observable in males compared to females. This indicates that in some cases at least, morphological traits could be under selection despite potential gene flow amongst the western lowland gorilla populations and other subspecies.

The hypotheses and predictions for this chapter stated that regional morphological variation in size and shape would be observed between groups A and B, to reflect the morphological demes (plateau and coastal, respectively) identified by Groves (1970). In terms of size, this research does not support the demes classification with significant results for regional morphological variation, however, differences were observed at non-significant levels. In terms of skull shape, this research does support the demes classification for both sexes. Groups A (plateau deme) and B (coastal deme) showed significant variation for both sexes thus supporting the demes classification, and for males significant regional variation found between groups A (plateau) and C (the Congo gorillas that would fall into the coastal deme) but not between groups B and C, again, supporting the demes classification. The statistical tests (ANOVA and MANOVA) support these findings as does the DFA whereas the PCA did not detect obvious regional clustering but showed considerable overlap between regions, which was also identified by Groves (1970). Additionally, it was predicted that there would be more observable variation in males than females based on previous literature (Albrecht *et al*. 2003), the results found here supported this also.

2.6.2. Further research

To improve and further this research there are several recommendations. Firstly, the lateral view of the skulls could be included to allow for the analysis of the sagittal crest among other prominent features of the skull. Morphology of the sagittal crest amongst the gorilla species has been the subject of many evolutionary, ecological and taxonomic investigations, e.g. Sarmiento & Oates (2000), Breuer *et al*. (2012) and Balolia *et al*. (2017). The contextual information at the PCM indicates that observations on regional crest size were noted "*All Gorilla coming from the forest NW and W. from here (Arteck 3¾ N. 14¼ E.) seem to have smaller crests in comparison with those coming from the E & S districts."* (PCM gorilla field notes). However, the sagittal crest in museum specimens could be damaged, or could show a lack of homologous landmarks across specimens, making this type of analysis potentially unproductive and time-consuming. Although no regional variation was observed for the lateral view of the mandibles in this research, the dorsal view may yield different results. Morphological variation in the masticatory apparatus of the skull could then be explored in terms of feeding habits and food items across the range of the Western lowland gorilla. Furthermore, analyses using 3D geometric morphometrics such as Fleagle *et al*. (2010) would likely capture more morphological data and analyses using computed tomography data such as Ito (2019) has increased benefits of investigating internal structures, which is absent from 2D morphometric techniques. These

further analyses would provide a more robust set of results and perhaps provide further insight into regional variation amongst the western lowland gorilla population.

Secondly, the most robust and preferred continuation of this research would be to repeat the research of Groves (1970) incorporating all the specimens he used but applying the geometric morphometrics methods which were not available fifty years ago. This would allow a fully comparable investigation using different methodologies. It would be interesting to see if the same results were obtained.

Thirdly, if analysis on the entire dataset were possible, the specimens used in this study could be 're-grouped' into their appropriate demes from the start. This was not possible with this study as the Sangha deme did not have enough specimens to represent a further group, thus they were combined into group A. In addition, 10 of the specimens used here were placed into group B (coastal deme) which was not strictly following the demes classification but occurred due to grouping individuals based on the specimens available for analysis representing the furthest geographical distances.

Chapter 3

Past and present populations of the western lowland gorilla: Mitochondrial Hypervariable Region I

3.1. Introduction

Until the 1960s, the evolutionary establishment of relationships among species, was primarily based on morphological data (Moritz 1995; Clifford *et al*. 2003). Genetic studies of gorillas, despite becoming more abundant in recent years, are comparatively still few in relation to other species (Garner & Ryder 1996; Clifford *et al*. 2003), but their importance in systematic studies is becoming increasingly significant (Clifford *et al*. 2003). Genetic studies on species in general often complement ecological and/or morphological findings, but there are exceptions (Harris & Disotell 1998). Mitochondrial DNA (mtDNA) research has exposed significant variation within western gorillas (Gagneux *et al*. 1999); however, the geographic data is often absent from such studies as few samples are of known geographic origin (Clifford *et al*. 2003).

Genetic and demographic management of captive populations is of vital importance for the conservation of endangered/critically endangered species (King & Courage 2008, King *et al.* 2012; King *et al*. 2014), as genetic variation is a crucial requisite for the long-term survival of a species, facilitating evolutionary adaptive change to the environment and future evolutionary potential (Brennan et al. 2019; Razgour *et al*. 2019). With natural populations continuing to decline (Tutin *et al*. 2005; Guschanski *et al*. 2009; Soto-Calderón *et al*. 2015; Estrada *et al.* 2017), the concern to maintain genetic variation in the captive populations is of utmost importance. Despite this concern, limited information is available regarding the scope to which genetic variation is portrayed in the captive populations compared to their wild counterparts (Soto-Calderón *et al*. 2015).

Mitochondrial DNA (mtDNA) has proven to be one of the most favoured markers used within the fields of population genetics and phylogeography (Hillis *et al*. 1996; Avise 2012). This is due to mtDNA representing one single nonrecombining locus from which ancestral relationships can be ascertained (Thalmann *et al*. 2005; Calvignac *et al.* 2011; Avise 2012). Inherited solely via the maternal lineage, and evolving faster than nuclear DNA, the mitochondrial control region (MCR) has demonstrated to be particularly effective at determining population structure (Clifford *et al*. 2003; Soto-Calderón *et al*. 2014; Popadin *et al.* 2017). Within the MCR lies the first hyper-variable region (HVI) which has been a commonly used molecular tool for a variety of taxa (Finnilä & Majamaa 2001). However, nuclear transfers of mtDNA copies (numts) have been revealed in studies on human and

non-human primates (Popadin *et al.* 2017); this unintentional amplification of numts is often associated with the HVI region and is prominent in gorilla studies using this approach. The prevalence of numts within the MCR of primates has been acknowledged in many studies, e.g. Jensen-Seaman *et al*. (2004), Thalmann *et al*. (2005), Anthony *et al*. (2007), Douadi *et al*. (2007), Soto-Calderón *et al*. (2014), Dayama *et al*. (2020). However, the genus *Gorilla* is affected the most extensively, which significantly impacts the reliability of using mtDNA for population genetics purposes (Jensen-Seaman *et al*. 2004; Thalmann *et al*. 2005; Soto-Calderón *et al*. 2014). DNA sequence data obtained from numts, and used as part of a data set in population genetics or phylogeographic studies, result in increased genetic diversity estimates within populations, and in disparate levels of genetic structure among populations due to having different mutational rates compared with true mtDNA sequences found within mitochondria (Hacia 2001; Clifford *et al*. 2003; Hlaing 2009). Despite the implications that numts can incur, there remains an abundance of literature relating to the HVI region, which makes it possible to use the data from other studies, and to compare genetic diversity and structure among different studies, if the DNA sequence data is used with caution.

Previous literature of wild (historic and contemporary populations) as well as captive populations have identified four major mtDNA haplogroups within the gorilla species. Haplogroup A is restricted solely to the eastern mountain gorillas, and haplogroup B is specific only to eastern lowland gorillas (Clifford *et al*. 2004). Haplogroups C and D are specific to western lowland gorillas, with each of them consisting of three haplogroup subgroups: C1, C2, C3, and D1, D2 and D3 (Soto-Calderón *et al*. 2015).

3.2. Gorilla populations, past and present

Captive populations are at risk from the loss of genetic diversity in the same ways in which small wild populations are at risk (Lande & Barrowclough 1987; Vrijenhoek 1994; Gooley et al. 2018; Ayala-Burbano *et al*. 2020). Therefore, studies which investigate the genetic diversity and genetic 'health' (see Chapter 1) of captive populations are essential to ensure the long-term survival of the species/subspecies, particularly of critically endangered species who are already facing small population numbers (Vrijenhoek 1994; Ayala-Burbano *et al*. 2020).

Clifford *et al*. (2004) investigated the mtDNA phylogeography of wild populations of the western lowland gorilla analysing 53 sequences, generated from their study, and an additional 30 sequences from GenBank, two of which were from museum specimens and one was from the Powell-Cotton Museum (PCM). Their study concluded, in general, that western lowland gorilla mitochondrial lineages, do display regional distinctions represented by haplogroups with some degree of admixture.

In what is perhaps the most comprehensive study on western and eastern gorillas, Soto-Calderón *et al*. (2015) used a total of 249 HVI sequences representing all the major mitochondrial lineages that were currently available, from which 42 sequences were generated from their research on US captive zoo population of western gorillas, 10 sequences were obtained from GenBank and 197 were reference sequences (from Anthony *et al*. 2007). Their study confirmed that the US captive zoo population had retained the entirety of the main mitochondrial lineages of western gorillas throughout their range. Additionally, a novel lineage (haplogroup C3) was observed in their analysis which had previously only been depicted in gorillas originating from Cameroon.

Previous literature regarding mtDNA research has recommended further investigations to assist with the conservation of western lowland gorillas. For example, Clifford *et al*. (2004) recommended further research focusing on nuclear loci in addition to behavioural and ecological data collection, and Soto-Calderón *et al*. (2015) recommended that the mitochondrial haplotypes of all remaining captive gorillas be determined to provide additional genetic information to guide and assist current and future breeding and conservation programs.

As human pressure on wild populations continues to rise, specimens held in museums have become increasingly important repositories of biodiversity, representing the biological past and permitting research by taxonomists, morphologists and anatomists alike (Burrell *et al*. 2015). Museum collections provide a resource which is historically unique and can contribute significantly in a variety of ways to molecular studies (Austin & Melville 2006; Flanagan *et al.* 2017). Retrieving DNA from specimens can be invaluable in conservation genetic studies where declining or extinct populations and species are the focus (Austin & Melville 2006). However, DNA from natural history collections (NHCs) is typically degraded (Burrell *et al.* 2015; Sproul & Maddison 2017), and/or sample size is small (Wandeler *et al*. 2007), and often specimens are of unknown origin or records are incomplete (Soberón & Peterson 2004; Pyke & Ehrlich 2010), which makes those with additional accurate contextual information/data even more valuable. Despite associated issues that can arise from working with NHCs, there are numerous successful molecular based studies encompassing a variety of taxa, e.g. Higuchi *et al*. (1984), Cooper *et al*. (1992), and Rohland *et al*. (2004), and include studies on primates, e.g. Boubli *et al*. (2008), Guschanski *et al*. (2013), and van der Valk *et al*. (2017).

3.3. Aims

The aims of this study were:

- To investigate the genetic diversity of the mtDNA HVI in relation to the contemporary UK captive population at the Aspinall Foundation and a historical sample of wild western lowland gorillas from the Powell-Cotton Museum NHC.
- To investigate regional genetic variation of western lowland gorillas, including in the analysis HVI sequences from GenBank (which include eastern gorillas), to study the evolutionary relationships among the haplogroups, and particularly identify the haplogroups for the contemporary UK captive population of western lowland gorillas, which may assist with future conservation efforts of the subspecies.
- To identify nuclear pseudogenes of mtDNA sequences (numts) from true mitochondrial sequences to decrease the unreliability of using data where numts are present.

3.4. Hypotheses and predictions

Previous literature has identified four major mtDNA haplogroups within the gorilla species, haplogroup A (eastern mountain gorillas), haplogroup B (eastern lowland gorillas), haplogroup C and D that are specific to western lowland gorillas, with each of the latter two haplogroups consisting of three haplogroup subgroups: C1, C2, C3, and D1, D2 and D3 (Clifford *et al*. 2004; Soto-Calderón *et al*. 2015). Additionally, numts are known to be prevalent in gorilla mtDNA studies and the result of not identifying numts can cause inaccurate phylogenetic inferences to be made (Tay et al. 2017; Bingpeng et al. 2018; Kunz *et al.* 2019) and worse still, can be reported as true mtDNA sequences. Therefore, it is expected that:

- The western lowland gorilla samples and specimens used in this study from a contemporary captive population (Aspinall Foundation), and a historical wild population (PCM) will belong to haplogroups C or D.
- Regional variation of the historical population will be present and reflected in haplogroup distribution which can be confirmed with the use of the biodiversity mapping of the historical populations used in previous chapters.
- The contemporary population will contain more known related individuals thus, there may be less genetic diversity and variation in the contemporary population compared with the historical population.
- Numts are likely to occur in this dataset and, if not identified, may produce unreliable results.

3.5. Methods

3.5.1. Sampling and DNA extraction

In total, there are 242 gorilla specimens in the PCM collection. Of those, 182 specimens have skins, while the remaining 60 specimens consist of skulls and/or skeletons only. Of the 182 skins, four are mounted in the museum and thus cannot be sampled. The remaining 178 gorilla skin specimens in the PCM collection had a small piece of skin (approx. 1 cm^2) removed for sampling, and the samples remain at Canterbury Christ Church University (CCCU).

Originally, it was anticipated that a small 1 cm^2 would be taken from a part of the skin which would cause the least amount of damage to the specimen. Hands, face and feet were to be left intact. However, due to the way in which the skins had been preserved, stored and folded, this proved problematic. Many of the skins were inflexible which meant taking a sample from the same area of each gorilla was not possible, a large proportion also had other damage from moths, and others were damaged from disease (primarily yaws, or gorilla treponematosis), which they would have suffered from prior to their death. The method in which the gorillas had been skinned in the field meant that, in most cases, an incision had been made from the back, at the base of the neck, straight down to the fork/upper legs. The skins were then folded and stored in a generally consistent manner, which meant that the most pliable area to sample was along the incision line towards the lower back/upper leg. To avoid cross contamination of samples, full Personal Protective Equipment (PPE) was worn. Using sterilised equipment consisting of a scalpel, forceps and scissors (submerged in ethanol then held in a flame), several thin scrapings of skin were removed. The first scraping was discarded as was often coated in the preservation substance (naphthalene), the following 3-4 scrapings, approximately 1-2 mm in thickness and 5-8 mm long, were stored in 2 ml microcentrifuge tubes and clearly labelled with the specimen number, sex and date of collection. Sterilisation of equipment took place between each specimen.

From the PCM samples, DNA was extracted from 74 gorilla skins and three extra skin/tissue samples from the collection at the Royal College of Surgeons (RCS). These 77 samples represent the historic population (N = 77) of western lowland gorillas and were collected over a duration of nine years from 1927 to 1936 from the wild in equatorial Africa. Geographical coordinates were recorded at the time of capture by Percy Powell-Cotton and/or Fred Merfield who were the hunters that contributed to the collection at the PCM. To visualise the PCM data and ascertain regional groupings, a map was produced in ArcGIS and the sequences were divided into three historical regional subgroups (A, B and C, Fig. 3.1) following the same methods as in the previous chapter. The three specimens from the RCS do not have geographical data recorded but archival evidence indicates that those specimens were originally part of the PCM collection and were donated by Percy Powell-Cotton to the RCS.

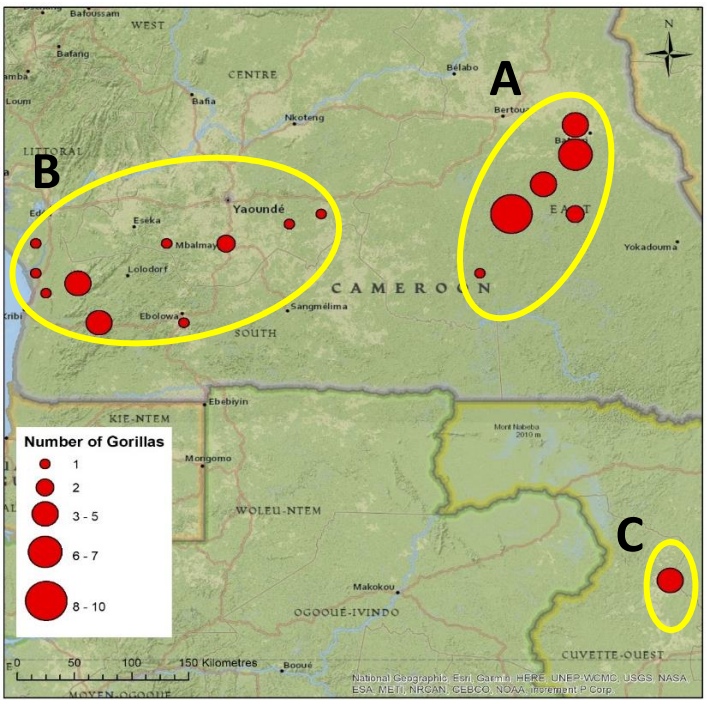


Figure 3.1 Map of the Powell-Cotton Museum (PCM) gorilla specimens used for mtDNA analysis showing the three defined subgroups A, B, C based on geographical origin.

The contemporary population of western lowland gorillas is represented by blood samples (Whatman FTA 'blood' cards) and tissue samples (N = 59) all previously collected by the Aspinall Foundation and includes family groups. Six silverbacks have sired many of the offspring present in the captive group, namely: Djala, Djanghou, Kouillou, Kifu, Kijo and Bitam. Although there are many other individuals that have reproduced and contributed genetically to the Aspinall group, e.g. Asato and Sammi, they have not sired as many offspring as the six silverbacks. Family trees for the six silverbacks are shown in Figure 3.2. A complete list of the individuals for which FTA/tissue samples were available for this study, and their relationships are given in Appendix 1.

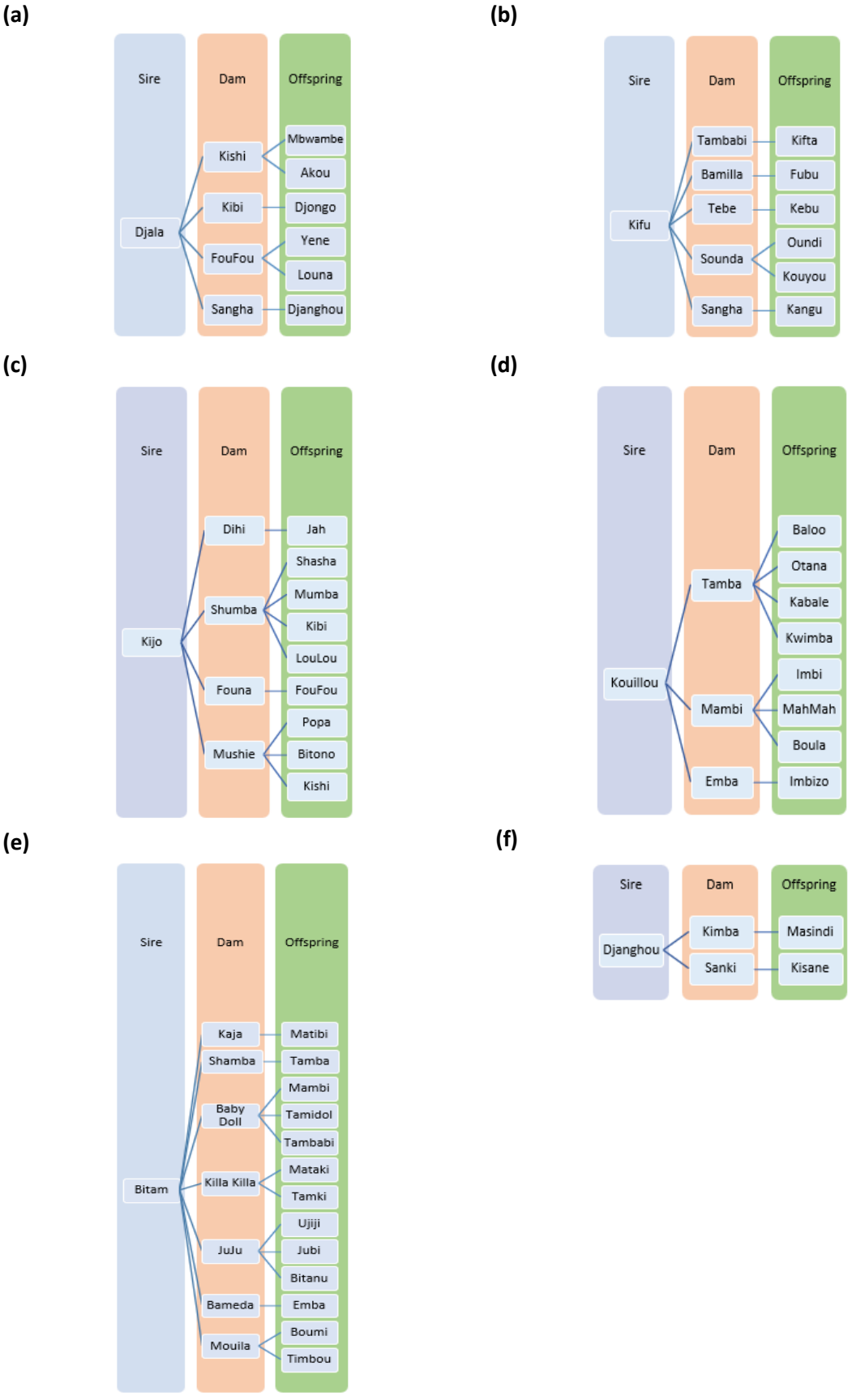


Figure 3.2 Family trees for the six main silverback males from the Aspinall Foundation used in this study, (a) Djala, (b) Kifu, (c) Kijo, (d) Kouillou, (e)Bitam, (f) Djanghou (Djanghou is Djalas' offspring).

The total sample including historic and contemporary DNA samples therefore totalled 136 western lowland gorillas (N = 136). DNA was extracted from the museum samples and for the contemporary tissue samples from the Aspinall Foundation using the QIAamp Fast DNA Tissue Kit (Qiagen, UK), following the standard protocol, with exceptions to step 5, where the incubation period was increased to 50 minutes, and step 11, where 50 µl of buffer ATE was added and incubation at room temperature was increased from 1 to 5 minutes.

DNA was originally extracted from the FTA cards by following the methods as described by Fowler *et al*. (2012). Gel electrophoresis was used to check for the presence of DNA (Fig. 3.3). Gel electrophoresis is a technique that is used to separate DNA fragments according to their size (Voytas 2000). The gels were made at a 1% concentration of agarose in 1X TAE buffer and stained with 4 µl of SYBR®safe (Invitrogen); 5 µl of DNA were combined with 2 µl of loading dye and loaded into the wells, the last well was reserved for 2 µl of the molecular weight marker which consists of DNA ladder, loading dye and molecular grade water; gels were run for approximately 30 minutes at 100 V. Visualisation of DNA was performed under UV-light on a Bio-Rad's Gel Doc XR+ (Bio-Rd, UK).

A Qubit fluorometer was used to quantify the DNA. However, the results were extremely variable with many samples not giving any reading due to insufficient quality, and others ranging from 0.59 ng/ml to a maximum of 199 ng/ml. The same samples were retested at separate times and yielded different results. Due to the inconsistency of the Qubit results the decision was made to proceed without quantification. Additionally, several amendments were made to the FTA DNA extraction protocol to increase DNA quality, the revised protocol is given in Appendix 2. From the FTA cards, 22 produced DNA that was viable for downstream analysis using the original protocol. The remaining FTA cards which had previously failed to produce viable DNA were subjected to the revised protocol and yielded results.

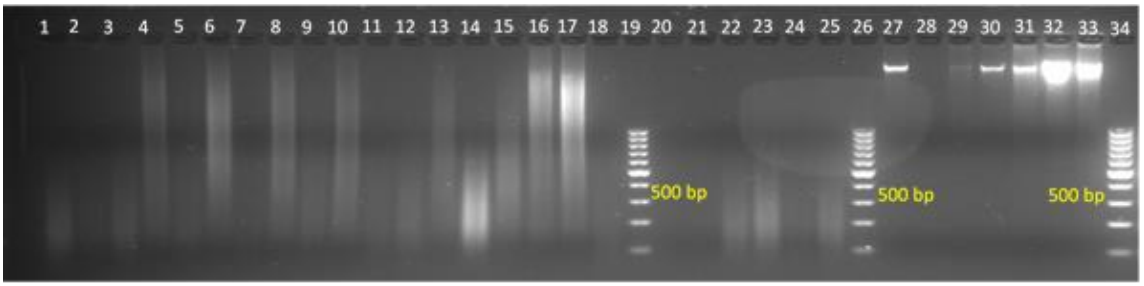


Figure 3.3 Agarose gel electrophoresis of western lowland gorilla DNA (1% agarose in 1X TAE buffer, stained with SYBR®safe, run for 30 minutes at 100 V, and visualised under UV-light). From left to right, wells 1-18 contain Powell-Cotton Museum (PCM) samples: FC.122, FC.124, CamI.44, CamI.45, CamI.46, CamI.14, ZVI.32, M342, MI.30, CamI.97, CamI.106, CamI.109, CamI.110, MII.6, CamI.95, M135, M879 and MII.25. Wells 20-25 contain repeated DNA extractions for the Royal College of Surgeons (RCS) samples: PA61-A, PA61-B, PA62-A, PA62-B, PA63-A and PA63-B. Wells 27-33 contain the extracted DNA from the Aspinall Foundation tissue samples, including 58-Masindi infant, 59-Kwimba 2nd infant, 60-

Kouilla, 61-Babydoll, 62-Louna, 63-Virginika infant and 64-Mouilla. Wells 19, 26 and 34 contain the molecular weight marker/ladder (100 bp DNA ladder). The gel shows high molecular weight and intact DNA (e.g. well 27), smeared DNA including high molecular weight fragments (e.g. well 33), and low molecular weight fragments (e.g. well 14), and of different intensities due to varying DNA quantities (relative brightness).

The PCM and RCS samples appeared as smears of DNA (Fig. 3.3); this is not uncommon for museum samples as they are often degraded. Where possible, fresh blood or tissue samples should be used (Knebelsberger & Stöger 2012), obviously, the use of fresh museum samples was not a possibility for this study. The tissue samples from the Aspinall Foundation (wells 27-33 of Fig. 3.3), showed high quality DNA for many of the samples. The FTA cards, although not as degraded as the museum specimens, still produced a smear indicative of DNA degradation. However, degraded DNA could still be used for the amplification of HV1 by Polymerase Chain Reaction (PCR; see below).

3.5.2. Mitochondrial DNA amplification and sequencing

Following Garner & Ryder (1996) and Clifford *et al*. (2004), nested primers were used to amplify a short 258 base-pair fragment of the mitochondrial hyper variable region I. First-round primers PDPF1 (5'-CACCATCAGCACCCAAAGCTAATAT-3') and PDPR2 (5'-TTGTGCGGGATATTGATTTCACGGA-3'), and the second-round primers L91-115 and H402-27 (Garner & Ryder 1996; Clifford *et al*. 2004) followed the same cycle conditions. Cycle conditions were as follows: 94°C for 3 minutes, followed by 50 cycles of 94°C for 30 seconds, 50-68°C (depending on the sample) for 30 seconds and 72°C for 30 seconds with a final step of 72°C for 10 minutes. Negative controls were run for every DNA extraction and PCR to control for potential contamination. Each first-round PCR contained a final volume of 13 µl consisting of 6.25 µl of DreamTaq Green PCR Master Mix 2x (Thermo Scientific), 1.25 µl of PDPF1 and PDPR2 (concentration of 0.02 µM), 3.25 µl of H_2O and 1 µl of genomic DNA. The nested PCRs with the second-round primers consisted of the same volumes, but 1 µl of first-round PCR product was used in place of 1 µl of genomic DNA.

PCR products were visualised via gel electrophoresis, following recommendations from Magdeldin (2012), and a 3% agarose gel was used as opposed to 1% for the DNA visualisation which is preferential for small fragment sizes. PCR products were purified using a GeneJET PCR Purification Kit (Thermo Scientific) and were directly sequenced with the second-round primers L91-115 and H402-27 by DBS Genomics (Durham University) using Applied Biosystems 3730 DNA analyser. Forward and reverse sequences were manually checked for errors, aligned and made into a consensus sequence in

Bioedit version 7.2.5 (Hall 2016). Consensus sequences were checked against a DNA sequence library using the NCBI Basic Local Alignment Search Tool (BLAST) to check for accuracy and contamination. Sequence data was of varied quality, as is often the case with museum specimens and degraded DNA. In general, the Aspinall Foundation (contemporary) samples yielded higher quality data than the PCM samples (Fig. 3.4).

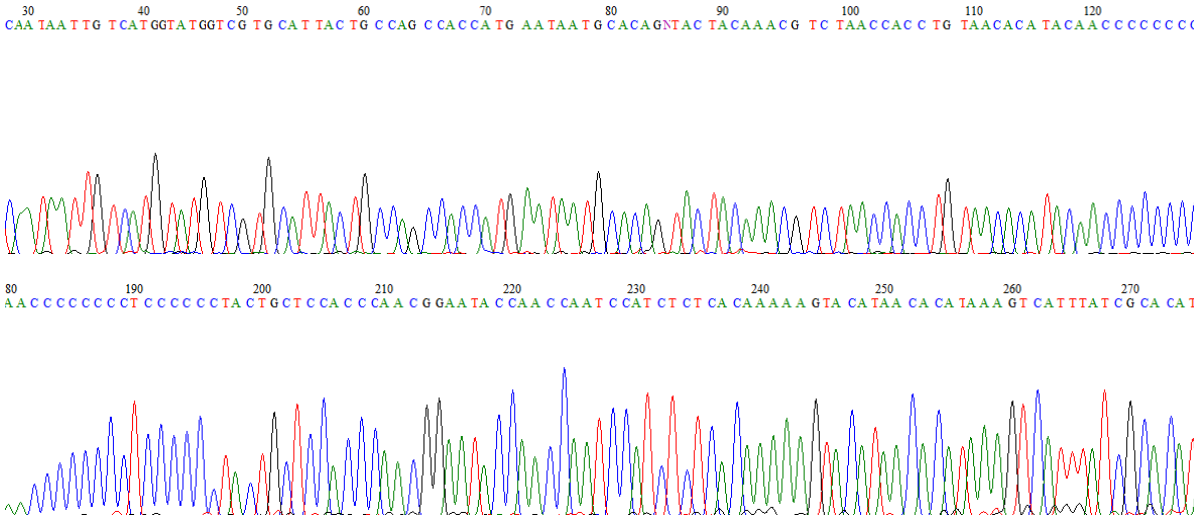


Figure 3.4 Chromatograph showing DNA sequence data for Aspinall Foundation gorilla named Jah; (top) forward sequence using second-round primer L91-115 and (bottom) reverse sequence using second-round primer H402-27.

Where DNA sequencing failed, modifications were made to the PCR protocol; increasing the annealing temperature improved the results significantly although the optimal temperature (68°C) was well outside the range of published protocols. A proportion of samples consistently produced poor quality data (primarily the PCM samples), and despite repeated attempts to improve the quality, they were removed from the analysis. Approximately 25% of samples were re-sequenced to ensure consistency of results. The final dataset used for downstream analysis consisted of a total of 256 DNA sequences of sufficient quality for analysis, including 59 individuals from the contemporary population and 51 individuals from the historic population (which included two of the three RCS specimens, including PA62 and PA63, who are known to be mother and offspring respectively, but excluding PA61 which was not possible to sequence). In addition, 146 HVI sequences were obtained from GenBank (Accession numbers are given in Appendix 3), which included eastern lowland gorillas, mountain gorillas, Cross River gorillas and captive and wild western lowland gorillas. One orangutan (*Pongo abelii*) DNA sequence was used as an outgroup (GenBank accession number AJ586558).

3.5.3. Sequence alignment and genetic diversity analyses

Multiple pairwise DNA sequence alignment of all specimens including the outgroup was performed in MEGA version 7.0.25 (Kumar *et al*. 2016). A 27 base-pair polymorphic C region was removed from all samples subsequent to alignment and prior to analysis, and the 5' and 3' ends of DNA sequences were trimmed to remove missing data from the multiple pairwise alignment. The removal of the poly-C region is common and is repeatedly reported in previous literature, e.g. Garner & Ryder (1996), Clifford *et al*. (2003), Jensen-Seaman *et al*. (2004), Anthony *et al*. (2007). This is due to the region being notoriously difficult to sequence using Sanger sequencing and normally results in the termination of the sequence at this point (Clifford *et al*. 2003). The removal of this region resulted in a total sequence length of 193 bp for each specimen.

For initial analyses, sequences were separated into 11 groups: the Aspinall sequences were treated as one group as were the PCM sequences, with the GenBank data set divided into nine groups. Eastern gorillas were separated into their two subspecies: eastern mountain gorillas (*Gorilla beringei beringei*) and eastern lowland gorillas (*Gorilla beringei graueri*). Sequences from captive individuals in American zoos were classed as one captive group, and the remaining sequences from wild populations were split into regional groups based on country of origin: Congo, Gabon, Central African Republic, Cameroon, Equatorial Guinea and Nigeria (which correspond to the Cross River gorillas). The 11 group names were abbreviated as follows: ASP (the western lowland captive gorillas from the Aspinall Foundation), PCM (the wild historic western lowland gorillas from the Powell-Cotton Museum), EM (wild contemporary eastern mountain gorillas), EL (wild contemporary eastern lowland gorillas), CAP (captive western lowland gorillas in US zoos), CAR (contemporary wild western lowland gorillas from the Central African Republic), CON (contemporary wild western lowland gorillas from the Republic of Congo), CAM (contemporary wild western gorillas from Cameroon), GAB (wild contemporary western lowland gorillas from Gabon), NIG (contemporary wild Cross River gorillas from Nigeria) and EQG (contemporary wild western lowland gorillas from Equatorial Guinea).

DnaSP version 6 (Rozas 2009) is a popular, exhaustive software for molecular population genetic analyses that measures levels of polymorphism within and between populations and divergence levels between species; in addition, DnaSP estimates variation and patterns of gene flow and recombination, and computes *P*-values for a multitude of neutrality tests based on coalescent simulations (Rozas 2009). DnaSP was used to perform the following analyses on the 11 predefined groups: number of polymorphic sites, number of parsimony informative sites, number of haplotypes and diversity, nucleotide diversity and K-sites in addition to mismatch analyses and the calculation of R_2 values.

The number of polymorphic (or segregating) sites is simply the number of variable positions in the sequence. Parsimony informative sites refers to sites that contain a minimum of two variations

of nucleotides with at least two occurring with a minimum frequency of two (Rozas 2009). A haplotype in general terms, is the variation of the markers observed in individuals, if two individuals match then they can be considered as sharing the same haplotype and are likely to be related (i.e. they share the same mtDNA type). A haplogroup refers to individuals which share the same or similar haplotypes and determines which clade they belong to (International Society of Genetic Genealogy 2020*)*. Genetic diversity can be calculated by haplotype diversity and nucleotide diversity. Haplotype diversity is also known as gene diversity and denotes the probability of differences between two randomly sampled alleles (Hamilton 2011). Nucleotide diversity is the mean number of nucleotide differences per site in pairwise comparisons of DNA sequences (Nei & Li 1979; Lawrence *et al.* 2015) and *K* is the average number of pairwise differences (Rozas 2009). Mismatch distribution is also known as the analysis of the distribution of pairwise differences and is used to infer demographic events (Ramos-Onsins & Rozas 2002). DnaSP employs the coalescent theory which is a population genetic model that focuses on neutral evolution, and it is considered the most powerful approach for interpreting sequence data (Ramos-Onsins & Rozas 2002). Many statistical tests have been developed to determine the reliability of mismatch analyses/population expansion processes (Harpending *et al*. 1993; Harpending 1994; Rogers *et al*. 1996) with Ramos-Onsins & Rozas (2002) concluding that the R_2 test is the most favourable for small sample sizes and *Fs* is more powerful for larger sample sizes. In addition, tests based on mismatch distribution are deemed very conservative due to the methods they employ (Ramos-Onsins & Rozas, 2002). Due to some of the gorilla groups used in this analysis containing very small samples sizes, the R_2 test was selected.

The concept of *F*-statistics was first introduced by Sewall Wright in 1921 and developed over the decades (Wright 1921; Neigel 1997). Wright defined the *F*-statistics (fixation index), as a set of correlation coefficients in regard to the correlation between gametes (Neigel 1997). Derived from *F* (the inbreeding coefficient), the fixation index contains the parameters F_{ST}, F_{IT} and F_{IS} and are related to levels of heterozygosity (Hamilton 2011). *F* values range from 0-1, if a population contains high levels of $F_{IS,}$ then it implies there is a substantial level of inbreeding (Hartl & Clark 1997; Hamilton 2011). *F*-statistics are estimated from genetic data representing multiple populations, where the hierarchical parameters F_{ST}, F_{IT} and F_{IS} enable genetic structure in populations to be summarised and compared (Weir & Cockerham 1984). Of the three measures which form the fixation index, the most commonly used measure of genetic differentiation is F_{ST} (Bird *et al*. 2011), and it is regarded as the most informative statistic of the three (Hartl & Clark 1997). F_{ST} makes comparisons between the least to most inclusive population hierarchy levels and measures all the effects of population substructure combined as a whole (Hartl & Clark 1997). A high F_{ST} value indicates substantial differentiation among populations (Neigel 1996; Shane 2005). To accurately interpret qualitative values of F_{ST}, Wright (1978) suggested the following: values of the range 0.0 to 0.05 may be considered as indicating little genetic

differentiation, from 0.05 to 0.15 indicating moderate genetic differentiation, from 0.15 to 0.25 indicating great genetic differentiation, and values of F_{ST} above 0.25 indicating very great genetic differentiation (Neigel 1996; Shane 2005). Pairwise F_{ST} is essentially a comparison of the F_{ST} values of the populations compared with each other (Shane 2005).

Developed by Excoffier *et al*. (1992), the analysis of molecular variance (AMOVA) is a method for detecting population differentiation of a single species based on a hierarchal model, and is one of the most favoured methods for estimating *F*-statistics (Bird *et al*. 2011; Meirmans 2012). It can estimate different types of *F*-statistics (e.g. F_{ST}, φ_{ST}, R_{ST}) and is also able to integrate additional hierarchical levels of population structure (Peakall *et al*. 1995; Meirmans & Liu 2018). In addition, it can be used to identify population clustering of genetic datasets (Dupanloup *et al*. 2002; Meirmans 2012; Meirmans & Liu 2018). For haplotypic data, AMOVA can be performed which estimates genetic structure using the allelic information based on a genetic distance matrix using Euclidean squared distances (Excoffier *et al*. 2005). Arlequin version 3.0 (Excoffier *et al*. 2005) was used to calculate pairwise F_{ST} and genetic variation within and among populations by means of AMOVA.

3.5.4. Phylogenetic analyses

Phylogenetic trees do not provide a definitive representation of the true phylogeny; they are estimates (hypotheses) which may or may not represent the ancestor-descendant evolutionary relationships (Hall 2011). However, a phylogenetic tree is a best estimate and the results are dependent on the assumptions of the model used to create the tree (Hillis *et al*. 1996; Hall 2011). The bootstrap method is the most commonly applied method to test the reliability of trees. It essentially pseudo-repeats data collecting and constructs a tree using the same method and parameters as depicted for the original tree, the new tree is then compared with the original tree and a bootstrap value (in %) is assigned to the clades or branches of the tree (Hillis *et al*. 1996; Hall 2011). Bootstrap replicates are typically run 100 to 2000 times, and the higher the bootstrap value, the more confident and reliable the tree (Hall 2011).

An initial phylogenetic tree was constructed in MEGA for the 11 defined groups, using the Neighbour Joining (NJ) method and 500 bootstraps, then edited in Figtree version 1.4.4 (Rambaut 2012). NJ is an algorithmic distance-based method that calculates and manipulates a genetic distance matrix of pairwise differences to produce a new matrix to construct the tree (Hall 2011). NJ is a minimum change method (as is parsimony) but it does not produce with absolute certainty the tree with the smallest overall distance, but it provides a computationally quick tree that can form the basis

for additional model-based analyses such as Maximum Likelihood and Bayesian methods (Hillis *et al*. 1996; Swofford *et al*. 2001; Hall 2011).

Unlike NJ, which is a distance-based method, Maximum Likelihood (MLH) is a character-based phylogenetic method which searches for a tree that maximises the likelihood/probability of observing the data. In most cases, a single tree is recovered (Swofford *et al*. 2001; Hall 2011). Bayesian analysis is a more recent variation of MLH; similar to MLH, it is a character-based method, and it searches for the best trees that are consistent with the alignment data and the evolutionary model selected (Ronquist & Huelsenbeck 2003). However, Bayesian approaches search for the best set of trees, rather than a single best tree (Huelsenbeck *et al*. 2001; Ronquist & Huelsenbeck 2003; Hall 2011). The advantage of this is that the MLH can become fixed on a single tree as it only considers a tree once, whereas Bayesian methods could sample the same tree repeatedly during the search (Hillis *et al*. 1996; Huelsenbeck *et al*. 2001; Hall 2011). Bayesian methods use the Markov Chain Monte Carlo (MCMC) method, which can be interpreted as independent tree searches that exchange information. This method allows the tree search to effectively 'jump around' the trees, sampling repeatedly a set of trees (Hillis *et al*. 1996; Hall 2011). The Bayesian approach can also become fixed on a tree in the same way that MLH methods can, however, the Bayesian analysis programs have an approach to overcome this by running several (usually four) independent searches in parallel starting with different trees (Hillis *et al*. 1996; Ronquist & Huelsenbeck 2003). When those searches have converged on the same set of trees, the final tree presented by the computer program is likely to represent the best tree (Ronquist & Huelsenbeck 2003; Hall 2011). In Bayesian phylogenetics, instead of bootstrap support, the distribution of trees is summarised by the majority-rule consensus tree annotated with the posterior probabilities (i.e. the updated probability of the branch occurring after taking into consideration the character-based data while building the tree) for each branch within the tree (Cranston & Rannala 2007).

Phylogenetic trees can be inferred from a variety of methods with each method consisting of its own assumptions (Weiss & von Haeseler 2003). To test which method best fits the sequence data in this study, a Maximum Composite Likelihood (MCL) method was applied to the dataset in MEGA. This method produces a substitution rate matrix. The MCL method recommends the evolutionary substitution model with the lowest BIC (Bayesian Information Criterion). The substitution model recommended by MEGA for the gorilla HVI dataset was the Hasegawa-Kishino-Yano (HKY) method with a discrete Gamma distribution (BIC value 8043.681869, Gamma value 0.37207761).

A MLH tree was constructed in MEGA using all sequences in their predefined 11 groups using the HKY+G method and with 1000 bootstraps. MrBayes version 3.2.5 (Ronquist & Huelsenbeck 2003) was used to construct a tree on the same dataset using 20 million MCMC steps, four independent runs (or MCMC chains), with a sampling interval every 1000 steps, and the first 25% of samples discarded

as burnin (i.e. the number of samples that will be discarded at the start of the run). Tracer version 1.7.1 (Rambaut *et al*. 2018) was used to check for convergence of the four chains of the Bayesian analyses. For further analyses of population diversity and structure, the PCM sequences were divided into their 3 regional subgroups (A, B and C) using ArcGIS version 9.3 to visualise the groupings as per the previous chapter.

Phylogenetic networks are an alternative method to phylogenetic trees for presenting the data, essentially, a network is a graph used to represent evolutionary relationships (Huson *et al*. 2010). Network version 5.0 (Bandelt *et al*. 1999) was used to create median-joining networks of the data. The greedy algorithm was used to search for the simplest haplotype network.

3.5.5. Detection and implications of Numt sequences

Discovered over 30 years ago, Numts are nuclear integrations of mitochondrial DNA, sections of the mitochondrial genome that have been interpolated into the nuclear genome (Jensen-Seaman *et al*. 2004; Gunbin *et al.* 2017) and are often inadvertently amplified during the PCR process, which can cause estimates of genetic diversity to be inflated, and lead to errors in phylogenetic inference (Song *et al*. 2008; Soto-Calderón *et al*. 2014). The gorilla genome has been considered to be the most prevalent in terms of numts, which has led to the reliability of mtDNA data in this species to be questioned (Soto-Calderón *et al*. 2014). Despite the problems associated with numts, numerous studies exist that have focused on the HVI region of mtDNA in gorillas and a few studies solely focusing on numts, e.g. Jensen-Seaman *et al*. (2004). For initial analyses, the whole of this dataset was considered (numts included) and further analyses were performed on the true mitochondrial sequences of which 30 were generated from this study.

Numts are not only extremely prevalent in the gorilla genome but appear to be more easily amplified in historic DNA such as museum specimens (Den Tex *et al.* 2010). To avoid numt amplification some methods include the isolation of the entire mtDNA genome, long range PCR, cloning, specific primers as opposed to universal primers and alternative sources of mtDNA such as tissue (e.g. muscle), however, numts can persist (Triant & DeWoody 2007; Den Tex *et al*. 2010). For historic samples some methods are not possible to employ. Using an alternative source such as muscle was not an option and due to DNA in historic samples being degraded, amplification of short fragments is more achievable and reliable, hence long-range PCR is not suitable, therefore specific nested primers were selected.

Numts can be identified by a variety of methods, initially they may present themselves as double bands on the electrophoresus PCR gels or as double peaks in the mtDNA sequence

chromatographs (Calvignac *et al.* 2011). Sequences can be compared to previously published data and those held on GENBANK. BLAST and LAST aid with identification (Tsuji *et al.* 2012) and various alignment tools such as CLUSTAL X can be used to identify true mtDNA or numt sequences (Clifford *et al.* 2004) as well as comparison of multiple fragments from the same invidual (Den Tex *et al.* 2010). The construction of phylogenetic trees using sequences with unknown and known origins for phylogenetic analysis is another method used for numt identification (Den Tex *et al.* 2010).

Numts in this research were identified initially by the presence of double bands in the electrophoresus gels and by double peaks present in the chromatographs. Sequences were compared to known numt sequences published in previous literature (e.g. Clifford *et al.* 2004) and compared to sequences held in GenBank. BLAST and the alignment tool (CLUSTAL X) were used to confirm/further identify numt sequences and finally, the construction and analysis of phylogenetic trees using known numt sequences was employed.

3.6. Results

3.6.1. Genetic diversity

The results of the 11 groups using all the DNA sequences revealed that the PCM sequences contained the greatest number of polymorphic and parsimony informative sites, 62 and 56 respectively. It also contained the greatest number of haplotypes, 35. Interestingly, the ASP sequences and the other captive individuals (CAP) were extremely similar, for example, the number of polymorphic sites was 51 for ASP and 48 for CAP, parsimony informative sites were 42 for ASP and 41 for CAP and the number of haplotypes was also similar at 20 for ASP and 21 for CAP. Haplotype diversity was greatest for the PCM group, followed by CAM (which consists of wild gorillas), then the ASP group. However, in terms of nucleotide diversity (π), whilst the PCM group showed the greatest result ($\pi = 0.09397$) here also, with EL and CAM following with similar levels ($\pi = 0.09015$ and $\pi = 0.09013$ respectively), whilst ASP and CAP showed lower levels of nucleotide diversity but were similar to each other, $\pi = 0.06868$ and $\pi = 0.06703$ respectively. The lowest nucleotide diversity was observed in the Cross River gorillas (NIG, $\pi = 0.00730$) and the eastern mountain gorillas (EM, $\pi = 0.00779$). A similar pattern can be seen for the number of *K*-sites with the PCM, CAM and EL groups showing the greatest number and NIG the lowest. Table 3.1 summarises the results for all the 11 groups.

Table 3.1 Summary of DNA polymorphism and genetic diversity of the 11 groups of gorillas

Population	No. of seq's	No. of poly. Sites	Parsimony Info. Sites	No. of haplotypes	Haplotype diversity	Nucleotide diversity (π)	Average no. of pairwise differences (*K*)
ASP	59	51	42	20	0.92928	0.06868	9.40970
PCM	51	62	56	35	0.97822	0.09397	12.87373
EM	6	2	2	3	0.80000	0.00779	1.06667
EL	22	45	39	11	0.91775	0.09015	12.35065
CAP	55	48	41	21	0.88822	0.06703	9.18316
CAR	8	6	2	5	0.78571	0.00860	1.17857
CON	5	26	3	15	0.70000	0.06277	8.60000
CAM	23	54	44	15	0.95257	0.09013	12.34783
GAB	17	39	27	10	0.79412	0.05303	7.26471
NIG	4	4	0	3	0.83333	0.00730	1.00000
EQG	5	4	0	2	0.40000	0.00876	1.20000

ASP: Aspinall Foundation, PCM: Powell-Cotton Museum, EM: Eastern mountain, EL: Eastern lowland, CAP: Captive (US), CAR: Central African Republic, CON: Congo, CAM: Cameroon, GAB: Gabon, NIG: Nigeria, EQG: Equatorial Guinea.

By pooling all the captive individuals (ASP and CAP) into one main 'captive' group and the remaining groups (excluding the eastern species) into one main 'wild' group of western gorillas, and then additionally separating the PCM (historic) group into its three regional groups (A, B and C), the genetic diversity analyses revealed that as a whole, the wild individuals possessed greater haplotype and nucleotide diversity than the captive individuals, but were still less diverse than the PCM (historic) group. Within the PCM (historic) group only, haplotype and nucleotide diversity were greatest in group C and lowest in group A, however, even the lowest results in group A were still greater than the in captive group. Table 3.2 summarises these results.

Table 3.2 Summary of DNA polymorphism and genetic diversity of captive and wild populations of gorillas

Population	No. of sequences	No. of poly. Sites	Parsimony Info. Sites	No. of haplotypes	Haplotype diversity	Nucleotide diversity (π)	Average no. of pairwise differences (*K*)
Captive	114	58	49	37	0.94877	0.07906	10.90964
Wild	113	60	53	56	0.97076	0.08847	12.20860
PCM (historic)	51	62	56	35	0.97822	0.09397	12.87373
PCM A	29	51	43	20	0.95567	0.08774	15.96798
PCM B	17	53	42	14	0.97059	0.09935	18.08088
PCM C	5	39	21	5	1.00000	0.11319	20.60000

The mismatch distribution (the distribution of pairwise differences) analyses were performed on each of the 11 groups, and the and R_2 and *P*-values were calculated for each. Figure 3.5.a-d shows the observed and expected frequencies of pairwise distances for the PCM, ASP, CAR and NIG groups. Table 3.3 summarises the R_2 and *P*-values for each of the 11 groups. The mismatch distribution plots for the PCM and ASP groups reveal a multimodal/ragged shape. The R_2 and *P*-values for each group were not significant with two exceptions: NIG and CAR. Likewise, it was these two groups whose mismatch distribution plots did not follow the same pattern as the other nine groups and produced unimodal distribution plots consistent with a model of sudden demographic population expansion.

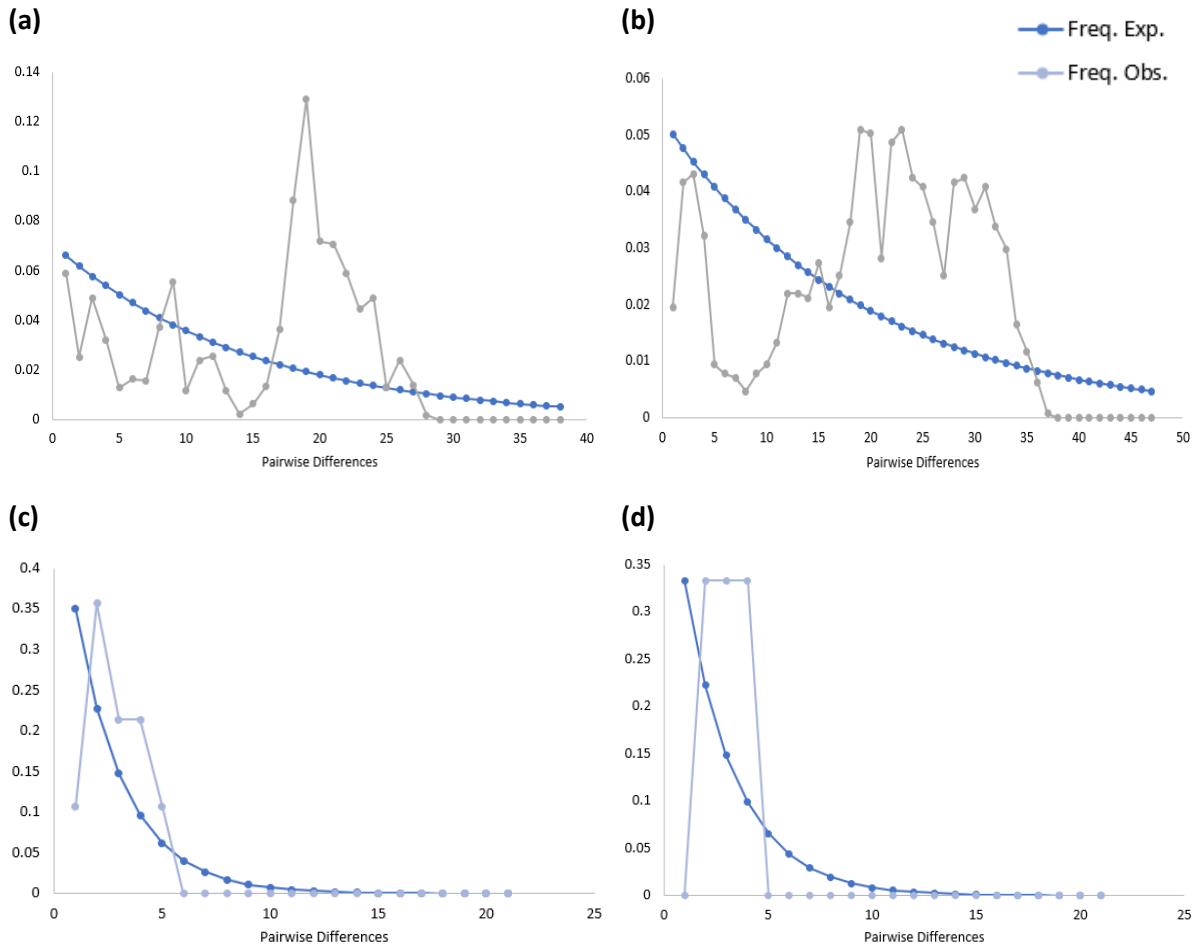


Figure 3.5 Mismatch distribution plots showing expected and observed frequencies of pairwise differences of (a) Aspinall Foundation (ASP), (b) Powell-Cotton Museum (PCM), (c) Central African Republic (CAR) and (d) Nigeria (NIG). ASP and PCM show multimodal distributions whereas CAR and NIG reveal unimodal distribution. The expected frequencies represent the expectations based on a neutral model of evolution for natural selection-genetic drift in a randomly reproducing theoretical population.

Table 3.3 Summary of R_2 and *P*-values for the 11 groups of gorillas

Population	R_2 Statistic	95% Confidence Interval Lower limit	Upper limit	*P*-Value
ASP	0.1350	0.05406	0.16214	0.87600
PCM	0.1501	0.05645	0.16274	0.93300
EM	0.2667	0.16667	0.37268	0.61348
EL	0.1906	0.07801	0.18780	0.98000
CAP	0.1036	0.05650	0.16706	0.53400
CAR	0.1098	0.12877	0.33072	**0.00413**
CON	0.2452	0.10672	0.35526	0.63828
CAM	0.1550	0.07511	0.19113	0.84800
GAB	0.1350	0.08590	0.20059	0.53900
NIG	0.1768	0.15635	0.43301	**0.04853**
EQG	0.4000	0.16330	0.40000	0.85493

** significant values shown in bold*
Acronyms are the same as in Table 3.1

The AMOVA conducted for the 11 groups showed that the percentage of variation among populations was 26.74% and within populations was 73.26%. Results for comparison between the captive group as a whole, and the wild population as a whole found 3.45% variation among populations and 96.55% within populations. Considering the ASP group and the PCM divided into its three regional subgroups (A, B and C), the AMOVA results showed 18.56% variation among populations and 81.44% within populations. Although the variation among populations was lower than within populations, the variation among populations was significant in all cases; 11 groups F_{ST} = 0.2674, $P < 0.001$; captive versus wild F_{ST} = 0.0345, $P < 0.001$; APS versus PCM subgroups F_{ST} = 0.1856, $P < 0.001$.

The pairwise F_{ST} results for the 11 groups showed significant *P*-value differences in almost all cases, with only a few pairwise comparisons showing non-significant (> 0.05) results: CON vs PCM, CAM vs PCM, CON vs CAP, CON vs GAB and CAM vs NIG. The point of interest here is that the NIG group (Cross River gorillas) did not produce a significant result with the CAM group, and given that they are considered a separate subspecies the expectation was that they would produce a significant result in all comparisons. Another interesting observation is that the ASP group was significantly different in all comparisons, the US captive group (CAP) revealed itself not to be significantly different from the CON group but was significantly different in all other comparisons. The F_{ST} pairwise differences using the distance method for the 11 groups analysis are shown in Table 3.4.

