# Genetic tools for the study and conservation of forest elephants with potential application to the geographical assignment of ivory

**Stéphanie Bourgeois**

**August 2018**

Thesis submitted for the degree of Doctor of Philosophy

Biological and Environmental Sciences, School of Natural Sciences

The University of Stirling

**Declaration**

I hereby declare that this thesis has been composed by myself and that it embodies the results of my own research. Where appropriate, I have acknowledged the nature and extent of work carried out in collaboration with others.

Signed: ……

## Summary

Elephant ivory is one of the most traded products and directly threatens the survival of this charismatic taxon. In Central African rainforests, alarming rates of decline of the elusive forest elephant (*Loxodonta cyclotis*) have been reported over the last fifteen years. There is a crucial need to develop novel tools to underpin traditional approaches that have failed to provide timely information for the rapid intervention of national wildlife law enforcement agencies to hinder the slaughter and stop fuelling the global illegal market.

Genetic data are appealing because they can provide real-time, reliable and shareable information for immediate response on the ground and increased collaboration between countries involved along the global ivory trade chain. Genetic resources for forest elephants are scarce and the available nuclear microsatellite genetic markers present limited opportunities for capacity building and data sharing due to the need for sophisticated equipment and calibration among laboratories. In chapter 2, I present the first genome wide set of 1,365 SNP markers generated for forest elephants and validate genotyping assays for a subset of 107 SNP loci. In chapter 3, I develop two new simple, inexpensive and reliable sexing assays that are suitable for non-invasive DNA samples and can be incorporated into larger SNP panels.

These new genetic resources offer ease of data sharing and technical portability and thus present great potential to provide routine tools. However, wildlife managers are often still reluctant to rely on faecal DNA surveys because there remains the uncertainty of success in recovering sufficient good quality data, while a high investment into fieldwork and laboratory costs is required. In chapter 4, I provide new tools and guidelines for sample collection, storage and preparation in order to increase the quality and cost-efficiency of non-invasive DNA genotyping.

In chapter 5, using an extensive elephant SNP dataset derived from elephant faecal DNA, I investigate the genetic structure of elephants in Gabon. I find that SNP markers reveals the existence of four groups, with additional patterns of genetic differentiation within one group. Moreover, by combining SNP genotyping data with hypervariable mitochondrial DNA control region sequences, I show that it is possible to increase the resolution of geographical assignment to populations defined by a

combination of observed genetic differentiation and investigative needs. This thesis links genetic approaches and management needs, in order to provide fast, accurate, cost-efficient and needs-driven tools to support National wildlife law enforcement agencies into elephant population management and investigation of the illegal ivory trade at the intra-national level.

## Acknowledgements

Many people have contributed to the completion of this project. First, I would like to thank my supervisors, Kathryn Jeffery, Nils Bunnefeld and Kate Abernethy at the University of Stirling, Helen Senn at the Royal Zological Society of Scotland, Rob Ogden and Ross McEwing at TRACE Widlife Forensic network for their support, inputs and advice throughout this project. They together formed a very complementary team from which I found support both in Scotland and Gabon for the very many aspects of this project. This included discussions about project ideas and conservation questions, inputs for data analyses, development of laboratory techniques, skills acquisition, as well as other but not minor aspects that are field logistics, planning and general support.

In Gabon, I deeply thank the Agence Nationale des Parcs Nationaux (ANPN) for hosting the study and providing huge institutional, logistical and financial support in particular Lee J. T. White and Aurélie Flore Koumba Pambo. I thank the Institut de Recherche en Ecologie Tropicale of CENAREST, in particular Alfred Ngomandah for hosting laboratory analyses in Libreville.

I am particularly grateful to ANPN field staff (all the park rangers and wardens) for their assistance in the field, in particular Jean-Paul Mondjo, my field partner during nine months of fieldwork. Forestry concessions (SEEF, Rougier, CEB, Toujours Vert, Foreex), OELO, SFM, WCS, SEGC and PPG kindly provided logistical support. I thank Christopher Orbell and Loïc Makaga for their help in collecting (and continuing to collect) additional samples in areas where I couldn't find time to go. I thank Claude Maganga Mboga for performing all DNA extractions at IRET.

In Scotland, I thank the Conservation and Science Department in the Royal Zoological Society of Scotland for hosting me for 6 months. I am particularly grateful to Jenny Kaden for her teaching, advice and availability in the laboratory work. I would also like to thank all members of the Tropical Ecology and Conservation group at Stirling university for warmly receiving me at each stay in Scotland.

I thank the Centre National de la Recherche Scientifique et Technologique (CENAREST; permit n°AR0016/14/MESRS/CENAREST/CG/CST/CSAR) for research authorization to conduct this study in Gabon, and the Direction Générale de la

Faune et des Aires Protégées (DGFAP) for delivering CITES sample export permit (n°0161).

Funds for the project were provided by CEEAC under the ECOFAC V Programme (Fragile ecosystems of Central Africa) funded by the European Union (contract n° FED/2013/332-349).

## Table of contents

## List of figures

## List of tables

## List of appendices

# Chapter 1
# General Introduction

## 1.1 Illegal wildlife trade

Human activity is responsible for the sixth and fastest biodiversity crisis, with species extinction rates 100 to 1000 times higher than previous crisis events (Barnosky et al., 2011; Pimm et al., 1995). Direct exploitation (i.e. hunting, trading and collecting) is the second major driver of biodiversity loss, after habitat destruction (Dirzo and Raven, 2003). Unsustainable levels of harvesting may drive numerous species to local and ultimately global extinction (Sodhi et al., 2004), with species targeted by the international illegal trade, such as elephants, rhinoceroses, big cats and pangolins facing the highest risk of extinction (Corlett, 2007). The shift from subsistence hunting towards transnational commercial trade of high value products is concomitant with the involvement of international well-organized criminal networks (United Nations Office on Drugs and Crime, 2010). Trade of the most lucrative illegal wildlife products, including elephant ivory, tiger bones, Tibetan antelope, bear gallbladders, rhino horns and exotic birds and reptiles (Haken, 2011) is often entangled with other trafficking crimes (Ratchford et al., 2013). Illegal wildlife trade is now considered as one of the most important illegal trades in the world, along with drugs, counterfeiting, and human trafficking (Haken, 2011).

The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) established in 1975 is the primary instrument to monitor and regulate international trade of wildlife. Commercial trade of species listed in Annex I (threatened by extinction) is prohibited by CITES and non-compliance with regulations by a signatory party may lead to the exclusion of that party from trade in some or all CITES-listed species with economic consequences. Parties are requested to submit annual reports on the legal and illegal trade in CITES-listed species. Analyses of seizure data provided to the CITES Secretariat by its Parties showed that a few regions are the main

suppliers for particular taxa (United Nations Office on Drugs and Crime, 2016). Birds are associated with Central and South America, while Sub-Saharan Africa and South-East Asia are the two major supply regions for the illegal trade of mammals. The main reason is that many species are endemic to these regions. Furthermore, illegal harvesting and export are exacerbated by weak governance and lack of capacity to control borders and vast tracts of wilderness in many source countries (United Nations Office on Drugs and Crime, 2010). The majority of seizures occur in South or South-East Asia (Rosen and Smith, 2010) and are often believed to be destined to China, where wildlife consumption for food, traditional medicines, ornaments or pets is high and growing (Zhang and Yin, 2014). However, even though seizure data are critical to understanding patterns of global trade, their utility has been limited by the variable data quality reported or obtained by each country, with key information lacking such as source, transit and destination of the shipment (Underwood et al., 2013; United Nations Office on Drugs and Crime, 2016). There is a need to develop investigative tools to contribute to a better understanding of the local, regional and transnational trade routes.

## 1.2 The elephant ivory crisis

Elephants are among the most illegally traded species (United Nations Office on Drugs and Crime, 2010, 2016). Poaching driven by the illegal ivory market has decimated African and Asian elephant populations, as levels of harvesting are now unsustainable for the species (Wittemyer et al., 2014). Asian elephants (*Elephas maximus*) have disappeared from 95% of their historical habitat range and habitat loss is currently the main threat to their survival (Sukumar, 2006). However, sustained levels of poaching throughout their range (Blake and Hedges, 2004) has led to strongly biased sex ratios (Sukumar et al., 1998), while alarming rates of poaching have been recently documented (Sampson et al., 2018). Over the last ten years, poaching of African elephants (*Loxodonta spp*.) has risen to an unprecedented level, which correlates with the increased ivory prices and seizures in China (Wittemyer et al., 2014). The weight of ivory seized globally (~40 tons in 2016) has increased three-fold since 2007 and a record number of 22 large scale ivory seizures (>500 kg) was reported in 2016 (Convention on International Trade in Endangered Species, 2017). East and Central Africa are the main sources fuelling the global illegal ivory trade (Underwood et al.,

2013). Eastern African elephants have been deeply affected by poaching with a massive 50% decline reported between 2006 and 2015 (Chase et al., 2016; Thouless et al., 2016). In Central Africa, accurately monitoring the extent of poaching in the dense forest habitat is particularly challenging. However, direct evidence of poaching has been reported in the region since the mid-1990's (Blake et al., 2007; Nishihara, 2003). Furthermore, a 60% decline in elephant numbers was recorded between 2002 and 2012 (Maisels et al., 2013), thus confirming that the current poaching crisis started earlier in Central Africa than in other regions (Wittemyer et al., 2014). Annual continent-wide population decline is estimated at 8% (Chase et al., 2016), which means there is an urgent need to tackle the illegal ivory trade and reverse population trends.

*Historical trade in African ivory*

Historically, the ivory trade has occurred for millenia (Guérin, 2010). For more than four centuries, African elephant populations experienced successive decline and recovery periods driven by levels of exploitation and habitat loss. The rise in the ivory market in Western Europe and the United States at the end of the eighteenth century drove the intensity of elephant hunting in Africa (Hakansson, 2004). As elephant populations rapidly declined and became locally extinct, ivory harvesting expanded from the coasts to the interior in Eastern and Southern Africa (Hakansson, 2004). Similar patterns of rapid local coastal extinctions have been reported as early as the seventeenth century in Central Africa where the ivory trade was entangled with the Atlantic slave trade (Hymas, 2016). West African elephant populations have been the most affected, with overexploitation combined with habitat reduction leading to eradication in most of their historic range in the early twentieth century (Douglas-Hamilton, 1987).

After a relative lull in ivory demand linked to the two world wars and economic depression, another elephant poaching crisis arose in Africa in the 1970s and 1980s due to increasing ivory prices and the proliferation of automatic rifles. This resulted in the rapid and dramatic loss of one half of all elephants in the continent (Douglas-Hamilton, 1987). By that time trade routes had been displaced to Asia, in particular Japan and, to a lesser extent, Hong Kong and China (Martin and Stiles, 2002). In 1989, CITES introduced the first international ban in ivory, which successfully resulted in reversing trends in most African elephant populations (Lemieux and Clarke, 2009). Elephants

were listed in Annex I of CITES which prohibited international trade, with the exceptions of two allowed sales in 1999 and 2008.

Despite the ban, an increase in demand in China has driven a new dramatic surge in illegal ivory trade since 2006. This demand stems from a growing middle and upper class linked to a long tradition of carving, making China the primary global consumer of ivory (Underwood et al., 2013; Vigne and Martin, 2017). It was recently shown that there is a direct correlation between elephant poaching and the volume of ivory auctioned in China until 2011 (Gao and Clark, 2014).. Increased business connections between Eastern Asia and Africa in the past ten years (Vigne and Martin, 2017) and the development of a large black market to fulfil the demand exceeding the legal production in China (Yu et al., 2017) have favoured the expansion of illegal ivory trade.

*Conservation status of African elephant populations*

Historical patterns of population trends and the current demand for ivory in Asia raised different conservation concerns about elephant populations across African regions. Very small and sparse populations remain in West Africa, with most of the historic range now being devoid of elephants in the region (Bouché et al., 2011). In contrast, the 1989 ivory ban was followed, until very recently, by the recovery of East African populations and increases in South African populations, leading to the downlisting of African elephant populations in four countries (Botswana, Namibia, South Africa, Zimbabwe) to CITES Appendix II (Douglas-Hamilton, 2009). Central African rainforests from the western Cameroon to the eastern border of the Democratic Republic of Congo are inhabited by forest elephants (Figure 1.1), which are populations currently listed as endangered by the IUCN (Blanc, 2008).

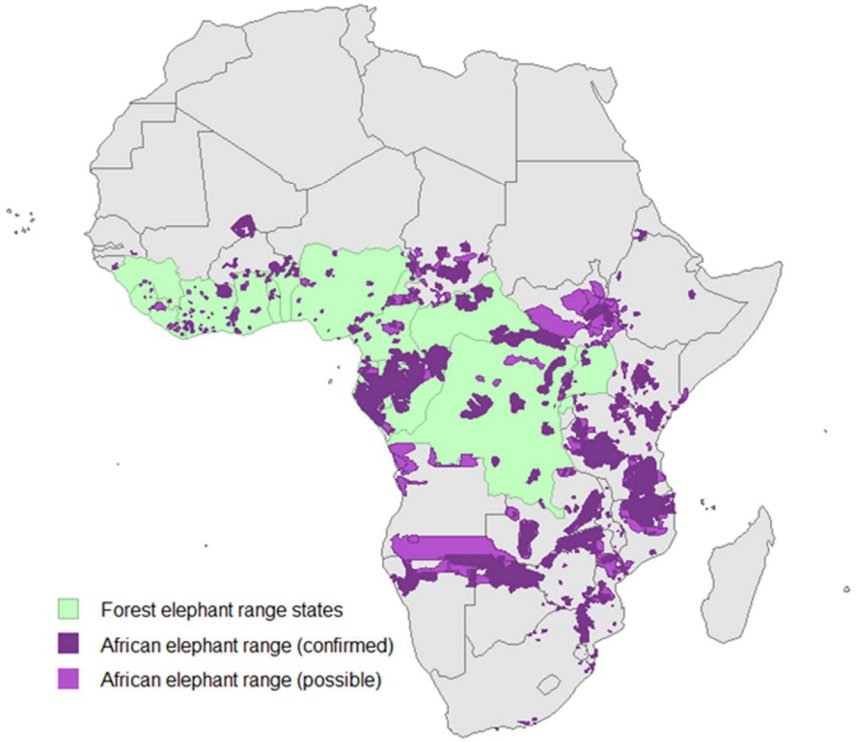


**Figure 1.1** Map of the current range of the African elephants. Central and West African rainforests host the elusive forest elephants (*Loxodonta cyclotis*), while savannah elephants (*Loxodonta Africana*) inhabit Eastern and Southern African plains and bushlands (data from Blanc 2008).

Forest elephants (*Loxodonta cyclotis*) have been recently split from the savannah elephants (*Loxodonta africana*) based on genetic evidence (Ishida et al., 2011; Rohland et al., 2010). While the two-species taxonomic status is supported by morphological (Grubb et al., 2000), behavioral (Schuttler et al., 2014) and ecological (Blake et al., 2009) differences, it has not yet been recognized as such by the IUCN African Elephant Specialist Group (AfESG). Due to a combination of their elusive nature, the remote and often inaccessible areas they inhabit and the dense vegetation that precludes direct counts, forest elephants have received little attention compared to savannah elephants. However, surveys using dung counts showed that they have experienced continuous declines in Central Africa despite the 1989 ivory ban (Lemieux and Clarke, 2009). Hunting pressure on forest elephant harvesting has been exacerbated by specific demand from Japan for their ivory which is harder than savannah elephant ivory

(Nishihara, 2012). Additionally, low annual birth rates result in slow recovery rates of the species (Turkalo et al., 2017). In the Democratic Republic of Congo, it is thought that 95% of the forests are likely to be almost empty of forest elephants (Maisels et al., 2013). Gabon, which has long being considered a stable forest refuge (Lemieux and Clarke, 2009; Michelmore et al., 1994) is now facing intense poaching in the North-East of the country, where an estimated 25,000 elephants have been killed in just a decade (Poulsen et al., 2017). The country was recently identified as one of the main sources for illegal ivory within Africa (Underwood et al., 2013).

The research presented in this thesis focusses on forest elephant as a case study because of the species' urgent conservation needs. The risk of their extinction in the wild is a reality. Urgent measures are required to reverse trends in illegal ivory trade in order to ensure the survival of this emblematic species. Traditional approaches for poaching monitoring and trade investigation on the ground have not been able to produce evidence rapidly enough for immediate intervention to stop poaching and the release of ivory into the global black market. Thus, novel techniques are essential to generate real-time and shareable data.

## 1.3 The growing field of wildlife forensic science

The growing field of wildlife forensic science has emerged from the need to support illegal wildlife trade monitoring and provide evidence to assist in wildlife law enforcement (Ogden et al., 2009). Wildlife forensics uses science to answer a legal question. This applied field is used to provide information to help investigation and evidence to prosecute and convict poachers and traffickers (Huffman and Wallace, 2012).

*Investigative questions*

The most common investigative questions are related to species identification, individual identification, parentage analyses, sexing, and geographic origin. Examples exist in a variety of taxa, including mammals, reptiles, birds, fishes and plants (Iyengar, 2014). Species identification is crucial to determine if the animal or wildlife product belongs to a species that is protected at national or international level, in other terms if it is illegal (Linacre and Tobe, 2011). Individual identification has been used to link a

suspect, a carcass and a poaching crime scene (Lorenzini et al., 2011; Thommasen et al., 1989). Parentage analyses applications have included verification of claims that an individual was captive bred, when regulations allow trade in captive specimen (Dawnay et al., 2009) and identification of an individual whose parents are known (Gupta et al., 2011a). Sexing is important in species where killing of only one sex is prohibited to ensure sustainable harvesting (An et al., 2007). Another increasingly used application is the identification of the geographic origin of a seized animal or wildlife product (Ogden and Linacre, 2015). This approach has been used to prove illegal translocation of game species (Frantz et al., 2006) or investigate the source of illegal wildlife products (Mondol et al., 2015).

*Wildlife DNA forensics*

A variety of methodological approaches are used in wildlife forensics including morphology, toxicology, isotope analyses, veterinary pathology and molecular biology (Bell, 2011; Linacre and Tobe, 2011). DNA techniques are the most widely employed (Ogden, 2010). As opposed to morphological approaches which require intact specimens, molecular markers are ideal because they are suitable on degraded or highly processed specimens (Alacs et al., 2010). In addition, a major contribution of DNA approaches to wildlife forensics is the possibility to examine trace materials (Tobe and Linacre, 2008). Recovering DNA from wildlife samples might be difficult due to environmental degradation by bacteria, UV light, or physical destruction (Jeffery et al., 2007), with processed samples being even more challenging. However, extraction techniques and DNA markers have been developed to work with a variety of materials, some of them containing relatively little DNA (Morin et al., 2007).

Genetic methods have been successfully employed for forensically determining species in a variety of illegally traded animals or animal parts, such as shark fins and bird eggs (Johnson, 2010), whale meat (Baker et al., 2010), duiker bushmeat and crocodile skin handbags (Eaton et al., 2010) and traditional medicines claiming to contain tiger (Tobe and Linacre, 2009). Identifying species or the broad geographic origin of a sample often involves mitochondrial DNA which is easier to recover from degraded or processed samples because it is present in multiple copies per cell (Tobe and Linacre, 2008). However, nuclear DNA markers are required for other applications. While microsatellites have traditionally been the marker of choice for individual

identification and parentage analyses (Dawnay et al., 2008, 2009), single nucleotide polymorphism (SNP) markers have more recently gained traction as promising markers in wildlife forensics (Ogden, 2011). SNP assays target a small amplicon size in comparison to DNA sequencing (Gettings et al., 2015) which increases genotyping success with degraded DNA samples. The use of SNPs has long been limited by low availability in non-model species, until increasing availability and decreasing costs of next generation sequencing enhanced the initial discovery of SNPs in a number of species (Kumar et al., 2012). SNP markers have a string of advantages, including ease of genotyping and data transfer compared to microsatellites which require allele size calibration across laboratories (Ogden, 2011). In addition, the bi-allelic nature of SNPs facilitates their formal forensic validation and make them a marker of choice for wildlife DNA forensic applications (Ogden, 2011).

*Forensic intelligence*

Genetic techniques need to be formally validated for use in forensic casework, with the ultimate goal of admissibility of evidence in the legal system (Ogden et al., 2009). Validation involves demonstration of the accuracy, precision, and reproducibility of a procedure in order to produce evidential data (Dawnay et al., 2009). Similar techniques, but with lower level of validation and quality control may be implemented at a lower cost and used to produce decisive data for intelligence purposes (Ogden, 2010). The importance of forensic intelligence to support crime investigation has been highlighted (Ribaux et al., 2006). A key example is the use of DNA techniques to provide invaluable intelligence information on the source of wildlife products to support investigation of the illegal wildlife trade.

A first approach based on phylogeography of mtDNA haplotypes has been used to identify broad geographic regions of origin of individuals, but these markers often display insufficient variation to distinguish between close populations (Chapman et al., 2009; Murray-Dickson et al., 2017). Population assignment is another approach, which estimates the probability of a nuclear genotype to belong to each putative population based on interpopulation differences in allele frequencies distribution, thus enabling the detection of finer-scale genetic patterns (Alacs et al., 2010; Manel et al., 2005). The development of geographical assignment tests is challenging because they are based on the identification of underlying spatial genetic patterns and therefore rely on marker

availability and the existence of a genetic reference database (Ogden and Linacre, 2015). As a consequence and even though its potential to detect poaching hotspots is well established (Manel et al., 2002), geographic assignment is a relatively new area for conservation genetics (Ogden and Linacre, 2015). Geographic assignment tests have been successfully used to source the origin of traded live animals (Ghobrial et al., 2010) or wildlife products such as leopard skins (Mondol et al., 2015). Major applications include controls of traded timber and fish products in order to detect illegal logging and fishing and false certification through the identification of the country or sea of origin (Degen et al., 2013; Nielsen et al., 2012). Another key example is the investigation of the source of illegally traded elephant ivory to understand poaching hotspots (Wasser et al., 2015), which was adopted as a resolution by CITES (2014).

## 1.4 Forensic tools available to investigate the ivory trade

The development of forensic tools over the last twenty years has helped to understand and investigate the global ivory trade of raw ivory and worked ivory items.

*Elephant species identification*

The first investigative question requested by law enforcement officers is the demonstration of the illegality of the trade. All three living elephant species are currently listed under CITES Appendix I, with the exception of four African elephant populations listed in Appendix II. Nonetheless, all commercial international trade in the ivory of African elephants is currently prohibited under CITES. Pre-convention ivory is an exemption, as is ivory from the Woolly mammoth (*Mammuthus primigenus*), an extinct species that is not listed on CITES. In addition, national ivory trade is regulated by national laws and Thailand allows trading of ivory from domesticated elephants (Krishnasamy et al., 2016). The distinction between extinct and living species and between Asian and African elephants is therefore crucial to distinguish legal and illegal ivory products.

Traditional morphological techniques using patterns of crossing lines in a transverse section of the tusk (Schreger lines) enable elephant ivory to be distinguished from substitutes and other species including mammoth (Espinoza and Mann, 1993). However, this technique may be more challenging to employ on worked ivory items.

Moreover, using morphology to distinguish between tusks belonging to each of the three living elephant species is challenging. Forest elephant tusks are usually thinner and straighter (Grubb et al., 2000) but individual variations are observed (*pers. obs*.). Even though forest elephant ivory is considered to be harder than savannah elephant ivory (Nishihara, 2012), species identification of worked ivory items using morphological examination would be extremely difficult.

DNA techniques offer a powerful alternative in the absence of morphological features to distinguish between species. DNA extraction and amplification has been demonstrated from raw ivory (Comstock et al., 2003) as well as processed ivory items (Gupta et al., 2011a). Effort has been invested into improving DNA yield obtained from ivory samples (Mailand and Wasser, 2007; Winters et al., 2018) and subsequent genotyping success (Lee et al., 2009). Mitochondrial DNA markers have been successfully used for species identification to establish whether the ivory was of mammoth, Asian elephant or African elephant origin using sequencing (Gupta et al., 2011b) or a real-time PCR assay (Wozney and Wilson, 2012).

*Individual-level genetic testing*

Genetic tests have been used to answer several other investigative questions. Individual identification using nuclear DNA microsatellite markers has been successfully used to link paired tusks to a carcass in order to prove a poaching case in the Asian elephant (Gupta et al., 2011c) and across multiple large ivory seizures in order to better understand links between trade networks in the African elephant (Wasser et al., 2015). Sexing tests suitable for degraded carcass samples such as hair, bone or skin samples have been developed as a useful indication of poaching in Asian elephants in which only males wear tusks (Gupta et al., 2006).

*Geographical assignment of ivory*

A major investigative question requested by law enforcement agencies is the geographic origin of seized ivory. Global monitoring protocols such as the Elephant Trade Information System (ETIS) used by CITES have proven to be very efficient at understanding main trade routes from Africa to consumer countries in Asia (Underwood et al., 2013). However, there is a lack of information about trade patterns from poaching sites to the main sea and airports of shipment from Africa. For instance, Cameroon,

Nigeria and Ghana were identified as major exporters in Central and West Africa between 2001 and 2008 (Milliken, 2013). However, Democratic Republic of Congo, Gabon, and Republic of Congo were identified by late population census as major source countries fuelling the traffic during that period (Maisels et al., 2013). This highlights the urgent need for investigative tools to identify countries of origin of ivory seizures.

A CITES resolution requests that all Parties forensically test all large seizures (>500 kg) to determine the origin of the ivory (Convention on International Trade in Endangered Species, 2014). Two approaches are currently available to source ivory products. Isotopic analyses have been proposed since 1990 (Van der Merwe et al., 1990). This technique has been successful at locating the origin of ivory at regional scale. However, high inter-individual variations limit its precision (median accuracy of 876 km) especially in the forest elephant (Ziegler et al., 2016). Alternatively, DNA techniques have been developed fifteen years ago (Comstock et al., 2003) and probably hold the greatest promise to understanding the main trade routes and origin of illegal wildlife products at varying spatial scales (Ogden et al., 2009; Wasser et al., 2008). Species identification tests using mitochondrial DNA markers have been expanded to the identification of all three living elephant species (Kitpipit et al., 2017; Lee et al., 2009) in order to better understand the source of ivory at broad scales. Microsatellite markers pioneered the genetic geographical assignment of African elephant ivory in 2004 and greatly improved the resolution within the African continent (Wasser et al., 2004). Using spatial smoothing allele frequency estimates from reference samples with a spatial correlation between neighbouring localities, this assignment test enables to identify the source of ivory with a precision of about 350 km for forest elephants and 490 km for savannah elephants. This approach has been successfully used to investigate the geographic origin of several major seizures during the last two decades and confirm Central Africa and East Africa as the main regions targeted by poachers within Africa (Wasser et al., 2008, 2015). In addition, mitochondrial DNA control region sequencing has been proposed as a complementary approach to microsatellites in order to improve the success of genetic assignment of ivory, because of its high variability and distinct geographic signal in comparison to nuclear DNA due to female matrilocality (Ishida et al., 2013).

*Limitations of the current methodological approaches*

Despite multiple tests being available to answer a variety of investigative questions about ivory legality, geographic origin and sample matching, a number of limitations exist. Firstly, not all investigative questions can be answered with the existing tests. Investigative questions might cover different geographical scales, from the species range to a specific population or area of interest. The existing assignment test has greatly improved the understanding of global patterns of the illegal ivory trade (Wasser et al., 2015) but is not informative enough for country-based law enforcement actions. National wildlife law enforcement agencies expect a much higher resolution in locating poaching hotspots to help guiding antipoaching activities at the land management units' level (i.e. Protected Areas or logging concessions). In addition, more understanding is needed about the structure of the ivory trade within Africa including collection at poaching sites and shipment between regions to the ports of exit from Africa (Milliken, 2013; United Nations Office on Drugs and Crime, 2013). As transnational ivory trade relies on regional trade, it has been highlighted that investigations should focus on the nodes, i.e. locations or middlemen connecting various networks and scales, in particular the ones who are fuelling the trade by connecting the local with the transnational networks (Titeca, 2018). There is an urgent need to develop intelligence tools that are adapted to a finer scale investigation of the local and regional ivory trade. This would be a major step contributing to a better allocation of limited antipoaching resources and reduction of illegal exportation.

Secondly, existing tests are not endorsed as routine practice by management and law enforcement agencies. Only 10% of large seizures have been genetically analysed between 2011 and 2014 (Convention on International Trade in Endangered Species, 2017). Time lag in delivering intelligence information, reluctance from countries to export samples to foreign forensic facilities and CITES requirements for export of samples all impede the use of forensic testing. Building in-country or regional forensic capacity is needed for timely response and increased access to forensic tools by law enforcement investigators (McEwing and Ahlers, 2016). There is a need to design accurate and reproducible forensic techniques that can be easily implemented in countries directly involved in the illegal wildlife trade and allow data sharing between source, transit and consumer countries. This is crucial to foster direct communication and collaboration through rapid exchange of intelligence information and data sharing

for reference database building and comparison of individual genotypes. This would be a major step towards a joint and strong commitment of all countries involved along the illegal ivory trade chain.

## 1.5 Justification of the study

This thesis proposes to address the limitations discussed above, by delivering tools to increase resolution of geographic assignment and ease of technical transfer and data exchange. The type of markers available (Hess et al., 2011), the existence of good quality reference datasets (Paetkau et al., 2004), and the ability to detect underlying genetic structure at the target scale (Dimauro et al., 2015; Puckett and Eggert, 2016) are factors influencing the resolution of assignment tests. Ease of technical portability and suitability for data sharing are inherent to the choice of markers and assays.

*Available genetic resources*

Genetic resources for elephants have been developed for *L. africana* and *E. maximus*, including numerous primers for presumed neutral genetic markers such as mitochondrial control region and microsatellites (Ishida et al., 2012; Nyakaana et al., 2005), while forest elephants have received little attention. Only very recently were species-specific microsatellite loci generated for *L. cyclotis* (Gugala et al., 2016). However, using microsatellite markers across species may result in null alleles and size homoplasies (Queloz et al., 2010). In addition, SNP resources for elephants are scarce (Sharma et al., 2012), despite their high conservation profile and the rapid development and growing interest in the use of SNP markers. The available genetic resources may have limited power to detect genetic differentiation and perform geographic assignment, because neutral markers usually underperform markers under selection (Landguth and Balkenhol, 2012). Novel genetic markers are urgently needed to better inform forest elephant conservation and management.

Existing geographic assignment tests for African elephants are based on microsatellite or mitochondrial DNA markers. It would be challenging to transfer these tests to countries lacking laboratory capacity because microsatellite genotyping and mitochondrial DNA sequencing still require expensive and specialized equipment such as a sequencer and often need to be outsourced. More importantly, the huge investment

in time and efforts required to transfer microsatellite-based tests among laboratories both for calibration of allele sizes among instruments and formal forensic validation (Ogden, 2011) is a major limitation of this approach. There is an urgent need to develop new genetic markers that are suitable for capacity building and data sharing. In that respect, SNP markers, whether they are derived from nuclear or mitochondrial DNA, would enable a great improvement because they allow the development of in-house genotyping assays.

In a species exhibiting high differences in social behaviour, sex of individuals is a very important variable in any population structure study. In elephants, dispersal is biased towards males, while females show high fidelity to their natal group and area (Archie et al., 2008). Any interpretation of assignment results should account for adult males being potential migrants, thus any genetic test will reveal their natal location (Ogden and Linacre, 2015). Two genetic sexing approaches have been developed in elephant species. The first test is based on restriction fragment length polymorphism within the orthologous sexual chromosome zinc finger protein genes (ZFX/ZFY) (Fernando and Melnick, 2001; Munshi-South et al., 2008). The second method targets a locus specific to chromosome Y (SRY or AMELY), along with an additional control test using mitochondrial DNA or chromosome X specific regions (Ahlering et al., 2011; Gupta et al., 2006). These techniques are based on fragment differentiation by gel electrophoresis and therefore require multiple steps and sufficient fragment sizes (~140-200 bp) for gel resolution of differences. For application to forensic samples, there is a need for fast, reliable and inexpensive sexing tests that are suitable for all types of samples including short degraded DNA fragments.

*Genetic reference databases*

The availability of an extensive and good-quality reference database is key to develop assignment tests (Dormontt et al., 2015; Ogden and Linacre, 2015). In elusive or endangered species, sample collection relies on non-invasive sampling, with faecal samples being the gold standard in elephant species. Faecal genotyping is associated with pitfalls including low DNA yield, DNA degradation and presence of PCR inhibitors that may lead to low genotyping success and high genotyping error rates. To our knowledge, few studies have applied SNP genotyping to large non-invasive faecal datasets yet (Goossens et al., 2016; Spitzer et al., 2016; von Thaden et al., 2017).

Substantial work is needed to refine sample collection and preparation in order to overcome cost and technical limitations of genetic analyses using degraded DNA samples such as ivory as well as faecal samples used to build large reference databases.

*Population genetics of forest elephants*

Forest elephants have been largely understudied compared to the savannah elephants. Two studies have investigated genetic structure at regional (Bawe-Johnson, 2008) and local scale (Eggert et al., 2014), both using microsatellites developed for savannah elephants. Very recently, another study investigated genetic structure within Central Africa (Ishida et al., 2018) using a novel set of microsatellite markers developed for forest elephants (Gugala et al., 2016). Results from these studies suggested very high gene flow across Central Africa, with evidence of weak genetic differentiation between the western and eastern sides of the Congolese forest block (Ishida et al., 2018), and between coastal and interior populations within the western group (Bawe-Johnson, 2008). In addition, population structure was weak but significant at the local scale within the coastal group (Eggert et al., 2014). There is a need for a more thorough investigation of the population structure of forest elephants at fine scale, that would benefit from increased sample size and development of SNP markers, which have more power than microsatellites to detect weak genetic differentiation (Mesnick et al., 2011; Morin et al., 2009).

*Study area*

The research presented in this thesis focusses on forest elephants within Gabon, because the country has demonstrated a strong commitment into forest elephant conservation. It was the first Central African country to burn its ivory stockpile in 2012. In 2014, in addition to its adherence to several international agreements and conventions, Gabon was a founding member of the Elephant Protection Initiative, and consequently, engaged into the design of a National Elephant Action Plan. Gabon hosts half of the remaining forest elephant population (Maisels et al., 2013), which represents one of the last quasi-continuous elephant population in Africa (Thouless et al., 2016). The country is still 85% covered by forest (Verhegghen et al., 2012), even though rapid on-going development of extractive industry and associated infrastructures may threaten

genetic connectivity. The National Park Agency (ANPN) was created in 2007 to manage and protect the new network of Protected Areas created in 2003 (Figure 1.2).