Table 3.4 Pairwise F_{ST} differences using the distance method for the 11 groups of gorillas

	ASP	PCM	EM	EL	CON	CAP	GAB	CAM	NIG	CAR	EQG
ASP	0.000										
PCM	**0.149**	0.000									
EM	**0.534**	**0.528**	0.000								
EL	**0.232**	**0.306**	**0.366**	0.000							
CON	**0.247**	0.088	**0.810**	**0.406**	0.000						
CAP	**0.261**	**0.107**	**0.651**	**0.451**	0.050	0.000					
GAB	**0.153**	**0.131**	**0.660**	**0.343**	0.005	**0.086**	0.000				
CAM	**0.208**	0.002	**0.575**	**0.337**	**0.136**	**0.138**	**0.189**	0.000			
NIG	**0.440**	**0.205**	**0.961**	**0.542**	**0.572**	**0.303**	**0.507**	0.111	0.000		
CAR	**0.393**	**0.312**	**0.953**	**0.523**	**0.459**	**0.283**	**0.436**	**0.380**	**0.896**	0.000	
EQG	**0.307**	**0.231**	**0.955**	**0.448**	**0.349**	**0.195**	**0.276**	**0.295**	**0.899**	**0.807**	0.000

**values that produced significant F_{ST} P-values are shown in bold*
Acronyms are the same as in Table 3.1

The F_{ST} pairwise differences using the distance method between captive and wild groups produced a value of 0.0345 and a significant *P*-value < 0.05. Population pairwise F_{ST}'s for the ASP and the three PCM subgroups (A, B and C) revealed significant *P*-values for ASP vs A, ASP vs B and A vs B. Sub-group C was not significantly different in any of the pairwise comparisons.

3.6.2. Phylogenetic analyses

The genetic distance matrix (Table 3.5) was used to create a neighbour-joining (NJ) tree in MEGA using the 11 groups dataset.

Table 3.5 Genetic distance matrix for the 11 groups of gorillas based on mitochondrial Hypervariable Region I (HVI) sequences

	ASP	PCM	EM	EL	CAP	CAR	CON	CAM	GAB	NIG	EQG
ASP	0.000										
PCM	0.018	0.000									
EM	0.096	0.104	0.000								
EL	0.023	0.036	0.047	0.000							
CAP	0.023	0.009	0.098	0.040	0.000						
CAR	0.049	0.044	0.133	0.068	0.029	0.000					
CON	0.024	0.010	0.092	0.040	0.000	0.023	0.000				
CAM	0.026	0.001	0.110	0.041	0.012	0.055	0.016	0.000			
GAB	0.013	0.017	0.086	0.029	0.006	0.043	0.000	0.022	0.000		
NIG	0.075	0.035	0.154	0.096	0.035	0.090	0.047	0.021	0.062	0.000	
EQG	0.037	0.034	0.116	0.045	0.021	0.038	0.020	0.040	0.026	0.074	0.000

Acronyms are the same as in Table 3.1

The NJ tree (Fig. 3.6) clearly shows the eastern species (EM and EL) and the Cross River gorillas (NIG) on distinctly separate branches. The EQG and CAR groups were also represented on different branches but with little bootstrap support. The PCM group was most closely related the CAM group which was not unexpected given the majority of the sequences from the PCM group were from gorillas in Cameroon, with only a small subset captured in the Congo, and no individuals from Gabon were included in the PCM group. The CAP group, interestingly, appeared to be more closely related to the CON group rather than the other captive group, ASP. The ASP group showed separation from the other regional groups including the other captive group. The GAB group also showed a degree of separation but not less so than the ASP group; it is prudent to note the low bootstrap support when making these inferences.

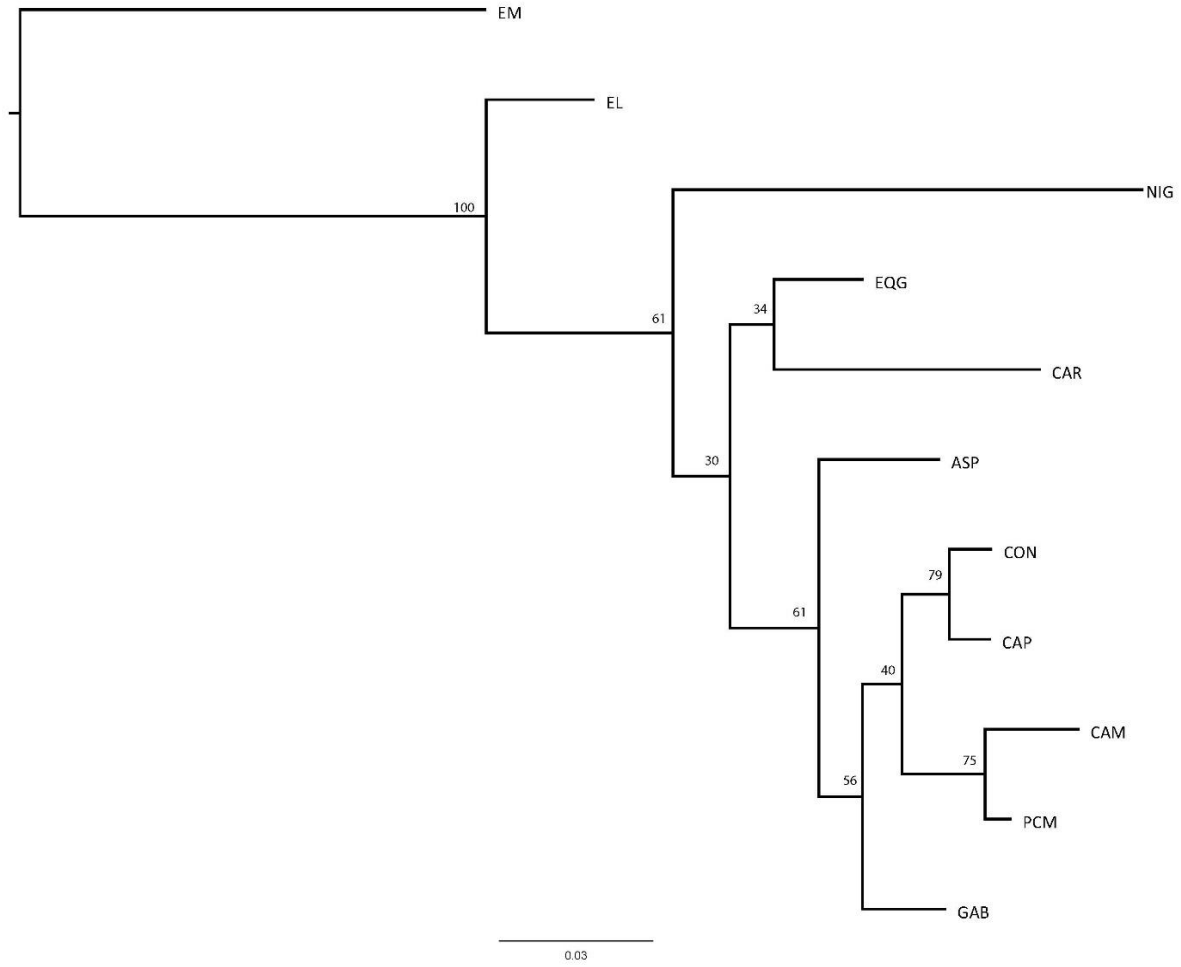
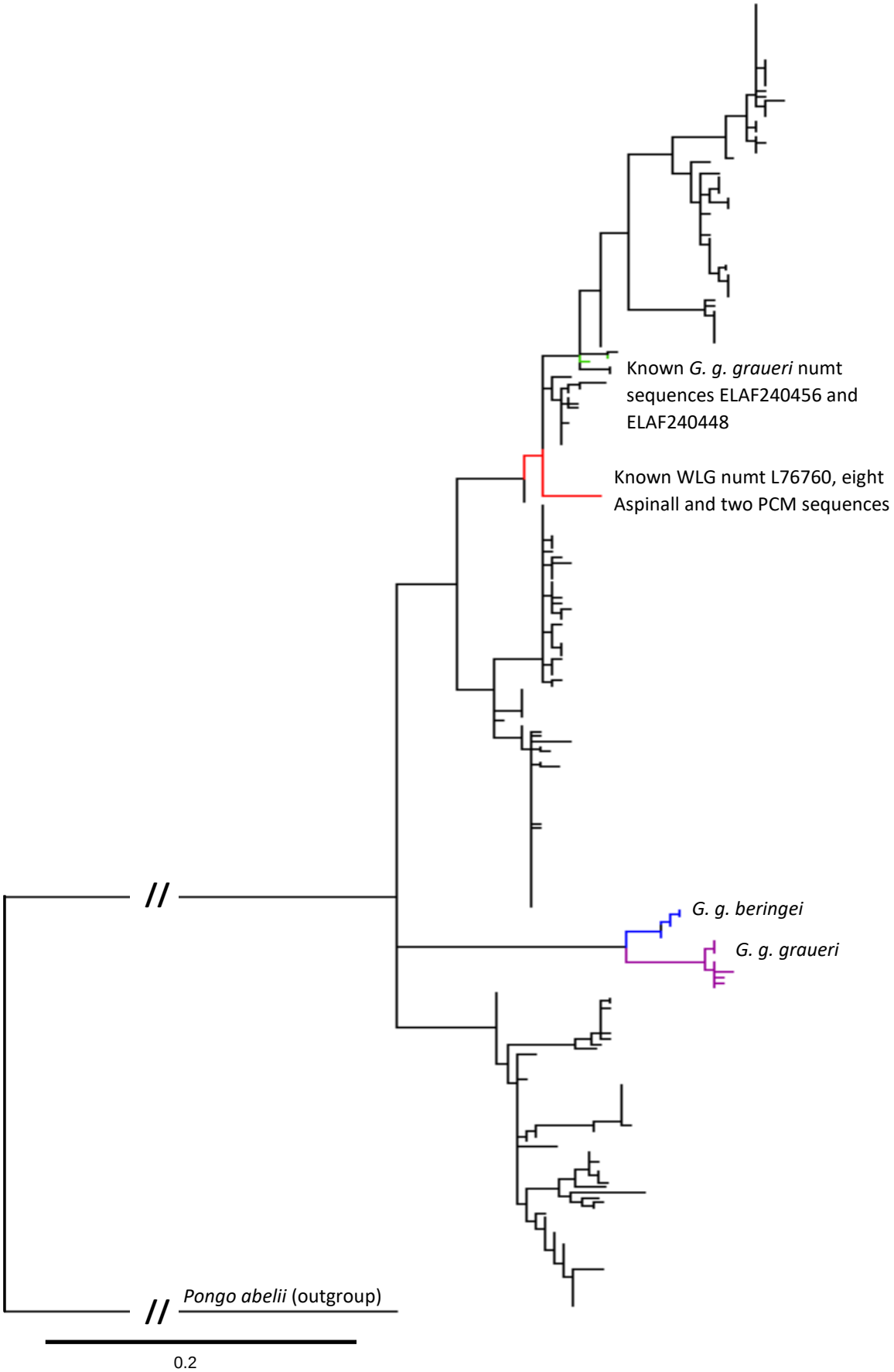


Figure 3.6 Neighbour Joining tree of the 11 groups of gorillas rooted with the eastern gorilla species; bootstrap values are shown on branches.

The MLH and Bayesian trees produced similar results (Fig. 3.7.a, b). Although the topology was similar, it was apparent that the inclusion of the numt sequences had a negative effect on the resolution of the phylogenetic trees. For example, on both the MHL and Bayesian trees, two of the eastern lowland gorilla sequences, identified in GenBank as numt sequences (accession numbers AF240448 and AF240456) did not cluster with the other eastern lowland sequences but clustered with other western lowland gorilla sequences, including Mer.139 (PCM) and Mayombe and Virginika Inf (ASP). GenBank numt sequence CAP L76760 clustered with many of the ASP gorillas, this phylogenetic representation on the tree indicates they formed an all-numt class. Other inconsistencies in the trees such as known related individuals appearing on different branches highlighted the unreliability and confirmed the presence of numts within the data generated from this study. The inclusion of numt sequences in the phylogenetic tree reconstructions was deliberate to ensure that all DNA sequences generated from this study could be accurately identified as numts or as true sequences.

(a)


Known G. g. graueri numt sequences ELAF240456 and ELAF240448
Known WLG numt L76760, eight Aspinall and two PCM sequences
G. g. beringei
G. g. graueri
Pongo abelii (outgroup)
0.2

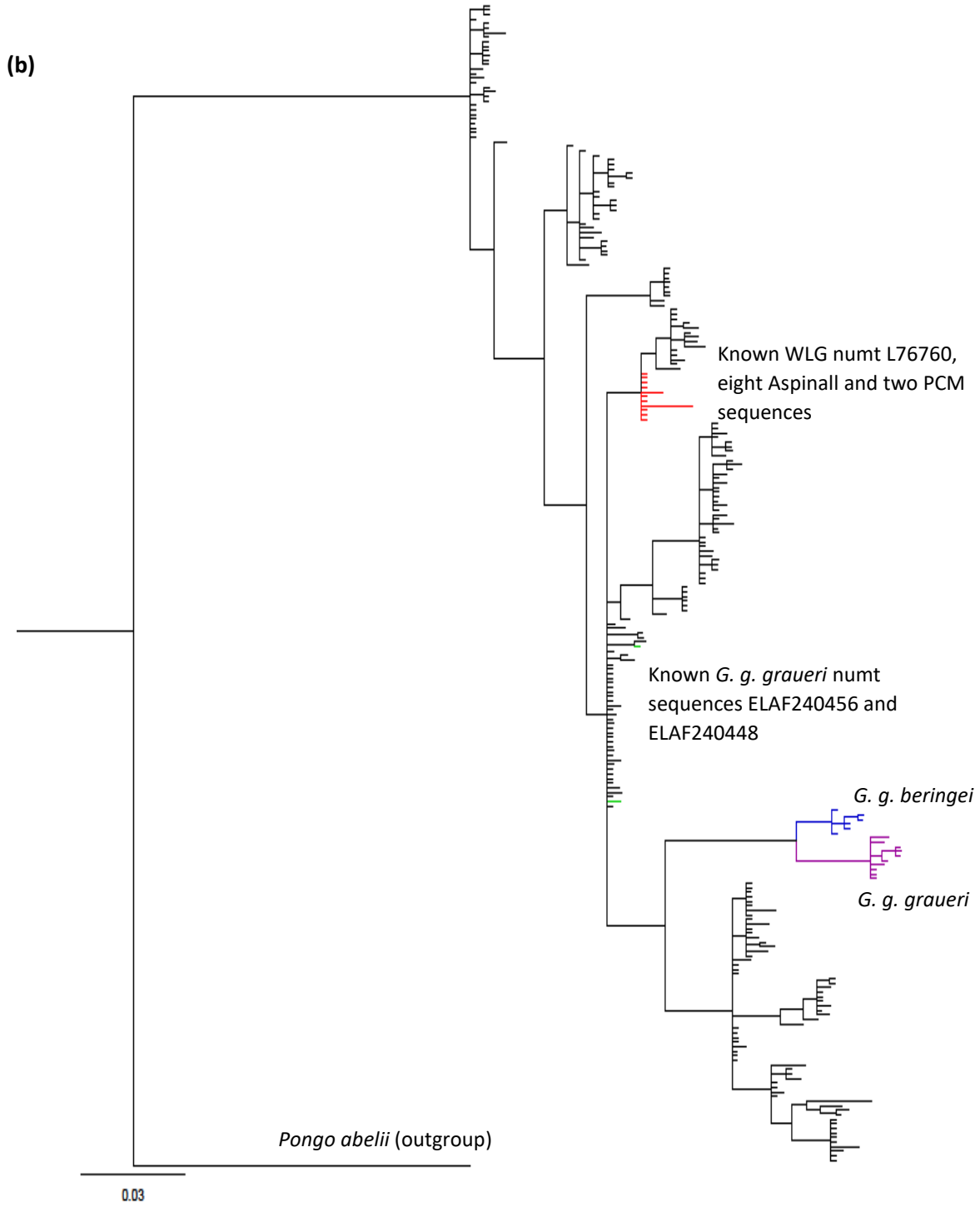
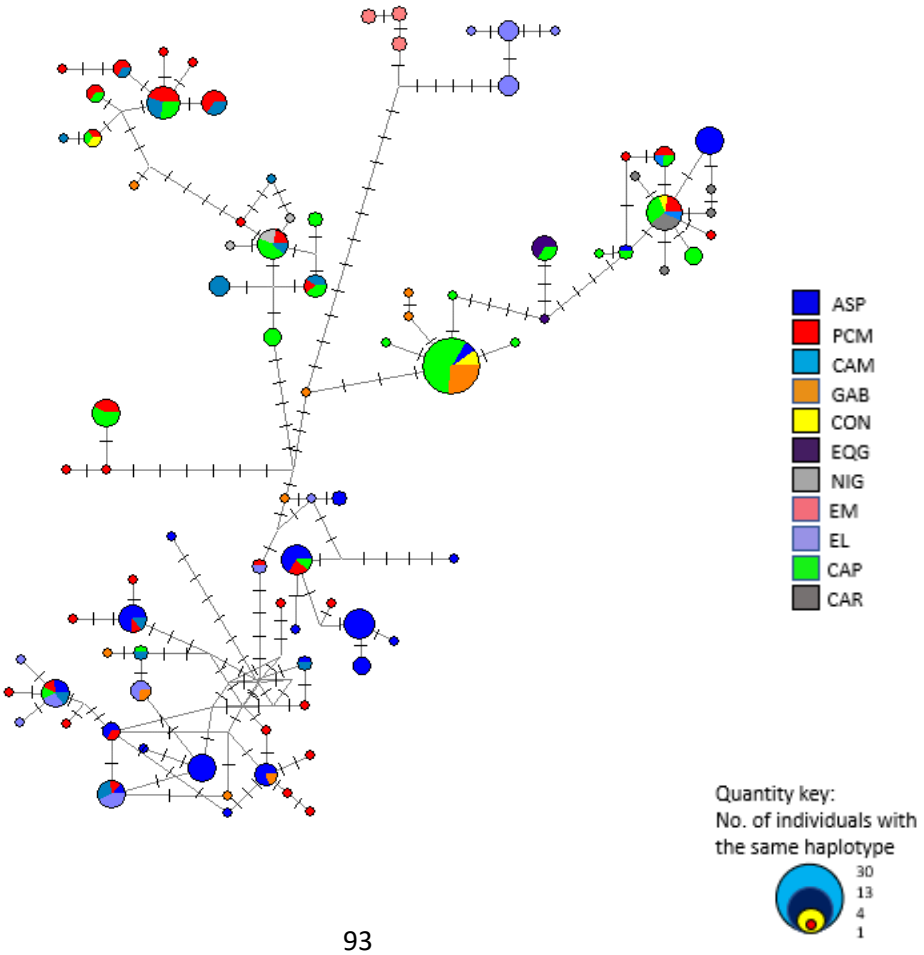


Figure 3.7 Phylogenetic trees for all sequences including nuclear inserts of mitochondrial DNA (numts), where (a) is the maximum likelihood tree and (b) is the Bayesian tree, both rooted with the outgroup. Coloured branches depict the issues of numt sequences which could lead to misinterpretation of phylogeny if undetected. The red branches depict the numt sequence CAP L76760 and includes eight of the ASP gorillas (Matibi, Bitanu, Jubi, Ujiji, Thirza, FouFou, Louna and Yene), and two of the PCM gorillas (Mer.387, FC.130). The green branches include two known eastern lowland numt sequences EL AF240456 and ELAF240448 that did not cluster with known eastern gorillas (*G. g. beringei* (blue branches) and *G. g. graueri* (purple branches). Both trees showed PCM Mer.138 clustered with EL240448. Known numt sequences were included in this data to assist with numt identification of data generated from this study.

Phylogenetic network analysis also revealed the problems associated with the inclusion of numt sequences in the data. Two networks were drawn that included all the data, with Figure 3.8.a representing the data for the 11 groups and Figure 3.8.b representing the same data but split into 'captive' or 'wild' populations. In both instances the distance of the eastern gorillas was observed depicting their separation from the western species, as was the distance between clusters in the western lowland gorillas. However, as in the case of the phylogenetic trees, eastern gorillas appeared amongst the western populations which signals inaccuracies with the data. Interestingly, the Cross River gorillas appear to share a haplogroup with both PCM and ASP sequences (Fig. 3.8.a). The phylogenetic network showed that the PCM sequences appeared more widely dispersed across groups with the ASP sequences predomominantly occupying one cluster. The captive vs wild phylogenetic network (Fig. 3.8.b) showed no clear groupings of captive and wild haplotypes. However, although these results provided an indication of haplotype diversity, they cannot be considered a reliable phylogenetic reconstruction due to the inclusion of numt sequences, but again ensured that all DNA sequences generated from this study could be accurately identified as numts or as true sequences.

(a)



(b)

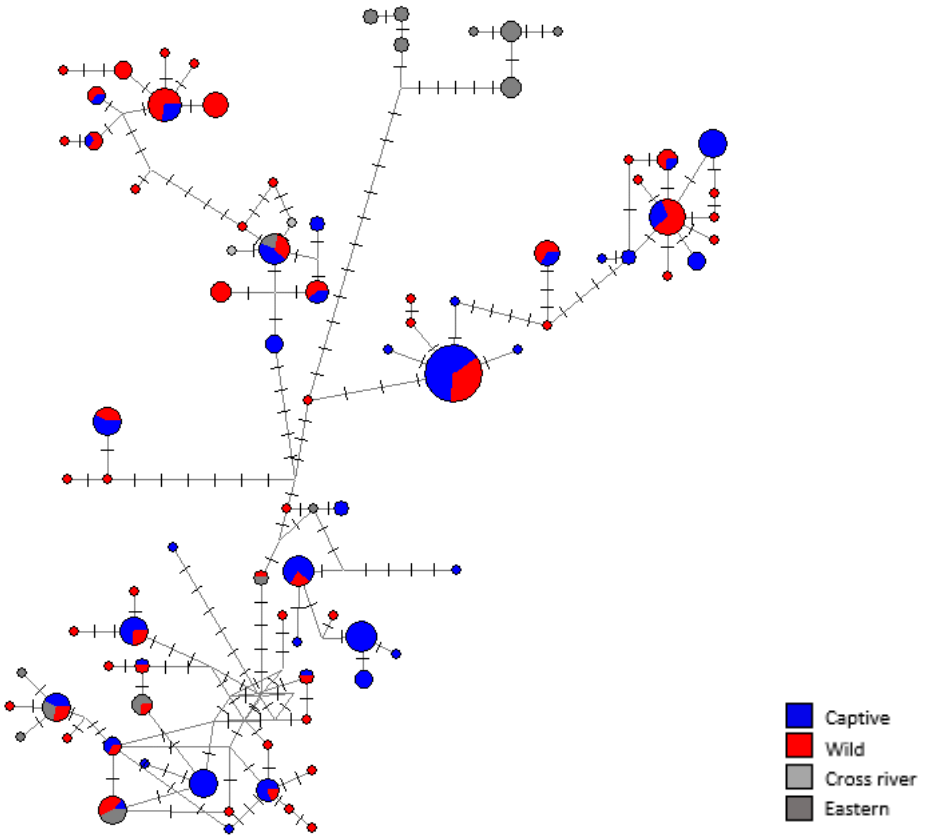


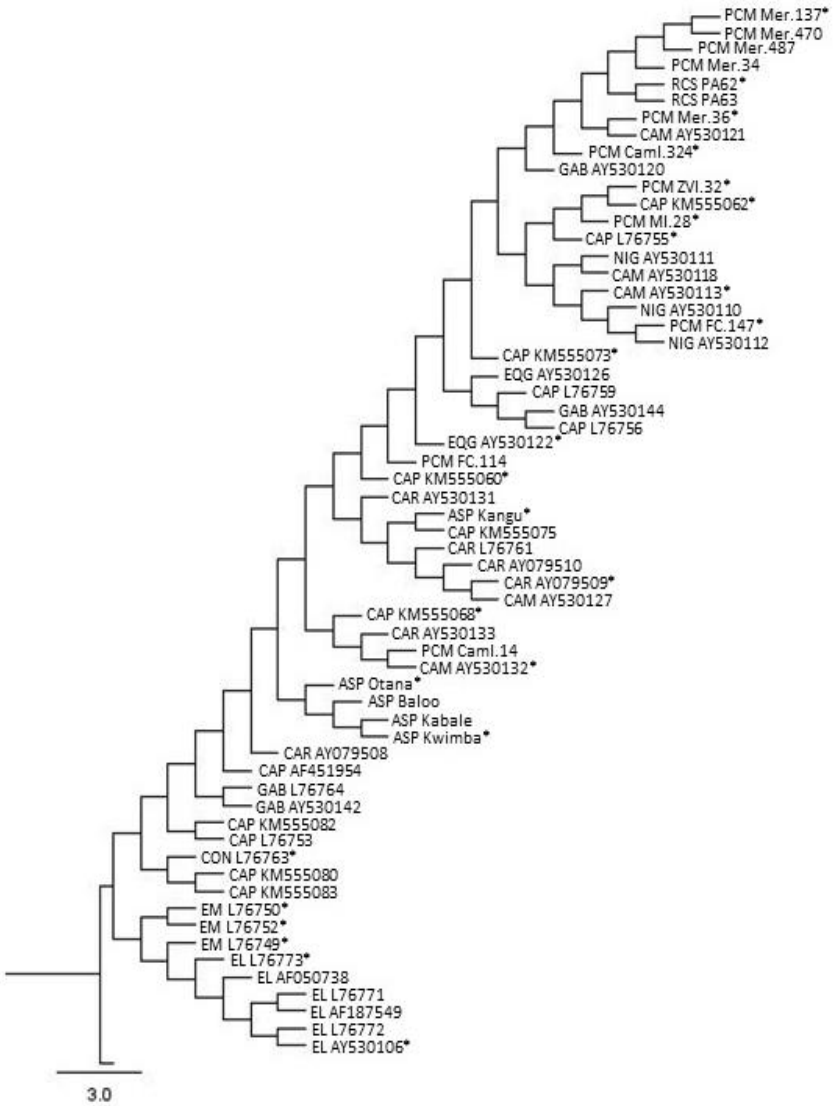
Figure 3.8 Phylogenetic networks of mitochondrial Hypervariable Region I (HVI) haplotypes of gorillas showing the haplotype distribution for: (a) the 11 groups and (b) captive versus wild populations.

3.6.3. True mtDNA sequences

Following the removal of the numt sequences, the final dataset included 149 mtDNA sequences, inclusive of the outgroup sequence, from which 30 were genuine mtDNA sequences of western lowland gorillas generated from this research. Only eight DNA sequences belonged to the ASP population, including the individuals Otana, Tamba, Baloo, Kabale, Kwimba (and her two infants 1 and 2), and Kangu. All except Kangu are F1 or F2 descendants of the mitochondrial founding individual Shamba (Cameroon origin). Kangus' mother is Sangha (Congo origin) and she is another mitochondrial founder of the ASP group. The other 22 genuine mitochondrial sequences generated from this study belonged to the PCM (and two from the RCS), and included the specimens CamI.324, CamI.325, Mer.34, Mer.58, Mer.136, Mer.137, Mer.264, Mer. 840, Mer.487, Mer.470, Mer.471, Mer.720, Mer.29, Mer.59, Mer.36, FC.147, MI.28, ZVI.32, CamI.14, FC.114, PA62 and PA63. Although the removal of numt sequences significantly reduced the dataset, it was a necessary requirement to ensure the reliability of the genetic analysis. The refined mtDNA data showed a total of 61 haplotypes

in the final dataset. Phylogenetic trees (MLH and Bayesian) of the haplotype data (Fig. 3.9.a, b) showed generally the same topology although there were observable dissimilarities. The eastern gorillas in both trees were clearly distinct from the western gorillas as would be expected. The Cross River gorillas clustered with the same western lowland gorillas in both trees which included GenBank sequences and PCM sequences. The novel C3 haplogroup revealed by Soto-Calderón *et al*. (2015) was clearly depicted in both trees. The EQG gorillas resided next to each other in the Bayesian tree but this was not the case for the MLH tree where they were shown to be further distanced from each other. The ASP gorillas clustered together in both trees with the exception of Kangu who was distanced further from the other seven ASP gorillas. In the Bayesian phylogenetic tree (Fig. 3.9.b), the ASP sequences belonged to haplogroup D2, whereas the 22 PCM/RCS sequences primarily resided in haplogroup C2 (17 sequences), three sequences belonged to haplogroup C1, and two sequences belonged to haplogroup D2.

(a)



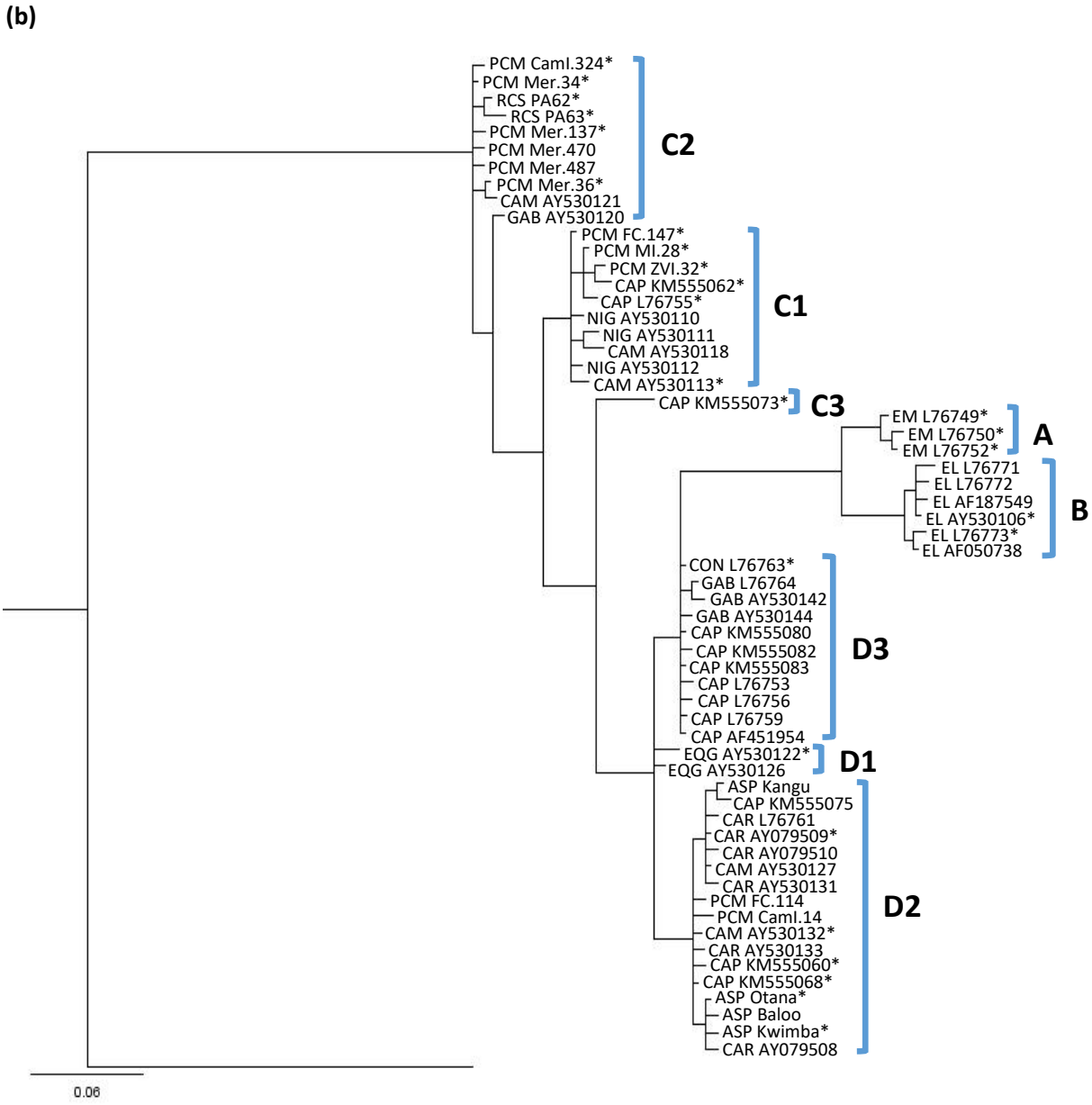


Figure 3.9 Phylogenetic haplotype trees of true mtDNA data showing their haplogroup distribution for (a) Maximum Likelihood (MLH) and (b) Bayesian methods. Both trees included 30 mtDNA sequences generated from this study as well as data from Clifford *et al*. (2004) and Soto-Calderón *et al*. (2015), and other additional sequences from GenBank. The trees were rooted at the outgroup. ***** indicates more than one individual shared the same haplotype; the list of haplotypes is given in Appendix 4.

The phylogenetic network (Fig. 3.10) reflected the findings from the Bayesian tree, with the eastern species being the most genetically distant from the western species and a split between the western species. All ASP individuals were present in haplogroup D2, and EQG and CAR groups also belonged to this haplogroup. Individuals from GAB group primarily resided in D3. Haplogroup C was

the most diverse with many of the groups present, including the Cross River gorillas which occupy C1, sharing the same haplotypes as individuals from the PCM, CAM and CAP groups.

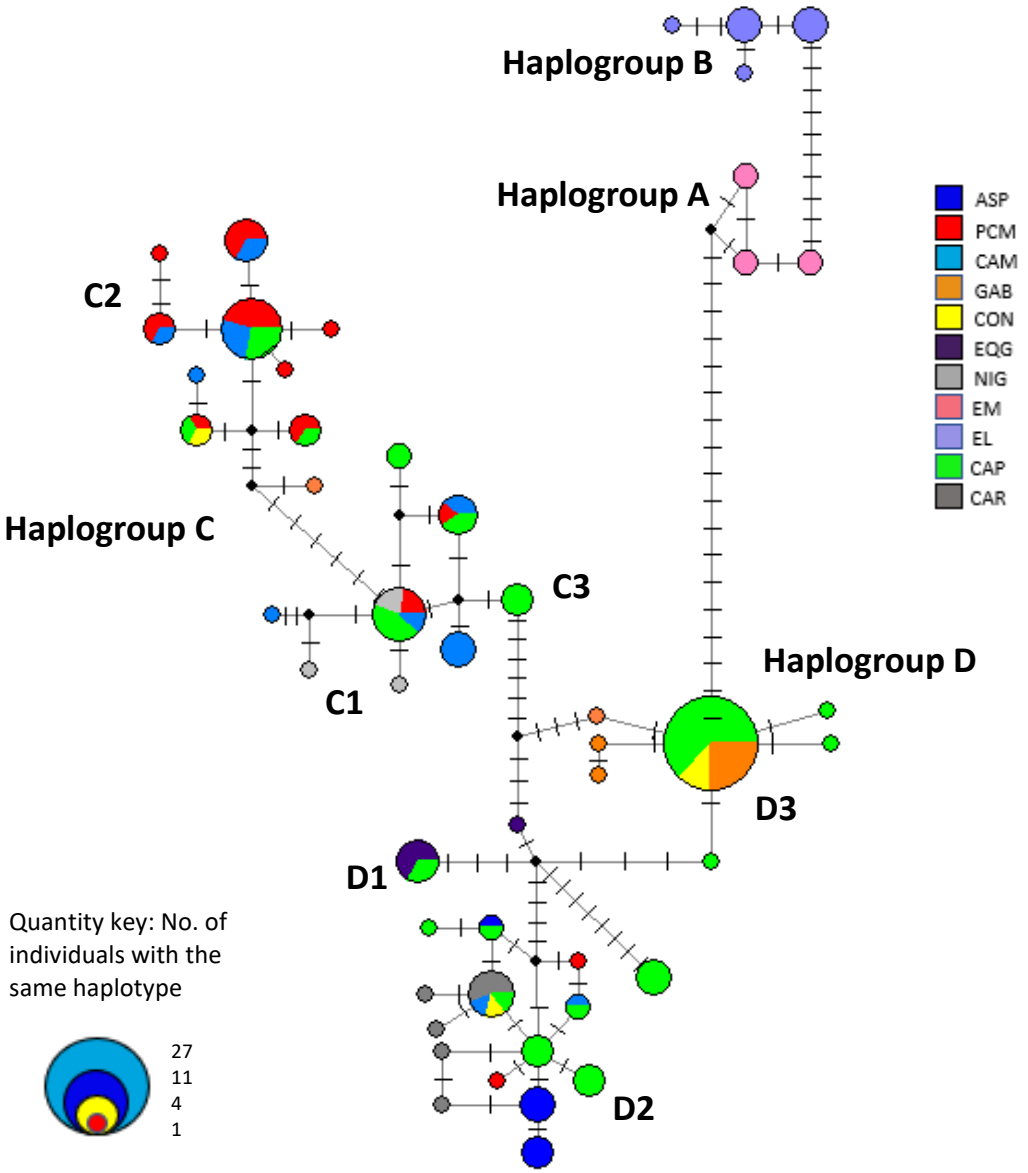


Figure 3.10 Network analysis showing the 11 groups of gorillas based on mitochondrial Hypervariable Region I (HVI) sequences with haplogroups A, B, C, D (and their respective subgroups C1, C2, C3, D1, D2, D3) identified based on Clifford *et al*. (2004).

Having established which haplogroup the PCM individuals belonged to and knowing their precise geographical locations, a map was generated to illustrate the geographical distribution of the haplotypes (Fig. 3.11). Individuals with haplogroup C2 were found in the regional subgroup A, and individuals from haplogroups C1 and D2 were found in subgroups regions B and C.

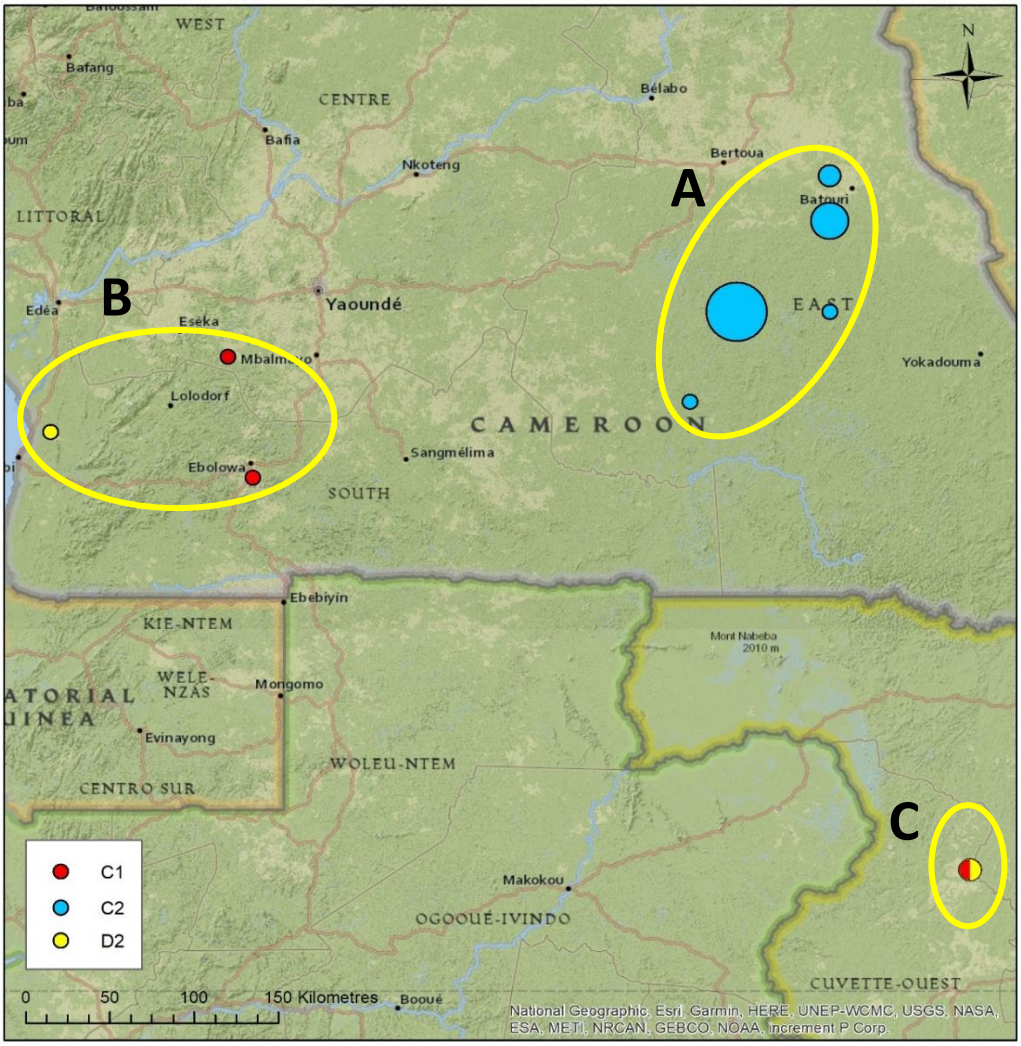


Figure 3.11 Map of the Powell-Cotton Museum (PCM) gorillas haplogroup locations. Circle size refers to the number of individuals at the same location (range 1-7). Haplogroup C2 was the only haplogroup present in region A, whereas haplogroups C1 and D2 were both present in regions B and C.

3.6.4. Genetic diversity of true mtDNA sequences

Genetic diversity analyses of the western lowland gorilla captive population and the wild population (i.e. the eastern species and Cross River subspecies removed from the data), revealed that nucleotide diversity was greatest in the wild population ($\pi = 0.07761$) compared with the captive population ($\pi = 0.05720$). The same was true for haplotype diversity, where the wild population had higher diversity

(h = 0.948) than the captive population (h = 0.902). Separating the wild population into historic (PCM) and contemporary wild groups, showed that nucleotide and haplotype diversity was greatest in the contemporary wild group (π = 0.07384, h = 0.930) compared to lower levels of diversity found in the historic (PCM) population (π = 0.04178, h = 0.922).

AMOVA for the captive and wild populations of western lowland gorillas showed that 9.77% variation was found among populations and 90.23% within populations. Pairwise F_{ST} was 0.09774 and produced a significant *P*-value ($P < 0.05$). Separating the wild population into historic and contemporary wild groups produced 19.92% variation among populations and 80.08% within populations, indicating genetic differentiation between historic and contemporary wild groups. Pairwise differences based on distance method among groups are summarised in Table 3.6.; pairwise F_{ST} values produced significant results for PCM vs Captive and PCM vs Wild (contemporary), but no significant difference was observed between the captive and contemporary wild populations.

Table 3.6 Pairwise differences among gorilla groups based on mitochondrial Hypervariable Region I (HVI) DNA sequences

	Captive	Wild (contemporary)	Powell-Cotton Museum (PCM)
Captive	0.00000		
Wild (contemporary)	0.01024	0.00000	
PCM	**0.39558**	**0.29517**	0.00000

** Significant P-values of F_{ST} values are shown in bold*

Mismatch analysis (Fig. 3.12.a) for the total 148 sequences showed a multimodal frequency distribution pattern indicative of population structure and complex evolutionary history, not consistent with a model of sudden demographic expansion. Mismatch distribution plots for the captive, contemporary wild and PCM (historic) populations (Fig. 3.12.b-d) revealed the same pattern. The R_2 values (Table 3.7) produced non-significant > 0.05 *P*-values in all three populations.

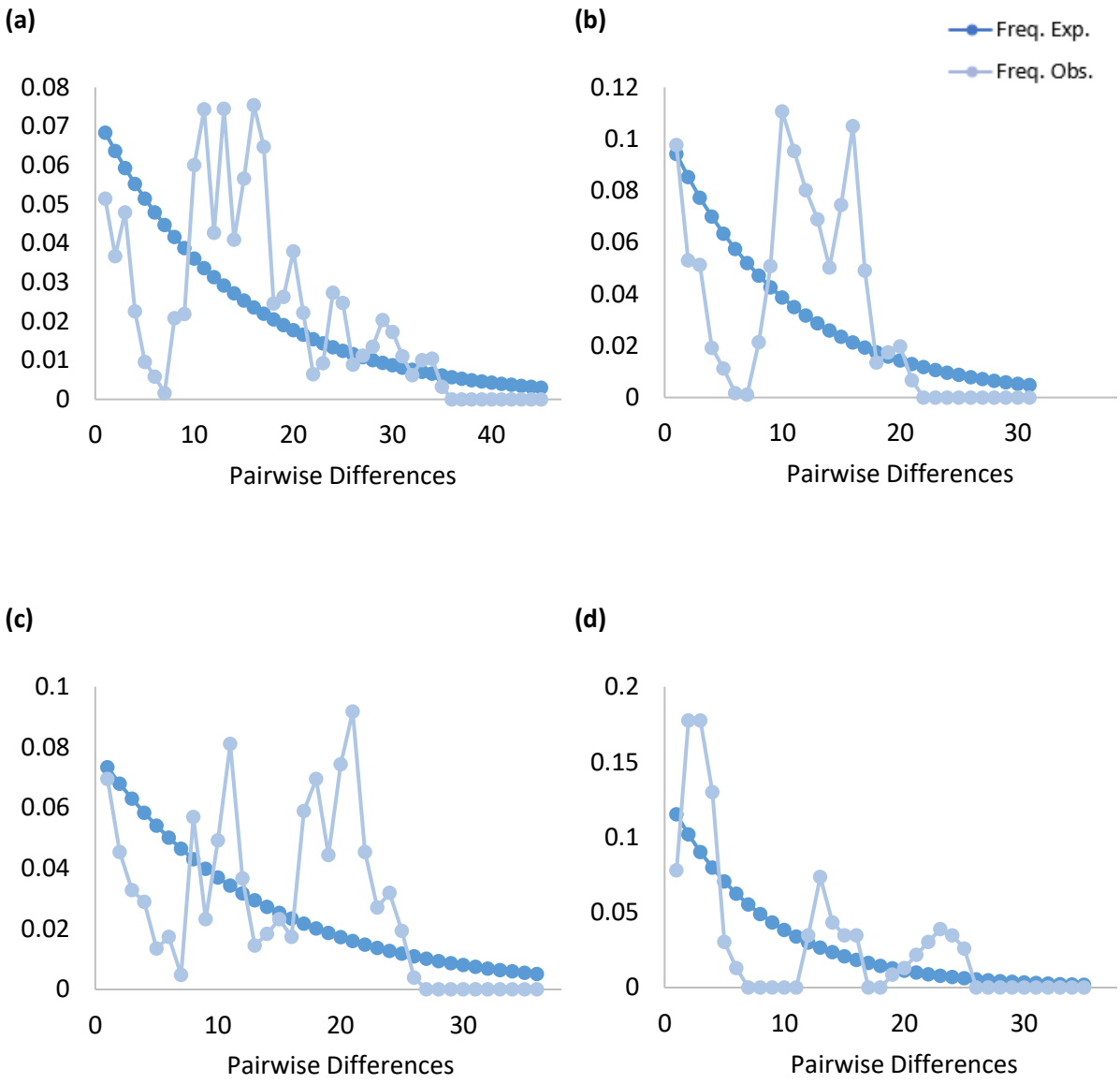


Figure 3.12 Mismatch distribution plots showing multimodal distribution for (a) the total dataset, (b) captive population, (c) contemporary wild population and (d) the Powell-Cotton Museum (PCM) groups of gorillas.

Table 3.7 R_2 statistics for gorilla groups based on mitochondrial Hypervariable Region I (HVI) DNA sequences

Pop.	R_2 Statistic	95% Confidence Interval Lower limit	Upper limit	*P*-Value
Captive	0.1356	0.05316	0.16056	0.87500
Wild (contemporary)	0.1515	0.05674	0.16805	0.93600
Powell-Cotton Museum (PCM)	0.1103	0.07495	0.19051	0.29900

Using the true mtDNA dataset and having established which haplogroups the 30 sequences generated from this study resided in, analyses of genetic diversity between haplogroups was possible. Analyses of haplogroups A, B, C, D and subgroups of haplogroups revealed that the greatest haplotype diversity and nucleotide diversity were present in haplogroup C (h = 0.936, π = 0.04908), with haplogroup D containing the second greatest nucleotide diversity (π = 0.03654) followed by haplogroup B and then A. Haplogroup B presented the second highest haplotype diversity (h = 0.867) followed by D then A (h = 0.837, 0.800 respectively). The haplogroup subgroups revealed that D2 contained the greatest haplotype diversity (h = 0.918) and the second greatest nucleotide diversity (π = 0.01285). C1 contained the second greatest haplotype diversity (h = 0.898) and the highest nucleotide diversity (π = 0.01419). Lowest haplotype diversity was observed in D1 (h = 0.286) which also had the third lowest nucleotide diversity (π = 0.00621). D3 displayed the lowest nucleotide diversity (π = 0.00272) and second lowest haplotype diversity (h = 0.384). C3 did not have adequate sample size for calculations to be made. Table 3.8 summarises the results and shows the number of polymorphic sites and parsimony informative sites.

Table 3.8 Genetic diversity of haplogroups and haplo-subgroups of gorillas based on mitochondrial Hypervariable Region I (HVI) DNA sequences

Haplogroup	No. of sequences	No. of poly. Sites	Parsimony Info. Sites	No. of haplotypes	Haplotype diversity	Nucleotide diversity (π)	Average no. of pairwise differences (K)
A	6	2	2	3	0.800	0.00583	1.067
B	10	6	2	6	0.867	0.00984	1.800
C	61	37	30	21	0.936	0.04908	8.933
C1	26	12	7	10	0.898	0.01419	2.597
C2	31	12	6	10	0.830	0.01025	1.875
C3	4	0	0	1	0.000	0.00000	0.000
D	71	23	16	24	0.837	0.03654	5.737
D1	7	4	0	2	0.286	0.00621	1.143
D2	31	12	7	14	0.918	0.01285	2.108
D3	33	7	1	8	0.384	0.00272	0.481

The AMOVA showed 69.73% variation among haplogroups, and 30.27% variation within haplogroups. Pairwise F_{ST} revealed the most genetic differentiation between haplogroups A and D (0.81176), followed by A and C (0.79017). The least differentiation was observed between C and D (0.61520), then A and B (0.63779). All pairwise F_{ST} values were significant ($P < 0.05$).

The AMOVA for the haplogroups separated into their subgroups showed 89.22% variation among subgroups and 10.78% within subgroups. Table 3.9. summarises the F_{ST} pairwise differences. The greatest genetic differentiation was found between haplogroup A and C3, followed by A and D3, and then C3 and D3. All pairwise F_{ST} comparisons were significant ($P < 0.05$).

Table 3.9 F_{ST} pairwise differences of haplogroups and haplo-subgroups of gorillas based on mitochondrial Hypervariable Region I (HVI) DNA sequences

	A	B	C1	C2	C3	D1	D2	D3
A	0.000							
B	0.638	0.000						
C1	0.936	0.888	0.000					
C2	0.954	0.915	0.831	0.000				
C3	0.981	0.834	0.859	0.917	0.000			
D1	0.962	0.817	0.870	0.913	0.952	0.000		
D2	0.921	0.861	0.860	0.896	0.870	0.728	0.000	
D3	0.978	0.923	0.923	0.940	0.971	0.913	0.833	0.000

Mismatch distributions for the haplogroups showed multi- or bimodal distributions in haplogroups C and D, indicative of population structure but not consistent with a model of sudden demographic expansion (Fig. 3.13). Haplogroups A and B however (eastern species), showed unimodal mismatch distributions consistent with a model of sudden demographic expansion (Fig 3.13.a, b).

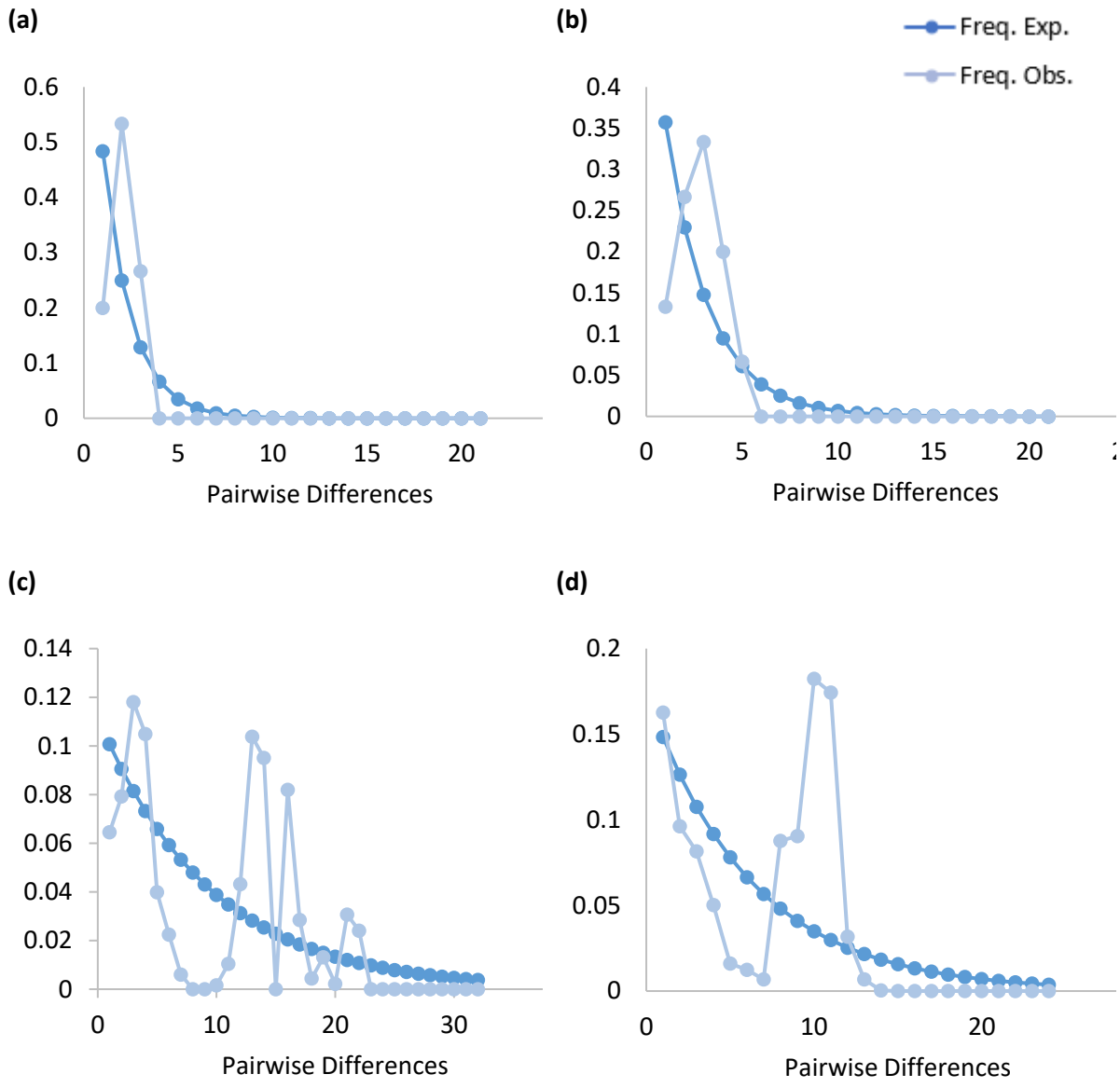


Figure 3.13 Mismatch distribution plots showing unimodal and multimodal distributions for the four haplogroups (A, B, C and D) of gorillas based on mitochondrial Hypervariable Region I (HVI) DNA sequences. (a) Haplogroup A is the eastern mountain gorillas, (b) haplogroup B is the eastern lowland gorillas, both showing unimodal distribution. (c) Haplogroup C with a multimodal distribution and (d) haplogroup D with a bimodal distribution, where both are the western species; haplogroup C contains the genetically differentiated Cross River gorillas.

Mismatch distributions of subgroups of haplogroups C and D showed bimodal distribution for C1 and D1, and unimodal distributions for C2, D2 and D3. However, haplogroup A and D1 should not be considered as they possess less than 10 sequences which is not considered appropriate for this analysis, likewise C3 was not able to produce results due to having only one haplotype. R_2 (Table 3.10.) showed non-significant *P*-values ($P > 0.05$), except for subgroup D3 with an $R_2 = 0.0557$ $P < 0.001$.

Table 3.10 R_2 values of haplogroups and subgroups of gorillas based on mitochondrial Hypervariable Region I (HVI) DNA sequences

Pop.	R_2 Statistic	95% Confidence Interval Lower limit	Upper limit	*P*-Value
A	0.2667	0.16667	0.37268	0.63015
B	0.1384	0.12288	0.30000	0.09480
C	0.1179	0.05120	0.16501	0.71300
C1	0.1005	0.07337	0.20508	0.25777
C2	0.0733	0.06497	0.20645	0.05506
C3	-	-	-	-
D	0.1211	0.04697	0.16315	0.76500
D1	0.3499	0.16410	0.34993	0.83017
D2	0.0804	0.06692	0.21290	0.10632
D3	0.0557	0.05871	0.25189	**< 0.001**

** significant p-values shown in bold*

3.7. Discussion

Previous studies on wild populations of gorillas focusing on the HVI mtDNA region have revealed four phylogenetic haplogroups: A, B, C and D, and within them further subdivisions in the haplogroups C (C1, C2 and C3) and D (D1, D2 and D3). Haplogroups A and B are restricted only to the eastern gorilla species, with haplogroup A restricted to the eastern mountain gorillas (*Gorilla beringei beringei)* and haplogroup B found within the eastern lowland subspecies (*G. b. graueri)*. Haplogroups C and D and their subgroups are found only in western lowland gorillas (Clifford *et al*. 2004). Only recently was the lineage C3 recognised, which was identified for the first time in US captive populations but was only considered previously to be present in wild populations (Soto-Calderón *et al*. 2015). In addition, several publications have acknowledged and investigated numt sequences and their impact on phylogenetic analyses and diversity measures amongst the gorilla species (Garner & Ryder 1996; Soto-Calderón *et al*. 2014).

Previous research that has investigated genetic diversity and phylogeny in contemporary wild gorilla populations has observed that nucleotide diversity is greatest in western lowland gorillas compared with eastern species (Garner & Ryder 1996; Jensen-Seaman & Kidd 2001; Clifford *et al*. 2004), and nucleotide diversity was equally distributed between haplogroups C and D (Clifford *et al*. 2004; Soto-Calderón *et al*. 2015). Within haplo-subgroups nucleotide diversity was equally distributed amongst haplogroup C, but this was not the case for the those in haplo-subgroups D. Haplo-subgroup D3, which is widely distributed geographically, displayed the lowest genetic variability (Soto-Calderón *et al*. 2015). The results from this study, including 30 new mtDNA sequences, support previous research, where nucleotide diversity was found to be greater in western gorillas than in eastern

gorillas, and nucleotide diversity was generally evenly distributed among haplo-subgroup C (C3 was the exception, possibly due to low sample size and containing only one haplotype). This research also found that nucleotide diversity in haplogroup D was not equally distributed among haplo-subgroups and found D3 to contain the least amount of genetic diversity (with the exception of C3).

In relation to captive populations, Soto-Calderón *et al*. (2015) was the first study to show the mitochondrial phylogeographic structure and diversity in the US captive populations in comparison to wild populations. Their study revealed all the haplogroups present in wild populations were also present in the US captive population, and they reported that haplotype diversity between captive and wild populations did not differ, but nucleotide diversity was significantly lower in the captive population. In addition, wild eastern gorillas displayed significantly lower nucleotide diversity in comparison to both wild and captive western gorillas. Their study reported that admixture between historic wild populations had not been investigated but may provide relevant information for conservation purposes of current captive populations and recommended that other captive populations would benefit from mitochondrial haplotype identification to assist with breeding programs. The research presented here attempted to address the recommendations by Soto-Calderón *et al*. (2015) by exploring haplotype diversity and distribution of a historic wild population (PCM) and a contemporary captive population (ASP). To date, no other genetic research has been conducted for the captive Aspinall Foundation gorillas, highlighting the relevance of this study on the conservation genetics of this species.

3.7.1. Data analysis with numt sequences

This research contained two datasets, one with and one without numts. With the numt sequences, the main observations were that wild populations contained greater nucleotide diversity compared with the total captive population ($\pi = 0.08847$ and $\pi = 0.07906$, respectively). This was also true for geographic subdivisions of the PCM group that still contained greater nucleotide diversity in all three regional groups in comparison to the all captive individuals, a pattern also consistent with haplotype diversity.

The two captive groups (ASP and CAP) were similar in terms of nucleotide diversity ($\pi = 0.06868$ and $\pi = 0.06703$, respectively). In some instances, these two groups displayed higher levels of nucleotide diversity than regional wild groups such as CAR ($\pi = 0.00860$) and GAB ($\pi = 0.05303$); however, the likely explanation for this is that the ASP and CAP groups contain individuals across all regions, so comparing them to a regionally defined group would likely yield these results. Of interest in this data, the PCM group (which primarily consists of DNA sequences of gorillas from Cameroon)

and the CAM group displayed similar levels of nucleotide and haplotype diversity (Table 3.1), and in pairwise F_{ST} comparisons (Table 3.4) they were one of a minority of comparisons that did not produce a significant result. This could be an indication that although population of western lowland gorillas have drastically decreased over the last few decades and genetic diversity levels are lower than in the past, they have not yet been significantly affected by genetic diversity loss, which is a positive indication for the long-term conservation of the subspecies.

The mismatch distribution plots for the PCM and ASP (Fig. 3.5.a, b) groups showed a multimodal/ragged shape, which is indicative of population structure and ragged distributions indicate that the lineage was widespread as reported in various other studies, e.g. Rogers & Harpending (1992); Ray *et al*. (2003); Joshi *et al*. (2013). This is in contrast to the CAR and NIG groups that showed unimodal distribution indicating their lineage was more restricted and/or less structured.

The phylogenetic network analysis (Fig. 3.8.b) showed captive and wild western lowland gorillas to be fairly equally distributed across clusters; however, Figure 3.8.a showed the ASP gorillas to mainly dominate one cluster, and the US captive gorillas (CAP) mainly dominating an adjacent cluster, although CAP was formed from a geographically more widespread origin than the ASP group, whilst the PCM group remains fairly widely distributed amongst the network. The MLH and Bayesian trees (Fig. 3.7.a, b) broadly reproduced the results expected in comparison to other research (e.g. Clifford *et al*. 2004), but also highlighted the errors and inconsistencies that occur due to the inclusion of numt sequences, where for example, known related individuals in the ASP group appeared on different branches on the tree and clustered with known numt sequences as identified in previous literature. The phylogenetic analyses indicated that the results obtained using the first dataset could not be considered reliable, although there were some interesting observations, and that numt sequences are known to inflate genetic diversity measures (Garner & Ryder 1996), which means caution must be given to these results.

3.7.2. True mtDNA data and phylogenetic relationships

Although the removal of the numt sequences significantly reduced the amount of sequences for analysis, it was a necessary step to increase the reliability of the results. A total of 149 sequences remained (inclusive of an outgroup) of which 8 belonged to the ASP group and 22 to the PCM, giving a total of 30 true mtDNA sequences generated from this research. This result is disappointing particularly for the ASP group, however, a few observations can be made and given the absence of any genetic research being present on the ASP group, any contribution to the current knowledge of gorilla mtDNA diversity and structure is considered scientific progress.

The phylogenetic trees constructed using the MLH and Bayesian methods did complement those produced in previous literature, e.g. Clifford *et al*. (2004). As this study used sequences from previous literature the replication of similar trees is a strong indication of the reliability of the results. However, this study found that the Bayesian method was superior to the MLH method in recovering a similar tree topology as in previous studies. Although the MLH tree (Fig. 3.9.a) was similar in topology to that of Clifford *et al*. (2004), there were notable dissimilarities: the EQG group was represented marginally different. However, the Bayesian tree (Fig. 3.9.b) reflected previous results more similarly and correctly placed all sequences. The difference in the MLH tree was likely due the limitations as explained above, where MLH methods can become 'stuck' on a single tree and cease to continue to search for other potential trees, whereas the Bayesian approach using MCMC and independent runs searches phylogenetic trees in parallel and continually relocates its search until all runs converge and a single consensus tree. It is likely that the 30 sequences generated from this study are indeed real mtDNA sequences as not only did they match genuine (non-numt) sequences via BLAST, they were each placed in a haplogroup as defined by the previous literature. From this analysis, it appears that the eight ASP gorillas belonged to haplogroup D2 and the 22 PCM specimens primarily belonged to haplogroup C2, although there are individuals within haplogroup C1 and D2. Phylogenetic network reconstruction (Fig. 3.10) showed the separation of the eastern species and a divide in the western species into the C and D haplogroups (with Cross River gorillas present in haplo-subgroup C1).

3.7.3. Genetic diversity and structure of true mtDNA data

In general, the results found here were consistent with previous studies but there were some differences and interesting findings to be highlighted. When considering the captive and wild populations as a whole (i.e. contemporary and historic wild pooled), the results complemented previous research with the wild population containing greater diversity than the captive population. Nucleotide diversity was greatest in the wild population compared with the captive population. However, when the wild population was separated into contemporary and historic, the results were not as expected; nucleotide diversity was greatest in the contemporary wild population, followed by the captive population, and the lowest diversity was found in the historic (PCM) population. Pairwise F_{ST} of the captive versus wild populations revealed that there was a significance difference between the two populations, however, pairwise comparisons between the captive and contemporary wild did not yield a significant difference, but in both cases, the PCM vs captive and PCM vs contemporary wild, there was a significant difference. An explanation for this is that the 22 PCM specimens were primarily from Cameroon (with only two from the Congo), whereas the captive and contemporary

wild populations were represented by gorillas from a wider geographical distribution including Gabon, Equatorial Guinea and the Central African Republic, thus the PCM population could be considered as representing one region. The PCM and CAM groups were one of a minority of pairwise F_{ST} comparisons which did not show significant genetic differentiation. This is encouraging for the captive population, because although genetic diversity was less than in the contemporary wild population, the genetic differentiation was not significant, indicating that the captive population can be considered genetically 'healthy' in comparison to its wild counterparts; this was also the conclusion of previous research for the captive US population (Soto-Calderón *et al*. 2015). Despite the fact that captive populations are considered to be genetically isolated, similar to fragmented or island populations (Baas *et al.* 2018), and introductions of wild gorillas to the captive population ceased in the 1970s (Endangered Species Act 1976; Hemley 1994; Nsubuga *et al*. 2010; Huxley 2013), the effects of this have not yet yielded a genetic diversity concern. Mace (1988), who studied the demographic and genetic status of captive gorilla populations, found that with appropriate management, it would be possible to preserve 90% of the current genetic diversity in the captive populations of Europe and North America for the next 200 years. It would appear that the captive population is being managed appropriately to preserve genetic diversity. Analyses of haplogroups A and subgroups of haplogroups revealed that nucleotide diversity was greatest in haplogroup C, followed by haplogroups D, B and then A. For the haplo-subgroups, C1 contained the greatest nucleotide diversity followed by D2, C2 and C1, and D3 with the lowest nucleotide diversity of all haplo-subgroups. These results were comparable with those found by Clifford *et al*. (2004).

All pairwise F_{ST} scores produced significant *P*-values ($P < 0.05$) confirming that genetic structure does exist regionally within the western lowland gorilla subspecies. The four haplogroups A, B, C and D for the contemporary wild population, did generally complement the four demes proposed by Groves (1967, 1970) regarding morphological variation which was investigated in the previous chapter (geometric morphometrics) and have been discussed and supported in other studies such as Clifford *et al*. (2004).

The mismatch distribution plots for the total true mtDNA dataset of western gorillas which included the combined captive population (ASP and US gorillas), the contemporary wild population and the PCM population all showed multimodal distribution (Fig. 3.12), again, reflecting widespread lineage and population structure. The mismatch distribution plots of haplogroups (Fig. 3.13) showed unimodal distribution for the eastern gorillas (Fig 3.13.a, b), whereas multi/bimodal distribution was observed for haplogroups C and D (Fig. 3.13.c, d). Haplogroup C was found to be the most genetically diverse haplogroup in terms of nucleotide diversity and the mismatch distribution plots support this showing haplogroup C with the most ragged/multimodal distribution and haplogroup D, with bimodal distribution, supporting widespread lineage and structure within the population.

3.7.4. Captive and historic populations

The US captive gorillas were widely distributed amongst the haplogroups and their subsequent subgroups, and this is encouraging for conservation purposes in terms of genetic diversity as all the haplotypes are present in the captive population, as reported by Soto-Calderón *et al*. (2015). The majority of the US western lowland captive gorillas belong to D3, which as mentioned earlier, is the most widespread haplo-subgroup but the least genetically diverse. Having ascertained that the eight ASP gorillas belong to haplo-subgroup D2 is potentially important for future breeding efforts. In addition, any of the females identified as being D2 that go on to reproduce will have D2 offspring. Although Djanghous' sequence was of too poor quality for analysis, Djanghou is half-brother to Kangu on the maternal side (Sangha); therefore, he is predicted to belong to the D2 haplogroup as well. The D2 haplogroup is not as geographically widespread as the D3 haplogroup but it is present in the Central African Republic, Congo, Cameroon and Gabon. The seven ASP individuals who are descendants of Shamba are of Cameroonian origins, and Kangu (who is a descendant of Sangha) is of Congolese origin; both of these regions have the D2 haplogroup present. Kangu (and three of his half-brothers from the paternal side) were released in the Republic of Congo in 2017 (T. King pers. comm, 2018). Although current breeding programs are managed by kinship (the international studbook), having knowledge of individual gorillas' genetic diversity is of benefit to maintain haplogroup diversity and promote genetic diversity of the captive population (Soto-Calderón *et al*. 2015).

The 22 PCM specimens occupied three of the previously identified haplogroups, C1, C2 and D2. Having mapped each individual PCM gorilla revealed some notable observations:

1) Several PCM gorillas belonged to region A/haplogroup C2 (which included specimens Mer.136, Mer.137 Mer.264, Mer. 840, Mer.487, Mer.470, CamI.324, CamI.325, Mer. 58, Mer.34, Mer.471, Mer.720, Mer.29, Mer.59, Mer.36, RCSPA62 and RCSPA63), which agreed geographically with previous haplogroup data of contemporary populations from Soto-Calderón *et al*. (2015).
2) Although the RCS specimens (PA62 and PA63, mother and infant, respectively) do not have known geographical locations, they both produced identical sequences to each other and to Mer.58 from the PCM collection, and also to KF029427 from GenBank. Mer.58 was hunted from the Batouri and Lomie regions in Cameroon. Given the exact matching of the sequences, the volume of sequences acquired from this haplogroup/area and previous research supporting the distribution of the C2 haplogroup in this region, it would be fair to assume that the RCS individuals could have come from the Batouri and Lomie regions in Cameroon, although this cannot be considered truly reliable as the C2 haplogroup does cover other geographical areas.

3) There were other PCM specimens for which the geographic location agreed with the published contemporary haplogroup geographic distribution data from Soto-Calderón *et al*. (2015): MI.28 and ZVI.32 were haplogroups C1 specimens in region B; FC.114 belonged to haplogroup D2 in region C; FC.147 belonged to haplogroup C1 also in region C; and CamI.14 was a D2 haplogroup individual located in region B. The C1 haplogroup has not been reported to be found that far south, and the D2 haplogroup has not been reported in the region C area, nor has it been reported that far West (region B) where typically the C1 haplogroup dominates.

Interestingly, Clifford *et al*. (2004) reported that their PCM specimen labelled CamI, was found to belong to haplogroup D2 and was in the south-west of Cameroon. Unfortunately, the exact specimen number was not given so it is not possible to determine whether this was specimen CamI.14. However, if it was the same specimen, then two independent researchers have found the same result. If their specimen was not CamI.14, then this study has found an additional specimen which also appears to contradict the proposed haplogroup distributions, plus an additional two specimens (FC.114 and FC.147) which also contradict previous distributions. Clifford *et al*. (2004) noted that the unexpected placement of the two museum specimens used in their research raised concerns regarding the reliability of using museum specimens for phylogeographic analysis of current populations. This study suggests the opposite, museum specimens can be of use for examining phylogeographic distributions, indeed, those haplogroup distributions (C1 and D2) may not be present in the current western lowland gorilla populations, but the evidence here suggests they were 100 years ago. These results also suggest that haplogroups may have been more widely distributed historically than they are at present. Soto-Calderón *et al*. (2015) noted that admixture between wild historic lineages has not yet been evaluated but could provide important information for conservation and breeding/release programs. These findings support the need for further investigations into historic wild populations and haplogroup distribution.

Soto-Calderón *et al*. (2015) also recommended that to prevent hybridisation and preserve phylogeographic structure in current populations of wild gorillas, captive-bred individuals should only be released to areas where their haplogroups are present. Mentioned previously was the introduction of Kangu (a D2 haplogroup individual) and his three half-brothers to the Republic of Congo in 2017. The Lesio Louna Wildlife Reserve was the site of the introductions which is much further south than the current haplotype distribution data available; however, this research has shown that the D2 haplogroup is likely to have had a wider geographical distribution in the past which did extend further south than the current distribution shows. Therefore, it is not unlikely that Kangu at least may have been released back to the former haplogroup D2 distribution.

The aims of this chapter were to investigate regional genetic diversity of the mitochondrial hyper-variable region 1 in western lowland gorillas of a contemporary captive and historic wild population, determine the haplogroups of individuals from those populations and compare the genetic diversity of those populations to each other, and to other wild populations as well as the US captive population based on the research of Soto-Calderón *et al*. (2015). Additionally, this study sought to identify the presence of numts from true mtDNA sequences to avoid misinterpretation of phylogenetic trees/ networks and avoid inflated genetic diversity results. The overall purpose for this research was to aid the future conservation planning of this critically endangered primate by adding to the literature of the western lowland gorilla captive population present in the UK.

This study has found comparable results to previously published literature, where for the most part, results are complementary to the findings of the US captive population, which is encouraging for the future of captive western lowland gorillas. Additionally, this chapter has reiterated the usefulness of incorporating museum specimens into ecological research for the purpose of determining the 'genetic health' of captive populations compared to historical populations which may be useful for future conservation planning of the subspecies. The identification of some of the haplogroups in the contemporary captive population is useful for conservation planning, the reintroduction of a group of western lowland gorillas to their former likely, haplogroup distribution in the wild is a positive step for the genetic health of the wild population long term and follows the recommendations of previous research by Soto-Calderón *et al*. (2015). The historical element of this research has largely complemented previous regional genetic investigations in terms of haplogroup distribution, however, it has revealed that in some cases at least, historical haplogroup distribution may have been more widely distributed than currently reported. Numts were identified in the dataset and 30 true mtDNA sequences were identified for more accurate analyses.

The hypotheses and predictions for this chapter stated that it was anticipated that the western lowland gorilla samples and specimens used in this study from a contemporary captive population (Aspinall Foundation) and a historical wild population (PCM) would belong to haplogroups C or D. This was an accurate statement, thus confirming all gorillas used in the dataset were western lowland gorillas and not another subspecies. Generally, the regional variation of the historical population was present in terms of haplogroup and confirmed with the use of biodiversity mapping of the historical populations. However, some interesting contradictions were revealed in terms of haplogroup distribution with indications of the historical haplogroup distributions potentially being more widely distributed and admixed than initially reported. Despite the contemporary population containing more known related individuals than the historical population, genetic diversity estimates for the contemporary captive population revealed that the captive population is genetically 'healthy' and has retained a large proportion of genetic diversity found in former historical populations. Numts

were prevalent in this study as predicted, with a larger number than anticipated being identified, the removal of numts was necessary and demonstrated by the analyses of two datasets (all data, and true mtDNA sequences), despite this, the study still produced thirty true mtDNA sequences which have added to the growing literature and research of this critically endangered primate.

To further this research, it would be of benefit to sequence as many remaining PCM specimens as possible, although there are associated issues such as the quality of DNA and subsequent sequences obtained, plus the prevalence of numts which is likely to continue to hinder analyses. However, the successful analyses of 22 of the PCM specimens highlighted that more information from natural history collections could be of significant value to the conservation of species, particularly critically endangered species whose wild populations are in decline and/or those whose historical distributions are unknown. Additionally, improvements to the methods such as cloning and long-range PCR, as described by Clifford *et al*. (2004), or other methods using high throughput sequencing techniques like Restriction Site-Associated DNA sequencing (RAD-seq or RAD-tags, Baird *et al*. 2008) may yield improved quality data. Complete mitochondrial genomes have been sequenced for gorillas and other primates by various methods (Xu & Arnason 1996; Das *et al.* 2014; van der Valk *et al.* 2017). Van der Valk *et al.* (2017) applied target capture and subsequent next generation sequencing methods to obtain sequences from fecal and museum specimens, these methods effectively overcome the issues of numts because true mtDNA fragments are present in vast numbers compared to the number of numts (Li *et al.* 2012; Guschanski *et al.* 2013), this research would benefit from such analyses. With the application of more advanced molecular techniques such as mitogenomic sequencing, microsatellite and Single Nucleotide Polymorphism (SNPs) analyses, further investigations into single markers, such as mtDNA HVI, may not be of the most value.

Chapter 4

Microsatellite genotyping of historic and contemporary populations of western lowland gorillas

4.1. Introduction

Primates (human and non-human) have been the subject of a large volume of genetic research and publications for many years (Enard & Pääbo 2004; Fischer *et al*. 2006; Hughes & Rozen 2012; Prado-Martinez *et al*. 2013). Due to their genetic similarity to humans and chimpanzees, gorillas have featured in many comparative genetic primate publications, e.g. Meyer *et al*. (1995), Rubinsztein *et al*. (1995), Bradley *et al*. (2000), Hacia (2001). However, Vigilant & Bradley (2004) noted there were relatively few studies investigating genetic variation in wild populations of gorillas, in comparison to chimpanzees. Moreover, much of the genetic research relating to historic populations of the two gorilla species and their subspecies lack precise geographical coordinates, which is a general issue of most studies investigating genetic variation from samples obtained from natural history collections (NHCs) (Ponder *et al*. 2001). Many NHCs have regions specified but few contain more precise geographical data which ultimately means investigations into genetic regional variation of historic populations is hindered (Murphey *et al*. 2004). The relative uniqueness of the specimens in the Powell-Cotton Museum (PCM) with regards to the additional geographical data available (see previous chapters) allows the investigation of genetic variation with a geographical context. Additionally, the accompaniment of other contextual information recorded with the specimens, such as sex, family groups, indicators of disease, etc., makes the collection invaluable for a variety of studies not limited to genetic investigations.

Microsatellites, also known as simple sequence repeats (SSRs), short tandem repeats (STRs) or simple sequence length polymorphisms (SSLPs), are chromosomal regions of short repetitive motifs of DNA, formed by two to six nucleotide base pairs in length, repeated between 5 to 50 times (He *et al*. 2003; Garner *et al*. 2005; Richard *et al*. 2008; Gulcher 2012). The existence of microsatellite loci in eukaryotic organisms was discovered in the 1970s, however, it was not until 1982 that the prevalent occurrence and vast quantity of these sequences was demonstrated by Hamada *et al*. (1982) when they identified the presence of multiple copies of the poly(Dt-Dg)$_n$ motif in yeast and in vertebrates (Bruford & Wayne 1993; Bruford *et al*. 1996; Ashworth *et al*. 2004).

In contrast to other tandem repeats known as minisatellites (Jeffreys *et al*. 1985; Paço *et al.* 2019), microsatellites commonly have allele sizes less than 500 bp making them well suited for polymerase chain reaction (PCR) combined with gel electrophoresis and genotyping (Bruford & Wayne

1993; Bruford *et al*. 1996; Neumann & Wetton 1996; Dumbovic *et al.* 2017). Initial publications on microsatellites in humans (Litt & Luty 1989; Weber & May 1989) and *Drosophila* spp. (Tautz 1989) led to the extensive use of microsatellites in a multitude of scientific applications making them one of the most prolific and significant genetic markers available (Bruford *et al*. 1996; Neumann & Wetton 1996; Dumbovic *et al.* 2017).

One comparative great ape study by Bradley *et al*. (2000) identified nine and ten microsatellite loci which they applied for genotyping wild chimpanzees (*Pan troglodytes verus*) and western lowland gorillas (*G. g. gorilla*), respectively, using faecal samples. Their study concluded that the loci selected were suited for comparative genetic analysis due to their high level of variability. Other studies, e.g. Bergl *et al*. (2008), Arandjelovic *et al*. (2010), Nsubuga *et al*. (2010), Simons *et al*. (2013), and Fünfstück *et al*. (2014) have all selected these loci (or a selection of them with a combination of other loci) for genetic variation analysis. These loci continue to be used in more recent studies, such Hagemann *et al*. (2018), indicating their usefulness for comparative studies.

4.2. Aims

The aims of this study were:

- To investigate genetic diversity and structure of the historic (Powell-Cotton Museum, PCM collection) and contemporary (Aspinall Foundation, ASP) populations of the western lowland gorilla (*G. g. gorilla*) using microsatellite loci.
- To ascertain whether the historical population of the western lowland gorilla demonstrates regional genetic variation.
- To compare the genetic diversity between the historical and the contemporary populations, and to confirm that all gorillas from the ASP and PCM were western lowland gorillas.
- To confirm the paternity of a few captive individuals where the sire is not confirmed or questionable.
- To generate genotypic information of the contemporary population to assist with the management of the captive population and aid in the planning and decision-making for conservation biology.

4.3. Hypotheses and predictions

Based on previous research and results from previous chapters of this thesis, the hypotheses and predictions for this study are as follows:

- Regional structure and variation will be present in the historical population.
- The historical population will likely reveal more genetic diversity and structure than the captive population.
- Due to the social structure and polygamous mating strategy, the silverback of the group will be confirmed as the paternal sire in cases where paternity is questioned.
- The gorillas used in this study will be confirmed as the western lowland subspecies and not the Cross River subspecies.
- Structure of the UK captive population studied here will be comparable to results of the US captive population.

4.4. Methods

The following methods were employed to conduct these analyses using a multidisciplinary approach which combines findings from previous chapters. The initial gorilla mapping in Chapter 1 (Fig. 1.5) was used to identify the regional groups (A, B and C), as seen in previous chapters.

4.4.1. DNA extraction, amplification and genotyping

The sampling and DNA extraction methods were the same as described in Chapter 3 for mtDNA. Although there were 182 skin specimens in the PCM available for DNA extraction and genotyping, a subset of 140 specimens were selected due to availability of resources for this research. Multilocus genotypes were produced for a total of 140 western lowland gorillas, 64 from the contemporary population at the Aspinall Foundation (ASP) and 76 for the historic population (PCM), which consisted of 74 gorillas from the PCM collection and 2 from the Royal College of Surgeons (RCS). The two RCS specimens were included because they are of scientific importance due to the amount of tissue preserved on the skeletons, which is unusual, and permitted histological analyses (Cooper & Hull 2017). The RCS specimens are originally thought to have been part of the PCM collection and there are indications from archival evidence that this was the case (J. E. Cooper pers. comm. 2016). Furthermore, one PCM specimen (MII.25) was of unknown origin but it was included with the

anticipation of locating the origins of the sample. The 73 specimens with precise geographic coordinates were split into the three geographical subgroups: 45 individuals in subgroup A, 26 individuals in subgroup B, and 5 individuals in subgroup C. Where individuals in subgroup A and B failed to produce viable DNA for genotyping, another individual from PCM was selected from preferably the exact same geographical coordinate to replace them.

Figure 4.1 was produced to investigate the dataset in terms of sex distribution to avoid bias and to define the regional groups for this chapter which are consistent with previous chapters. Many studies not limited to primatological ones, require information on sex dispersal and demographics. Each subgroup was represented by individuals from both sexes although at eleven precise unique geographical coordinates only one sex was represented, however, of those eleven, nine of those locations only had one individual representing that coordinate. In total, 41 individuals were female, including the RCS specimen PA62, while the sex of RCS PA63 is unknown. MII.25 is a female gorilla, but the geographical coordinates are unknown. Males are represented by thirty-three individuals with known geographical coordinates.

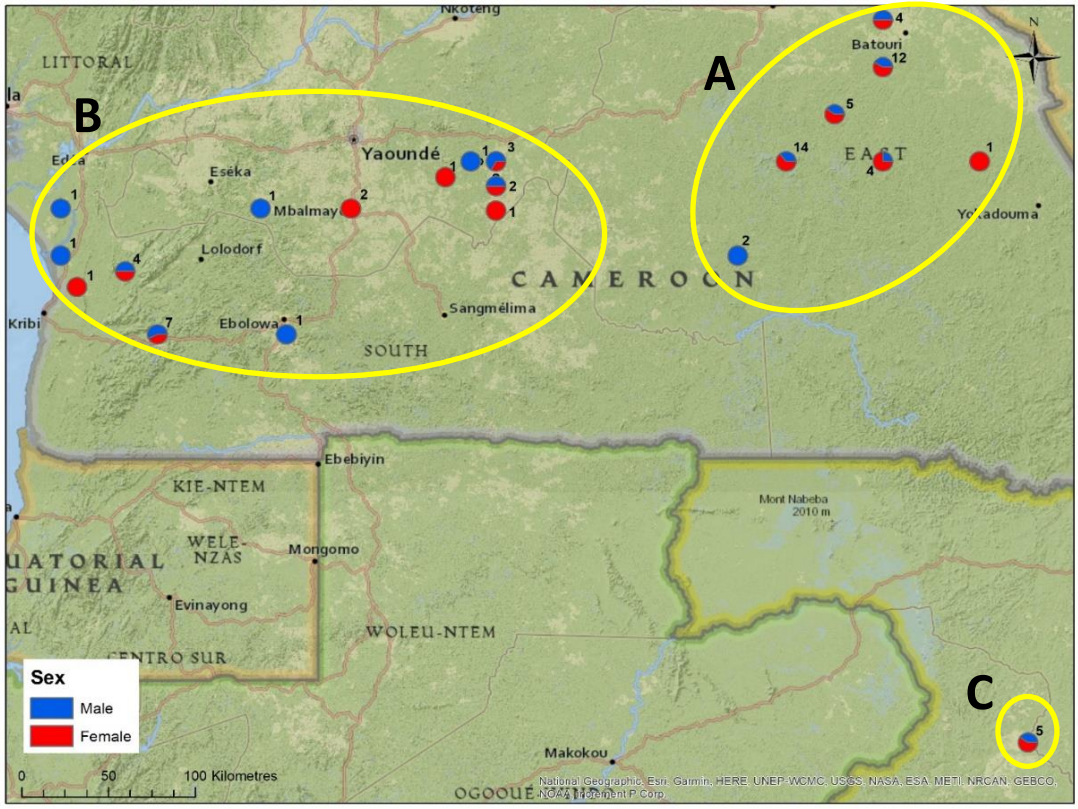


Figure 4.1 The precise geographical locations of the 73 gorilla genotypes from the Powell-Cotton Museum (PCM; historic) population used in this study. Geographical subgroups/regions are depicted by the yellow ovals and their designated subgroup name is indicated by the uppercase letter (A, B and C). The circles indicated the locations of the gorillas (blue for male, red for female) and the number next to each location refers to the number of gorillas captured from that precise location.

A panel of 10 pairs of primers (Table 4.1) were used to amplify polymorphic autosomal microsatellite loci (namely, D2s1326, D10s1432, D16s2624, D4s1627, D7s817, D5s1470, vWF, D7s2204, D1s550 and D8s1106; Bradley *et al*. 2000). The forward primers were fluorescently labelled at the 5' end for downstream fragment analysis of microsatellites. The ten microsatellite loci were selected due to showing high proportion of amplification success and polymorphism (Bradley *et al*. 2000). Additionally, the use of these loci would allow comparisons with other results, such as Arandjelovic *et al*. (2010), Bradley *et al*. (2010) and Simons *et al*. (2013). All the loci selected were tetranucleotide repeats (4 bp) except for D5s1470, which was a tetranucleotide repeat with a 2 bp indel (insertion/deletion).

Fluorescent dyes for the forward primers were chosen according to the allele size ranges of the microsatellite loci and to allow multiplexing in downstream fragment analysis (i.e. combining multiple loci in one fragment analysis step); loci with similar or overlapping allele size ranges had forward primers with different fluorescent dyes, but if allele size ranges were non-overlapping and sufficiently different in size the same fluorescent dye could be selected (Table 4.1). Due to low quality and quantity of genomic DNA, as expected for museum samples, two multiplexing reactions (named Multiplexing Group A and Group B) were performed per DNA sample targeting five different loci each (Table 4.1). Microsatellite fragment analysis was performed by DBS Genomics, Durham University using the Applied Biosystems 3730 DNA Analyser with the DS-33 filter set consisting of the fluorescent dyes: 6-FAM (blue), VIC (green), NED (yellow) and PET (red), and LIZ (orange), the latter reserved for the sizing standard.

Table 4.1 Panel of 10 microsatellite loci used for this study (based on Bradley *et al*. 2000).

Locus	Label	Primer sequences (5'-3')	Allele size range	Multiplexing Group
vWF	PET	F: CCCTAGTGGATGATAAGAATAATC R: GGACAGATGATAAATACATAGGATGGATGG	144-160 bp	A
D4s1627	NED	F: AGCATTAGCATTTGTCCTGG R: GACTAACCTGACTCCCCCTC	230-246 bp	A
D8s1106	VIC	F: TTGTTTACCCCTGCATCACT R: TTCTCAGAATTGCTCATAGTGC	123-151 bp	B
D10s1432	FAM	F: CAGTGGACACTAAACACAATCC R: TAGATTATCTAAATGGTGGATTTCC	156-176 bp	B
D2s1326	FAM	F: AGACAGTCAAGAATAACTGCCC R: CTGTGGCTCAAAAGCTGAAT	250-286 bp	A

D5s1470	PET	F: CATGCACAGTGTGTTTACTGG R: TAGGATTTTACTATATTCCCCAGG	170-202 bp	B
D7s2204	VIC	F: TCATGACAAAACAGAAATTAAGTG R: AGTAAATGGAATTGCTTGTTACC	217-249 bp	B
D7s817	NED	F: TTGGGACCTCTTATTTTCCA R: GGGTTCTGCAGAGAAACAGA	160-196 bp	B
D1s550	VIC	F: CCTGTTGCCACCTACAAAAG R: TAAGTTAGTTCAAATTCATCAGTGC	170-194 bp	A
D16s2624	FAM	F: TGAGGCAATTTGTTACAGAGC R: TAATGTACCTGGTACCAAAAACA	128-144 bp	A

Singleplex (individual) PCR reactions were carried out for each sample and each loci in a final volume of 13 µl which contained the following reagents: 6.25 µl DreamTaq Green Hot Start Green DNA Polymerase master mix (Thermo Scientific, UK), 1.25 µl of forward Primer (10 µM), 1.25 µl of reverse primer (10 µM), 3.25 µl of dH_2O, and 1 µl of DNA template.

Three DNA samples (FC.130 and ZII.63 from the PCM, and Bitono from ASP) were chosen for initial testing with primers for loci vWF, D4s1627, D8s1106 and D10s1432, based on the presence of genomic DNA on agarose gel electrophoresis for mtDNA. Negative controls (with dH_2O instead of DNA) were included in all PCRs. Two thermocycler protocols were tested for amplification success and quality of PCR bands (Table 4.2).

Table 4.2 Thermocycler protocols tested for the amplification of 10 microsatellite loci.

Msat Thermocycler Protocol	**MtDNA Thermocycler Protocol**
Initial denaturing at 95°C for 15 minutes	Initial denaturing at 95°C for 4 minutes
45 cycles of 94°C denaturing for 30 seconds, 50°C annealing for 90 seconds, 72°C extension for 90 seconds	50 cycles of 95°C denaturing for 30 seconds, 50°C annealing for 30 seconds, 72°C extension for 30 seconds
Final extension of 72°C for 10 minutes	Final extension of 72°C for 10 minutes

PCR products were separated by size following electrophoresis in 3% agarose gels in 1X TAE buffer and stained with SYBR safe (Invitrogen), set at 80 V for 60-80 minutes, and visualised under UV-light on a Bio-Rad's Gel Doc XR+ (Bio-Rd, UK). Although both PCR protocols produced visible bands on the gels (Fig. 4.2), the thermocycler protocol used originally for the mtDNA produced the most favourable results. Primers for the remaining loci (D2s1326, D5s1470, D7s2204, D7s817, D1s550 and D16s2624) were tested using the same three DNA samples and also produced bands at the desired molecular weight.

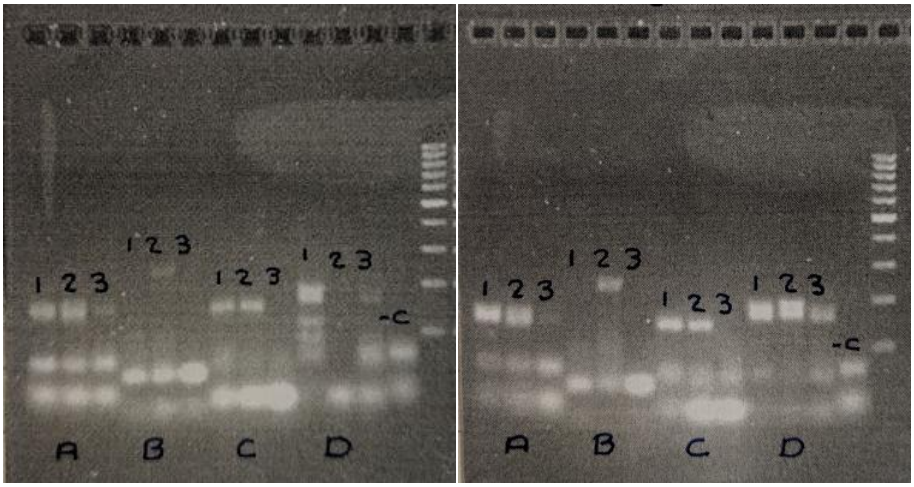


Figure 4.2 Gel electrophoresis results for primer tests using DNA from three gorillas 1) FC.130 (PCM), 2) ZII.63 (PCM), and 3) Bitono (ASP) using four primers for loci: A) vWF, B) D4s1627, C) D8s1106 and D) D10s1432. Two thermocycler protocols were tested, the microsatellite protocol (left image) and the mitochondrial protocol from Chapter 3 (right image). The mtDNA protocol produced more favourable results indicating more PCR bands at the desired molecular weights.

For microsatellite fragment analysis, three more individuals from ASP (Mataki, Kwimba Infant 1 and Baloo), and five individuals from PCM (FC.114, FC.115, FC.124, FC.147 and Mer.720) were processed with the mtDNA thermocycler protocol for all ten loci. The presence of bands was observed after gel electrophoresis for many of the loci across all individuals except for Baloo.

For all individuals (including Baloo) 1 µl of PCR products for each Group A loci and 1 µl of PCR products for each Group B loci were multiplexed and sent for microsatellite fragment analysis at DBS genomics to trial multiplexing of PCR products and ascertain whether the quality and quantity of PCR product was sufficient for genotyping. Baloo was included to test whether genotyping was still possible even in the absence of a visual result from the agarose gels. For the other samples, 5 µl of PCR product for each individual and each locus was pipetted into separate wells to test whether singleplex would be more favourable than multiplex reactions.

The multiplexing trial showed promising results, and Baloo, showed good quality data for genotyping (Fig. 4.3.a, b).

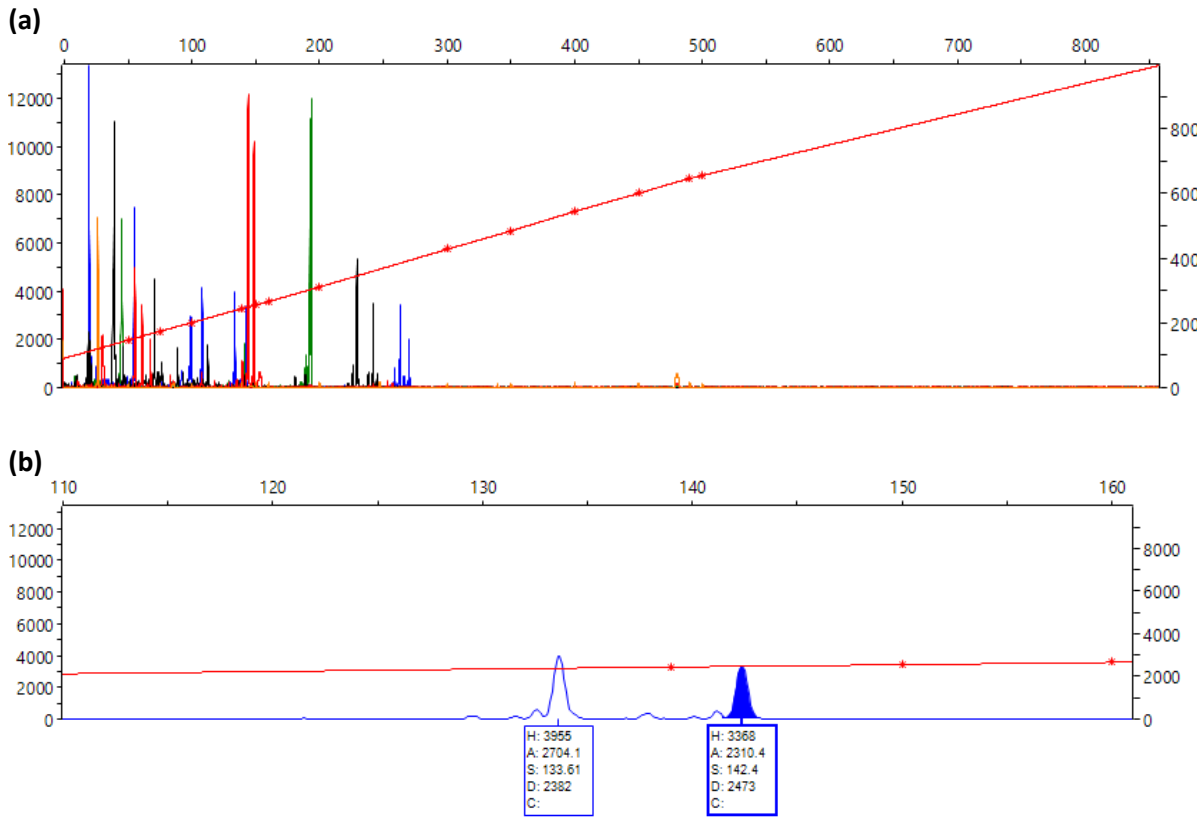


Figure 4.3 Multiplexing result for Baloo: (a) for loci vWF, D4s1627, D2s1326, D1s550 and D16s2624 and (b) singleplex for loci D16s2624.

The multiplexing method was selected for application to the other samples. All 64 Aspinall and 76 museum samples were subjected to this method. Where a locus failed to be genotyped in the multiplex method, 5 µl of the same PCR product was sent individually, and if this method still failed to produce a result, the PCR was repeated. After a third attempt, the whole process was repeated with 2 µl of DNA template. If those attempts still did not yield results, the annealing temperature was increased to 55°C and repeated three more times. All ASP samples were successfully amplified and genotyped. The museum samples were not as successful, likely due to the increased degradation of the DNA compared to the fresher contemporary blood samples. However, most loci across most individuals amplified successfully.

To check for reproducibility and to avoid genotyping errors, at least 25% of all samples were re-amplified and genotyped. All alleles which appeared to be homozygous were re-amplified and genotyped at least three times. No contamination was observed in any of the negative controls. Due

to the DNA similarity with humans and potential cross amplification, I genotyped myself with all 10 primer pairs. Although a few of my alleles fell within the allele size ranges of the gorilla genotypes, as expected, my genotype data did not match any of the samples. Because of this cautionary step, coupled with the negative controls and the data quality checks, it was assumed that the microsatellite data obtained from gorilla DNA was correct and free from contamination.

Further quality checks were conducted manually using known gorilla family groups primarily within the Aspinall Foundation, and one of the known family groups from the PCM collection, and the RCS specimens as they were also known to be a family group consisting of one adult male and female and their one offspring. Additionally, the genotype scoring results from two different DNA extraction methods were possible from the same individual. The gorilla Louna (ASP) had an FTA card and a tissue sample and the same genotype was obtained.

4.4.2. Assessing genotyping errors

Genotype data was examined for genotyping errors, allelic dropout and null alleles using Microchecker version 2.2.3 (van Oosterhout *et al*. 2004). Allelic dropout refers to an observed source of missing microsatellite genotype data and is a reasonably common occurrence amongst genotype data; it is due to one or both alleles at a locus in a diploid individual, failing to amplify during the PCR (Taberlet *et al*. 1996; Soulsbury *et al*. 2007). Allelic dropout is a random event in which either of the two alleles are likely to dropout and in general, occurs when DNA quality is low (Wang *et al*. 2012; Séré *et al*. 2014). Allelic dropouts are usually locus-specific, a result of primers not being a perfect match to the flanking sequences (Séré *et al*. 2014). If undetected, allelic dropout leads to an increased bias in estimates of inbreeding and a downwards bias to observed levels of heterozygosity, due to heterozygotes being mistakenly classified as homozygotes because one allele is not observed (Taberlet *et al*. 1996; Wang *et al*. 2012). To avoid allelic dropout, one method is to repeat genotyping of any homozygous loci which minimises experimental error (Taberlet *et al*. 1996; Wang *et al*. 2012).

A microsatellite null allele can be defined as any allele at a locus that consistently and repeatedly fails to amplify to detectable levels via the PCR (Dakin & Avise 2004). Similarly, to allelic dropout, null alleles can result from low quality DNA and poor primer annealing. Additionally, shorter length alleles commonly amplify more effectively than larger ones, this implication means that only the smaller of the two alleles at a locus might be detected from a heterozygous individual (Dakin & Avise 2004). For the same reasons as allelic dropout, undetected null alleles lead to a bias of increased homozygosity and a decrease of heterozygosity.

To check for allelic dropout and null alleles, all suspected homozygote loci were repeatedly genotyped as described previously, and approximately 25% of all loci were repeatedly amplified by PCR and genotyped.

4.4.3. Linkage disequilibrium and the Hardy-Weinberg equilibrium

Linkage disequilibrium (LD), also called allelic association or gametic disequilibrium (Morton 2001) is the non-random association of alleles at different loci (Slatkin 2008). It is the difference between the expected co-occurrence of two alleles at two loci, assuming no selection, random mating and independent segregation (Ovenden *et al*. 2004). For a population to be considered in linkage equilibrium, the alleles of the specified loci are required to be independently distributed on the chromosomes (Hudson 2004; Ovenden *et al*. 2004). LD is caused by mutation, genetic drift, selection and admixture (Ovenden *et al*. 2004). In addition, it assumes no subpopulation structure, no migration and a random sampling of the population in its entirety (Hill 1981).

Arlequin version 3.5 (Excoffier & Lischer 2010), was used to estimate LD (i.e. significant association) between all pairs of loci. Because the gametic phase was unknown, a likelihood ratio test of LD was performed. This test obtains the haplotype frequencies from the allele frequencies and employs the Expectation-Maximisation (EM) algorithm to estimate the haplotype frequencies (Excoffier & Lischer 2010). The test was run for the global population, with 10,000 MCMCs and 100,000 dememorisation steps, in addition to 10,000 permutations and initial conditions for EM set at 2, with a significance level of 0.05. The Holm-Bonferroni method was applied to the results, a statistical correction method used to rectify familywise error rates (FWER or FWE) for multiple hypothesis tests (Hartl & Clark 1997; Hamilton 2011). FWER is the probability of performing at least one Type I error (incorrectly rejecting the null hypothesis) during multiple hypothesis tests (Hartl & Clark 1997; Hamilton 2011). The likelihood ratio test of LD assumes Hardy-Weinberg equilibrium (HWE) proportions of genotypes. Departure from the HWE could lead to rejection of this test (Excoffier & Lischer 2010).

HWE is used to describe and predict genotype and allele frequencies in a population that is not evolving and is often the first analysis performed in population genetics studies (Waples 2014). The principle is that the genetic variation of a population from one generation to the next, will remain constant if the assumptions are not violated (Hartl & Clark 1997). It is an important null model as it allows the genetic structure of a population to be compared over time, with the genetic structure of the population as would be expected if evolution was not occurring (Hartl & Clark 1997). As a null model, HWE makes predictions based on an ideal or simplified situation and assumes that there are

no biological processes in play and that the genotype frequencies result from random combination of gametes (Hartl & Clark 1997).

HWE has five basic assumptions, if any of these assumptions are violated the population will not be in equilibrium. The five assumptions are: no mutation occurs, mating is random, there is no gene flow among populations, the population size is infinite, and natural selection is not occurring (Masel 2012). The method to calculate HWE consists of two equations, both of which must equal one because they are frequencies. One equation calculates allele frequencies and the other calculates genotype frequencies. The equation for calculating allele frequencies is:

$$p + q = 1$$

Where, p = the frequency of one of two alleles, and q = the frequency of the other allele.

In diploid organisms with alleles *A* and *a* for a specified locus, there are three possible genotypes; *AA, Aa* and *aa*. The equation that represents a population in HWE is:

$$p^2 + 2pq + q^2 = 1$$

Where p^2 = the frequency of homozygotes for one allele, $2pq$ = the frequency of heterozygotes for both alleles, and q^2 = the frequency of homozygotes for the second allele.

There two ways in which to test for HWE: a goodness of fit test or an exact test. GenAlEx version 6.5 (Peakall & Smouse 2012), uses a goodness of fit test and was used to test for deviations from HWE for all loci in the global population and the Holm-Bonferroni correction applied to adjust for multiple comparisons.

4.4.4. Genetic diversity analysis

The genetic diversity of populations was assessed using measures of allelic diversity and heterozygosity as in other studies, e.g. Bergl *et al*. (2008) and Simons *et al*. (2012). Estimates of genetic diversity per locus and per population were performed in GenAlEx version 6.5 (Peakall & Smouse 2012). This included estimators of heterozygosity, allele frequencies, effective number of alleles and private alleles.

Genetic diversity was assessed in terms of allelic richness, allelic frequencies, and heterozygosity. The number of alleles (*NA*), allele frequencies and number of effective alleles (*AE*) were calculated in GenAlEx for the contemporary and historic populations (as well as the three regional subgroups, A, B and C) for all 10 loci. Caution must be given to the results as 11% of data was

missing at locus D4s1627 and 6% at D5s1470 for the historic population. The contemporary population had complete genotype data for all 64 gorillas.

Allele frequencies (also referred to as gene frequencies) describe the proportion (fraction or percentage) of copies of genes for a particular allele in a defined population (Silver 2001). Allele frequency is a measure of the relative frequency of an allele on a genetic locus in a population (Hamilton 2011) they show the genetic diversity of a population and the richness of its gene pool and it is usually expressed as a fraction, a percentage or decimal (Hartl & Clark 1997; Hamilton 2011). Allele frequencies vary among population groups (Butler 2015). In a population, there are twice as many alleles as there are diploid individuals (Stephenson 2016), because each individual possesses two alleles per locus, one allele inherited maternally, and one allele inherited paternally (Stephenson 2016). An allele frequency is calculated by the total number of copies of that allele in the population, divided by the total number of copies of all alleles of the gene (Hartl & Clark 1997; Hamilton 2011).

Observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated in GenAlEx. Heterozygosity is often one of the first parameters presented in a data set. It is used to infer information about population structure and their history (Smith & Wayne 1996). Values of high heterozygosity indicate a higher level of genetic variability, in contrast, low levels of heterozygosity indicate lower amounts of genetic variation (Fowler *et al*. 2013). Comparisons of observed heterozygosity (H_O) and expected heterozygosity (H_E) are often made to surmise theories regarding the population structure of the species/subspecies in question. If observed heterozygosity is higher than expected, we can infer theories such as the hybridisation of previously isolated populations (isolate-breaking effect) and if heterozygosity is lower than expected we can make assumptions that other forces, such as inbreeding are occurring (Smith & Wayne 1996).

There are several measures of heterozygosity, ranging from 0 (no heterozygosity) to 1 (high proportion of equally frequent alleles) (Ashcroft & Pereira 2002; Fowler *et al*. 2013). The simplest method to calculate H_E is:

$$1 - \sum_{i=1}^{k} p_i^2$$

Where, p_i = is the frequency of the i^{th} of k alleles.

To calculate the gene diversity over several loci, the formula needs to be doubled as:

$$1 - \frac{1}{\mathrm{m}} \sum_{l=1}^{\mathrm{m}} \sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{p_i}^2$$

Where, the first summation is for the l^{th} allele of m loci, and the second part of the formula remains the same as in the first H_E equation (McDonald 2008).

Individual heterozygosity refers to the proportion of heterozygous loci within an individual, and mean individual heterozygosity was calculated as the mean number of heterozygous loci for each individual gorilla divided by the total number of loci. This is the same method as employed by Nsubuga *et al*. (2008) and Simons *et al*. (2013).

The number of effective alleles (*AE*) refers to the number of alleles that can be present in a population. It is a measure of the evenness of the allele frequency distribution, averaged across all loci, and it provides an indication of the number of alleles that would be expected in a locus in each population. *AE* can be affected by sample size (Simons *et al*. 2013). The calculation for *AE* is as follows:

$$AE = 1/(1 - h) = 1/\sum p_i^2$$

Where, p_i = frequency of the i^{th} allele in a locus, and h is the heterozygosity in a locus (De Vicente *et al*. 2004).

Allelic richness (*AR*) is a genetic diversity measure indicative of the long-term potential for adaptation in a population (Greenbaum *et al*. 2014); the measure controls for differences in sample size which *AE* does not (Simons *et al*. 2013), hence, allowing comparison between different sample sizes. Allelic richness was calculated in FSTAT version 2.9.4 (Goudet 2005) for all loci in both populations and the subgroups.

T-tests were performed in SPSS version 23.0 (IBM Corp., Released 2015, Armonk, NY) to test for significance in allelic diversity values (*NA, AE, AR*) between the contemporary and historic populations, and an analysis of variance (ANOVA) was performed to test for significance among the three historic subgroups.

The population parameter θ (Weir & Cockerham 1984; Neigel 1996) was calculated in Arlequin under the stepwise mutation model (SMM). θ refers to the correlation of genes in the same population belonging to different individuals (Weir & Cockerham 1984), while SMM is a mathematical model for analysing microsatellite mutations (Fan & Chu 2007). The original model has undergone various revisions (Valdes *et al*. 1993) and the SMM now follows these assumptions: small changes in

repeat number, equal probabilities of increasing or decreasing repeat number, unlimited allele size, independence of the rate of size of mutations from the repeat number (Fan & Chu 2007). Alleles can mutate up or down in the SMM by one repeat unit (or a small number). The strict (single step) SMM is the model that changes only by one repeat unit each time. If the change is more than one unit at a time, it is known as the TPM (two-phase mutation model) (Fan & Chu 2007).

Private alleles can be defined as alleles that are only present in a single population among a wider expanse of populations (Szpiech & Rosenberg 2011), unlike rare alleles that have been traditionally defined in relation to their relative frequencies (Kimura 1983). Private alleles have demonstrated to be informative for a multitude of population genetic studies in fields of conservation genetics and molecular ecology (Szpiech & Rosenberg 2011) given their effectiveness for investigating migration patterns and population structure (Neel & Thompson 1978; Calafell *et al.* 1998; Schroeder *et al.* 2007; Szpiech *et al*. 2008;). In this study, private alleles for all populations were determined in GenAlEx.

4.4.5. Genetic structure of populations

There are a variety of methods that can be applied to ascertain the genetic structure of populations, including *F*-statistics, analysis of molecular variance (AMOVA), and genetic clustering with Bayesian methods and non-Bayesian methods. These methods were selected for their comparative capabilities and their common use in population genetic studies.

In addition to the 140 western lowland gorilla genotypes generated from this study, to address questions relating to species population structure, an additional 34 Cross River gorilla genotypes were included in parts of the analyses. These 34 genotypes also represent a contemporary population (N = 20) and a historic population (N = 14) data from Arandjelovic *et al*. (2015) and Thalmann *et al*. (2011), respectively. However, only four loci (D5s1470, D8s1106, D16s2624 and vWF) could be used for comparative purposes because they were comparable across the data, whereas the other 6 loci used in this study for western lowland gorillas were either not present in the Cross River gorilla data or the genotype scoring was not comparable.

The population genetic structure was assessed using *F*-statistics. As mentioned in the previous chapter, *F*-statistics (fixation index) was developed by Wright (1921) and is used to quantify the inbreeding effect of population substructure (Hartl & Clark 1997). F_{IS} (the inbreeding coefficient) measures the extent of genetic inbreeding within subpopulations and ranges from –1.0 to 1.0, representing total heterozygosity and total homozygosity, respectively (Shane 2005). The F_{ST} index is the most inclusive population structure measure (Hartl & Clark 1997) and measures the extent of genetic differentiation among subpopulations with ranges from 0.0 to 1.0 indicating no differentiation

to complete differentiation, respectively (Hartl & Clark 1997; Shane *et al*. 2005). F_{ST} has advantages over traditional *F*-ratio based statistical methods as it allows for comparisons in variation across taxonomic units and among species and can also compare divergence among variables e.g. genetic F_{ST} vs morphological differentiation (Relethford 1994; Leigh *et al*. 2003). *F-statistics* for this study, including pairwise F_{ST} were calculated in GenAlEx.