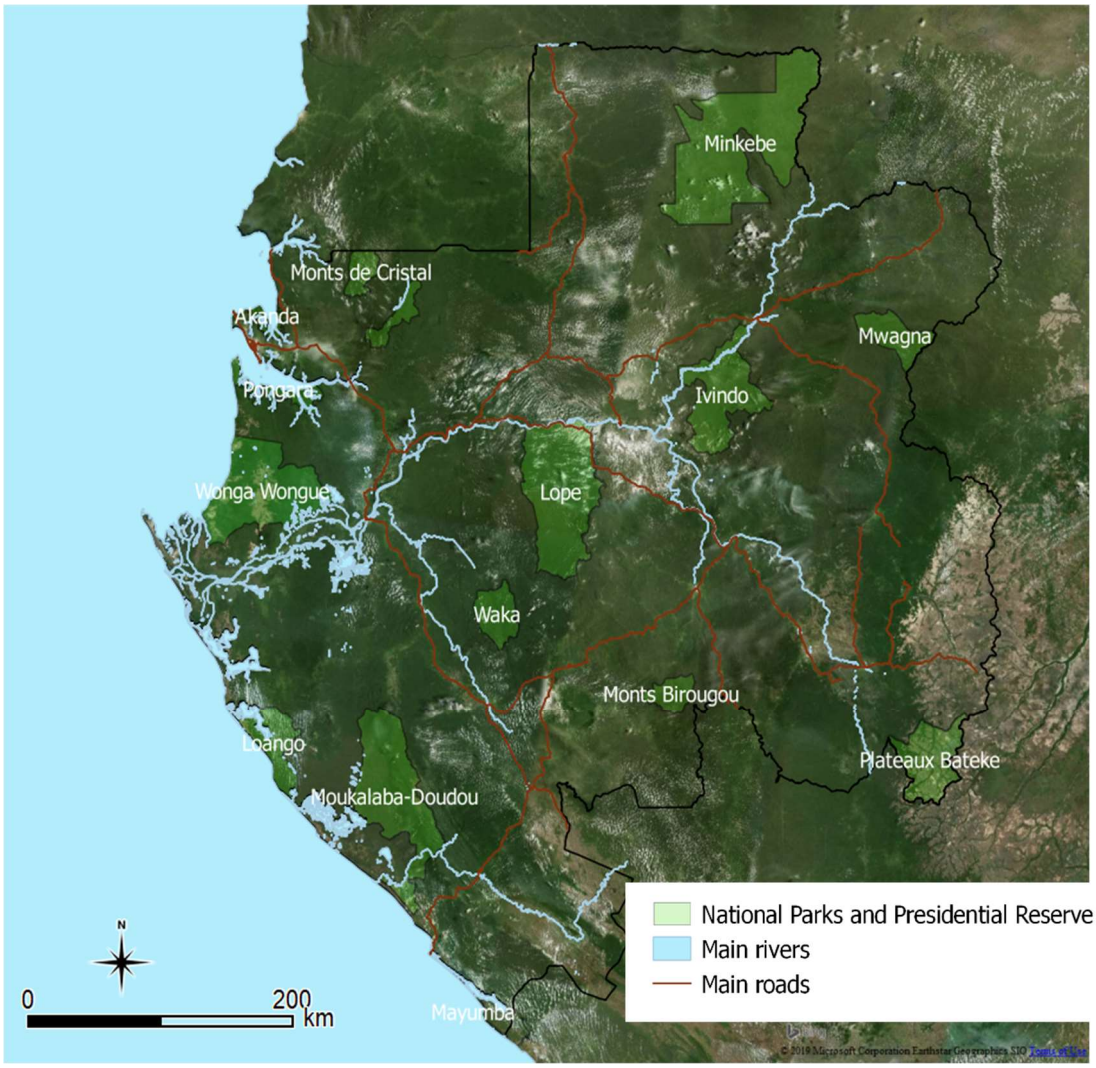


Figure 1.2 Map of Gabon's network of National Parks and Presidential Reserve.

In 2013, a routine monitoring study revealed that a major poaching crisis was occurring in the north-east of the country. However, by the time the wildlife authorities were aware of the scale of the crisis, an estimated 25,000 elephants had been lost to poachers (Maisels et al 2013). Frustrated by a lack of tools to enable earlier poaching detection and more rapid intervention, ANPN developed partnerships to explore novel techniques in wildlife forensics and how they could be adapted to the situation in Gabon. In this context, genetic methods were appealing because they were not only fast and cost-efficient tools, but can produce robust data. This study was designed to produce the methods and data that would respond to urgent, real-time needs of the national authorities and be more effective in informing Gabon's forest elephant

management strategy. It is urgent for the survival of the forest elephant species that pressure from hunting is relieved in Gabon. In order to achieve this, success will also depend on more efficient targeting of limited anti-poaching resources as well as the engagement of all parties concerned in stopping the illegal exportation of ivory from the country. Reliably and rapidly tracing the exact origin, and trade routes, of illegally hunted Gabonese ivory will ensure a significant step forward in the effectiveness of national and international protection measures.

*Aims and outline of this thesis*

The overall objective of this study is to increase genetic resources available for forest elephants in order to provide fast, accurate, cost-efficient and needs-driven tools to support National wildlife law enforcement agencies into elephant population management and investigation of the illegal ivory trade at the intra-national level. I combine intensive genetic sampling with last-generation genetic techniques to achieve the following objectives:

- Develop new SNP (chapter 2) and sexing (chapter 3) markers for forest elephants
- Develop tools to improve the quality and cost-effectiveness of a non-invasive genetic dataset (chapter 4)
- Understand genetic differentiation of forest elephants within Gabon (chapter 5)
- Maximize power of geographic assignment using a combination of genetic markers (chapter 5)

In chapter 2, I develop the first genome-wide set of SNP markers generated for forest elephants and validate genotyping assays for a subset of 107 SNPs that are suitable for degraded DNA samples and offer technical portability for capacity building. In chapter 3, I develop two new simple, inexpensive and reliable sexing assays that are suitable for non-invasive DNA samples and can be incorporated into larger SNP panels. This work was based on tissue samples (blood, skin, muscles). In chapter 4, I develop tools and guidelines to improve cost-efficiency of genotyping using degraded samples. This work is based on faecal samples because they are the samples of choice to create large scale reference databases in elusive and endangered species. Even though no ivory samples were included in this thesis, this approach is also suitable for this type of samples which contain relatively little DNA that may be degraded. In chapter 5, I use

SNP markers and techniques developed in chapter 2, 3 and 4 and an extensive faecal sample set to investigate genetic variation of forest elephant throughout Gabon. As I expect the population to be poorly structured, I combine SNP genotyping data with hypervariable mitochondrial DNA control region sequences to maximize the power of geographic assignment across populations defined by a combination of observed genetic differentiation and investigative needs. In chapter 6, I discuss how the results of this thesis are used to underpin rapid conservation actions, planning and responses and present future directions for research in forest elephant conservation genetics and forensic science.

# Chapter 2

# SNP discovery and panel characterization in the African forest elephant

An adapted version of this chapter has been published as:

Bourgeois S.[1,2,3], H. Senn[2], J. Kaden[2], J. B. Taggart[4], R. Ogden[5,6], K. J. Jeffery[1,3,7], N. Bunnefeld[1], K. Abernethy[1,7], R. McEwing[5]. SNP discovery and panel characterization in the African forest elephant. Ecology and evolution 8, 2207–2217.

[1]Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom; [2]WildGenes laboratory, The Royal Zoological Society of Scotland, RZSS Edinburgh Zoo, Edinburgh, United Kingdom; [3]Agence Nationale des Parcs Nationaux, Libreville, Gabon; [4]Aquaculture Pathfoot Building, University of Stirling, Stirling, United Kingdom; [5]TRACE Wildlife Forensics Network, Edinburgh, United Kingdom; [6]Royal (Dick) School of Veterinary Studies & The Roslin Institute, University of Edinburgh, United Kingdom; [7]Institut de Recherche en Écologie Tropicale, Libreville, Gabon.

**Contributions**: S.B., H.S., R.O. and R.M. designed the study. S.B. collected the samples. S.B. and J.K. performed laboratory work. J.B.T. designed ddRAD protocol and performed the sequencing run. R.M. supervised laboratory work. H.S. performed bioinformatic analyses. S.B. and H.S. performed statistical analyses. S.B. and H.S. wrote the manuscript. H.S., J.K., R.O., K.J., N.B., K.A., J.B.T and R.M. revised the manuscript. H.S., K.J., N.B., K.A. and R.M. supervised the study.

## 2.1 Abstract

The continuing decline in forest elephant (*Loxodonta cyclotis*) numbers due to poaching and habitat reduction is driving the search for new tools to inform management and conservation. For dense rainforest species, basic ecological data on populations and threats can be challenging and expensive to collect, impeding conservation action in the field. As such, genetic monitoring is being increasingly implemented to complement or replace more burdensome field techniques. Single nucleotide polymorphisms (SNPs) are particularly cost-effective and informative markers that can be used for a range of practical applications, including population census, assessment of human impact on social and genetic structure, and investigation of the illegal wildlife trade. SNP resources for elephants are scarce, but next generation sequencing provides the opportunity for rapid, inexpensive generation of SNP markers in non-model species. Here we sourced forest elephant DNA from 23 samples collected from 10 locations within Gabon, Central Africa, and applied double-digest restriction-site-associated DNA (ddRAD) sequencing to discover 31,851 tags containing SNPs that were reduced to a set of 1365 high-quality candidate SNP markers. A subset of 115 candidate SNPs was then selected for assay design and validation using 56 additional samples. Genotyping resulted in a high conversion rate (93%) and a low per allele error rate (0.07%). This study provides the first panel of 107 validated SNP markers for forest elephants. This resource presents great potential for new genetic tools to produce reliable data and underpin a step-change in conservation policies for this elusive species.

## 2.2 Introduction

Evidences of lack of nuclear gene flow and high genetic divergence were used to split African elephants into two species, with the forest elephant (*Loxodonta cyclotis*) now established as a distinct species from the savannah elephant (*Loxodonta africana*) (Roca et al., 2015), even if not yet recognized as such by the IUCN African Elephant Specialist Group (AfESG). Due to its elusive nature and remote tropical rainforest habitat, compounded by a lack of species-level recognition, the African forest elephant has largely been understudied compared to the savannah elephant. Within the last

decade, intense poaching and habitat reduction have caused a decline of more than 60% in Central African elephant numbers (Maisels et al., 2013). Gabon now hosts half of the remaining global population of *L. cyclotis*, but the north-east of the country suffered the steepest declines recorded for the decade 2004-2014 (Poulsen et al., 2017) and was revealed to be a major source of illegal ivory within Africa (Wasser et al., 2015). To respond to this conservation crisis, there is a desperate and immediate need to develop efficient tools to monitor forest elephant populations and threats.

Genetic tools have been widely used to understand elephant ecology and inform their management and conservation (Archie and Chiyo, 2012) and have shown tremendous potential to help understanding of the illegal ivory trade (Wasser et al., 2015). Numerous primers for presumed neutral genetic markers, including mitochondrial control region and microsatellites are available in the literature for *L. africana* and the Asian elephant (*Elephas maximus*) (Ishida et al., 2012; Nyakaana et al., 2005). However, nuclear genetic studies of *L. cyclotis* have all used microsatellite markers developed for *L. africana* (Bawe-Johnson, 2008; Eggert et al., 2003, 2014; Munshi-South, 2011; Schuttler et al., 2014). While it is widely recognized that null alleles and size homoplasies may occur as a result of using microsatellite markers across species (Queloz et al., 2010), only very recently were species-specific microsatellite loci generated for *L. cyclotis* (Gugala et al., 2016).

Microsatellites have long been the most widely used genetic markers in ecological studies, primarily due to their high mutation rate and polymorphism (Ellegren, 2004; Slatkin, 1995). However technological advances are driving a shift in the field of molecular genetics from microsatellite to single nucleotide polymorphism (SNP) markers. Numerous studies have revealed the great potential for SNPs to be cost-effective and highly informative markers (Helyar et al., 2011; Morin et al., 2004; Vignal et al., 2002), with a string of advantages including low error rates (Ranade et al., 2001), small amplicon sizes (<100 bp) (Senge et al., 2011), and technical portability and reproducibility across laboratories (Seeb et al., 2011). However, SNP resources for elephants are scarce, despite their high conservation profile and genome data being available for their development (Dastjerdi et al., 2014; Elephant genome project, 2009). To date, SNP markers have been used for species differentiation in African elephants (Ishida et al., 2011; Roca et al., 2001) and to study genetic diversity and structure of the highly endangered Bornean elephant (*Elephas maximus borneensis*) (Goossens et al.,

2016; Sharma et al., 2012). However, novel genetic markers are urgently needed to better inform forest elephant conservation and management. The application of SNP markers to understand forest elephant population status and connectivity, and the illegal ivory trade would tackle some priority areas of research.

The use of SNPs has been limited by the cost and availability of SNP discovery techniques, especially in non-model organisms. Recently, advances in next-generation sequencing technologies and bioinformatics analyses have revolutionized the development of large numbers of genetic markers followed by the selection of a reduced high-quality panel for a wide variety of species (Davey et al., 2011). Reduced representation genome sequencing approaches, where a subset of the genome is partitioned and sequenced, have arisen as inexpensive and simple methods for *de novo* SNP discovery in model and non-model species (Van Tassell et al., 2008). One of these approaches is the restriction site-associated DNA (RAD) sequencing, which targets short fragments of DNA adjacent to a particular restriction enzyme site (Baird et al., 2008). The simplification of the procedure in the double-digest RAD (ddRAD) approach, through the elimination of random shearing and the use of two-enzyme digestion followed by strict size selection (Peterson et al., 2012), has allowed discovery of targeted panels of a few thousand SNPs in a number of non-model species (Adenyo et al., 2017; Cruz et al., 2016). Notably, RAD methodologies permit simultaneous SNP discovery and genotyping. Where required, allele frequency data generated for multiple individuals from different locations can be exploited to better inform a subsequent targeted SNP assay design phase, reducing potential ascertainment bias (Clark et al., 2005; Nielsen, 2004).

In this study, we used ddRAD to discover thousands of potential SNP loci in the endangered forest elephant. Our aims were to (1) generate and identify potential SNP loci in forest elephants and (2) validate a subset of around a hundred SNP markers on a larger sample set via genotyping assays and comparison between genotyping and sequencing data.

## 2.3 Materials and methods

*Samples*

Sixty-four samples from 58 forest elephants in Gabon were available for the SNP discovery phase. Blood, muscle and skin samples were collected, as available, from 14 elephants immobilized for collaring operations in 2003 (Blake et al., 2008) and 44 elephant carcasses found in 14 locations (Figure 2.1). Samples were selected from a range of geographic locations across Gabon to reduce possible ascertainment bias (Nielsen, 2004). A second batch of 20 samples was added for candidate SNP validation. These samples were collected from 6 poached elephants in Gabon and 8 elephants immobilized for collaring operations in the adjacent Odzala-Kokoua National Park in Congo in 2014 (Figure 2.1). DNA was extracted primarily using the Qiagen DNeasy Blood and Tissue kit according to the manufacturer's protocol. In order to assess genotyping errors, 13 individuals were repeated using two different sample types and 8 blood samples were extracted twice independently.

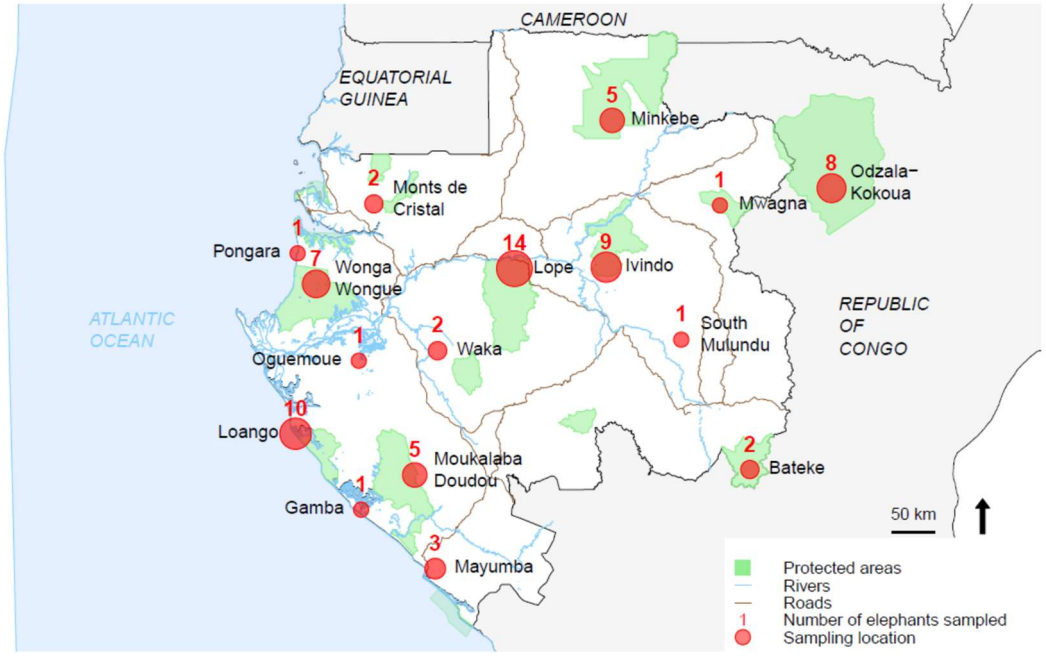


**Figure 2.1** Distribution of elephant sampling localities throughout Gabon. The circles are proportional to the number of elephant sampled (with the total number indicated above). The number and location of samples used for the ddRAD analysis are given in Table 2.1.

*ddRADseq library preparation*

DNA quality was assessed via agarose gel electrophoresis on a 1% gel and only non-degraded DNA (as judged by a tight high molecular weight band against a lambda standard) was selected for the library preparation stage. DNA was quantified using a Qubit Broad Range dsDNA Assay (Thermofisher Scientific) according to the manufacturer's instructions and normalized to c. 7 ng µl-1.

A double digest RAD (ddRAD) library was constructed according to a modified protocol of the original Peterson et al. (2012) methodology. This is described in detail elsewhere (Brown et al., 2016; Manousaki et al., 2016). High quality DNA suitable for ddRAD library preparation was obtained for 23 elephants. An additional positive control (repeated individual, LOC0279_d) was included to allow for quality control of the experimental process and for assessment of genotyping error-by-read depth. Furthermore, each sample was processed in quadruplicate to enhance evenness of coverage of samples within the library. Briefly, individual genomic DNAs (24 x 4 replicates; 21 ng each) were restriction digested by *SbfI* and *SphI* and then Illumina specific sequencing adaptors (P1 & P2) were ligated to fragment ends. The pooled samples were size selected (320-590bp fragments) by gel electrophoresis, PCR amplified (15 cycles) and the resultant amplicons (ddRAD library) were purified and quantified. Combinatorial inline barcodes (5 or 7 bases long) included in the P1 &P2 adaptors allowed each sample replicate to be identified post sequencing. The ddRAD library was sequenced on the Illumina MiSeq Platform (a single paired-end run; v2 chemistry, 2 x 160 bases).

*Bioinformatics*

The sequences were quality assessed using FastQC (Andrews, 2010) and the reads demultiplexed by barcode using the *process_radtags* module (default parameters) of the stacks bioinformatics pipeline (Catchen et al., 2013). This module also filtered out low quality reads. The retained reads, now missing variable length barcodes, were then trimmed to a standard 148 bases in length. Demultiplexed read files were concatenated into read files for each individual (four barcode combinations per individual, see above). For each individual, matching forward and reverse reads were then concatenated into a single longer "artificial" read using a custom perl script. This

was to allow for tracking of the closely linked read 1 and 2 loci in subsequent bioinformatics analyses.

The individual data were then processed using the *denovo_map.pl* module of stacks (m 10 -M 2 -n 1) to assemble and create a catalogue of genetic loci contained in the data. The Stacks scripts *export_sql.pl* and populations and five filtering steps were used to retain all loci that fulfilled the following criteria:

1. Contained exactly 1 SNP (in the concatenated forward and reverse reads) to remove physically linked markers and ensure availability of a constant sequence surrounding the target SNP to facilitate primer design;

2. Contained exactly 2 alleles, as the presence of more than two alleles might represent repeat sequence found at multiple sites within the genome;

3. Were present in the data for ≥ 10 elephants and had a read depth of ≥ 10 reads per individual to maximize the likelihood of the SNP being real;

4. Were heterozygous in at least one individual but not in all individuals in the dataset; both the lack or apparent fixation of heterozygotes could be indicative of variation between repeat sequence found at more than one locus;

5. Had a minimum of 50 bases flanking sequence either side of the SNP to ensure that the sequence meets the requirements for the design of a genotyping probe assay (LGC Genomics, 2014).

*SNP validation*

In order to validate the results from the bioinformatics pipeline, two sets of SNPs were tested for validation using different approaches. The default parameters were used for all programs, unless otherwise specified. First, a random subset of 22 SNP loci were selected as candidates for assay design and ordered from LGC Genomics using the Kompetitive allele-specific PCR (KASP) system to evaluate the conversion rate, i.e. the proportion of successful assays that resulted in distinct genotyping clusters. They were run on a StepOne real-time PCR machine (Applied Biosystems) on the DNA samples used to generate the library. PCR was carried out in 8 µl single-locus reactions following thermal cycling conditions recommended in the KASP User guide (LGC Genomics, 2013). The quality of the genotyping cluster plots was visually assessed.

When the probe did not produce distinct clusters, further examination of the SNP containing sequences was conducted by aligning them against the *L. africana* genome (Elephant genome project, 2009) using NCBI's Basic Local Alignment Search Tool (BLAST) to investigate any repetition within the genome.

Second, a genotyping panel was selected among the candidate SNP markers using a combination of measures of genetic diversity and divergence, in order to validate assay performance and select potentially informative markers with the aim to explore genetic variation among individuals and populations. The filtered matrix of sequencing genotype data at 1365 loci was examined for "missingness" using Plink (Purcell et al., 2007). A principle components analysis was run using the package adegenet (Jombart, 2008) in R (R Core Team, 2016) to examine structure in the data matrix (results not shown). Three population clusters were then defined based on a mixture of the geographic and genetic information: North-East (South Mulundu, Ivindo, Minkébé, Monts de Cristal), Central (Lopé, Waka), Coastal (Wonga Wongue, Mayumba, Loango, Moukalaba Doudou) (Figure 2.1). These groups were used to calculate and rank loci according to expected heterozygosity ($H_E$), global $F_{ST}$ and $F_{ST}$ in the three pairwise population combinations. Loci were then given an unweighted joint rank across all five categories and the highest ranking 266 SNPs were chosen. Finally, loci were excluded that had zero or > 1 BLAST matches against the *L. africana* genome using a discontiguous megablast of the 148 bases sequence containing the SNP. The cut-off e-value was set at $10^{-10}$ with a minimal alignment length of 100 bp including the SNP site. Sequences with no matches based on these criteria were excluded on the basis that they could be from a different organism, while multiple matches revealed that the sequence was duplicated within the genome and therefore not suitable for assay design. The 30 bp flanking sequences either side of the SNP were also independently searched against the savannah elephant genomic data (cut-off e-value <0.00001 and length >27 bp) to minimize the chance of designing primers that may anneal at multiple sites. This step was added following validation of 22 probes from the pipeline (see above).

Sequence information for 115 SNP loci that passed the above criteria were submitted to LGC Genomics service labs for KASP assay design and genotyping of 74 forest elephant DNA samples that included both the samples used to generate the library and all additional samples that yielded suitable DNA (as revealed by DNA quality and quantity tests) even if they were not suitable for the ddRAD library construction. The

stringent parameters used by LGC Genomics for automatic allele calling usually result in a high proportion of unassigned genotype calls (Semagn et al., 2014). Therefore, the genotype plots of each assay were visually checked using SNPviewer 2 software (LGC Genomics) and rescored manually if individuals that clearly belonged to a cluster had not been called automatically. The proportions of manually rescored genotypes and missing data (no calls) were calculated for each locus as indices of assay quality. Genotype profiles obtained from the KASP assays were compared to the genotype data from the ddRAD pipeline to ensure that matching genotypes were recovered. We distinguished two types of mismatches: (a) category 1 – a SNP scored as heterozygote by KASP genotyping assay but homozygote by sequencing and (b) category 2 – a SNP scored as homozygote by KASP genotyping assay but heterozygote or a different homozygote by sequencing. A proportion of category 1 mismatches were to be expected because allelic dropout usually occurs during RAD sequencing (Gautier et al., 2013) and increase for low read coverage loci (Pelak et al., 2010). Category 2 mismatches were likely due to sequencing artefacts or assay design failure and these SNP loci were removed from consideration. For all converted assays, the allelic error rate, including false alleles and allelic dropout, was estimated from mismatches between the genotypes of repeated individuals. Two positive controls were genotyped seven times. In addition, 12 individuals were repeated twice using DNA extractions from both tissue and blood or saliva samples, and DNA was extracted twice independently from 8 blood samples. Preliminary measures of polymorphism and population differentiation were estimated using the dataset of 57 individuals attributed to one of the three pre-defined populations (North-East, Central and Coastal). Minor allele frequency (MAF), expected ($H_E$) and observed heterozygosity ($H_O$) were estimated for each population using the R package *adegenet* (Jombart, 2008) and overall $F_{ST}$ was calculated in the R package *pegas* (Paradis, 2010).

*Characterization of the loci*

In the absence of a reference genome for forest elephants, the selected loci were searched against the African savannah elephant *L. africana* assembly. A megablast of the 148 bp sequences containing the SNP (e-value cut-off = $10^{-40}$) was used to match the sequences to scaffolds and determine if the SNPs were located within a gene locus, and in particular within a coding region. Pairwise linkage disequilibrium was tested for using the R package *LDheatmap* (default parameters) (Shin et al., 2006).

## 2.4 Results

Approximately one third of the samples yielded DNA of sufficiently high molecular weight to attempt ddRAD library preparation. In total 17,378,607 raw sequencing reads, were generated from the 24 samples library, representing individuals from 10 locations (Table 2.1). Three individuals (LOC0044_a, LOC0225_a and LOC0394_a) had very low read numbers (<12,000) and were removed from further bioinformatic analyses at this point. Another individual (LOC201_a) was excluded because, despite exhibiting the highest read depth, it had missing data at all loci, which was likely due to pre-DNA extraction contamination of the sample (bacterial decay). The average read depth per individual for the remaining samples was 656,955 (range: 112,534-1,259,614). The data for each individual are deposited in the NCBI Short Read Archive under accession numbers SRR6371502-21. A catalogue of 31,851 tags was assembled, of which 4,749 contained exactly 1 SNP with exactly 2 alleles and 1,365 met the chosen population coverage and read depth requirements. A further 161 of these SNPs were removed from consideration because of a lack of heterozygotes and 784 were not suitable for assay design (the SNP was less than 50bp from either end of the read). This resulted in a dataset of 420 single SNP loci for 19 elephants.

**Table 2.1** Sampling locality and number of ddRAD reads generated per individual, following quality filtering and concatenation. All samples used for discovery were tissue (skin and muscle) samples, except LOCO279_d which is a duplicate blood sample used as a positive control in the library.

| Sample ID | Population | Number of reads |
|---|---|---|
| LOC0279_b | South Mulundu | 659 295 |
| LOC0279_d (positive control) | South Mulundu | 788 139 |
| LOC0049_a | Ivindo | 735 621 |
| LOC0050_b | Ivindo | 908 474 |
| LOC0051_a | Ivindo | 566 824 |
| LOC0225_a | Loango | 11 450 |
| LOC0274_a | Loango | 791 494 |
| LOC0037_a | Lopé | 1 159 937 |
| LOC0038_a | Lopé | 1 088 247 |
| LOC0088_a | Lopé | 633 191 |
| LOC0044_a | Mayumba | 128 |
| LOC0201_a | Mayumba | 2 264 818 |
| LOC0309_a | Mayumba | 501 070 |
| LOC0035_a | Minkébé | 453 030 |
| LOC0121_a | Minkébé | 112 534 |
| LOC0122_a | Minkébé | 566 704 |
| LOC0311_a | Monts de Cristal | 595 430 |
| LOC0127_a | Moukalaba Doudou | 120 598 |
| LOC0151_a | Moukalaba Doudou | 1 002 779 |
| LOC0310_a | Moukalaba Doudou | 133 832 |
| LOC0041_a | Waka | 683 264 |
| LOC0263_a | Wonga Wongue | 1 259 614 |
| LOC0394_a | Wonga Wongue | 1 095 |
| LOC0040_a | Wonga Wongue | 379 030 |

A moderate conversion rate of 68 % was achieved with the first set of 22 randomly chosen SNP loci. Fifteen KASP assays yielded scorable profiles, whereas 7 produced diffuse clusters that could not be confidently resolved into genotypes (Figure 2.2). BLAST alignment against the *L. africana* genome revealed that this could generally be explained by the likely presence of potential multiple primer binding sites in the genome.

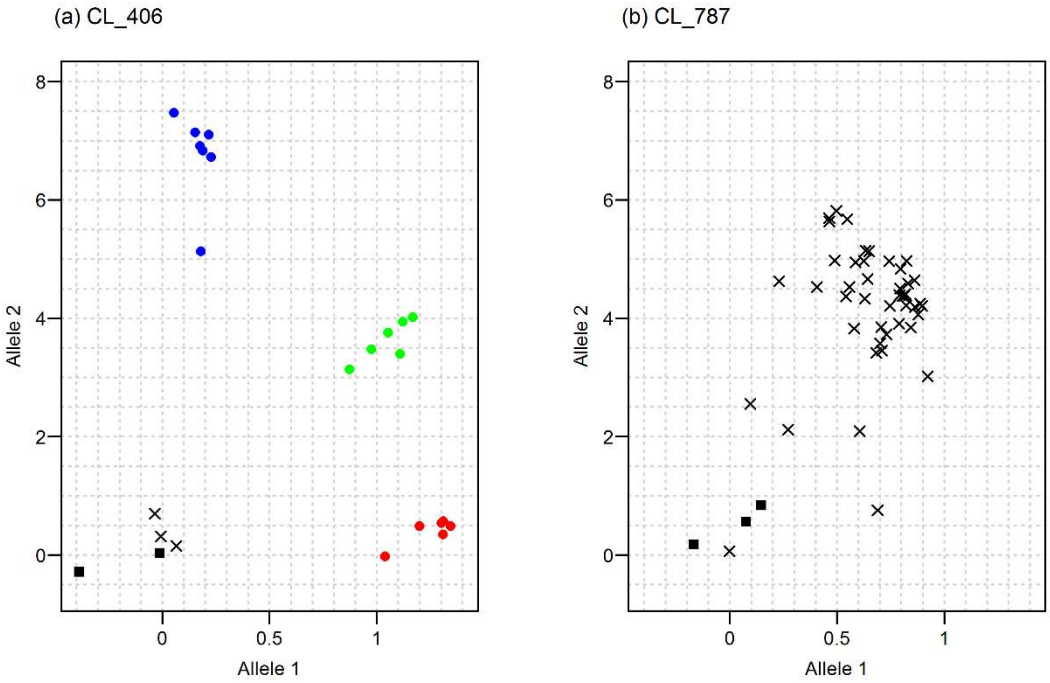


**Figure 2.2** Examples of genotype plots using validated and failed KASP assays. The fluorescence for the two alleles is plotted along the x- and y-axes. a) Samples were well separated into three clusters using assay CL_406, with the green, blue and red dots representing the heterozygous and the two homozygous genotypes, respectively, black squares are negative controls, crosses are ungenotyped samples; b) The second assay CL_787 produced a single diffuse cluster and failed to define genotypes. BLAST searches against *Loxodonta africana* genome produced a unique match for CL_406 and multiple matches for CL_787.

A further three individuals (LOC0121_a, LOC0127_a and LOC0310_a) were removed from the dataset at this stage due to having high levels of missing data in the matrix (>70%), leaving a dataset of 420 SNPs and 16 individuals with > 60% of the loci genotyped. A list of 266 highest ranking SNPs was then selected according to measures of genetic diversity and divergence (see above). A BLAST search of the whole sequence and of the flanking regions of the SNP against the *L. africana* genomic data produced no matches for 36 of these loci and multiple matches for 39 others. The search identified a unique match based on selected criteria for 191 loci, of which a random subset of 115 SNPs was subsequently chosen for KASP assay design and genotyping.

Following genotyping of 74 samples, six SNPs (CL_2059, CL_2174, CL_3260, CL_5749, CL_6220, CL_10063) failed to provide distinct clusters in the signal intensity plot and were excluded from further analysis. When comparing the genotypes obtained from the KASP assays to the 19 ddRAD profiles, the proportion of missing data was higher in the ddRAD pipeline (23.0%) than in the LGC genotyping data (1.7%). The proportion of category 1 and category 2 mismatches were 1.40% and 0.15% respectively. Only three loci yielded category 2 mismatches, of which one (CL_340) was rescored as the discrepancies were due to KASP scoring error caused by low quality plots, namely little separation between the heterozygous group and one of the homozygous groups. The two other loci (CL_3004 and CL_10172) were removed from consideration because of a high proportion of category 2 errors (9.26 and 6.82% respectively). This resulted in an estimated conversion rate of 93% (107/115).

In total 2.6% of the genotypes were manually rescored. The allelic error rate among replicates was 0.07%. The overall quality of the genotyping plots was good (i.e. clearly segregated clusters), as even though 73% of SNPs (78/107) needed to be rescored for at least one sample, only 16 were rescored for more than 5% of the samples (range: 0-17.2%). The proportion of missing genotype data per locus was <15% for all except 13 loci (overall range: 2.2-44.1) (Appendix 1). Mean MAF for individual loci was 0.213, and 30.3% of SNPs were highly polymorphic (MAF>0.3). Fifteen loci were monomorphic in at least one of the three populations. Mean overall $H_O$ and $H_E$ per locus were 0.27 and 0.31 respectively. Mean overall $F_{ST}$ was 0.015, suggesting low genetic differentiation, but ranged from 0.03 to 0.162 for 31 SNPs, indicating substantial differences in allele frequencies at these loci (Appendix 2). However, these measures are preliminary due to the small sample size.

*SNP characterization*

Following assay design, the median length of the targeted sequence, as obtained from matching forward and reverse primers to the 148 bp sequences containing the SNPs, was 54 (range: 41-104) (Figure 2.3 and Appendix 1). All 107 SNP sequences were successfully mapped to one of 60 *L. africana* unplaced scaffolds (sequence similarity from 97 to 100 %), of which 78 SNPs (71.6%) matched the same scaffold as one to five other SNPs suggesting that they could be linked (Appendix 3). However, linkage disequilibrium was not detected between most loci. Only four pairs were in

weak linkage disequilibrium ($r^2$>0.3), but the two loci in each pair didn't belong to the same scaffolds. In total, 50 sequences (46.7%) returned a match against a functional region of the *L. africana* genome, of which only 7 SNPs occurred within the coding DNA sequence of the gene (Appendix 3).

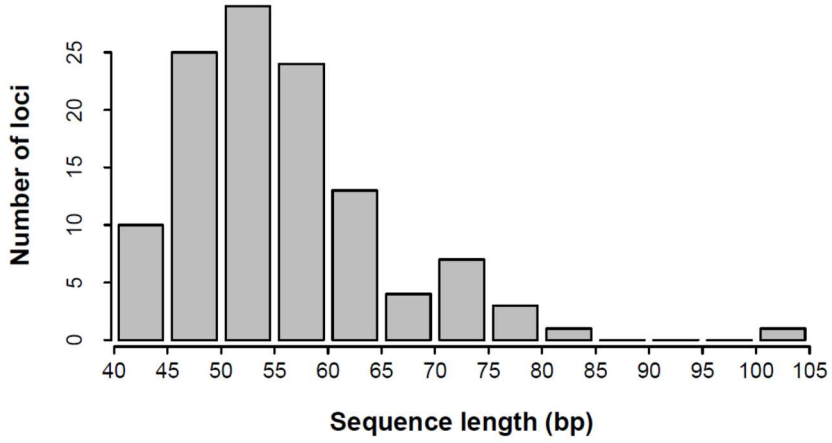


**Figure 2.3** Distribution of sequence length following assay design for the 107 validated SNPs. The median length was 54 bp and ranged from 41 to 104 bp. Only two assays targeted a sequence of more than 80 bp.

## 2.5 Discussion

After quality filtering, we have generated a new genetic resource of 1365 SNP loci which is available for further studies. As this is the first genome-wide set of SNP markers generated for African elephants, it represents a major advance for the genetic study of this taxon.

In this study, ddRAD was demonstrated to be effective for the rapid discovery of a large number of SNPs in the forest elephant. Due to double restriction digestion and precise size selection, ddRAD sequencing produces only the subset of fragments generated by cuts with both restriction enzymes and close to the target size. Therefore, ddRAD libraries are expected to provide less coverage than the original RAD method (Peterson et al., 2012). In addition, we used concatenated tags during the filtering process in order to preserve linkage information from both reads and create a high-quality dataset. This approach reduced the final number of SNPs generated compared to studies handling forward and reverse sequences separately, and was compounded by the strict first filtering criterion to allow just a single SNP per tag. As a result, the two first

filtering steps led to a sharp reduction of 85.1% in the number of loci retained. As a comparison, ddRAD sequencing and SNP filtering using restrictive criteria similar to ours generated 3,060 SNPs in koala (Kjeldsen et al., 2016) and 2,381 in an Oriental fruit bat (Chattopadhyay et al., 2016). Differences are likely linked to lower number of individuals and read depth in the forest elephant discovery panel. Both the above-mentioned studies used a large sample size (46 and 171 respectively) and reported an average of approximately 1.8 million reads per individual, which is three times higher than in our study.

A major limitation for the preparation and success of this library was the difficulty in obtaining high quality DNA samples from an endangered and elusive species. Whereas other studies used fresh blood and tissue samples, we used tissue samples obtained from carcasses of elephants poached for ivory, killed accidentally or shot during crop raiding to generate the library. Tropical environments often lead to high degradation rates of genetic material in carcasses. Thus, even though 64 samples were available at the stage of the library preparation, 41 were removed from consideration due to poor DNA quality. In order to obtain a good-quality set of SNP markers, a major component of the SNP discovery phase is to choose a panel of samples of diverse origin to minimize any ascertainment bias (Clark et al., 2005). The use of a narrow sample size from selected populations for a discovery process may result in a bias toward highly polymorphic SNPs or SNPs that segregate within particular populations, especially if population structure is pronounced (Clark et al., 2005). Our final selection of 23 samples was therefore a compromise between DNA quality and sample location across the country in order to avoid as much as possible any ascertainment bias toward particular populations whilst retaining overall sample size. However, a further four individuals were removed from consideration due to DNA degradation, as suggested by a high rate of missing data from ddRAD.

A high proportion (~70%) of the loci containing exactly one SNP were removed from consideration because of generally low read depth per individual at a locus, leading to a high rate of missing data among individuals. In retrospect, as the elephant genome is large, with a size between 3.1 and 4.01 Gb (Elephant genome project, 2009; Kasai et al., 2013), a narrower size selection or more sequencing effort might have produced better read depth per locus and resulted in more loci kept in the filtering stages. Strict filtering criteria decreases the genotyping error rate but also tends to

reduce the amount of data retained. Previous studies recommended the use of a sequence read depth of between 30-35X for accurate genotyping due to the high risk of sequencing errors, mainly allelic dropout, when the read depth decreases (Pelak et al., 2010). In a *de novo*-assembled dataset, increasing the coverage threshold from 5x to 30x decreased the frequency of genotyping errors from 0.11 to 0.04, but also led to a 13-fold decline in the number of loci detected across individuals(Fountain et al., 2016). The coverage threshold should be a balance between acceptable risk of errors and amount of data generated, in light of the objectives of the study. Our study used sequencing data to discover potential SNPs, but not for estimating some population genetic parameters, except for the purpose of selecting a reduced SNP panel. Therefore, the major challenge was not to reduce the amount of allelic dropout within the data but to avoid selecting false SNPs. The chosen threshold of 10x coverage appeared to be a sensible balance that retained about 30% of the potential SNPs while generating a low allelic error rate (1.52%). It was combined with a subsequent laboratory validation of a subset of SNPs to confirm them being real.

We validated genotyping assays for a subset of 107 SNP loci. KASP assays have been successfully used in a variety of crop and animal species (Hiremath et al., 2012; Senn et al., 2013), and they generally demonstrate high conversion rates and low error rates among replicates. The allelic error rate among replicates for the elephant SNPs was particularly low (0.07%), in contrast to the 0.7-1.6% reported for other studies using this technology (Semagn et al., 2014). Conversion rate was high, with the additional BLAST alignment check against *L. africana* genomic data improving the conversion rate from 68% to 93%. This illustrates the value of whole genome data for assisting with such studies and pointed to variation between sequence repeats found at multiple sites within the genome being probably the main factor explaining SNP conversion failure. Two SNP assays (CL_3004 and CL_10172) were removed from consideration because they did not cluster as expected genotypes. Monomorphic results were observed in the cluster plots, whereas all three genotypes existed in the ddRAD data. This was likely due to sequence repeats that were not detected using the incomplete *L. africana* genomic data. Even though ddRAD sequencing is suitable for non-model organisms, these results highlighted the advantages of using genetic resources from a closely related species to detect sequence repeats. *L. africana* genomic data have also successfully been used to characterize SNP markers in the Bornean

elephant (*Elephas maximus borneensis*) (Sharma et al., 2012) and microsatellites in the forest elephant (Gugala et al., 2016). If no related genome is available, the number of loci selected for assay design should be increased in order to take account of expected lower conversion rate.

One major challenge was to find SNPs that were appropriate for assay design, as our criterion (50 bp flanking region upstream and downstream of the target SNP) removed almost 58% of loci from consideration. A similar issue has been raised by another study that reported that as many as 75% of potential SNPs were unsuitable for assay design (Sharma et al., 2012). We followed LGC Genomics recommendations for KASP assay design, but these criteria are stricter than other genotyping platforms. A minimum of 50 bases of sequence on either side of the target SNP is required for submission for KASP assay design, similar to Illumina GoldenGate, compared with 40 bases for Applied Biosystems TaqMan assays and down to 30 bases with Sequenom iPlex assays for instance. Following assay design, the median length of the targeted sequence was as small as 54, meaning that if it was possible to relax this filtering parameter, more potentially assayable SNPs could be retained. Alternatively, using longer sequencing read technology, e.g. 250 base paired-end sequencing, would generate more SNPs with 50 bases flanking regions around the SNP position.

From a practical perspective, potential useful applications for this new set of 1365 markers include individual identification, parentage analysis, population genetics analysis and identification of the source of seized ivory. Genetic tools are particularly attractive for individual-level studies in elusive forest species. In addition, a thorough understanding of population genetic structuring of forest elephants is essential to effectively manage populations across the species range. Given the limited sample size, using $F_{ST}$ on populations of 5 to 6 individuals potentially introduced bias in SNP panel selection. However, this method was used to identify markers that might be showing population differentiation. The 107 validated SNPs will be re-assessed for utility in future population structure analysis, which may require the validation of additional loci to reach enough power. Particular attention will be paid to several of the newly developed SNP markers that were located within the coding region of genes, as markers associated with gene under selection may increase the power to detect population differentiation (Landguth and Balkenhol, 2012). Preliminary analyses of MAF and heterozygosity (Appendix 2) indicated that many of the 107 SNP markers will be useful

for individual identification and parentage analysis within Gabon. However, further investigation is needed to explore the extent of genetic variability at these new SNP markers in other forest elephant populations. Ascertainment bias is a major challenge in the widespread use of SNP panels, even though corrections have been proposed (Albrechtsen et al., 2010). The samples used in this study were widely distributed throughout Gabon, but the SNP markers developed in Gabon are expected to underestimate genetic diversity in other range countries, so they should be applied to examination of population structure with care. However, the genetic structure of forest elephant populations in Central Africa is expected to be weak (Bawe-Johnson, 2008) due to relatively high mobility of individuals, suggesting that with some further testing on populations outside of Gabon these markers may have wider use for individual ID across the species range. In contrast, preliminary testing of our 107 SNPs in two African savannah elephant samples and BLAST alignment of these alleles to the published *L. africana* assembly found only two markers to be polymorphic (data not shown), which is consistent with the species separation (Ishida et al., 2011).

## 2.6 Conclusion

We generated the first genome-wide SNP resources for forest elephants that are available for further studies. In addition, we validated KASP assays for a subset of 107 SNPs to allow in-house genotyping in local laboratories that have limited access to sequencing technologies. The use of this novel SNP panel on a wider range of samples will provide the foundation for new practical tools and in-depth information for the conservation and management of forest elephants. Given the urgency of conservation and management interventions for this species, we believe that research on the population status, genetic structure and the illegal ivory trade of forest elephants would greatly benefit from a shift towards use of SNP markers to increase potential for data sharing between researchers and allow the rapid expansion of databases in time and space required for timely response to the current crisis in this species' survival prospects.

## 2.7 Acknowledgements

Funds for the project were provided by CEEAC under the ECOFAC V Programme (Fragile ecosystems of Central Africa) funded by the European Union (contract n° FED/2013/332-349). We thank the Centre National de la Recherche Scientifique et Technologique (CENAREST; permit n°AR0016/14/MESRS/CENAREST/CG/CST/CSAR) for research authorization to conduct this study in Gabon, and the Direction Générale de la Faune et les Aires Protégées (DGFAP) for delivering CITES sample export permits (n°0161, 0210, 0215 and 0222). We thank the Agence Nationale des Parcs Nationaux (ANPN), in particular Lee White, for hosting the study and providing institutional and logistical support. We are particularly grateful to ANPN field staff (all the park rangers and wardens) and to Christopher Orbell, Philipp Henschel, Diane Savarit, Olly Griffin, Anne Meyer and Dietrich Ian Lafferty for their assistance in collecting samples. We thank Steve Blake, Sharon Deem and Odzala-Kokoua National Park (Republic of Congo), in particular Guillaume Le Flohic, for providing samples from collared elephants.

# Chapter 3

# One-step sexing tests for elephant species and their application to faecal DNA

An adapted version of this chapter is being prepared for submission with following co-authors:

Bourgeois, S.[1,2,3], K. Ouitavon[4], P. Kongmee[4], T. Veeramaethaphan[4], J. Kaden[2], R. McEwing[5]. One-step sexing tests for elephant species and their application to faecal DNA.

[1]Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom; [2]WildGenes laboratory, The Royal Zoological Society of Scotland, RZSS Edinburgh Zoo, Edinburgh, United Kingdom; [3]Agence Nationale des Parcs Nationaux, Libreville, Gabon; [4]Department of National Parks, Wildlife and Plant Conservation, Bangkok, Thailand; [5]TRACE Wildlife Forensics Network, Edinburgh, United Kingdom.

**Contributions**: S.B. and R.M. designed the study. R.M. designed the primers. S.B., J.K., P.K. and T.V. performed laboratory work. R.M. and K.O. supervised laboratory work. S.B. performed statistical analyses. S.B. wrote the manuscript. All co-authors revised the manuscript.

## 3.1 Abstract

We developed two real-time PCR assays for fast sexing in all three elephant species, which amplify small fragments of the Y-specific gene SRY (56 bp) and the orthologous sexual chromosome zinc finger protein genes ZFX/ZFY (65 bp), respectively. These two assays are simple, inexpensive and reliable tools that are suitable for non-invasive DNA samples and can be incorporated into larger SNP panels for individual identification and population genetic studies.

## 3.2 Introduction

Monitoring sex ratios is crucial as it might affect population viability. Intense male-preferred poaching for ivory is decimating elephant species throughout their range and may lead to strongly biased sex ratio (Sukumar et al., 1998). Molecular sexing is a key tool when direct sightings is impeded in forest environments (Vidya et al., 2003) or by poaching-induced elusive behaviors, and to understand the illegal ivory trade dynamics (Chelliah et al., 2013; Mondol et al., 2014).

The two genetic sexing approaches used in elephants targeted either the orthologous sexual chromosome zinc finger protein genes (ZFX/ZFY) with a restriction fragment length polymorphism (Fernando and Melnick, 2001; Munshi-South et al., 2008), or a locus specific to chromosome Y (SRY or AMELY), along with an additional control test using mitochondrial DNA or chromosome X specific regions (Ahlering et al., 2011; Gupta et al., 2006). As these techniques are based on fragment differentiation by gel electrophoresis, they require multiple steps and sufficient fragment sizes (~140-200 bp) for gel resolution of differences. There is a need for fast and reliable test that can be applied to degraded DNA samples.

Real-time PCR assays are single-step, sensitive, suitable for short fragment size (<100 bp), less prone to error and contamination, and have been successfully used in other species (Cain et al., 2013; Chang et al., 2008; Matejusová et al., 2013; Morin et al., 2005). Here we developed two real-time PCR assays for elephant DNA sexing and tested their suitability across all three elephant species and for non-invasive samples.

## 3.3 Methods

From an alignment of two published Asian elephant *Elephas maximus* and a single African savannah elephant *Loxodonta africana* (GenBank Accession: DQ078276 / AF180946 and KP141784 respectively) SRY gene regions, a 56-bp consensus sequence was observed from which a forward (ElSRY_F 5'ACTGGTATCCCAGCAGCTTGCT 3') and reverse primers (ElSRY_F 5' GCTAGAGAATCCCCAAATGCGCAA 3') were designed, as a KASP probe, with the forward primer incorporating a FAM fluorophore at the terminal nucleotide. For the ZFX/Y region, a single published example of the ZFX and ZFY genes (GenBank Accession: AF393752 and AF393751 respectively) from E. maximus were aligned and a T/G single nucleotide polymorphism targeted using a three primer KASP probe (EMZFY_F 5'ACAAAATGGTGCATAAGGAAAAGGGAT-FAM 3'/ EMZFX_F 5' CAAAATGGTGCATAAGGAAAAGGGAG-VIC 3' and EMZFR 5'CAAAATGGTGCATAAGGAAAAGGGAG 3') amplifying a 65-bp sequence.

Samples were obtained from individuals of known sex, including tissue, blood and four faecal samples from 64 African forest elephants *Loxodonta cyclotis* from Gabon, blood samples from 94 Asian elephants from Thailand, and blood samples from 2 African savannah elephants. An additional 497 faecal samples of *L. cyclotis* were collected from unknown individuals.

DNA was extracted from tissue and blood samples using the QIAGEN Blood and tissue extraction kit, Purelink Genomic DNA Mini kit or Nucleospin Blood kit, following manufacturer's instructions. DNA was extracted from faecal samples using the QIAGEN Fast Stool Mini Kit, following a modified protocol (fully described in chapter 4). Elephant DNA yield was quantified using a species-specific quantitative PCR assay (chapter 4) and non-faecal samples were normalized to ~5 ng/µl.

PCR reactions were performed by LGC Genomics, or in-house using 2-3 µl of DNA, 5 µl of KBIO High ROX mastermix and 0.1 µl primer mix on a StepOne or Quantstudio 3 real time PCR machine (Thermofisher), following manufacturer's instructions. The allelic discrimination amplification plot obtained using EMZFX/Y primers assigned individuals as male (heterozygotes T:G) or female (homozygotes T:T), while the ElSRY primers assigned individuals as males (homozygotes T:T) or not males

(no amplification) (Figure 3.1). In order to quantify genotyping errors, we used 2-4 replicates of 30 faecal samples.

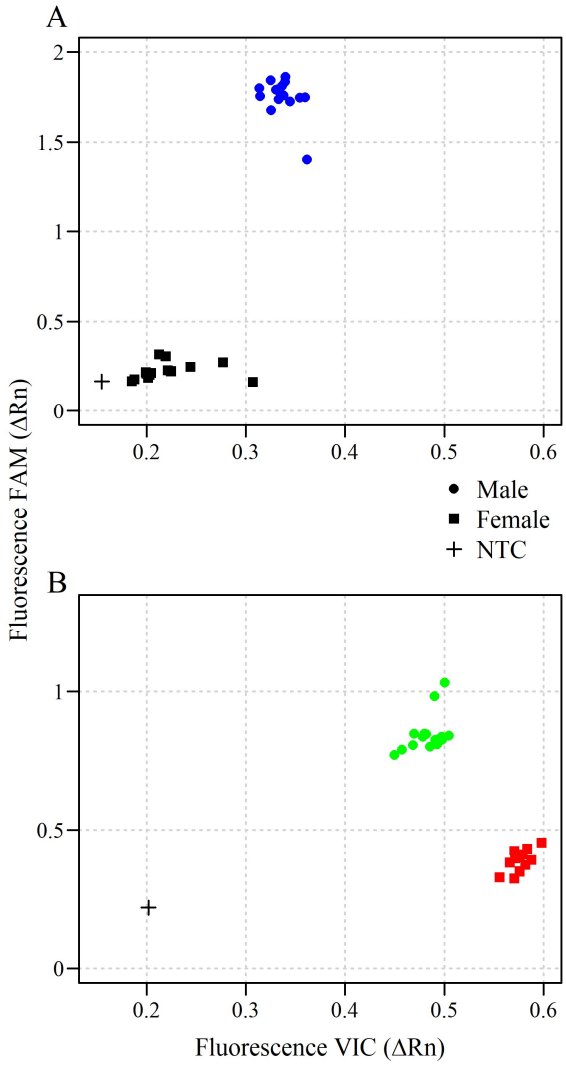


**Figure 3.1** Genotype plots of fluorescence values obtained using ElSRY (A) and EMZFX/Y assay (B). Fluorescence values are normalized with ROX (normalized reporter value ΔRn). The crosses are negative controls.

## 3.4 Results and discussion

All results from the EMZFX/ZFY test were consistent with known sexes (accuracy = 100%), while two samples failed to be identified as males by the ElSRY test (accuracy = 99.6%). In faecal samples, the genotyping error rate between duplicate samples was 0% for both tests. With the EMZFX/Y test, no allelic dropout, i.e. male sexed as female, was observed. Mean amplification success was 74.5% because we used a range of DNA concentrations in order to assess the test sensitivity. Amplification started with DNA concentrations <10 pg per reaction but 70.5 pg were required to reach a 90% probability of genotyping success (Figure 3.2). Below this threshold, 28 males (23.0%) were detected only by one of the two assays. DNA concentration did not explain preferential amplification between the two assays (Wilcoxon-test, p = 0.53).

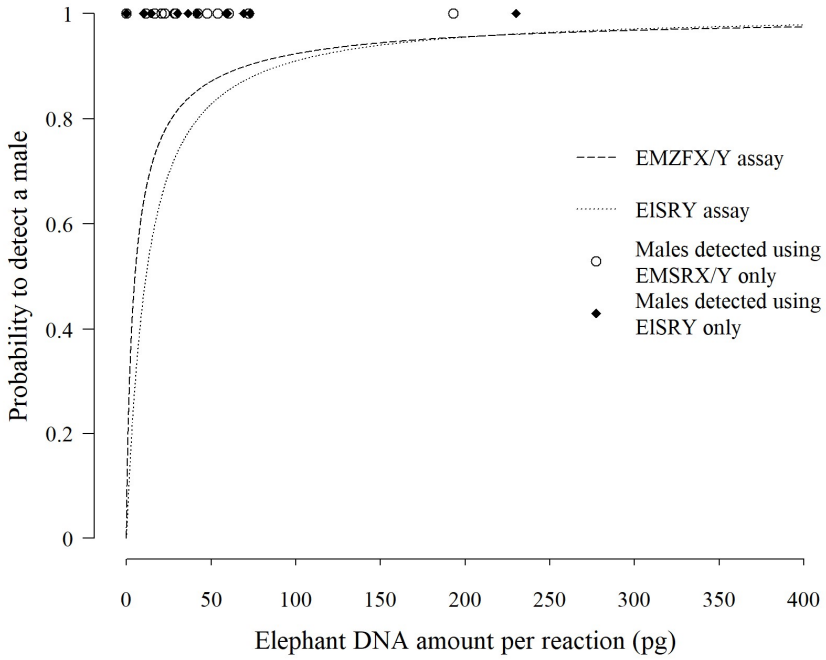


**Figure 3.2** Probability to detect a male using EMZFX/Y and ElSRY assays as a function of elephant DNA concentration. Dots represent males that were detected by one of the two tests only.

Both assays allowed for fast and highly accurate sexing of both good quality and degraded elephant DNA samples. The SRY assay does not distinguish between true females and PCR failure (type 1 error) due to low DNA yield or PCR inhibitors, both

issues being common with faecal samples. A systematic verification of all females by a second, independent test has been suggested (Robertson and Gemmell 2006). Here, the value of combining the two assays was demonstrated by the identification of a substantial number of additional males. Therefore, we recommend to combine the two assays when elephant DNA quantity is below 70.5 pg per reaction. Having been designed as standard allelic discrimination assays, these two sex determining assays have the potential to be run alongside nuclear SNP panels for population genetic studies, or as a powerful marker for individual identification in mark recapture studies or DNA registration projects in relation to African or Asian elephants.

## 3.5 Acknowledgments

We thank the Centre National de la Recherche Scientifique et Technologique (CENAREST; permit n°AR0016/14/MESRS/CENAREST/CG/CST/CSAR) for research authorization to conduct this study in Gabon. We thank the Agence Nationale des Parcs Nationaux, in particular L. J. T. White, and the Institut de Recherche en Ecologie Tropicale of CENAREST for hosting the study and providing institutional and logistical support. We thank West Midlands Safari Park for providing samples of savannah elephant collected from routine health monitoring. Funds for the project were provided by CEEAC under the ECOFAC V Programme (Fragile ecosystems of Central Africa) funded by the European Union (contract n° FED/2013/332-349) and by DEFRA, UK as part of their commitment to tackling Illegal Wildlife Trade.

# Chapter 4

# Improving cost-efficiency of faecal genotyping: new tools for elephant species

An adapted version of this chapter has been published as:

Bourgeois S.[1,2,3], J. Kaden[2], H. Senn[2], N. Bunnefeld[1], K. J. Jeffery[1,3,4], E. F. Akomo-Okoue[4], R. Ogden[5], R. McEwing[5]. 2019. Improving cost-efficiency of faecal genotyping: new tools for elephant species. PloS One 14(1).

[1]Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, United Kingdom; [2]WildGenes laboratory, The Royal Zoological Society of Scotland, RZSS Edinburgh Zoo, Edinburgh, United Kingdom; [3]Agence Nationale des Parcs Nationaux, Libreville, Gabon; [4]Institut de Recherche en Écologie Tropicale, Libreville, Gabon; [5]TRACE Wildlife Forensics Network, Edinburgh, United Kingdom.

**Contributions**: S.B., H.S. and R.M. designed the study. S.B. collected the samples. S.B. and J.K. performed laboratory work. R.M. supervised laboratory work. S.B. and H.S. performed statistical analyses. N.B. supervised statistical analyses. S.B. wrote the manuscript. H.S., J.K., N.B., K.J., K.A., R.O. and R.M. revised the manuscript. H.S., N.B., K.J., K.A., R.O. and R.M. supervised the study.

## 4.1 Abstract

Despite the critical need for non-invasive tools to improve monitoring of wildlife populations, especially for endangered and elusive species, faecal genetic sampling has not been adopted as regular practice, largely because of the associated technical challenges and cost. Substantial work needs to be undertaken to refine sample collection and preparation methods in order to improve sample set quality and provide cost-efficient tools that can effectively support wildlife management. In this study, we collected an extensive set of forest elephant (*Loxodonta cyclotis*) faecal samples throughout Gabon, Central Africa, and prepared them for genotyping using 107 single-nucleotide polymorphism assays. We developed a new quantitative polymerase chain reaction (PCR) assay targeting a 130-bp nuclear DNA fragment and demonstrated its suitability for degraded samples in all three elephant species. Using this assay to compare the efficacy of two sampling methods for faecal DNA recovery, we found that sampling the whole surface of a dung pile with a swab stored in a small tube of lysis buffer was a convenient method producing high extraction success and DNA yield. We modelled the influence of faecal quality and storage time on DNA concentration in order to provide recommendations for optimized collection and storage. The maximum storage time to ensure 75% success was two months for samples collected within 24 hours after defecation and extended to four months for samples collected within one hour. Lastly, the real-time quantitative PCR assay allowed us to predict genotyping success and pre-screen DNA samples, thus further increasing the cost-efficiency of our approach. We recommend combining the validation of an efficient sampling method, the build of in-country DNA extraction capacity for reduced storage time and the development of species-specific quantitative PCR assays in order to increase the cost-efficiency of routine non-invasive DNA analyses and expand the use of next-generation markers to non-invasive samples.

## 4.2 Introduction

Since the early 1990's, the use of non-invasive DNA analysis has evolved rapidly, allowing the study of species, individuals, gender, kinship and genetic variation (Höss et al., 1992; Morin et al., 1994), with clear ethical and practical advantages in

endangered or elusive species (Kohn and Wayne, 1997). With the decrease in laboratory costs per analysis and development of powerful analytical tools, non-invasive genetic population surveys have become increasingly accessible for wildlife management (Johnson et al., 2016; Rodgers and Janečka, 2013; Waits and Paetkau, 2005). Population censuses based on non-invasive DNA individual identification are more precise and accurate than estimates from indirect signs for a variety of elusive, low-density or wide-ranging species (Guschanski et al., 2009; Hedges et al., 2013; Mondol et al., 2009). Cost-effectiveness of non-invasive DNA surveys has also been demonstrated (Hedges et al., 2013) , but strongly relies on the ability to overcome technical challenges inherent in the use of faecal DNA samples.

The two main technical limitations of faecal sampling are the difficulty of recovering good quality DNA and the high risk of genotyping errors (Morin et al., 2001; Taberlet et al., 1999; Vigilant, 2002). Faecal samples often contain polymerase chain reaction (PCR) inhibitors and low quantities of target DNA, and are prone to DNA degradation and co-recovery of non-target DNA. All of these parameters are strongly influenced by the diet of the sampled individual (Murphy et al., 2007; Panasci et al., 2011) and the environmental conditions affecting the faecal sample in the field. In particular, DNA degrades rapidly in tropical environments due to heat, humidity and a high diversity of microorganisms (Jeffery et al., 2007; Wultsch et al., 2015).

Attempts to compensate for low DNA extraction success may include increasing the number of faecal samples collected to counteract low success rates (Rodgers and Janečka, 2013) and optimizing collection, preservation or extraction protocols (Renan et al., 2012). The choice of sampling method and storage conditions (particularly storage media, duration and temperature) strongly influences the quality and quantity of DNA that might be recovered from samples (Frantzen et al., 1998). Numerous sampling and preservation techniques have been extensively tested in a range of species with varying success (Frantzen et al., 1998; Nsubuga et al., 2004; Soto-Calderon et al., 2009), however empirical comparisons have led to a consensus that techniques targeting the outer layer of the dung are generally more efficient (Stenglein et al., 2010). Widely used storage methods include desiccation in silica beads and a variety of liquid storage media, but their efficacy for preserving genomic DNA differs across species and habitat (Waits and Paetkau, 2005). A two-step protocol consisting of a short period of storage in ethanol followed by silica desiccation has been successfully reported with ungulate

and primate samples collected from Central African rainforests (Arandjelovic et al., 2009; Nsubuga et al., 2004). In the field of human forensic science, swabs are widely used to collect touched evidence from crime scenes (Beja-Pereira et al., 2009) and have proved to be very promising for faecal sampling in a few other taxa and environments (Akomo-Okoue et al., 2015; Hayaishi and Kawamoto, 2006; Ramón-Laca et al., 2015).

Several approaches have been developed to decrease error rates associated with low quality DNA during the amplification process. For example, replicated genotyping (the multiple tube approach) became the gold standard for microsatellite genotyping to minimize allelic dropout in the 1990's (Taberlet et al., 1996), but is costly and incurs significant effort. More recently, single-nucleotide polymorphism (SNP) markers have become widely available (Helyar et al., 2011), with SNP assays less susceptible to genotyping error reducing the need to repeat analysis (Ranade et al., 2001). Because of this, they are well-suited to non-invasive samples and present a viable alternative to microsatellites (von Thaden et al., 2017). Another approach to balance cost and effort with sample size and error rate is through assessment of faecal DNA samples prior to amplification. The quantification of total DNA alone is not informative enough because faecal samples contain both host and exogeneous DNA, nor is the amplification of one robust marker (e.g. sex marker or 500bp of mitochondrial DNA) sufficient to filter poor quality samples (Fernando et al., 2003; Paetkau, 2003). Instead, species-specific quantitative PCR has been developed as a more informative approach to quantify host DNA yield in order to predict the risk of errors and provide critical thresholds for PCR and genotyping replicates (Beja-Pereira et al., 2009; Morin et al., 2001). In addition, methods to enrich host DNA from faeces have been proposed (Chiou and Bergey, 2018).

Despite these advances in molecular techniques and the variety of tools available, there is little objective evaluation on how to choose between sampling and laboratory methods (Renan et al., 2012), which precludes the spread of new tools for routine non-invasive genetic analyses. The use of swabs for faecal sampling remains anecdotal among the vast published literature on conservation genetics studies (Ramón-Laca et al., 2015), while quantitative PCR assays have been developed in only a limited number of species (Lampa et al., 2013). To date, relatively few studies have applied SNP genotyping to faecal samples (Fabbri et al., 2012; Fitak et al., 2015; Goossens et al., 2016; Norman and Spong, 2015; Schultz et al., 2018). The underuse of these new

techniques is one reason why non-invasive genetic approaches arise slowly as routine tools to support conservation management and decision-making (Arandjelovic and Vigilant, 2018; Johnson et al., 2016). Managers are still reluctant to commit resources to faecal DNA surveys because there remains the uncertainty of success in recovering enough good quality data, while a high investment into fieldwork and laboratory costs is required (Sarre and Georges, 2009). Substantial work needs to be undertaken to refine sample collection and preparation methods in order to increase the accuracy and success of routine non-invasive DNA surveys and facilitate their implementation for conservation and management.

This chapter proposes guidelines to optimize the quality of faecal DNA samples for accurate and cost-effective genotyping. We conducted a non-invasive genetic study using a panel of SNP markers and faecal samples of the endangered forest elephant (*Loxodonta cyclotis*), a relatively understudied species where non-invasive approaches are desirable due to the scarcity of direct observations in a rainforest environment. Our goals were to develop tools for all three elephant species, using an approach that can be applied to multiple taxa, as follows:

(1) Development of a quantitative PCR assay;

(2) Validation of a new convenient field sampling method with recommendations for sample storage and suitable extraction protocol;

(3) Prescreening of a faecal sample set using the quantitative PCR assay and DNA threshold determination for accurate genotyping with a panel of SNP markers.

## 4.3 Materials and methods

*Sample Collection and Storage*

We conducted fieldwork between June 2014 and January 2015 at 26 study sites in Gabon, Central Africa (Figure 4.1). Gabon is mainly covered by tropical forests and 10% of the land has been classified as National Parks. The long rainy season extends from October to April, with a variable short dry season in December and January. The long dry season extends from May to September, although variations occur within the country. Average monthly precipitation ranged from 62 mm to 420 mm. Average

monthly temperatures fluctuated between 28°C to 31°C and the mean relative humidity between 88% to 92%. Gabon hosts half of the remaining forest elephant population (~ 50,000 individuals) (Maisels et al., 2013) but faces an unprecedented poaching crisis (Poulsen et al., 2017). The study sites included both National Parks and forestry concessions believed to host high numbers of elephants.

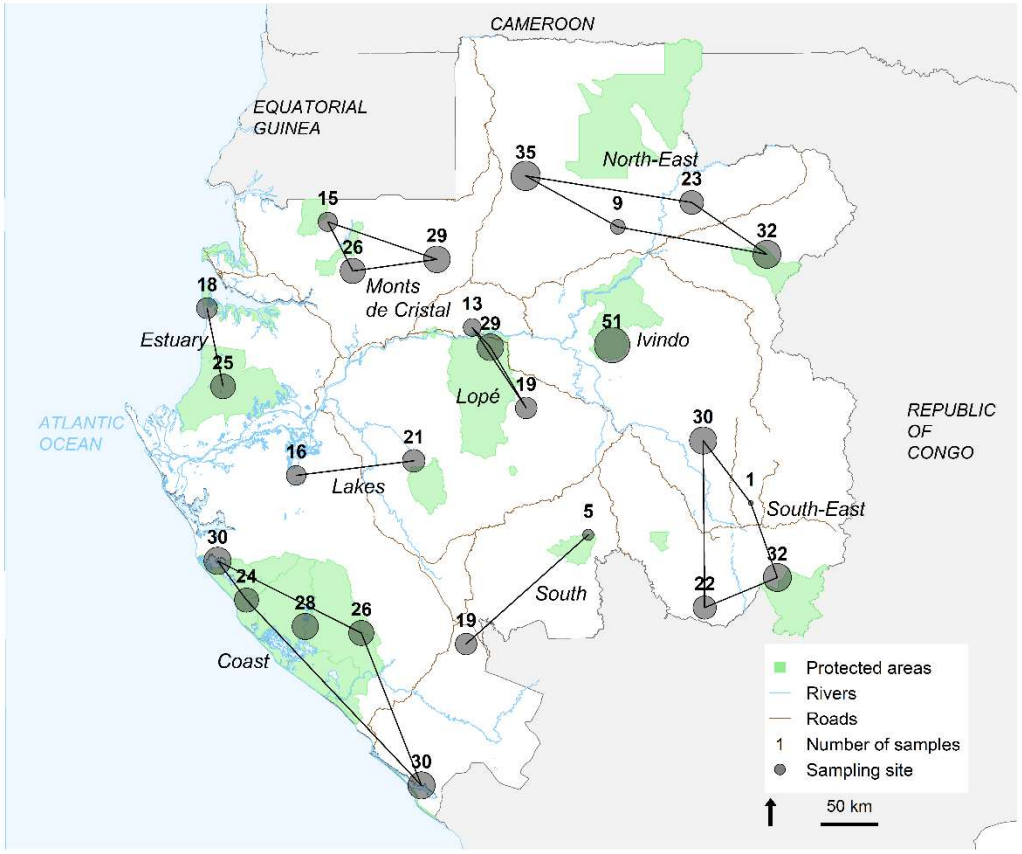


**Figure 4.1**Distribution of elephant sampling locations throughout Gabon. The circles are proportional to the number of faecal samples collected in each sample site (with the total number indicated above). Samples sites were grouped into nine sampling locations (represented by polygons).

This research was undertaken by the Gabon National Parks Agency (ANPN). We received permissions to conduct this research from the Centre National de la Recherche Scientifique et Technologique (permit AR0016/14) and the Direction Générale de la Faune et des Aires Protégées (certificate of origin 005/15). We obtained access permits from forestry concessions when applicable.

We conducted 1-2 weeks of field surveys within each study site to collect fresh elephant faeces. Faeces were considered "fresh" if they were estimated to be less than 24 hours old, were protected from sunlight by forest cover and had not been exposed to heavy rain. Fresh dung piles were characterized by a shiny colour, mostly intact boli (unless very humid or destroyed by insects) and strong odour (Schuttler et al., 2014). Presence of urine, small flies and elephant footprints in close proximity were other strong indices of freshness. A subset of the fresh faeces was reclassified as "very fresh" (i.e. < one hour old), when the elephant was directly seen or heard, and the dung pile was warm.

To evaluate the influence of dung pile quality on DNA extraction efficacy, we also collected samples from faeces that presented a "reduced surface" suitable for sampling (i.e. those classed as less than 24 hours old but partly destroyed by insects or directly exposed to sunlight), and from potentially "degraded" dung piles (i.e. those classed as between 24 and 48 hours old and those of any age that were found after rain or partly immersed in water). For the two latter categories, only the intact shiny surface was swabbed.