The AMOVA is a flexible method as it can be implemented for different types of genetic data, e.g. haploid data (mtDNA sequences as in the previous chapter), as well as diploid data (microsatellites) (Michalakis & Excoffier 1996). In the previous chapter, AMOVA was performed using the allelic information based on a genetic distance matrix using Euclidean squared distances (Excoffier *et al*. 2005). For microsatellite markers the within and between-groups sums of squares are also calculated from a squared Euclidean genetic distance matrix but takes a locus-by-locus approach (Michalakis & Excoffier 1996). For this study, the AMOVAs were performed in GenAlEx for the subspecies (Cross River and western lowland gorillas), for the western lowland gorillas divided into their respective contemporary and historic populations, and for the historic dataset (PCM) divided into the three regional subgroups (A, B, and C).

The software Structure version 2.3.3 (Pritchard *et al*. 2000) was used to infer the population genetic structure of 140 individuals across 10 loci, assuming two populations (contemporary and historic) as prior information. The historic population was also analysed separately to detect any regional sub-structure which may have been initially undetected from the first analysis. In addition, genotype data for Cross River gorillas from a historic population (Thalmann *et al*. 2011) and a wild contemporary population Arandjelovic *et al*. (2015) were compared with the genotype data generated from this study across four loci (D5s1470, D8s1106, D16s2624 and vWF) to ascertain/confirm that all gorillas used in this study from the contemporary (ASP) and wild (PCM) populations were actually western lowland gorillas.

Structure implements a Bayesian clustering algorithm and Markov Chain Monte Carlo (MCMC) simulations to assign individuals to genetic clusters (*K*), defined by allele frequency variations, even in the absence of predefined genetic population and/or spatial information (Gilbert *et al*. 2012; Janes *et al*. 2017). Structure also assumes that all loci are unlinked, populations are in HWE, there is free recombination between loci (Pritchard *et al* 2000; Gilbert *et al*. 2012) and random mating within populations (Gilbert *et al*. 2012). During analysis, a membership coefficient matrix is generated (i.e. the individual *Q*-matrix) and assigned to each individual for each group (Porras-Hurtado *et al*. 2013). As recommended, and investigated by previous research, e.g. Simons *et al*. (2013) and Gilbert *et al*. (2012), the optimal number of *K* was inferred by conducting multiple independent replicates (25 in total for each *K* value set). Values of *K* were set for *K* = 1-20, then *K* = 1-10 and finally *K* = 1-5. Individual replicates within each *K* value were tested with 100,000 burn-in steps followed by 1,000,000 MCMC

replications. The admixture model with correlated allele frequencies and the LOCPRIOR (contemporary and historic populations) options were selected.

There has been much debate regarding the selection of the optimal number of *K*. Tests which include high *K* value runs have the potential to lead to an overestimation of *K* as discussed by Rosenberg *et al*. (2001), Nsubuga *et al*. (2008) and Simons *et al*. (2013). The delta *K* (Δ*K*) statistical method was introduced by Evanno *et al*. (2005) and has been recommended to facilitate the identification of the true number of *K* (Janes *et al*. 2017). However, Evanno *et al*. (2005) also concluded that Δ*K* detects the genetic structure patterns of the highest hierarchy, thus, subsequent hierarchical analysis is recommended to be performed on identified nested clusters of *K,* and the Δ*K* method should not be the sole method employed for defining the true value of *K* (Janes *et al*. 2017). Additionally, Δ*K* does not allow a result of *K* = 1, thus, combined with the aforementioned likelihood of only the upper most level of genetic structure being detected, further analysis should be performed in addition to the *ΔK* method (Janes *et al*. 2017).

Janes *et al*. (2017) investigated the numerous publications that report *K* = 2 and the consequences of misinterpreting the true value of *K* in relation to conservation. Of the 1,264 studies investigated they observed an increase over time of studies reporting *K* = 2 with 37% of those only reporting the Δ*K* method. Recent alternative statistics have been developed by Puechmaille (2016), which consist of four new estimators (median of means, maximum of means, median of medians and maximum of medians). These new estimators, known as the Puechmaille method, are being portrayed as being more efficient and reliable methods to estimate *K*, however, they have yet to be tested widely (Janes *et al*. 2017). In addition, various authors, e.g. Hubisz *et al*. (2009) and Janes *et al*. (2017), have recommended that the LOCPRIOR parameter is used if the data is suspected to have weak genetic structure.

Results from Structure were uploaded into Structure Selector (Li & Liu 2018), which is a web-based program for visualising the optimal number of clusters using the Δ*K* method using Structure Harvester and the Puechmaille method. The individual *Q* matrix for the selected *K* number of clusters was inputted into Structure Plot (Ramasamy *et al*. 2014) to generate the individual probability assignment bar plots.

A different approach to assign individuals to clusters is with the Principal Coordinates Analysis (PCoA) methods. PCoA is an ordination method that can be used to investigate similarities or dissimilarities in genotyping data, and to visualise them graphically (Legendre & Legendre 2012). Like all ordination methods, PCoA requires a distance matrix (similarity or dissimilarity matrix) and it then calculates a series of eigenvalues (also known as “latent values”) with an equal number of eigenvectors (Guiller *et al*. 1998; Palmer 2004). Typically, the first eigenvalue is called the “leading” or “dominant” eigenvalue and they are usually ranked from the largest to smallest (Camiz 1999). PCoA

is primarily used to visualise group and/or individual differences (Legendre & Legendre 2012). GenAlEx was employed in this study to implement the PCoA method for the contemporary and historical populations and for the historical subgroups A, B and C using data generated from this study. Additionally, the Cross River gorilla data (Thalmann *et al*. 2011; Arandjelovic *et al*. 2015) was incorporated for further investigation.

4.4.6. Genetic bottlenecks, effective population size, migration and mutation

The likelihood of population extinction is increased with the reduction of genetic diversity which can be caused by populations subjected to a bottleneck (Kuo & Janzen 2003). A bottleneck is described as when a population undergoes a severe, but temporary, population decrease/restriction which causes a significant deviation from the mutation-drift equilibrium (Piry *et al*. 1999; Forsdick *et al.* 2017; Murphy *et al*. 2018; Wang *et al.* 2019). Bottleneck version 1.2 is computer program that detects a recent reduction in effective population size (*Ne*) based on allele frequency data (Piry *et al*. 1999). The contemporary population and the historic population were analysed without splitting the historic population into its three subgroups in order to meet the minimum requirements of this program.

Essentially, Bottleneck tests for an excess of heterozygosity (*He*) compared with expected equilibrium heterozygosity (*Heq*) calculated from the sample size and number of alleles, rather than testing for the number of heterozygotes observed (*Ho*) (Piry *et al*. 1999). Bottleneck uses three mutation models: the Infinite Allele Model (IAM), the SMM, and the Two Phase Mutation Model (TPM). In addition, it performs three statistical tests: the Wilcoxon Sign Rank test, Wilcoxon test and the Standardised Differences test which are used to test for significant differences in *Heq* and *He* (Simons *et al*. 2013). The SMM is the most conservative model (Cornuet & Luikart 1996) and a minority of loci follow the strict SMM (Piry *et al*. 1999). The SMM and the IAM are two extremes of mutation models (Piry *et al*. 1999). The SMM is favoured over the IAM for testing microsatellite data as the IAM is recommended for allozyme data (Piry *et al*. 1999). However, the TPM is considered even more favourable than the SMM (Piry *et al*. 1999) and is a combination of the IAM and SMM (Simons *et al*. 2013). Regarding the three statistical analysis tests, the Standardised Differences test is recommended for datasets compromising of more than 20 loci (Cornuet & Luikart 1996). For datasets containing fewer than 20 loci, Wilcoxon's test is deemed the most powerful and is recommended (Piry *et al*. 1999).

For the contemporary and historic populations used in this study, demographic history was inferred using Bottleneck, all three models were applied in addition to all three statistical tests. Following the methods of Weckworth *et al*. (2005) and Simons *et al*. (2013), runs were performed with

stepwise changes of the SMM contribution to the TPM of 70%, 75%, 80%, 85% and 90%. In addition, a run of 95% was made as recommended by Piry *et al*. (1999) and is reported.

The effective population size (*Ne*) can be defined as the size of an ideal Wright-Fisher population (i.e. one that meets all the assumptions of the HWE) that exhibits an equal amount of genetic drift and inbreeding corresponding to the rate of the population under examination (Wright 1931; Ferchaud *et al*. 2016). The effective population size over a generation (*Ne*) is a critical variable to be estimated in natural populations because it ascertains the rate of genetic variability loss (Ferchaud *et al*. 2016), while the effective number of breeders over a reproductive cycle (*Nb*) is directly associated to *Ne* because *Nb* multiplied by the generation time is approximately equal to *Ne* (Waples 1989; Ferchaud *et al*. 2016). *Ne* usually differs from the census size (*N*) due to a variety of factors including temporal fluctuations in population size, unequal sex ratio and certain mating structures. *Ne* must be considered when modelling the effects of genetic drift because the smaller the population size, the greater the effects of drift (Andrews 2010). To estimate *Ne*/*Nb* in this study, NeEstimator version 2.0 (Do *et al*. 2014) was used employing the linkage disequilibrium method.

GeneClass version 2 (Piry *et al*. 2004) is a computer software program used to detect first-generation migrants based on their genotypic data (Piry *et al*. 2004). Analyses were run in GeneClass, with the historic population data only separated into its three regional subgroups. GeneClass was used with the Likelihood ration computation = set as L_home / L_max using the Rannala & Mountain (1997) criterion, the number of simulated individuals set to 10,000, the simulation algorithm employed was Paetkau *et al*. (2004), and the threshold (*p*-value) was set as 0.05.

4.5. Results

4.5.1. Geographic regions and sampling

Being able to visualise the contextual information in map form is more user friendly and useful as a quick reference guide than numerical data. The visualisation of the historic gorilla population that were successfully genotyped for this study and their regional distribution (Fig. 4.1) showed that subgroup C (which was composed by the gorillas located in the Congo) was by far the most isolated; those five individuals were approximately 350 km from the closest two individuals in subgroup A, and 450-600 km from the closest individuals in subgroup B.

4.5.2. Genotyping errors, data quality checks and paternity testing

The genotyping results from an FTA sample and a tissue sample for the same individual (Louna from the contemporary population) confirmed that the two methods of DNA extraction, from FTA cards and tissue samples, yielded the same results and could therefore be deemed as reliable (Table 4.3).

The data quality checks of genotyping data revealed that the loci were tetranucleotide repeats except for D5s1470, which was a tetranucleotide repeat with a 2 bp indel. Moreover, the data quality checks on family groups in the contemporary population (Aspinall Foundation) confirmed that genotype scoring was accurate and confirmed the paternity of several family groups. Table 4.4 shows examples of genotyping results for several family groups, including: 1) from the contemporary population, Kouillou (male) who bred with Mambi (female) and produced three offspring: Boula, Imbi and MahMah, where for each locus the alleles of the offspring could be matched to the parents (one allele from each parent as they are diploid organisms); 2) from the historic population, Mer.135 (male), Mer.136 (female) and their infant Mer.137, where the alleles at each locus could be matched against the parental alleles; 3) from the RCS, where only two of the three specimens could be genotyped (the mother and infant), and Locus D7s2204 of the mother failed to amplify, but despite the missing data the infant possesses one allele from its mother for all other loci.

All 10 loci and both populations (contemporary and historic) did not produce any evidence of genotyping/scoring error, allelic dropout or null alleles. These results combined with the data quality checks of known related individuals (primarily from the contemporary population) indicated the data was of good quality for downstream analysis.

The family data quality checks revealed an interesting and important finding regarding the paternity of two infant gorillas in the contemporary population (Table 4.5): Djanghou was the intended sire and the expectation was that he was to breed with Kwimba (dam); however, it was speculated that Kisane (Djanghou's offspring), had sired both of Kwimba's infants. Based on the genotyping data, Djanghou could not be the biological father of either of the infants since five of both infants' loci (vWF, D7s2204, D7s817, D1s550 and D16s2624) only possessed an allele from Kwimba and the other allele did not match either of Djanghou's alleles. However, Kisane's genotype confirmed that he was the biological parent to both infants, with all alleles matching across all 10 loci (Table 4.5).

Table 4.3 Genotype scoring data for Louna for two methods of DNA extraction (tissue and FTA card).

	Locus																			
Sample/Specimen	vWf		D4s1627		D8s1106		D10s1432		D2s1326		D2s1470		D7s2204		D7s817		D1s550		D16s2624	
ASP-Louna (FTA)	153	157	237	241	141	145	162	178	259	263	194	198	244	248	173	177	181	185	138	142
ASP-Louna (tissue)	153	157	237	241	141	145	162	178	259	263	194	198	244	248	173	177	181	185	238	142

Table 4.4 Data quality checks for known family groups. Male (sire) alleles are indicated in blue and female (dam) in pink.

Sample/Specimen		Locus																			
Aspinall example	S/D/O	VWf		D4s1627		D8s1106		D10s1432		D2s1326		D2s1470		D7s2204		D7s817		D1s550		D16s2624	
Kouilou	Sire	145	157	229	249	137	141	162	166	259	263	194	198	236	240	181	189	193	197	134	142
Mambi	Dam	149	161	233	242	133	141	170	178	247	267	194	198	236	240	169	177	181	185	138	142
Boula	Offspring	149	157	233	249	137	141	166	170	259	267	198	198	236	240	169	181	181	193	134	142
Imbi	Offspring	145	161	229	242	133	141	162	170	259	267	198	198	236	240	169	189	181	197	134	138
MahMah	Offspring	145	161	229	233	133	141	166	170	247	263	198	198	236	240	177	181	185	193	134	142
PCM example																					
Mer.135	Sire	153	157	241	245	141	153	174	186	251	263	192	198	240	244	177	181	181	185	138	146
Mer.136	Dam	153	157	233	237	149	153	170	174	263	267	192	196	240	244	177	189	185	193	138	142
Mer.137	Offspring	153	157	233	245	149	153	174	186	263	267	192	196	240	244	177	181	181	185	138	142
RCS example																					
PA62	Dam	149	157	233	233	137	149	166	170	267	271	186	198	-	-	185	189	181	193	138	142
PA63	Offspring	153	157	233	242	137	141	170	178	255	267	186	194	240	248	173	185	181	185	134	138

Table 4.5 Paternity testing of Kwimba's infants between two sires (Djanghou and Kisane). Green alleles indicate a mismatch in paternity.

Sample/Specimen		Locus																			
Aspinall example	S/D/O	vWf		D4s1627		D8s1106		D10s1432		D2s1326		D2s1470		D7s2204		D7s817		D1s550		D16s2624	
Djanghou	Sire (exp.)	149	153	237	242	145	149	158	174	259	263	194	206	236	240	177	181	189	193	138	146
Kisane	Sire	153	161	237	249	141	145	170	174	259	263	198	206	240	248	173	181	185	193	138	142
Kwimba	Dam	153	157	242	249	141	141	162	178	259	259	194	198	240	244	181	185	189	193	134	142
Kwimba Inf 1	Offspring	157	161	237	249	141	145	174	178	259	263	194	198	240	248	173	181	185	193	134	138
Kwimba Inf 2	Offspring	157	161	237	249	141	145	174	178	259	263	198	206	240	248	173	181	185	193	134	142

4.5.3. Hardy-Weinberg equilibrium and linkage disequilibrium

GenAlEx results for the global population for the Chi-squared test revealed three loci that deviated from the HWE at significant levels (D2s1326, D8s1106 and D7s817) after the Holm-Bonferroni statistical correction to adjust for type 1 errors. Linkage disequilibrium (LD) was observed in the global population for eight pairs of loci at significant levels after adjusting for type 1 errors. These loci included: D4s1627 and D2s1326, D2s1326 and D5s1470, vWf and D7s2204, D4s1627 and D7s2204, D2s1326 and D7s2204, D7s2204 and D1s550, D2s1326 and D16s2624, D7s2204 and D16s2624.

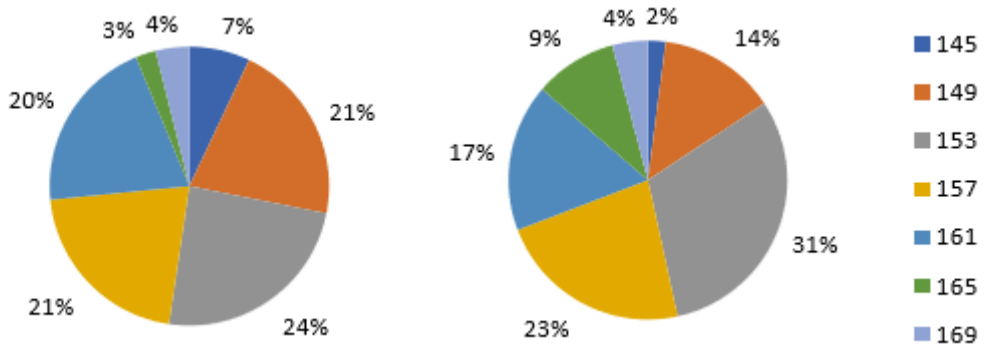
Although significant deviations from HWE and LD were found, based on previous literature and considering the sampling limitations of this study (see Discussion), all loci were retained for analyses.

4.5.4. Genetic diversity

The allele frequencies of contemporary and historic populations and subgroups at each locus are presented visually for comparative purposes (Fig. 4.4.a-j). The results showed that some allele frequencies at some loci, e.g. D4s1627, D7s2204, D7s817, were very similar between the two populations. However, there were loci with clear dissimilarities among populations in terms of allele frequencies. For example, in the contemporary population for loci Ds51470 allele 186 was present at a frequency of 36%, allele 190 at 8% and allele 194 at 25%, while in the historic population the same alleles represented 15%, 17% and 10% of the total allele frequencies, respectively.

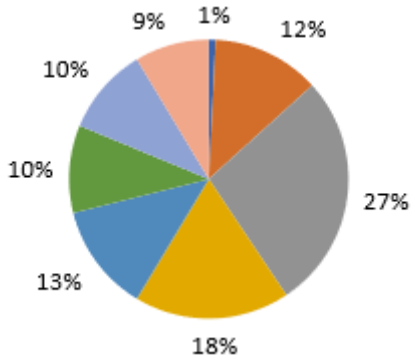
(a) Allele Frequency at vWF

Contemporary Population (N = 64) Historic Population (N = 73)

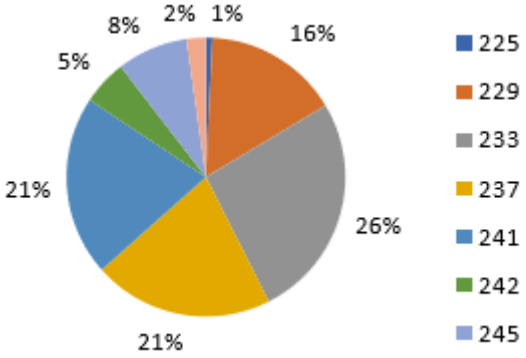


(b) Allele Frequency at D4s1627

Contemporary Population (N = 64)

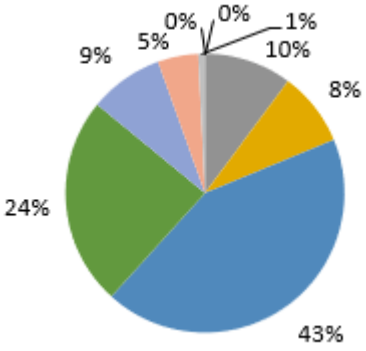


Historic Population (N = 67)

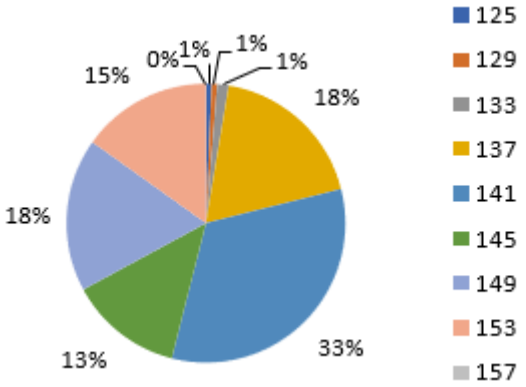


(c) Allele Frequency at D8s1106

Contemporary Population (N = 64)

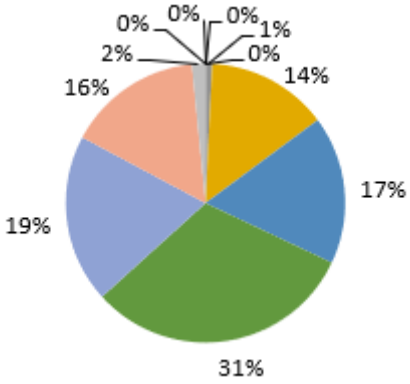


Historic Population (N = 76)

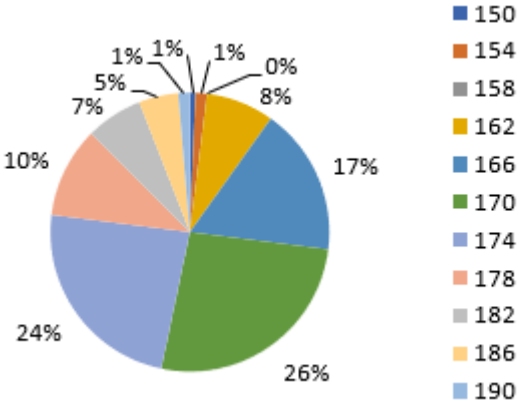


(d) Allele Frequency at D10s1432

Contemporary Population (N = 64)

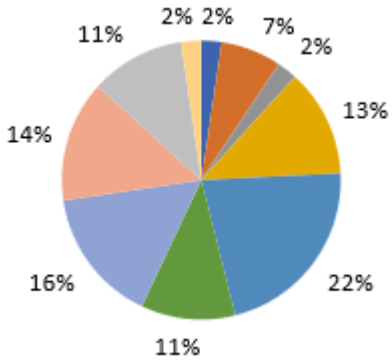


Historic Population (N = 76)

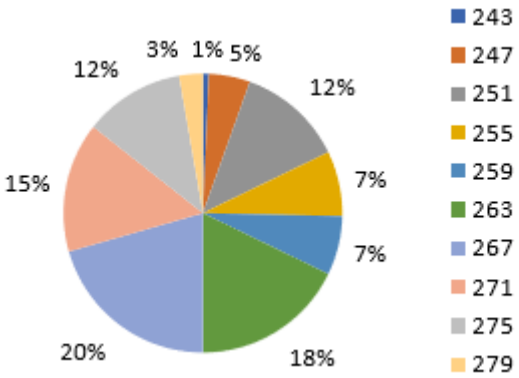


(e) Allele Frequency at D2s1326

Contemporary Population (N = 64)

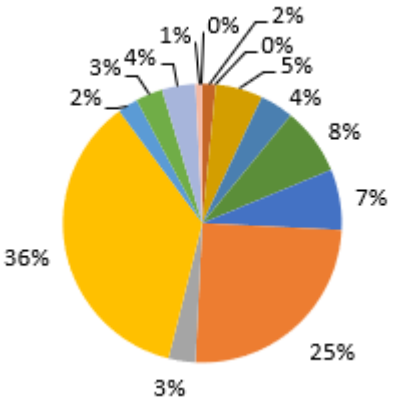


Historic Population (N = 73)

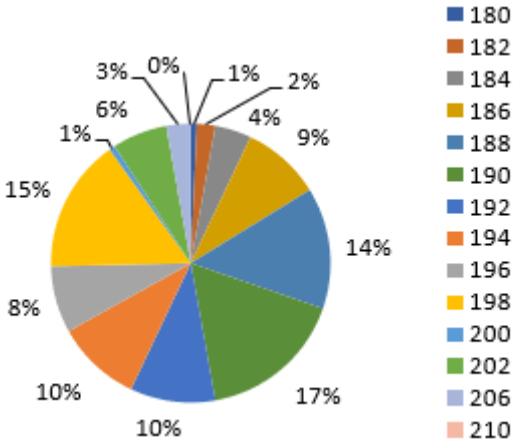


(f) Allele Frequency at D5s1470

Contemporary Population (N = 64)

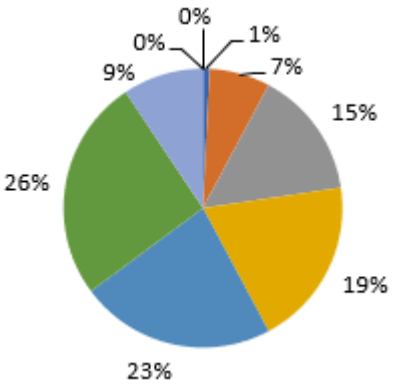


Historic Population (N = 71)

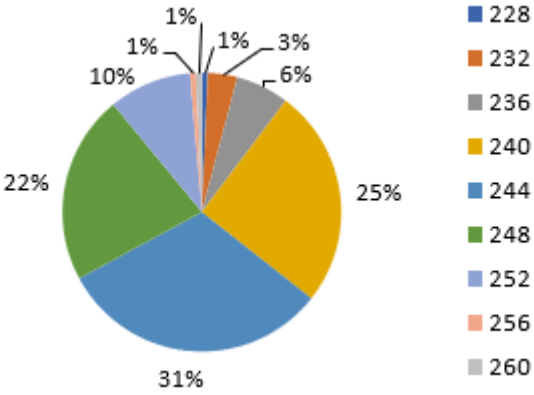


(g) Allele Frequency at D7s2204

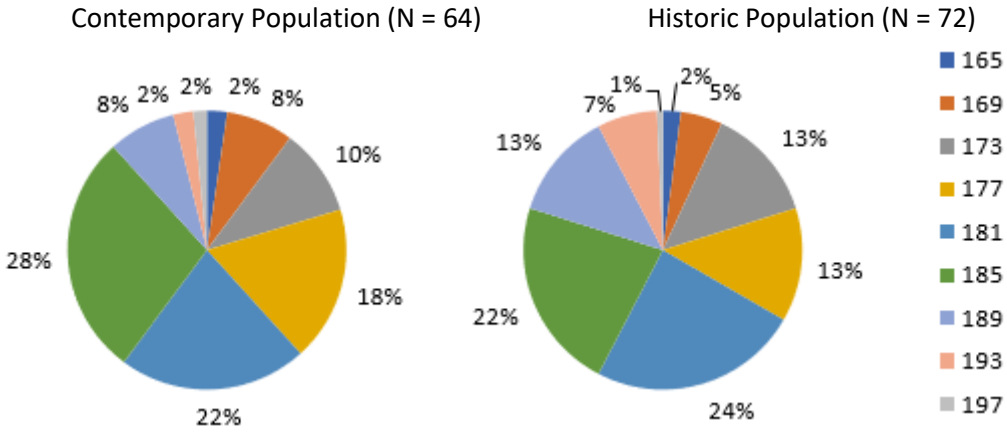
Contemporary Population (N = 64)



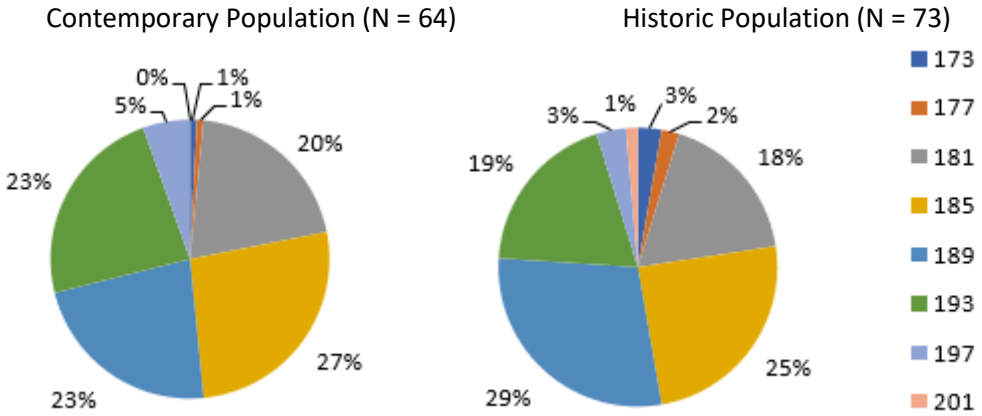
Historic Population (N = 73)



(h) Allele Frequency at D7s817



(i) Allele Frequency at D1s550



(j) Allele Frequency at D16s2624

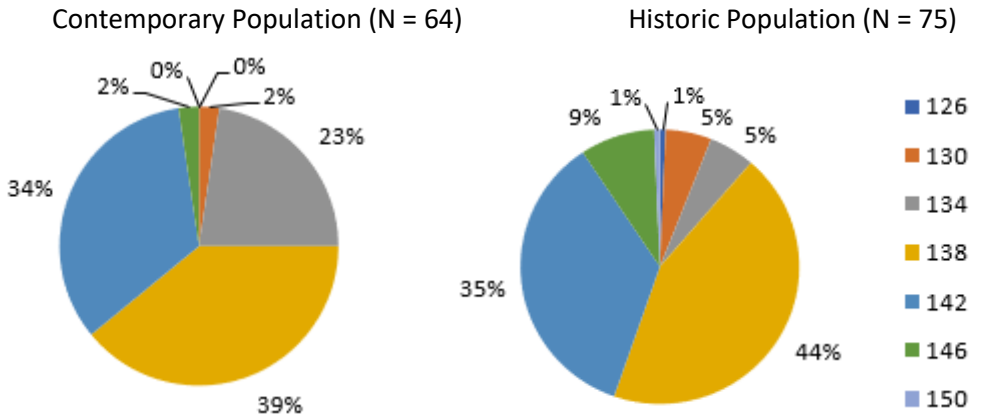


Figure 4.4 Comparisons of allele frequencies between the contemporary and historic populations of gorillas for all 10 loci: (a) vWF, (b) D4s1627, (c) D8s1106, (d) D10s1432, (e) D2s1326, (f) D5s1470, (g) D7s2204, (h) D7s817, (i) D1s550, (j) D16s2624.

The *NA* (number of alleles), *AR* (allelic richness), *Ae* (effective number of alleles), H_O (observed heterozygosity), individual heterozygosity (H_I) and H_E (expected heterozygosity) by loci are reported in Table 4.6 for the contemporary and historical populations, and for the three subgroups A, B and C. In all populations and subgroups, the observed heterozygosity was greater than the expected heterozygosity at each of the ten loci. There were observable differences amongst the populations in terms of the number of alleles, this was in part due probably to several loci not being successfully genotyped in the historic population and to the small sample size of sub-group C (N = 5).

Table 4.6 Number of alleles (*NA*), allelic richness (*AR*), effective number of alleles (*Ae*), observed heterozygosity (H_O), individual heterozygosity (H_I) and expected heterozygosity (H_E) by loci for the contemporary and historic populations and subgroups of gorillas

Locus	*NA*	*AR*	*Ae*	H_I	H_O	H_E
Contemporary Population (N = 64)						
VWf	7.00	7.00	5.10	-	0.97	0.80
D4s1627	8.00	7.96	6.01	-	0.97	0.83
D8s1106	7.00	7.66	3.70	-	0.97	0.73
D10s1432	7.00	9.79	4.77	-	0.98	0.79
D2s1326	10.00	9.88	7.24	-	0.97	0.86
D5s1470	12.00	12.80	4.73	-	0.92	0.79
D7s2204	7.00	8.63	5.21	-	0.95	0.81
D7s817	9.00	8.89	5.46	-	1.00	0.82
D1s550	7.00	7.98	4.52	-	0.95	0.78
D16s2624	5.00	6.71	3.15	-	0.84	0.68
Mean	**7.90**	**8.73**	**4.99**	**0.95**	**0.95**	**0.79**
Historic Population (N = 76)						
VWf	7.00	7.00	4.87	-	0.97	0.80
D4s1627	8.00	7.74	5.26	-	0.90	0.81
D8s1106	8.00	7.37	4.67	-	0.97	0.79
D10s1432	10.00	9.31	5.59	-	0.99	0.82
D2s1326	10.00	9.91	7.20	-	0.97	0.86
D5s1470	13.00	12.81	8.82	-	1.00	0.89
D7s2204	9.00	7.65	4.43	-	0.96	0.77
D7s817	9.00	8.83	6.00	-	0.93	0.83
D1s550	8.00	7.60	4.66	-	0.97	0.79
D16s2624	7.00	5.92	3.01	-	0.89	0.67
Mean	**8.90**	**8.41**	**5.45**	**0.96**	**0.96**	**0.80**
Historic Sub pop A (N = 45)						
vWF	6.00	4.53	4.34	-	0.96	0.77

Continued from previous page

D4s1627	6.00	4.59	5.19	-	0.88	0.81
D8s1106	6.00	4.17	4.67	-	0.96	0.79
D10s1432	9.00	4.66	5.09	-	0.98	0.80
D2s1326	9.00	5.38	6.68	-	1.00	0.85
D5s1470	12.00	5.87	8.44	-	1.00	0.88
D7s2204	7.00	4.39	3.99	-	0.98	0.75
D7s817	9.00	4.54	6.47	-	0.95	0.85
D1s550	8.00	4.16	4.68	-	0.96	0.79
D16s2624	5.00	3.36	2.96	-	0.91	0.66
Mean	**7.90**	**4.56**	**5.25**	**0.94**	**0.96**	**0.79**
Historic Sub pop B (N = 26)						
Vwf	7.00	2.80	5.16	-	1.00	0.81
D4s1627	7.00	3.79	5.09	-	0.95	0.80
D8s1106	6.00	3.60	4.27	-	1.00	0.77
D10s1432	8.00	3.79	5.32	-	1.00	0.81
D2s1326	10.00	5.36	7.20	-	0.96	0.86
D5s1470	12.00	2.98	9.00	-	1.00	0.89
D7s2204	8.00	3.76	4.58	-	1.00	0.78
D7s817	7.00	4.73	5.19	-	0.88	0.81
D1s550	7.00	5.00	4.43	-	1.00	0.77
D16s2624	6.00	3.58	3.06	-	0.92	0.67
Mean	**7.80**	**3.94**	**5.33**	**0.89**	**0.97**	**0.80**
Historic Sub pop C (N = 5)						
Vwf	3.00	4.33	2.38	-	1.00	0.58
D4s1627	4.00	4.48	3.33	-	0.80	0.70
D8s1106	4.00	4.21	2.94	-	1.00	0.66
D10s1432	4.00	4.69	3.33	-	1.00	0.70
D2s1326	6.00	5.19	5.00	-	0.80	0.80
D5s1470	3.00	5.62	2.63	-	1.00	0.62
D7s2204	4.00	5.19	2.94	-	0.60	0.66
D7s817	5.00	4.81	4.55	-	1.00	0.78
D1s550	5.00	4.18	4.00	-	1.00	0.75
D16s2624	4.00	3.32	2.38	-	0.60	0.58
Mean	**4.20**	**4.60**	**3.45**	**0.88**	**0.88**	**0.68**

Summarising the allelic patterns visually across populations showed that firstly, subgroups A and B were very similar. Secondly, there was a noticeable difference between *NA* of the contemporary population compared with the historic population (Fig. 4.5). The most noticeable and interesting difference was observed with identified private alleles which is discussed later.

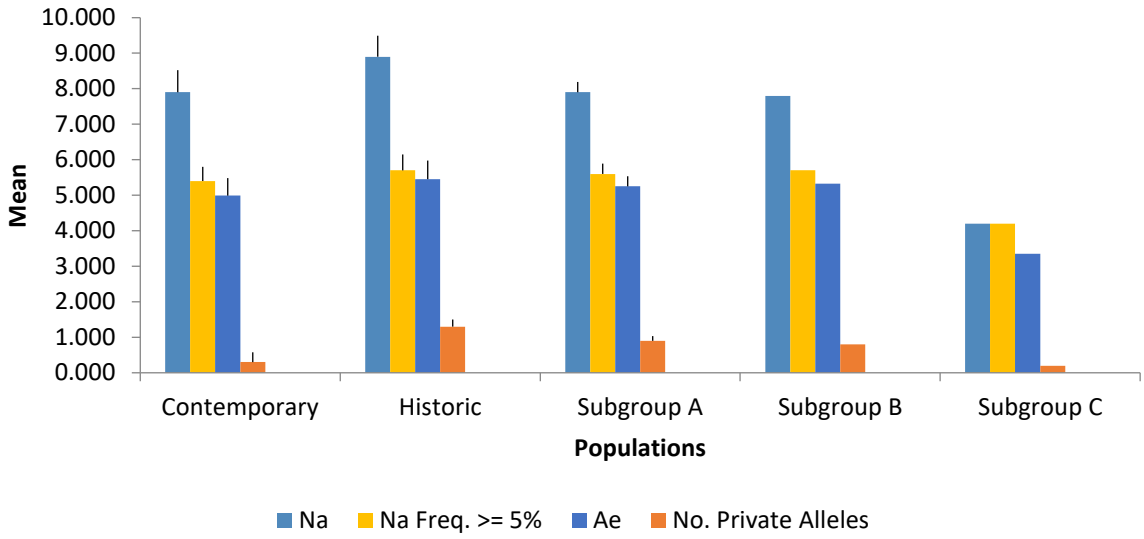


Figure 4.5 Allelic patterns for the contemporary and historic populations of gorillas summarising the number of alleles (*Na*), the number of effective alleles (*Ae*) and the number of private alleles.

Table 4.7 summarises the mean results of the genetic diversity methods and includes effective population size (*Ne*), number of migrants (*Nm*) and the inbreeding coefficient (F_{IS}). These latter three results are discussed later in this chapter. There was an absence of a result from some loci in the historic population due to nonamplification, despite this, the mean number of alleles were higher in the historic population compared with the contemporary population (*Na* = 8.90 and 7.90, respectively). *AR*, which is deemed as the most informative diversity measure and takes into consideration sample size, was higher in the contemporary population tha historic (*AR* = 8.73 and 8.41, respectively), although *Ae* was lower in the contemporary population (*Ae* = 4.99). The mean observed heterozygosity values were similar for all populations except for subgroup C. Likewise, expected heterozygosity were similar, except for subgroup C. Significance using ANOVA was tested for *NA, AR, Ae* and H_E and there were no significant results except for *NA* and H_E among the subgroups. Individual heterozygosity results were similar in the contemporary and historic populations and subgroup A (which had the largest sample size of the subgroups, N = 45). Subgroup B and C were similar in terms of their H_I.

The results for *Theta (ϑ)* (under the stepwise mutation model) showed little differentiation between contemporary and historic populations (ϑ = 2.574 and 2.718, respectively). However, there was some differentiation between subgroups A (ϑ = 2.808) and B (ϑ = 2.664) with subgroup C (ϑ = 2.025).

Table 4.7 Mean results of genetic diversity for populations of gorillas for number of alleles (*Na*), allelic richness (*AR*), effective number of alleles (*Ae*), observed heterozygosity (H_O), expected heterozygosity (H_E) and individual heterozygosity (H_I). Results for the effective population size (*Ne*), number of migrants (Nm), ϑ (under the step-wise mutation model) and the inbreeding coefficient (F_{IS}) are also included.

Population	*N*	*NA*	*AR*	*AE*	H_O	H_E	H_I	*Ne*	*Nm*	ϑ	F_{IS}
Contemporary	64	7.90	8.73	4.99	0.95	0.79	0.95	21.8	-	2.574	-0.205
Historic	76	8.90	8.41	5.45	0.96	0.80	0.96	1171.7	11	2.718	-0.236
Historic sub pop A	45	7.90	4.56	5.25	0.96	0.79	0.94	254.0	7	2.808	-0.210
Historic sub pop B	26	7.80	3.94	5.33	0.97	0.80	0.89	520.6	2	2.664	-0.222
Historic sub pop C	5	4.20	4.49	3.35	0.89	0.68	0.88	Infinity	2	2.025	-0.298

Of the 140 genotypes used in this study, 28 individuals were found to possess private alleles, meaning that 20% of all the individuals genotyped were carrying at least one or more private alleles. The historic population had 25 individuals each possessing one private allele, and the contemporary population had three individuals carrying one private allele. Despite the slight variation in number of individuals between the two populations (N = 76 and 64 for the historic and contemporary populations, respectively), the historic population contained approximately nine times the number of private alleles compared to the contemporary population. Figure 4.6 shows a graphical representation of the private allele distributions among the two populations and identifies the individual gorillas that possessed at least one private allele in their individual genotype.

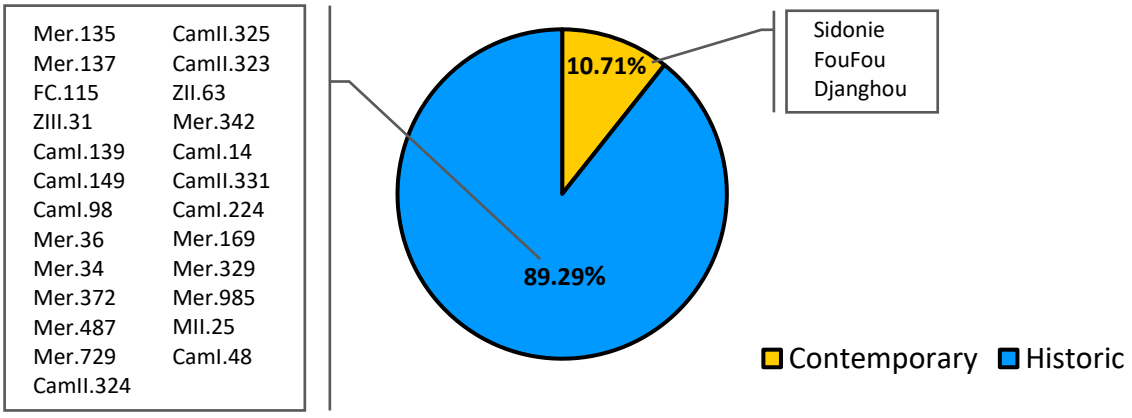


Figure 4.6 Comparison between the contemporary and historic populations of western lowland gorillas in private allele distributions (shown as percentages). The historic population contained 25 individuals with private alleles and the contemporary population contained three.

Of the ten loci used in this research, six had private alleles associated with them (D5s1470, D8s1106, D10s1432, D16s2624 and D7s2204) and four (vWF, D7s817, D2s1326 and D4s1627) did not

have any individuals containing private alleles. Table 4.8 summarises the number of individuals from each population, the number of loci with private alleles and the specific loci.

Table 4.8 Individuals from each population of gorillas with private alleles and the specific loci.

Gorilla ID	*Population*	*No. of loci with private alleles*	*Loci*
Sidonie	Contemporary	1	D5s1470
FouFou	Contemporary	1	D8s1106
Djanghou	Contemporary	1	D10s1432
Mer.135	Historic	1	D10s1432
Mer.137	Historic	1	D10s1432
FC.115	Historic	1	D16s2624
ZIII.31	Historic	1	D5s1470
CamI.149	Historic	1	D10s1432
CamI.139	Historic	1	D10s1432
CamI.98	Historic	1	D10s1432
Mer.36	Historic	1	D10s1432
Mer.34	Historic	1	D10s1432
Mer.372	Historic	1	D5s1470
Mer.487	Historic	1	D10s1432
Mer.729	Historic	1	D7s2204
CamII.324	Historic	1	D1s550
CamII.325	Historic	1	D5s1470
CamII.323	Historic	1	D1s550
ZII.63	Historic	1	D7s2204
Mer.342	Historic	1	D5s1470
CamI.14	Historic	1	D5s1470
CamII.331	Historic	1	D5s1470
CamI.224	Historic	1	D10s1432
Mer.169	Historic	1	D10s1432
Mer.329	Historic	1	D10s1432
Mer.985	Historic	1	D10s1432
MII.25	Historic	1	D8s1106
CamI.48	Historic	1	D16s2624

Adding these results to a geographical representation (Fig. 4.7), showed that the distribution of the private alleles in the historic population was not restricted to any one of the subgroups. However, most (eight of the twenty-five private alleles in the historic population) were found in

subgroup A (N = 45), eight in subgroup B (N = 26) and one in subgroup C (N = 5). One specimen (MII.25) is not shown on the map as geographical coordinates were not available for that individual.

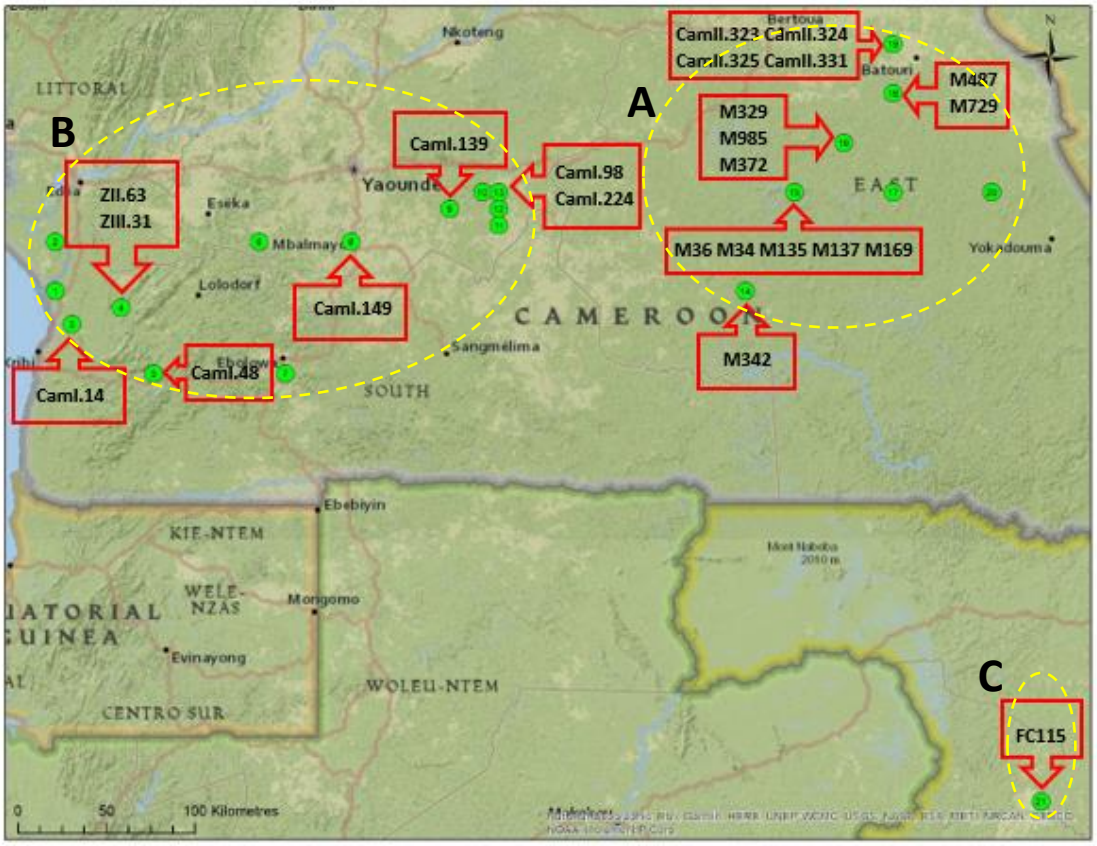


Figure 4.7 The geographic distribution of private alleles for the historic population of western lowland gorillas. A total of 25 gorillas in the historic population were identified as carrying a private allele, however, MII.25 is not shown on map as coordinates were not available. The individuals with private alleles are depicted by the red boxes, and green circles indicate their exact geographical location. The number in the green circles identifies each geographical location and the gorillas at that specified location (Appendix 5).

Within the contemporary population, Sidonie, FouFou and Djanghou were the only three individuals found to carry private alleles. Djanghou is carrying a private allele (158) at loci D10s1432.

4.5.4. Genetic structure

The AMOVA for the two western lowland gorilla populations [contemporary (ASP) and historic (PCM)] plus the two Cross River populations (contemporary and historic) showed an F_{ST} value of 0.114 corresponding to moderate genetic differentiation between the four groups.

Considering the western lowland gorillas only (contemporary and historic), the F_{ST} value was 0.015 corresponding to low genetic differentiation. The greatest variation was observed within

individuals, for both subspecies (western lowland and Cross River) results showed 90% variation (Fig. 4.8.a) and for the western lowland subspecies (contemporary and historic populations) the results showed 99% variation within individuals (Fig. 4.8.b). Among populations variation were 10% for western and Cross River species and 1% for the western lowland subspecies (contemporary and historic).

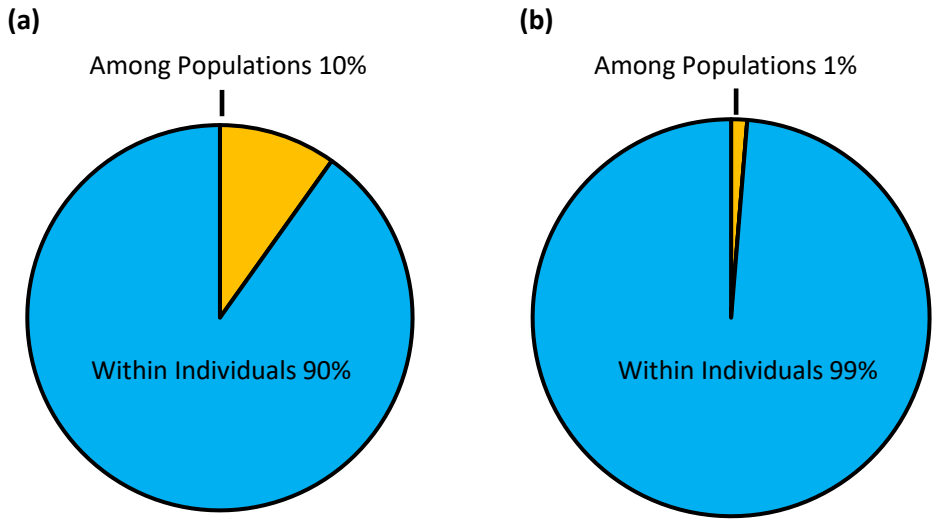


Figure 4.8 Analysis of molecular variance (AMOVA) results shown as percentages for comparisons between (a) the four populations of gorillas consisting of two Cross River gorilla populations (contemporary and historic) and two western lowland gorilla populations (contemporary and historic), and (b) comparison between the contemporary and historic western lowland gorilla subspecies only.

F_{IS} (inbreeding coefficient) results across the four loci (D5s1470, D8s1106, D16s2624 and vWF) for the Cross River gorillas were F_{IS} = -0.045 and -0.105 (for the contemporary and historic populations, respectively), and for the western lowland subspecies were F_{IS} = -0.346 and -0.218 (for the contemporary and historic populations, respectively). However, the results for F_{IS} for the western lowland subspecies (ASP and PCM) were different when the analyses were performed across 10 loci, with the contemporary (ASP) and historic (PCM) populations showing F_{IS} = -0.205 and -0.236, respectively.

The AMOVA results for the historic (PCM) subgroup analysis only provided the same visual representation as the contemporary and historic populations (Fig 4.8.b). The F_{ST} value, however, gave a result of 0.016. This is a difference in F_{ST} of 0.0001 from the contemporary population and indicates little genetic differentiation.

Pairwise F_{ST} analysis considering the subspecies across four loci, showed the most differentiation of F_{ST} = 0.225 between both the contemporary populations (Cross River and ASP), considered as high genetic differentiation. High genetic differentiation was also observed between

the two Cross River gorilla populations (F_{ST} = 0.179). The lowest differentiation was observed between the western lowland gorilla populations (ASP and PCM) with a result of F_{ST} = 0.059.

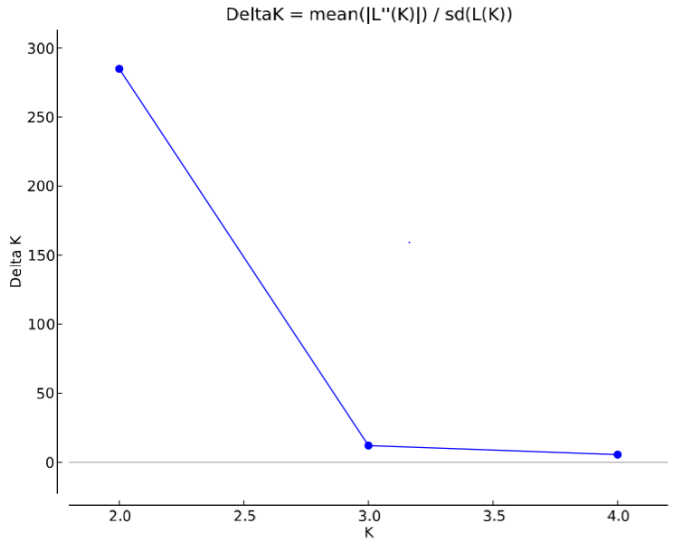
Analyses across all 10 loci between the contemporary population (ASP) and the three historic (PCM) subgroups A, B and C, showed that the highest differentiation was found between subgroup C and the other groups, while differentiation among contemporary and any of the subgroups was comparable with the pairwise F_{ST} values among subgroups A and B.

Table 4.9 Pairwise genetic differentiation (F_{ST}) between subspecies of western gorillas, and among western lowland gorilla populations.

Cross River & western lowland				
	Contemporary (ASP)	Historic (PCM)	CR Contemporary	CR Historic
Contemporary (ASP)	0.000			
Historic (PCM)	0.059	0.000		
CR Contemporary	0.225	0.165	0.000	
CR Historic	0.115	0.093	0.179	0.000
Western lowland				
	Contemporary (ASP)	Sub-group A	Sub-group B	Sub-group C
Contemporary (ASP)	0.000			
Subgroup A	0.014	0.000		
Subgroup B	0.012	0.012	0.000	
Subgroup C	0.038	0.032	0.042	0.000

The first Structure analysis performed included the 140 genotypes from this study and 20 genotypes from contemporary Cross River gorillas and 14 genotypes from historical Cross River gorillas, across four loci: D5s1470, D8s1106, D16s2624 and vWF, to ascertain that the gorillas used in this study (ASP and PCM) were all western lowland gorillas. The Evanno method using Δ*K* detected two clusters, whereas the Puechmaille method detected four or five clusters (Fig. 4.9.a, b).

(a)



(b)

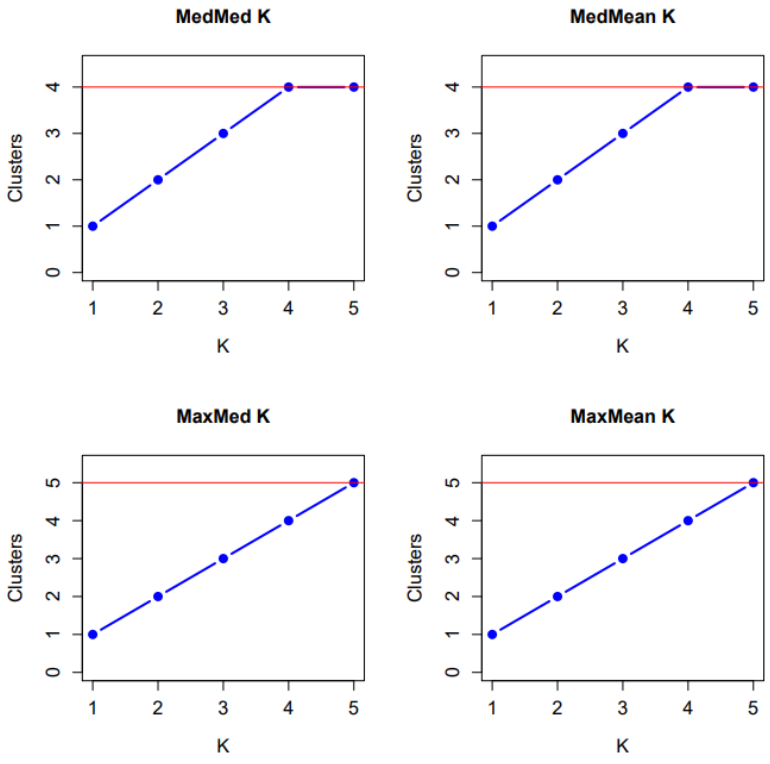


Figure 4.9 Results from Structure analysis of the Cross River gorilla and western lowland gorilla populations, each containing a contemporary and historic population. (a) Structure Harvester results with the Evanno method (Δ*K*) detected 1 or 2 clusters. (b) Structure Selector using the Puechmaille method detected four or five clusters.

Individual probability plots (based on probability of assignment to a *K*-cluster) revealed that the Evanno method effectively identified the two subspecies, separating them into western lowland and Cross River gorillas, whereas the Puechmaille method had identified further substructure within each subspecies (Fig. 4.10.a, b). Both methods confirmed that the ASP and PCM gorillas were western lowland gorillas with no individuals identified as belonging to the Cross River subspecies.