Faecal samples were collected using a buccal swab (Isohelix, Cell projects) previously moistened with storage buffer (500 µl of LS buffer and 25 µl of proteinase K, Stabilizing Kits, Isohelix, Cell Projects). The entire shiny, mucous surface of every bolus belonging to a dung pile was gently scrubbed with the swab to target the mucous layer coating the dung pile and care was taken to avoid collecting actual faecal material (Figure 4.2). The swab tip was then snapped and immersed into storage buffer in a labelled 2 ml light-protective Eppendorf safe-lock tube. Samples were stored at ambient temperature in the dark for 1 to 4 weeks before being transferred to the laboratory for immediate DNA extraction or storage at -20°C. As a comparison, we collected duplicate samples from a subset of 78 dung piles using a different sampling method and a two-step preservation protocol (Figure 4.2). In this method, a small piece of faeces was taken from the outer layer of a bolus and stored in 96% ethanol (20 ml) for 24 hours at ambient temperature before being transferred into silica beads (30 g) (Nsubuga et al., 2004).


**Figure 4.2** Sample collection from an elephant dung-pile using two sampling methods. Samples were collected using (A) a swab stored in lysis buffer in a 2-ml light protective tube or (B) following a two-step protocol in which a small piece of faeces is stored in ethanol in a 50-ml tube during 24 hours before being transferred into another 50-ml tube with silica beads. The swabbing material was more convenient and easy to carry in the field and allowed to scrub the entire surface of the dung pile.

*DNA Extraction*

We used the QIAamp Fast Stool Mini kit protocol (QIAGEN) to extract DNA from samples preserved using the two-step method, following the manufacturer's instructions. We modified this protocol to extract DNA from the swabbed samples, as follows: (i) the initial sample (swab tip in buffer solution) was vortexed and centrifuged for 2 minutes (14,100 g) before discarding the swab, (ii) 250 µl of Inhibitex were added to the supernatant, (iii) samples were incubated with proteinase K for 1 hour at 56°C, (iv) 500 µl of CT capture buffer (Isohelix extraction kit, Cell Project) were added to the sample (replacing ethanol), and (v) DNA was eluted in 75 µl of buffer ATE (Appendix

4). For every batch of samples, we used DNA extraction blanks to monitor contamination. All DNA extracts were purified using OneStep PCR inhibitors Removal Kits (Zymo research).

*DNA Quantification*

The concentration of elephant DNA in all samples was measured using a quantitative PCR assay. We designed primers 2804 F (5'CCTGGCAGAGCTCAGCA GAT-3') and 2804 R (5'GGATGAGGGCCAGAGTGTCC-3') using Primer3 (Rozen and Skaletsky, 2000) in Geneious version 9 (Kearse et al., 2012) to amplify a short nuclear sequence (130 bp) of the transmembrane protein 184A gene previously demonstrated to be conserved in forest elephants (Bourgeois et al., 2018). We choose the length of the targeted sequence to ensure its suitability for degraded samples and similarity to SNP amplicon size. Faecally-derived DNA samples from two African savannah (*Loxodonta africana*) and five Asian (*Elephas maximus*) elephants were included in the analysis to test for efficiency of the primers in these species. We used BLAST (Basic Local Alignment Search Tool) to confirm that the primers did not amplify human DNA.

Seven serial dilutions of DNA extracted from a forest elephant tissue sample provided standards to calibrate absolute quantification. The serial dilution ranged from 20 to 0.0013 ng/µl with a serial factor of 5. The four highest standards (20, 4, 0.8 and 0.16 ng/µl) were stored at 7°C for 48 hours to ensure homogenisation and quantified by fluorometry (using QuBIT DNA Broad Range and High Sensitivity Assay kits, Invitrogen, Thermo Fisher). The three lowest standards were freshly prepared before the experiment by serial dilution and vortexed to ensure homogenisation before the subsequent dilution. Standards and negative controls were included in duplicate in all plates. All quantitative PCR experiments were performed over a period of four days in order to minimize the variation of standards between plates and two positive controls were repeated across plates to check for variability. A subset of faecal samples were re-run in pairs of swabbed samples with duplicated two-step preserved samples over a two-day period with fresh standards. In addition, we quantified a subset of 27 samples by fluorometry (using QuBIT DNA Broad Range and High Sensitivity Assay kits, Invitrogen, Thermo Fisher) in order to compare total and elephant DNA yield.

Quantitative real-time PCR reactions were conducted in 10 µl reactions containing 1 µl of DNA, 5 µl of SYBR Green I Master mix, 1 µl of QN ROX Reference Dye (QuantiNova SYBR Green PCR Kit, Qiagen) and 0.7 µl of each primer (10 µM). To dilute inhibitors (Goossens et al., 2016), faecal samples were diluted 1 in 20 with double distilled water before the experiment. Quantitative PCR reactions were carried out on a StepOne Real-time PCR system (Applied Biosystems) with an initial holding step of 2 min at 95°C, followed by 40 cycles of 95°C for 5 s, 60°C for 10 s and a final melt curve stage gradually increasing from 60 to 95°C for 15 minutes. Standard curves were used to calculate elephant DNA concentration in the 20 x diluted samples (Morin et al., 2001). The converted concentration of the neat DNA extracts was used for further analyses, unless otherwise stated. Efficiency of the standard curves (correlation coefficient $r^2$) and melt curve profiles were examined. Any standard or sample generating non-specific amplification (i.e. PCR products that melt at temperatures above or below the desired product 84.7°C) were discarded from the analysis.

*Genotyping*

In order to assess genotyping success, samples were sent to LGC Genomics for SNP genotyping using 107 KASP assays developed and validated for forest elephants (Bourgeois et al., 2018). A pilot study was performed using four SNP assays (CL_370, CL_406, CL_2831 and CL_2968) and several dilutions (5, 10, 20, 40) of a subset of 88 samples selected over a wide range of concentrations (0 to 12.2 ng/µl). In order to determine the optimal dilution, we classified the samples into four categories based on target DNA concentration: [0-0.01), [0.01-0.1), [0.1-0.6) and ≥ 0.6 ng/µl. We estimated the mean genotyping success at four loci at each dilution factor for all categories. Based on this preliminary testing, further genotyping was performed using 10 x dilutions of all faecal samples and all samples that yielded a concentration above 0.01 ng/µl were selected for genotyping (Figure 4.3). To test if elephant DNA concentration predicted genotyping success, a random subset of samples with very low DNA yield (0-0.01 ng/µl) were also selected for genotyping. Genotype scoring was conducted by automatic allele calling (LGC Genomics). In order to control for quality, two negative controls were included in each 96-well plate and 14 samples were replicated two or three times in different plates. We assessed the allelic error rate directly as the proportion of allelic dropout and false alleles within the positive controls.

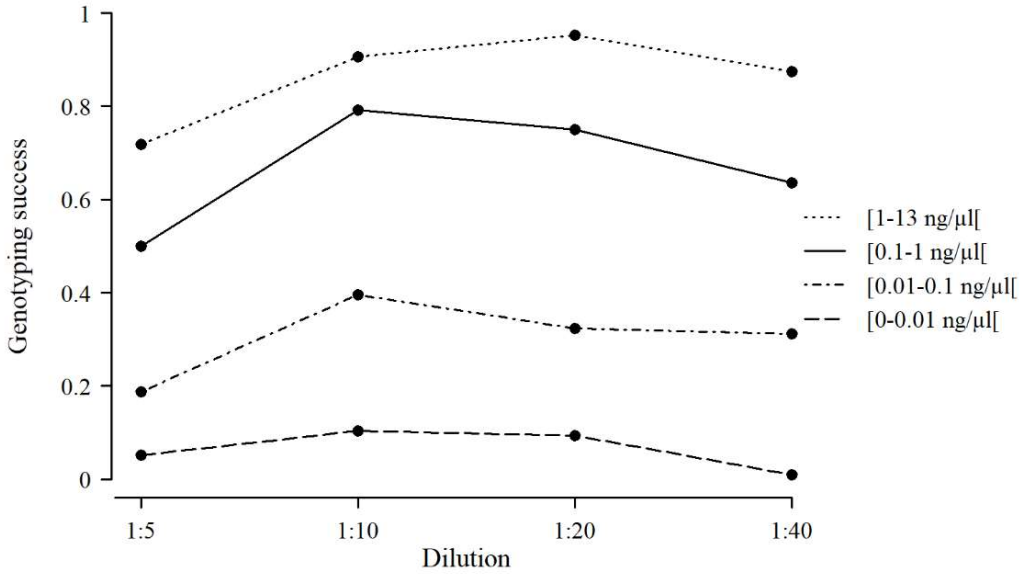


**Figure 4.3** Genotyping success at 4 SNP loci at four dilutions (1:5, 1:10, 1:20 and 1:40). The pilot study included 88 faecal DNA extracts DNA extracts classified into four categories based on target DNA concentration.

*Data Analyses*

We estimated the extraction success as the proportion of samples with a detectable elephant DNA yield using the quantitative PCR assay and the genotyping success per sample as the proportion of loci for which an unambiguous genotype was assigned. Using the subset of 78 duplicate dung samples, we compared elephant DNA concentrations from samples collected by the swab and two-step protocols using a nonparametric Mann-Whitney-Wilcoxon test. We also evaluated statistical differences across the two sampling methods within the different faecal quality groups.

We used generalized linear mixed models to test the influence of storage time and faecal quality as independent predictor variables on elephant DNA concentration. As the frequency plot suggested zero-inflation (Figure 4.4), we used a two-part model in order to investigate the influence of storage time and quality on both DNA presence and concentration (Zuur et al., 2009). In the first part, we used a binomial distribution to model the probability that a zero value is observed and we used the model to predict extraction success against storage time for different DNA qualities. In the second part, we fitted a truncated negative binomial distribution to the non-zero data to account for

over-dispersion and we used the model to test the influence of storage time and faecal quality on elephant DNA concentration.

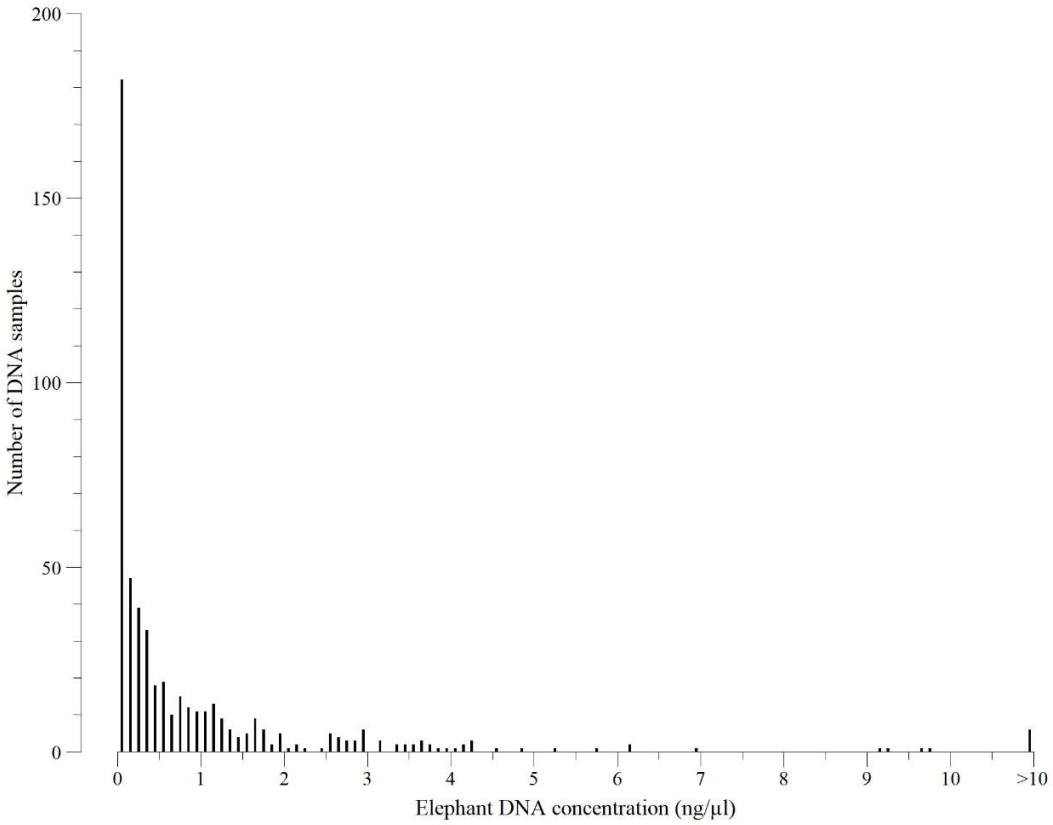


**Figure 4.4** Distribution of elephant DNA concentration in 496 fecal samples collected using a swab. DNA concentrations ranged between 0 and 28.0 ng/µl. Samples were preserved into a lysis buffed and elution volume was 75 µl for all samples.

The response variable was the absolute value of elephant DNA concentration in pg/µl. Quality types included "very fresh", "fresh", "reduced surface" and "degraded" faeces. Fresh quality was used as the reference category. We used storage time (in weeks) as a continuous variable (standardized). We also included an interaction between storage time and quality in the model to test if the influence of storage time varied with faecal quality. Storage time was highly correlated with season due to logistical constraints so we excluded the latter from the model. Study sites were grouped into nine locations when they were close (Figure 4.1) and visited at the same season. All samples from one location were collected, transported to the laboratory and extracted simultaneously as a batch. Therefore, to correct for the lack of independence between samples collected within the same location and account for other possible effects (e.g. weather, diet, habitat type, transport conditions), we treated sampling

location as a random effect. We used the Akaike Information Criterion (AIC) to compare candidate models and choose the minimal adequate model (Burham and Anderson, 2002).

We used quasi-binomial generalized linear models to examine the influence of target DNA concentration on genotyping success for different panels of 15, 50 and all 107 SNPs and determine concentration thresholds for genotyping. DNA concentrations were log transformed for statistical analyses. Panels of 15 and 50 SNPs were selected based on highest genotyping success per locus. All analyses were conducted using R version 3.3.1 (R Core Team, 2016), using the packages *lme4* (Bates et al., 2014) and *glmmADMB* (Fournier et al., 2012; Skaug et al., 2016).

## 4.4 Results

In total, 572 faecal samples, including 458 fresh dung samples were collected using the swabbing technique. Median storage time between sample collection and DNA extraction was 7.6 weeks (range: 0.7-18.9). Following quantitative PCR, all standard curves showed good accuracy ($r^2 > 0.95$). All three elephant species amplified successfully using the 2804 primers demonstrating the conserved nature of this fragment. Faecal DNA concentrations for fresh swab samples ranged from 0.0 to 26.99 ng/µl (mean= 0.97 ng/µl, n=458). The proportion of endogenous to total DNA ranged from 0.001 to 29.5% (mean=2.93%, n=27). The overall extraction success for fresh samples was 65.9% (n=458). It rose to 74.5% (n=47) for very fresh samples collected within one hour of defecation and 84.7% (n=261) for fresh samples extracted within 8 weeks. Following DNA extraction, the colour of 76 DNA eluates was brown and failure of quantitative PCR reactions indicated the presence of inhibitors. These samples were excluded from further analyses. In total, 382 samples yielded a target elephant DNA concentration above 0.01 ng/µl and were genotyped at all SNP loci, along with 121 samples that didn't reach this threshold. Following genotyping at 107 loci, the error rate was 0.0029.

*Comparison of Sampling Methods*

The elephant DNA concentration in swabbed samples was 42.9 times higher than in silica-preserved samples and the difference was statistically significant

(V=1631, $p < 0.001$) (Table 4.1). Higher target DNA concentration was also obtained with the swabbing technique in all categories of faecal quality ($p < 0.05$). Median concentration was between 29.5 (< 1 hour) and 505.4 (> 24 hours) times higher in swabbed samples than in silica gel-preserved samples. Maximum elephant DNA concentration obtained from samples preserved using the two-step method was as low as 0.47 ng/µl and only 5 samples reached the DNA concentration threshold of 0.01 ng/µl (Figure 4.5).

**Table 4.1** Results of the Mann-Whitney-Wilcoxon test on mean elephant DNA concentration (ng/µl) between two faecal DNA sampling methods. The results are based on 79 faecal samples collected in duplicates using a swab or a two-step protocol. Samples were classified into four categories based on faecal quality: very fresh, fresh, reduced surface and degraded.

| Quality | Mean two-step | Mean swab | V | p-value |
|---|---|---|---|---|
| All (n=79) | 0.011±0.544 | 0.471±1.032 | 1631 | <0.001 |
| Very fresh (n=10) | 0.056±0.145 | 1.668±2.286 | 45 | 0.009 |
| Fresh (n=33) | 0.007±0.021 | 0.478±0.713 | 290 | <0.001 |
| Reduced surface (n=22) | 0.004±0.011 | 0.165±0.226 | 150 | <0.001 |
| Degraded (n=14) | 0.000±0.000 | 0.095±0.102 | 28 | 0.022 |

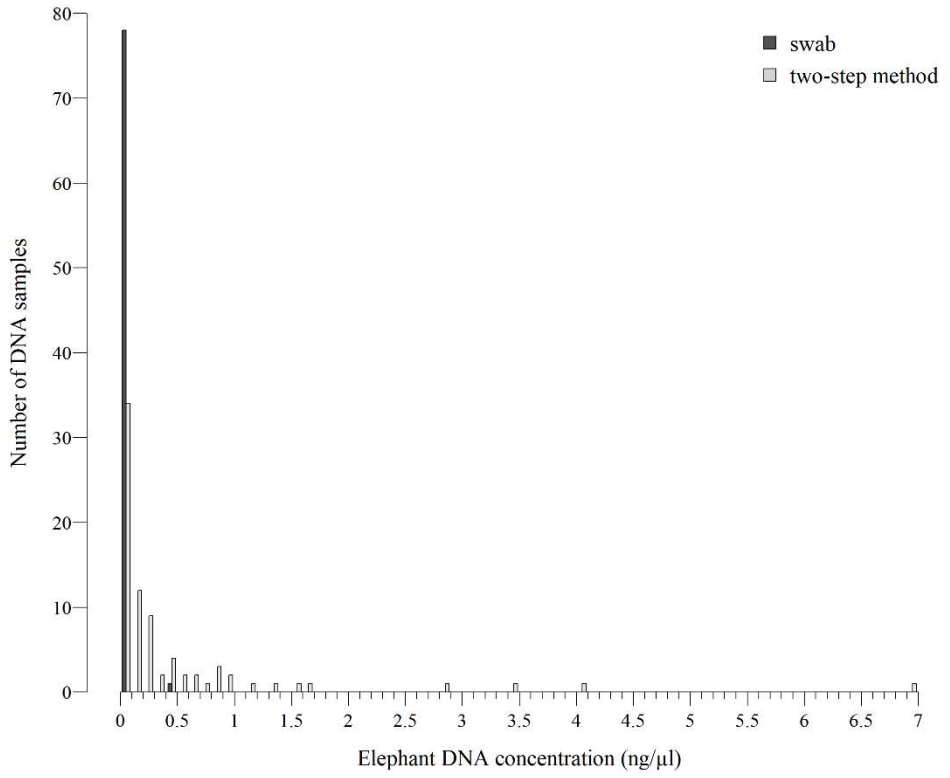


**Figure 4.5** Distribution of elephant DNA concentration in 79 faecal samples collected in duplicates using a swab or a two-step method. Elution volume was 75 µl for all samples.

*Influence of Storage Time on Target DNA Concentration*

In the binomial model explaining DNA presence, the two best models based on AIC included only storage time or both storage time and quality effects (ΔAIC<2). The interaction term did not significantly improve the model (ΔAIC=-2.1) (Table 4.2). We used the model with storage time and quality to model DNA presence because it had the lowest AIC and the "degraded" category was significantly different ($p < 0.05$) (Table 4.3). Storage time of the faecal sample had a significant influence on the probability of DNA presence in the extract ($p < 0.001$) (Table 4.3). Degraded dung piles were 2.13 times less likely to provide DNA than fresh dung piles ($p < 0.05$). The difference between very fresh, fresh and reduced surface faeces was not significant. The random effects explained 14.1% of the variance. Extraction success was 11.3% and 12.3% lower in samples collected in two of the locations (South and Coast) (Figure 4.6). The model predicted that the extraction success declined to 75% after 9.5 weeks of storage. The predicted success dropped to 50% after 19.5 weeks of storage for samples collected

from fresh faeces, against 12.6 weeks from degraded faeces and > 6 months from very fresh faeces (Figure 4.7). The prediction fitted well to observed data with the exception of a batch of 47 samples from the Coast, for which success was only 31.9% (Appendix 5).

**Table 4.2** Comparison of candidate binomial models for the prediction of elephant DNA presence in faecal samples. Variables are storage time (t) (standardized) and faecal quality categorized into four groups: very fresh (Qvf), fresh (reference category), reduced surface (Qs), degraded (Qd).

| Model | Intercept | t | Qvf | Qs | Qd | t*Qvf | t*Qs | t*Qd | AIC |
|---|---|---|---|---|---|---|---|---|---|
| Intercept-only | 1.34* | | | | | | | | 521 |
| Storage time | 1.27* | -0.62* | | | | | | | 510 |
| Faecal quality | 1.40* | | 0.58 | -0.30 | -0.71* | | | | 521 |
| Storage time +faecal quality | 1.29* | -0.65* | 0.75 | -0.12 | -0.78* | | | | 508 |
| Storage time *faecal quality | 1.29* | -0.72* | 1.36 | -0.19 | -0.75* | -0.77 | 0.32 | 0.35 | 510 |

AIC, Akaike information criterion.

*Parameter values of candidate models are marked by an asterisk if significant at the 5% level.

**Table 4.3** Summary of the best binomial generalized linear mixed model for the effects of storage time and faecal quality on elephant DNA extraction success. Faecal quality of 496 faecal DNA extracts was categorized into four groups: very fresh, fresh (reference category), reduced surface and degraded. Sampling location was included as random effect.

| **Variable** | **Coeff. (±SE)** | **Z** | **p-value** |
|---|---|---|---|
| **Fixed effects** | | | |
| Intercept | 1.288 ±0.210 | 6.133 | <0.001 |
| Storage time | -0.653 ±0.156 | -4.179 | <0.001 |
| Very fresh | 0.751 ±0.502 | 1.496 | 0.135 |
| Reduced surface | -0.124 ±0.376 | -0.331 | 0.741 |
| Degraded | -0.757 ±0.362 | -2.088 | 0.034 |
| **Random effects** | | | |
| No. groups | 9 | | |
| Variance | 0.164 | | |
| SD | 0.405 | | |

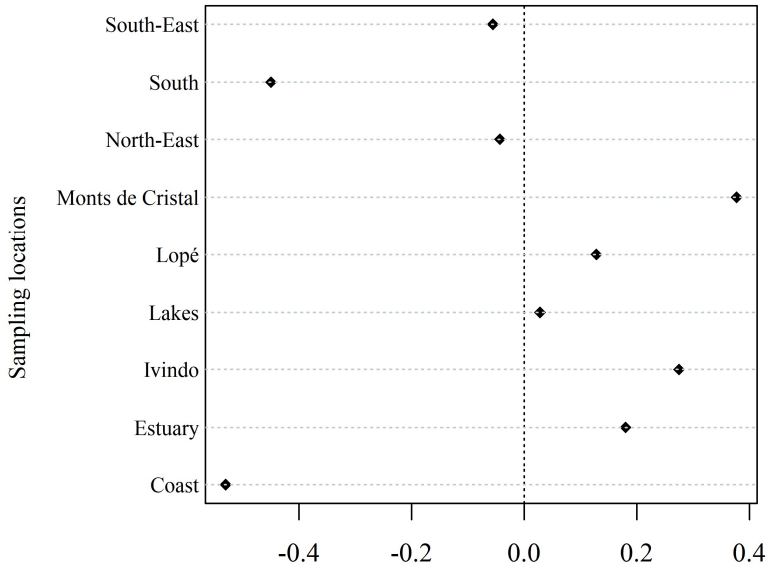


**Figure 4.6** Coefficients of random effects for the 9 sampling locations in the best binomial generalized linear mixed model for the effects of storage time and faeces quality on elephant DNA extraction success.

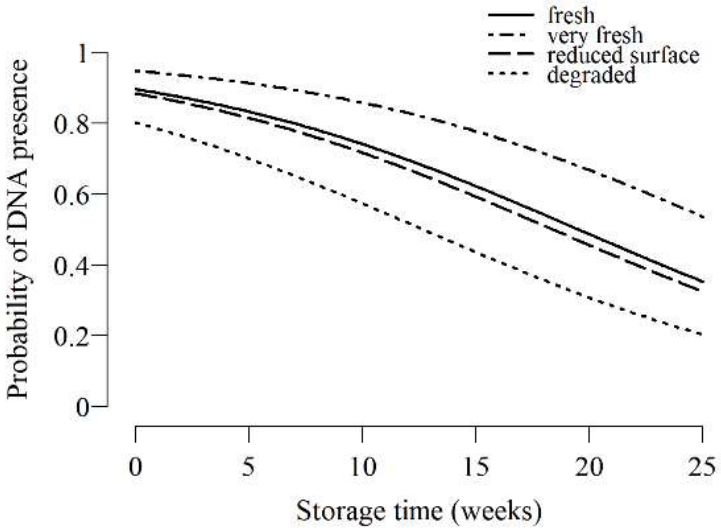


**Figure 4.7** Predicted probability to extract elephant DNA from faeces per week of storage for different faecal qualities. Faecal quality categories included: very fresh (collected within 1 hour after defecation), fresh (collected within 24 hours after defecation), reduced surface (less than 24 hours old but partly destroyed by insects or directly exposed to sunlight), and degraded (collected between 24 and 48 hours after defecation or found after rain or partly immersed in water). Details of the model are given in Table 4.3.

In the model containing data above zero, the model with the lowest AIC indicated that elephant DNA concentration was influenced by storage time, faecal quality and an interaction effect between the two variables (Table 4.4). The results were less strong than the binomial model due to small sample size in three quality categories and noise, but confirmed similar patterns to the first part of the model (results presented in Appendix 6).

**Table 4.4** Comparison of candidate truncated negative binomial models for the prediction of elephant DNA concentration in faecal samples. Variables are storage time (t) (standardized) and faecal quality categorized into four groups: very fresh (Qvf), fresh (reference category), reduced surface (Qs), degraded (Qd).

| Model | Intercept | t | Qvf | Qs | Qd | t*Qvf | t*Qs | t*Qd | AIC |
|---|---|---|---|---|---|---|---|---|---|
| Intercept-only | 6.52* | | | | | | | | 6339 |
| Storage time | 6.47* | -0.28 | | | | | | | 6327 |
| Faecal quality | 6.43* | | 1.02* | -0.59* | 0.05 | | | | 6251 |
| Storage time +faecal quality | 6.36* | -0.30* | 1.12* | -0.53* | 0.02 | | | | 6234 |
| Storage time *faecal quality | 6.43* | -0.13 | 0.96* | -0.63* | -0.18 | -0.44* | -0.23 | -0.56* | 6217 |

AIC, Akaike information criterion.

*Parameter values of candidate models are marked by an asterisk if significant at the 5% level.

*Influence of DNA Concentration on Genotyping Success*

Genotyping success was significantly correlated with elephant DNA concentration ($p < 0.001$) (Figure 4.8 A). The model predicted that a concentration of 4.65 ng/µl (698 ng per reaction) resulted in a 80% genotyping success with the panel of all 107 SNPs. Target concentration thresholds were lower for smaller SNP panels (Figure 4.8 B). Minimum concentrations of 0.285 and 0.115 ng/µl (42.8 and 17.3 pg DNA per reaction) were required to reach a genotyping success of 80% with a panel of 40 and 15 SNPs, respectively. Our threshold of 0.010 ng/µl (1.5 pg per reaction) resulted in a 49.3% genotyping success with a panel of 15 SNPs.

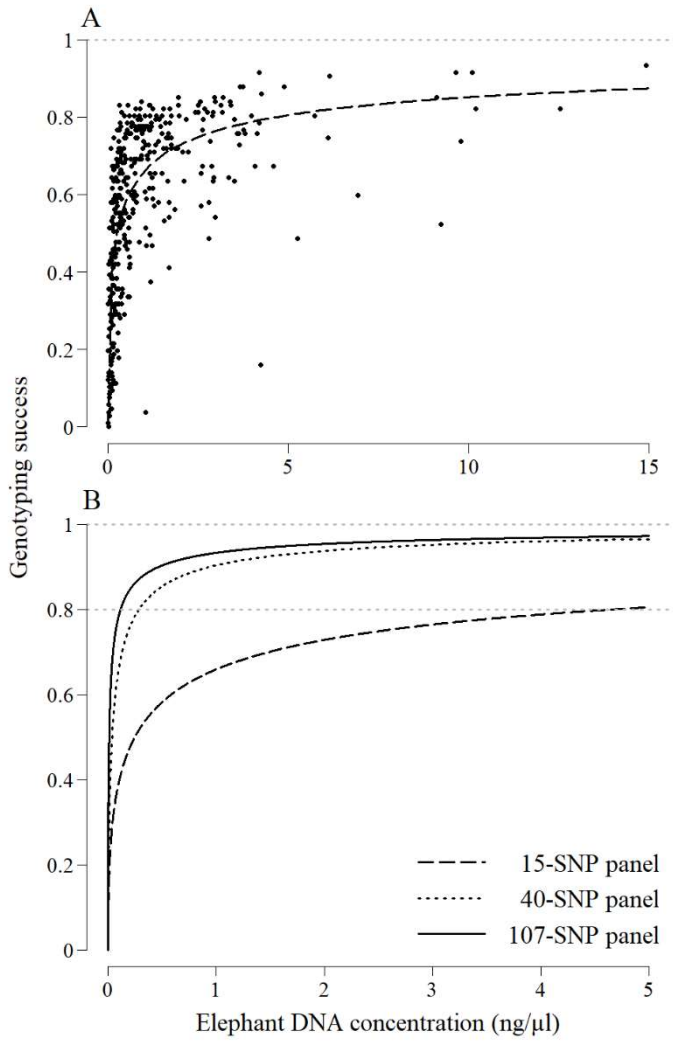


**Figure 4.8** Relationship between the genotyping success using different SNP panels and elephant DNA concentration. (A) The relationship between the genotyping success at 107 SNP loci and elephant DNA concentration measured using a real-time quantitative PCR assay was established using a dataset of 521 faecal DNA extracts (represented by points). (B) This relationship was compared to smaller panels of 15 and 40 SNPs. Genotyping was performed for each locus using 1.5 µl of a 1:10 dilution of DNA extracts.

## 4.5 Discussion

Despite the need for non-invasive tools to monitor wildlife populations, faecal genetic sampling is not routinely used as a wildlife management tool, largely because of the associated technical challenges and cost. Optimization work is required at all steps

from sample collection to DNA preparation for genotyping in order to improve cost-efficiency and dataset quality. In this study, we collected an extensive set of forest elephant faecal samples and assessed their suitability for genetic analyses. Through a newly developed quantification assay, we demonstrated the efficiency of new sampling and extraction protocols in elephants. As expected, the real-time quantitative PCR assay allowed us to predict genotyping success and pre-screen DNA samples.

*Optimizing field sampling protocols*

Choice of sampling technique and storage medium are crucial for subsequent genotyping success. Our results show that swabbing the dung surface followed by storage into a lysis buffer was an effective sampling technique, consistent with previous studies of other species (Hayaishi and Kawamoto, 2006; Lampa et al., 2008; Rutledge et al., 2009). Despite these promising results and their convenience in the field, swabs have been relatively little-used in faecal genetic sampling of wildlife (Beja-Pereira et al., 2009; Ramón-Laca et al., 2015) To our knowledge, this study is the first to report the use of swabs for faecal sampling in elephants and the observed DNA extraction success for samples collected within 24 hours of defecation was high (85% within 8 weeks of storage). This is higher than reported in other studies of forest elephants, where 60 to 80% of faecal samples stored in ethanol or in Queen's college buffer, which is recommended by the CITES MIKE (Monitoring the Illegal Killing of Elephants) programme (Hedges and Lawson, 2006), were successfully used for microsatellite genotyping (Eggert et al., 2003, 2014; Gray et al., 2014; Hedges et al., 2013; Schuttler et al., 2014).

We found that target DNA yield was more than 40 times higher in swab samples compared to samples preserved using a two-step method, irrespective of faecal quality. Based on elephant DNA concentration, we would therefore have discarded 79.5% of the samples collected using the two-step method before genotyping. Only two other studies have made a direct comparison between swabbing and other sampling techniques. Similar to our results, in equids, genotyping success was nearly zero with the two-step method and almost 100% with swabs (Renan et al., 2012). Higher target DNA yield has been reported with swabs compared to ethanol storage in several species, especially in herbivores (Ramón-Laca et al., 2015).

The high efficacy of the swabbing method demonstrated in our study might be explained by the sample collection technique. We used the swab to scrub the entire surface of the dung, thus yielding more DNA per sample than techniques targeting a small piece of the outer layer of the dung. This is especially true in species with large scats or numerous pellets providing a greater surface area (Cullingham et al., 2010). In addition, our findings illustrated the efficiency of the swabbing technique to target host cells, as the proportion of endogenous DNA was high compared to values reported in other studies using faecal samples collected with other techniques (Perry et al., 2010; Ramón-Laca et al., 2015). Swabs target sloughed intestinal epithelial cells at the surface of the dung more specifically than other collection techniques, thus reducing the simultaneous collection of diet or microbial material (Lampa et al., 2008). The relatively low proportion of endogenous DNA in our study (up to 29.5%) compared to values around 50% reported in swab samples in another study (Ramón-Laca et al., 2015) might be limited by a higher concentration of microorganisms at the surface of the dung in a tropical environment.

The choice of dung piles that are suitable for sampling is another crucial step determining the success of sample collection. We found that the extraction success of swab samples was influenced by dung pile freshness and exposure to various environmental factors, which included UV light, humidity, as well as unmeasured factors such as temperature and microorganisms, as expected from previous studies (Brinkman et al., 2010; Murphy et al., 2007; Santini et al., 2007; Wultsch et al., 2015). We showed that freshness had a major impact on elephant DNA concentration and, therefore, sample quality for DNA studies was optimal within one hour after defecation. This was contrary to previous findings in otter (*Lutra lutra*), where no variation was detected within 20 hours after defecation (Lampa et al., 2008), but likely due to the tropical environment, as degradation happens quicker than in dry or very cold environments (Murphy et al., 2007). In our study, the extraction success declined due to the humid environment (rainfall or partial immersion into water) and the age of dung sample exceeding 24 hours, which was also reported in tigers (*Panthera tigris*) (Reddy et al., 2012). In contrast, exposure to direct sunlight significantly reduced DNA concentration but not DNA presence, and these samples were suitable for genotyping. This outcome might be explained by our sampling technique, as we swabbed only the sides of the dungs that were shaded from direct UV light.

In our study, differences among locations and individuals also explained a part of the variability in both models of DNA presence and concentration. These differences could be explained by variations in diet, which is known as a factor influencing genotyping success (Panasci et al., 2011). Previous studies suggested that diet quality influences the digestion time, and thus the abrasion of intestinal cells that contain host DNA (Maudet et al., 2004), and that some plants or fruits contain PCR inhibitors (Monteiro et al., 2001; Schrader et al., 2012). In our study, fruit species and the proportion of grass in elephant faeces varied among sampling sites and seasons (S. Bourgeois, personal observation). More research is needed to help select dung piles that are most suitable for DNA studies based on elephant diet.

A major outcome of our study is the reduction in cost and effort for generating a high quality faecal genetic dataset. The swabbing material was more convenient and easy to carry, requiring minimal space in the field, thus allowing to collect more samples in one field trip, representing a 50% reduction in field man-days in remote areas. These are strong advantages for remote and difficult to access field sites, such as tropical rainforests. In addition, the high extraction success reduces the targeted number of dung samples usually necessary to compensate for analytical failure (Rodgers and Janečka, 2013), which further decreases field costs and effort by about 15%. In species with low density and/or daily defecation rate, the reduction in field costs might be limited by the difficulty to find fresh dung samples (< 24 hours). In addition, the age of dung piles might be difficult to evaluate in the field (Piggott, 2005). In these species, it may be necessary to collect older dung samples in order to increase the number of samples collected, even though this leads to an increase of laboratory costs due to a decreased extraction rate. A pilot study including dung samples of various ages would allow to set reasonable thresholds for dung age in these species, as a balance between laboratory costs and field efforts.

In the laboratory, DNA extraction from swab samples was fast and straightforward, as the tube was simply vortexed for 10 seconds and centrifuged for 2 minutes before the swab was discarded. In contrast, DNA extraction from samples collected using the two-step protocol was time consuming and involved a higher risk of contamination, due to the need to scrape or choose a piece of faeces prior to the extraction. The number of swab samples that could be extracted per day per person was 48 with the swab samples, compared to only 16 with the two-step protocol, representing

a 66% reduction in labor costs. This was similar to results from a previous study showing that DNA extraction from swabs was associated with faster processing times and allowed to work with larger batch sizes (Quasim et al., 2018).

The marked advantages of sampling fresh faecal material in terms of laboratory success should also be considered in relation to the increased effort in finding sufficient samples of this type, as opposed to more relaxed criteria for collecting faecal material in any condition. There is always a trade-off in terms of project cost between sample collection and laboratory analysis. Laboratory analysis is easier, quicker and cheaper when using reliable DNA sample sources, but while this results in a preference for invasive samples types over non-invasive samples, and fresh non-invasive faecal material over older material, lab efficiencies due to high sample quality may be offset by elevated field costs. However, the trade-off has some hard borders. Just as it is considered completely impractical (financially and ethically) to tranquilize wild forest elephants to get the best possible quality of DNA sample, it is simply not possible to perform DNA analysis on samples in which the DNA is completely degraded. As this point is approached, the cost of DNA analysis increases, but also, importantly, the quality of the resulting genetic data and its utility in biological inference decrease. This issue of data quality is often overlooked in a simple cost trade-off between lab and field expenses. We would therefore argue that higher search effort in the field to find fresh samples is actually a requirement, rather than a balanced choice, if the alternative is the collection of samples which are not only very expensive to process in the lab, but also only yield data of marginal biological value. It is important that this issue is widely understood to improve fieldwork planning and to manage expectations of wildlife managers and donors when embarking on conservation genetic projects.

*Optimizing sample preparation*

We highlighted the importance of sample preparation, including faecal sample storage before DNA extraction and DNA sample dilution prior to genotyping, by investigating the effects of storage time and dilution rate on genotyping success. We showed that storage time negatively influenced DNA extraction success and we used this relationship to provide recommendations for maximum storage time. Elephant DNA concentration also decreased with increased storage time even if there was high variability among samples. Predicted faecal DNA extraction success declined to below

75% after two months and 50% after five months. This finding was similar to other studies that show a significant reduction in genotyping success after one to three months of storage irrespective of storage medium (Murphy et al., 2002; Soto-Calderon et al., 2009). PCR success rates of 75% were obtained with DNA extracts stored for up to four months, and 50% for those stored for more than six months, when dung piles were sampled within 1 hour of deposition. This highlights the importance of selecting the freshest dung possible, although admittedly this is not always practical for elusive species. Some authors have suggested removing the cotton swab for long-term storage (Rutledge et al., 2009). Storage of samples at lower temperature, such as -80°C, might also slow DNA degradation. However, we believe that a short storage time is a key factor in the success of genetic surveys.

Careful planning for laboratory analyses prior to conducting fieldwork is paramount in order to limit storage time and increase DNA extraction success. Building in-country capacity for DNA extraction in a source country would allow to process samples as they are collected, which is especially important in studies involving a long fieldwork period where regular export of samples is impractical. The required investment in basic equipment and training is reasonable. A DNA extraction laboratory may be set up in one room equipped with a bench, a set of pipettes, a centrifuge, an incubator, a vortexer and a freezer (total cost < 6,000 USD) and ready-to-use DNA extraction kits. Training of a lab technician in DNA extractions may be possible within a couple of weeks.

Beside a low extraction success leading to absence or insufficient target DNA yield, the presence of inhibitors is the second most common cause of amplification failure in faecal samples, under validated PCR conditions (Kohn and Wayne, 1997). Our study highlighted the need to conduct a pilot study to determine the optimal dilution prior to genotyping. The pilot study showed that a 10x dilution increased the genotyping success, which was similar to a previous study in Asian elephants (Goossens et al., 2016). The optimal dilution was a compromise between the appropriate dilution of inhibitors in samples with a high DNA yield while minimizing the risk of diluting DNA in samples with a low DNA yield. However, in our study, a substantial subset of DNA eluates (14.5%) exhibited a brown colour, which is often associated with the presence of inhibitors such as humic contaminants (Matheson et al., 2010). These samples could not be quantified using the PCR assay at any dilution rate. As the

provenance of these samples were concentrated in 5 sites (2 sites in the Estuary, 1 site along the Coast, Lopé and Lakes), we believe this was due to variations in diet and not to the sampling method. The swabbing technique was rather found to minimize PCR inhibitors (Ramón-Laca et al., 2015). Future research should be directed to improve extraction protocols, in particular purification steps in order to optimize the removal of inhibitors (Costa et al., 2017).

*Prescreening DNA samples prior to genotyping*

We found that despite optimized sample collection, preservation and extraction protocols, the quality and quantity of DNA extracted from dung piles varied greatly across samples. Therefore, a prior assessment of samples was needed to increase the overall genotyping success and decrease the risk of errors. Target DNA yield was a good predictor of genotyping success, as shown in previous studies (Arandjelovic et al., 2009; Campbell and Narum, 2009; Hausknecht et al., 2010; Morin et al., 2001). When the species of origin is difficult to confirm by visual examination of the dung (e.g. in carnivores), prior identification of the species is required and often involves mitochondrial DNA sequencing (Cullingham et al., 2010; Wultsch et al., 2015). A two-step approach starting with mitochondrial DNA sequencing to inform the subsequent choice of an appropriate species-specific quantitative PCR assay, may be a cost-effective technique for prescreening the samples based on concentration. This would require thorough testing of primer specificity to ensure they do not amplify DNA from related species (Kanthaswamy et al., 2012; Ng et al., 2014). Alternatively, a single-step option would be to differentiate among multiple species (e.g. carnivores) by combining carnivore-generic PCR primers with species discriminatory melt-curve analysis in a single qPCR assay.

By simulating two different reduced panels of 15 and 40 SNPs using high quality loci and numbers of loci commonly used for individual identification or parentage analyses (Kidd et al., 2006; Pakstis et al., 2010), we showed that the relationship between target DNA concentration and genotyping success varied across number of markers and individual loci. Our approach was conservative, as we didn't rescore the genotype plot manually. This would have increased the genotyping success, because automatic allele calling results in a high proportion of unassigned genotype calls (Semagn et al., 2014). Despite this, we found that very low amounts of DNA per

reaction was required to achieve 80% genotyping success (22.5 or 45 pg DNA per reaction with a panel of 15 or 40 SNPs, respectively). These values were lower than cut-offs reported for microsatellite and SNP genotyping (50-200 pg per reaction) in previous studies (Arandjelovic et al., 2009; Cullingham et al., 2010; Morin et al., 2001; Nussberger et al., 2014). Differences in thresholds between studies are explained by variation in the type of markers (Campbell and Narum, 2009), choice of genotyping assay (Fabbri et al., 2012) and species of interest (Hausknecht et al., 2010). As a consequence, thresholds for sample categorization need to be established on a case-by-case basis for each species and set of markers. Our study re-emphasized the need to conduct a pilot study (Lampa et al., 2013; Taberlet et al., 1999) in order to set reasonable thresholds. A pilot study would also allow estimating the proportion of bad quality samples and decide if they should be discarded or genotyped in replicates, as a balance between costs and the need of a suitable sample size.

The use of quantitative PCR has long been limited by equipment and reagent costs (Lampa et al., 2013) but this technique is now affordable (Beja-Pereira et al., 2009). Discarding samples unlikely to produce viable results before genotyping reduces the genotyping costs and the risk of errors. DNA quantification (reagents and plates) costs approximately US$ 1 per sample, excluding labour costs, and allowed us to reduce overall genotyping costs by more than a third. The need for a prescreening of samples is even higher in species where the age of dung piles is difficult to evaluate in the field (Piggott, 2005), thus leading to a higher proportion of unsuitable samples. In other studies, up to 50-60% of non-invasive samples have been discarded based on target DNA quantification (Ball et al., 2007; Ebert et al., 2012; Hausknecht et al., 2010). The cost reduction is even greater when compared to the multi-tube approach advocated for microsatellite studies (Taberlet et al., 1999), where the recommended seven replications for homozygous loci is often prohibitively expensive. For example, a quantitative PCR assay was used to reduce the number of replications required for accurate genotyping down to 2 for samples above DNA quantity thresholds (Morin et al., 2001). We believe that the development and use of species-specific quantification assays would strongly increase the cost-efficiency of faecal DNA surveys.

## 4.6 Conclusions

We demonstrated the efficiency of our tools in generating a good quality faecal DNA dataset in elephants. Therefore, we recommend the collection of faecal DNA samples within 24 hours of defecation for elephant species using a swab preserved in lysis buffer. DNA extraction should be performed as soon as possible after collection or within two months to ensure 75 % extraction success. The use of the quantitative PCR assay, that was validated in all three elephant species, to pre-screen the DNA samples is valuable to reduce the cost of genotyping.

The same approach might be used by managers to improve the cost-efficiency of routine faecal DNA surveys in a wide variety of species. In order to optimize the quality of faecal DNA samples from the field to the laboratory for accurate and cost-effective genotyping, we recommend to:

(1) Validate an efficient and convenient sampling technique in the species and environment of interest. We strongly recommend testing the swabbing technique and expect that its use will rise in future studies of elephants and other species;

(2) Perform DNA extraction as soon as possible after sample collection to ensure suitable DNA yield. In many cases, the development of an in-country capacity for DNA extraction would be instrumental in reducing storage time;

(3) Conduct a pilot study to assess optimal dilution to minimize the effects of inhibitors and determine a threshold for successful and accurate genotyping using a chosen set of markers;

(4) Quantify target DNA in all samples and discard poor quality samples before genotyping.

We believe this approach will help managers widely embrace faecal DNA surveys and contribute to a shift towards the field of genomics using faecal DNA.

## 4.7 Acknowledgments

We thank the Centre National de la Recherche Scientifique et Technologique (CENAREST; permit n°AR0016/14/MESRS/CENAREST/CG/CST/CSAR) for

research authorization to conduct this study in Gabon, and the Direction Générale de la Faune et des Aires Protégées (DGFAP) for delivering CITES sample export permit (n°0161). We thank the Agence Nationale des Parcs Nationaux (ANPN), in particular L. J. T. White, and the Institut de Recherche en Ecologie Tropicale of CENAREST for hosting the study and providing institutional and logistical support. We are particularly grateful to ANPN field staff (all the park rangers and wardens), in particular J.-P. Mondjo, for their assistance in the field. Forestry concessions (SEEF, Rougier, CEB, Toujours Vert, Foreex), OELO, SFM, WCS, SEGC and PPG provided logistical support. C. Orbell and L. Makaga collected additional samples. We are grateful to Dr K. Ouithavon from the Wildlife Forensic Science Unit, Department of National Parks, Wildlife and Plant Conservation, Thailand for assisting with the analysis of Asian elephant samples. We thank West Midlands Safari Park for providing samples of savannah elephants collected from routine health monitoring. We are grateful to M. Sigaud and J. Cusack for providing useful comments and contributions to the manuscript. Funds for the project were provided by CEEAC under the ECOFAC V Programme (Fragile ecosystems of Central Africa) funded by the European Union (contract n° FED/2013/332-349).

# Chapter 5

# Geographic assignment of ivory in an important source country

An adapted version of this chapter is being prepared for submission with following co-authors:

Bourgeois S.[1,2,3], H. Senn[2], R. McEwing[4], K. Abernethy[1,5], K. J. Jeffery[1,3,5], N. Bunnefeld[1], R. Ogden[4,6].

[1]Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, United Kingdom; [2]WildGenes laboratory, The Royal Zoological Society of Scotland, RZSS Edinburgh Zoo, Edinburgh, United Kingdom; [3]Agence Nationale des Parcs Nationaux, Libreville, Gabon; [5]TRACE Wildlife Forensics Network, Edinburgh, United Kingdom. [4]Institut de Recherche en Écologie Tropicale, Libreville, Gabon; [6]Royal (Dick) School of Veterinary Studies & The Roslin Institute, University of Edinburgh, United Kingdom;

**Contributions**: S.B., H.S., R.M. and R.O. designed the study. S.B. collected the samples and performed laboratory work. R.M. supervised laboratory work. S.B., H.S. and R.O. performed statistical analyses. S.B. wrote the manuscript. H.S., R.M., K.A., K.J.J., N.B., and R.O. revised the manuscript. H.S., R.M., K.A., K.J.J., N.B., and R.O. supervised the study.

## 5.1 Abstract

The illegal ivory trade is the main conservation threat to the endangered forest elephant. It is urgent for the survival of the species that hunting pressure is reduced in Gabon, which hosts 50% of the remaining global population. However, existing tools have failed to detect poaching events in real-time or provide critical data to wildlife authorities for timely law enforcement responses and investigations. Reliably and rapidly tracing the exact origin, and trade routes, of illegally hunted Gabonese ivory will enable a huge advance in the effectiveness of national and international protection measures. Here, we selected a set of informative nuclear SNP markers and combined them with mitochondrial DNA polymorphism information gathered from an intensive reference sample collection. We used this to develop a new test that can assign individual elephants to protected areas or groups of protected areas. This novel test will produce real-time, reliable and shareable data to support fine-scale investigation of the illegal ivory trade at the intra-national level and foster collaboration between countries involved all along the chain.

## 5.2 Introduction

Illegal ivory trade decimates elephant populations worldwide (Wittemyer et al., 2014). African forest (*Loxodonta cyclotis*) and savannah elephant (*Loxodonta africana*) populations have declined by an estimated 62% (162,000 individuals) between 2002 and 2011 (Maisels et al., 2013) and 30% (144,000 individuals) between 2007 and 2014 (Chase et al., 2016), respectively. Yet the international ivory trade appears to be booming. The weight of ivory seized globally (~40 tons in 2016) has increased three-fold since 2007 and a record number of 22 large scale ivory seizures (>500 kg) was reported in 2016 (Convention on International Trade in Endangered Species, 2017), followed by a record seizure of more than 7 tons of ivory in 2017 (TRAFFIC, 2017). While understanding the source of large seizures is key to identifying patterns of national and international trade routes and ultimately the poaching hotspots on the ground and, it is not straightforward as most ivory is covertly traded through organized criminal networks (Underwood et al., 2013). Ivory seizures in the original source

countries represent only a small part of global seizures (Nkoke et al., 2017) and are thus not useful for estimating the importance of a country as a global trade source.

On-the-ground monitoring of poaching is extremely difficult in most elephant ranges, but particularly in Central African dense rainforest environments (Blake et al., 2007). Rangers patrol large and poorly accessible areas on foot and most elephant carcasses remain undetected. Other approaches are needed for early warnings of population declines such as periodic population census (Hedges, 2012) and investigation of the source of illegal products (Manel et al., 2002). Region-wide surveys of central African forest elephants show that they have suffered steep population declines over the past 15 years (Maisels et al., 2013). Investigation of the national origin of traded ivory could elucidate the distribution of this poaching pressure more rapidly than on the ground surveys in such difficult terrain. Genetic assignment of samples to their population of origin has been proposed as a novel approach to monitor illegal wildlife trade. This approach is based on interpopulation differences in allele frequencies distribution. (Manel et al., 2005; Ogden and Linacre, 2015). Thus, we tested the use of DNA to rapidly identify the origin of ivory samples, with the aim of expediting national efforts to regulate poaching and reduce the level of illegal ivory entering the global black markets.

DNA techniques pioneered the geographical assignment of ivory in 2004 (Mailand and Wasser, 2007; Wasser et al., 2004) with the successful use of a panel of microsatellite markers to investigate the geographic origin of several major seizures at the African continent scale (Wasser et al., 2008, 2015). This assignment test has greatly improved the understanding of global patterns of the illegal ivory trade, but is not informative enough for country-based law enforcement actions. Improved assignment methods now reach a precision of about 350 km for forest elephants (Wasser et al., 2015), which is too large a geographic area for effective law enforcement responses, usually encompassing land in different management categories (i.e. conflating logging concessions, National Parks and agricultural lands), and often crossing national boundaries. Finer-scale information is crucial for national wildlife law enforcement agencies to guide antipoaching activities and better understand national illegal trade routes. Moreover, despite a CITES decision urging countries to determine the geographic origin of large ivory seizures of more than 500 kg (Convention on International Trade in Endangered Species, 2014), only 10% of large seizures made

between 2011 and 2014 were genetically analyzed (Convention on International Trade in Endangered Species, 2017).

One reason is the difficulty and reluctance of countries involved in the illegal ivory trade chain to export samples abroad. Samples from endangered species are controlled by CITES requirements, which can further complicate and slow the international movement of seizure samples for testing and reference samples required for method development. Laboratory techniques using microsatellites offer limited opportunity for data sharing as an alternative to sample exchange, due to the need for calibration of allele sizes among laboratories. There is an urgent need to develop tools that can be transferred to build identical analytical capacity in source, transit and consumer countries in order to foster direct collaboration and reliable, trusted data sharing between them, for timely law enforcement response (McEwing and Ahlers, 2016).

Mitochondrial DNA control region (D-loop) sequencing has been proposed as a complementary approach to microsatellites in order to improve the success of genetic assignment of ivory, because of its high variability and distinct geographic signal in comparison to nuclear DNA, due to female matrilocality (Ishida et al., 2013). Importantly, mitochondrial DNA nucleotide sequence data are much more amenable to inter-laboratory transfer than microsatellites as there is no requirement for calibration among instruments. However, their use has been limited to the identification of locality-specific haplotypes with relatively limited reference data per range country preventing comparison of haplotype frequencies (Ishida et al., 2013). Single nucleotide polymorphism (SNP) markers have more recently shown themselves to be promising tools to detect population structure even when genetic differentiation is low (Glover et al., 2010; Morin et al., 2009). The use of markers under selection further increases the power to detect population differentiation (Landguth and Balkenhol, 2012). This is crucial for the development of geographical assignment tests whose accuracy relies on the existence of an underlying population genetic structure and its detectability at the target scale (Jones and Wang, 2012). Other advantages of SNP markers include low error rates and technical portability and reproducibility across laboratories (Ogden, 2011; Seeb et al., 2011). Various sizes of SNP panels (13 – 3,000) have been successfully used for geographic assignment in marine fishes (Nielsen et al., 2012), lobsters (Benestan et al., 2015) and bears (Puckett and Eggert, 2016). The recent

discovery of SNP markers for timber trees (Degen et al., 2017) and forest elephants (Bourgeois et al., 2018) offers potential to increase the resolution of geographical assignment, which has valuable implications for investigating illegal wildlife trade in Central Africa. In this paper, we will use a novel combination of a panel of SNP markers and mitochondrial DNA sequences in order to investigate forest elephant genetic structure for geographical assignment of ivory within an important Central African source country, Gabon.

Gabon is home to half of the remaining forest elephants (~50,000) (Maisels et al., 2013), but is considered to be one of the major sources of illegal ivory in Africa (Underwood et al., 2013; Wasser et al., 2015). In particular, the North-East of the country is experiencing a dramatic poaching crisis, with an estimated 2,500 individuals killed per year over the past 10 years in the Minkébé region (Poulsen et al., 2017). With a mean weight of 4.6 kg of ivory per elephant in that area (Gabon National Park Agency, unpublished data), this represents nearly one ton of illegal ivory covertly exported out of Gabon per month, presumably mostly through forest trails and rivers across the border to Cameroon (Nkoke et al., 2017). Reliably and rapidly tracing the exact origin, and trade routes, of illegally hunted Gabonese ivory is urgently needed to support national protection efforts.

We expect the population of elephants in Gabon to be poorly structured, because forest elephants are a long-lived species and their habitat is quasi-continuous in Gabon despite recent on-ongoing development of extractive industries (Laporte et al., 2007). This was supported by previous investigations of forest elephant population structure using microsatellites markers that detected only weak genetic differentiation across Central Africa (Bawe-Johnson, 2008; Ishida et al., 2018). The objective of this study was to maximise geographic assignment power across multiple markers in forest elephants for traceability to land management areas' level within Gabon in order to support national anti-poaching actions and law enforcement investigations. By generating a large reference database throughout Gabon and combining newly developed and potentially powerful SNP markers (chapter 2) with hypervariable mitochondrial DNA control region sequences, we aimed to develop a new tool that would increase the resolution of geographic assignment of ivory for the purposes of wildlife law enforcement.

## 5.3 Materials and methods

*Sample collection and preparation*

We collected forest elephant faecal samples at 24 sampling sites throughout Gabon between June 2014 and February 2015. We also collected tissue samples opportunistically from elephant carcasses and during collaring operations in Gabon (Blake et al., 2008) and the adjacent Odzala-Kokoua National Park in Republic of Congo. The study sites were selected to include all National Parks and forestry concessions believed to host high numbers of elephants (Maisels et al., 2013) in order to create a comprehensive reference database (Figure 2.1). Within the National Parks, we searched for forest elephant faeces along trails created by elephants through the rainforest and around suspected elephant hotspots, such as forest clearings, swamps and rivers, and fruiting areas (Blake and Inkamba-Nkulu, 2004). Outside National Parks, we searched for elephant faeces in areas that are known to attract elephants such as forestry roads and around crops. To avoid re-sampling the same individuals more than once, we spent only 2-8 days in each sampling site and used reconnaissance walks between itinerant camps. When we found multiple dung piles from a group of elephants, we collected only one sample per group to minimize chances of sampling closely related individuals.

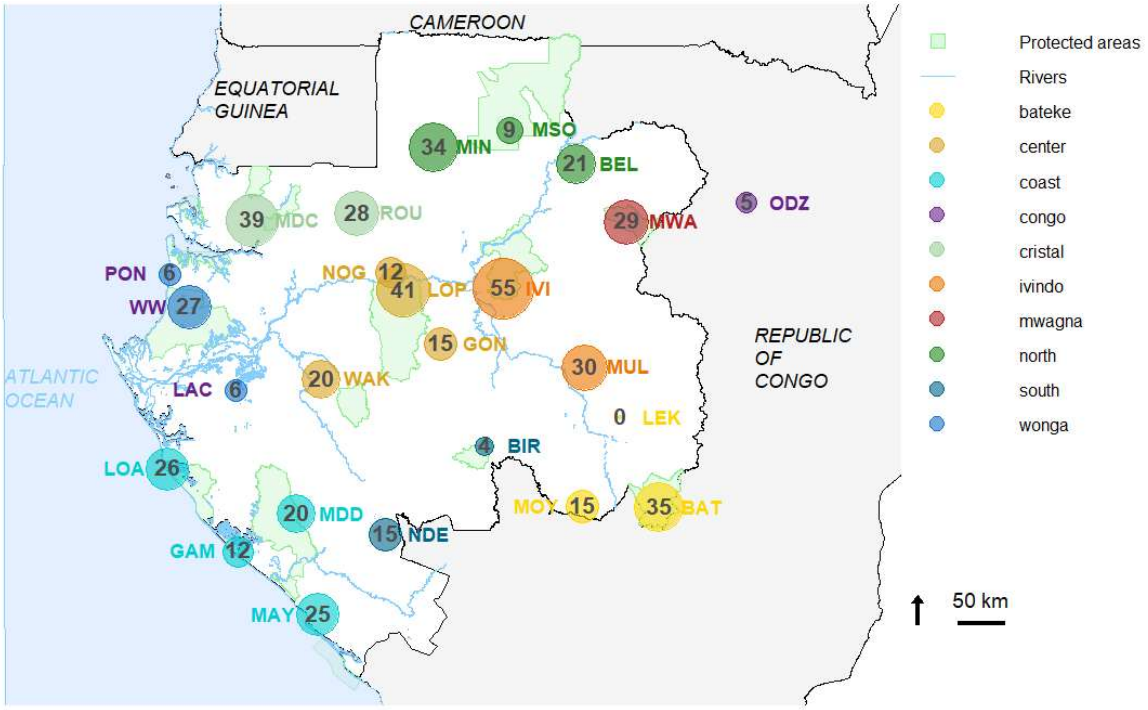


**Figure 5.1** Distribution of 24 sampling sites throughout Gabon and one sampling site in Republic of Congo. The circles are proportional to the number of elephants sampled (with the total number indicated inside). Sampling sites were grouped into ten locations (represented by different colours) based on low $F_{ST}$ values between sites and land management units. Our sampling sites were as follows: BAT, Batéké; BEL, Belinga; GAM, Gamba; GON, Gongue; IVI, Ivindo; LAC, Lac Oguemoue; LOA, Loango; LOP, Lopé; MAY, Mayumba; MDC, Monts de Cristal; MDD, Moukalaba Doudou; MIN, Minkébé West; MOY, Moyabi; MSO, Minkébé South; MUL, South Mulundu; MWA, Mwagna; NDE, Ndéndé; NOG, North Ogooue; ODZ, Odzala; PON, Pongara; ROU, Haut Abanga; WAK, Waka; WW, Wonga Wongue.

We collected fresh faecal samples (approximately less than 24 hours after defecation) by gently scrubbing the whole surface of all boli with a single buccal swab (Isohelix, Cell Projects Ltd., Harrietsham, UK) which was immediately placed into storage buffer (500 µl of LS buffer and 25 µl of proteinase K, Isohelix Stabilizing Kits) in a 2 ml light-protective tube (following chapter 4). We extracted DNA from faecal samples using the Isohelix DNA Isolation or QIAGEN (Hilden, Germany) Fast Stool Mini kits following a modified protocol (detailed in chapter 4). We purified all faecal DNA extracts using OneStep PCR Inhibitor Removal Kits (Zymo research, Irvine, CA, USA) and quantified them using a species-specific quantitative PCR assay (chapter 4).

We stored tissue samples (skin, muscle) in 96% ethanol before DNA extraction, which was performed using the QIAGEN Blood and Tissue Kit, following manufacturer's instructions.

*Mitochondrial DNA sequencing and analyses*

We amplified samples using elephant-specific primers MDL3 and MDL5 targeting a 630 bp fragment of the mitochondrial DNA control region (D-Loop) (Fernando et al., 2000). We performed PCR reactions with 7 µl of Maxima HotStart mastermix (Thermo Fisher Scientific, Waltham, USA), 1 µl of each primer at 10 µM and 1 µl of DNA. We diluted faecal samples 1 in 20 before PCR to dilute inhibitors (Schrader et al., 2012). We further diluted 1 in 50 all samples that failed to amplify at first attempt. PCR conditions consisted of an initial activation step of 95°C for 5 min, followed by 35 cycles of 30 sec denaturation at 95°C, 30 sec annealing at 60°C and 1 min elongation at 72 °C, with a final extension of 10 min at 72 °C. We purified PCR products using 1 µl of a 1:1 Exo1/FastAP solution (Thermo Fisher Scientific) and sequenced them in the forward direction on an ABI3730 capillary sequencer (Edinburgh Genomics GenePool facility, Edinburgh, UK) using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific), following manufacturer's instructions. To control for dataset quality, we amplified and sequenced 250 samples 2-10 times and re-sequenced a subset of 173 samples in the reverse direction. We visually checked and edited all sequences in MEGA version 7 (Kumar et al., 2016). We then aligned sequences in Geneious version 10.2.2 (Kearse et al., 2012) and manually verified and edited them where necessary.

We trimmed sequences to a short fragment of 316 bp that was homologous to sequences previously published in Gabon (Debruyne, 2005; Eggert et al., 2014; Ishida et al., 2013; Johnson et al., 2007; Munshi-South, 2011; Schuttler et al., 2014) and deposited in Genbank. We examined the haplotypes and their geographic distribution using a median joining network constructed using PopArt software (Leigh and Bryant, 2015).

*SNP genotyping and population differentiation analyses*

We selected all samples above a threshold of 0.01 ng/µl for further analyses (following chapter 4) and sent them to LGC Genomics for genotyping using a panel of

107 SNP markers (chapter 2) and two sex markers (chapter 3). In order to assess allelic dropout rate and control for quality, 14 faecal samples and 2 tissue samples were repeated 2 to 6 times across different plates. We only included loci with < 40 % missingness and samples with < 50 % missingness in the analyses. We also excluded monomorphic (minor allele frequency < 0.01) (Roesti et al., 2012) and dimorphic (fixed homozygotes) SNPs from further analyses. We estimated minor allele frequency (MAF), expected ($H_E$) and observed heterozygosity ($H_O$) for each locus using the R package *adegenet* (Jombart, 2008). We tested loci for deviations from Hardy-Weinberg equilibrium using the Chi-square test in the R package *pegas* (Paradis, 2010).

We calculated pairwise $F_{ST}$ between pairs of sampling sites and bootstrapped confidence intervals (95%) to test for significance (i.e. significantly greater from zero) using the package R *hierfstat* (Goudet, 2005) in order to quantify genetic differentiation. We then grouped the sampling sites based on low $F_{ST}$ values ($F_{ST}$ <0.05) between sites. We distinguished between protected areas with varying types and levels of threats (ANPN management plans 2015, Gabon National Elephant Action Plan 2018) to further divide the inferred groups into ten locations (Figure 5.1) that were used for further analyses. We first investigated genetic differentiation across Gabon using a Bayesian clustering approach implemented in STRUCTURE v. 2.3.4 (Pritchard et al., 2000), using an admixture model with information on geographic location (Hubisz et al., 2009). We performed 6 runs for each value of K=1-6 with a burn-in period of 100 000 repetitions followed by 1 000 000 Markov chain Monte Carlo iterations.

We used multivariate analyses to identify groups of genetically similar individuals without any assumption about the underlying genetic model and implemented in R using the package *adegenet* (Jombart, 2008). First, we ran a standard principal component analysis (PCA) which uses no sampling locality information. We then implemented a discriminant analysis of principal components (DAPC), which highlights the genetic separation between groups, while minimizing variation within groups (Jombart et al., 2010). We used information on sampling location to define groups *a priori*. We also used a spatially explicit multivariate method, the spatial principal components analysis (sPCA) which takes into account the spatial auto-correlation of genetic variability in order to distinguish global and local patterns. This approach has been developed as a tool to investigate cryptic spatial genetic structuring (Jombart et al., 2008). We implemented sPCA using the neighbourhood by distance

algorithm to form the connection network. We set up maximum distance between neighbours was to 110 km, which is the maximum distance between two known locations of the same individual in forest elephants (Eggert et al., 2014). We tested the inferred patterns using sPCA global and local tests (Montano and Jombart, 2017).

*Population assignment*

We calculated global and pairwise genetic differentiation ($F_{ST}$) between pairs of locations for each locus using the R package hierfstat (Goudet, 2005). SNPs were ranked according to global and pairwise $F_{ST}$ values and subset of loci with the highest values were selected for geographic assignment tests. We calculated the probability of assigning individual genotypes to each of the locations using the Bayesian assignment method (Rannala and Mountain, 1997) implemented in the GENECLASS2 program (Piry et al., 2004). We used the leave-one-out test to assign one sample at a time using the rest of the samples as reference populations. We performed assignment tests both at the regional scale, using populations inferred from Bayesian and multivariate population differentiation analyses, and all or a subset of SNPs selected based on highest global $F_{ST}$ values. Then, we ran assignment tests at the local scale, using the ten locations defined above that are meaningful for management authorities and all or a subset of SNPs selected based on highest pairwise $F_S$ values between all ten locations.

## 5.4 Results

We collected 607 faecal and 63 tissue samples (range: 1-55 individuals per sampling location). A total of 553 individuals were successfully sequenced for the 316 bp fragment of the mitochondrial DNA control region. A total of 402 faecal DNA samples had a concentration above 0.01 ng/µl and were selected for SNP genotyping at 107 SNP loci. Two sampling locations were excluded from SNP (Léké and Birougou) and mitochondrial (Léké) analyses because the sample size was zero. Three individuals were identified as recaptures based on identical genotypes and thus removed from the dataset.

*Mitochondrial DNA*

Using the 316 bp fragment, 27 variable sites and 37 haplotypes were identified of which 9 have not been published before. The median joining network showed two distinct haplogroups that differed by at least 6 bp (Figure 5.2) and overlapped the West-central and North-central subclades previously defined in African elephants (Ishida et al 2012). The number of haplotypes per sampling location varied between 1 (Ndéndé) and 10 (Minkébé West and Monts de Cristal). A high number of haplotypes per sampling location (8-10) were found in the Centre, Northern and Eastern parts of the country. The two most internal haplotypes in the median-joining network (Gab01, Gab36) were found in 30.7% of individuals and widespread throughout Gabon (Figure 5.3 A). By contrast, 23 haplotypes were specific to one or two neighbouring sampling sites and found in 10.0% of individuals (Figure 5.3 B). The other 11 haplotypes were found in 59.2% of individuals. They were shared between groups of 3 to 11 sampling sites but showed restricted geographic distribution (Figure 5.3 C). The geographically restricted haplotypes that were found in neighbouring sampling locations grouped together in the median joining network, thus revealing five groups of haplotypes corresponding to partly overlapping geographical zones within Gabon (Figure 5.2).