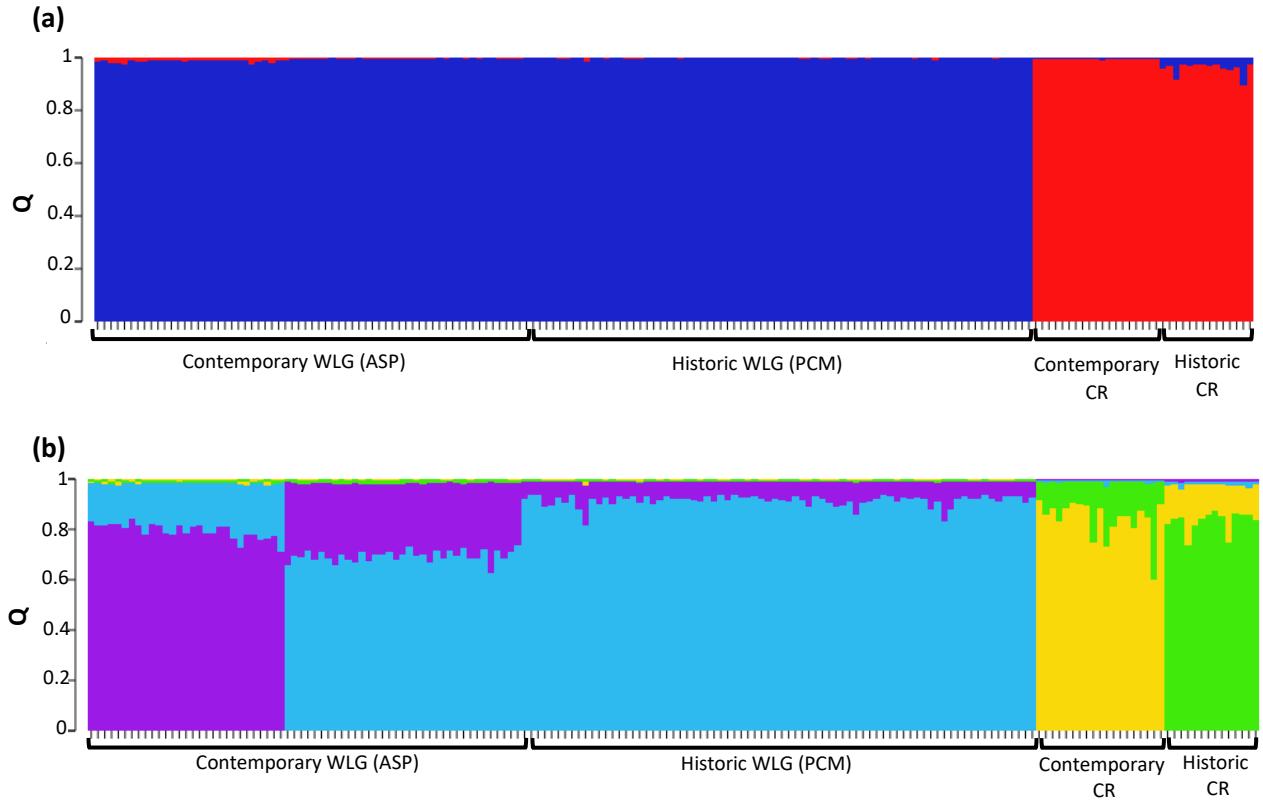
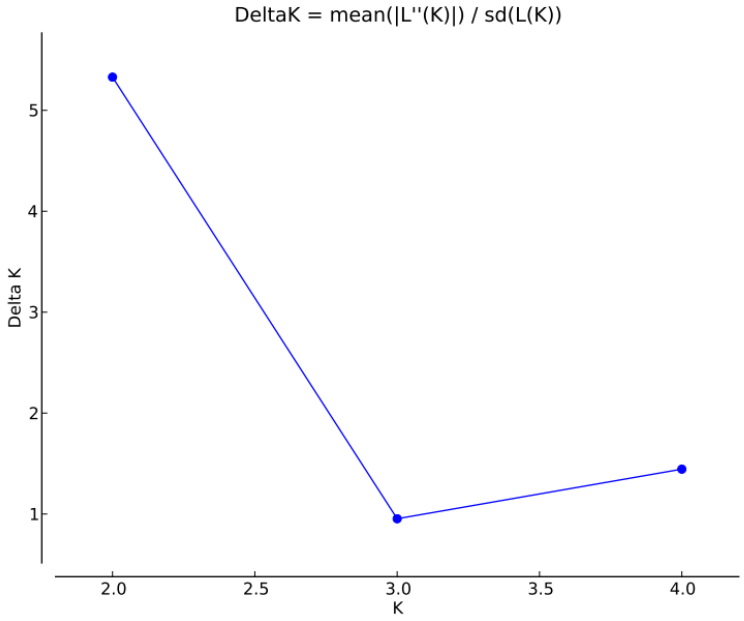


Figure 4.10 Individual probability plots from Structure for the Cross River and western lowland gorilla analysis showing (a) the Δ*K* method identifying two clusters corresponding to the western lowland and Cross River subspecies, and (b) the Puechmaille method identifying four or five clusters (four shown) corresponding to the two subspecies plus additional substructure clusters in contemporary Cross River and in Contemporary (ASP) gorillas.

Having established that all gorilla genotypes generated from this research were all western lowland gorillas, the Structure analysis was repeated to include only the western lowland gorilla subspecies (ASP and PCM) and data from all 10 loci. The Evanno method indicated that two clusters were present in the data as did the Puechmaille method for all four statistical tests (MedMedk, MedMeak, MaxMedk and MaxMeak) (Fig. 4.11.a, b).

(a)



(b)

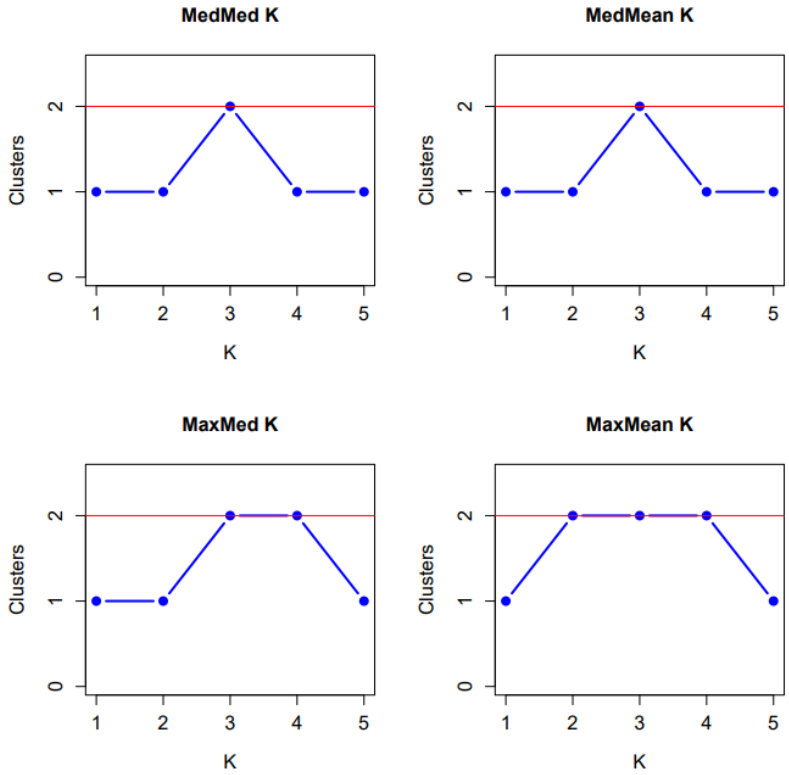


Figure 4.11 Results from Structure analysis of the 140 western lowland gorilla genotypes generated from this study for (a) the Evanno method which indicated that two clusters were present in the data as did (b) the Puechmaille method for all four statistical tests (MedMedk, MedMeak, MaxMedk and MaxMeak).

The visual representation of the data showed that the contemporary gorillas contained more mixed lineage individuals than the historic gorillas. When the data was ordered by *Q*, it did not cluster

readily into historic and contemporary populations, but showed a clear break in the data revealing one smaller cluster which consisted entirely of individuals from the contemporary population, and a much larger cluster which consisted of all the historic individuals plus many of the contemporary gorillas (Fig. 4.12). There were also individuals in both clusters that shared a proportion of probability of belonging to both clusters, here referred to as 'hybrids' for the purpose of this study. Additionally, the most geographically distinct individuals from within the historic population (FC115, FC147, FC130, FC114 and FC124 from the Republic of Congo), did not cluster together, they were dispersed throughout cluster 2, but not one of those five individuals clustered with another one from that same geographical region.

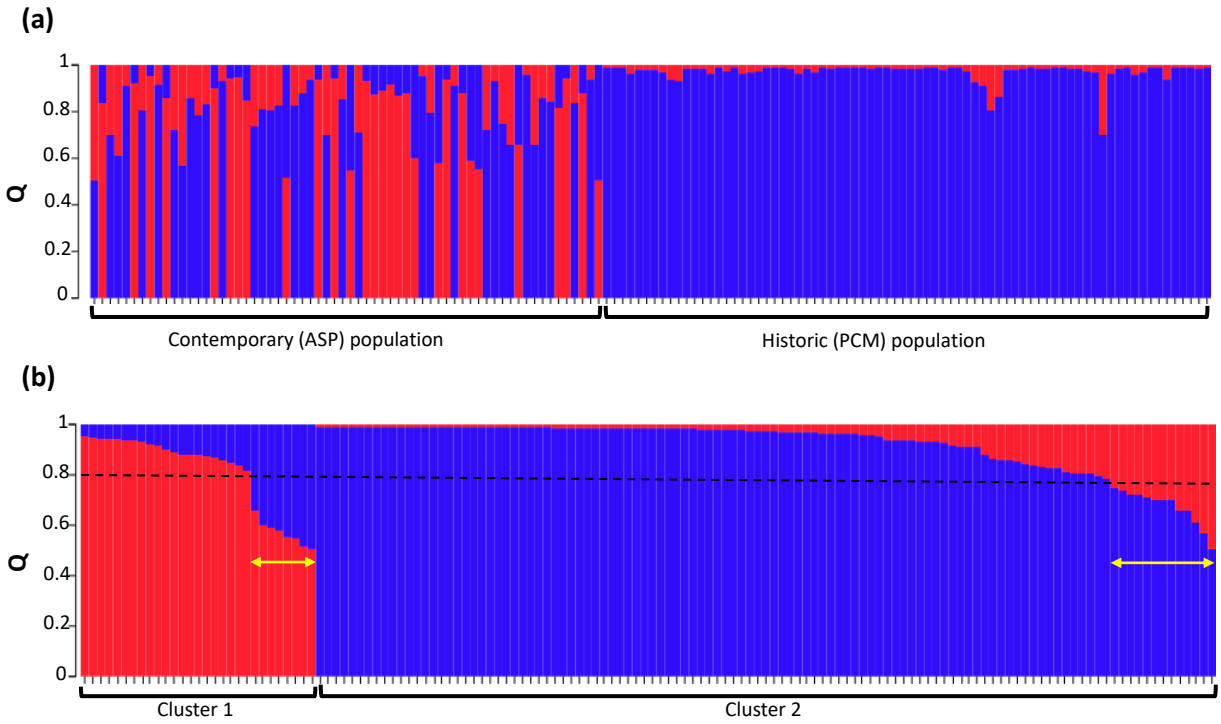


Figure 4.12 Individual probability plots from Structure for the western lowland gorilla. (a) Original order (contemporary and historic populations) of individuals. (b) Individuals ordered by *Q*. Each individual bar represents an individual gorilla. The 140 individuals were divided into two clusters with a clear break in the data. Hybrid individuals are depicted by the yellow arrows (their *Q* proportion fell below 80%). Cluster 1 contained only individuals from the contemporary population whereas Cluster 2 contained all the historical gorillas plus 35 gorillas from the contemporary population.

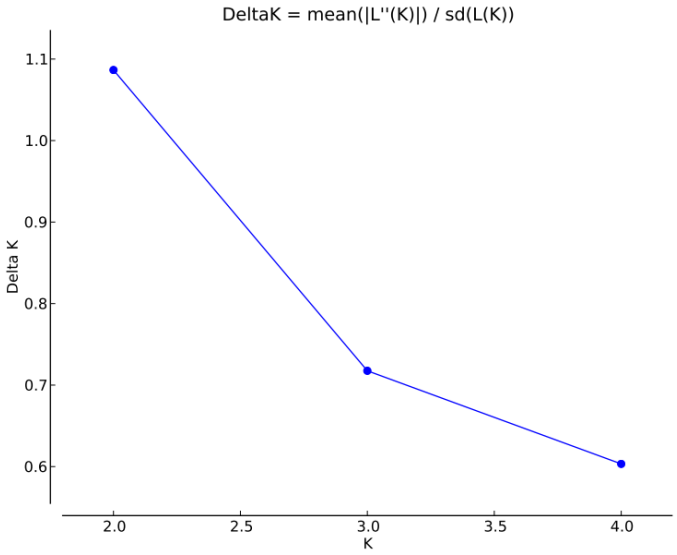
Cluster 1 contained 29 individuals from the contemporary population only. Of those, eight individuals were deemed as 'hybrids' because their *Q* proportion fell below 80% (Q = ≤ 0.8). Cluster 2 contained the remaining 35 individuals from the contemporary population and all 76 gorillas from

the historic population. From this cluster, 13 individuals were identified as 'hybrids' including 12 contemporary gorillas and only one historical gorilla (Mer.95) showing mixed genetic information.

Upon further investigation of Cluster 1 (29 gorillas including hybrids), the cluster mainly contained individuals that are related to three silverbacks (see Fig. 3.2 mtDNA family trees). Cluster 1 included the silverback Kouillou (who was born of wild living parents), his offspring (and theirs), it also includes all the offspring of another silverback, Bitam, however, there was not a sample available to sequence Bitam. Kouillou and his descendants, plus Bitams descendants, make up a high proportion of the individuals in Cluster 1 (21 in total). The remaining eight individuals in Cluster 1, include another silverback, Djanghou [who is a hybrid and the last individual on the plot in Cluster 1 (Fig 4.12.b) with an almost 50/50 split of lineage] and his offspring Kisane (who is known from this research to have sired Kwimba's infants). Additionally, Cluster 1 contained Louna (who was sired by Djala, as was Djanghou) and Kouyou, Oundi and Fubu (who were sired by Kifu). The remaining two individuals in Cluster 1, were Virginika and her infant (sired by Kouillou). Virignika was sired by Ngola and was the only individual in the contemporary population to have been sired by him. Interestingly, Djala resided in Cluster 2 with all the historic population but was one of the 13 contemporary 'hybrids' in that cluster. Djala is an F1 descendant from wild parents, both maternally and paternally. Kifus' (another silverback) genotype was not available as there was no sample for him, but he has sired a total of six individuals that were included in this study, three of them: Kouyou, Oundi and Fubu resided in Cluster 1 but his other three offspring: Kifta, Kangu and Kebu resided in Cluster 2 with the historic population genotypes, although Kangu was one of the 13 hybrids in Cluster 2.

The historical population was subjected to analysis independently to ensure further substructure had not gone undetected. The results from the Evanno method indicated that there were two clusters, not three as had been previously defined on a regional basis. The Puechmaille method produced mixed results of one or two clusters (Fig. 4.13). However, the ΔK method is unable to produce a result of $K = 1$, therefore, in terms of subgroup analysis, the interpretation was one or two clusters.

(a)



(b)

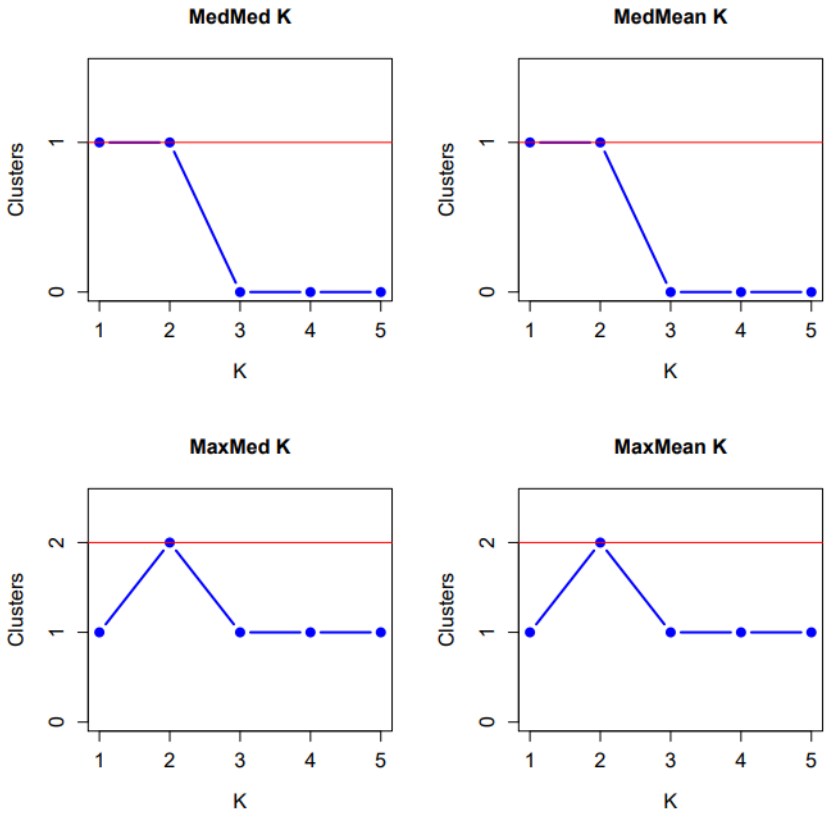


Figure 4.13 Results from Structure analysis of the 76 historical western lowland gorilla genotypes generated from this study for (a) the Evanno method which indicated that two clusters were present in the data as did (b) the Puechmaille method for all four statistical tests (MedMedk, MedMeak, MaxMedk and MaxMeak).

The visual representation of the historical population data analysis performed in Structure Plot (Fig. 4.14), confirmed no substructure within the historical population and therefore, should be considered as one cluster. There was no clear break in the dataset, which is indicative of no genetic substructure. Figure 4.14 shows the historical population in its original order, i.e. in regional groups (not arranged by *Q* values). This result, plus Evanno's *ΔK* and Puechmaille methods, indicated that there was no regional genetic variation amongst the historic population, despite the historic population also containing known related/family groups (e.g. Mer 135, 136 and 137) as well as larger family groups.

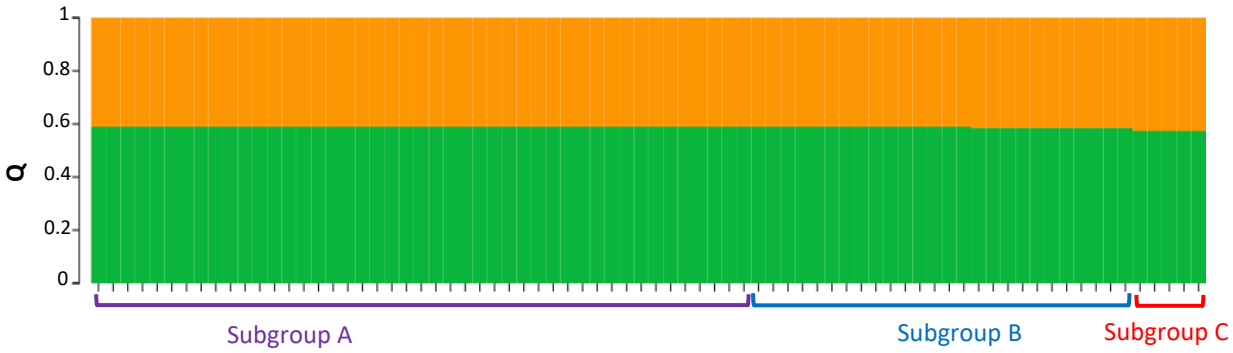
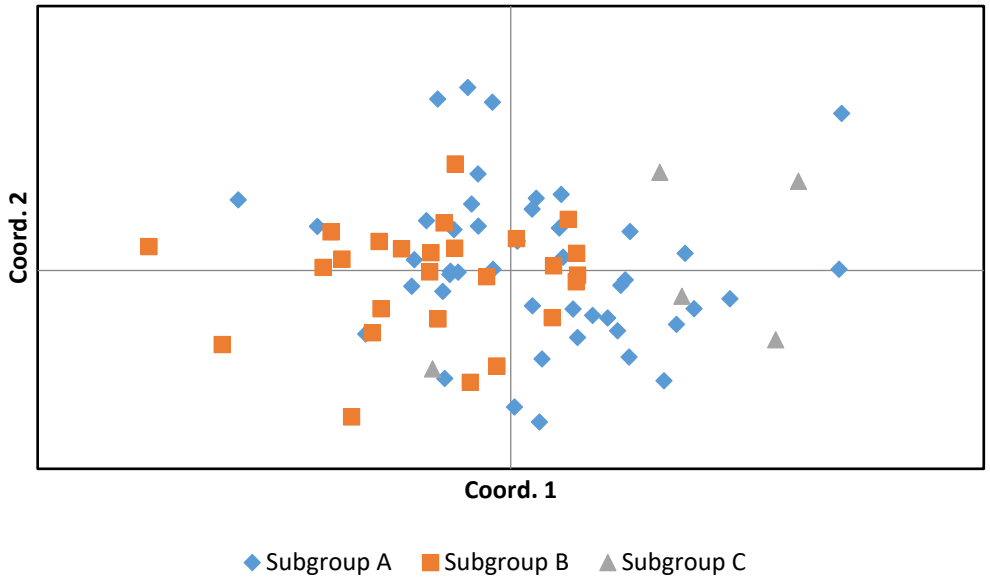


Figure 4.14 Individual probability plot of the 76 historical western lowland gorillas ordered by their original defined geographical regions (A, B, and C). This visual representation is indicative of no substructure. There is a slight observable difference in *Q* of subgroup C, but it does not indicate regional substructure as there is not a clear break in the data as observed in previous analyses.

Principal Coordinates Analysis (PCoA) was initially performed on the historical population data set separated into their regional subgroups. The results (Fig. 4.15.a) indicated that there was no genetic substructure, consistent with the Bayesian Structure analysis. The five individuals from subgroup C were dispersed throughout the coordinates and did not cluster together, and the three regions in general were admixed with no clearly defined clusters or groupings.

PCoA for the contemporary and historic populations showed in general there was not any distinct clustering between the historic and contemporary populations, and the Congo individuals were dispersed and not clustered together. However, there was a genetically distant cluster (Fig. 4.15.b) of individuals all from the contemporary population, except for one gorilla from the historic (PCM) population. These individuals were primarily belonging to Cluster 1 identified from the Structure analysis; thus, the analyses appear to complement each other.

(a)



(b)

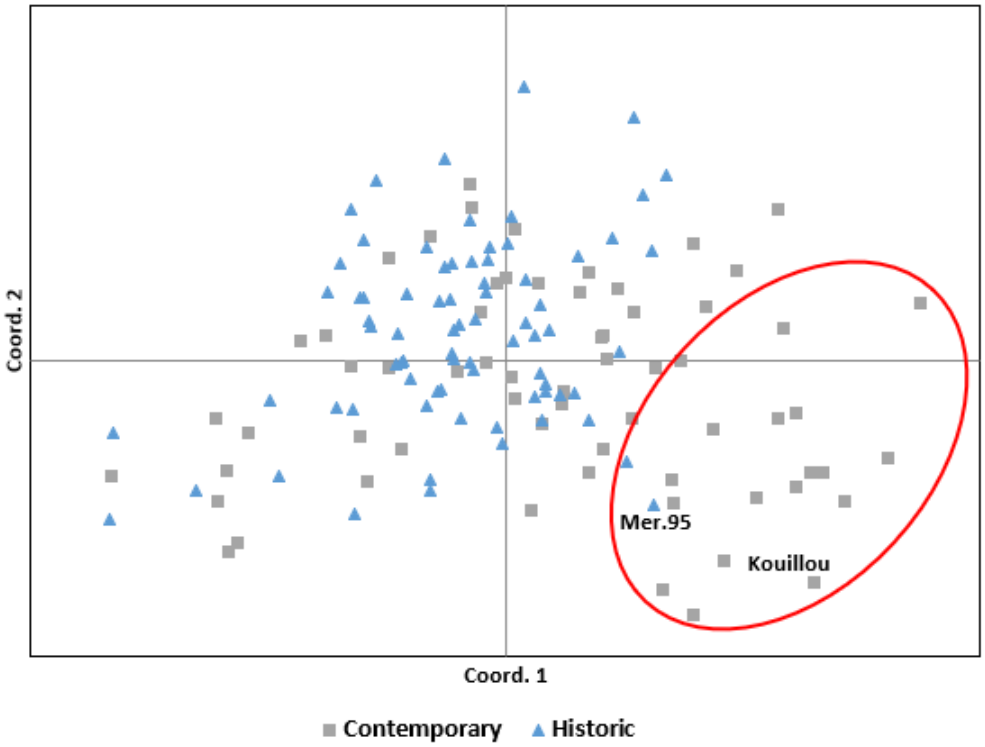


Figure 4.15 Principal Coordinates Analysis (PCoA) for (a) the historical subgroup analysis showing no apparent substructure within the population, with the five gorillas from Congo (group C) not clustering together, and (b) PCoA for the historic and contemporary gorillas in the same analysis. There is a genetically distant cluster depicted by the red oval which includes individuals from Cluster 1 as identified in the Structure analysis. Kouillou has been labelled as he is one of the most genetically distant individuals as well as Mer.95, a gorilla from the historic (PCM) population who clustered with these contemporary individuals.

For further analyses and visualisation purposes, a PCoA was performed separating the data into ASP Cluster 1 (29 contemporary western lowland gorilla genotypes), ASP Cluster 2 (35 contemporary western lowland gorillas), PCM cluster (historic western lowland gorillas), CR contemporary cluster (the Cross River gorilla contemporary population) and CR historic (the Cross River gorilla historic population), as per the results from the Structure analysis. The ASP contemporary Cluster 2 and PCM cluster formed one cluster from the Structure results, but they were deliberately separated for this analysis to visualise the genetic distance (Fig. 4.16). A pairwise F_{ST} of these clusters showed ASP Cluster 2 and the PCM cluster to contain the lowest genetic differentiation (F_{ST} = 0.012) and the highest genetic differentiation was observed between ASP Cluster 1 and CR contemporary (F_{ST} = 0.137) followed by the two Cross River gorilla clusters (contemporary and historic) with a pairwise F_{ST} value of F_{ST} = 0.119 (Table 4.10).

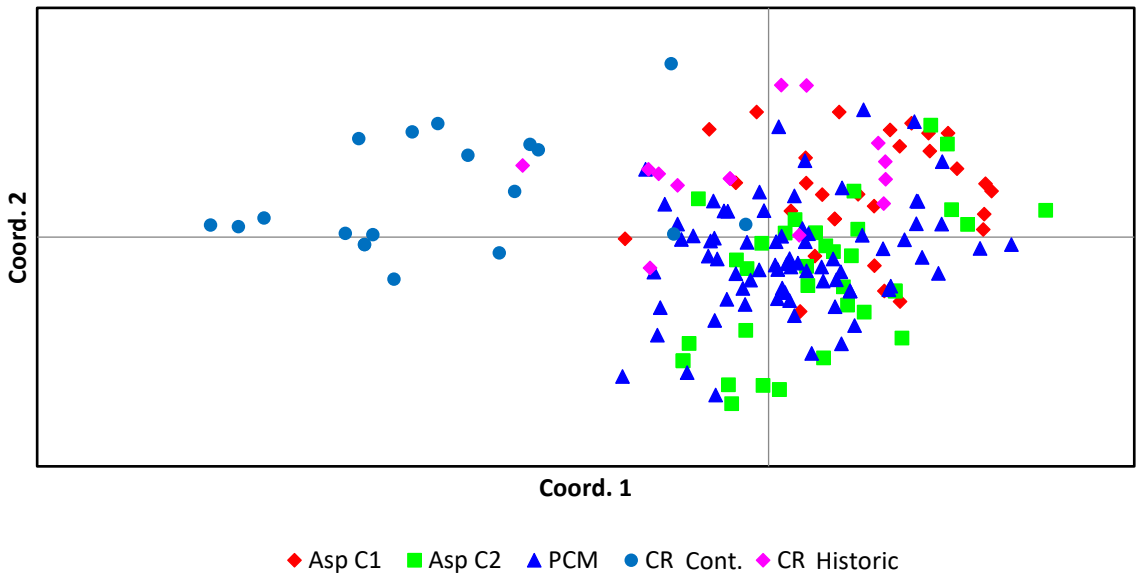


Figure 4.16 Principal Coordinates Analysis (PCoA) of all 140 gorillas from the Aspinall Foundation (ASP) and Powell-Cotton Museum (PCM) plus an additional 34 Cross River individuals (from Thalmann *et al*. 2011; Arandjelovic *et al*. 2015) across four loci. The cluster groups identified were based on the Structure results where the (ASP) contemporary population was divided into two clusters. ASP Cluster 2 and the PCM cluster shown here formed one cluster in Structure analysis but were divided here for visual purposes. The CR (Cross River gorilla) contemporary population is notably the most genetically distant cluster whereas the CR historic cluster shows considerable genetic similarity primarily with ASP Cluster 1, but also ASP Cluster 2 and the PCM cluster.

Table 4.10 Pairwise genetic differentiation (F_{ST}) among gorilla clusters identified in Structure analysis between subspecies and populations.

	ASP cluster 1	ASP cluster 2	Historic (PCM)	CR Contemp.	CR Historic
ASP cluster 1	0.000				
ASP cluster 2	0.035	0.000			
Historic (PCM)	0.040	0.012	0.000		
CR Contemporary	0.137	0.115	0.100	0.000	
CR Historic	0.071	0.069	0.065	0.119	0.000

4.5.6. Genetic bottlenecks, effective population size and migration

To assess whether either the contemporary or historic population had undergone genetic bottleneck, all three mutation models were run (IAM, SMM and TPM) (Table 4.11).

Table 4.11 Genetic bottleneck results for the contemporary and historic gorilla populations under three mutation models.

	Sign test			Standardised diff. test			Wilcoxon test		
Population	IAM	TPM	SMM	IAM	TPM	SMM	IAM	TPM	SMM
Contemporary	**0.005**	0.153	0.149	**<0.001**	0.059	0.322	**<0.001**	0.053	0.097
Historic	**0.006**	0.145	0.408	**<0.001**	0.095	0.463	**<0.001**	0.053	0.423

Significant results shown in bold

Only the IAM in all three models gave significant results but this model is not deemed appropriate for this dataset, so it was disregarded. The SMM and TPM with both the Standardised Differences test and the Wilcoxon's test did not produce any significant results. Given the SMM is the more conservative of the two models (SMM and TPM), and neither have produced any significant results, this data indicates that neither the historic nor contemporary population have undergone a population bottleneck severe enough to be detected genetically.

The effective population size (*Ne*) results of the linkage disequilibrium method are reported. This method was selected as the most suitable for this dataset. The estimated *Ne* for the contemporary population was 21.8 whereas the historic population had an estimated *Ne* of 1171.1 individuals. The three subgroups A, B and C produced estimated *Ne*'s of 254.0, 520.6 and 'infinity' (not calculable) number of individuals, respectively.

The number of migrants (*Nm*) were obtained from analysis in GeneClass for the historic population data separated into its three regional subgroups (A, B and C). The regional analysis identified a total of 11 first-generation (F1) migrants: M264, M139, M470, M387, CamII.331, M691, M532, CamI.14, CamII.224, FC115 and FC147. The visualisation of the identified migrants from the subgroup analysis and their geographical locations are shown in Figure 4.17. Of the 11 individual migrants identified, the majority of them were found in subgroup A (7 individuals: CamII.331, M532, M387, M139, M691, M470, M264) and four further migrants in subgroups B and C (two in each region).

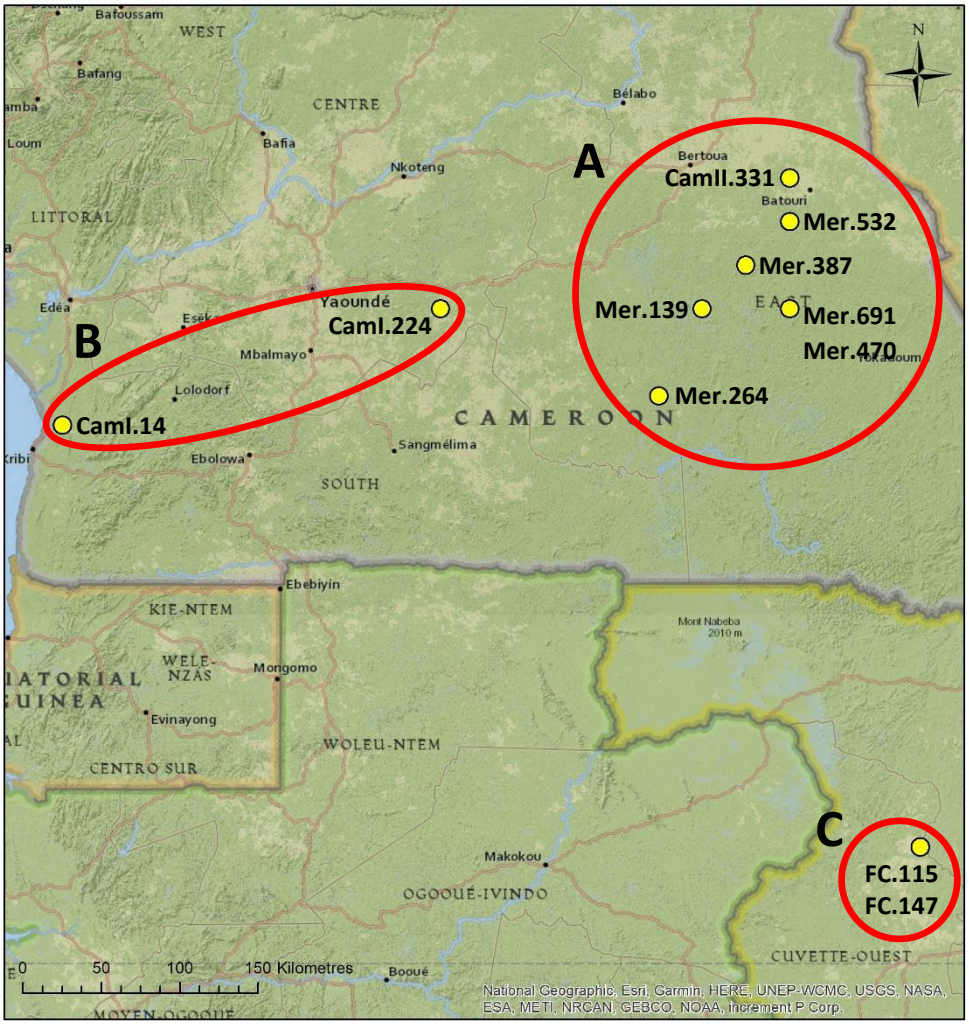


Figure 4.17 Identified migrant gorillas depicted by the yellow circles from each geographic region. Subgroup A contains seven migrant gorillas whereas subgroups B and C both contain two migrant gorillas each.

4.6. Discussion

The aims of this chapter were to study the population genetic diversity and structure of western gorillas. Specifically, it was intended to ascertain whether there was any geographical difference in genetic diversity amongst the historical western lowland gorilla population, to confirm whether the contemporary population consisted only of western lowland gorillas, and to identify potential genetic issues in relation to the conservation of this critically endangered sub-species. Furthermore, the microsatellite genotyping data could also be indicative of parentage within the captive population and could be used to hypothesise the origin of samples without geographical information.

4.6.1. Reliability of the data and pre-analysis observations

The genotyping data was found to be free of null alleles and allelic drop out. This coupled with repeated PCRs and genotyping of all homozygote loci, at least a 25% of the entire dataset being subjected to repeated PCRs and genotyping, as well as the inclusion of family genotype checks, revealed the data to be of reliable quality.

In terms of parentage analysis, the discovery of Djanghou (a silverback in the contemporary population) not being the sire to Kwimba's two offspring was an important finding for the Aspinall Foundation. It confirmed the sexual maturity of his offspring Kisane, who was determined as the biological sire of both the offspring. Typically, in gorilla family groups, the dominant silverback is the only male to breed (Forcina *et al.* 2019). This result would have informed the recommendation that Kisane should be considered for relocation to another group. However, during this study, and before this result was known, Kisane was removed from the group on 27th March 2019 and was quarantined until 30th April 2019 where he was then translocated to Mogo Wildlife Park in Australia (V. Mathieson pers. comm 2020). Nonetheless, this genotyping result validates the decision to translocate Kisane to a different gorilla group.

Common to most population genetics studies, one of the first few analyses performed is to test whether populations are in HWE and to test for LD (Hamilton 2011). These tests were performed, and deviations from the HWE were found and some loci were in LD. It is rare to find natural populations with whole genotypes in HWE, most populations are under natural selection at least, thus violating at least one of the assumptions (Hartl & Clark 1997; Hamilton 2011). These results were not unexpected because deviations from HWE and LD have been reported in previous studies regarding gorillas, e.g. Clifford *et al*. (2003), Bergl & Vigilant (2007), Bergl *et al*. (2008), Simons *et al*. 2013, Fünfstück & Vigilant (2015). Inclusion of family groups in the dataset is likely to have caused the deviations in this research just as they have done in previous literature (Bergl & Vigilant 2007; Bergl *et al*. 2008; Simons *et al*. 2013). The contemporary population is well documented and family

relationships amongst the dataset are known except for those individuals who were born to wild parents. Removing individuals from analysis, however, would affect the overall assessment of genetic diversity in captive or endangered populations with very few individuals available for genotyping. Furthermore, the non-random association of alleles in a few loci in the global analysis of LD is likely not to be due to physical linkage (Bradley et al. 2000; Clifford *et al*. (2003); Bergl & Vigilant (2007); Bergl *et al*. (2008); Simons *et al*. 2013; Fünfstück & Vigilant (2015).), but likely due to genetic drift, inbreeding, selection and/or gene flow. If microsatellites are assumed to be neutral markers, the most likely explanations for LD observed here would be genetic drift acting on small populations, inbreeding and possible historical levels of gene flow among distant populations. LD tested on the historical group produced only one significant result (data not shown), likely reflecting the effects of inbreeding when including the captive animals. Nonetheless, LD affecting analyses would result in an overstatement of population genetic structure, which was not the case in this study where only weak genetic structure was found in the western lowland gorilla. Deviations from the HWE are increased for populations considered to be partially inbred (Wang *et al.* 1998), the captive data did reveal a higher level of inbreeding than the historic, but neither population was considered to be inbred, however, this could account for the deviations observed. The Structure program generates clusters according to LD and HWE deviations caused by admixture of populations, thus, the presence of LD improves clustering results, but overestimation of clusters can occur if LD is 'strong' or if HWE departure is present (Falush *et al.* 2003; Kaeuffer *et al*. 2007).

Unusually for a museum collection, the gorilla specimens in the Powell-Cotton Museum also contain considerable detail in relation to family groups, temporal information, geographical locations and other biological information. For example, the largest group of historical specimens captured at the same location in this dataset consisted of 14 individuals, nine females (M36, M170, M29, M35, M58, M95, M136, M138 and M139) and five males (M59, M34, M135, M137 and M169), all hunted in the Batouri and Lomie region, the colonial region of the Cameroons (03.35°N and 13.45°E), over a 9-month period from 26th March (specimen M29) through to 10th December of 1935. The second largest grouping consisted of 12 individuals, seven females (M799, M840, M841, M855, M865, M877 and M532) and five males (M471, M505, M720, M487 and M729), again from the Batouri region, in the colonial region of the Cameroons but at 04.15°N and 14.15°E. These individuals were hunted over a longer period than the previous group spanning 20 months from the first capture on 2nd August 1932 (M487) to 26th April 1934 (M877). Other groups at the same location span even longer hunting trips, for example: ZII.63, ZII.64, ZIII.31 and ZVI.33 were all captured at 03.10°N and 10.20°E but over a duration of nearly three years, from 14th May 1930 to 13th March 1933. The seven individuals from the SE of Kribi, Cameroon (CamI.41, CamI.42, CamI.43, CamI.44, CamI.45, CamI.46 and CamI.48) at a location of 02.50°N and 10.30°E consisted of three females and four males, all

captured on the same day, 29th March 1929. It would be a fair to assume that this group is one family group given the social structure of western lowland gorillas, coupled with the extensive detail of the contextual information. For example, the five French Congo gorillas used in this dataset (FC.114, FC.115, FC.124, FC.130 and FC.147) hunted between 26th April and 2nd June 1927, at a location of 00.40°N and 15.30°E, includes a male (FC.115) and a female (FC.114) that were both captured on the same day (26th April). It could be assumed that these two individuals were part of the same family group, however, the following two inserts from Major Powell-Cottons field notes prove otherwise:

114. Female yg *Taken near the village of Mambili 0' 40'N. 15' 30'E.*
Height 43". Span 63". Chest 36". Hand 7¼" x 3½". Foot 9" x 3¾". Weight clean with heart, lungs and skin 60lbs. Stomach much larger than chest. One of two or more.
115. Male yg *Taken near the village of Mambili 0' 40'N. 15' 30'E.*
Height 42". Span 82". Chest 43". Hand 9¼" x 4¼". Foot 11¼" x 4¾". Weight clean but with heart, lungs and skin 94lbs. Not same family as 114.

Given the evidence from previous research regarding deviations from the HWE and LD, this research is not invalidated due to the inclusion of family groups and there was justification to continue with further analysis following the literature (Lukas *et al*. 2004; Bergl *et al*. 2008; Simons *et al*. 2013). The removal of all family groups would not have been possible as it would have resulted in the dataset becoming unfeasibly small.

4.6.2. Species and subspecies determination and population structure

This study was the first to perform a genetic analysis of this scale on the contemporary population at the Aspinall Foundation and the Powell-Cotton Museum. Thus, confirmation that the gorillas in the Aspinall Foundation captive population are all western lowland gorillas, has not been previously confirmed via genetic analysis. The Eastern gorilla species are notably distinguished morphologically from the western species as they are larger and have longer hair than western gorillas (Caillaud *et al*. 2008), whereas the two-western subspecies (western lowland and Cross River) are less easily distinguished due to their very similar morphological and phenotypic appearance (Sarmiento & Oates 2000). It is important to confirm the taxonomic identity and that the individuals are not of mixed subspecies because conservation and breeding programs within captivity operate at the subspecies level for primates (Lindburg *et al*. 1984). The Cross River gorilla is also the most critically endangered of the gorilla subspecies with less than 250 individuals in the wild (IUCN 2019); therefore, subspecies confirmation of captive individuals is of utmost importance.

To test for subspecies identification, the 140 gorilla genotypes generated in this study were analysed with an additional 34 Cross River gorilla genotypes, 20 from a contemporary (wild)

population (Arandjelovic 2015) and 14 from a historic (museum) population (Thalmann *et al*. 2011). The population structure results based on Bayesian methods confirmed all 140 gorilla genotypes generated from this study (contemporary and historic populations) belonged to the western lowland subspecies. Different methods were employed to check the optimal number of *K*-clusters. The Δ*K* method identified two clusters (Fig. 4.10.a) and the Puechmaille method identified four or five clusters (Fig. 4.10.b), corresponding to the two subspecies plus additional substructure within the clusters. The Δ*K* method is the most commonly applied method used to detect population structure (Li & Liu 2018), however, this method tends to only detect the uppermost hierarchical structure which could lead to a misinterpretation of population structure (Puechmaille 2016; Li & Liu 2018). Relatively recently, Puechmaille (2016) developed four alternative statistics (MedMedk, MedMeak, MaxMedk and MaxMeak) with results that suggest these methods detect hierarchical substructure more accurately (Puechmaille 2016; Li & Liu 2018). The results obtained here support the finding that the Puechmaille method does appear to detect further hierarchical substructure than the Δ*K* method. Hence, both methods were applied to all population structure analyses to ensure all hierarchical substructure was detected. Additional subsequent analyses of subsets were also performed in Structure to determine any hidden within-group substructure, as recommended in the literature (Evanno *et al*. 2005). Evanno *et al*. (2005) additionally recommended that the Δ*K* method should not be used exclusively to detect population structure, therefore, there were justifiable reasons to apply both methods.

The two methods confirmed that the contemporary and historical genotype data generated from this study were all western lowland gorillas by identifying the Cross River gorillas as belonging to one cluster and the western lowland gorillas as a separate cluster. Removing the Cross River gorillas and performing further population structure analyses showed (by both Δ*K* and Puechmaille methods) that the western lowland gorillas contained two subclusters but they were not defined by contemporary and historic populations, but consisted of one small cluster of 29 individuals all from the contemporary population, and one much larger cluster of 111 individuals which contained all of the historical gorillas plus 35 of the contemporary population (Fig. 4.12).

Additionally, the historic population was subjected to further analysis as a sub dataset to ensure additional (perhaps regional) structure had not been overlooked by the initial analysis of all 140 genotypes. The results from the analysis of historic population (Fig. 4.14) confirmed that it did not contain any further substructure. Thus, it can be concluded that no geographic regional structure was identified in the historic population. The *ΔK* method is not able to find the best *K* if the actual (real) *K* = 1 (Evanno *et al*. 2005). Therefore, a visual representation as the one shown in Figure 4.14 is what would be expected for a population showing no substructure. Additionally, there was the

same visual representation from the Puechmaille method of the western lowland gorillas which was performed with the Cross River gorilla analyses (Fig. 4.10.b).

The absence of regional genetic clustering of the historic population in this study, was an unexpected result. Although the specimens were primarily from Cameroon (there were not any specimens from Gabon), they were geographically widespread, thus, observing just one genetic cluster in the wild as opposed to two found in the captive population is of interest. There remain over 70 skin samples that can be analysed from the historic population, and it would be of interest to genotype the remaining samples to ascertain whether the historic population does contain further lineages not identified in this current dataset, which is a possibility. One hypothesis could be that the Cluster 1 individuals may have come from lineages further north in Cameroon than those sampled from the historical collection or in the north central region where specimens were also absent. Although the Sangha River separates the Cross River and western lowland gorillas, the western lowland gorilla distribution does extend further north than the regions sampled in this study in the historic population and there was a notable absence of specimens in the central regions of Cameroon. Mer.95 (the only hybrid historical gorilla) is located on the outskirts of region A close to the central region where specimens are absent. Perhaps Mer.95 was part of a separate Cameroon lineage from this central/norther region, and perhaps this is the region the other Cluster 1 individuals in the contemporary have originated from. The PCoA (Fig. 4.16) does support this hypothesis somewhat, given the ASP Cluster 1 individuals appear to be more genetically similar to the historical Cross River population than the ASP cluster and PCM group are, and it would indicate that the ASP Cluster 1 was geographically closer allowing gene flow before the Cross River gorillas became entirely genetically distinct. Cross River gorillas diverged from their ancestral population ~17, 800 years ago, however, Thalmann *et al.* (2011) observed substantial gene flow between the two western subspecies that continued after initial divergence, and ceased to continue approximately 420 years ago which coincided with the population bottleneck they detected in the Cross River gorillas. The demographic events observed by Thalmann *et al.* (2011) within the Cross River subspecies are consistent with the climate change scenario that has occurred over the last tens of thousands of years. A scenario which has led to forests repeatedly expanding and contracting, thus facilitating gene flow between the western gorilla species at periods of forest expansion as well as population expansion, and then the isolation of the Cross River species and reduction in population at periods of forest contraction, the latter has been compacted by anthropogenic impact.

The western gorilla species historically had an almost continuous distribution ranging from southern Central African Republic to the Congo River and to the coast in the west (IUCN 2019). The Sangha River is the barrier between the two western subspecies and habitat fragmentation has played an increasing role in the fragmentation of populations for this species (IUCN 2019). The

northern limit for the western lowland gorilla is the boundary where forest habitat becomes savannah and the Sanaga River is the north western boundary (IUCN 2019). Therefore, it is plausible that this lineage may be from closer proximity to a boundary line with Cross River gorillas. If so, this would explain the observed similarity is genetic distance from the PCoA and account for the 'hybrid' individuals present in both clusters.

However, caution must be given to the interpretation of the clusters observed due to deviations from the HWE and LD. It is possible that the two clusters observed from the Structure analysis, are a result of the inclusion of family members in the dataset, thus, family genetic structure is being identified rather than population genetic structure. The 29 gorillas in Cluster 1 were all from the captive population and primarily included two silverbacks (Kouillou and Djanghou plus their desecenants) and Bitams desecendants, but it did also include individuals unrelated to them.

Bayesian clustering methods, which the Strutcure program employs, infers population structure by minimising HWE and LD within subpopulations (Rodríguez-Ramilo & Wang 2012). An investigation into the effect of the inclusion close relatives in Bayesian clustering programs was performed by Rodríguez-Ramilo & Wang (2012) additionally, Pritchard *et al*. (2000) has previously warned that the number of clusters could be overestimated by Structure when close relatives are included in the dataset. Guinard *et al.* (2006) and Anderson & Dunham (2008) both found that an overestimation of *K* was found by the Structure program when data included closely related individuals. Anderson & Dunham (2008) noted that the overestimation of *K* is most apparent when the sample contains large groups of full siblings.

Both the contemporary and historical populations used in this research are known to include related individuals and there is software available such as Colony (Jones & Wang 2010) and ML-Relate (Kalinowski *et al.* 2006) which can detect related individuals. The ideal scenario would be to remove all known related individuals from both of the populations, however, as discussed previously, the removal of such individuals would leave the dataset unfeasibly small and other studies such as Clifford *et al*. (2003), Bergl & Vigilant (2007), Bergl *et al*. (2008), Simons *et al*. 2013, Fünfstück & Vigilant (2015) also included known related individuals and observed deviations in HWE and LD.

The inclusion of related individuals, however, cannot be dismissed and therefore caution must be exercised when interpreting the Structure results. It is possible that the two clusters observed represent genetic family structure, rather than population genetic structure; however, if that were the case, the expectation would be that the historical population, when analysed independently to detect further substructure, would have presented more than the one observed cluster given that related individuals including entire family groups were included in the data. The results of the Structure analyses also complemented those of previous studies such as Nsubuga *et al*. (2010) who also found two lineages in the North American captive gorilla population.

Additionally, as discussed by Waples & Anderson (2017), the removal/purging of putative siblings can have a deleterious effect on downstream analysis. It is considered 'best practice' and often routinely performed to remove sibling groups from the data (Peterman *et al.* 2016). However, Waples & Anderson (2017) noted at least three significant problems when attempts are made to do so in datasets. The first issue is that siblings do occur naturally in populations, thus, removing sibling groups risks eradication of evolutionary signal, thus, making populations appear larger, or infinitely large. Secondly, it also reduces sample size, which is the issue that would have been encountered in this research if sibling groups had been purged. Thirdly, the methods used to detect sibling relationships have their own limitations, particularly for detection of half-siblings and distant relatives. However, Waples & Anderson (2017) do note that there are some other clustering methods that are not as sensitive to family groups, and do not make assumptions about HWE and LD, both of which are a factor in the Structure program and other Bayesian clustering methods. They conclude the issue is complex, void of a single solution that can be applied to all data and the issue requires further extensive research. This study acknowledges the inclusion of related individuals in the data and gives caution to the two clusters observed with regards to the captive population, however, it follows suit with previous research and has produced comparable data.

Ideally, obtaining samples from the rest of the UK and European captive gorilla population, applying sibling identification methods and increasing the number of microsatellites from a panel of ten to thirty or more would provide a more robust dataset to make inferences from. The application of other non-Bayesian clustering methods in addition to Structure would add reliability if similar inferences emerged. Here, a diversity of analyses were performed on the microsatellite data set, including Structure, PCoA and pairwise F_{ST}, and overall results indicated weak population genetic structure in the western lowland gorilla.

Most genetic studies of wild western lowland gorillas are based on contemporary populations, e.g. Lukas *et al*. (2004), Jeffery *et al*. (2007), Nsubuga *et al*. (2008). This study, however, also focused on a historic population and was the first study to genotype the PCM collection to this extent. Arandjelovic & Thalmann (2001) noted how few temporal studies existed given the abundance of museum specimens available in natural history collections globally and those using microsatellites for historical analysis were very limited. This study would benefit from the inclusion of western lowland gorilla genotypes from contemporary wild populations across the distribution range, to make further temporal and geographical comparisons, as in the case of Nsubuga *et al*. (2010) who included six wild western lowland gorilla genotypes from a contemporary wild population from Cameroon.

The PCoA, Structure analysis and F_{ST} pairwise comparisons demonstrated the genetic differentiation observed between historical and contemporary Cross River gorilla populations. The

Cross River gorilla contemporary population is severely fragmented with less than 250 individuals remaining in the wild (IUCN 2019). Although western lowland gorillas are a critically endangered species, their estimated numbers in the wild are 100,000 individuals (IUCN 2019), considerably higher than the Cross River. Despite their higher population numbers, they have suffered a population decrease of more than 80% over three generations, which has led to their critically endangered status (IUCN 2019). This means that 100-150 years ago, the number of western lowland gorillas was much higher, their habitat would not have been exposed to the pressures it faces today from anthropogenic activities such as deforestation leading to the fragmentation of populations, and gene flow/migration would have occurred more easily throughout the continuous forest habitat. This could explain why the results found here reflect one lineage in the more abundant, genetically interconnected, historic population, even though regional genetic variation has been reported in wild contemporary populations and was observed in the contemporary captive population. Moritz (1995) and Clifford *et al*. (2003) noted that nuclear genes are expected to retain ancestral polymorphism for longer time periods than mtDNA, and Seaman *et al*. (1999) reported for gorillas that mtDNA divergence is greater than in nuclear divergence, while Kaessmann *et al*. (1999) observed much less variation in nuclear genes than in mtDNA for chimpanzees.

Nsubuga *et al*. (2010) investigated the genetic structure of the North American captive gorilla population using 32 microsatellite loci in 144 western lowland gorillas. The results reported here are directly comparable with theirs. Nsubuga *et al*. (2010) identified two genetic clusters in the North American captive population which they reported as a surprising result, and Simons *et al*. (2013) identified two clusters in the North American captive-born western lowland gorillas; therefore, the results shown here appear to be consistent with previous research on captive western lowland gorillas (despite the inclusion of related individuals). As noted by Nsubuga *et al*. (2010), regional genetic clusters have been identified in wild gorilla populations for the Cross River gorilla (Bergl & Vigilant 2007), and among the gorilla species (Guschanski *et al*. 2008), which may have indicated that this would be the result for captive populations also. However, the analysis of the historic wild population presented in this chapter, did not find any evidence of genetic regional clustering. Nsubuga *et al*. (2010) surmised that the largest genetic cluster from their data was likely of Cameroon and the Republic of Congo origins, and Cluster 1 of their study may have represented another genetic cluster of Cameroon origin. The same is true for this data, where Cluster 2 contained all the historic wild population gorillas which were primarily from Cameroon, with the exception of five individuals from the Republic of Congo. Cluster 2 also contained 35 individuals from the contemporary population, and of those 35 gorillas, Mouilla, Babydoll and Sidonie, were all born of wild parents from Cameroon, thus supporting the idea that Cluster 2 represents individuals from Cameroon.

However, the gorilla Tebe who is a descendent of wild Gabon gorillas also clustered in Cluster 2 (as a hybrid) which could indicate that Cluster 2 could be more geographically widespread.

The Cluster 1 individuals included the silverback Kouillou, a direct descendent of wild Congo gorillas. However, Cluster 1 did not represent a Congo cluster exclusively. It contained Kouillou and his offspring, but also contained most of Bitams' offspring: Boumi, Mambi, Bitanu, Tamba, Ujiji, Jubi, Matibi, Timbou and Tamki. Bitam was a direct descendent of wild Gabon gorillas and died in 2006 (Wilms & Bender 2010). Tambabi was sired by Bitam and she resided in Cluster 2 with her dam, Babydoll (of Cameroon origins). The other females Bitam bred with are mainly of Cameroon origins (maternal lineage) and includes Mouilla (a mitochondrial founder of the contemporary population) of Cameroon origin, yet their offspring reside in Cluster 1. Therefore, there does not appear to be a distinguishable genetic clustering between Gabon/Cameroon or Congo as might have been expected, but does appear to support Nsubuga *et al*. (2010) with the possibility that the genetic clustering represents two Cameroon/Congo lineages.