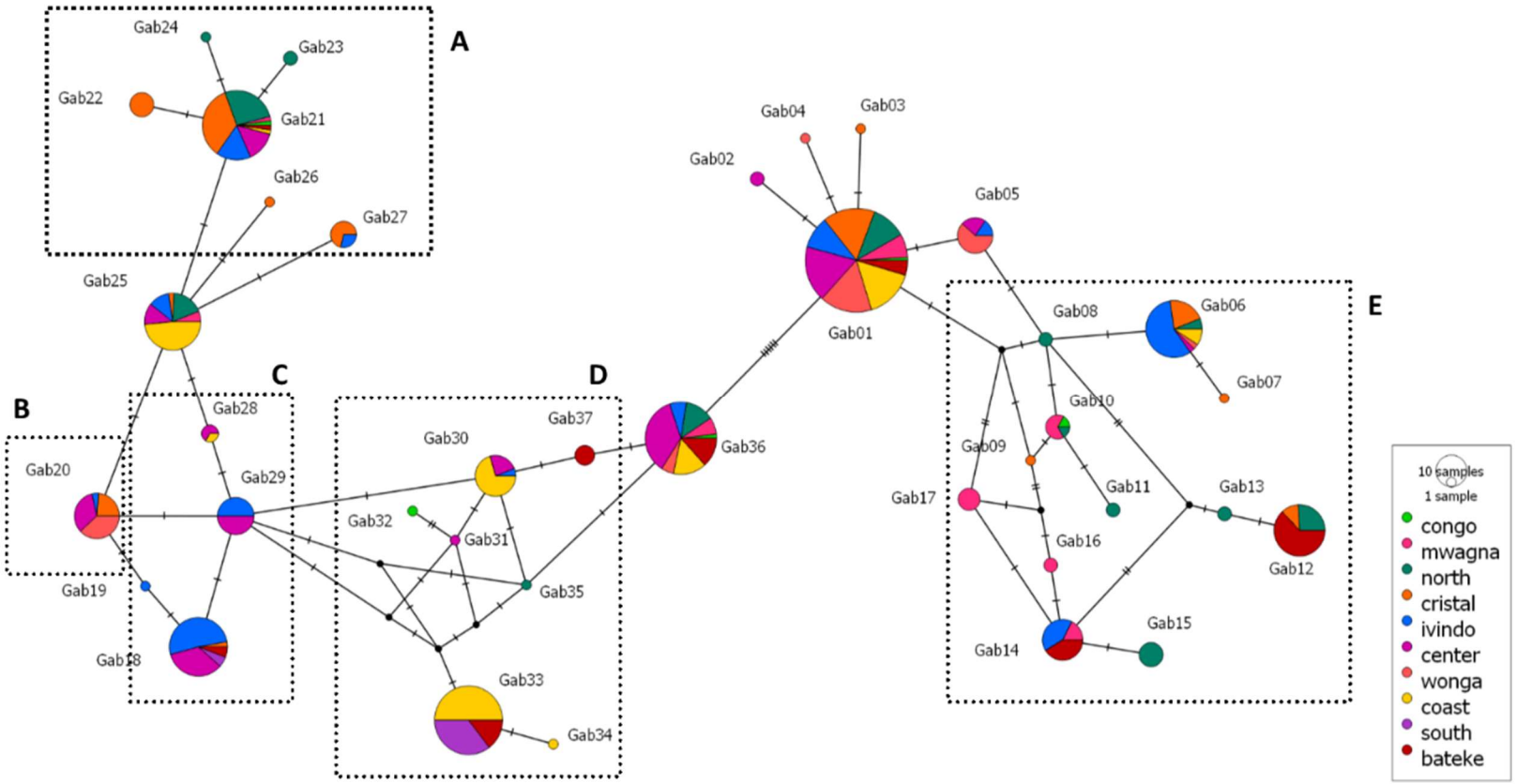


**Figure 5.2** Median joining networks constructed using 316-bp sequences of the mitochondrial DNA control region (D-loop). Dots are sized proportionally to the frequency of each haplotype in our dataset of 553 individuals and coloured according to sampling locations. Haplotypes can be grouped into 5 geographically restricted areas, as follows: A, Northern to North-Eastern Gabon; B, Wonga to Centre; C, Centre to Ivindo; D, South-Western to South-Eastern Gabon; and -E, North-Eastern to South-Eastern Gabon.

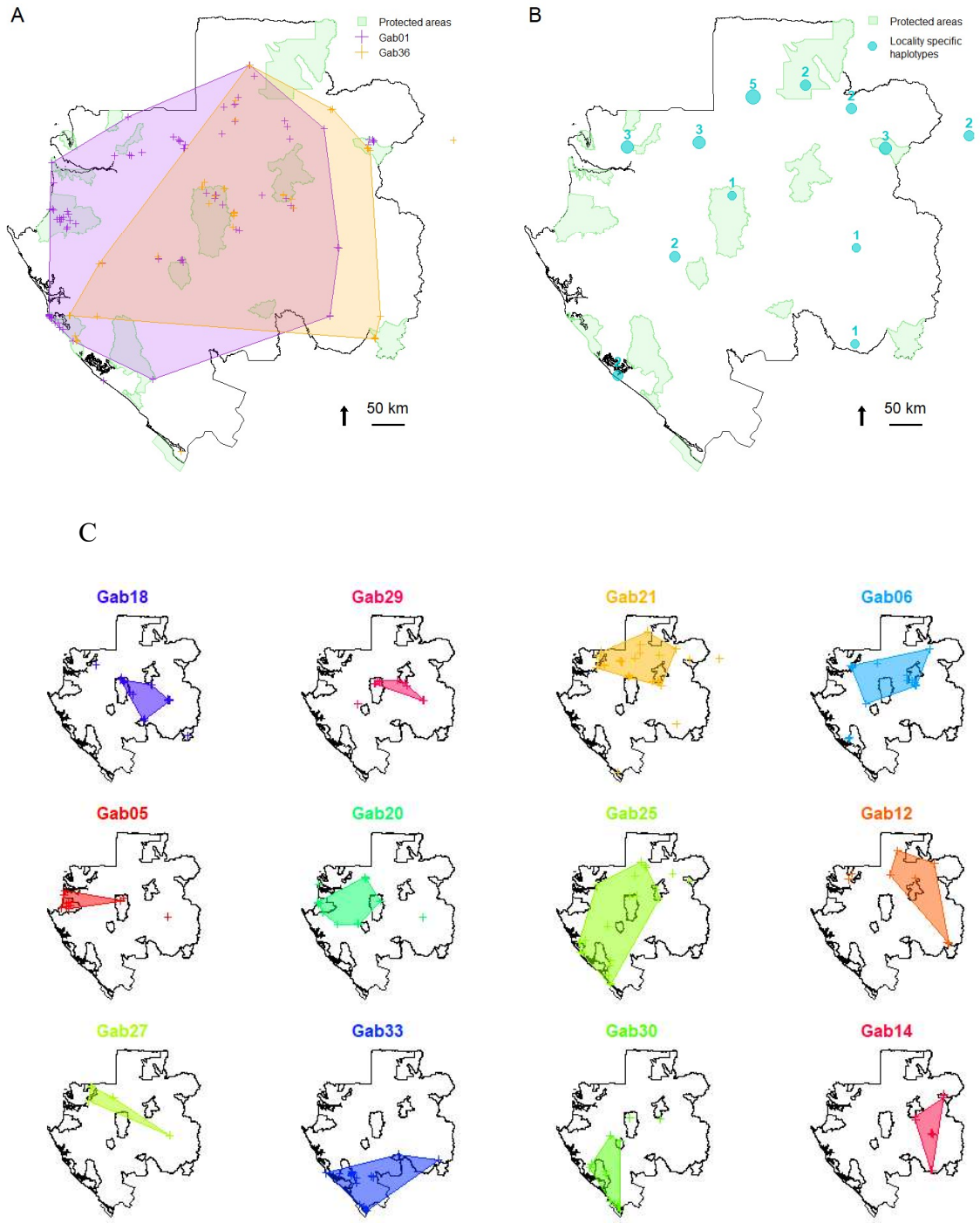


**Figure 5.3** Geographical distribution of 37 forest elephant mitochondrial DNA haplotypes across Gabon. (A) Distribution of samples carrying two haplotypes that were widely shared across Gabon. (B) Distribution of 23 haplotypes that were locality specific (shared between one or two neighbouring sampling sites). The circles are proportional to the number of locality specific haplotypes found in each sampling site. (C) Distribution of samples carrying one of 12 haplotypes that were geographically restricted within Gabon.

*SNP analyses*

Following genotyping of 467 individuals at 107 SNP loci, the genotyping error rate was estimated to be 0.475%. In total 21 loci and 100 samples were discarded from further analyses due to high proportions of missing data. In total, 86 loci were retained for further analyses. Mean overall $H_O$ and $H_E$ per locus were 0.202 ±0.133 and 0.245 ±0.160, respectively. MAF ranged from 0.008 to 0.500 (mean=0.171). Most loci evidenced Hardy-Weinberg equilibrium, except thirty-two loci that showed a heterozygote deficit (p-value < 0.01).

Pairwise $F_{ST}$ values indicated low to moderate genetic differentiation between sampling sites (average = 0.0460, range 0.0000 – 0.1957) and 68.0% were significantly greater than zero based on bootstrapped confidence intervals. Eight sampling locations (MSO, MIN, ROU, MDC, IVI, MUL, LOP and WAK) clustered together as shown by low $F_{ST}$ values between them ($F_{ST}$ < 0.03) and higher $F_{ST}$ values with other sampling locations (Figure 5.4). In order to fulfil management needs for higher resolution, this group was further divided into four locations (North, Cristal, Ivindo, Center). Six additional groups were defined based on geographical location and genetic differentiation, of which three were unique sampling sites (ODZ, MWA and NDE). Notably, Mwagna (mean $F_{ST}$ = 0.0413) and Belinga (mean $F_{ST}$ = 0.0463) significantly differed from other sampling sites in Northern and Eastern Gabon (except BEL-MSO). Gongue showed significant moderate to high differentiation (mean $F_{ST}$ = 0.0940) from other sampling sites, including the very close site of Lopé. Although genetically distinguishable from other sites, the four sampling sites clustering along the coast (LOA, GAM, MDD, MAY) still exhibited low to moderate differentiation between them. By contrast Ivindo (mean $F_{ST}$ = 0.0226) and the adjacent South Minkébé (mean $F_{ST}$ = 0.0144) were characterized by low pairwise $F_{ST}$ values with other sampling locations.

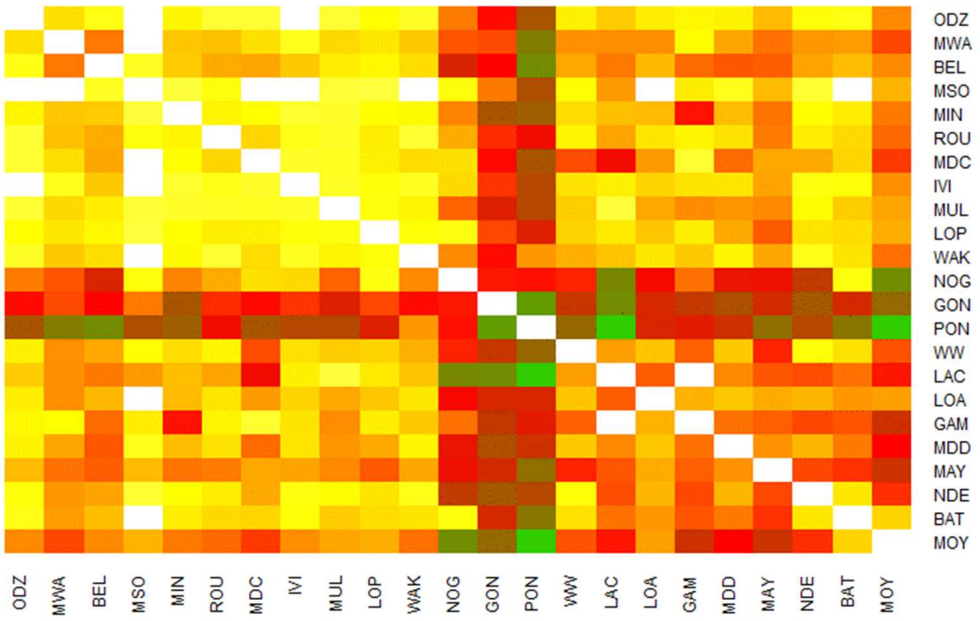


**Figure 5.4** Heatmap based on pairwise $F_{ST}$ among the 24 elephant sampling sites across Gabon. The heatmap is coloured according to three groups of $F_{ST}$ values: $F_{ST} < 0.05$ (yellow), $0.05 < F_{ST} < 0.10$ (red) and $F_{ST} > 0.10$ (green). Colour intensity increases gradually within each group according to $F_{ST}$ value between pairs of locations.

The STRUCTURE analyses suggested two distinct groups as supported by STRUCTURE Harvester (K=2). Batéké and Coastal clusters differed from a main cluster pooling all other sampling locations, even though partitioning was incomplete (Figure 5.5). PCA analyses involving all individuals without prior information on sampling location did not reveal any strong pattern (Appendix 8). The first principal component revealed slight differentiation of two locations (Coast and Centre) but explained only 2.9% of the variance. A third location (Congo) separated along the second principal component axis accounting for 2.7% of the variance. DAPC showed weak genetic differentiation between locations, with samples from the Coast separating along the first principal component which accounted for 25.7% of the variance. Four locations (Wonga, South, Congo and Batéké) were differentiated by the second principal component explaining 16.9% of the variance, although groups partly overlapped (Figure 5.6).

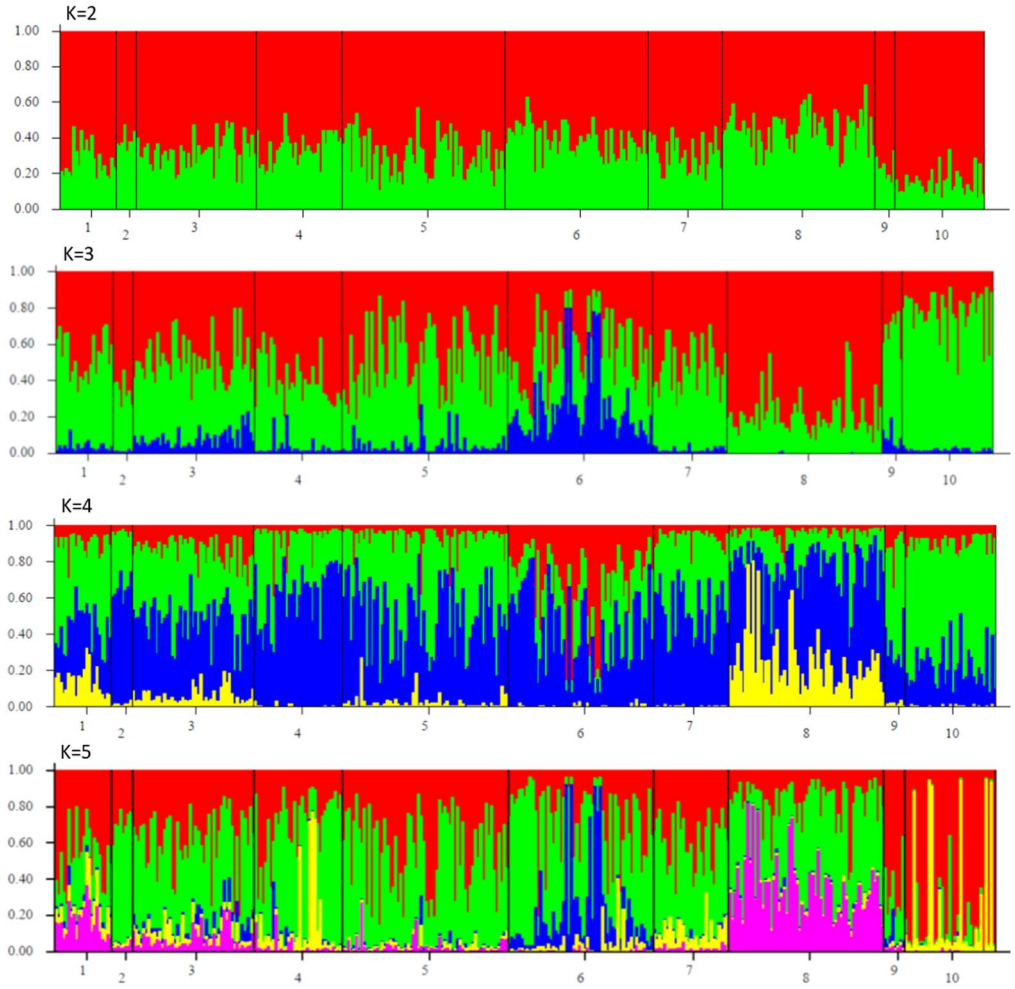


**Figure 5.5** Individual membership to K=2-6 genetic clusters detected using 77 SNPs and a Bayesian clustering approach implemented in STRUCTURE (Pritchard et al 2000). Individuals are sorted by sampling location. The best supported value of K based on STRUCTURE Harvester (Earl and vonHoldt, 2012) was calculated as K=2. Sampling locations were as follows: 1, Mwagna; 2, Congo; 3, North; 4, Cristal; 5, Ivindo; 6, Centre; 7, Wonga; 8, Coast; 9, South; 10, Batéké.

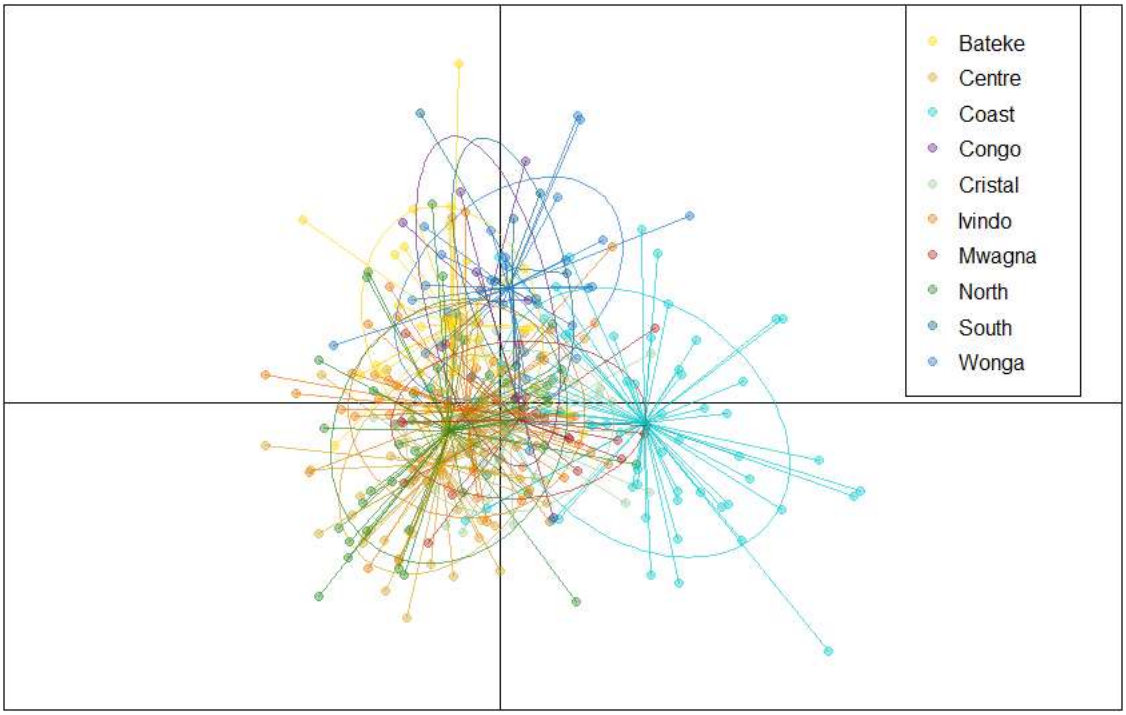


**Figure 5.6** Scatterplot of discriminant analysis of principal component using a priori defined populations for the genetic structure of forest elephants in Gabon using 86 SNP markers. Each sample is represented by a dot and coloured according to its sampling location. Inertia ellipses include 67% of the samples for each location.

In sPCA, three global components were retained (Appendix 7). The existence of a global pattern was confirmed by a significant sPCA global test (observed value = 0.0083, p-value = 0.01), while a local test was not significant (observed value = 0.0054, p-value = 0.95). Inconsistent genetic patterns were observed with the randomized datasets, thus confirming the robustness of the inferred patterns. The first sPCA scores indicated a progressive south-west/north-east differentiation with a coastal group strongly separated from other locations (Appendix 7). The second PCA scores revealed a sharp separation between two groups, with the first group distributed along the coast from Wonga Wongue to Moukalaba Doudou and the second group spanning all the interior locations and Mayumba on the Southern coast (Appendix 7). The third global scores showed a north-south separation as well as a weak differentiation of Moukalaba Doudou within the coastal location and Mwagna within the North-East (Appendix 7). The interpretation of the sPCA scores through the combination of the three interpolated lagged scores suggested the existence of four genetic groups corresponding to four geographical zones, including “Batéké”, “Coast”, “Wonga” and “Main” groups (Figure

5.7). In addition, weak genetic structure from South to North was observed within the coastal cluster.

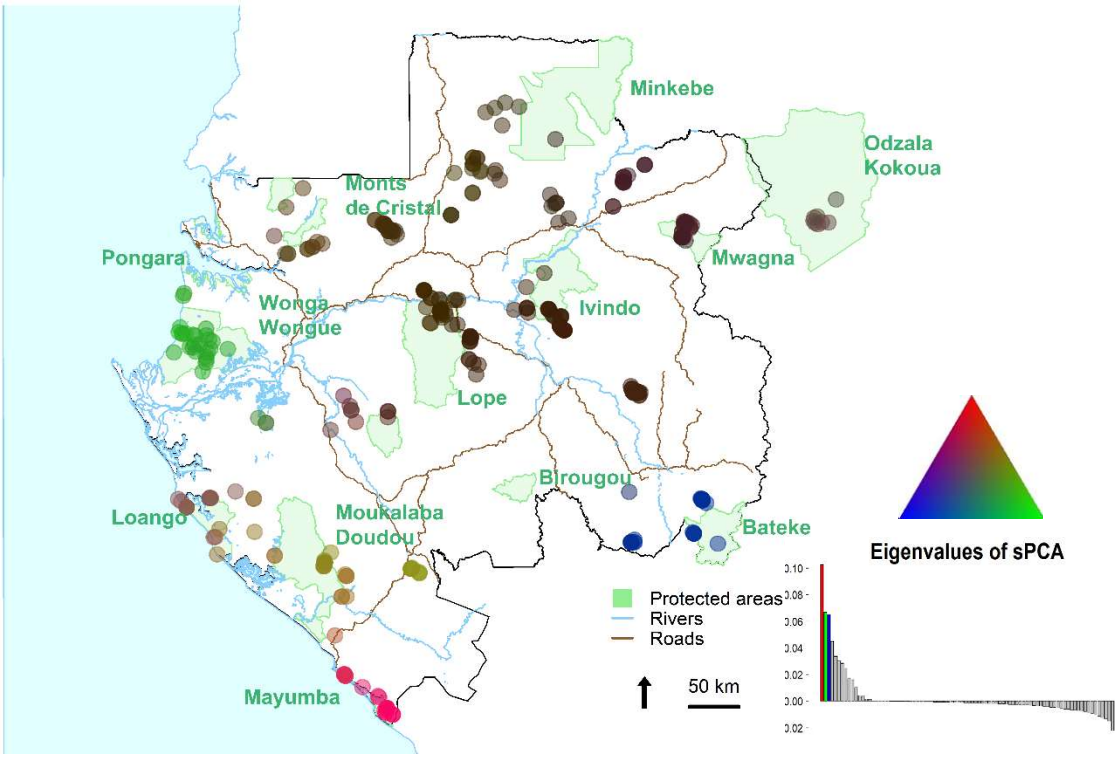


**Figure 5.7** Colorplot of spatial principal component analysis for the genetic structure of forest elephants in Gabon using 86 SNP markers. Each sample is represented by a dot and is coloured using a gradient that synthetized its coordinates on the first three principal components translated into a channel of colour: red, green and blue (as shown in the barplot of eigenvalues). Maps of each principal components are presented in Annex 7. The connection network is based on neighbourhood by distance (dmax=110 km).

*Assignment*

At the regional scale, 57.9% of the samples were unambiguously assigned (probability score > 70%) to one of the four groups identified by population structure analyses (Coast, Batéké, Main and Wonga) using all retained 86 SNP loci, of which 68.6% were correctly assigned to their sampling group (**Table 5.1** A). The average assignment success decreased to 63.1% using a selected panel of 35 highest ranking loci based on pairwise $F_{ST}$ values. Using all 86 SNP loci, the assignment success per group was higher for "Batéké" (86.4%) and "Wonga" (81.8%) than for "Coast" and "Main" groups (70.5 and 63.9%, respectively). When individuals were assigned to one of five

locations ("Main" group split into "North-East" and "Center"), the proportion of individuals unambiguously assigned and average assignment success decreased to 49.3 and 60.0% respectively (Table 5.1 B). At the local scale, 50% of unambiguously assigned individuals were correctly assigned to their sampling location (Table 5.2). However, 65.3% of the samples were ambiguous (probability score <70%). The assignment success per sampling location ranged from 11.1 (Ivindo) to 82.4% (Batéké), with four (Batéké, Mwagna, Center, Coast) out of ten sampling locations having a success >50%.

**Table 5.1** Assignment success in GENECLASS2, using 86 SNPs and four (A) or five (B) reference populations. Individuals were considered to be unambiguously assigned when the probability score was >70%.

A

| | | Assigned population | | | | | |
|---|---|---|---|---|---|---|---|
| | | Batéké | Coast | Main | Wonga | % success | Unassigned samples |
| Sampling population | Batéké | 19 | 0 | 1 | 2 | 0.864 | 13 |
| | Coast | 3 | 31 | 4 | 6 | 0.705 | 24 |
| | Main | 18 | 15 | 85 | 15 | 0.639 | 98 |
| | Wonga | 1 | 0 | 1 | 9 | 0.818 | 18 |

B

| | | Assigned population | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Batéké | Coast | Centre | Northeast | Wonga | % success | Unassigned samples |
| Sampling population | Batéké | 17 | 0 | 0 | 1 | 1 | 0.895 | 16 |
| | Coast | 3 | 25 | 1 | 4 | 5 | 0.079 | 30 |
| | Centre | 3 | 1 | 19 | 6 | 3 | 0.094 | 24 |
| | Northeast | 11 | 8 | 17 | 35 | 11 | 0.134 | 93 |
| | Wonga | 1 | 0 | 0 | 1 | 6 | 0.125 | 21 |

**Table 5.2** Assignment success in GENECLASS2, using 86 SNPs and ten reference populations. Individuals were considered to be unambiguously assigned when the probability score was >70%.

| | | Assigned population | | | | | | | | | | % success | Unassigned samples |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Batéké | Centre | Coast | Congo | Cristal | Ivindo | Mwagna | North | South | Wonga | | |
| Sampling population | Batéké | 14 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0.824 | 18 |
| | Centre | 3 | 14 | 0 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 0.583 | 32 |
| | Coast | 2 | 0 | 15 | 2 | 3 | 0 | 2 | 1 | 1 | 1 | 0.556 | 33 |
| | Congo | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0.333 | 5 |
| | Cristal | 1 | 1 | 1 | 1 | 4 | 1 | 0 | 0 | 0 | 1 | 0.400 | 24 |
| | Ivindo | 1 | 4 | 1 | 2 | 3 | 2 | 2 | 2 | 1 | 0 | 0.111 | 46 |
| | Mwagna | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 0 | 0 | 0 | 0.800 | 17 |
| | North | 4 | 1 | 0 | 0 | 1 | 0 | 3 | 6 | 0 | 0 | 0.400 | 32 |
| | South | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NA | 8 |
| | Wonga | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 3 | 0.429 | 22 |

We were able to combine SNP assignment and mitochondrial DNA haplotype distribution for 343 samples. At the regional scale, 11.5% and 56.6% of the samples that were not assigned (score < 70%) to one of five locations using the SNP panel carried a haplotype specific to a locality or a region, respectively. Among the 21.9% of samples misassigned using the SNP panel to a location that was different to their sampling location, 8.0 and 54.6% carried a locality specific or geographically restricted haplotype, respectively. In total, 75.8% of the samples were assigned to a specific area using one or both tests. At the local scale, 25.9% of the samples were assigned to one of ten locations using one or both tests, while another 49.5% carried geographically restricted haplotypes.

## 5.5 Discussion

We conducted the first study using nuclear SNP markers in forest elephants and detected more genetic structuring than previously detected in this species (Bawe-Johnson, 2008; Ishida et al., 2018). Four groups (Coast, Batéké, Wonga and Main) and patterns of genetic differentiation from North to South within the "Coast" group were detectable. Bayesian clustering, multivariate analyses and assignment tests all supported the existence of a weak population structure, although $F_{ST}$ values were moderate, partitioning from Bayesian clustering was incomplete and clusters identified by multivariate analyses partly overlapped. Using a large dataset of mitochondrial DNA control region (D-Loop) sequences, we found high genetic differentiation with a number of haplotypes exhibiting restricted geographic distributions. Selection of the most informative nuclear SNP markers, combination with mitochondrial DNA polymorphism information and intensive reference sample collection allowed us to develop a new test able to source individuals to five groups of protected areas and, in some cases, to 10 protected areas or group of protected areas.

*Nuclear genetic patterns*

We found limited nuclear genetic structuring in forest elephants in Gabon using 86 SNP markers. This finding was consistent with previous studies using microsatellites in Central Africa that showed similar patterns of incomplete partitioning between groups using multivariate analyses and Bayesian clustering approaches (Bawe-Johnson,

2008; Ishida et al., 2018). In elephants, male-biased dispersal results in a high level of gene flow in unfragmented populations. While most elephant populations are now fragmented by habitat loss and intense poaching leading to local extinctions, Gabon hosts one of the last quasi-continuous elephant populations in Africa. As a comparison, the continuous *L. africana* population in the Kavango-Zambezi transfrontier conservation area, which occupies an area of similar size to Gabon (~270,000 km2) across Northern Botswana and neighbouring countries, displays similar patterns of low nuclear DNA structure with some evidence of isolation by distance (De Flamingh et al., 2015). Moreover, even in elephant populations isolated due to anthropogenic disturbances, previous attempts failed to show clear genetic differentiation (Okello et al., 2008), because landscape changes typically only result in detectable spatial genetic structure after a time lag of five to ten generations (Anderson et al., 2010).

Our study revealed the existence of four groups within Gabon, with evidence of genetic differentiation within one group. Our SNP markers were more powerful than microsatellites panels previously used (Bawe-Johnson, 2008; Ishida et al., 2018) for detecting cryptic genetic structuring. Using 12 polymorphic microsatellite loci identified from the African savannah elephant, a previous study suggested that coastal populations in Gabon were distinct from other western Congolese forest populations (Bawe-Johnson, 2008). Using 18 new microsatellites markers developed for forest elephants (Gugala et al., 2016), a recent larger-scale study revealed only two groups representing the western and eastern sides of the Congolese forest block (Ishida et al., 2018). However, no sample from the Gabonese coastal populations were included in this latter analysis. Isolation by distance was not detected across the western side (Bawe-Johnson, 2008) but possible across the whole Congolese forest block (Ishida et al., 2018). Our finding corroborated previous studies stating that SNPs have more power than microsatellites to detect weak population structure because they are distributed across the entire genome (Glover et al., 2010), with markers under selection further increasing power to detect population differentiation (Helyar et al., 2011; Landguth and Balkenhol, 2012).

One of the most structured populations according to our study was along the coast of Gabon. This was consistent with previous study (Bawe-Johnson, 2008). This pattern might be the result of natural forces, as the region is separated from the interior of the Ogooue basin by a series of low mountains, or historical overharvesting that led

to local extinction along the coast during the seventeenth century followed by recolonization (Hymas, 2016). Interestingly, a similar pattern has been observed in crocodiles of the genera *Mecistops* and *Osteolaemus* (Shirley and Austin, 2017) and was associated with geological features. The other differentiated forest elephant population in our study was from Batéké Plateau, an area characterized by distinct geological features (Vande Weghe et al., 2016). We found that gene flow was high in the Central to North-Eastern part of the country which corresponds to the interior of the Ogooue basin Lowlands and was, until recently, characterized by high elephant densities (Maisels et al., 2013).

Our ability to resolve genetic structuring in a continuous population might have been limited by the relatively small size of the 86-SNP panel used to investigate genetic differentiation. Previous studies highlighted the high resolution obtained from very large panels of SNPs (Benestan et al., 2015; Bovine HapMap Consortium, 2009). Increasing the number of SNP markers from a hundred to a thousand allowed successful detection of weak genetic structure in several studies (Benestan et al., 2015; Jones and Wang, 2012; Puckett and Eggert, 2016). However, thorough panel selection can provide more gain in power to detect genetic structuring than increased numbers of loci (Ogden, 2011; Wilkinson et al., 2011). For instance, as many as 200 SNPs were needed for accurate and precise spatial assignment in bears (Puckett and Eggert, 2016). In contrast, panels of 8 to 50 SNPs were able to correctly assign fish to area of origin (Nielsen et al., 2012). We used a panel of 107 SNP markers previously selected based on $F_{ST}$ as an indicator of potential informativeness to investigate genetic differentiation (chapter 2, Bourgeois et al. 2018). In order to produce a fast, cost-effective and reliable assignment test, it is crucial to maximize its power while reducing panel size. The assignment success was not affected when the SNP panel decreased from 86 to 35 loci selected based on global $F_{ST}$ values, suggesting that our approach was successful at selecting most informative markers. However, the power of the test decreased when individuals were assigned to five compared to four locations. There is potential for the identification of additional powerful markers within the SNP resource available for forest elephants (chapter 2, Bourgeois et al 2018) to help better resolve fine-scale genetic structure.

*Mitochondrial DNA haplotypes*

We found patterns of high genetic differentiation based on mitochondrial DNA sequences, with pairwise $F_{ST}$ values higher than values calculated using nuclear DNA data. This was consistent with previous studies of forest elephants both at regional (Ishida et al., 2013, 2018) and local scales (Eggert et al., 2014). Discordant patterns of nuclear and mitochondrial DNA structure are found in species with male-biased dispersal behaviour (Toews and Brelsford, 2012) and have been fully described in elephants (Debruyne, 2005; Roca et al., 2007). Social behaviour is well known in *L. africana*, in which male dispersed across long distances while female philopatry limits dispersal of mitochondrial DNA (Archie et al., 2008). Savannah elephants live in fission-fusion societies where maternal genetic relatedness predict associations between females (Archie et al., 2006). Social behaviour of the forest elephants has been less studied due to their elusive nature. However, direct observation at savannah patches or forest clearings (Fishlock and Lee, 2013; Morgan and Lee, 2007; Turkalo et al., 2013) and genetic studies (Schuttler et al., 2014) supported the existence of similar fission-fusion patterns, with associations between related females sharing the same mitochondrial DNA haplotype being consistent over time even though social groups were smaller than in savannah elephants.

We found two haplogroups that overlapped with west-central and north-central haplogroups previously described in Gabon (Ishida et al., 2013). The most internal haplotype of each haplogroup were widely shared across the country, which was consistent with the hypothesis that forest elephant mitochondrial haplogroups originated from separate glacial forest refugia and their distribution followed the expansion of populations from these refugia during post-glacial periods (Brandt et al., 2012; Ishida et al., 2018). In contrast, external haplotypes showed clear restricted geographic distribution. This could be explained by restricted mitochondrial DNA gene flow due to female philopatry combined with high levels of mutation observed in the mitochondrial DNA control region (Sigurḥardottir et al., 2000). As a consequence, the potential mitochondrial DNA hypervariable D-Loop (control region) to identify mitochondrial DNA lineages corresponding to specific locations for geographical assignment is high (Ogden and Linacre, 2015).

While the use of mitochondrial DNA to determine the geographical provenance of ivory has been demonstrated before (Ishida et al., 2013), our study provided the highest density of reference sequences available across the forest elephant range. By increasing the number of reference data from 290 to 828, we found higher resolution from the country to the subregional level. The clear benefit for geographical assignment was enhanced by low cost and high success associated with mitochondrial DNA sequencing using non-invasive and degraded samples. The combination of tests increased overall assignment success, while the precision was increased when source areas identified by both tests partly overlapped. In addition, haplotype data helped detecting misassignments in most cases, thus reinforcing strength of our conclusions. Therefore, we strongly recommend increasing the collection of reference samples across forest elephant range to support investigation of the origin of ivory with a higher resolution.

*Conservation implications*

We found that it was possible to increase resolution of geographic assignment in forest elephants using novel SNP panels combined with mitochondrial DNA haplotypes. Assigning elephant samples to one of four groups across Gabon increased precision compared to a previous study reporting a mean precision of 349 km for forest elephant (Wasser et al., 2015), which corresponds to half the size of Gabon. In addition, the 69% assignment success in our study was comparable to the success reported from previous assignment test using microsatellites, i.e. 50% of samples correctly assigned within 301 km and 75% within 416 km of their sampling location (Wasser et al., 2015). Increasing resolution of geographic assignment from countries up to land management units, such as Protected Areas, or logging concessions' level is highly relevant for the conservation of species targeted by the illegal wildlife trade. Because of the difficulty in monitoring large wild areas, identifying local protected areas specifically targeted by poachers is crucial to help better allocation of anti-poaching efforts. Moreover, fine scale investigation of the illegal ivory trade, which is urgently needed to improve understanding of trade routes within Africa from poaching sites to the ports of exit (Milliken, 2013) would also benefit from the application of high resolution assignment tools to determine the origin of seized ivory. As transnational ivory trade relies initially on regional trade, it has been highlighted that law enforcement investigations should

focus on locations and on middlemen who are fuelling the trade by connecting the local with the transnational networks (Titeca, 2018). Although genetic tools are already available to investigate global trade routes (Wasser et al., 2015), our study constitutes a step forward toward delivering intelligence tools that are adapted to the finer scale investigation of the local and regional ivory trade required for national or local authorities to react on the ground.

Our sampling design and definition of locations was largely based on non-genetic management units, i.e. national parks and other protected areas. Many studies investigate population genetic structure in order to provide management recommendations (Funk et al., 2012; Laikre et al., 2005; Poulakakis et al., 2008). However, ecological findings do not necessarily align with management realities. In particular, genetic clusters may not match management units (Degen et al., 2017). To circumvent this inevitable limitation, we used a needs-driven approach to maximize power to distinguish between *a priori* defined locations corresponding to land management units in a continuous population with weak genetic structure. However, we obtained uneven resolution across Gabon, with lower resolution in the "Main" group corresponding to the central and North-Eastern part of the country where genetic differentiation was the lowest. Future research should prioritise the 'Main' area due to its high conservation urgency (Poulsen et al., 2017), to further increase resolution by increasing sizes of reference sample set and SNP panel.

Geographical assignment is a relatively new and undeveloped area of research within the growing field of wildlife forensic science (Ogden and Linacre, 2015). Despite being a common request by crime investigators, relatively few examples exist (Mondol et al., 2015; Wasser et al., 2008) because test development is challenging and specific to one investigative question and species of interest (Ogden and Linacre, 2015). A significant investment of both time and resources was required in developing new SNP markers for forest elephants (chapter 2) and building an extensive and high-quality reference database. The endangered status and elusive nature of forest elephants drove the need to collect non-invasive samples. Access to samples was difficult as elephants live in remote and hardly accessible locations and because limited local knowledge was available to guide dung searching. In addition, SNP genotyping on a large faecal sample set was technically challenging due to typical low DNA yield, presence of PCR inhibitors and DNA degradation in non-invasive samples (Morin et al., 2001; Taberlet

et al., 1999). Only recently has SNP genotyping been applied to large non-invasive sample sets due to the need of high DNA quality and quantity (Goossens et al., 2016; Spitzer et al., 2016; von Thaden et al., 2017). Moreover, a main limitation of assignment test development is that its accuracy relies on the existence and detectability of underlying genetic structure. One of the main challenges is the incongruence between ecological patterns and political boundaries (e.g. countries, national parks) of the geographical region of interest for law enforcement (Degen et al., 2017; Ogden and Linacre, 2015). Weak genetic structure in a continuously distributed species may force to reduce expectations from wildlife trade investigators and protected area managers. While we believe that geographical assignment is a promising powerful tool to help tackling the illegal wildlife trade, we strongly recommend prior discussions between researchers and managers to ensure research being needs-driven and managers being aware of promises and limitations of this approach.

## 5.6 Acknowledgements

We thank the Centre National de la Recherche Scientifique et Technologique (CENAREST; permit n°AR0016/14/MESRS/CENAREST/CG/CST/CSAR) for research authorization to conduct this study in Gabon, and the Direction Générale de la Faune et des Aires Protégées (DGFAP) for delivering CITES sample export permit (n°0161, 0210, 0215 and 0222). We thank the Agence Nationale des Parcs Nationaux (ANPN), in particular L. J. T. White, and the Institut de Recherche en Ecologie Tropicale of CENAREST for hosting the study and providing institutional and logistical support. We are particularly grateful to ANPN field staff (all the park rangers and wardens), in particular J.-P. Mondjo, for their assistance in the field. Forestry concessions (SEEF, Rougier, CEB, Toujours Vert, Foreex), OELO, SFM, WCS, SEGC and PPG provided logistical support. C. Orbell and L. Makaga collected additional samples. We thank Steve Blake, Sharon Deem and Odzala-Kokoua National Park (Republic of Congo), in particular Guillaume Le Flohic, for providing samples from collared elephants. Funds for the project were provided by CEEAC under the ECOFAC V Programme (Fragile ecosystems of Central Africa) funded by the European Union (contract n° FED/2013/332-349).

# Chapter 6
# General discussion

The illegal trade in wildlife products, including traditional medicine, ornaments, pets and food has become a major driver of extinction for high value species, with elephant ivory being among the most traded products (United Nations Office on Drugs and Crime, 2016; Wittemyer et al., 2014). The steep decline of forest elephants (*Loxodonta cyclotis*) due to poaching and habitat reduction (Maisels et al., 2013; UNEP et al., 2013) is driving an urgent need for new tools to better inform management and conservation. The survival of this charismatic species requires strong commitment and collaboration between all countries involved in the ivory supply chain to stop poaching and illegal exportation, as well as to control illegal importation and reduce demand in consumer countries. It is essential to develop novel approaches to underpin traditional field techniques that have failed to provide real-time data for immediate intervention in trade chains or poaching campaigns on the ground. Genetic approaches offer great potential to produce reliable evidence for timely monitoring of poaching pressure and insights into the illegal ivory trade patterns.

I set out to increase genetic resources available for forest elephants in order to provide fast, accurate, cost-efficient and needs-driven tools to support National wildlife law enforcement agencies into elephant population management and investigation of the illegal ivory trade at the intra-national level. In this thesis, I presented the first genome-wide resource of 1365 single SNP loci generated for African forest elephants and validated genotyping assays for a subset of 107 SNPs (Chapter 2, Bourgeois et al. 2018). This represents a major advance for the genetic study of this taxon and was essential for the development of accurate assignment tests. I also presented new fast sexing tests that are suitable for all three elephant species (Chapter 3). Having been designed as standard allelic discrimination assays, these two sex determining assays can be incorporated into larger nuclear SNP panels. These new SNP and sexing markers allowed Gabon National Parks Agency (ANPN) to develop in-country laboratory capacity to produce real-time genetic data. Then, I provided guidelines for wildlife

managers to improve the cost-efficiency and quality of routine faecal DNA surveys, including validation of a very convenient and efficient sampling method and the development of a new quantitative PCR assay (Chapter 4). These guidelines have been implemented by ANPN for direct application of extensive faecal DNA sampling to elephant management, such as population census. In addition, I followed these guidelines to create a high-quality reference dataset of SNP genotypes across Gabon for geographical assignment, using elephant faecal samples. Finally, I developed a new test for the geographical assignment of elephants to land management units within Gabon (Chapter 5). This test allows to source ivory at fine scale within an important source country, thus directly addressing ANPN's need for tools to support identification of poaching hotspots and local trade routes at the intra-national scale.

Results and genetic tools developed in this thesis provide the foundation for a range of practical tools and in-depth information for timely response to the current crisis in this species' survival prospects and are made available for both wildlife managers and the elephant research community. Chapter 2 has been published in an open access Journal and Chapter 4 submitted for open access publication. The genetic sampling of poached elephant carcasses using field kits and protocols elaborated for this thesis has been implemented in 2014, with the ultimate goal to link tusks to carcasses. Since 2015, the systematic DNA sampling of seized ivory has been an important part of the national ivory management system aiming at preventing leakage from the stockpiles. In total, over a hundred rangers and field assistants have already been trained in ivory, carcass or faecal sampling.

Results from this thesis directly support the newly developed Gabon National Elephant Action Plan (NEAP). The use of genetic approaches to determine the location of origin of national ivory seizures, develop data exchange networks within Africa and Asia, collect genetic evidences on elephant crime scene in order to provide forensic evidence in court, and set up a qualified wildlife DNA laboratory have been stated as clear objectives for elephant conservation. Given that geographical assignment of illegal wildlife products is becoming a common investigative request, results from this thesis can also be used to provide recommendations and highlight limitations for the application of similar approaches in a wide variety of non-model organisms targeted by the illegal wildlife trade. ANPN has already uptaken the project approach in its strategy for pangolin conservation, which now includes genetic mapping of wild populations for

traceability. Here, I discuss implications of the results of this thesis for conservation and management and identify important directions for future research.

## 6.1 Conservation implications

A main aim of this thesis was to provide efficient tools to support national wildlife law enforcement agencies in their efforts to address the current elephant conservation crisis. I discuss the value of the results of this thesis for the traceability of ivory within a country of origin for tackling the illegal ivory trade at source. The novel tools developed in this thesis are already contributing to enhance the use of genetic surveys for better-informed management of forest elephants through the availability of new genetic resources, increased cost-efficiency and ease of transfer for capacity building.

*Novel SNP resources for forest elephants*

This thesis presents the first genome wide SNP resource for the forest elephant. As SNP resources were scarce across elephant species (Sharma et al., 2012) and not available for African elephant species, this is a major step in the study of this taxon. SNP markers are associated with a string of advantages derived from their bi-allelic nature, including low error rates and potential for automatic high-throughput genotyping, rendering large scale studies realistic (Ranade et al., 2001).

Another valuable application of SNP markers is their capacity to reveal weak genetic population structuring that can be undetected by other markers such as microsatellites, as supported by the findings of this thesis. The novel SNP panel was able to reveal more genetic structure than was previously detected in Gabon (Bawe-Johnson, 2008; Ishida et al., 2018). However, the SNP library was generated using samples collected in various locations across Gabon, which means that ascertainment bias (Clark et al., 2005) is a possibility and may limit the ability of this novel SNP panel to investigate genetic differentiation at the regional scale (Nielsen, 2004). Further investigation is needed to explore the extent of genetic variability at these new SNP markers in other forest elephant populations. The guidelines presented in this thesis (Chapter 2) will be useful for the expansion of the SNP discovery using a larger number of samples in the future to mitigate the effects of ascertainment bias in other regions.

*Traceability of ivory*

Testing ivory seizures to determine their geographical provenance is a CITES requirement for all Parties involved in the illegal ivory trade (Convention on International Trade in Endangered Species, 2014). While tests exist to investigate global trade patterns (Wasser et al., 2015; Ziegler et al., 2016), this thesis contributes to address the gap in the understanding of local and national trade patterns (Milliken, 2013; Titeca, 2018), through the development of a new test able to identify the geographical provenance of ivory within a source country. Analyses of all seizures made in Gabon during the last couple of years are on-going using this test and will allow ANPN to better understand national trade routes and ivory supply for Gabon onto the global illegal market.

Developing tests at the intra-national level is challenging. Firstly, underlying genetic structure is generally weak and difficult to detect at fine scale in highly mobile or long-lived species (Latch et al 2014). Secondly, developing accurate tests requires an extensive and good quality reference dataset at the target geographical scale. Increased reference sample size has been shown to improve accuracy of assignment when the population structure is weak (Benestan et al 2015). However, lack of communication between scientists developing tests and national wildlife agencies having access to the resource is one reason limiting the collection of reference samples, while CITES regulations might further slow or hinder the exchange of reference samples. As a consequence, most tests rely on limited number of samples per country (~30) not allowing the development of fine-scale analyses. This thesis demonstrated how the establishment of strong bonds between research institutions and national wildlife law enforcement agencies, along with joint involvement in the development of new tools, facilitated access to samples for the creation of reference databases at the intra-national level.

Lastly, land management units, such as National Parks or forestry concessions may not necessarily reflect gene flow, thus creating a discrepancy between boundaries defined by management realities and genetic differentiation. When genetic markers are unable to detect genetic structuring, assignment tests between arbitrary units might lead to low assignment power or misassignment of individuals to a different population from their true origin. The approach used in this thesis was based on a combination of the use

of management boundaries with the investigation of genetic differentiation. This approach allowed to develop a test able to source elephant samples to five or in some cases ten sampling locations within Gabon. These findings corroborated results from a previous study that consisted of the selection of panels of SNP loci based on statistical power to address different policy-led scenarios of illegal fishing and false certification and led to the develop of successful assignment tests (Nielsen et al., 2012).

*Limitations of the geographical assignment tests*

A main limitation of the assignment test developed in this thesis was the low resolution and assignment success in a large area encompassing six National Parks located in the Central, Northern and Eastern parts of Gabon that was characterized by high levels of gene flow (Chapter 5). However, the combination of another independent test based on mitochondrial DNA haplotype with the SNP genotyping was successfully used to increase resolution. Mitochondrial DNA sequences are useful because their geographical distribution does not reflect gene flow but rather female philopatry (Archie et al., 2008). Data from this thesis has been contributed to the continent scale reference database (Ishida et al., 2011). The generation of an extensive database throughout Gabon increased the resolution of geographical assignment through the identification of several new rare haplotypes and a better understanding of the geographical distribution of haplotypes. However, a third of individuals carrying common haplotypes might not be assigned to a specific area. Further research is needed to increase power to detect nuclear genetic differentiation at fine scale in the North-East part of Gabon due to its conservation urgency (Poulsen et al., 2017). The use of haplotype frequency differentiation across populations to better infer geographical provenance using this dataset and the development of mitochondrial DNA SNP assays for fast in-country analyses are potential avenues to improve the accuracy and implementation of the test.

Another limitation of assignment tests is the risk of over-interpreting results (Ogden and Linacre, 2015) due to incomplete information on the behaviour of forest elephants. Dispersal behaviour is poorly investigated in forest elephants, but it is most likely that males move long distances as observed in savannah elephants (Archie et al., 2008). This may mislead conclusions about poaching locations for individuals that might have travelled far from their natal group (Manel et al., 2005). The tools developed

in this thesis will enable researchers to combine geographical assignment with sexing tests in order to better guide interpretation of assignment results.

*Forensic validation*

It is crucial that available penalties for wildlife crime are applied for an effective impact of law enforcement efforts. There is an urgent need for forensic tools to support the legal system by providing evidences for prosecutions. The transfer of research tools to forensic science requires formal validation of the test for acceptance of evidence in court (Ogden and Linacre 2015). Among the genetic markers available, SNPs are well-suited for forensic validation due to their binary nature (Ogden, 2011). Results from this thesis has allowed ANPN to develop national wildlife DNA capacity in Gabon. The next step for the application of the new available tools is their forensic validation. Validation of the quantification test for species identification from blood remains will be used to prove the involvement of a suspect into an elephant poaching case. In addition, selection and forensic validation of a SNP panel suitable for individual identification of forest elephants will have practical implications for law enforcement by establishing links between elephant meat or tusks, carcasses and/or suspects, with the ultimate goal to provide evidence for wildlife crime.

The application of these genetic tools to investigate illegal ivory trade and provide evidences require prior optimization of DNA extraction protocols from a variety of samples that are expected to be degraded or contain low DNA yield such as decomposing tissue and bones from carcasses or ivory. Such work is on-going in Gabon (~ 150 DNA samples from elephant carcasses and ~ 200 ivory DNA samples have been tested) and will benefit from guidelines developed for non-invasive samples and presented in this thesis.

*Towards routine genetic surveys for elephant management*

Genetic approaches are now essential tools for the understanding of species' ecology, especially in elusive species. Genetic data can be incorporated into conservation planning, with an emphasis on the need to preserve genetic diversity within species and connectivity between fragmented populations. Genetic approaches also have potential to replace more burdensome field techniques, such as transects methods which are costly and provide only low precision, especially in forest environments. An important aim of this thesis was to address the urgent need for novel

genetic tools to better inform conservation and management of the understudied forest elephant. This thesis delivered a set of tools for forest elephant genetic surveys, with validation of protocols and guidelines for all steps of genetic analyses, including sample collection, sample storage, DNA extraction, pre-screening of DNA yield and genotyping. These tools have already been uptaken by ANPN in Gabon and are now in use to support the implementation of several objectives within the NEAP. Furthermore, these tools are available for the whole elephant research community and will support the increasing implementation of genetic surveys of elephants worldwide driven by the urgency of conservation and management interventions for these species.

Improving cost-efficiency and convenience in the field was crucial to promote the use of novel tools by wildlife managers. Results from this thesis showed that protocols were successful in providing higher DNA yield than traditional sampling techniques, suitable for non-invasive and degraded samples, and allowed to pre-screen DNA samples before genotyping in order to decrease laboratory costs and efforts (Chapter 4). The technique was also successful in sampling elephant faecal DNA in savannah environments, which highlighted its potential for use in savannah elephants *(Loxodonta Africana*). The only limitation in savannah environments was a shortened optimal collection period since only the sides of the dungs that were shaded from direct UV light were suitable for DNA sampling. Dung survey standards for elephant species have been developed by the Monitoring the Illegal Killing of Elephants programme (MIKE) more than a decade ago (Hedges and Lawson, 2006) and remained unchanged despite the need for more efficient techniques. Based on findings of this thesis, I strongly recommend updating the guidelines for elephant faecal DNA sampling using collection and quantification protocols developed in this thesis. Protocols derived from this thesis were delivered and successfully used to ANPN and partners, including World Wildlife Fund, Wildlife Conservation Society, Panthera, Smithsonian Institution, Fauna and Flora International and Duke university for elephant sample collection in Africa. In addition, results of this thesis have been submitted to an open access journal for wide dissemination.

In-country laboratory capacity building is critical for the implementation of routine genetic surveys to inform elephant management. The genetic tools developed in this thesis can be transferred to range, transit and consumer countries for direct application of genetic tools to elephant monitoring and investigation of the illegal ivory

trade. Advantages of SNP markers over other microsatellites are the ease of technical portability and data sharing with no requirement for calibration among instruments. Moreover, SNP assays offer opportunity for in-house genotyping without the need to access specialized equipment such as a sequencer which is rarely available in-country. Laboratory capacity building for forest elephant genetic analyses has been undertaken by ANPN since 2014, as a direct application of techniques developed in this thesis. Basic laboratory techniques and equipment required for DNA extraction were relatively cheap and fast to acquire. A DNA extraction laboratory was set up in one room equipped with a bench, a set of pipettes, a centrifuge, an incubator, a vortexer and a freezer (total cost < 6,000 USD) and ready-to-use DNA extraction kits. Training of a lab technician in DNA extractions was possible within a couple of weeks. Further capacity transfer required a higher investment in terms of equipment and training for DNA quantification and genotyping. However, real-time PCR machines have become affordable (cost ~25,000 USD and as low as 5,000 USD for a basic second-hand real-time PCR machine), as did genotyping chemistries for genotyping and quantitative PCR. Gabon has become fully autonomous for the DNA extraction and genotyping from a variety of elephant samples, including faeces, ivory and carcass samples. SNP genotyping is routinely carried out using KASPar probes (LGC Genomics) in single-locus reactions on a StepOne Real-time PCR system (Applied Biosystems). This novel national capacity offers strong advantages such as increased success due to short storage time, cost reduction and absence of need for export permits. Furthermore, the current capacity building in Gabon offers new avenues for future research on forest elephant genetics.

## 6.2 Research perspectives

*Individual identification and parentage analyses*

Genetic tools can be particularly efficient for individual identification and parentage analysis in elusive species. Except at a few forest clearings (Turkalo et al., 2013), direct observations of forest elephants are rare and in-depth sociality studies or estimation of population sizes require indirect non-invasive methods. Novel SNP resources presented in this thesis are contributing to the development of new tools for elephant monitoring. Testing of the use of a reduced SNP panel for the identification of

individuals combined with mark-recapture analyses is underway in Gabon. This represents a promising area of research for the replacement of more burdensome and costly field techniques for population census (Eggert et al., 2003; Hedges et al., 2013) and will provide insights into social structure (Schuttler et al., 2014) and dispersal patterns of this elusive species. Importantly, DNA-based population surveys will provide an efficient tool for early warning of population declines and identification of poaching hotspots.

Results from this thesis indicated that many of the novel SNP markers will be useful for individual identification and parentage analysis (Chapter 2). A subset of potentially informative SNP loci was selected from the library with the aim of providing management tools and addressing different conservation questions. The idea of selecting a cost-effective SNP panel able to address different questions is very attractive, but the main challenge is the trade-off between best markers for discriminating between individuals or differentiating populations. SNPs that are useful for individual identification have high heterozygosity. In contrast, the power to discriminate between populations is linked to their ability to reveal any existing allele frequency differences between populations, and thus maximized when markers are fixed for alternative alleles in each population. However, even if the extreme values are mutually exclusive, some SNPs providing information for population assignment might still show some levels of heterozygosity within populations (Hou et al. 2011). Therefore, the 107 SNP assays presented in this thesis were developed for a subset of SNPs selected using a combination of measures of diversity and divergence in order to filter a subset of SNP markers that has substantial power for analyses at both individual and population scales.

*Genetic connectivity*

Genetic structure and connectivity analyses are crucial to inform forest elephant management plans in a context of rapid on-going economic diversification, where mining and agriculture add to the logging industry, alongside associated infrastructure development. Direct findings from this thesis and further investigation of the genetic structure at multiple scales across Gabon will provide important data for better-informed management decisions. This will contribute to objectives defined in the NEAP that includes studies on genetic structure and applications to corridor design among

research priorities. Genetic methods have great potential to infer functional connectivity at spatial scales where ecological surveys cannot be used (Holderegger and Wagner, 2008). In addition, genetic approaches might be used to understand historical patterns and thus better understand the need to restore lost connectivity. Observed genetic structure results from the cumulative effects of gene flow that occurred during both recent and historical processes. There is a time lag of 5 to 10 generations before landscape changes result in detectable effects in the observed spatial genetic structure (Anderson et al., 2010; Cushman and Landguth, 2010).

Results from this thesis revealed patterns of genetic differentiation within the quasi-continuous forest elephant population distributed across Gabon, even though genetic structure was weak (Chapter 5). Investigation of population structuring using a panel of 86 SNPs revealed the existence of four groups ("Coast", "Batéké", "Wonga" and "Main"). Additional patterns of genetic differentiation from North to South within the "Coast" group were detectable. A direction of research would be to better understand causes of genetic differentiation between coastal and interior populations. Furthermore, the use of markers under selection to increase power has been demonstrated (Landguth and Balkenhol 2012) and future research should be directed towards identifying non-neutral SNPs within the novel SNP library presented in this thesis.

## 6.3 Final conclusions

The main objective of this thesis was to provide new elephant genetic resources and methods to support National wildlife law enforcement agencies 'efforts to stop poaching and supplying of the illegal ivory trade at source. The study was designed in close collaboration with ANPN in order to address specific needs and understand the conservation context and constraints. As a result, novel tools have already been uptaken by ANPN and national capacity building was implemented as a direct result of this thesis. The new laboratory capacity for elephant DNA genotyping in Gabon brings perspectives to better address the need for genetic data that was identified in the NEAP. Genetic data provides real-time information for immediate actions to tackle the illegal ivory trade, but also important insights into population connectivity for the long-term conservation of the species. Findings from this thesis highlight the importance of good

communication between conservation scientists and wildlife managers to identify investigative questions at a scale at which managers will be able to be responsive to results, and thus develop effective tests that can be applied in practice.

## References


Adenyo, C., Ogden, R., Kayang, B., Onuma, M., Nakajima, N., and Inoue-Murayama, M. (2017). Genome-wide DNA markers to support genetic management for domestication and commercial production in a large rodent, the Ghanaian grasscutter (*Thryonomys swinderianus*). Animal Genetics *48*, 113–115.

Ahlering, M.A., Hailer, F., Roberts, M.T., and Foley, C. (2011). A simple and accurate method to sex savannah, forest and asian elephants using noninvasive sampling techniques. Molecular Ecology Resources *11*, 831–834.

Akomo-Okoue, E.F., Inoue, E., Atteke, C., Nakashima, Y., Hongo, S., Inoue-Murayama, M., and Yamagiwa, J. (2015). Noninvasive genetic analysis for assessing the abundance of duiker species among habitats in the tropical forest of Moukalaba, Gabon. Mammal Research *60*, 375–384.

Alacs, E.A., Georges, A., FitzSimmons, N.N., and Robertson, J. (2010). DNA detective: a review of molecular approaches to wildlife forensics. Forensic Science, Medicine, and Pathology *6*, 180–194.

Albrechtsen, A., Nielsen, F.C., and Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. Molecular Biology and Evolution *27*, 2534–2547.

An, J., Lee, M., Min, M.-S., Lee, M.-H., and Lee, H. (2007). A molecular genetic approach for species identification of mammals and sex determination of birds in a forensic case of poaching from South Korea. Forensic Science International *167*, 59–61.

Anderson, C.D., Epperson, B.K., Fortin, M.-J., Holderegger, R., James, P.M., Rosenberg, M.S., Scribner, K.T., and Spear, S. (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. Molecular Ecology *19*, 3565–3575.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Arandjelovic, M., and Vigilant, L. (2018). Non-invasive genetic censusing and monitoring of primate populations. American Journal of Primatology *80*, e22743.

Arandjelovic, M., Guschanski, K., Schubert, G., Harris, T.R., Thalmann, O., Siedel, H., and Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. Molecular Ecology Resources *9*, 28–36.

Archie, E.A., and Chiyo, P.I. (2012). Elephant behaviour and conservation: social relationships, the effects of poaching, and genetic tools for management. Molecular Ecology *21*, 765–778.

Archie, E.A., Moss, C.J., and Alberts, S.C. (2006). The ties that bind: genetic relatedness predicts the fission and fusion of social groups in wild African elephants. Proceedings of the Royal Society of London B: Biological Sciences *273*, 513–522.

Archie, E.A., Maldonado, J.E., Hollister-Smith, J.A., Poole, J.H., Moss, C.J., Fleischer, R.C., and Alberts, S.C. (2008). Fine-scale population genetic structure in a fission–fusion society. Molecular Ecology *17*, 2666–2679.

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*, e3376.

Baker, C.S., Steel, D., Choi, Y., Lee, H., Kim, K.S., Choi, S.K., Ma, Y.-U., Hambleton, C., Psihoyos, L., and Brownell, R.L. (2010). Genetic evidence of illegal trade in protected whales links Japan with the US and South Korea. Biology Letters rsbl20100239.

Ball, M.C., Pither, R., Manseau, M., Clark, J., Petersen, S.D., Kingston, S., Morrill, N., and Wilson, P. (2007). Characterization of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template. Conservation Genetics *8*, 577–586.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., and Maguire, K.C. (2011). Has the Earth's sixth mass extinction already arrived? Nature *471*, 51.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R Package Version *1*, 1–12.

Bawe-Johnson, M.B. (2008). Genetic variation of the forest elephant *Loxodonta africana cyclotis* across Central Africa. PhD Thesis. Cardiff University.

Beja-Pereira, A., Oliveira, R., Alves, P.C., Schwartz, M.K., and Luikart, G. (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. Molecular Ecology Resources *9*, 1279–1301.

Bell, L.S. (2011). Forensic Science in Support of Wildlife Conservation Efforts—Morphological and Chemical Approaches (Global Trends)—. Forensic Science Review *23*, 29–35.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., and Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). Molecular Ecology *24*, 3299–3315.

Blake, S., and Hedges, S. (2004). Sinking the flagship: the case of forest elephants in Asia and Africa. Conservation Biology *18*, 1191–1202.

Blake, S., and Inkamba-Nkulu, C. (2004). Fruit, minerals, and forest elephant trails: do all roads lead to Rome? Biotropica *36*, 392–401.

Blake, S., Strindberg, S., Boudjan, P., Makombo, C., Bila-Isia, I., Ilambu, O., Grossmann, F., Bene-Bene, L., De Semboli, B., and Mbenzo, V. (2007). Forest elephant crisis in the Congo Basin. PLoS Biology *5*, e111.

Blake, S., Deem, S.L., Strindberg, S., Maisels, F., Momont, L., Isia, I.-B., Douglas-Hamilton, I., Karesh, W.B., and Kock, M.D. (2008). Roadless wilderness area determines forest elephant movements in the Congo Basin. PloS One *3*, e3546.

Blake, S., Deem, S.L., Mossimbo, E., Maisels, F., and Walsh, P. (2009). Forest elephants: tree planters of the Congo. Biotropica *41*, 459–468.

Blanc, J. (2008). Loxodonta africana. The IUCN Red List of Threatened Species 2008: e.T12392A3339343. http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T12392A3339343.en. Downloaded on 24 February 2019.

Bouché, P., Douglas-Hamilton, I., Wittemyer, G., Nianogo, A.J., Doucet, J.-L., Lejeune, P., and Vermeulen, C. (2011). Will elephants soon disappear from West African savannahs? PLoS One *6*, e20619.

Bourgeois, S., Senn, H., Kaden, J., Taggart, J.B., Ogden, R., Jeffery, K.J., Bunnefeld, N., Abernethy, K., and McEwing, R. (2018). Single-nucleotide polymorphism discovery and panel characterization in the African forest elephant. Ecology and Evolution *8*, 2207–2217.

Bovine HapMap Consortium, B.H. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science *324*, 528–532.

Brandt, A.L., Ishida, Y., Georgiadis, N.J., and Roca, A.L. (2012). Forest elephant mitochondrial genomes reveal that elephantid diversification in Africa tracked climate transitions. Molecular Ecology *21*, 1175–1189.

Brinkman, T.J., Schwartz, M.K., Person, D.K., Pilgrim, K.L., and Hundertmark, K.J. (2010). Effects of time and rainfall on PCR success using DNA extracted from deer fecal pellets. Conservation Genetics *11*, 1547–1552.

Brown, J.K., Taggart, J.B., Bekaert, M., Wehner, S., Palaiokostas, C., Setiawan, A.N., Symonds, J.E., and Penman, D.J. (2016). Mapping the sex determination locus in the hāpuku (*Polyprion oxygeneios*) using ddRAD sequencing. BMC Genomics *17*, 448.

Burham, K.P., and Anderson, D.R. (2002). Model selection and multimodel inference: A practical information-theoretic approach (Springer Science & Business Media).

Cain, B., Watts, P.C., Wandera, A.B., Stevens-Wood, B., and Shawcross, S.G. (2013). A reliable, single-step method for gender determination in black rhinoceros from low-copy template DNA. Conservation Genetics Resources *5*, 655–657.

Campbell, N.R., and Narum, S.R. (2009). Quantitative PCR assessment of microsatellite and SNP genotyping with variable quality DNA extracts. Conservation Genetics *10*, 779–784.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology *22*, 3124–3140.

Chang, H.-W., Gu, D.-L., Su, S.-H., Chang, C.-C., Cheng, C.-A., Huang, H.-W., Yao, C.-T., Chou, T.-C., Chuang, L.-Y., and Cheng, C.-C. (2008). High-throughput gender identification of Accipitridae eagles with real-time PCR using TaqMan probes. Theriogenology *70*, 83–90.

Chapman, D.D., Pinhal, D., and Shivji, M.S. (2009). Tracking the fin trade: genetic stock identification in western Atlantic scalloped hammerhead sharks *Sphyrna lewini*. Endangered Species Research *9*, 221–228.

Chase, M.J., Schlossberg, S., Griffin, C.R., Bouché, P.J., Djene, S.W., Elkan, P.W., Ferreira, S., Grossman, F., Kohi, E.M., and Landen, K. (2016). Continent-wide survey reveals massive decline in African savannah elephants. PeerJ *4*, e2354.

Chattopadhyay, B., Garg, K.M., Kumar, A.V., Doss, D.P.S., Rheindt, F.E., Kandula, S., and Ramakrishnan, U. (2016). Genome-wide data reveal cryptic diversity and genetic introgression in an Oriental cynopterine fruit bat radiation. BMC Evolutionary Biology *16*, 41.

Chelliah, K., Bukka, H., and Sukumar, R. (2013). Modeling harvest rates and numbers from age and sex ratios: A demonstration for elephant populations. Biological Conservation *165*, 54–61.

Chiou, K.L., and Bergey, C.M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Scientific Reports *8*, 1975.

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., and Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. Genome Research *15*, 1496–1502.

Comstock, K.E., Ostrander, E.A., and Wasser, S.K. (2003). Amplifying nuclear and mitochondrial DNA from African elephant ivory: a tool for monitoring the ivory trade. Conservation Biology *17*, 1840–1843.

Convention on International Trade in Endangered Species (2014). Elephant conservation, illegal killing and ivory trade (Geneva, Switzerland: 65th Meeting of the Standing Committee).

Convention on International Trade in Endangered Species (2017). Status of elephant populations, levels of illegal killing and the trade in ivory: A report to the CITES standing committee (Geneva, Switzerland: 69th Meeting of the Standing Committee).

Corlett, R.T. (2007). The impact of hunting on the mammalian fauna of tropical Asian forests. Biotropica *39*, 292–303.

Costa, V., Rosenbom, S., Monteiro, R., O'Rourke, S.M., and Beja-Pereira, A. (2017). Improving DNA quality extracted from fecal samples—a method to improve DNA yield. European Journal of Wildlife Research *63*, 3.

Cruz, V.P., Vera, M., Pardo, B.G., Taggart, J., Martinez, P., Oliveira, C., and Foresti, F. (2016). Identification and validation of single nucleotide polymorphisms as tools to

detect hybridization and population structure in freshwater stingrays. Molecular Ecology Resources.

Cullingham, C.I., Curteanu, M., Ball, M.C., and Manseau, M. (2010). Feasibility and recommendations for swift fox fecal DNA profiling. Journal of Wildlife Management *74*, 849–859.

Cushman, S.A., and Landguth, E.L. (2010). Scale dependent inference in landscape genetics. Landscape Ecology *25*, 967–979.

Dastjerdi, A., Robert, C., and Watson, M. (2014). Low coverage sequencing of two Asian elephant (*Elephas maximus*) genomes. GigaScience *3*, 12.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics *12*, 499–510.

Dawnay, N., Ogden, R., Thorpe, R.S., Pope, L.C., Dawson, D.A., and McEwing, R. (2008). A forensic STR profiling system for the Eurasian badger: a framework for developing profiling systems for wildlife species. Forensic Science International: Genetics *2*, 47–53.

Dawnay, N., Ogden, R., Wetton, J.H., Thorpe, R.S., and McEwing, R. (2009). Genetic data from 28 STR loci for forensic individual identification and parentage analyses in 6 bird of prey species. Forensic Science International: Genetics *3*, e63–e69.

De Flamingh, A., Sole, C.L., and Van Aarde, R.J. (2015). Genetic evidence for spatial structuring in a continuous African elephant (*Loxodonta africana*) population. Conservation Genetics *16*, 613–623.

Debruyne, R. (2005). A case study of apparent conflict between molecular phylogenies: the interrelationships of African elephants. Cladistics *21*, 31–50.

Degen, B., Ward, S.E., Lemes, M.R., Navarro, C., Cavers, S., and Sebbenn, A.M. (2013). Verifying the geographic origin of mahogany (*Swietenia macrophylla King*) with DNA-fingerprints. Forensic Science International: Genetics *7*, 55–62.

Degen, B., Blanc-Jolivet, C., Stierand, K., and Gillet, E. (2017). A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. Forensic Science International: Genetics *27*, 132–141.