The PCoA analyses supports the Structure results and found no regional genetic clustering among the historic population (Fig. 4.15.a) and identified a partially genetically distinct cluster of contemporary gorillas (which were Cluster 1 individuals), with the remaining individuals of both the historic and contemporary population not demonstrating any genetic clustering (Cluster 2 individuals) (Fig. 4.15.b). The PCoA which included the Cross River gorillas (Fig. 4.16) demonstrated the genetic distinctiveness of the contemporary Cross River gorilla population in comparison to all other groups. Interestingly, Cluster 1 individuals from this study, did show considerable genetic similarity with the historic Cross River gorilla population, however, there was considerable overlap with all groups with the exception of the Cross River gorilla contemporary population. The pairwise F_{ST} comparisons confirmed the Cross River gorilla contemporary population to be the most genetically differentiated of the groups, while the ASP Cluster 2 and PCM group showing the least genetic differentiation (Table 4.10), confirming the genetic clustering of these latter two groups as one cluster in the Structure analysis. The AMOVA results showed that the historic and contemporary western lowland gorillas populations have similar levels of genetic variation and little genetic differentiation. The F_{ST} value was 0.015 corresponding to low genetic differentiation between the two populations.

4.6.3. Regional genetic diversity comparisons of the historic western lowland gorillas

The historic population when regionally defined into the three subgroups A, B and C, showed little genetic diversity between them. The only exception was subgroup C, which produced lower results for all genetic diversity measures for *NA, AE,* H_O, H_E, and H_I, except for *AR,* where it had a higher level

(4.49) than sub-group B (3.94), but a lower level than subgroup A (4.56). Allelic richness is deemed to be the most informative measure and takes into consideration sample size (Simons *et al*. 2013). Amongst the subgroups there were no significant results in terms of genetic diversity except for *NA* and H_E which could be attributed to the inclusion of the small sample size of subgroup C (*N = 5*).

The Fixation Index complemented the previous findings somewhat, in that subgroup C had an F_{IS} value of -0.298 deeming it the least inbred of the subgroups. Subgroup B, had an F_{IS} value of -0.222 and the 'most' inbred subgroup was A, despite it having the largest sample size (*N = 45*) with an F_{IS} = -0.210, however, subgroup A did have the highest *NA, AR,* and H_I levels of all three subgroups.

Of the subgroup analysis, theta *(ϑ)* under the stepwise mutation model produced mean results of 2.808 for subgroup A, 2.664 for subgroup B and 2.025 for subgroup C. These results showed that subgroup A was the most diverse/least inbred and subgroup C was the most inbred/least diverse. The results of subgroup C here contradict those of the F_{IS} results, questioning the reliability of the small sample size for subgroup C.

4.6.4. Genetic diversity and demographic comparisons of the contemporary and historic western lowland gorillas

The inbreeding coefficient (F_{IS}) showed that the contemporary population was considered slightly more inbred than the historic population (F_{IS} = -0.205 and -0.236 respectively). Theta *(ϑ)* under the stepwise mutation model produced mean results of 2.574 and 2.718 for the contemporary and historic populations, respectively. This indicates that the historic population is the least inbred/most genetically diverse (and complements the F_{IS} results) but there is not great genetic differentiation between the two populations.

The historic population had marginally higher levels of *NA, AE,* H_o and $H_{I.}$ Interestingly, the contemporary population had a higher level of *AR,* although none of the results were significant. These results complement the research by Simons *et al*. (2013), who reported on the genetic diversity of captive born gorillas in North America. Their results found that *AR* was significantly higher in the captive population compared to all wild populations in their study, and allelic diversity and heterozygosity was higher. This study finds that the contemporary population also has higher *AR* levels (although not significant) but the other diversity measures were extremely similar. Mean *AR* values were higher than mean *AE* values in both populations which is indicative of the influence of private alleles (Nsubuga *et al*. 2010) and has also been observed in previous studies e.g. Nsubuga *et al*. (2010). The *AR* results are particularly interesting, not only is *AR* considered the most informative genetic diversity measure as it accounts for sample size (Simons *et al*. 2013), the *AR* results here

support the genetic population structure results in that the contemporary population was found to contain two lineages indicating population structure, whereas the historic population did not contain population structure. The population structure (two genetic clusters) is likely due to the presence of a diversity of alleles and their frequency in the contemporary population.

A significant finding of this study was the results from the private alleles. The reduction of private alleles from a population are an indication of genetic diversity loss (Szpiech *et al*. 2008; Szpiech & Rosenberg 2011). There was a total of 28 individuals carrying private alleles with only 3 of them present in the contemporary population, belonging to Djanghou, Sidonie and FouFou. In addition, the private alleles amongst the historic population were not restricted to one subgroup. Given the sample sizes of the historic and contemporary populations were very similar, one could expect to see a similar distribution of private alleles amongst the two populations, but this was not the case. From the data quality checks, it was established the Djanghou had not sired Kwimbas' two infants, but Kisane (Djanghous' offspring) had. Djanghou and Kisane were in the same family group and Djanghou has sired many offspring but none have inherited his private allele. As mentioned earlier, this research would have recommended Kisane be removed from the family group, if Djanghou were expected to sire further offspring, a move which has already been implemented by the Aspinall Foundation (without knowing the genotypic constitution).

Sidonie was born in 1972 and is no longer kept in a family group and has passed her reproductive years. FouFou was part of Djalas family group who were reintroduced to Gabon in June 2013 (V. Mathieson pers. comm, 2019), and this was the first time that a large family group of gorillas was reintroduced (King 2013). Tragically, in September 2014, five of Djalas family group were killed by an attack from another gorilla. Kishi, Tamki, Mumba and Kibi were all killed (all females) and Akou, the young male of the group (Djalas offspring). FouFou went missing and is presumed dead (V. Mathieson pers. comm, 2019).

The genetic bottleneck analysis confirmed that the neither the historic population or the contemporary population underwent a genetic bottleneck event, this research revealed the same results as the Simons *et al*. (2013) study in that only the only the IAM model produced a significant result but can be disregarded as it is not a suitable analysis for this dataset, with the other two models, the SMM and TPM, not producing a significant result. This is surprising considering the persecution of gorillas during the 19th and 20th century. The regional migration results for the historical population (Fig. 4.17) are highly unlikely to be accurate and thus require further analysis. GeneClass detects F1 and F2 individuals only, the ecology of the western lowland gorilla shows that although both sexes disperse (Fünfstück *et al*. 2014) homeranges would not extend over the distance required to allow the results from GeneClass to be realistic. Other software such LAMARC (Beerli &

Felsenstein 2001; Kuhner 2006) and Migrate (Beerli *et al.* 2019) also detect migration so there is potential for further analysis in this area, although it was not a main aim of this research.

Previous literature has indicated that female gorillas shape the genetic structure of populations (Guschanski *et al.* 2008). Using microsatellite data and the program, Structure, a population of the remaining two mountain gorilla (*G. g. beringei*) populations was genotyped and investigated by Guschanski *et al.* (2008) for population structure. Their results found that the Bwindi gorilla population were geographically and genetically structured which was attributed to the non-random movement by the female gorillas. There was a lack of genetic and geographical structure in the males which suggested that the dispersal of the males was great enough to eradicate a geographical signal to be detected. Their findings were consistent with those of Douadi *et al*. (2007) who investigated western lowland gorilla dispersal via microsatellite and mtDNA markers and found sex-biased dispersal among the population, with females showing more structure than males. In contrast, Fünfstück *et al*. (2014) found using nuclear data that both sexes of western lowland gorillas disperse and migrate over similar distances. Whilst male gorillas may disperse over longer distances initially, females will transfer to neighbouring groups but may do so multiple times, thus dispersing over similar longer distances as observed in the males.

These contrasting results provoke interesting questions. With further sampling of individuals, it would be interesting to investigate genetic diversity and structure in terms of sex distribution, as well as differences in genetic and geographical clustering. Separating the contemporary and historical populations by sex would be possible as data already exists and other individuals in the PCM have not been genotyped. Further Structure analysis (or alternative clustering software programs) may reveal regional clusters when a larger data is analysed in relation to sex. The historical population showed no regional clustering but perhaps separating the data by sex would reveal some level of genetic structure, where males could show weak structure due to regional movements from and to other groups as young silverbacks while females may reveal genetic structure due to philopatry (Douadi *et al*. 2007; Guschanski *et al.* 2008). As mentioned previously in Chapter 3, the historical population could be considered as one population with no regional differentiation due to many specimens coming from Cameroon with no obvious physical barriers such as rivers which could influence regional genetic diversity and population structure as discussed in Anthony *et al*. (2007). Whereas the contemporary population contains individuals with origins covering a much wider geographical area which includes Cameroon, Gabon and the Republic of Congo and which could account for the two clusters observed in this chapter as opposed to the one cluster found in the historical population.

4.6.5. Conservation implications

One of the main aims of this research was to establish whether the historical population displayed any genetic structure regionally. It has been problematic assessing genetic regional variation in western lowland gorillas because their habitat is primarily swamp and rainforests, among other reasons; thus, it is difficult to make complete surveys which is why their numbers in the wild remain a vast estimate (Magliocca *et al*. 1999). In addition, samples that are non-invasively collected (i.e. faeces and shed hairs) often produce poor quality DNA which hampers genetic analysis (Clifford *et al*. 2003) and historic samples typically do not contain enough geographical information and often a broad region such as "Gabon" is the only information present. This is what makes this research unique, the PCM gorilla collection contains skin samples which yielded sufficient DNA for downstream analysis plus virtually every sample had a region defined and a specific geographical coordinate which allowed for within-region analysis.

These initial findings indicate a positive outlook for the conservation of western lowland gorillas in terms of genetic diversity. Despite wild population numbers being drastically reduced in the last two decades, resulting in them currently being classified by the IUCN as critically endangered (IUCN 2019), the results presented here indicate that the genetic diversity of the captive population remains similar to the gorilla population 100-150 years ago.

Most current conservation efforts of critically endangered species focus on species and sub-species management and not on regional variation management. The contemporary population is currently managed in this way and is applied internationally for the western lowland gorilla conservation program. Many of the gorillas in the contemporary population have already bred with individuals from other regions. For example, Djanghou who has a Congolese ancestry, has bred with Kimba of Cameroon origin. Kifu, who is born from Cameroon origins (maternal line) (Wilms & Bender 2010), has bred with several females including Tambabi (Cameroon maternal line), Sounda (Congolese maternal line) and Tebe (Gabonese maternal line). Comparably to Nsbuga *et al*. (2010), the results shown here indicate that gene flow among the two lineage clusters was occurring in the wild population. This was shown here by individuals who are F1 descendants of wild founders that appeared as 'hybrid' gorillas in the Structure analysis (Tebe and Djala).

This study concludes that although there has not been any recruitment of wild western lowland gorillas into the captive population since the 1970s (Nsbuga *et al*. 2010), there is no evidence to indicate the captive population is at risk from genetic diversity loss, and at present there is no recommendation that the two genetic clusters should be managed separately. Therefore, as in the case of Nsubuga *et al*. (2010) and Simons *et al*. (2013), this study can also conclude the genetic management via kinship and the International studbook, is maintaining the genetic diversity of the

captive population, and that continuing to manage it as a single population is supported by these findings.

This study has also revealed some important findings in relation to gorilla conservation. To further this research, genotyping the remainder of historic population specimens should be performed, there are still over 70 skin samples that can be analysed for this dataset. In addition, obtaining samples/datasets for other captive gorillas in the UK would assist, firstly by building the dataset further, and secondly it would help to identify any potential further lineages and any undetected Cross River gorillas. With more samples from historic populations and current captive populations plus data from wild contemporary western lowland gorillas a fully comprehensive account of regional population structure could be formed, and further conclusions could be drawn regarding the captive population of the UK gorillas. In addition to enlarging the dataset in terms of individuals, increasing the microsatellite panel to 30 would be desirable although Nsubuga *et al*. (2010) used 32 microsatellites and this study found the same results with 10 loci.

Single Nucleotide Polymorphism (SNP) analyses, which are superseding microsatellite analysis, would also be of benefit and is certainly underway for many gorilla studies. If possible, SNP analysis of this dataset would certainly be of benefit and it would be interesting to compare and combine the results from the two methods. Advances in molecular technology means there are now several marker methods that can be applied to data for constructing genetic maps which include, allozymes, Restriction Fragment Length Polymorphisms (RFLPs), Random Amplified Polymorphic DNA (RAPDs), Sequence-Tagged sites (STSs) and Amplified Fragment Length Polymorphisms (AFLPs), in addition to microsatellites and SNPs (Ball *et al.* 2010). Comparatively, SNPs provide less information per locus than microsatelittes because they are usually biallelic, the consequence of this is more markers are required than the highly polymorphic microsatellites, but SNPs tend to have a lower error rate than microsatellites (Hoffman & Amos 2005; Ball *et al.* 2010). Additionally, SNPs evolve slower than microsatellites but can detect expansion from a common ancestor and demographic inferences over thousands of generations, whereas microsatellites can make inferences regarding more recent demographics relating to gene flow and population structure (McManus *et al.* 2015).

Studies of gorillas employing SNPs include Kennedy *et al.* (2003), Yu *et al.* (2004), McManus *et al.* (2015), Xue *et al*. (2015) and Das *et al.* (2019). McManus *et al.* (2015) found evidence of a one population model for the western lowland gorilla via SNPs analysis and referred to this contradiction when compared to Nsubuga *et al.* (2010) and Fünfstück *et al.* (2014), both of which found evidence for multiple population clusters, as did this study for the captive population. This study recommends the use of both microsatellite and SNPs data, acknowledging the advantages and disadvantages of each method to allow for more robust and reliable datasets for demographic and population

structure analyses of gorilla species, with this study providing microsatellite data for a historic and captive population, which has not been previously investigated.

Expanding this study wider than the western lowland gorilla subspecies would also be of benefit to many endangered species. The Powell-Cotton Museum holds such a well-documented collection with one of the best primate collections in the world, there is a great deal of research that could be achieved, and not just with primates as they have vast samples of many duiker, gazelle and antelope species as well as buffalo and waterbuck, just to name a few. Genetic and geographical investigations into the family Galagidae (bush babies) would be of specific interest given the lack of data available for them. Additionally, a geographic/genetic comparison of primates occupying different ecological niches would be of conservation interest, including an extension of this research on the western lowland gorilla, and also Galagidae and de Brazza monkeys.

The aims of this chapter were to investigate genetic diversity and structure of the historic (Powell-Cotton Museum, PCM collection) and contemporary (Aspinall Foundation, ASP) populations of the western lowland gorilla (*G. g. gorilla*) using microsatellite loci to ascertain whether the historical population of the western lowland gorilla demonstrates regional genetic variation. This aim was achieved and genetic diversity estimates were comparable among the captive and historic populations for *Na* and *Ae* and for *AR*, the latter, which is deemed the greatest genetic diversity measure as it accounts for differences in sample size (Simons *et al.* 2013), was greater for the captive population. Subgroup C of the historic population had some questionable results which can likely be attributed to the small sample size (N = 5). Nonetheless, the contemporary population appears to be genetically 'healthy' in comparison to wild historic populations and the US captive population. No regional genetic structure was observed in the historic population but was found in the UK captive population.

The paternity of Kwimbas' infants in the captive population was confirmed and genotypic information of both the contemporary and historic populations was generated and the gorillas used in this research were confirmed to be western lowland gorillas. The generation of the genetic data for the captive population will form a genetic database which can be used to assist with the management of the captive population and aid in the planning and decision-making for conservation biology. Two genetic clusters were observed in the captive population which reflect the results found by Nsubuga et al. (2010) for the US captive population, and likewise, no recommendations are put forward to manage the two clusters identified separately.

The hypotheses and predictions stated regional structure and variation will be present in the historical population. This was based on previous literature and the results of previous chapters; however, this was found not be the case. The historical population was predicted to reveal more genetic diversity and structure than the captive population. Again, genetic diversity estimates, found

greater diversity in the captive population in terms of allelic richness and marginally greater diversity estimates were found in the historic population for number and effective number of alleles. Heterozygosity estimates were extremely similar throughout all the populations, apart from Subgroup C in the historical population which had the small population size and it was the captive population which revealed more population structure than the historical population.

Despite the social structure and polygamous mating strategy of western lowland gorillas, the silverback (Djanghou) was confirmed not to be the sire of two infants in the family group which was a surprising result. It was predicted that the gorillas used in this study would be confirmed as the western lowland subspecies and not the Cross River subspecies. This prediction was accurate, however, and it was interesting that the Cluster 1 individuals in the captive population grouped closer to the historic Cross River gorillas than any other population, including the contemporary Cross River population. Finally, the results found here were comparable with the US captive population and were in fact, extremely similar, indicating genetically diverse and healthy populations in captivity.

Chapter 5

General Discussion

The broader aims of this study were to investigate regional variation (morphological and genetic) of western lowland gorillas using a combination of population genetics, phylogeography, geographical information systems (GIS) and geometric morphometrics approaches. More specifically, the main aims were to compare the population genetic diversity and structure within and among populations of gorillas from a historic collection from the Powell-Cotton Museum and from captive individuals from the Aspinall Foundation, as well as to study the phylogeographic patterns of these historic and contemporary populations of the western lowland gorillas in comparison with previously published genetic data.

The purpose of this multidisciplinary study was to 'bridge the gap' in the scientific research of this critically endangered primate by combining geographical, morphological and genetic data from various sources, and reiterate the importance of museum natural history collections and captive breeding programmes for conservation purposes. The genetic data obtained from microsatellite and mtDNA markers also allowed to confirm whether the captive gorillas used in this study were effectively (from the genetics point of view) western lowland gorillas and not hybrids with any other gorilla subspecies, particularly the Cross River gorilla subspecies. Furthermore, the genetic data allowed to confirm the relatedness and parentage of gorilla individuals in captivity through paternity testing.

This project has only been possible to carry out through the use of historic samples that have been carefully preserved and made available for research by the Powell-Cotton Museum, as well as the collaboration with the Aspinall Foundation, a conservation organisation with the mission to stop the extinction of rare and endangered species in the wild.

5.1. Main findings and conservation implications

Three of the four subspecies of gorilla are critically endangered, the western lowland, the eastern lowland and the Cross River gorillas. The mountain gorilla is currently listed as endangered, but it was listed as critically endangered until 2018 (IUCN 2019). Habitat loss, fragmentation and degradation as well as other anthropogenic pressures such as bush meat hunting have all contributed to the decline in their populations as well as infectious disease such as Ebola (Junker *et al*. 2012; Baas

et al. 2018). Of all the four subspecies, the western lowland gorilla is the most numerous and occupies the largest range, the relatively large continuous habitat permits higher levels of gene flow (Fünfstück & Vigilant 2015) than those of the other gorilla subspecies whose populations are significantly smaller and more fragmented. This is particularly true for the Cross River gorilla whose wild population is assumed to be less than 250 individuals surviving in severely fragmented populations (Baas *et al.* 2018; IUCN 2019). The genetic consequences of small isolated populations increase levels of inbreeding and reduce genetic diversity thus threatening the long-term survival of the population by reducing fertility, increasing susceptibility to disease and the inability to adapt to environmental changes (Baas *et al.* 2018).

The findings of this study support that high levels of gene flow were occurring in the wild population of western lowland gorillas (in Cameroon and the Republic of Congo), at least up until 100-150 years ago as demonstrated by the lack of population structure at the level of nuclear DNA observed in the historical population. The contemporary captive population investigated in this study, does not appear to be suffering from the effects of genetic diversity loss and inbreeding compared with other captive populations in the US and with the historical wild population. This is a positive outcome in terms of gorilla conservation, not only is the captive population genetically 'healthy' (not showing high levels of inbreeding , inbreeding depression, and very few alleles compared with a historical population), it implies that the gorillas in the captive population could be introduced to the wild without genetic consequences (outbreeding depression) and additionally, although regional differentiation has been observed morphologically and genetically in terms of mtDNA (this study and others), the analysis of the historic wild population has indicated that those variations may have been less significant in the past and that populations may have been more continuous with increased overlap than what is currently observed.

Conservation efforts are limited by resources and financial implications for a multitude of species. For the western lowland gorilla, efforts to limit habitat loss and degradation and to maintain connectivity between populations such as habitat corridors is of considerable importance to permit the continuation of gene flow and reduce genetic diversity loss for future generations, thus maintaining the historic demographics of the subspecies. A recent study regarding the eastern gorillas noted the importance on maintaining habitat connectivity to facilitate gene flow rather than concentrating conservation efforts on the core areas of genetic diversity (van der Valk *et al*. 2018). However, the practicalities of such a widescale effort are often not feasible. The reduction in the bush meat trade and protecting the wild population from further population decreases is paramount in preventing genetic diversity loss, which could result in the western lowland gorilla facing increased pressures as has been observed in the other gorilla subspecies. The IUCN (2019) reports that just 22% of western lowland gorillas reside in protected areas which cover a mere 14% of their geographic

range. A further 21% reside in certified logging concession areas which equates to a further 8% of their distribution (Strindberg *et al*. 2018). Over half (58%) of western lowland gorillas and 78% of their range is unprotected and thus, highly vulnerable to poachers (IUCN 2019). The IUCN 2015-2025 action plan recommends (amongst other criteria) that the maintenance of large, intact and well-protected forest areas will be crucial for long term gorilla (and other great apes) population maintenance (IUCN 2019).

There is evidence and hope that gorillas, given the correct protection and management, are able to recover from the anthropogenic factors that have been inflicted upon them. In 2018, the eastern mountain gorillas were relisted from critically endangered to endangered. The mountain gorilla population exists in two isolated populations (Virunga and Bwindi) (Granyon *et al.* 2018). The total population size was 1004, significantly smaller than the estimated western lowland gorillas, however, the mountain gorillas have received extensive monitoring and conservation efforts since the 1950s (this subspecies is the famous 'Gorillas in the Mist' gorillas which Dian Fossey studied), which as seen their numbers increase from 620 in 1989 to their present numbers (https://wwf.panda.org/knowledge_hub/endangered_species/great_apes/gorillas/mountain_gorilla/). This increase in population size warrants their relisting.

In Chapter 1 the importance of museum collections to biodiversity conservation was investigated incorporating a multidisciplinary approach such as the inclusion of GIS. This research has highlighted and reiterated the abundance of primary biodiversity species data that is held in natural collections and has found the Powell-Cotton Museum to be exceptional in terms of its additional contextual information, allowing for specific biogeographical investigations to be performed for a variety of species including the western lowland gorilla. There is no doubt that further investigations into the collections for other species will be of interest to scientists, and the addition of the biodiversity mapping for each species in the collection will act as an initial visual guide for researchers in terms of geographical locations of the species or area for which they are interested in investigating.

Museum samples are becoming increasingly utilised, van der Valk *et al*. (2018) noted the essential role museum specimens play by providing a window into the past which allows for temporal analyses and the assessment of anthropogenic factors. Their study quantitatively assessed genetic diversity of past and present Grauer's gorillas spanning a few generations. Yeates *et al.* (2016) noted that with the advancement of molecular tools and techniques such as next-generation-sequencing and analyses, the genetic value of museum specimens may become more widely appreciated. This study supports the value of museum specimens and recognises the scope and wealth such collections can bring to species conservation.

In terms of regional morphological variation of a historic western lowland gorilla population studied by means of geometric morphometric analyses of skulls and mandibles, in Chapter 2 it was shown that the results complemented previous research that used traditional morphometric analysis (e.g. Groves 1967, 1970). Size (centroid size) comparisons revealed that all three regions were similar, with skulls and mandibles being smaller, albeit not significantly, in the Republic of Congo (region C) specimens (Fig. 2.8). Shape analyses, however, did reveal significant differences regionally with regards to the skulls, although mandibles did not show any significant differences. The multivariate analysis of variance (MANOVA) identified significant differences for both sexes in relation to skull shape (Table 2.5) and further pairwise Hotelling T^2 tests showed significant regional variation for both sexes between regions A and B, and also between regions A and C for males only. Visualisation analysis via discriminant function analysis (DFA) did discriminate among the regions, whereas principal components analysis (PCA) and relative warps plots did not reveal any clear regional clustering. The research from this study confirmed that the sub-species *G. g. gorilla* showed high morphological variability throughout its geographic distribution with considerable 'overlap' among individuals from various regions (Groves 1970; Uchida 1998; Stumpf *et al*. 2002, Albrecht *et al.* 2003 and Leigh *et al.* 2003). Despite the high variability and overlap, significant regional morphological variation was observed, which was consistent with the contemporary designation of populations into demes (Groves 1970; Leigh *et al.* 2003). Regions A and B (the plateau and coastal demes, respectively) were significantly different, and no regional variation was found between individuals in region C compared with region A and B. Region C individuals were most similar to region B (the coastal deme) which is the deme Groves (1970) had originally assigned them to. Female gorillas were more homogeneous in their skull morphology compared with males, as found in previous studies (Albrecht *et al.* 2003). Although there does not appear to be highly significant levels of regional morphological variation, there was a certain degree of variation more notable in males than females and particularly with reference to male skulls, not mandibles. This indicates that morphological traits in these gorilla demes could be under selection despite gene flow among populations across the region studied, or there could be a phenotypic plastic response to slightly different environments or food items.

With regards to the genetic analyses, Chapters 3 and 4 revealed some important findings relevant for western lowland gorilla conservation. The principal findings of the mtDNA analysis of the Hypervariable Region I (HVI) highlighted the difficulties of obtaining true mtDNA sequences and detecting nuclear inserts of mitochondrial DNA (numt) sequences which hinder phylogenetic or phylogeographic analyses. Despite the relatively low number of true mtDNA sequences (30 in total) obtained compared with numt sequences identified, there were still important findings. Firstly, the refined mtDNA data which included the 30 sequences generated from this study and GenBank data,

showed a total of 61 haplotypes in the final dataset of 149 sequences. Phylogenetic analysis revealed that the eight true mtDNA sequences from the contemporary population belonged to haplogroup D2, and that the 22 historic sequences were more widely dispersed throughout the haplogroups with 17 individuals in haplogroup C2, three individuals in C1 and two individuals in haplogroup D2.

Mapping the PCM haplogroups (Fig. 3.11) showed consistencies and inconsistencies with previous research. The 17 C2 haplogroup specimens reflected the geographical range as reported previously by Soto-Calderón *et al*. (2015). However, evidence was found of three specimens (CamI.14, FC.114 and FC.147) which have raised questions regarding the historical distribution of haplogroups. The D2 haplogroup (specimens CamI.14 and FC.114) were recorded further south (in the Republic of Congo; specimen FC.114) and much further west (specimen CamI.14) than has been previously reported. Specimen FC.147, a C1 haplogroup individual was found to be further south than the reported C1 distribution. These findings indicate the historical population may have contained more geographically widespread haplogroups than has been observed in contemporary populations, and that haplogroup distribution may have overlapped more than current data suggests. This is not an unreasonable conclusion, given that Clifford *et al*. (2004) also found a PCM specimen to contradict haplogroup distribution (exact specimen unknown) and this study had identified an additional three individuals (or two if the same specimen CamI.14 was used). These unexpected phylogeographical results highlight the importance of using museum specimens for current conservation purposes.

Surprisingly, the historical (PCM) population showed lower nucleotide diversity than current contemporary wild populations and captive populations, but this result may not be as unexpected as initially hypothesised. Given the PCM population is primarily from Cameroon (with a handful of specimens from the Republic of Congo and none from Gabon) and the captive and contemporary wild populations encompass individuals from all three of those regions (and all haplogroups) this could be the explanation. Additionally, the pairwise F_{ST} tests showed no significant difference between the PCM and CAM (contemporary wild) group, further supporting the PCM population to represent a single Cameroon population with little genetic variation despite the observation of three mtDNA haplogroups.

With regards to the contemporary captive population (ASP), the most relevant results found here for mtDNA analysis were the identification of eight individuals in the D2 haplogroup, seven of which were closely related with the exception of the gorilla Kangu. This is an important contribution in terms of genetic diversity management of the contemporary captive population as it is the first recording of any genetic analysis for the Aspinall Foundation gorillas. Although the mtDNA analysis was severely hindered by numt sequences, this is the start of building genetic profiles for each individual. Kangu was released back into the wild population and it would appear (according to the

historic DNA analysis) that he was released into the region where the D2 haplogroup may have been present.

Significant findings from the microsatellite analysis firstly confirmed that all gorillas from the contemporary and historic population used in this study were western lowland gorillas and did not contain any Cross River individuals (Fig. 4.10). Secondly, the analysis of parentage of Kwimbas' two infants (Table 4.5) confirmed Kisane as a mature silverback and highlighted that changes needed to be implemented in the family group if Djanghou was expected to remain the dominant silverback and sire further offspring. This was a change in the captive gorilla population implemented during the course of this research, and that this research supports and validates. Furthermore, the paternity of other individuals in the captive population was confirmed during the data quality checks and did not reveal any other paternity discrepancies, thus all captive breeding records regarding the paternity of those individuals sampled can be confirmed as accurate.

This research confirmed the presence of two genetic clusters within the contemporary population, a result that was comparable to research performed on the US captive gorillas (Nsubuga *et al*. 2010; Simons *et al*. 2013). This is encouraging for the UK captive population and the international captive population as it demonstrates that the genetic diversity of the captive population is being maintained at an international level despite no wild introductions to the captive population since the mid-seventies (Nsubuga *et al*. 2010). Again, this is the first genetic analysis of the Aspinall Foundation gorillas, therefore providing the foundations of a genotype database for the individuals present in the UK captive population.

Mace (1988) concluded that the loss of genetic diversity was not an initial threat to captive populations for at least 200 years, and this research supports those findings. Probably the most significant finding of this study was that the contemporary population appears to be representative of the wild population in terms of genetic diversity. Genetic diversity measures based on microsatellite data of the captive population generally considered it to be less diverse than the historic population but not significantly so. Allelic richness (*AR*) was in fact higher in the contemporary population compared with the historic population (*AR* = 8.73 and 8.41, respectively) and *AR* is deemed to be the most informative genetic diversity measure (Simons *et al.* 2013). The inbreeding coefficient (F_{IS}) found the contemporary population to be slightly more inbred than the historic population (-0.205, -0.236, respectively) but values were relatively comparable and because of their negative sign, they indicate that the populations are outbred, possibly reflecting the little genetic structure and high gene flow estimates. Upon further investigation, the allelic richness result was likely due to the contemporary population containing two genetic clusters whereas the historic population was found to contain only one (Fig. 4.12.b and Fig. 4.14, respectively), the latter was an unexpected result given previous research and the findings of previous chapters in this research

which has found regional variation in terms of morphology and mitochondrial analyses. Additionally, this research found no evidence to suggest the captive population needs to be managed differently from its current management plan, nor does it find any evidence to suggest that individuals from Cluster 1 be managed differently to those in Cluster 2, a similar conclusion also reached for the US captive population (Nsubuga *et al.* 2010).

Several observations have become apparent when viewing the genetic results in their entirety. The Cluster 1 individuals from the population structure analysis of genotype data (Chapter 4) contained all the haplogroup D2 individuals from the mtDNA research with the exception of Kangu, who was considered a 'hybrid' in the microsatellite analysis and was the only unrelated individual to the other seven D2 haplogroup gorillas. The population structure analysis did confirm the presence of two genetic clusters in the historic population as evidenced by one individual specimen (Mer.95) which was identified as a 'hybrid' individual in Cluster 2. However, Mer.95 was the only historic specimen considered as a 'hybrid', but there is evidence of further historic specimens (CamI.97 and CamI.109) that contained both genetic clusters but they were assigned to Cluster 2 as their proportional group membership (*Q*) exceeded 80% for Cluster 2. Given there are over 70 individuals in the historic population for which genotype data was not obtained, it would be of interest to genotype the remaining gorillas and reanalyse the population genetic structure to verify if further historic specimens are revealed as hybrids or Cluster 1 members. If this were to be confirmed, then the results for the contemporary and historic populations would both reveal two genetic clusters.

The genetic clustering of the contemporary western lowland gorilla population appears to support two lineages from Cameroon/Congo origins as also found by Nsubuga *et al.* (2010). Additionally, the 'hybrid' individuals identified here between Clusters 1 and 2 imply a level of gene flow between populations in the wild (Nsubuga *et al.* 2010). The lack of population structure observed in the historic population for the microsatellite analysis could be reflective of the 'sampling' technique by the Major Powell-Cotton which did cover a widespread area but primarily focused on Cameroon and did not encompass other regions such Central African Republic and Equatorial Guinea where population structure has been observed in other studies using contemporary samples (Anthony *et al.* 2007; Soto-Calderón *et al*. 2014).

The lack of population genetic structure in the historic sample, could be an indication that the historic population had more gene flow than contemporary wild populations. The decrease of population size which western lowland gorillas have been subjected to in the last three decades and the increasing fragmentation of their habitat (IUCN 2019) may be the reason contemporary populations show regional differentiation. The western lowland gorilla is the most widespread geographically and still remains the most numerous of gorilla subspecies (IUCN 2019), historically their numbers would have been much greater and gene flow between populations would have

occurred more easily across the landscape. Gorillas are one of only a few mammal species for which males and females disperse from the natal group (Stoinski *et al*. 2009; Fünfstück *et al.* 2014). Guschanski *et al*. (2008) reported that females dictate gorilla genetic population structure more so than males. Interestingly, the results from GeneClass detected 11 migrants in the historic population, from which four were males and seven were females. Further to this, two female specimens (CamI.14 and FC.147) were detected as migrants and were also two of the three individuals in the mtDNA analyses to be out of their haplogroup distribution, however, as discussed previously, the GeneClass results are extremely dubious given that ecologically those individuals would not have migrated that great a distance. However, there is future scope to investigate sex dispersal in gorillas based on genetic markers and other ecological techniques; if there is differential dispersal among the sexes there would be implications for conservation of habitats and for reintroduction programmes.

There were nearly nine times as many individuals carrying private alleles in the historic population compared to the contemporary population, but the identification of private alleles in three of the contemporary individuals is an important point to mention. One of those individuals was Djanghou (a dominant silverback) engaged in an active conservation breeding program. The results found here have identified that as of yet, Djanghou has not passed on this private allele to his offspring. Private alleles are an important indicator of genetic diversity (Szpiech *et al*. 2008; Szpiech & Rosenberg 2011), and captive breeding programmes intend to preserve the genetic diversity and evolutionary potential of the species. Thus, the conclusion can be made that Djanghou is an important gorilla. Further research may identify more individuals with private alleles which may then also be deemed as important individuals in captive breeding programmes but how can the importance of a specific gorilla be assessed? Is one gorilla 'more' important than any other in terms of breeding programmes? If the expectation is that Djanghou is to breed with more females on an international scale because he carries a unique allele, should he be translocated to other zoos to maximise the chances of having offspring with females and passing on this allele? There are many risks to transporting wild and captive animals (Linhart *et al.* 2008). If one individual has an 'important' label does that mean the females should be brought to him thus risking their safety and welfare instead? It is important to remember that gorillas are sentient beings and although genetic diversity is of utmost importance for the conservation of species (Garner *et al*. 2005), where do we draw the line in the conservation efforts? Artificial insemination is a possibility and has been achieved for gorillas (Pope *et al.* 1997). This reduces the translocation risk, but reduces normal reproductive behaviour, social interactions and bonding which are an important part of gorilla family life. This is an ethical debate outside the scope of this thesis but does warrant acknowledgement.

5.2. Complementary or contradictory? A summarised synthesis of findings

This study essentially comprised of three investigations, one morphological and two genetic (mitochondrial and nuclear) and focused specifically on two western lowland gorilla populations, one wild historic population from the PCM and a contemporary captive population of the UK at the Aspinall Foundation. General aims of this research were to investigate regional variation of western lowland gorilla populations via morphological investigations of the historic population and genetic investigations of both the historic and captive populations. Additional aims were to reiterate the importance of museum natural history collections specifically those which contain valuable underutilised contextual information and to provide information relating to the genetic diversity and variation of the captive population for future conservation planning.

Each experimental Chapter (2-4) had its own specific aims, predictions and hypotheses and are addressed at the end of each chapter accordingly. However, it is important to synthesise the findings and look at the bigger picture of how different results complement each other in this multidisciplinary research, and how other findings show contradictory information. One of the main aims of this research was to investigate if regional variation exists among western lowland gorillas. Typically, in population genetics there is not strict answer due to the biogeographical complexity, evolutionary factors and animal behaviour affecting populations and the movement, reproduction and survival of individuals. Morphological regional variation has been evidenced in numerous studies, perhaps the most well-known study done by Groves (1970) on which all systematic gorilla taxonomy has been based. The four demes classification arose from the work of Groves (and supported by others) using traditional morphometrics to distinguish morphological regional differences. This study used a more powerful geometric morphometric approach and for the most part, complemented and supported regional demes variation. Significant regional differences in morphological shape were observed and, although non-significant, regional variation in size was also present. As scientific methods advance, it is not only important to look forward and use the most recent methods available, but also to look back, and to make comparisons which allow a more comprehensive and robust analyses of all the methods available. The results from the morphological chapter in this study, which used 2D geometric morphometrics, are a step further from traditional morphometrics, but there are now other more advance methods such as 3D geometric morphometrics and scanning (Adams *et al*. 2013; MacLeod 2017) which would be the future direction and logical next step in this research. Therefore, this morphological research fills a gap between traditional methods and the latest methods. Each method has its advantages and disadvantages, and it would be interesting to investigate whether the latest 3D methods also complement traditional and 2D geometric morphometric methods.

Having observed evidence of regional variation in skull shape from morphological analysis, the following chapters moved on to genetic investigations. The first focused on the mtDNA Hypervariable Region 1 which has been considered the workhorse of population genetics studies (Zink & Barrowclough 2008; DeSalle *et al*. 2017; Burgos *et al.* 2019). Although mtDNA only focuses on the maternal lineage and is becoming considered less useful in the presence of more advanced methods, it is still used in population genetics studies and has been used in several gorilla studies to investigate regional variation. This is one of the factors why this method was selected, to compare with previous findings and also, because no genetic research had been performed for either of the populations used in this study, to this extent, therefore, any information is new and likely, as in the case of the morphological research, to fill a gap between past methods and the most current.

The mtDNA investigations of Chapter 3 found historical regional variation in the form of haplotypes and their distribution, thus complementing the morphological research for the most part. However, many mitochondrial studies in gorillas use contemporary populations whereas this one specifically aimed to investigate a historical population as well as a captive contemporary population. The historical results indicate, more than once, that there is evidence for haplotype distribution to have been more widely distributed than previously reported with a few specimens containing a haplotype out of their current reported range. Clifford *et al*. (2003) also found this anomaly with one the PCM specimens they investigated (they did not investigate the collection to the extent of this study). This is interesting because it means the results found here were not a special case relating to one sample. This occurrence occurred for three of the specimens investigated in this study and for one investigated by Clifford et al. (2003), therefore there is growing evidence that historical western lowland gorilla population haplotype distributions were indeed more widely distributed. This is not an unfeasible assumption given that the number of gorillas in the wild 100 years ago was significantly larger than the contemporary populations (which are the focus of most gorilla studies), and as historical datasets of such detail are rare, and DNA from historical samples is often too degraded for successful amplification.

The mtDNA investigation were not as successful as expected with the presence of numts causing unreliability in the data and therefore, having to remove all numt sequences resulting in a significantly decreased sample size for both populations, but more so for the captive population, which was disappointing. However, for the eight remaining true mtDNA sequences of the captive population, their haplogroup was determined as D2 which is useful for conservation purposes. A group of gorillas were reintroduced to the wild by the Aspinall Foundation in 2017 and it is likely from the research conducted here, that those gorillas were also D2 haplogroup individuals and were in fact returned to their haplogroup distribution, a recommendation put forward by Soto-Calderón *et al*. (2015).

The final investigations focused on nuclear DNA rather than mitochondrial and therefore considers both parental lineages and not just the maternal line. Again, there are more recent techniques such as next-generation sequencing (NGS) approaches which use genome-wide markers such as single nucleotide polymorphisms (SNPs) (Šarhanová *et al.* 2018; Roques *et al.* 2019), but the advantages of using microsatellites in this research were the data could be directly compared with other datasets for the US captive populations (Nsubuga *et al*. 2010; Simons *et al*. 2013) and also for a contemporary (Arandjelovic *et al*. 2015) and historical population (Thalmann *et al*. 2011) of Cross River gorillas. Additionally, as with the morphological chapter, the microsatellite data can be used to bridge the gap between older and more recent techniques allowing for a comprehensive and robust dataset, building entire genetic profiles for individuals encompassing a broad range of methods.

The microsatellite analyses produced the most unexpected results in comparison to the two previous chapters on regional morphology and mtDNA, but it was the most comprehensive of the three experimental chapters. Unlike the previous chapters, Chapter 4 did not observe any regional distinction in the historical population at the PCM, which was predicted to occur given previous literature and the results of the two previous chapters in this thesis. However, the findings were directly comparable with research by Nsubuga *et al*. (2010) and Simons *et al*. (2013) for the US captive gorilla population, in finding two population clusters in the UK contemporary population, thus adding credibility to the results found here and genetic diversity levels were similar to those of the US captive population.

As discussed previously, there is evidence to suggest that female gorillas influence the genetic structure of populations, where it was absent from males (Guschanski *et al.* 2008). As population structure was observed in the mtDNA (maternally inherited, albeit being less informative than nuclear markers) analyses but absent from the microsatellite (nuclear markers) analyses, this research may support sex-based population structure, which is why it would be of interest to continue this research and perform further investigations of the microsatellite data using historical and contemporary populations but splitting it further into males and females, as was done for the morphological chapter. However, the morphological results did reveal males to show more regional variation than females, the latter were more homogenous, but this could be due to non-neutral evolutionary factors.

The observation of two clusters in the UK contemporary population included both sexes and still found population structure, as did the US captive population (Nsubuga *et al*. 2010; Simons *et al*. 2013). The inclusion of related family groups could be the reason for this distinction in the UK captive population, as discussed previously due to HW disequilibrium and LD, however, the results mirrored those of the US captive population so there is evidence to suggest that it is not reflecting family structure or biased analyses but indeed, true population structure, especially as the UK captive

population includes individuals from lineages with wider geographical distribution than those in the historical population.

The inclusion of the Cross River gorillas (contemporary and historical) was of benefit to the microsatellite analyses of the contemporary and historical populations of western lowland gorillas. Firstly, this allowed confirming that all the gorillas in this research belong to the western lowland subspecies, but also by identifying that Cluster 1 of the contemporary population was genetically more similar to the historical Cross River gorillas (Fig. 4.16) although considerable overlap was present in all the groups with exception of the contemporary Cross River population. This could support the observation of the two clusters in the UK captive population with the smaller Cluster 1 perhaps having historical origins and past levels of admixture with the Cross River gorillas, at a time of divergence, a hypothesis which may be supported by the findings of Thalmann *et al.* (2011) who found that substantial gene flow between the two subspecies of western gorillas continued after divergence and only ceased approximately 420 years ago.

5.3. Future work and impact case studies

This study was the first to provide a genetic analysis of the captive population of western lowland gorillas held at the Aspinall Foundation and as such provides a basis to build from. To further this research, expanding the genotyping analysis to encompass further individuals from the Aspinall Foundation (when samples become available) and to incorporate analysis from other institutions from around the world that hold populations of captive gorillas would be of benefit as it may identify further clusters and alleles currently not represented in the captive population sampled, and it would provide a more comprehensive dataset of the global captive population of gorillas.

Likewise, expanding the morphological research to encompass more specimens from different regions using other natural history collections would be of benefit to provide a more robust dataset. Additionally, including the remaining PCM samples for genotyping analysis may reveal further genetic clustering and also expanding to dataset to other natural history collections which include specimens from other regional areas such as Gabon and Equatorial Guinea, however, as demonstrated, one of the issues with museum specimens is the uncertainty or generalist information regarding precise geographical origins of the specimens.

Other methods could be applied for mtDNA analysis such as long-range PCR and cloning as have been employed in other mtDNA studies e.g. Clifford *et al.* (2004) and Soto-Caldrón *et al*. (2015). However, the prevalence of numt sequences will always be an issue for the mtDNA HVI, particularly for gorillas (Jensen-Seaman *et al*. 2004; Thalmann *et al*. 2005; Soto-Calderón *et al*. 2014), although sequencing of the whole mitochondrial genome, as discussed previously, removes the issues

associated with numts, and with the emergence of more sophisticated analyses such as whole genome sequencing and SNP analysis, mtDNA HVI sequence data may not be of considerable value. Restriction Site Associated DNA sequencing (RADseq) methods for obtaining SNPs, however, require high quality DNA (McMichael *et al.* 2009; Graham *et al.* 2015) which is notoriously difficult to obtain from degraded museum samples, as was the case in this study. However, if microsatellite genotyping is possible with low quality DNA, as demonstrated here, further research could focus on increasing the number of loci targeted from 10 to 30 plus, although Nsbuga *et al.* (2010) employed a panel of 32 microsatellite markers and obtained similar results.

Considering species conservation in the wider context, the PCM holds an abundance of unsampled and underutilised specimens which could yield valuable insights into past populations. For example, the genus *Galago* (bush babies) collection is relatively numerous and contains 107 specimens for *Euoticus elegantulus* and 61 specimens for *Galagoides demidovii,* and biogeographic and genetic analyses of these species would be of interest given relatively little-known information is available for them in comparison to many other primate species (S. Bearder, pers. comm, 2014).

The research presented in this thesis lends itself for further potential impact studies. One potential impact case relates to the genetic results revealed in Chapters 3 and 4. The western lowland gorilla UK captive population is managed via the international stud book which uses kinship for captive breeding management. To complement this, additional genetic data such as the results found in this research could be incorporated into the studbook and built upon to create genetic profiles of each individual. Although the studbook is obviously working as a captive management tool, proven by the UK and US captive populations genetic diversity being comparable to each other, and wild populations, it would be prudent and useful to create a genetic database linked to the studbook information. For example, gorillas identified as carrying private alleles or belonging to a specific haplogroup may be identified and selected for breeding or reintroduction programs to maintain genetic diversity in captive and wild populations, this information would not be apparent with the current studbook management. This would not only apply to the western lowland gorilla but there is potential for this to be achieved for any species or subspecies in captivity and is particularly relevant for endangered or critically endangered species where genetic diversity loss is a significant threat to the long-term survival of the species either in captivity or the wild.

Additionally, further impact studies are numerous in relation to the historical data. This research focused solely on the western lowland gorilla but the entirety of the PCM specimens were mapped and contextual information is available for many of the specimens in the collection. The methods applied here could easily be transferred to other endangered species and not necessarily just focused on endangered species. Biodiveristy as a whole is under threat (UN 2019), with 75% of primate species under pressure with decreasing population trends and 60% now considered

threatened with extinction (Estrada *et al*. 2017). Would it not be prudent to investigate the genetic diversity of historical populations of species not yet under threat? Where possible, create genetic databases of historic populations and all captive individuals and take a proactive approach rather than reactive, and not wait for numbers to dwindle to critical levels before we act, which inevitably, for some species, will be too late. There is a plethora of studies waiting to be investigated, not only in the PCM but in natural history collections globally. This research is just one example of such investigations but has identified potential for further research focusing on the western lowland gorillas and further potential for many more species. A wise Professor once said, "a good PhD thesis ends up asking more questions than it answers", this thesis did achieve many of the aims intended but throughout its duration, has raised many further questions which warrant investigation and provides future scope for research and impact studies.

5.4. General conclusions

In conclusion, the findings of this multidisciplinary research indicate that despite anthropogenic pressures and activities that are consistently and increasingly leading to the decline of biodiversity as a whole, the western lowland gorilla, despite being critically endangered, is genetically well placed to sustain future generations both in the wild and in captivity. Natural history collections like the PCM contain untapped biological resources which, in combination with adequate geographical information about the specimens and Geographical Information Systems, could complement ongoing conservation genetics studies as well as to inform current conservation projects by giving contextual information. Studying the morphological diversity of critically endangered organisms is also relevant for determining any population structure or regional differences. Coupled with genetic analyses, morphological studies help provide a broader picture of the biological diversity as well as generate hypotheses for future research on the causes of morphological differentiation. There is still a big gap in our understanding between morphology, genetics and adaptation to local environments, but with the development of more robust molecular, genetic and morphological techniques and analyses the gap is closing. Studying the neutral genetic variation, the genetic basis of adaptation and morphological variation of populations would further and significantly progress the conservation of the evolutionary potential of critically endangered species, like gorillas, under the current scenario of habitat fragmentation and loss.

References

Adams, D.C., Rohlf, F.J. and Slice, D.E., 2004. Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology*, *71*(1), pp.5-16.

Adams, D.C., Rohlf, F.J. and Slice, D.E., 2013. A field comes of age: geometric morphometrics in the 21st century. *Hystrix, 24*(1), p.7.

Akeley, C.E., 1923. *In Brightest Africa*. Garden City Publishing, Doubleday, New York.

Albrecht, G.H. and Miller, J.M., 1993. Geographic variation in primates. In *Species, species concepts and primate evolution* (pp. 123-161). Springer, Boston, MA.

Albrecht, G.H., Gelvin, B.R. and Miller, J.M., 2003. The hierarchy of intraspecific craniometric variation in gorillas: a population-thinking approach with implications for fossil species recognition studies. *Cambridge Studies in Biological and Evolutionary Anthropology*, pp.62-103.

Andersen, B.J., 2008. Research in the journal Landscape Ecology, 1987–2005. *Landscape Ecology*, *23*(2), pp.129-134.

Anderson, E.C. and Dunham, K.K., 2008. The influence of family groups on inferences made with the program Structure. *Molecular ecology resources, 8*(6), pp.1219-1229.

Andrews, C.A., 2010. Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. *Nature Education Knowledge, 3*(10), pp.5

Anthony, N.M., Clifford, S.L., Bawe-Johnson, M., Abernethy, K.A., Bru-ford, M.W. and Wickings, E.J., 2007. Distinguishing gorilla mitochondrial sequences from nuclear integrations and PCR recombinants: guidelines for their diagnosis in complex sequence databases. *Molecular Phylogenetics and Evolution, 43*, pp.553–566.

Arandjelovic, M. and Thalmann, O., 2001. Microsatellites in historic and ancient DNA. *e LS*.

Arandjelovic, M., Bergl, R.A., Ikfuingei, R., Jameson, C., Parker, M. and Vigilant, L., 2015. Detection dog efficacy for collecting faecal samples from the critically endangered Cross River gorilla (*Gorilla gorilla diehli*) for genetic censusing. *Royal Society Open Science*, *2*(2), pp.140423.

Arandjelovic, M., Head, J., Kuehl, H., Boesch, C., Robbins, M.M., Maisels, F. and Vigilant, L., 2010. Effective non-invasive genetic monitoring of multiple wild western gorilla groups. *Biological Conservation*, *143*(7), pp.1780-1791.

Aristide, L., Strauss, A., Halenar-Price, L.B., Gilissen, E., Cruz, F.W., Cartelle, C., Rosenberger, A.L., Lopes, R.T., dos Reis, S.F. and Perez, S.I., 2019. Cranial and endocranial diversity in extant and fossil atelids (Platyrrhini: Atelidae): A geometric morphometric study. *American Journal of Physical Anthropology*, *169*(2), pp.322-331.

Ashcroft, S.J. and Pereira, C., 2002. *Practical statistics for the biological sciences: simple pathways to statistical analyses*. Palgrave. UK.

Ashton, E.H. and Zuckerman, S., 1951. The influence of geographic isolation on the skull of the 12green monkey (*Cercopithecus aethips sabaeus*) IV. The degree and speed of dental 13 differentiation in the St Kitts green monkey. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *138*(892), pp.354-374.

Ashton, E.H. and Zuckerman, S., 1952. Age changes in the position of the occipital condyles in the chimpanzee and gorilla. *American Journal of Physical Anthropology*, *10*(3), pp.277-288.

Ashworth, V.E.T.M., Kobayashi, M.C., De La Cruz, M. and Clegg, M.T., 2004. Microsatellite markers in avocado (*Persea americana Mill.*): development of dinucleotide and trinucleotide markers. *Scientia Horticulturae*, *101*(3), pp.255-267.

Austin, J.J. and Melville, J., 2006. Incorporating historical museum specimens into molecular systematic and conservation genetics research. *Molecular Ecology Notes*, *6*(4), pp.1089-1092.

Avise, J.C., 1994. Conservation genetics. In *Molecular Markers, Natural History and Evolution* (pp. 361-398). Springer, Boston, MA.

Avise, J.C., 2012. *Molecular markers, natural history and evolution*. Springer Science & Business Media.

Ayala-Burbano, P.A., Junior, P.M.G., Wormell, D., Pissinatti, A., Marques, M.C. and de Freitas, P.D., 2020. Studbook and molecular analyses for the endangered black-lion-tamarin; an integrative approach for assessing genetic diversity and driving management in captivity. *Scientific reports, 10*(1), pp.1-11.

Baas, P., van der Valk, T., Vigilant, L., Ngobobo, U., Binyinyi, E., Nishuli, R., Caillaud, D. and Guschanski, K., 2018. Population-level assessment of genetic diversity and habitat fragmentation in critically endangered Grauer's gorillas. *American Journal of Physical Anthropology*, *165*(3), pp.565-575.

Bailey, S.E., 2002. A closer look at Neanderthal postcanine dental morphology: the mandibular dentition. *The Anatomical Record*, *269*(3), pp.148-156.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, *3*(10).

Ball, A.D., Stapley, J., Dawson, D.A., Birkhead, T.R., Burke, T. and Slate, J., 2010. A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (Taeniopygia guttata). *BMC genomics, 11*(1), pp.1-15.

Balolia, K.L., Soligo, C. and Wood, B., 2017. Sagittal crest formation in great apes and gibbons. *Journal of Anatomy*, *230*(6), pp.820-832.

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

Beerli, P. and Felsenstein, J., 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach*. Proceedings of the National Academy of Sciences, 98*(8), pp.4563-4568.

Beerli, P., Mashayekhi, S., Sadeghi, M., Khodaei, M., & Shaw, K. (2019). Population genetic inference with MIGRATE. *Current Protocols in Bioinformatics, 68*(1).

Beissinger, S.R. and Peery, M.Z., 2007. Reconstructing the historic demography of an endangered seabird. *Ecology*, *88*(2), pp.296-305.

Bender, J. and Ziegler, S., 2011. GIS based threat analysis of gorilla habitat in the Congo Basin. *Zentralblatt für Geologie und Paläontologie, Teil I*, 1(2), pp.97-111.

Berge, C. and Penin, X., 2004. Ontogenetic allometry, heterochrony, and interspecific differences in the skull of African apes, using tridimensional Procrustes analysis. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 124*(2), pp.124-138.

Bergl, R.A. and Vigilant, L., 2007. Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (*Gorilla gorilla diehli*). *Molecular Ecology*, *16*(3), pp.501-516.

Bergl, R.A., Bradley, B.J., Nsubuga, A. and Vigilant, L., 2008. Effects of habitat fragmentation, population size and demographic history on genetic diversity: the Cross River gorilla in a comparative context. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *70*(9), pp.848-859.

Bertram, B., 2004. Misconceptions about zoos. *Biologist*, *51*(4), pp.199-206.

Bignon, O., Baylac, M., Vigne, J.D. and Eisenmann, V., 2005. Geometric morphometrics and the population diversity of Late Glacial horses in Western Europe (*Equus caballus arcelini*): phylogeographic and anthropological implications. *Journal of Archaeological Science*, *32*(3), pp.375-391.

Bingpeng, X., Heshan, L., Zhilan, Z., Chunguang, W., Yanguo, W. and Jianjun, W., 2018. DNA barcoding for identification of fish species in the Taiwan Strait. *PloS one, 13*(6), p.e0198109.

Bird, C.E., Karl, S.A., Smouse, P.E. and Toonen, R.J., 2011. Detecting and measuring genetic differentiation. *Phylogeography and Population Genetics in Crustacea*, *19*(3), pp.1-55.

Blackith, R.E. and Reyment, R.A., 1971. *Multivariate morphometrics*. Academic Press. London-New York.

Boas, F., 1905. The horizontal plane of the skull and the general problem of the comparison of variable forms. *Science*, *21*(544), pp.862–863

Bond, W.J., 1994. Keystone species. In *Biodiversity and ecosystem function* (pp. 237-253). Springer, Berlin, Heidelberg.

Bookstein F. L., 1991. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, Cambridge.

Bookstein, F.L., 1996. Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, *58*(2), pp.313.

Bookstein, F.L., Chernoff, B., Elder, R.L., Humphries, J.M., Smith, G.R. and Strauss, R.E., 1985. *Morphometrics in evolutionary biology: the geometry of size and shape change, with examples from fishes*. Academy of Natural Sciences of Philadelphia; 1st Edition.

Borrell, J., 2016. Eight reasons why zoos are good for conservation. *Biologist*, *63*(5), p.9.

Boubli, J.P. and De Lima, M.G., 2009. Modeling the geographical distribution and fundamental niches of *Cacajao* spp. and *Chiropotes israelita* in Northwestern Amazonia via a maximum entropy algorithm. *International Journal of Primatology*, *30*(2), pp.217-228.

Boubli, J.P., da Silva, M.N.F., Amado, M.V., Hrbek, T., Pontual, F.B. and Farias, I.P., 2008. A taxonomic reassessment of *Cacajao melanocephalus* Humboldt (1811), with the description of two new species. *International Journal of Primatology*, *29*(3), pp.723-741.

Bradley, B.J., Boesch, C. and Vigilant, L., 2000. Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (*Pan troglodytes verus*) and gorilla (*Gorilla gorilla gorilla*) DNA from faeces. *Conservation Genetics*, *1*(3), pp.289-292.

Brennan, R.S., Garrett, A.D., Huber, K.E., Hargarten, H. and Pespeni, M.H., 2019. Rare genetic variation and balanced polymorphisms are important for survival in global change conditions. *Proceedings of the Royal Society B, 286*(1904), p.20190943.

Breuer, T., Robbins, A.M., Boesch, C. and Robbins, M.M., 2012. Phenotypic correlates of male reproductive success in western gorillas. *Journal of Human Evolution*, *62*(4), pp.466-472.

Breuer, T., Manguette, M. and Groenenberg, M., 2018. Gorilla Gorilla spp conservation–from zoos to the field and back: examples from the Mbeli Bai Study. *International Zoo Yearbook, 52*(1), pp.137-149.

Brimacombe, C.S., Kuykendall, K.L. and Nystrom, P., 2015. Analysis of correlations between measures of skeletal development and dental mineralization in Pan troglodytes. *American journal of physical anthropology, 158*(2), pp.279-287.

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C., 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *Journal of Molecular Evolution*, *18*(4), pp.225-239.

Bruford, M.W. and Wayne, R.K., 1993. Microsatellites and their application to population genetic studies. *Current Opinion in Genetics & Development*, *3*(6), pp.939-943.

Bruford, M.W., Cheesman, D.J., Coote, T., Green, H.A.A., Haines, S.A., O'Ryan, C. and Williams, T.R., 1996. Microsatellites and their application to conservation genetics. In: Smith TB and Wayne RK (eds) *Molecular Genetic Approaches in Conservation* (pp 278-297). Oxford University Press, New York.

Buckingham, F. and Shanee, S., 2009. Conservation priorities for the Peruvian yellow-tailed woolly monkey (*Oreonax flavicauda*): a GIS risk assessment and gap analysis. *Primate Conservation*, *24*(1), pp.65-72.

Burgman, M.A., Grimson, R.C. and Ferson, S., 1995. Inferring threat from scientific collections. *Conservation Biology*, *9*(4), pp.923-928.

Burgos, G., Simão, F., Flores-Espinoza, R., Yepez-Santos, J.I., Garzón-Salazar, A., Paz-Cruz, E., Freire-Paspuel, B., Carvalho, E.F. and Gusmão, L., 2019. An approach to maternal ancestry in a sample of Ecuadorian “mestizo” population by sequencing the control region of mtDNA. *Forensic Science International: Genetics Supplement Series, 7*(1), pp.537-538.

Burrell, A.S., Disotell, T.R. and Bergey, C.M., 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, *79*, pp.35-44.

Caillaud, D., Levréro, F., Gatti, S., Ménard, N. and Raymond, M., 2008. Influence of male morphology on male mating status and behavior during interunit encounters in western lowland gorillas. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *135*(4), pp.379-388.

Calafell, F., Shuster, A., Speed, W.C., Kidd, J.R. and Kidd, K.K., 1998. Short tandem repeat polymorphism evolution in humans, *European Journal of Human Genetics, 6*(1), pp.38-49.

Calvignac, S., Konecny, L., Malard, F. and Douady, C.J., 2011. Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion, 11*(2), pp.246-254.

Camiz, S., 1999. Comparison of Euclidean approximations of non-Euclidean distances. In *Classification and Data Analysis* (pp. 139-146). Springer, Berlin, Heidelberg.

Campbell, G., Junker, J., Boesch, C. and Kühl, H., 2012. Global A.P.E.S. status report. A report with information from the A.P.E.S. project.

Canhos, V.P., de Souza, S., De Giovanni, R. and Canhos, D.A.L., 2004. Global Biodiversity Informatics: setting the scene for a "new world" of ecological forecasting. *Biodiversity Informatics*, *1*, pp.1-13.

Cardini, A., Jansson, A.U. and Elton, S., 2007. A geometric morphometric approach to the study of ecogeographical and clinal variation in vervet monkeys. *Journal of Biogeography*, *34*(10), pp.1663-1678.

Cardini, A., Diniz Filho, J.A.F., Polly, P.D. and Elton, S., 2010. Biogeographic analysis using geometric morphometrics: clines in skull size and shape in a widespread African arboreal monkey. In *Morphometrics for nonmorphometricians* (pp. 191-217). Springer, Berlin, Heidelberg.

Cassalett, S. and Rothman, J.M., 2018. Diets (primates). *The International Encyclopedia of Biological Anthropology,* pp.1-10.

Chiszar, D., Murphy, J.B. and Iliff, W., 1990. For zoos. *The Psychological Record*, *40*(1), pp.3-13.

Cipolletta, C., 2003. Ranging patterns of a western gorilla group during habituation to humans in the Dzanga-Ndoki National Park, Central African Republic. *International Journal of Primatology*, *24*(6), pp.1207-1226.

Citek, J., Rehout, V., Hajkova, J. and Pavkova, J., 2006. Monitoring of the genetic health of cattle in the Czech Republic. *Veterinary Medicine*, *51*(6), pp.333-339.

Clark, W.L.G., 1950. Hominid characters of the australopithecine dentition. *Journal of the Anthropological Institute of Great Britain and Ireland*, *80*, pp.37-54.

Clifford, S.L., Abernethy, K.A., White, L.J.T., Tutin, C.E.G. and Bruford, M.W., 2003. Genetic studies of western gorillas, In *Gorilla Biology: A Multidisciplinary Perspective* (pp. 269–292), edited by Taylor AB, Goldsmith ML.

Clifford, S.L., Anthony, N.M., Bawe-Johnson, M., Abernethy, K.A., Tutin, C.E., White, L.J., Bermejo, M., Goldsmith, M.L., McFarland, K., Jeffery, K.J. and Bruford, M.W., 2004. Mitochondrial DNA phylogeography of western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology*, *13*(6), pp.1551-1565.

Cole, T.M., 1996. Historical note: early anthropological contributions to ''geometric morphometrics." *American Journal of Physical Anthropology, 101*(2), pp.291–296.

Conniff, R., 2009. Discovering gorilla. *Evolutionary Anthropology: Issues, News, and Reviews*, *18*(2), pp.55-61.

Consorte-McCrea, A., Bainbridge, A., Fernandez, A., Nigbur, D., McDonnell, S., Morin, A. and Grente, O., 2016. Understanding attitudes towards native wildlife and biodiversity in the UK: the role of zoos. *World Sustainability Series*. pp. 295-311. ISSN 2199-7373.

Cooke, S.B. and Terhune, C.E., 2015. Form, function, and geometric morphometrics. *The Anatomical Record, 298(*1), pp.5-28.

Coolidge Jr, H.J., 1929. A revision of the genus Gorilla. *Memoirs Museum Comparative Zoology at Harvard College*, *50*, pp.292-381.

Cooper, A., Mourer-Chauviré, C., Chambers, G.K., von Haeseler, A., Wilson, A.C. and Pääbo, S., 1992. Independent origins of New Zealand moas and kiwis. *Proceedings of the National Academy of Sciences*, *89*(18), pp.8741-8744.

Cooper, J.E. and Hull, G., 2017. *Gorilla Pathology and Health: With a Catalogue of Preserved Materials*. Academic Press.

Cooper, J.E., 2016. Personal Communication.

Cornuet, J.M. and Luikart, G., 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, *144*(4), pp.2001-2014.

Coutu, A.N., 2015. The elephant in the room: mapping the footsteps of historic elephants with big game hunting collections. *World Archaeology, 47*(3), pp.486-503.

Coyne, J.A. and Orr, H.A., 2004. Speciation: a catalogue and critique of species concepts. In *Philosophy of biology: an anthology* (pp.272-292). Wiley-Blackwell, Oxford.

Cranston, K.A. and Rannala, B., 2007. Summarizing a posterior distribution of trees using agreement subtrees. *Systematic Biology*, *56*(4), pp.578-590.

Crosilla, F., Beinat, A., Fusiello, A., Maset, E. and Visintini, D., 2019. Orthogonal Procrustes Analysis. In *Advanced Procrustes Analysis Models in Photogrammetric Computer Vision* (pp. 7-28). Springer, Cham.

d'Huart, J. and Grubb, P., 2003. Distribution of the common warthog (*Phacochoerus africanus*) and the desert warthog (*Phacochoerus aethipicus*) in the Horn of Africa. *African Journal of Ecology, 39*, pp. 156-169.

Dakin, E.E. and Avise, J.C., 2004. Microsatellite null alleles in parentage analysis. *Heredity*, *93*(5), pp.504.

Das, R., Hergenrother, S.D., Soto-Calderón, I.D., Dew, J.L., Anthony, N.M. and Jensen-Seaman, M.I., 2014. Complete mitochondrial genome sequence of the eastern gorilla (Gorilla beringei) and implications for African ape biogeography. *Journal of Heredity, 105*(6), pp.846-855.

Das, R., Roy, R. and Venkatesh, N., 2019. Using ancestry informative markers (AIMs) to detect fine structures within gorilla populations. *Frontiers in genetics, 10*, p.43.

Dayama, G., Zhou, W., Prado-Martinez, J., Marques-Bonet, T. and Mills, R., 2020. Characterization of nuclear mitochondrial insertions in the whole genomes of primates. *bioRxiv*.

De Vere, R.A., Warren, Y., Nicholas, A., Mackenzie, M.E. and Higham, J.P., 2011. Nest site ecology of the Cross River gorilla at the Kagwene Gorilla Sanctuary, Cameroon, with special reference to anthropogenic influence. *American Journal of Primatology, 73*(3), pp.253-261.

De Vicente, M.C., Lopez, C. and Fulton, T., 2004. *Genetic diversity analysis with molecular marker data: learning module*. International Plant Genetic Resources Institute (IPGRI), Rome, Italy.

Debray, R., Snyder-Mackler, N., Kohn, J., Wilson, M., Barreiro, L. and Tung, J., 2018. Social integration predicts mitochondrial DNA copy number in rhesus macaques. *bioRxiv,* p.408849.

DeLuycker, A.M., Heymann, E.W., Mittermeier, R.A., Ratsimbazafy, J., Rylands, A.B., Williamson, E.A., Oates, J.F., Mbora, D., Ganzhorn, J.U., Rodriguez-Luna, E. and Palacios, E., 2007. Peruvian yellow-tailed woolly monkey *Oreonax flavicauda* (Humboldt, 1812). *Primates in peril: the world's*, *25*, pp.2006-2008.

Den Tex, R.J., Maldonado, J.E., Thorington, R. and Leonard, J.A., 2010. Nuclear copies of mitochondrial genes: another problem for ancient DNA. *Genetica, 138*(9-10), pp.979-984.

DeSalle, R., Schierwater, B. and Hadrys, H., 2017. MtDNA: The small workhorse of evolutionary studies. *Frontiers in Bioscience*, *22*, pp.873-887.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J. and Ovenden, J.R., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), pp.209-214.

Dominy, N.J. and Duncan, B., 2002. GPS and GIS methods in an African rain forest: applications to tropical ecology and conservation. *Conservation Ecology*, *5*(2).

Dorado, G., Luque, F., Morales, P.P., Jiménez, I., Sánchez-Cañete, F.J.S., Raya, P., Martínez-Acitores, J.S., Sánchez, A., Tham, T.E.R., Sánchez, V.F.V. and Hernández, P., 2018. Evolution from first hominids to modern humans: philosophy, bioarchaeology and biology-Review. *Archaeobios,* (12), pp.69-82.

Doran, D.M., 1997. Ontogeny of locomotion in mountain gorillas and chimpanzees. *Journal of Human Evolution*, *32*(4), pp.323-344.

Doran, D.M. and McNeilage, A., 1998. Gorilla ecology and behavior. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *6*(4), pp.120-131.

Douadi, M.I., Gatti, S., Levréro, F., Duhamel, G., Bermejo, M., Vallet, D., Ménard, N. and Petit, E.J., 2007. Sex-biased dispersal in western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology*, *16*(11), pp.2247-2259.

Dryden, I. L. and Mardia, K. V., 1998. *Statistical Shape Analysis*. New York: John Wiley and Sons.

Dunmore, C.J., Kivell, T.L., Bardo, A. and Skinner, M.M., 2019. Metacarpal trabecular bone varies with distinct hand-positions used in hominid locomotion. *Journal of anatomy, 235*(1), pp.45-66.

Dupanloup, I., Schneider, S. and Excoffier, L., 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, *11*(12), pp.2571-2581.

E Crews, D. and M Gerber, L., 2003. Reconstructing life history of hominids and humans. *Collegium antropologicum*, *27*(1), pp.7-22.

Effenberger, S. and Suchentrunk, F., 1999. RFLP analysis of the mitochondrial DNA of otters (*Lutra lutra*) from Europe—implications for conservation of a flagship species. *Biological Conservation*, *90*(3), pp.229-234.

Enard, W. and Pääbo, S., 2004. Comparative primate genomics. *Annual Review of Genomics and Human Genetics*, *5*, pp.351-378.

Endangered Species (Import and Export) Act 1976 c.72. Available at: http://www.legislation.gov.uk/ukpga/1976/72 accessed 21st April 2020

Estrada, A., Garber, P.A., Rylands, A.B., Roos, C., Fernandez-Duque, E., Di Fiore, A., Nekaris, K.A.I., Nijman, V., Heymann, E.W., Lambert, J.E. and Rovero, F., 2017. Impending extinction crisis of the world's primates: Why primates matter. *Science Advances*, *3*(1), p.e1600946.

Evanno, G., Regnaut, S. and Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), pp.2611-2620.

Excoffier, L. and Lischer, H.E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), pp.564-567.

Excoffier, L., Laval, G. and Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, *1*, p.117693430500100003.

Excoffier, L., Smouse, P.E. and Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), pp.479-491.

Falush, D., Stephens, M. and Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics, 164*(4), pp.1567-1587.

Fan, H. and Chu, J.Y., 2007. A brief review of short tandem repeat mutation. *Genomics, Proteomics & Bioinformatics*, *5*(1), pp.7-14.

Fatica, L.M., Almécija, S., McFarlin, S.C. and Hammond, A.S., 2019. Pelvic shape variation among gorilla subspecies: Phylogenetic and ecological signals. *Journal of Human Evolution*, *137*, p.102684.

Ferchaud, A.L., Perrier, C., April, J., Hernandez, C., Dionne, M. and Bernatchez, L., 2016. Making sense of the relationships between Ne, Nb and Nc towards defining conservation thresholds in Atlantic salmon (*Salmo salar*). *Heredity*, *117*(4), p.268.

Feulner, P.G.D., Bielfeldt, W., Zachos, F.E., Bradvarovic, J., Eckert, I. and Hartl, G.B., 2004. Mitochondrial DNA and microsatellite analyses of the genetic status of the presumed subspecies *Cervus elaphus montanus* (Carpathian red deer). *Heredity*, *93*(3), p.299.

Field, A., 2013. *Discovering statistics using IBM SPSS statistics*. SAGE Publications Ltd, UK.

Finnilä, S., Lehtonen, M.S. and Majamaa, K., 2001. Phylogenetic network for European mtDNA. *The American Journal of Human Genetics*, *68*(6), pp.1475-1484.

Fiorenza, L. and Bruner, E., 2018. Cranial shape variation in adult howler monkeys (*Alouatta seniculus*). *American Journal of Primatology*, *80*(1), p.e22729.

Fischer, A., Pollack, J., Thalmann, O., Nickel, B. and Pääbo, S., 2006. Demographic history and genetic differentiation in apes. *Current Biology*, *16*(11), pp.1133-1138.

Flanagan, S.P., Forester, B.R., Latch, E.K., Aitken, S.N. and Hoban, S., 2018. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary applications, 11*(7), pp.1035-1052.

Fleagle, J.G., Gilbert, C.C. and Baden, A.L., 2010. Primate cranial diversity. *American Journal of Physical Anthropology*, *142*(4), pp.565-578.

Foody, G.M., 2008. GIS: biodiversity applications. *Progress in Physical Geography*, *32*(2), pp.223-235.

Forcina, G., Vallet, D., Le Gouar, P.J., Bernardo-Madrid, R., Illera, G., Molina-Vacas, G., Dréano, S., Revilla, E., Rodríguez-Teijeiro, J.D., Ménard, N. and Bermejo, M., 2019. From groups to communities in western lowland gorillas. *Proceedings of the Royal Society B, 286*(1896), p.20182019.

Fossey, D., 1979. Development of the mountain gorilla (*Gorilla gorilla beringei*): The first thirty-six months. In (DA Hamburg & ER McCown, Eds.) *The Great Apes*. Benjamin/Cummings, MP.

Fowler, J., Cohen, L. and Jarvis, P., 2013. *Practical statistics for field biology*. John Wiley & Sons.

Fowler, K.E., Reitter, C.P., Walling, G.A. and Griffin, D.K., 2012. Novel approach for deriving genome wide SNP analysis data from archived blood spots. *BMC Research Notes*, *5*(1), p.503.

Frankham, R., 1995. Conservation genetics. *Annual Review of Genetics*, *29*(1), pp.305-327.

Frankham, R., Briscoe, D.A. and Ballou, J.D., 2002. *Introduction to Conservation Genetics*. Cambridge university press.

Frankham, R., Ballou, J.D., Ralls, K., Eldridge, M., Dudash, M.R., Fenster, C.B., Lacy, R.C. and Sunnucks, P., 2017. *Genetic management of fragmented animal and plant populations.* Oxford University Press.

Franklin, D., Oxnard, C.E., O'Higgins, P. and Dadour, I., 2007. Sexual dimorphism in the subadult mandible: quantification using geometric morphometrics. *Journal of Forensic Sciences*, *52*(1), pp.6-10.

Fredsted, T., Pertoldi, C., Olesen, J.M., Eberle, M. and Kappeler, P.M., 2004. Microgeographic heterogeneity in spatial distribution and mtDNA variability of gray mouse lemurs (*Microcebus murinus, Primates: Cheirogaleidae*). *Behavioral Ecology and Sociobiology*, *56*(4), pp.393-403.

Fünfstück, T. and Vigilant, L., 2015. The geographic distribution of genetic diversity within gorillas. *American Journal of Primatology*, *77*(9), pp.974-985.

Fünfstück, T., Arandjelovic, M., Morgan, D.B., Sanz, C., Breuer, T., Stokes, E.J., Reed, P., Olson, S.H., Cameron, K., Ondzie, A. and Peeters, M., 2014. The genetic population structure of wild western lowland gorillas (*Gorilla gorilla gorilla*) living in continuous rain forest. *American Journal of Primatology*, *76*(9), pp.868-878.

Gagneux, P., Wills, C., Gerloff, U., Tautz, D., Morin, P.A., Boesch, C., Fruth, B., Hohmann, G., Ryder, O.A. and Woodruff, D.S., 1999. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proceedings of the National Academy of Sciences*, *96*(9), pp.5077-5082.

Garibaldi, A. and Turner, N., 2004. Cultural keystone species: implications for ecological conservation and restoration. *Ecology and society*, *9*(3).

Garner, A., Rachlow, J.L. and Hicks, J.F., 2005. Patterns of genetic diversity and its loss in mammalian populations. *Conservation Biology*, *19*(4), pp.1215-1221.

Garner, K.J. and Ryder, O.A., 1996. Mitochondrial DNA diversity in gorillas. *Molecular Phylogenetics and Evolution*, *6*(1), pp.39-48.

Garrod, B., 2017. *Primates of the Caribbean: using historical-era introduction of monkeys in the Lesser Antilles to understand patterns of island evolution* (Doctoral dissertation, UCL (University College London)).

Garson, G.D., 2012. *Testing statistical assumptions*. Statistical Associates Publishing, Asheboro, NC.

Gaubert, P., Papeş, M. and Peterson, A.T., 2006. Natural history collections and the conservation of poorly known taxa: ecological niche modeling in central African rainforest genets (*Genetta* spp.). *Biological Conservation*, *130*(1), pp.106-117.

Gayzik, F.S., Mao, M.Y., Danelson, K.A., Slice, D.E. and Stitzel, J.D., 2008. Quantification of age-related shape change of the human rib cage through geometric morphometrics. *Journal of Biomechanics*, *41*(7), pp.1545-1554.

Genton, C., Cristescu, R., Gatti, S., Levrero, F., Bigot, E., Caillaud, D., Pierre, J.S. and Menard, N., 2012. Recovery potential of a western lowland gorilla population following a major Ebola outbreak: results from a ten year study. *PloS One*, *7*(5), e37106

Gilbert, C.C., 2011. Phylogenetic analysis of the African papionin basicranium using 3-D geometric morphometrics: The need for improved methods to account for allometric effects. *American Journal of Physical Anthropology*, *144*(1), pp.60-71.

Gilbert, K.J., Andrew, R.L., Bock, D.G., Franklin, M.T., Kane, N.C., Moore, J.S., Moyers, B.T., Renaut, S., Rennison, D.J., Veen, T. and Vines, T.H., 2012. Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Molecular Ecology*, *21*(20), pp.4925-4930.

Gilbert, T. and Soorae, P.S., 2017. The role of zoos and aquariums in reintroductions and other conservation translocations. *Int. Zoo Yb. 51*: 9–14

Gómez-Robles, A., Martinón-Torres, M., De Castro, J.B., Margvelashvili, A., Bastir, M., Arsuaga, J.L., Pérez-Pérez, A., Estebaranz, F. and Martínez, L.M., 2007. A geometric morphometric analysis of hominin upper first molar shape. *Journal of Human Evolution*, *53*(3), pp.272-285.

Gooley, R.M., Hogg, C.J., Belov, K. and Grueber, C.E., 2018. The effects of group versus intensive housing on the retention of genetic diversity in insurance populations. *BMC Zoology, 3*(1), pp.1-12.

Gordon, A.D., Marcus, E. and Wood, B., 2013. Great ape skeletal collections: making the most of scarce and irreplaceable resources in the digital age. *American Journal of Physical Anthropology, 152*, pp.2-32.

Gott, T. and Weir, K., 2013. *Gorilla*. Reaktion Books Ltd, London, UK.

Goudet, J., 2005. FSTAT (version 2.9. 4): a program to estimate and test population genetics parameters.

Graham, C.F., Glenn, T.C., McArthur, A.G., Boreham, D.R., Kieran, T., Lance, S., Manzon, R.G., Martino, J.A., Pierson, T., Rogers, S.M. and Wilson, J.Y., 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, *15*(6), pp.1304-1315.

Granjon, A.C., Robbins, M.M., Arinaitwe, J., Cranfield, M.R., Eckardt, W., Mburanumwe, I., Musana, A., Robbins, A.M., Roy, J., Sollmann, R. and Vigilant, L., 2020. Estimating abundance and growth rates in a wild mountain gorilla population. *Animal Conservation*.

Greenbaum, G., Templeton, A.R., Zarmi, Y. and Bar-David, S., 2014. Allelic richness following population founding events–a stochastic modeling framework incorporating gene flow and genetic drift. *PloS One*, *9*(12).

Gregory, T., Mullett, A. and Norconk, M.A., 2014. Strategies for navigating large areas: a GIS spatial ecology analysis of the bearded saki monkey, *Chiropotes sagulatus*, in Suriname. *American Journal of Primatology*, *76*(6), pp.586-595.

Groves, C.P. and Stott Jr, K.W., 1979. Systematic relationships of gorillas from Kahuzi, Tshiaberimu and Kayonza. *Folia Primatologica, 32*(3), pp.161-179.

Groves, C.P., 1967. Ecology and taxonomy of the gorilla. *Nature*, *213*(5079), pp.890-893.

Groves, C.P., 1970. Population systematics of the gorilla. *Journal of Zoology*, *161*(3), pp.287-300.

Groves, C.P., 2002. *A history of gorilla taxonomy*. Cambridge: Cambridge University Press.

Grubb, P., Butynski, T.M., Oates, J.F., Bearder, S.K., Disotell, T.R., Groves, C.P. and Struhsaker, T.T., 2003. Assessment of the diversity of African primates. *International Journal of Primatology, 24*, pp.1301-1357.

Guiller, A., Bellido, A. and Madec, L., 1998. Genetic distances and ordination: the land snail Helix aspersa in North Africa as a test case. *Systematic Biology*, *47*(2), pp.208-227.

Guinand, B., Scribner, K.T., Page, K.S., Filcek, K., Main, L. and Burnham-Curtis, M.K., 2006. Effects of coancestry on accuracy of individual assignments to population of origin: examples using Great Lakes lake trout (Salvelinus namaycush). *Genetica, 127*(1-3), pp.329-340.

Gulcher, J., 2012. Microsatellite markers for linkage and association studies. *Cold Spring Harbor Protocols*, *2012*(4), pp.pdb-top068510.

Gunbin, K., Peshkin, L., Popadin, K., Annis, S., Ackermann, R.R. and Khrapko, K., 2017. Integration of mtDNA pseudogenes into the nuclear genome coincides with speciation of the human genus. A hypothesis. *Mitochondrion, 34*, pp.20-23.

Guschanski, K., Caillaud, D., Robbins, M.M. and Vigilant, L., 2008. Females shape the genetic structure of a gorilla population. *Current Biology*, *18*(22), pp.1809-1814.

Guschanski, K., Vigilant, L., McNeilage, A., Gray, M., Kagoda, E. and Robbins, M.M., 2009. Counting elusive animals: comparing field and genetic census of the entire mountain gorilla population of Bwindi Impenetrable National Park, Uganda. *Biological Conservation*, *142*(2), pp.290-300.

Guschanski, K., Krause, J., Sawyer, S., Valente, L.M., Bailey, S., Finstermeier, K., Sabin, R., Gilissen, E., Sonet, G., Nagy, Z.T. and Lenglet, G., 2013. Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology*, *62*(4), pp.539-554.

Hacia, J.G., 2001. Genome of the apes. *Trends in Genetics*, *17*(11), pp.637-645.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D. and Cook, W.M., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science advances, 1*(2), p.e1500052.

Haddow, A.J. and Ross, R.W., 1951. A critical review of Coolidge's measurements of gorilla skulls. In *Proceedings of the Zoological Society of London* (Vol. 121, No. 1, pp. 43-54). Oxford, UK: Blackwell Publishing Ltd.

Hagemann, L., Boesch, C., Robbins, M.M., Arandjelovic, M., Deschner, T., Lewis, M., Froese, G. and Vigilant, L., 2018. Long-term group membership and dynamics in a wild western lowland gorilla population (*Gorilla gorilla gorilla*) inferred using non-invasive genetics. *American Journal of Primatology*, *80*(8), pp.e22898.

Hager, L.D., 1996. Sex differences in the sciatic notch of great apes and modern humans. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *99*(2), pp.287-300.

Haikal, M., 2020. *Master Pongo: A Gorilla Conquers Europe* (Vol. 17). Penn State Press.

Hale, S.L. and Koprowski, J.L., 2018. Ecosystem-level effects of keystone species reintroduction: a literature review. *Restoration Ecology, 26*(3), pp.439-445.

Hall, B.G., 2011. *Phylogenetic trees made easy: A how to manual*. Sinauer.

Hall, T., 2016. BioEdit v7. 2.5: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, pp.95-98.

Hamada, H., Petrino, M.G. and Kakunaga, T., 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences*, *79*(21), pp.6465-6469.

Hamilton, M., 2011. *Population genetics*. John Wiley & Sons.

Hammer, Ø., Harper, D.A. and Ryan, P.D., 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, *4*(1), p.9.

Harcourt, A.H. and Stewart, K.J., 2007. Gorilla society: What we know and don't know. Evolutionary Anthropology: *Issues, News, and Reviews, 16*(4), pp.147-158.

Harpending, H.C., 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, pp.591-600.

Harpending, H.C., Sherry, S.T., Rogers, A.R. and Stoneking, M., 1993. The genetic structure of ancient human populations. *Current Anthropology*, *34*(4), pp.483-496.

Harris, E.E. and Disotell, T.R., 1998. Nuclear gene trees and the phylogenetic relationships of the mangabeys (Primates: Papionini). *Molecular Biology and Evolution*, *15*(7), pp.892-900.

Hartl, D.L. and Clark, A.G. 1997. *Principles of population genetics* (Vol. 116). Sinauer associates, Sunderland, MA.

Hasegawa, M., Kishino, H., Hayasaka, K. and Horai, S., 1990. Mitochondrial DNA evolution in primates: transition rate has been extremely low in the lemur. *Journal of Molecular Evolution*, *31*(2), pp.113-121.

Haurez, B., Tagg, N., Petre, C.A., Brostaux, Y., Boubady, A. and Doucet, J.L., 2018. Seed dispersal effectiveness of the western lowland gorilla (Gorilla gorilla gorilla) in Gabon. *African Journal of Ecology, 56*(2), pp.185-193.

Hayasaka, K., Gojobori, T. and Horai, S., 1988. Molecular phylogeny and evolution of primate mitochondrial DNA. *Molecular Biology and Evolution*, *5*(6), pp.626-644.

He, G., Meng, R., Newman, M., Gao, G., Pittman, R.N. and Prakash, C.S., 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biology*, *3*(1), p.3.

Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H. and Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*, *101*(41), pp.14812-14817.

Hedrick, B.P., Heberling, J.M., Meineke, E.K., Turner, K.G., Grassa, C.J., Park, D.S., Kennedy, J., Clarke, J.A., Cook, J.A., Blackburn, D.C. and Edwards, S.V., 2020. Digitization and the future of natural history collections*. BioScience, 70*(3), pp.243-251.

Hemley, G., 1994. *International wildlife trade: a CITES sourcebook*. Island Press.

Hennessy, R.J. and Stringer, C.B., 2002. Geometric morphometric study of the regional variation of modern human craniofacial form. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *117*(1), pp.37-48.

Herzfeld, C., 2017. *The great apes: a short history*. Yale University Press.

Hey, J., Waples, R.S., Arnold, M.L., Butlin, R.K. and Harrison, R.G., 2003. Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution*, *18*(11), pp.597-603.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A. and Wilson, A.C., 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, *312*(5991), pp.282-284.

Hill, A., Guralnick, R., Smith, A., Sallans, A., Gillespie, R., Denslow, M., Gross, J., Murrel, Z., Conyers, T., Oboyski, P., Ball, J., Thomer, A., Prys-Jones, R., de la Torre, J., Kociolek, P. and Fortson, L., 2012. The notes from nature tool for unlocking biodiversity records through citizen science. *Zookeys*, 209, pp.219-233

Hill, W.G., 1981. Estimation of effective population size from data on linkage disequilibrium. *Genetics Research*, *38*(3), pp.209-216.

Hillis, D.M., Moritz, C. and Mable, B.K. eds., 1996. *Molecular systematics* (Vol. 23). Sinauer Associates, Sunderland, MA.

Hlaing, T., Tun-Lin, W., Somboon, P., Socheat, D., Setha, T., Min, S., Chang, M.S. and Walton, C., 2009. Mitochondrial pseudogenes in the nuclear genome of Aedes aegypti mosquitoes: implications for past and future population genetic studies. *BMC Genetics*, *10*(1), p.11.

Hoffman, J.I. and Amos, W., 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular ecology, 14*(2), pp.599-612.

Hofreiter, M., Siedel, H., Van Neer, W. and Vigilant, L., 2003. Mitochondrial DNA sequence from an enigmatic gorilla population (*Gorilla gorilla uellensis*). *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *121*(4), pp.361-368.

Hogan, L.A. and Tribe, A., 2007. Prevalence and cause of stereotypic behaviour in common wombats (*Vombatus ursinus*) residing in Australian zoos. *Applied Animal Behaviour Science*, *105*(1-3), pp.180-191.

Holmes, M.W., Hammond, T.T., Wogan, G.O., Walsh, R.E., LaBarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M. and Nachman, M.W., 2016. Natural history collections as windows on evolutionary processes. *Molecular Ecology*, *25*(4), pp.864-881.

Holton, N.E., Nicholas, C.L., Marshall, S.D., Franciscus, R.G. and Southard, T.E., 2015. The effects of altered maxillary growth on patterns of mandibular rotation in a pig model. *Archives of Oral Biology*, *60*(6), pp.933-940.

Honnay, O. and Jacquemyn, H., 2007. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conservation Biology*, *21*(3), pp.823-831.

Hopkins, M.E., 2016. Mantled howler monkey spatial foraging decisions reflect spatial and temporal knowledge of resource distributions. *Animal Cognition*, *19*(2), pp.387-403.

Horovitz, I. and Meyer, A., 1995. Systematics of New World monkeys (*Platyrrhini*, Primates) based on 16S mitochondrial DNA sequences: a comparative analysis of different weighting methods in cladistic analysis. *Molecular Phylogenetics and Evolution*, *4*(4), pp.448-456.

Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. and Ladle, R.J., 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics, 46*, pp.523-549.

Hubisz, M.J., Falush, D., Stephens, M. and Pritchard, J.K., 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*(5), pp.1322-1332.

Hudson, R.R., 2004. Linkage disequilibrium and recombination. *Handbook of statistical genetics*.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R. and Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, *294*(5550), pp.2310-2314.

Hughes, J.F. and Rozen, S., 2012. Genomics and genetics of human and primate Y chromosomes. *Annual Review of Genomics and Human Genetics*, *13*, pp.83-108.

Huson, D.H., Rupp, R. and Scornavacca, C., 2010. *Phylogenetic networks: concepts, algorithms and applications*. Cambridge University Press.

Hutchins, M., Smith, B. and Allard, R., 2003. In defense of zoos and aquariums: the ethical basis for keeping wild animals in captivity. *Journal of the American Veterinary Medical Association*, *223*(7), pp.958-966.

Huxley, C., 2013, November. CITES: the vision. In *Endangered Species Threatened Convention* (pp. 23-32). Routledge.

International Society of Genetic Genealogy 2020 available at: https://isogg.org/wiki/Genetics_Glossary accessed on 22nd April 2020

IPBES. (2019). IPBES Global Assessment Preview. Retrieved 15th June, 2019, from https://www.ipbes.net/news/ipbes-global-assessment-preview.

Irish, J.D., 1998. Ancestral dental traits in recent Sub-Saharan Africans and the origins of modern humans. *Journal of Human Evolution*, *34*(1), pp.81-98.

Isaac, N.J., Mallet, J. and Mace, G.M., 2004. Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution*, *19*(9), pp.464-469.

Ito, T. and Koyabu, D., 2018. Biogeographic variation in skull morphology across the Kra Isthmus in dusky leaf monkeys. *Journal of Zoological Systematics and Evolutionary Research*, *56*(4), pp.599-610.

IUCN 2019. The IUCN Red List of Threatened Species. Version 2019-3. http://www.iucnredlist.org. Downloaded on 10 December 2019

Jamieson, D., 1985. Against zoos. *Environmental Ethics: Readings in Theory and Application*, *5*, pp.97-103.

Janes, J.K., Miller, J.M. and Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I. and Andrew, R.L., 2017. The K= 2 conundrum. *Molecular Ecology*, *26*(14), pp.3594-3602.

Jeffery, K.J., Abernethy, K.A., Tutin, C.E. and Bruford, M.W., 2007. Biological and environmental degradation of gorilla hair and microsatellite amplification success. *Biological Journal of the Linnean Society*, *91*(2), pp.281-294.

Jeffreys, A.J., Wilson, V. and Thein, S.L., 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature*, *314*(6006), pp.67-73.

Jenkins, P.D., 1990. *Catalogue of Primates in the British Museum (Natural History) and elsewhere in the British Isles. Part 5: Superfamily Hominoidea*. British Museum (Natural History), London.

Jensen-Seaman, M.I. and Kidd, K.K., 2001. Mitochondrial DNA variation and biogeography of eastern gorillas. *Molecular Ecology*, *10*(9), pp.2241-2247.

Jensen-Seaman, M.I., Sarmiento, E.E., Deinard, A.S. and Kidd, K.K., 2004. Nuclear integrations of mitochondrial DNA in gorillas. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *63*(3), pp.139-147.

Jones, O.R. and Wang, J., 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular ecology resources, 10*(3), pp.551-555.

Joshi, J., Salar, R.K. and Banerjee, P., 2013. Genetic variation and phylogenetic relationships of Indian buffaloes of Uttar Pradesh. *Asian-Australasian Journal of Animal Sciences*, *26*(9), p.1229.

Jungers, W.L., Falsetti, A.B. and Wall, C.E., 1995. Shape, relative size, and size-adjustments in morphometrics. *American Journal of Physical Anthropology*, *38*(21), pp.137-161.

Junker, J., Blake, S., Boesch, C., Campbell, G., Toit, L.D., Duvall, C., Ekobo, A., Etoga, G., Galat-Luong, A., Gamys, J. and Ganas-Swaray, J., 2012. Recent decline in suitable environmental conditions for A frican great apes. *Diversity and Distributions*, *18*(11), pp.1077-1091.

Jurmain, R., 1997. Skeletal evidence of trauma in African apes, with special reference to the Gombe chimpanzees. *Primates, 38*(1), pp.1-14.

Kaessmann, H., Wiebe, V. and Pääbo, S., 1999. Extensive nuclear DNA sequence diversity among chimpanzees. *Science*, *286*(5442), pp.1159-1162.

Kaeuffer, R., Réale, D., Coltman, D.W. and Pontier, D., 2007. Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity, 99*(4), pp.374-380.

Kalinowski, S.T., Wagner, A.P. and Taper, M.L., 2006. ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes, 6*(2), pp.576-579.

Kaliontzopoulou, A., Carretero, M.A. and Llorente, G.A., 2007. Multivariate and geometric morphometrics in the analysis of sexual dimorphism variation in Podarcis lizards. *Journal of Morphology*, *268*(2), pp.152-165.

Kamilar, J.M., Muldoon, K.M., Lehman, S.M. and Herrera, J.P., 2012. Testing Bergmann's rule and the resource seasonality hypothesis in Malagasy primates using GIS-based climate data. *American Journal of Physical Anthropology*, *147*(3), pp.401-408.

Kanthaswamy, S., Kurushima, J.D. and Smith, D.G., 2006. Inferring Pongo conservation units: a perspective based on microsatellite and mitochondrial DNA analyses. *Primates*, *47*(4), pp.310-321.

Kapoor, B.G. and Khanna, B., 2004. *Ichthyology handbook*. Springer Science & Business Media.

Kendall, K.C., Stetz, J.B., Boulanger, J., Macleod, A.C., Paetkau, D. and White, G.C., 2009. Demography and genetic structure of a recovering grizzly bear population. *The Journal of Wildlife Management*, *73*(1), pp.3-16.

Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W.M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J. and Liu, W., 2003. Large-scale genotyping of complex DNA. *Nature biotechnology, 21*(10), pp.1233-1237.

Kharouba, H.M., Lewthwaite, J.M., Guralnick, R., Kerr, J.T. and Vellend, M., 2019. Using insect natural history collections to study global change impacts: challenges and opportunities. *Philosophical Transactions of the Royal Society B*, *374*(1763), pp.20170405.

Kimura, M., 1983. *The neutral theory of molecular evolution*. Cambridge University Press.

King, T. and Courage, A., 2008. Western gorilla re-introduction to the Batéké Plateau region of Congo and Gabon. In *Global re-introduction perspectives: Re-introduction case-studies from around the globe* (pp.217-220). IUCN/SSC Reintroduction Specialist Group (RSG), Gland, Switzerland and Environment Agency, Abu Dhabi, UAE.

King, T., 2013. The Aspinall Foundation: 30 years of captive breeding, reintroduction and conservation. *Wild Conservation*, *1*, pp.76-89.

King, T., Chamberlan, C. and Courage, A., 2012. Assessing initial reintroduction success in long-lived primates by quantifying survival, reproduction, and dispersal parameters: western lowland gorillas (*Gorilla gorilla gorilla*) in Congo and Gabon. *International Journal of Primatology*, *33*(1), pp.134-149.

King, T., Chamberlan, C. and Courage, A., 2014. Assessing reintroduction success in long-lived primates through population viability analysis: western lowland gorillas *Gorilla gorilla gorilla* in Central Africa. *Oryx*, *48*(2), pp.294-303.

King, T., 2019. Personal Communication.

Kirchshofer, R., 1985. *International register and studbook of the gorilla*. Frankfurt AM Zoological Garden, Frankfurt, Germany.

Kitchener, A.C., 1997. The role of museums and zoos in conservation biology. *International Zoo Yearbook*, *35*(1), pp.325-336.

Knebelsberger, T. and Stöger, I., 2012. DNA extraction, preservation, and amplification. In *DNA Barcodes* (pp. 311-338). Humana Press, Totowa, NJ.

Knigge, R.P., Tocheri, M.W., Orr, C.M. and Mcnulty, K.P., 2015. Three-dimensional geometric morphometric analysis of talar morphology in extant gorilla taxa from highland and lowland habitats. *The Anatomical Record*, *298*(1), pp.277-290.

Kolleck, J., Yang, M., Zinner, D. and Roos, C., 2013. Genetic diversity in endangered Guizhou snub-nosed monkeys (*Rhinopithecus brelichi*): contrasting results from microsatellite and mitochondrial DNA data. *PLoS One*, *8*(8).

Krishtalka, L. and Humphrey, P.S., 2000. Can natural history museums capture the future? *BioScience*, *50*(7), pp.611-617

Kuhner, M.K., 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics, 22*(6), pp.768-770.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), pp.1870-1874.

Kunz, D., Tay, W.T., Elfekih, S., Gordon, K.H.J. and De Barro, P.J., 2019. Take out the rubbish–Removing NUMTs and pseudogenes from the Bemisia tabaci cryptic species mtCOI database. *bioRxiv*, p.724765.

Kuo, C.H. and Janzen, F.J., 2003. bottlesim: a bottleneck simulation program for long-lived species with overlapping generations. *Molecular Ecology Notes*, *3*(4), pp.669-673.

Lacy, R.C., 1997. Importance of genetic variation to the viability of mammalian populations. *Journal of Mammalogy*, *78*(2), pp.320-335.

Lande, R. and Barrowclough, G.F., 1987. Effective population size, genetic variation, and their use in population management. In *Viable Populations for Conservation*, ed. ME Soule (pp. 87- 123). Cambridge University Press, New York.

Lande, R., 1988. Genetics and demography in biological conservation. *Science*, *241*(4872), pp.1455-1460.

Lawing, A.M. and Polly, P.D., 2010. Geometric morphometrics: recent applications to the study of evolution and development. *Journal of Zoology, 280*(1), pp.1-7.

Lawrence, D.P., Travadon, R. and Baumgartner, K., 2015. Diversity of Diaporthe species associated with wood cankers of fruit and nut crops in northern California. *Mycologia, 107*(5), pp.926-940.

Legendre, P. and Legendre, L.F., 2012. *Numerical ecology* (Vol. 24). Elsevier.

Leigh, S.R., Relethford, J.H., Park, P.B. and Konigsberg, L.W., 2003. Morphological differentiation of Gorilla subspecies. *Gorilla Biology: A Multidisciplinary Perspective*, pp.104-131.

Leigh, D.M., Hendry, A.P., Vázquez-Domínguez, E. and Friesen, V.L., 2019. Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications, 12*(8), pp.1505-1512.

Lemopoulos, A., Prokkola, J.M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M.L., Koskiniemi, J. and Vainikka, A., 2019. Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution, 9(*4), pp.2106-2120.

Li, M., Schroeder, R., Ko, A. and Stoneking, M., 2012. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic acids research, 40*(18), pp.e137-e137.

Li, Y.L. and Liu, J.X., 2018. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, *18*(1), pp.176-177.

Lin, D., 2014. Arguments For and Against Zoos. *About News, December*, *10*.

Lindburg, D.G., Berkson, J.M. and Nightenhelser, L.K., 1984. Primate breeding in zoos: a ten year summary. In *One Medicine* (pp. 162-170). Springer, Berlin, Heidelberg.

Linhart, P., Adams, D.B. and Voracek, T., 2008. The international transportation of zoo animals: conserving biological diversity and protecting animal welfare. *Veterinaria Italiana*, *44*, pp.49-57.

Lippe, C., Dumont, P. and Bernatchez, L., 2006. High genetic diversity and no inbreeding in the endangered copper redhorse, *Moxostoma hubbsi* (Catostomidae, Pisces): the positive sides of a long generation time. *Molecular Ecology*, *15*(7), pp.1769-1780.

Lister, A.M. and Climate Change Research Group, 2011. Natural history collections as sources of long-term datasets. *Trends in Ecology & Evolution*, *26*(4), pp.153-154.

Litt, M. and Luty, J.A., 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, *44*(3), pp.397.

Lodwick, J.L. and Salmi, R., 2019. Nutritional composition of the diet of the western gorilla (Gorilla gorilla): Interspecific variation in diet quality. *American journal of primatology, 81*(9), p.e23044.

Losos, J.B., Arnold, S.J., Bejerano, G., Brodie III, E.D., Hibbett, D., Hoekstra, H.E., Mindell, D.P., Monteiro, A., Moritz, C., Orr, H.A. and Petrov, D.A., 2013. Evolutionary biology for the 21st century. *PLoS Biology*, *11*(1), pp.e1001466.

Lukas, D., Bradley, B.J., Nsubuga, A.M., Doran-Sheehy, D., Robbins, M.M. and Vigilant, L., 2004. Major histocompatibility complex and microsatellite variation in two populations of wild gorillas. *Molecular Ecology*, *13*(11), pp.3389-3402.

Lycett, S.J., von Cramon-Taubadel, N. and Gowlett, J.A., 2010. A comparative 3D geometric morphometric analysis of Victoria West cores: implications for the origins of Levallois technology. *Journal of Archaeological Science*, *37*(5), pp.1110-1117.

Mace, G.M., 1988. The genetic and demographic status of the Western lowland gorilla (*Gorilla g. gorilla*) in captivity. *Journal of Zoology*, *216*(4), pp.629-654.

MacLeod, N., 2017. Morphometrics: History, development methods and prospects. *动物分类学报 , 42*(1), pp.4-33.

Magdeldin, S. ed., 2012. *Gel electrophoresis: Principles and basics*. BoD–Books on Demand.

Magliocca, F., Querouil, S., and Gautier Hion, A. (1999). Population structure and group composition of Western Lowland Gorillas in North-Western Republic of Congo. *American Journal of Primatology, 48*, pp.1–14.

Maigret, T.A., 2019. Snake scale clips as a source of high quality DNA suitable for RAD sequencing. *Conservation Genetics Resources, 11*(4), pp.373-375.

Maisels, F., Bergl, R.A. and Williamson, E.A., 2018. Gorilla gorilla (amended version of 2016 assessment). The IUCN Red List of Threatened Species 2018: e. T9404A136250858.

Marcus, L., Hingst-Zaher, E. and Zaher, H., 2000. Application of landmark morphometrics to skulls representing the orders of living mammals. *Hystrix, The Italian Journal of Mammalogy*, *11*(1), pp. 27-47.

Marcus, L.F., 1990. Traditional morphometrics. In *Proceedings of the Michigan morphometrics workshop* (Vol. 2, pp. 77-122). Special Publication.

Martellucci, C.A., Sah, R., Rabaan, A.A., Dhama, K., Casalone, C., Arteaga-Livias, K., Sawano, T., Ozaki, A., Bhandari, D., Higuchi, A. and Kotera, Y., 2020. Changes in the spatial distribution of COVID-19 incidence in Italy using GIS-based maps. *Annals of Clinical Microbiology and Antimicrobials, 19*(1), pp.1-4.

Martinez-Maza, C., Freidline, S.E., Strauss, A. and Nieto-Diaz, M., 2016. Bone growth dynamics of the facial skeleton and mandible in *Gorilla gorilla* and *Pan troglodytes*. *Evolutionary Biology*, *43*(1), pp.60-80.

Masel, J., 2012. Rethinking Hardy–Weinberg and genetic drift in undergraduate biology. *BioEssays*, *34*(8), pp.701-710.

Mathieson, V., 2019. Personal Communication.

Mathieson, V., 2020. Personal Communication.

Matthee, C.A. and Robinson, T.J., 1999. Mitochondrial DNA population structure of roan and sable antelope: implications for the translocation and conservation of the species. *Molecular Ecology*, *8*(2), pp.227-238.

Mayr, E., 1957. *The Species Problem* (pp.395). American Association for the Advancement of Science, Washington, DC, USA.

Mayr, E., 1963. *Animal Species and Evolution* (pp.797). Harvard University Press, Cambridge, Massachusetts, USA.

Mayr, E., 1969. The biological meaning of species. *Biological Journal of the Linnean Society*, *1*, pp.311-320.

McDonald, D.B., 2008. Heterozygosity, HExp (or gene diversity, D). *Lecture 4. Population Genetics II*.

McHenry, H.M., 1982. The pattern of human evolution: studies on bipedalism, mastication, and encephalization. *Annual Review of Anthropology, 11*, pp.151-73.

McLain, A.T. and Faulk, C., 2018. The evolution of CpG density and lifespan in conserved primate and mammalian promoters. *Aging (Albany NY), 10*(4), p.561.

McManus, K.F., Kelley, J.L., Song, S., Veeramah, K.R., Woerner, A.E., Stevison, L.S., Ryder, O.A., Ape Genome Project, G., Kidd, J.M., Wall, J.D. and Bustamante, C.D., 2015. Inference of gorilla demographic and selective history from whole-genome sequence data. *Molecular biology and evolution, 32*(3), pp.600-612.

McMichael, G.L., Gibson, C.S., O'Callaghan, M.E., Goldwater, P.N., Dekker, G.A., Haan, E.A., MacLennan, A.H. and South Australian Cerebral Palsy Research Group, 2009. DNA from buccal swabs suitable for high-throughput SNP multiplex analysis. *Journal of Biomolecular Techniques: JBT*, *20*(5), p.232.

Meirmans, P.G. and Liu, S., 2018. Analysis of molecular variance (AMOVA) for autopolyploids. *Frontiers in Ecology and Evolution*, *6*, p.66.

Meirmans, P.G., 2012. AMOVA-based clustering of population genetic data. *Journal of Heredity*, *103*(5), pp.744-750.

Mekonnen, A., Fashing, P.J., Bekele, A. and Stenseth, N.C., 2020. Distribution and conservation status of Boutourlini's blue monkey (Cercopithecus mitis boutourlinii), a Vulnerable subspecies endemic to western Ethiopia. *Primates, 61*, pp.785-796.

Meyer, E., Wiegand, P., Rand, S.P., Kuhlmann, D., Brack, M. and Brinkmann, B., 1995. Microsatellite polymorphisms reveal phylogenetic relationships in primates. *Journal of Molecular Evolution*, *41*(1), pp.10-14.

Michalakis, Y. and Excoffier, L., 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, *142*(3), pp.1061-1064.

Mittermeier, R.A., Rylands, A.B. and Wilson, D.E., 2013. *Handbook of the mammals of the world*. Vol. 3. Primates. Lynx Edicons, Barcelona.

Mitteroecker, P. and Gunz, P., 2009. Advances in geometric morphometrics. *Evolutionary Biology*, *36*(2), pp.235-247.

Mitteroecker, P., Gunz, P., Windhager, S. and Schaefer, K., 2013. A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix, the Italian Journal of Mammalogy*, *24*(1), pp.59-66.

Moraes, B., Razgour, O., Souza-Alves, J.P., Boubli, J.P. and Bezerra, B., 2020. Habitat suitability for primate conservation in north-east Brazil. *Oryx, 54*(6), pp.803-813.

Morgan, D., 2007. *Best practice guidelines for reducing the impact of commercial logging on great apes in Western Equatorial Africa*. IUCN.

Morgan, K.N. and Tromborg, C.T., 2007. Sources of stress in captivity. *Applied Animal Behaviour Science*, *102*(3-4), pp.262-302.

Moritz, C., 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, *3*(4), pp.401-411.

Moritz, C., 1995. Uses of molecular phylogenies for conservation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *349*(1327), pp.113-118.

Morton, N. E. (2001). *Linkage Disequilibrium. Encyclopedia of Genetics, 1105.* doi:10.1006/rwgn.2001.0767

Muniz, A.C., Lemos-Filho, J.P., de Oliveira Buzatti, R.S., Ribeiro, P.C.C., Fernandes, F.M. and Lovato, M.B., 2019. Genetic data improve the assessment of the conservation status based only on herbarium records of a Neotropical tree. *Scientific Reports*, *9*(1), pp.1-11.

Murphey, P.C., Guralnick, R.P., Glaubitz, R., Neufeld, D. and Ryan, J.A., 2004. Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the Mountain and Plains Spatio-Temporal Database-Informatics Initiative (MaPSTeDI). *PhyloInformatics*, *3*, pp.1-29.

Mutai, S. and Chang, L., 2020. Post-fire hazard detection using ALOS-2 radar and landsat-8 optical imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 6,* pp.75-82.

Neel, J.V. and Thompson, E.A., 1978. Founder effect and number of private polymorphisms observed in Amerindian tribes. *Proceedings of the National Academy of Sciences*, 75, pp.1904-1908.

Nei, M. and Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), pp.5269-5273.

Neigel, J.E., 1996. Estimation of effective population size and migration parameters from genetic data. In *Molecular Genetic Approaches in Conservation* (pp.329-346). Oxford University Press, Oxford.

Neigel, J.E., 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics*, *28*(1), pp.105-128.

Neumann, K. and Wetton, J.H., 1996. Highly polymorphic microsatellites in the house sparrow Passer domesticus. *Molecular Ecology*, *5*(2), pp.307-309.

Newbold, T., 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, *34*, pp.3-22.

Niewoehner, W.A., 2005. A geometric morphometric analysis of Late Pleistocene human metacarpal 1 base shape. In *Modern morphometrics in physical anthropology* (pp. 285-298). Springer, Boston, MA.

Nishimura, T., Morimoto, N. and Ito, T., 2019. Shape variation in the facial part of the cranium in macaques and African papionins using geometric morphometrics. *Primates*, *60*(5), pp.401-419.

Nowak, M.M., Dziób, K., Ludwisiak, Ł. and Chmiel, J., 2020. Mobile GIS applications for environmental field surveys: A state of the art. *Global Ecology and Conservation*, p.e01089.

Nsubuga, A.M., Holzman, J., Chemnick, L.G. and Ryder, O.A., 2010. The cryptic genetic structure of the North American captive gorilla population. *Conservation Genetics*, *11*(1), pp.161-172.

Nsubuga, A.M., Robbins, M.M., Boesch, C. and Vigilant, L., 2008. Patterns of paternity and group fission in wild multimale mountain gorilla groups. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *135*(3), pp.263-274.

Oates, J.F., Tutin, C.E.G., Humle, T., Wilson, M.L., Baillie, J.E.M., Balmforth, Z., Blom, A., Boesch, C., Cox, D., Davenport, T. and Dunn, A., 2008. IUCN red list of threatened species: Version 2011.2. *International Union for Conservation of Nature and Natural Resources.*

O'Brien, R.G. and Kaiser, M.K., 1985. MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychological Bulletin*, *97*(2), pp.316.

Oliveira, L.R.D., Hingst-Zaher, E. and Morgante, J.S., 2005. Size and shape sexual dimorphism in the skull of the South American fur seal, Arctocephalus australis (Zimmermann, 1783)(Carnivora: Otariidae). *Latin American Journal of Aquatic Mammals*, *4*(1), pp.27-40.

Outomuro, D. and Johansson, F., 2017. A potential pitfall in studies of biological shape: Does size matter? *Journal of Animal Ecology*, *86*(6), pp.1447-1457.

Ovenden, J., Street, R., Peel, D., Peel, S., Courtney, T., Podlich, H., Basford, K. and Dichmont, C., 2004. *A new data source for fisheries resource assessment: genetic estimates of the effective number of spawners. Final Report to the Fisheries Research and Development Corporation*. Department of Primary Industries & Fisheries, Queensland.

Paetkau, D., Slade, R., Burden, M. and Estoup, A., 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular ecology*, *13*(1), pp.55-65.

Paľčo, B., Sharma, A.K., Petrželková, K.J., Vlčková, K., Todd, A., Yeoman, C.J., Wilson, B.A., Stumpf, R., White, B.A., Nelson, K.E. and Leigh, S., 2019. Gut microbiome composition of wild western lowland gorillas is associated with individual age and sex factors. *American journal of physical anthropology, 169*(3), pp.575-585.

Paine, R.T., 1974. Intertidal community structure. *Oecologia*, *15*(2), pp.93-120.

Palagi, E., Antonacci, D. and Cordoni, G., 2007. Fine-tuning of social play in juvenile lowland gorillas (gorilla gorilla gorilla). *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 49*(4), pp.433-445.

Paliy, O. and Shankar, V., 2016. Application of multivariate statistical techniques in microbial ecology. *Molecular ecology*, *25*(5), pp.1032-1057.

Palmer, M.W., 2004. *Ordination Methods - An Overview*. Botany Department, Oklahoma State University, Stillwater, Oklahoma.

Pardini, R., Nichols, E. and Püttker, T., 2017. Biodiversity response to habitat loss and fragmentation. *Encyclopedia of the Anthropocene, 1*, pp.1-12.

Pavlinov, I.Y., 2001. Geometric morphometrics, a new analytical approach to comparison of digitized images. *Zoological Journal of Moscow*, *79*, pp.1-27

Peakall, R., Smouse, P.E. and Huff, D.R., 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. *Molecular Ecology*, *4*(2), pp.135-148.

Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics*, *28*(19), p.2537.

Pearman, T.L. and Jabbour, R., 2015. Testing Functional Hypotheses About Variation in African Ape Scapulae Using 3D Geometric Morphometrics.

Perez, S.E., Raghanti, M.A., Hof, P.R., Kramer, L., Ikonomovic, M.D., Lacor, P.N., Erwin, J.M., Sherwood, C.C. and Mufson, E.J., 2013. Alzheimer's disease pathology in the neocortex and hippocampus of the western lowland gorilla (*Gorilla gorilla gorilla*). *Journal of Comparative Neurology*, *521*(18), pp.4318-4338.

Perez, S.I., Bernal, V. and Gonzalez, P.N., 2006. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *Journal of Anatomy*, *208*(6), pp.769-784.

Peterman, W., Brocato, E.R., Semlitsch, R.D. and Eggert, L.S., 2016. Reducing bias in population and landscape genetic inferences: the effects of sampling related individuals and multiple life stages. *PeerJ, 4*, p.e1813.

Petre, C.A., Tagg, N., Haurez, B., Beudels-Jamar, R., Huynen, M.C. and Doucet, J.L., 2013. Role of the western lowland gorilla (*Gorilla gorilla gorilla*) in seed dispersal in tropical forests and implications of its decline. *Biotechnologie Agronomie Société et Environnement*, *17*(3), pp.517-526.

Phelps, E.M., 1932. A critique of the principle of the horizontal plane of the skull. *American Journal of Physical Anthropology, 17,* pp.71-98.

Phillips, K.A., Elvey, C.R. and Abercrombie, C.L., 1998. Applying GPS to the study of primate ecology: A useful tool? *American Journal of Primatology*, *46*(2), pp.167-172.

Pigott, D.M., Golding, N., Mylne, A., Huang, Z., Henry, A.J., Weiss, D.J., Brady, O.J., Kraemer, M.U., Smith, D.L., Moyes, C.L. and Bhatt, S., 2014. Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife*, *3*, p.e04395.

Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L. and Estoup, A., 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, *95*(6), pp.536-539.

Piry, S., Luikart, G. and Cornuet, J.M., 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, *90*, pp.502-503.

Pitra, C., Vazpinto, P., O'Keeffe, B.W.J., Willows-Munro, S., Jansen van Vuuren, B. and Robinson, T.J., 2006. DNA-led rediscovery of the giant sable antelope in Angola. *European Journal of Wildlife Research, 52*, pp.145-152.

Plumptre, A.J., Nixon, S., Kujirakwinja, D.K., Vieilledent, G., Critchlow, R., Williamson, E.A., Nishuli, R., Kirkby, A.E. and Hall, J.S., 2016. Catastrophic decline of world's largest primate: 80% loss of Grauer's gorilla (Gorilla beringei graueri) population justifies Critically Endangered status. *PloS one, 11*(10), p.e0162697.

Ponder, W.F., Carter, G.A., Flemons, P. and Chapman, R.R., 2001. Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology*, *15*(3), pp.648-657.

Popadin, K., Gunbin, K., Peshkin, L., Annis, S., Kraytsberg, Y., Markuzon, N., Ackermann, R.R. and Khrapko, K., 2017. Mitochondrial pseudogenes suggest repeated inter-species hybridization in hominid evolution. *BioRxiv*, p.134502.

Pope, C.E., Dresser, B.L., Chin, N.W., Liu, J.H., Loskutoff, N.M., Behnke, E.J., Brown, C., McRae, M.A., Sinoway, C.E., Campbell, M.K. and Cameron, K.N., 1997. Birth of a western lowland gorilla (*Gorilla gorilla gorilla*) following in vitro fertilization and embryo transfer. *American Journal of Primatology*, *41*(3), pp.247-260.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á. and Lareu, M., 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, *4*, pp.98.

Porter L.M. and Garber, P.A., 2014. Navigating in small-scale space: The role of landmarks and resource monitoring in understanding saddleback tamarin travel. *American Journal of Primatology*, *76*(5), pp.447-459.

Power, M.E., Tilman, D., Estes, J.A., Menge, B.A., Bond, W.J., Mills, L.S., Daily, G., Castilla, J.C., Lubchenco, J. and Paine, R.T., 1996. Challenges in the quest for keystones: identifying keystone species is difficult—but essential to understanding how loss of species will affect ecosystems. *BioScience*, *46*(8), pp.609-620.

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G. and Cagan, A., 2013. Great ape genetic diversity and population history. *Nature*, *499*(7459), pp.471-475.

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

Puechmaille, S.J., 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), pp.608-627.

Püschel, T.A., Marcé-Nogué, J., Gladman, J.T., Bobe, R. and Sellers, W.I., 2018. Inferring locomotor behaviours in Miocene New World monkeys using finite element analysis, geometric morphometrics and machine-learning classification techniques applied to talar morphology. *Journal of the Royal Society Interface*, *15*(146), pp.20180520.

Pyke, G.H. and Ehrlich, P.R., 2010. Biological collections and ecological/environmental research: a review, some observations and a look to the future. *Biological Reviews*, *85*(2), pp.247-266.

Qeadan, F., 2015. On MANOVA using STATA, SAS & R. *A short course in biostatistics for the Mountain West Clinical Translational Research Infrastructure Network.* University of New Mexico Health Sciences Center.

Quéméré, E., Louis, E.E., Ribéron, A., Chikhi, L. and Crouau-Roy, B., 2010. Non-invasive conservation genetics of the critically endangered golden-crowned sifaka (*Propithecus tattersalli*): high diversity and significant genetic differentiation over a small range. *Conservation Genetics*, *11*(3), pp.675-687.

Ramasamy, R.K., Ramasamy, S., Bindroo, B.B. and Naik, V.G., 2014. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, *3*(1), p.431.

Rambaut, A., 2012. *FigTree v1. 4. Molecular evolution, phylogenetics and epidemiology*. University of Edinburgh, Institute of Evolutionary Biology, Edinburgh, UK.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. and Suchard, M.A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), pp. 901-904.

Ramos-Onsins, S.E. and Rozas, J., 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, *19*(12), pp.2092-2100.

Rannala, B. and Mountain, J.L., 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences*, *94*(17), pp.9197-9201.

Ray, N., Currat, M. and Excoffier, L., 2003. Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, *20*(1), pp.76-86.

Razgour, O., Forester, B., Taggart, J.B., Bekaert, M., Juste, J., Ibáñez, C., Puechmaille, S.J., Novella-Fernandez, R., Alberdi, A. and Manel, S., 2019. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proceedings of the National Academy of Sciences, 116*(21), pp.10418-10423.

Redmond, I., 2008a. *Primates of the World*. New Holland Publishers.

Redmond, I., 2008b. *The primate family tree: the amazing diversity of our closest relatives*. Firefly Books Limited.

Regan, T., 1996. Are zoos morally defensible? In *Ethics on the Ark: zoos, animal welfare, and wildlife conservation* (pp.38-51). Smithsonian Institution Press, Washington DC.

Relethford, J.H., 1994. Craniometric variation among modern human populations. *American Journal of Physical Anthropology*, *95*(1), pp.53-62.

Remis, M.J., 1997. Western lowland gorillas (*Gorilla gorilla gorilla*) as seasonal frugivores: use of variable resources. *American Journal of Primatology*, *43*(2), pp.87-109.

Reyment, R.A., 1991. *Multidimensional palaeobiology* (Vol. 7). Pergamon Press, Oxford.

Richard, G.F., Kerrest, A. and Dujon, B., 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews*, *72*(4), pp.686-727.

Richtsmeier, J.T. and Cheverud, J., 1986. Finite element scaling analysis of normal growth of the human craniofacial complex, *Journal of Craniofacial Genetics and Developmental Biology, 6*(3), pp.289.

Richtsmeier, J.T., Burke Deleon, V. and Lele, S.R., 2002. The promise of geometric morphometrics. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *119*(35), pp.63-91.

Robbins, M.M., Bermejo, M., Cipolletta, C., Magliocca, F., Parnell, R.J. and Stokes, E., 2004. Social structure and life-history patterns in western gorillas (*Gorilla gorilla gorilla*). *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *64*(2), pp.145-159.

Rocatti, G., Aristide, L., Rosenberger, A.L. and Perez, S.I., 2017. Early evolutionary diversification of mandible morphology in the New World monkeys (Primate, Platyrrhini). *Journal of Human Evolution*, *113*, pp.24-37.

Rodríguez-Ramilo, S.T. and Wang, J., 2012. The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. *Molecular Ecology Resources, 12*(5), pp.873-884.

Rogers, A.R. and Harpending, H., 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, *9*(3), pp.552-569.

Rogers, A.R., Fraley, A.E., Bamshad, M.J., Watkins, W.S. and Jorde, L.B., 1996. Mitochondrial mismatch analysis is insensitive to the mutational process. *Molecular Biology and Evolution*, *13*(7), pp.895-902.

Rogers, M.E., Abernethy, K., Bermejo, M., Cipolletta, C., Doran, D., McFarland, K., Nishihara, T., Remis, M. and Tutin, C.E., 2004. Western gorilla diet: a synthesis from six sites. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *64*(2), pp.173-192.

Rogers, M.E., Voysey, B.C., McDonald, K.E., Parnell, R.J. and Tutin, C.E.G., 1998. Lowland gorillas and seed dispersal: the importance of nest sites. *American Journal of Primatology*, *45*(1), pp.45-68.

Rohland, N., Siedel, H. and Hofreiter, M., 2004. Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *Biotechniques*, *36*(5), pp.814-821.

Rohlf, F.J. and Marcus, L.F., 1993. A revolution morphometrics. *Trends in Ecology & Evolution*, *8*(4), pp.129-132.

Rohlf, F.J., 2015. The tps series of software. *Hystrix*, *26*(1), pp.9-12.

Rommel, S.A., Pabst, D.A. and McLellan, W.A., 2009. Skull anatomy. In *Encyclopedia of marine mammals* (pp. 1033-1047). Academic Press.

Ronquist, F. and Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), pp.1572-1574.

Roques, S., Chancerel, E., Boury, C., Pierre, M. and Acolas, M.L., 2019. From microsatellites to single nucleotide polymorphisms for the genetic monitoring of a critically endangered sturgeon. *Ecology and evolution, 9*(12), pp.7017-7029.

Rosenberg, N.A., Burke, T., Elo, K., Feldman, M.W., Freidlin, P.J., Groenen, M.A., Hillel, J., Mäki-Tanila, A., Tixier-Boichard, M., Vignal, A. and Wimmers, K., 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics*, *159*(2), pp.699-713.

Rowe, N., 1996. *Pictorial guide to the living primates* (pp.263). Pogonias Press, East Hampton, NY.

Rozas, J., 2009. DNA sequence polymorphism analysis using DnaSP. In *Bioinformatics for DNA sequence analysis* (pp. 337-350). Humana Press.

Rubinsztein, D.C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S.H., Margolis, R.L., Ross, C.A. and Ferguson-Smith, M.A., 1995. Microsatellite evolution—evidence for directionality and variation in rate between species. *Nature Genetics*, *10*(3), pp.337.

Safina, C., 2018. Where Are Zoos Going—or Are They Gone?. *Journal of Applied Animal Welfare Science, 21*(sup1), pp.4-11.

Saltonstall, K., Amato, G. and Powell, J., 1998. Mitochondrial DNA variability in Grauer's gorillas of Kahuzi-Biega National Park. *Journal of Heredity*, *89*(2), pp.129-135.

Šarhanová, P., Pfanzelt, S., Brandt, R., Himmelbach, A. and Blattner, F.R., 2018. SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. *Ecology and Evolution, 8*(22), pp.10817-10833.

Sarmiento, E.E. and Oates, J.F., 2000. The Cross River gorillas: a distinct subspecies, *Gorilla gorilla diehli* Matschie 1904. *American Museum novitates*; no. 3304.

Sarmiento, E.E., Butynski, T.M. and Kalina, J., 1996. Gorillas of Bwindi-Impenetrable Forest and the Virunga Volcanoes: Taxonomic implications of morphological and ecological differences. *American Journal of Primatology*, *40*(1), pp.1-21.

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T. and McCarthy, S., 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature*, *483*(7388), pp.169-175.

Schaller, G.E., 1963. *The mountain gorilla: Ecology and behavior*. University of Chicago Press.

Scholten, H.J. and de Lepper, M.J., 1991. The benefits of the application of geographical information systems in public and environmental health. *World Health Stat Q*, *44*(3), pp.160-170.

Schroeder, K.B., Schurr, T.G., Long, J.C., Rosenberg, N.A., Crawford, M.H., Tarskaia, L.A., Osipova, L.P., Zhadanov, S.I. and Smith, D.G., 2007. A private allele ubiquitous in the Americas. *Biology Letters*, *3*(2), pp.218-223.

Schwarz, E., 1928. Die Sammlung afrikanischer Affen in Congo-Museum. *Revue de Zoologie et Botanique Africaine, 16*(2), pp.1-48.

Seaman, M.I., Deinard, A.S. and Kidd, K.K., 1999, January. Incongruence between mitochondrial and nuclear DNA estimates of divergence between Gorilla subspecies. In *American Journal of Physical Anthropology* (pp. 247-247). John Wiley & Sons Inc, New York, USA.

Séré, M., Kaboré, J., Jamonneau, V., Belem, A.M.G., Ayala, F.J. and De Meeûs, T., 2014. Null allele, allelic dropouts or rare sex detection in clonal organisms: simulations and application to real data sets of pathogenic microbes. *Parasites & Vectors*, *7*(1), pp.331.

Shaffer, H.B., Fisher, R.N. and Davidson, C., 1998. The role of natural history collections in documenting species declines. *Trends in Ecology & Evolution*, *13*(1), pp.27-30.

Shane, S., 2005. *Shane's simple guide to F-statistics*. University of Auckland, Auckland, New Zealand.

Shanee, S., Tello-Alvarado, J.C., Vermeer, J. and Bóveda-Penalba, A.J., 2013. GIS risk assessment and GAP analysis for the Andean titi monkey (*Callicebus oenanthe*). *Primate Conservation*, *26*(1), pp.17-23.

Shutt, K., Setchell, J.M. and Heistermann, M., 2012. Non-invasive monitoring of physiological stress in the Western lowland gorilla (Gorilla gorilla gorilla): validation of a fecal glucocorticoid assay and methods for practical application in the field. *General and Comparative Endocrinology, 179*(2), pp.167-177.

Siegel, S., George, O., Ellisor, A., Summerville, K.S. and Renner, M.J., 2020. A Population Survey of Golden Monkeys and L'Hoest's Monkeys in Gishwati Forest, Rwanda. African Primates, 14, pp.55-60.Pontzer, H., 2020. Ranging Ecology: The View from Above. *Current Biology, 30*(22), pp.R1378-R1380.

Siljander, M., Kuronen, T., Johansson, T., Munyao, M.N. and Pellikka, P.K., 2020. Primates on the farm–spatial patterns of human–wildlife conflict in forest-agricultural landscape mosaic in Taita Hills, Kenya. *Applied Geography*, *117*, pp.102185.

Simons, E.A., Frost, S.R. and Singleton, M., 2018. Ontogeny and phylogeny of the cercopithecine cranium: A geometric morphometric approach to comparing shape change trajectories. *Journal of Human Evolution*, *124*, pp.40-51.

Simons, N.D., Wagner, R.S. and Lorenz, J.G., 2013. Genetic diversity of North American captive-born gorillas (*Gorilla gorilla gorilla*). *Ecology and Evolution*, *3*(1), pp.80-88.

Simpson, G.G., 1961. *Principles of Animal Taxonomy* (pp.247). Columbia University Press, New York, USA.

Slatkin, M., 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*(6), p.477.

Slice, D.E., 2007. Geometric morphometrics. *Annual Review of Anthropology, 36.*

Smith, T.B. and Wayne, R.K. eds., 1996. *Molecular genetic approaches in conservation*. Oxford University Press.

Smith, Z.J., 2014. *Mapping the Spatial Movements, Behaviors, and Interactions of Captive Orangutans using Terrestrial Laser Scanning and GIS*. University of South Florida, USA.

Sneath, P.H.A., 1967. Trend-surface analysis of transformation grids. *Journal of Zoology, 151*, pp.65-122.

Sneath, P.H.A., and Sokal, R.R., 1973. *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. W.H. Freeman and Company, San Francisco, California, USA.

Soberón, J. and Peterson, T., 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1444), pp.689-698.

Sodhi, N.S., Butler, R., Laurance, W.F. and Gibson, L., 2011. Conservation successes at micro-, meso- and macroscales. *Trends in Ecology & Evolution*, *26*(11), pp.585-594

Sokal, R.R., 1973. The species problem reconsidered. *Systematic Zoology, 22*, pp.360-374.

Song, H., Buhay, J.E., Whiting, M.F. and Crandall, K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, *105*(36), pp.13486-13491.

Soto-Calderón, I.D., Clark, N.J., Wildschutte, J.V.H., DiMattio, K., Jensen-Seaman, M.I. and Anthony, N.M., 2014. Identification of species-specific nuclear insertions of mitochondrial DNA (numts) in gorillas and their potential as population genetic markers. *Molecular Phylogenetics and Evolution*, *81*, pp.61-70.

Soto-Calderón, I.D., Dew, J.L., Bergl, R.A., Jensen-Seaman, M.I. and Anthony, N.M., 2015. Admixture between historically isolated mitochondrial lineages in captive western gorillas: recommendations for future management. *Journal of Heredity*, *106*(3), pp.310-314.

Soulsbury, C.D., Iossa, G., Edwards, K.J., Baker, P.J. and Harris, S., 2007. Allelic dropout from a high-quality DNA source. *Conservation Genetics*, *8*(3), pp.733-738.

Spear, D.M., Pauly, G.B. and Kaiser, K., 2017. Citizen science as a tool for augmenting museum collection data from urban areas. *Frontiers in Ecology and Evolution, 5*, p.86.

Sproul, J.S. and Maddison, D.R., 2017. Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources, 17*(6), pp.1183-1201.

Stegmann, M.B. and Gomez, D.D., 2002. A brief introduction to statistical shape analysis. *Informatics and mathematical modelling, Technical University of Denmark, DTU*, *15*(11).

Stoinski, T.S., Vecellio, V., Ngaboyamahina, T., Ndagijimana, F., Rosenbaum, S. and Fawcett, K.A., 2009. Proximate factors influencing dispersal decisions in male mountain gorillas, Gorilla beringei beringei. *Animal Behaviour*, *77*(5), pp.1155-1164.

Strindberg, S., Maisels, F., Williamson, E.A., Blake, S., Stokes, E.J., Aba'a, R., Abitsi, G., Agbor, A., Ambahe, R.D., Bakabana, P.C. and Bechem, M., 2018. Guns, germs, and trees determine density and distribution of gorillas and chimpanzees in Western Equatorial Africa. *Science Advances*, *4*(4), p.eaar2964.

Stumpf, R.M., Fleagle, J.G. and Groves, C.P., 1997. Sexual dimorphism and geographic variation among subspecies of Gorilla. *American Journal of Physical Anthropology, Supplement, 24*, pp.223-224.

Stumpf, R.M., Polk, J.D., Oates, J.F., Jungers, W.L., Heesy, C.P., Groves, C.P. and Fleagle, J.G., 2002. *Patterns of diversity in gorilla cranial morphology. Gorilla Biology: A Multidisciplinary Perspective* (pp.35-61). Cambridge University Press, Cambridge.

Sukmak, M., Malaivijitnond, S., Schülke, O., Ostner, J., Hamada, Y. and Wajjwalku, W., 2014. Preliminary study of the genetic diversity of eastern Assamese macaques (*Macaca assamensis assamensis*) in Thailand based on mitochondrial DNA and microsatellite markers. *Primates*, *55*(2), pp.189-197.

Sundberg, P. 1989. Shape and size-constrained principal components analysis. *Systematic Zoology, 38*, pp.166.

Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O. and Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology*, *50*(4), pp.525-539.

Szpiech, Z.A. and Rosenberg, N.A., 2011. On the size distribution of private microsatellite alleles. *Theoretical Population Biology*, *80*(2), pp.100-113.

Szpiech, Z.A., Jakobsson, M. and Rosenberg, N.A., 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, *24*(21), pp.2498-2504.

Taberlet, P. and Bouvet, J., 1994. Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear *Ursus arctos* in Europe. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *255*(1344), pp.195-200.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P. and Bouvet, J., 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*(16), pp.3189-3194.

Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, *17*(16), pp.6463-6471.

Tay, W.T., Elfekih, S., Court, L.N., Gordon, K.H., Delatte, H. and De Barro, P.J., 2017. The trouble with MEAM2: Implications of pseudogenes on species delimitation in the globally invasive Bemisia tabaci (Hemiptera: Aleyrodidae) cryptic species complex. *Genome biology and evolution, 9*(10), pp.2732-2738.*reports*, 10(1), pp.1-11.

Taylor, A.B. and Goldsmith, M.L. eds., 2002. *Gorilla biology: a multidisciplinary perspective* (Vol. 34). Cambridge University Press, Cambridge.

Taylor, A.B., 2006. Size and shape dimorphism in great ape mandibles and implications for fossil species recognition. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *129*(1), pp.82-98.

Terhune, C.E., Kimbel, W.H. and Lockwood, C.A., 2007. Variation and diversity in Homo erectus: a 3D geometric morphometric analysis of the temporal bone. *Journal of Human Evolution*, *53*(1), pp.41-60.

Thalmann, O., Hebler, J., Poinar, H.N., Pääbo, S. and Vigilant, L., 2004. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Molecular Ecology*, *13*(2), pp.321-335.

Thalmann, O., Serre, D., Hofreiter, M., Lukas, D., Eriksson, J. and Vigilant, L., 2005. Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. *Molecular Ecology*, *14*(1), pp.179-188.

Thalmann, O., Wegmann, D., Spitzner, M., Arandjelovic, M., Guschanski, K., Leuenberger, C., Bergl, R.A. and Vigilant, L., 2011. Historical sampling reveals dramatic demographic changes in western gorilla populations. *BMC Evolutionary Biology*, *11*(1), p.85.

Tobias, P.V. and Cooper, J.E., 2003. The mountain gorilla: a little known chapter of pioneering studies. *Transactions of the Royal Society of South Africa, 58*(1), pp.75-77.

Todd, S. and Maurer, G., 2020. Bushfire recovery where it matters most. Impacts and actions in Key Biodiversity Areas affected by the 2019/20 Bushfire Crisis. *BirdLife Australia, Melbourne*.

Toder, R., Grützner, F., Haaf, T. and Bausch, E., 2001. Species-specific evolution of repeated DNA sequences in great apes. *Chromosome Research*, *9*(6), pp.431-435.

Tofilski, A., 2008. Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie*, *39*(5), pp.558-563.

Towle, I., Irish, J.D. and De Groote, I., 2018. Amelogenesis imperfecta in the dentition of a wild chimpanzee. *Journal of medical primatology, 47*(2), pp.117-119.

Triant, D.A. and DeWoody, J.A., 2007. The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. *Journal of Mammalogy, 88*(4), pp.908-920.

Tribe, A. and Booth, R., 2003. Assessing the role of zoos in wildlife conservation. *Human Dimensions of Wildlife*, *8*(1), pp.65-74.

Troudet, J., Vignes-Lebbe, R., Grandcolas, P. and Legendre, F., 2018. The increasing disconnection of primary biodiversity data from specimens: how does it happen and how to handle it?. Systematic *Biology, 67*(6), pp.1110-1119.

Tsuji, J., Frith, M.C., Tomii, K. and Horton, P., 2012. Mammalian NUMT insertion is non-random. *Nucleic acids research, 40*(18), pp.9073-9088.

Tutin, C., Stokes, E., Boesch, C., Morgan, D., Sanz, C., Reed, T., Blom, A., Walsh, P., Blake, S. and Kormos, R., 2005. Regional action plan for chimpanzees and gorillas in west equatorial Africa.

Tutin, C.E. and Vedder, A., 2001. Gorilla conservation and research in Central Africa. *African rainforest ecology and conservation: An Interdisciplinary Perspective*, pp.429-448.

Tutin, C.E., Parnell, R.J., White, L.J. and Fernandez, M., 1995. Nest building by lowland gorillas in the Lopé Reserve, Gabon: environmental influences and implications for censusing. *International Journal of Primatology*, *16*(1), p.53.

Tuttle, R.H., 2003. An Introductory Perspective: Gorillas—How Important, How Many, How Long? *Gorilla Biology: A Multidisciplinary Perspective*, pp.11-14.

Uchida, A., 1998. Variation in tooth morphology of Gorilla gorilla. *Journal of Human Evolution*, *34*(1), pp.55-70.

United Nations, "UN Report: Nature's Dangerous Decline "Unprecedented"; Species Extinction Rates "Accelerating", United Nations Sustainable Development, 2019 accessed 15th June 2019 https://www.un.org/sustainabledevelopment/blog/2019/05/nature-decline-unprecedented-report

Valdes, A.M., Slatkin, M. and Freimer, N.B., 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, *133*(3), pp.737-749.

van Casteren, A., Wright, E., Kupczik, K. and Robbins, M.M., 2019. Unexpected hard-object feeding in Western lowland gorillas. *American journal of physical anthropology*, *170*(3), pp.433-438.

van der Valk, T., Lona Durazo, F., Dalén, L. and Guschanski, K., 2017. Whole mitochondrial genome capture from faecal samples and museum-preserved specimens. *Molecular Ecology Resources*, *17*(6), pp.e111-e121

van der Valk, T., Sandoval-Castellanos, E., Caillaud, D., Ngobobo, U., Binyinyi, E., Stoinski, T., Gilissen, E., Sonet, G., Semal, P., Kalthoff, D.C. and Dalén, L., 2018. Significant loss of mitochondrial

diversity within the last century due to extinction of peripheral populations in eastern gorillas. *Scientific Reports*, *8*(1), pp.1-10.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P. and Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), pp.535-538.

Van Schaik, C.P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C.D., Singleton, I., Suzuki, A., Utami, S.S. and Merrill, M., 2003. Orangutan cultures and the evolution of material culture. *Science*, *299*(5603), pp.102-105.

Vega Bernal, R.R., 2010. *Phylogeographic and morphometric studies on the Eurasian pygmy shrew* Sorex minutus*: insights into its evolutionary history and postglacial colonisation in Europe* (Doctoral dissertation, University of York).

Vega, R., Mcdevitt, A.D., Kryštufek, B. and Searle, J.B., 2016. Ecogeographical patterns of morphological variation in pygmy shrews *Sorex minutus* (Soricomorpha: Soricinae) within a phylogeographical and continental-and-island framework. *Biological Journal of the Linnean Society*, *119*(4), pp.799-815.

Vigilant, L. and Bradley, B.J., 2004. Genetic variation in gorillas. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *64*(2), pp.161-172.

Viscosi, V. and Cardini, A., 2011. Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners. *PloS One*, *6*(10), pp.e25630.

Voytas, D., 2000. Agarose gel electrophoresis. *Current Protocols in Molecular Biology*, *51*(1), pp.2-5.

Vrijenhoek, R. C. 1994. Genetic diversity and fitness in small populations. Pages 37–53 in Tomiuk, J. and Jain, S.K., editors. Conservation genetics. Birkhauser Verlag, Basel, Switzerland. *Philosophical Transactions of the Royal Society of London Series B, 359*, pp.689–698.

Wandeler, P., Hoeck, P.E. and Keller, L.F., 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, *22*(12), pp.634-642.

Wang, J., Caballero, A. and Hill, W.G., 1998. The effect of linkage disequilibrium and deviation from Hardy–Weinberg proportions on the changes in genetic variance with bottlenecking. *Heredity, 81*(2), pp.174-186.

Wang, C., Schroeder, K.B. and Rosenberg, N.A., 2012. A maximum-likelihood method to correct for allelic dropout in microsatellite data with no replicate genotypes. *Genetics*, *192*(2), pp.651-669.

Waples, R.S., 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, *121*(2), pp.379-391.

Waples, R.S., 2014. Testing for Hardy–Weinberg proportions: have we lost the plot? *Journal of Heredity*, *106*(1), pp.1-19.

Waters, S.S. and Ulloa, O., 2007. Preliminary survey on the current distribution of primates in Belize. *Neotropical Primates*, *14*(2), pp.80-82.

Weber, J.L. and May, P.E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*, *44*(3), p.388.

Webster, M.A.R.K. and Sheets, H.D., 2010. A practical introduction to landmark-based geometric morphometrics. *The Paleontological Society Papers*, *16*, pp.163-188.

Weckworth, B.V., Talbot, S., Sage, G.K., Person, D.K. and Cook, J., 2005. A signal for independent coastal and continental histories among North American wolves. *Molecular Ecology*, *14*(4), pp.917-931.

Wegmann, M., Leutner, B. and Dech, S. eds., 2016. *Remote sensing and GIS for ecologists: using open source software*. Pelagic Publishing Ltd.

Weir, B.S. and Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), pp.1358-1370.

Weiss, G. and von Haeseler, A., 2003. Testing substitution models within a phylogenetic tree. *Molecular Biology and Evolution*, *20*(4), pp.572-578.

Whiten, A., Goodall, J., McGrew, W.C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C.E., Wrangham, R.W. and Boesch, C., 1999. Cultures in chimpanzees. *Nature*, *399*(6737), pp.682-685.

Whiten, A., Spiteri, A., Horner, V., Bonnie, K.E., Lambeth, S.P., Schapiro, S.J. and De Waal, F.B., 2007. Transmission of multiple traditions within and between chimpanzee groups. *Current Biology*, *17*(12), pp.1038-1043.

Wiley, E.O., 1978. The evolutionary species concept reconsidered. *Systematic Zoology*, *27*(1), pp.17-26.

Wilms, T. and Bender, U., 2010. International Studbook for the western lowland gorilla *Gorilla g. gorilla* (Savage & Wyman, 1847).

Wimmer, B., Tautz, D. and Kappeler, P.M., 2002. The genetic population structure of the gray mouse lemur (*Microcebus murinus*), a basal primate from Madagascar. *Behavioral Ecology and Sociobiology*, *52*(2), pp.166-175.

Wood, B., 2019. Personal Communication.

Wood, B.A., 1976. The nature and basis of sexual dimorphism in the primate skeleton. *Journal of Zoology*, *180*, pp.15-34.

Wood, B.A., 1979. Relationship between body size and long bone lengths in Pan and Gorilla. *American Journal of Physical Anthropology*, *50*(1), pp.23-25.

Wright, S., 1921. Systems of mating. *Genetics, 6*, pp.111-178.

Wright, S., 1931. Evolution in Mendelian populations. *Genetics*, *16*(2), p.97.Wright, S., 1978. *Evolution and the genetics of populations: a treatise in four volumes: Vol. 4: variability within and among natural populations*. University of Chicago Press.

Wright, E., Galbany, J., McFarlin, S.C., Ndayishimiye, E., Stoinski, T.S. and Robbins, M.M., 2020. Dominance rank but not body size influences female reproductive success in mountain gorillas. *PloS one*, *15*(6), p.e0233235.

Xu, X. and Arnason, U., 1996. A complete sequence of the mitochondrial genome of the Western lowland gorilla. *Molecular biology and evolution, 13*(5), pp.691-698.

Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N. and De Manuel, M., 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science, 348*(6231), pp.242-245.

Yamagiwa, J., Kahekwa, J. and Basabose, A.K., 2003. Intra-specific variation in social organization of gorillas: implications for their social evolution. *Primates*, *44*(4), pp.359-369.

Yamaguchi, N., Kitchener, A.C., Driscoll, C.A. and Macdonald, D.W., 2009. Divided infraorbital foramen in the lion (*Panthera leo*): its implications for colonisation history, population bottlenecks, and conservation of the Asian lion (*P. l. persica*). *Contributions to Zoology, 78*, pp.77-83.

Yaxley, K.J. and Foley, R.A., 2019. Reconstructing the ancestral phenotypes of great apes and humans (Homininae) using subspecies-level phylogenies. *Biological Journal of the Linnean Society*, *128*(4), pp.1021-1038.

Yeates, D.K., Zwick, A. and Mikheyev, A.S., 2016. Museums are biobanks: unlocking the genetic potential of the three billion specimens in the world's biological collections. *Current Opinion in Insect Science*, *18*, pp.83-88.

Yu, N., Jensen-Seaman, M.I., Chemnick, L., Ryder, O. and Li, W.H., 2004. Nucleotide diversity in gorillas. *Genetics, 166*(3), pp.1375-1383.

Zahidin, M.A., Abd Jalil, N., Naharuddin, N.M., Abd Rahman, M.R., Gani, M. and Abdullah, M.T., 2019. Partial mtDNA sequencing data of vulnerable Cephalopachus bancanus from the Malaysian Borneo. *Data in brief, 25,* p.104133.

Zeisset, I. and Beebee, T.J.C., 2010. Larval fitness, microsatellite diversity and MHC class II diversity in common frog (*Rana temporaria*) populations. *Heredity*, *104*(5), pp.423-430.

Zelditch, M.L., Swiderski, D.L., Sheets, H.D. & Fink, W.L. (2004) *Geometric Morphometrics for Biologists: A Primer*. Elsevier, New York.

Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Cao, Z., Wang, J., Yuan, W., Zhu, Y. and Song, C., 2020. COVID-19: challenges to GIS with big data. *Geography and sustainability, 1*(1), pp.77-87.

Zink, R.M. and Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, *17*(9), pp.2107-2121.

Zinner, D., Atickem, A., Beehner, J.C., Bekele, A., Bergman, T.J., Burke, R., Dolotovskaya, S., Fashing, P.J., Gippoliti, S., Knauf, S. and Knauf, Y., 2018. Phylogeography, mitochondrial DNA diversity, and demographic history of geladas (Theropithecus gelada). *PloS one, 13*(8), p.e0202303.

Zipfel, B., DeSilva, J.M. and Kidd, R.S., 2009. Earliest complete hominin fifth metatarsal—Implications for the evolution of the lateral column of the foot. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *140*(3), pp.532-545.

Zoo Licensing Act 1981 http://www.legislation.gov.uk/ukpga/1981/37

Appendix 1.

List of gorillas from the Aspinall Foundation samples that were used for in this study and their relationships, plus the mitochondrial founder of each.

NAME	FTA card	M/F	Sire	Dam	mtFounder
Mataki	Y	M	Bitam	Killa Killa	H61001-Mouila (Cameroon)
Kifta	Y	F	Kifu	Tambabi	H61000-Baby Doll (Cameroon)
Imbi	Y	F	Kouillou	Mambi	H61000-Baby Doll (Cameroon)
Boumi	Y	M	Bitam	Mouila	H61001-Mouila (Cameroon)
Sidonie	Y	F	wild	Wild	Sidonie (Cameroons)
Timbou	Y	M	Bitam	Mouila	H61001-Mouila (Cameroon)
Masindi	Y	F	Djanghou	Kimba	H69000-Mushie (Cameroon)
Yene	Y	F	Djala	Foufou	H72001-Founa (Gabon)
Shasha	Y	F	Kijo	Shumba	H69000-Mushie (Cameroon)
Ujiji	Y	M	Bitam	JuJu	H62000-JuJu (Cameroon)
Bitono	Y	M	Kijo	Mushie	H69000-Mushie (Cameroon)
Jah	Y	M	Kijo	Dihi	285-Mintha (Cameroon)
Bitanu	Y	M	Bitam	JuJu	H62000-JuJu (Cameroon)
Popa	Y	M	Kijo	Mushie	H69000-Mushie (Cameroon)
Baloo	Y	M	Kouillou	Tamba	H60000-Shamba (Cameroon)
Imbizo	Y	M	Kouillou	Emba	M2/JERSEY (Cameroon)
Tamki	Y	F	Bitam	Killa Killa	H61001-Mouila (Cameroon)
Fou Fou	Y	F	Kijo	Founa	H72001-Founa (Gabon)
Mbwambe	Y	F	Djala	Kishi	H69000-Mushie (Cameroon)
Djongo	Y _ DJANGA	M	Djala	Kibi	H69000-Mushie (Cameroon)
Akou	Y	F	Djala	Kishi	H69000-Mushie (Cameroon)
Louna	Y	M	Djala	FouFou	H72001-Founa (Gabon)
Djala	Y	M	wild	Wild	Djala mother (Congo)
Mumba	Y	F	Kijo	Shumba	H69000-Mushie (Cameroon)
Kishi	Y	F	Kijo	Mushie	H69000-Mushie (Cameroon)
Kibi	Y	F	Kijo	Shumba	H69000-Mushie (Cameroon)
Emmie	Y	F	Kibobo	Aline/Sabrina	Martha (Palmyre)
Mah Mah	Y	F	Kouillou	Mambi	H61000-Baby Doll (Cameroon)
Boma	Y	F	Tam Tam	Hyasmina	S91066/LA PLAINE-Hyasmina (Cameroon)
Otana	Y	M	Kouillou	Tamba	H60000-Shamba (Cameroon)
Lou Lou	Y	F	Kijo	Shumba	H69000-Mushie (Cameroon)
Oundi	Y	F	Kifu	Sounda	H87005-Sounda (Congo)
Tebe	Y	F	wild	Wild	H82000-Tebe (Gabon)
Emba	Y	F	Bitam	Bamenda	M2/JERSEY (Cameroon)
Boula	Y	F	Kouillou	Mambi	H61000-Baby Doll (Cameroon)
Jubi	Y	F	Bitam	JuJu	H62000-JuJu (Cameroon)
Kabale	Y	M	Kouillou	Tamba	H60000-Shamba (Cameroon)
Mambi	Y	F	Bitam	Baby Doll	H61000-Baby Doll (Cameroon)
Matibi	Y	F	Bitam	Kaja	?H62000-JuJu
Tamba	Y	F	Bitam	Shumba	H60000-Shamba (Cameroon)
Tamidol	Y	F	Bitam	Baby Doll	H61000-Baby Doll (Cameroon)
Kush	Y	M	H75000	Mushie	H69000-Mushie (Cameroon)

Tambabi	Y	F	Bitam	Baby Doll	H61000-Baby Doll (Cameroon)
Viringika	Y	F	Ngola (Durrell)	Inge (Frankfurt)	26FRAN-Dorret (unk)
Kwimba	Y	F	Kouillou	Tamba	H60000-Shamba (Cameroon)
Kwimba infant	Y	infant		Kwimba	H60000-Shamba (Cameroon)
Matadi	Y		Sekondi	Ozala	28/TWY-Biddy (Unk)
Shumba	Y	F	Mumbah	Mushie	H69000-Mushie (Cameroon)
Thirza	Y	F	Bokito	Tamani	285-Mintha (Cameroon)
Kouyou	Y	M	Kifu	Sounda	H87005-Sounda (Congo)
Fubu	Y	M	Kifu	Bamilla	H61001-Mouila (Cameroon)
Kangu	Y	M	Kifu	Sangha	P21320-Sangha (Congo)
Kebu	Y	M	Kifu	Tebe	H82000-Tebe (Gabon)
Mumba	Y	F	Kijo	H80001	
Kisane	Y	M	Djanghou	Sanki	H87005-Sounda (Congo)
Sammi	Y	M	Samson (WILD)	Minnie	Martha (Palmyre)
Kuimba	Y	F	Asato	Tamarilla	H61001-Mouila (Cameroon)
Mayombe	Y	F	Asato	Inge	26FRAN-Dorret (unk)
Mouila		F	wild	Wild	H61001-Mouila (Cameroon)
Baby Doll		F	wild	Wild	H61000-Baby Doll (Cameroon)
Kouillou		M	wild	Wild	Kouillou mother (Congo)
Kwimba infant 2				Kwimba	H60000-Shamba (Cameroon)
Masindi infant		F	Sammi	Masindi	H69000-Mushie (Cameroon)
Viringika infant				Viringika	26FRAN-Dorret (unk)
Louna		M	Djala	FouFou	H72001-Founa (Gabon)
Djanghou		M	Djala	Sangha	P21320-Sangha (Congo)
Sammi	Y	M	Samson (WILD)	Minnie	Martha (Palmyre)
Masindi	Y	F			H69000-Mushie (Cameroon)
Infant of Masindi & Sammi		F	Sammi	Masindi	H69000-Mushie (Cameroon)

Appendix 2.

Revised protocol for DNA extraction from FTA cards.

- Sterilise all equipment prior to use with ethanol and passing through a flame.
- Cut a small piece of FTA card (instead of using the punch) approximately 3-4 mm squared and place in a PCR tube, always sterilise scissors with ethanol and pass through flame between samples.
- Add 100µl of sterile H_2O to each sample.
- Heat in thermocycler at 96° for 15 minutes.
- Using a pipette, remove all the water.
- Add 100µl of sterile H_2O to each sample.
- Heat in thermocycler at 96° for 20 minutes.
- Do not remove card, leave in, use and store at 4° between use.

Appendix 3.

List of all GenBank sequences used in this study, including their group ID assigned for analyses and indication of whether the sequence is a numt sequence.

GenBank Accession Number	Group ID	Numt sequence
L76749	EM	No
L76750	EM	No
L76751	EM	No
L76752	EM	No
L76771	EL	No
L76772	EL	No
L76773	EL	No
AF187549	EL	No
AF050738	EL	No
L76754	CAP	Yes
L76760	CAP	Yes
L76761	CAR	No
L76763	CON	No
L76764	GAB	No
L76766	CAP	Yes
AY079508	CAR	No
AY079509	CAR	No
AY079510	CAR	No
AF250888	GAB	Yes
AY530102	EM	No
AY530103	EM	No
AY530104	EL	No
AY530105	EL	No
AY530106	EL	No
AY530107	EL	No
AY530108	EL	No
AY530109	NIG	No
AY530110	NIG	No
AY530111	NIG	No
AY530112	NIG	No
AY530113	CAM	No
AY530114	CAM	No
AY530115	CAM	No
AY530116	CAM	No
AY530117	CAM	No

AY530118	CAM	No
AY530119	CAM	No
AY530120	GAB	No
AY530121	CAM	No
AY530122	EQG	No
AY530123	EQG	No
AY530124	EQG	No
AY530125	EQG	No
AY530126	EQG	No
AY530127	CAM	No
AY530128	CAR	No
AY530129	CON	No
AY530130	CAR	No
AY530131	CAR	No
AY530132	CAM	No
AY530133	CAR	No
AY530134	GAB	No
AY530135	CON	No
AY530136	GAB	No
AY530137	GAB	No
AY530138	GAB	No
AY530139	GAB	No
AY530140	GAB	No
AY530141	CON	No
AY530142	GAB	No
AY530143	GAB	No
AY530144	GAB	No
AY530145	GAB	Yes
AY530146	GAB	Yes
AY530147	EL	Yes
AY530148	CAM	Yes
AY530149	CAM	Yes
AY530150	GAB	Yes
AY530151	CAM	Yes
AY530152	EL	Yes
AY530153	GAB	Yes
AY530154	CAM	Yes
KM555059	CAP	No
KM555060	CAP	No
KM555061	CAP	No
KM555062	CAP	No
KM555063	CAP	No
KM555064	CAP	No
KM555065	CAP	No
KM555066	CAP	No

KM555067	CAP	No
KM555068	CAP	No
KM555069	CAP	No
KM555070	CAP	No
KM555071	CAP	No
KM555072	CAP	No
KM555073	CAP	No
KM555074	CAP	No
KM555075	CAP	No
KM555076	CAP	No
KM555077	CAP	No
KM555078	CAP	No
KM555079	CAP	No
KM555080	CAP	No
KM555081	CAP	No
KM555082	CAP	No
KM555083	CAP	No
KM555084	CAP	No
KM555085	CAP	No
KM555086	CAP	No
KM555087	CAP	No
KM555088	CAP	No
KM555089	CAP	No
KM555090	CAP	No
KM555091	CAP	No
KM555092	CAP	No
KM555093	CAP	No
KM555094	CAP	No
KM555095	CAP	No
KM555096	CAP	No
KM555097	CAP	No
KM555098	CAP	No
KM555099	CAP	No
L76753	CAP	No
L76755	CAP	No
L76756	CAP	No
L76757	CAP	No
L76758	CAP	No
L76759	CAP	No
L76765	CAP	No
L76767	CAP	No
AF250887	GAB	Yes
AF451971	CAP	No
AF451968	CAP	No
AF240448	EL	Yes

AF240449	EL	Yes
AF240450	EL	Yes
AF240451	EL	Yes
AF240452	EL	Yes
AF240453	EL	Yes
AF240455	EL	Yes
AF240456	EL	Yes
AF240457	EL	Yes
AF240458	EL	Yes
AJ422244	CON	No
AF250891	CAM	Yes
AF451954	CAP	No
AF250890	CAM	Yes
AM392424	CAM	No
KF029427	CAM	No
KF029423	CAM	No
AM392422	CAM	No
AM392417	CAM	No
AM392409	CAM	No
AM392415	CAM	No
AJ586558	Outgroup	No

Appendix 4.

List of the 61 haplotypes from the numt free sequences, showing their haplotype, frequency, sequence ID and haplogroup.

Haplotype	Frequency	Sequences	Haplogroup
Hap 1	3	ASP Otana, ASP Tamba, ASP Kabale	D2
Hap 2	1	ASP Baloo	D2
Hap 3	3	ASP Kwimba, ASP KwimbaInf, ASP KwimbaInf2	D2
Hap 4	2	ASP Kangu, CAP KM555087	D2
Hap 5	3	PCM CamI 324, PCM CamI 325, PCM CamI 325	C2
Hap 6	11	PCM Mer 34, PCM Mer 471, PCM Mer 720, PCM Mer 29, PCM Mer 59, CAM AY530119, CAP KM555096, CAP KM555097, CAP L76758, CAM AM392422, CAM AM392417	C2
Hap 7	3	PCM Mer 58, RCS PA62, CAM KF029427	C2
Hap 8	6	PCM Mer 137, PCM Mer 136, PCM Mer 264, PCM Mer 840, CAM KF029423, CAM AM392415	C2
Hap 9	1	PCM Mer 470	C2
Hap 10	1	PCM Mer 487	C2
Hap 11	3	PCM FC 147, NIG AY530109, CAM AM392424	C1
Hap 12	5	PCM MI 28, CAP KM555070, CAP KM555084, CAP KM555092, CAP AF451971	C1
Hap 13	1	RCS PA63	C2
Hap 14	3	PCM Mer 36, CAP KM555074, CON AJ422244	C2
Hap 15	5	PCM ZVI 32, CAM AY530117, CAP KM555059, CAP KM555094, CAM AM392409	C1
Hap 16	1	PCM FC 114	D2
Hap 17	1	PCM CamI 14	D2
Hap 18	2	EM L76749, EM AY530103	A
Hap 19	2	EM L76750, EM L76751	A
Hap 20	2	EM L76752, EM AY530102	A
Hap 21	1	EL L76771	B
Hap 22	1	EL L76772	B
Hap 23	3	EL L76773, EL AY530104, EL AY530105	B
Hap 24	1	EL AF187549	B
Hap 25	1	EL AF050738	B
Hap 26	1	CAR L76761	D2

Hap 27	23	CON L76763, GAB AY530134, CON AY530135, GAB AY530136, GAB AY530137, GAB AY530138, GAB AY530139, GAB AY530140, CON AY530141, GAB AY530143, CAP KM555064, CAP KM555065, CAP KM555067, CAP KM555069, CAP KM555072, CAP KM555077, CAP KM555079, CAP KM555081, CAP KM555085, CAP KM555086, CAP KM555088, CAP KM555098, CAP L76757	D3
Hap 28	1	GAB L76764	D3
Hap 29	1	CAR AY079508	D2
Hap 30	5	CAR AY079509, CAR AY530128, CON AY530129, CAR AY530130, CAP KM555078	D2
Hap 31	1	CAR AY079510	D2
Hap 32	3	EL AY530106, EL AY530107, EL AY530108	B
Hap 33	1	NIG AY530110	C1
Hap 34	1	NIG AY530111	C1
Hap 35	1	NIG AY530112	C1
Hap 36	4	CAM AY530113, CAM AY530114, CAM AY530115, CAM AY530116	C1
Hap 37	1	CAM AY530118	C1
Hap 38	1	GAB AY530120	C2
Hap 39	1	CAM AY530121	C2
Hap 40	6	EQG AY530122, EQG AY530123, EQG AY530124, EQG AY530125, CAP KM555071, CAP L76767	D1
Hap 41	1	EQG AY530126	D1
Hap 42	1	CAM AY530127	D2
Hap 43	1	CAR AY530131	D2
Hap 44	2	CAM AY530132, CAP KM555066	D2
Hap 45	1	CAR AY530133	D2
Hap 46	1	GAB AY530142	D3
Hap 47	1	GAB AY530144	D3
Hap 48	3	CAP KM555060, CAP KM555061, CAP KM555063	D2
Hap 49	2	CAP KM555062, CAP KM555095	C1
Hap 50	3	CAP KM555068, CAP KM555093, CAP L76765	D2
Hap 51	4	CAP KM555073, CAP KM555076, CAP KM555089, CAP AF451968	C3
Hap 52	1	CAP KM555075	D2
Hap 53	1	CAP KM555080	D3
Hap 54	1	CAP KM555082	D3
Hap 55	1	CAP KM555083	D3
Hap 56	3	CAP KM555090, CAP KM555099, CAP L76755	C1
Hap 57	1	CAP L76753	D3
Hap 58	1	CAP L76756	D3
Hap 59	1	CAP L76759	D3
Hap 60	1	CAP AF451954	D3
Hap 61	1	Outgroup AJ586558	

Appendix 5.

List of the PCM specimens used in the microsatellite analyses showing specimen ID number, sex, locality of capture, geographical coordinates and subgroup allocation. The exact location (loc) refers to Fig. 4.7 where each specimen can be identified by the number in the green circle.

Specimen ID	Sex	Locality	N	E	Subgroup	Loc
ZII.65	M	Azija Bakoko/Cameroon	03.15'	10.00'	B	1
ZI.17	M	Bakoko/N'Jong/Cameroon	03.30'	10.00'	B	2
CAMI.14	F	Belar/Cameroon	03.05'	10.05'	B	3
ZII.63	F	River Mlonking/Bulu Bush/Cameroon	03.10'	10.20'	B	4
ZVI.33	F	Bikiango Rd/Bipindi/Cameroon	03.10'	10.20'	B	4
ZII.64	M	River Bikiango/Bulu Bush/Cameroon	03.10'	10.20'	B	4
ZIII.31	M	Bipindi/Cameroon	03.10'	10.20'	B	4
CAMI.44	F	SE of Kribi/Cameroon	02.50'	10.30'	B	5
CAMI.42	F	SE of Kribi/Cameroon	02.50'	10.30'	B	5
CAMI.43	F	SE of Kribi/Cameroon	02.50'	10.30'	B	5
CAMI.45	M	SE of Kribi/Cameroon	02.50'	10.30'	B	5
CAMI.46	M	SE of Kribi/Cameroon	02.50'	10.30'	B	5
CAMI.41	M	SE of Kribi/Cameroon	02.50'	10.30'	B	5
CAMI.48	M	SE of Kribi/Cameroon	02.50'	10.30'	B	5
MI.28	M	Yaounde-Kribi Rd/Cameroon	03.30'	11.02'	B	6
ZVI.32	M	Ebodonka Rd/Ebolowa/Cameroon	02.50'	11.10'	B	7
CAMI.149	F	S of Yaounde/N'Yong/Cameroon	03.30'	11.30'	B	8
CAMI.150	F	S of Yaounde/N'Yong/Cameroon	03.30'	11.30'	B	8
CAMI.139	F	Beycar/Cameroon	03.40'	12.00'	B	9
MII.6	M	Akonolinga District/Cameroon	03.45'	12.10'	B	10
CAMI.97	F	Olangina/Cameroons	03.35'	12.15'	B	11
CAMI.109	F	Akonolinga/Cameroon	03.40'	12.15'	B	12
CAMI.107	M	Akonolinga/Cameroon	03.40'	12.15'	B	12
CAMI.134	M	Olangina/Cameroons	03.45'	12.15'	B	13
CAMI.98	F	Olangina/Cameroons	03.45'	12.15'	B	13
CAMI.224	M	Olangina/Cameroons	03.45'	12.15'	B	13
M264	M	Lomie District/Cameroon	03.15'	13.30'	A	14
M342	M	Lomie District/Cameroon	03.15'	13.30'	A	14
M36	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M170	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M29	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M35	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M58	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M95	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15

M136	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M138	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M139	F	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M59	M	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M34	M	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M135	M	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M137	M	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M169	M	Batouri and Lomie/Cameroon	03.45'	13.45'	A	15
M329	F	Meyoss/Batouri District/Cameroon	04.00'	14.00'	A	16
M387	F	Meyoss/Batouri District/Cameroon	04.00'	14.00'	A	16
M985	F	Meyoss/Batouri District/Cameroon	04.00'	14.00'	A	16
M184	M	Gadji/SW of Batouri/Cameroon	04.00'	14.00'	A	16
M372	M	Meyoss/Batouri District/Cameroon	04.00'	14.00'	A	16
M409	F	Obala/Batouri District/Cameroon	03.45'	14.15'	A	17
M470	F	Obala/Batouri District/Cameroon	03.45'	14.15'	A	17
M691	F	Obala/Batouri District/Cameroon	03.45'	14.15'	A	17
M631	M	Obala/Batouri District/Cameroon	03.45'	14.15'	A	17
M799	F	Batouri District/Cameroon	04.15'	14.15'	A	18
M840	F	Obala/Batouri District/Cameroon	04.15'	14.15'	A	18
M841	F	Obala/Batouri District/Cameroon	04.15'	14.15'	A	18
M855	F	Batouri District/Cameroon	04.15'	14.15'	A	18
M865	F	Batouri District/Cameroon	04.15'	14.15'	A	18
M877	F	Batouri District/Cameroon	04.15'	14.15'	A	18
M532	F	Lelo/Batouri District/Cameroon	04.15'	14.15'	A	18
M471	M	Obala/Batouri District/Cameroon	04.15'	14.15'	A	18
M505	M	Obala/Batouri District/Cameroon	04.15'	14.15'	A	18
M720	M	Obala/Batouri District/Cameroon	04.15'	14.15'	A	18
M487	M	Lelo/Batouri District/Cameroon	04.15'	14.15'	A	18
M729	M	Obala/Batouri District/Cameroon	04.15'	14.15'	A	18
CAMII.324	F	Beri/Batouri District/Cameroon	04.30'	14.15'	A	19
CAMII.325	F	Beri/Batouri District/Cameroon	04.30'	14.15'	A	19
CAMII.323	M	Beri/Batouri District/Cameroon	04.30'	14.15'	A	19
CAMII.331	M	Beri/Batouri District/Cameroon	04.30'	14.15'	A	19
M460	F	Obala/Batouri District/Cameroon	03.45'	14.45'	A	20
FC.114	F	Mambili/French Congo	00.40'	15.00'	C	21
FC.124	F	Keba/French Congo	00.40'	15.00'	C	21
FC.147	F	Keba/French Congo	00.40'	15.00'	C	21
FC.115	M	Mambili/French Congo	00.40'	15.00'	C	21
FC.130	M	Mambili/French Congo	00.40'	15.00'	C	21