Dimauro, C., Nicoloso, L., Cellesi, M., Macciotta, N.P.P., Ciani, E., Moioli, B., Pilla, F., and Crepaldi, P. (2015). Selection of discriminant SNP markers for breed and geographic assignment of Italian sheep. Small Ruminant Research *128*, 27–33.

Dirzo, R., and Raven, P.H. (2003). Global state of biodiversity and loss. Annual Review of Environment and Resources *28*, 137–167.

Dormontt, E.E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J.C., and Koch, G. (2015). Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biological Conservation *191*, 790–798.

Douglas-Hamilton, I. (1987). African elephants: population trends and their causes. Oryx *21*, 11–24.

Douglas-Hamilton, I. (2009). The current elephant poaching trend. Pachyderm *45*, 154–157.

Earl, D.A., and vonHoldt, B. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources *4*, 359–361.

Eaton, M.J., Meyers, G.L., Kolokotronis, S.-O., Leslie, M.S., Martin, A.P., and Amato, G. (2010). Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. Conservation Genetics *11*, 1389–1404.

Ebert, C., Knauer, F., Spielberger, B., Thiele, B., and Hohmann, U. (2012). Estimating wild boar Sus scrofa population size using faecal DNA and capture-recapture modelling. Wildlife Biology *18*, 142–152.

Eggert, L.S., Eggert, J.A., and Woodruff, D.S. (2003). Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Molecular Ecology *12*, 1389–1402.

Eggert, L.S., Buij, R., Lee, M.E., Campbell, P., Dallmeier, F., Fleischer, R.C., Alonso, A., and Maldonado, J.E. (2014). Using genetic profiles of African forest elephants to infer population structure, movements, and habitat use in a conservation and development landscape in Gabon. Conservation Biology *28*, 107–118.

Elephant genome project (2009). LoxAfr 3.0, Genbank Assembly ID: GCA_000001905.1.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics *5*, 435–445.

Espinoza, E.O., and Mann, M.-J. (1993). The history and significance of the Schreger pattern in proboscidean ivory characterization. Journal of the American Institute for Conservation *32*, 241–248.

Fabbri, E., Caniglia, R., Mucci, N., Thomsen, H.P., Krag, K., Pertoldi, C., Loeschcke, V., and Randi, E. (2012). Comparison of single nucleotide polymorphisms and microsatellites in non-invasive genetic monitoring of a wolf population. Archives of Biological Sciences *64*, 321–335.

Fernando, P., and Melnick, D.J. (2001). Molecular sexing eutherian mammals. Molecular Ecology Resources *1*, 350–353.

Fernando, P., Pfrender, M.E., Encalada, S.E., and Lande, R. (2000). Mitochondrial DNA variation, phylogeography and population structure of the Asian elephant. Heredity *84*, 362–372.

Fernando, P., Vidya, T.N.C., Rajapakse, C., Dangolla, A., and Melnick, D.J. (2003). Reliable noninvasive genotyping: fantasy or reality? Journal of Heredity *94*, 115–123.

Fishlock, V., and Lee, P.C. (2013). Forest elephants: fission–fusion and social arenas. Animal Behaviour *85*, 357–363.

Fitak, R.R., Naidu, A., Thompson, R.W., and Culver, M. (2015). A new panel of SNP markers for the individual identification of North American pumas. Journal of Fish and Wildlife Management *7*, 13–27.

Fountain, E.D., Pauli, J.N., Reid, B.N., Palsbøll, P.J., and Peery, M.Z. (2016). Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. Molecular Ecology Resources *16*, 966–978.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software *27*, 233–249.

Frantz, A.C., Pourtois, J.T., Heuertz, M., Schley, L., Flamand, M.-C., Krier, A., Bertouille, S., Chaumont, F., and Burke, T. (2006). Genetic structure and assignment tests demonstrate illegal translocation of red deer (*Cervus elaphus*) into a continuous population. Molecular Ecology *15*, 3191–3203.

Frantzen, M.A.J., Silk, J.B., Ferguson, J.W.H., Wayne, R.K., and Kohn, M.H. (1998). Empirical evaluation of preservation methods for faecal DNA. Molecular Ecology *7*, 1423–1428.

Funk, W.C., McKay, J.K., Hohenlohe, P.A., and Allendorf, F.W. (2012). Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution *27*, 489–496.

Gao, Y., and Clark, S.G. (2014). Elephant ivory trade in China: Trends and drivers. Biological Conservation *180*, 23–30.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J.-M., and Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular Ecology *22*, 3165–3178.

Gettings, K.B., Kiesler, K.M., and Vallone, P.M. (2015). Performance of a next generation sequencing SNP assay on degraded DNA. Forensic Science International: Genetics *19*, 1–9.

Ghobrial, L., Lankester, F., Kiyang, J.A., Akih, A.E., De Vries, S., Fotso, R., Gadsby, E.L., Jenkins, P.D., and Gonder, M.K. (2010). Tracing the origins of rescued chimpanzees reveals widespread chimpanzee hunting in Cameroon. BMC Ecology *10*, 2.

Glover, K.A., Hansen, M.M., Lien, S., Als, T.D., Høyheim, B., and Skaala, Ø. (2010). A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. Bmc Genetics *11*, 2.

Goossens, B., Sharma, R., Othman, N., Kun-Rodrigues, C., Sakong, R., Ancrenaz, M., Ambu, L.N., Jue, N.K., O'Neill, R.J., Bruford, M.W., et al. (2016). Habitat

fragmentation and genetic diversity in natural populations of the Bornean elephant: Implications for conservation. Biological Conservation *196*, 80–92.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes *5*, 184–186.

Gray, T.N., Vidya, T.N.C., Potdar, S., Bharti, D.K., and Sovanna, P. (2014). Population size estimation of an Asian elephant population in eastern Cambodia through non-invasive mark-recapture sampling. Conservation Genetics *15*, 803–810.

Grubb, P., Groves, C.P., Dudley, J.P., and Shoshani, J. (2000). Living African elephants belong to two species: *Loxodonta africana* (Blumenbach, 1797) and *Loxodonta cyclotis* (Matschie, 1900). Elephant *2*, 3.

Guérin, S.M. (2010). Avorio d'ogni ragione: the supply of elephant ivory to northern Europe in the Gothic era. Journal of Medieval History *36*, 156–174.

Gugala, N.A., Ishida, Y., Georgiadis, N.J., and Roca, A.L. (2016). Development and characterization of microsatellite markers in the African forest elephant (*Loxodonta cyclotis*). BMC Research Notes *9*, 364.

Gupta, S.K., Thangaraj, K., and Singh, L. (2006). A simple and inexpensive molecular method for sexing and identification of the forensic samples of elephant origin. Journal of Forensic Sciences *51*, 805–807.

Gupta, S.K., Bhagavatula, J., Thangaraj, K., and Singh, L. (2011a). Establishing the identity of the massacred tigress in a case of wildlife crime. Forensic Science International: Genetics *5*, 74–75.

Gupta, S.K., Thangaraj, K., and Singh, L. (2011b). Identification of the source of ivory idol by DNA analysis. Journal of Forensic Sciences *56*, 1343–1345.

Gupta, S.K., Kumar, A., and Nigam, P. (2011c). Multilocus Genotyping of Asian Elephant Ivory: A Case Study in Suspected Wildlife Crime, J. Forensic Res. S *1*, 002–04.

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*, 290–300.

Hakansson, N.T. (2004). The human ecology of world systems in East Africa: the impact of the ivory trade. Human Ecology *32*, 561–591.

Haken, J. (2011). Transnational crime in the developing world (Global Financial Integrity).

Hausknecht, R., Bayerl, H., Gula, R., and Kuehn, R. (2010). Application of quantitative real-time polymerase chain reaction for noninvasive genetic monitoring. Journal of Wildlife Management *74*, 1904–1910.

Hayaishi, S., and Kawamoto, Y. (2006). Low genetic diversity and biased distribution of mitochondrial DNA haplotypes in the Japanese macaque (*Macaca fuscata yakui*) on Yakushima Island. Primates *47*, 158–164.

Hedges, S. (2012). Monitoring elephant populations and assessing threats: a manual for researchers, managers and conservationists (Universities Press (India)).

Hedges, S., and Lawson, D. (2006). Dung survey standards for the MIKE programme.

Hedges, S., Johnson, A., Ahlering, M., Tyson, M., and Eggert, L.S. (2013). Accuracy, precision, and cost-effectiveness of conventional dung density and fecal DNA based survey methods to estimate Asian elephant (*Elephas maximus*) population size and structure. Biological Conservation *159*, 101–108.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T., Cariani, A., Maes, G.E., Diopere, E., and Carvalho, G.R. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular Ecology Resources *11*, 123–136.

Hess, J.E., Matala, A.P., and Narum, S.R. (2011). Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Molecular Ecology Resources *11*, 137–149.

Hiremath, P.J., Kumar, A., Penmetsa, R.V., Farmer, A., Schlueter, J.A., Chamarthi, S.K., Whaley, A.M., Carrasquilla-Garcia, N., Gaur, P.M., Upadhyaya, H.D., et al. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnology Journal *10*, 716–732.

Holderegger, R., and Wagner, H.H. (2008). Landscape genetics. AIBS Bulletin *58*, 199–207.

Höss, M., Kohn, M., Pääbo, S., Knauer, F., and Schröder, W. (1992). Excrement analysis by PCR. Nature *359*, 199.

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*, 1322–1332.

Huffman, J.E., and Wallace, J.R. (2012). Wildlife forensics: methods and applications (John Wiley & Sons).

Hymas, O.R. (2016). L'Okoumé, fils du manioc: Post-logging in remote rural forest areas of Gabon and its long-term impacts on development and the environment. PhD Thesis. UCL (University College London).

Ishida, Y., Demeke, Y., van Coeverden de Groot, P.J., Georgiadis, N.J., Leggett, K.E., Fox, V.E., and Roca, A.L. (2011). Distinguishing forest and savanna African elephants using short nuclear DNA sequences. Journal of Heredity esr073.

Ishida, Y., Demeke, Y., van Coeverden de Groot, P.J., Georgiadis, N.J., Leggett, K.E., Fox, V.E., and Roca, A.L. (2012). Short amplicon microsatellite markers for low quality elephant DNA. Conservation Genetics Resources *4*, 491–494.

Ishida, Y., Georgiadis, N.J., Hondo, T., and Roca, A.L. (2013). Triangulating the provenance of African elephants using mitochondrial DNA. Evolutionary Applications *6*, 253–265.

Ishida, Y., Gugala, N.A., Georgiadis, N.J., and Roca, A.L. (2018). Evolutionary and demographic processes shaping geographic patterns of genetic diversity in a keystone species, the African forest elephant (*Loxodonta cyclotis*). Ecology and Evolution *8*, 4919–4931.

Iyengar, A. (2014). Forensic DNA analysis for animal protection and biodiversity conservation: a review. Journal for Nature Conservation *22*, 195–205.

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*, 281–294.

Johnson, R.N. (2010). The use of DNA identification in prosecuting wildlife-traffickers in Australia: do the penalties fit the crimes? Forensic Science, Medicine, and Pathology *6*, 211–216.

Johnson, M.B., Clifford, S.L., Goossens, B., Nyakaana, S., Curran, B., White, L.J., Wickings, E.J., and Bruford, M.W. (2007). Complex phylogeographic history of central African forest elephants and its implications for taxonomy. BMC Evolutionary Biology *7*, 244.

Johnson, W.E., Torres-Florez, J.P., Galetti Jr, P.M., Godoy, J.A., and Marín, J.C. (2016). Integrated genetic approaches for the conservation of tropical and subtropical species. In Tropical Conservation: Perspectives on Local and Global Priorities, (Aguirre AA, Sukumar, R), pp. 226–237.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics *24*, 1403–1405.

Jombart, T., Devillard, S., Dufour, A.B., and Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity *101*, 92.

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics *11*, 94.

Jones, O.R., and Wang, J. (2012). A comparison of four methods for detecting weak genetic structure from marker data. Ecology and Evolution *2*, 1048–1055.

Kanthaswamy, S., Premasuthan, A., Ng, J., Satkoski, J., and Goyal, V. (2012). Quantitative real-time PCR (qPCR) assay for human–dog–cat species identification and nuclear DNA quantification. Forensic Science International: Genetics *6*, 290–295.

Kasai, F., O'Brien, P.C., and Ferguson-Smith, M.A. (2013). Afrotheria genome; overestimation of genome size and distinct chromosome GC content revealed by flow karyotyping. Genomics *102*, 468–471.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., and Duran, C. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Kidd, K.K., Pakstis, A.J., Speed, W.C., Grigorenko, E.L., Kajuna, S.L., Karoma, N.J., Kungulilo, S., Kim, J.-J., Lu, R.-B., Odunsi, A., et al. (2006). Developing a SNP panel for forensic identification of individuals. Forensic Science International *164*, 20–32.

Kitpipit, T., Thongjued, K., Penchart, K., Ouithavon, K., and Chotigeat, W. (2017). Mini-SNaPshot multiplex assays authenticate elephant ivory and simultaneously identify the species origin. Forensic Science International: Genetics *27*, 106–115.

Kjeldsen, S.R., Zenger, K.R., Leigh, K., Ellis, W., Tobey, J., Phalen, D., Melzer, A., FitzGibbon, S., and Raadsma, H.W. (2016). Genome-wide SNP loci reveal novel insights into koala (*Phascolarctos cinereus*) population variability across its range. Conservation Genetics *17*, 337–353.

Kohn, M.H., and Wayne, R.K. (1997). Facts from feces revisited. Trends in Ecology & Evolution *12*, 223–227.

Krishnasamy, K., Milliken, T., and Savini, C. (2016). In Transition: Bangkok's Ivory Market - An 18-month survey of Bangkok's ivory market (Petaling Jaya, Selangor, Malaysia: TRAFFIC, Southeast Asia Regional Office).

Kumar, S., Banks, T.W., and Cloutier, S. (2012). SNP discovery through next-generation sequencing and its applications. International Journal of Plant Genomics *2012*.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution *33*, 1870–1874.

Laikre, L., Palm, S., and Ryman, N. (2005). Genetic population structure of fishes: implications for coastal zone management. AMBIO: A Journal of the Human Environment *34*, 111–119.

Lampa, S., Gruber, B., Henle, K., and Hoehn, M. (2008). An optimisation approach to increase DNA amplification success of otter faeces. Conservation Genetics *9*, 201.

Lampa, S., Henle, K., Klenke, R., Hoehn, M., and Gruber, B. (2013). How to overcome genotyping errors in non-invasive genetic mark-recapture population size estimation—A review of available methods illustrated by a case study. The Journal of Wildlife Management *77*, 1490–1511.

Landguth, E.L., and Balkenhol, N. (2012). Relative sensitivity of neutral versus adaptive genetic data for assessing population differentiation. Conservation Genetics *13*, 1421–1426.

Laporte, N.T., Stabach, J.A., Grosch, R., Lin, T.S., and Goetz, S.J. (2007). Expansion of industrial logging in Central Africa. Science *316*, 1451–1451.

Lee, J.C.-I., Hsieh, H.-M., Huang, L.-H., Kuo, Y.-C., Wu, J.-H., Chin, S.-C., Lee, A.-H., Linacre, A., and Tsai, L.-C. (2009). Ivory identification by DNA profiling of cytochrome b gene. International Journal of Legal Medicine *123*, 117–121.

Leigh, J.W., and Bryant, D. (2015). popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution *6*, 1110–1116.

Lemieux, A.M., and Clarke, R.V. (2009). The international ban on ivory sales and its effects on elephant poaching in Africa. The British Journal of Criminology *49*, 451–471.

LGC Genomics (2013). KASP genotyping chemistry User guide and manual.

LGC Genomics (2014). KASP assay design. Guide to submission of sequence information.

Linacre, A., and Tobe, S.S. (2011). An overview to the investigative approach to species testing in wildlife forensic science. Investigative Genetics *2*, 2.

Lorenzini, R., Cabras, P., Fanelli, R., and Carboni, G.L. (2011). Wildlife molecular forensics: identification of the Sardinian mouflon using STR profiling and the Bayesian assignment test. Forensic Science International: Genetics *5*, 345–349.

Mailand, C., and Wasser, S.K. (2007). Isolation of DNA from small amounts of elephant ivory. Nature Protocols *2*, 2228.

Maisels, F., Strindberg, S., Blake, S., Wittemyer, G., Hart, J., Williamson, E.A., Aba'a, R., Abitsi, G., Ambahe, R.D., Amsini, F., et al. (2013). Devastating decline of forest elephants in Central Africa. PloS One *8*, e59469.

Manel, S., Berthier, P., and Luikart, G. (2002). Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. Conservation Biology *16*, 650–659.

Manel, S., Gaggiotti, O.E., and Waples, R.S. (2005). Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology & Evolution *20*, 136–142.

Manousaki, T., Tsakogiannis, A., Taggart, J.B., Palaiokostas, C., Tsaparis, D., Lagnel, J., Chatziplis, D., Magoulas, A., Papandroulakis, N., Mylonas, C.C., et al. (2016). Exploring a Nonmodel Teleost Genome Through RAD Sequencing—Linkage Mapping in Common Pandora, *Pagellus erythrinus* and Comparative Genomic Analysis. G3: Genes, Genomes, Genetics *6*, 509–519.

Martin, E.B., and Stiles, D. (2002). The ivory markets of East Asia (Save the Elephants Nairobi, Kenya).

Matejusová, I., Bland, F., Hall, A.J., Harris, R.N., Snow, M., Douglas, A., and Middlemas, S.J. (2013). Real-time PCR assays for the identification of harbor and gray

seal species and sex: A molecular tool for ecology and management. Marine Mammal Science *29*, 186–194.

Matheson, C.D., Gurney, C., Esau, N., and Lehto, R. (2010). Assessing PCR inhibition from humic substances. The Open Enzyme Inhibition Journal *3*, 38–45.

Maudet, C., Luikart, G., Dubray, D., Von Hardenberg, A., and Taberlet, P. (2004). Low genotyping error rates in wild ungulate faeces sampled in winter. Molecular Ecology Notes *4*, 772–775.

McEwing, R., and Ahlers, N. (2016). Is wildlife DNA forensics delivering as an illegal trade enforcement tool? TRAFFIC Bulletin *28*, 56–58.

Mesnick, S.L., Taylor, B.L., Archer, F.I., Martien, K.K., Treviño, S.E., Hancock-Hanser, B.L., Moreno Medina, S.C., Pease, V.L., Robertson, K.M., and Straley, J.M. (2011). Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. Molecular Ecology Resources *11*, 278–298.

Michelmore, F., Beardsley, K., Barnes, R.F.W., and Douglas-Hamilton, I. (1994). A model illustrating the changes in forest elephant numbers caused by poaching. African Journal of Ecology *32*, 89–99.

Milliken, T. (2013). Progress in implementing the Elephant Trade Information System (ETIS)/Avancement dans la mise en oeuvre du Système d'Information sur le Trafic des éléphants (ETIS). Pachyderm *1*, 85–90.

Mondol, S., Karanth, K.U., Kumar, N.S., Gopalaswamy, A.M., Andheria, A., and Ramakrishnan, U. (2009). Evaluation of non-invasive genetic sampling methods for estimating tiger population size. Biological Conservation *142*, 2350–2360.

Mondol, S., Mailand, C.R., and Wasser, S.K. (2014). Male biased sex ratio of poached elephants is negatively related to poaching intensity over time. Conservation Genetics *15*, 1259–1263.

Mondol, S., Sridhar, V., Yadav, P., Gubbi, S., and Ramakrishnan, U. (2015). Tracing the geographic origin of traded leopard body parts in the Indian subcontinent with DNA-based assignment tests. Conservation Biology *29*, 556–564.

Montano, V., and Jombart, T. (2017). An Eigenvalue test for spatial principal component analysis. BMC Bioinformatics *18*, 562.

Monteiro, L., Gras, N., Vidal, R., Cabrita, J., and Mégraud, F. (2001). Detection of *Helicobacter pylori* DNA in human feces by PCR: DNA stability and removal of inhibitors. Journal of Microbiological Methods *45*, 89–94.

Morgan, B.J., and Lee, P.C. (2007). Forest elephant group composition, frugivory and coastal use in the Réserve de Faune du Petit Loango, Gabon. African Journal of Ecology *45*, 519–526.

Morin, P.A., Moore, J.J., Chakraborty, R., Jin, L., Goodall, J., and Woodruff, D.S. (1994). Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science *265*, 1193–1193.

Morin, P.A., Chambers, K.E., Boesch, C., and Vigilant, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Molecular Ecology *10*, 1835–1844.

Morin, P.A., Luikart, G., Wayne, R.K., and others (2004). SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution *19*, 208–216.

Morin, P.A., Nestler, A., Rubio-Cisneros, N.T., Robertson, K.M., and Mesnick, S.L. (2005). Interfamilial characterization of a region of the ZFX and ZFY genes facilitates sex determination in cetaceans and other mammals. Molecular Ecology *14*, 3275–3286.

Morin, P.A., Hedrick, N.M., Robertson, K.M., and Leduc, C.A. (2007). Comparative mitochondrial and nuclear quantitative PCR of historical marine mammal tissue, bone, baleen, and tooth samples. Molecular Ecology Notes *7*, 404–411.

Morin, P.A., Martien, K.K., and Taylor, B.L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. Molecular Ecology Resources *9*, 66–73.

Munshi-South, J. (2011). Relatedness and demography of African forest elephants: inferences from noninvasive fecal DNA analyses. Journal of Heredity *102*, 391–398.

Munshi-South, J., Tchignoumba, L., Brown, J., Abbondanza, N., Maldonado, J., Henderson, A., and Alonso, A. (2008). Physiological indicators of stress in African forest elephants (*Loxodonta africana cyclotis*) in relation to petroleum operations in Gabon, Central Africa. Diversity and Distributions *14*, 995–1004.

Murphy, M.A., Waits, L.P., Kendall, K.C., Wasser, S.K., Higbee, J.A., and Bogden, R. (2002). An evaluation of long-term preservation methods for brown bear (*Ursus arctos*) faecal DNA samples. Conservation Genetics *3*, 435–440.

Murphy, M.A., Kendall, K.C., Robinson, A., and Waits, L.P. (2007). The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. Conservation Genetics *8*, 1219–1224.

Murray-Dickson, G., Ghazali, M., Ogden, R., Brown, R., and Auliya, M. (2017). Phylogeography of the reticulated python (*Malayopython reticulatus ssp.*): Conservation implications for the worlds' most traded snake species. PloS One *12*, e0182049.

Ng, J., Satkoski, J., Premasuthan, A., and Kanthaswamy, S. (2014). A nuclear DNA-based species determination and DNA quantification assay for common poultry species. Journal of Food Science and Technology *51*, 4060–4065.

Nielsen, R. (2004). Population genetic analysis of ascertained SNP data. Human Genomics *1*, 218.

Nielsen, E.E., Cariani, A., Mac Aoidh, E., Maes, G.E., Milano, I., Ogden, R., Taylor, M., Hemmer-Hansen, J., Babbucci, M., and Bargelloni, L. (2012). Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nature Communications *3*, 851.

Nishihara, T. (2003). Elephant poaching and ivory trafficking in African tropical forests with special reference to the Republic of Congo. Pachyderm *34*, 66–74.

Nishihara, T. (2012). Demand for forest elephant ivory in Japan. Pachyderm *52*, 55–65.

Nkoke, S.C., Lagrot, J.F., Ringuet, S., and Milliken, T. (2017). Ivory Markets in Central Africa – Market Surveys in Cameroon, Central African Republic, Congo, Democratic Republic of the Congo and Gabon: 2007, 2009, 2014/2015 (Yaoundé, Cameroon and Cambridge, UK: TRAFFIC).

Norman, A.J., and Spong, G. (2015). Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. Ecology and Evolution *5*, 3056–3065.

Nsubuga, A.M., Robbins, M.M., Roeder, A.D., Morin, P.A., Boesch, C., and Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. Molecular Ecology *13*, 2089–2094.

Nussberger, B., Wandeler, P., and Camenisch, G. (2014). A SNP chip to detect introgression in wildcats allows accurate genotyping of single hairs. European Journal of Wildlife Research *60*, 405–410.

Nyakaana, S., Okello, J., Muwanika, V., and Siegismund, H.R. (2005). Six new polymorphic microsatellite loci isolated and characterized from the African savannah elephant genome. Molecular Ecology Notes *5*, 223–225.

Ogden, R. (2010). Forensic science, genetics and wildlife biology: getting the right mix for a wildlife DNA forensics lab. Forensic Science, Medicine, and Pathology *6*, 172–179.

Ogden, R. (2011). Unlocking the potential of genomic technologies for wildlife forensics. Molecular Ecology Resources *11*, 109–116.

Ogden, R., and Linacre, A. (2015). Wildlife forensic science: a review of genetic geographic origin assignment. Forensic Science International: Genetics *18*, 152–159.

Ogden, R., Dawnay, N., and McEwing, R. (2009). Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. Endangered Species Research *9*, 179–195.

Okello, J.B., Masembe, C., Rasmussen, H.B., Wittemyer, G., Omondi, P., Kahindi, O., Muwanika, V.B., Arctander, P., Douglas-Hamilton, I., and Nyakaana, S. (2008). Population genetic structure of savannah elephants in Kenya: conservation and management implications. Journal of Heredity *99*, 443–452.

Paetkau, D. (2003). An empirical exploration of data quality in DNA-based population inventories. Molecular Ecology *12*, 1375–1387.

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*, 55–65.

Pakstis, A.J., Speed, W.C., Fang, R., Hyland, F.C., Furtado, M.R., Kidd, J.R., and Kidd, K.K. (2010). SNPs for a universal individual identification panel. Human Genetics *127*, 315–324.

Panasci, M., Ballard, W.B., Breck, S., Rodriguez, D., Densmore, L.D., Wester, D.B., and Baker, R.J. (2011). Evaluation of fecal DNA preservation techniques and effects of sample age and diet on genotyping success. The Journal of Wildlife Management *75*, 1616–1624.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics *26*, 419–420.

Pelak, K., Shianna, K.V., Ge, D., Maia, J.M., Zhu, M., Smith, J.P., Cirulli, E.T., Fellay, J., Dickson, S.P., Gumbs, C.E., et al. (2010). The characterization of twenty sequenced human genomes. PLoS Genetics *6*, e1001111.

Perry, G.H., Marioni, J.C., Melsted, P., and Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. Molecular Ecology *19*, 5332–5344.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS One *7*, e37135.

Piggott, M.P. (2005). Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. Wildlife Research *31*, 485–493.

Pimm, S.L., Russell, G.J., Gittleman, J.L., and Brooks, T.M. (1995). The future of biodiversity. Science *269*, 347–350.

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*, 536–539.

Poulakakis, N., Antoniou, A., Mantziou, G., Parmakelis, A., Skartsi, T., Vasilakis, D., Elorriaga, J., De La Puente, J., Gavashelishvili, A., and Ghasabyan, M. (2008). Population structure, diversity, and phylogeography in the near-threatened Eurasian black vultures *Aegypius monachus* (Falconiformes; Accipitridae) in Europe: insights from microsatellite and mitochondrial DNA variation. Biological Journal of the Linnean Society *95*, 859–872.

Poulsen, J.R., Koerner, S.E., Moore, S., Medjibe, V.P., Blake, S., Clark, C.J., Akou, M.E., Fay, M., Meier, A., Okouyi, J., et al. (2017). Poaching empties critical Central African wilderness of forest elephants. Current Biology *27*, R134–R135.

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

Puckett, E.E., and Eggert, L.S. (2016). Comparison of SNP and microsatellite genotyping panels for spatial assignment of individuals to natal range: a case study using the American black bear (*Ursus americanus*). Biological Conservation *193*, 86–93.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics *81*, 559–575.

Quasim, S., McDonald, A.J., and Sarre, S.D. (2018). Towards more efficient large-scale DNA-based detection of terrestrial mammal predators from scats. Mammal Research *63*, 387–393.

Queloz, V., Duo, A., Sieber, T.N., and Grünig, C.R. (2010). Microsatellite size homoplasies and null alleles do not affect species diagnosis and population genetic analysis in a fungal species complex. Molecular Ecology Resources *10*, 348–367.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramón-Laca, A., Soriano, L., Gleeson, D., and Godoy, J.A. (2015). A simple and effective method for obtaining mammal DNA from faeces. Wildlife Biology *21*, 195–203.

Ranade, K., Chang, M.-S., Ting, C.-T., Pei, D., Hsiao, C.-F., Olivier, M., Pesich, R., Hebert, J., Chen, Y.-D.I., Dzau, V.J., et al. (2001). High-throughput genotyping with single nucleotide polymorphisms. Genome Research *11*, 1262–1268.

Rannala, B., and Mountain, J.L. (1997). Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences *94*, 9197–9201.

Ratchford, M., Allgood, B., and Todd, P. (2013). Criminal Nature-The Global Security Implications of the Illegal Wildlife Trade (International Fund for Animal Welfare (IFAW), Washington DC).

Reddy, P.A., Bhavanishankar, M., Bhagavatula, J., Harika, K., Mahla, R.S., and Shivaji, S. (2012). Improved methods of carnivore faecal sample preservation, DNA extraction and quantification for accurate genotyping of wild tigers. PLoS One *7*, e46732.

Renan, S., Speyer, E., Shahar, N., Gueta, T., Templeton, A.R., and Bar-David, S. (2012). A factorial design experiment as a pilot study for noninvasive genetic sampling. Molecular Ecology Resources *12*, 1040–1047.

Ribaux, O., Walsh, S.J., and Margot, P. (2006). The contribution of forensic science to crime analysis and investigation: forensic intelligence. Forensic Science International *156*, 171–181.

Roca, A.L., Georgiadis, N., Pecon-Slattery, J., and O'Brien, S.J. (2001). Genetic evidence for two species of elephant in Africa. Science *293*, 1473–1477.

Roca, A.L., Georgiadis, N., and O'Brien, S.J. (2007). Cyto-nuclear genomic dissociation and the African elephant species question. Quaternary International *169*, 4–16.

Roca, A.L., Ishida, Y., Brandt, A.L., Benjamin, N.R., Zhao, K., and Georgiadis, N.J. (2015). Elephant natural history: a genomic perspective. Annu. Rev. Anim. Biosci. *3*, 139–167.

Rodgers, T.W., and Janečka, J.E. (2013). Applications and techniques for non-invasive faecal genetics research in felid conservation. European Journal of Wildlife Research *59*, 1–16.

Roesti, M., Salzburger, W., and Berner, D. (2012). Uninformative polymorphisms bias genome scans for signatures of selection. BMC Evolutionary Biology *12*, 94.

Rohland, N., Reich, D., Mallick, S., Meyer, M., Green, R.E., Georgiadis, N.J., Roca, A.L., and Hofreiter, M. (2010). Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. PLoS Biology *8*, e1000564.

Rosen, G.E., and Smith, K.F. (2010). Summarizing the evidence on the international trade in illegal wildlife. EcoHealth *7*, 24–32.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In Bioinformatics Methods and Protocols, (Humana Press, Totowa, NJ), pp. 365–386.

Rutledge, L.Y., Holloway, J.J., Patterson, B.R., and White, B.N. (2009). An improved field method to obtain DNA for individual identification from wolf scat. Journal of Wildlife Management *73*, 1430–1435.

Sampson, C., McEvoy, J., Oo, Z.M., Chit, A.M., Chan, A.N., Tonkyn, D., Soe, P., Songer, M., Williams, A.C., and Reisinger, K. (2018). New elephant crisis in Asia—Early warning signs from Myanmar. PloS One *13*, e0194113.

Santini, A., Lucchini, V., Fabbri, E., and Randi, E. (2007). Ageing and environmental factors affect PCR success in wolf (*Canis lupus*) excremental DNA samples. Molecular Ecology Resources *7*, 955–961.

Sarre, S.D., and Georges, A. (2009). Genetics in conservation and wildlife management: a revolution since Caughley. Wildlife Research *36*, 70–80.

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR inhibitors–occurrence, properties and removal. Journal of Applied Microbiology *113*, 1014–1026.

Schultz, A.J., Cristescu, R.H., Littleford-Colquhoun, B.L., Jaccoud, D., and Frère, C.H. (2018). Fresh is best: Accurate SNP genotyping from koala scats. Ecology and Evolution *8*, 3139–3151.

Schuttler, S.G., Philbrick, J.A., Jeffery, K.J., and Eggert, L.S. (2014). Fine-scale genetic structure and cryptic associations reveal evidence of kin-based sociality in the African forest elephant. PloS One *9*, e88074.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L.W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Molecular Ecology Resources *11*, 1–8.

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Molecular Breeding *33*, 1–14.

Senge, T., Madea, B., Junge, A., Rothschild, M.A., and Schneider, P.M. (2011). STRs, mini STRs and SNPs–a comparative study for typing degraded DNA. Legal Medicine *13*, 68–74.

Senn, H., Ogden, R., Cezard, T., Gharbi, K., Iqbal, Z., Johnson, E., Kamps-Hughes, N., Rosell, F., and McEwing, R. (2013). Reference-free SNP discovery for the Eurasian beaver from restriction site–associated DNA paired-end data. Molecular Ecology *22*, 3141–3150.

Sharma, R., Goossens, B., Kun-Rodrigues, C., Teixeira, T., Othman, N., Boone, J.Q., Jue, N.K., Obergfell, C., O'Neill, R.J., and Chikhi, L. (2012). Two different high throughput sequencing approaches identify thousands of *de novo* genomic markers for the genetically depleted Bornean elephant. PLoS One *7*, e49533.

Shin, J.-H., Blay, S., McNeney, B., Graham, J., and others (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. Journal of Statistical Software *16*, 1–10.

Shirley, M.H., and Austin, J.D. (2017). Did Late Pleistocene climate change result in parallel genetic structure and demographic bottlenecks in sympatric Central African crocodiles, Mecistops and Osteolaemus? Molecular Ecology *26*, 6463–6477.

Sigurḥardottir, S., Helgason, A., Gulcher, J.R., Stefansson, K., and Donnelly, P. (2000). The mutation rate in the human mtDNA control region. The American Journal of Human Genetics *66*, 1599–1609.

Skaug, H., Fournier, D., Nielsen, A., Magnusson, A., and Bolker, B. (2016). glmmadmb: Generalized linear mixed models using'ad model builder' (R package version 0.8.3.3).

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. Genetics *139*, 457–462.

Sodhi, N.S., Koh, L.P., Brook, B.W., and Ng, P.K. (2004). Southeast Asian biodiversity: an impending disaster. Trends in Ecology & Evolution *19*, 654–660.

Soto-Calderon, I.D., Ntie, S., Mickala, P., Maisels, F., Wickings, E.J., and Anthony, N.M. (2009). Effects of storage type and time on DNA amplification success in tropical ungulate faeces. Molecular Ecology Resources *9*, 471–479.

Spitzer, R., Norman, A.J., Schneider, M., and Spong, G. (2016). Estimating population size using single-nucleotide polymorphism-based pedigree data. Ecology and Evolution *6*, 3174–3184.

Stenglein, J.L., De Barba, M., Ausband, D.E., and Waits, L.P. (2010). Impacts of sampling location within a faeces on DNA quality in two carnivore species. Molecular Ecology Resources *10*, 109–114.

Sukumar, R. (2006). A brief review of the status, distribution and biology of wild Asian elephants Elephas maximus. International Zoo Yearbook *40*, 1–8.

Sukumar, R., Ramakrishnan, U., and Santosh, J.A. (1998). Impact of poaching on an Asian elephant population in Periyar, southern India: a model of demography and tusk harvest. In Animal Conservation Forum, (Cambridge University Press), pp. 281–291.

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*, 3189–3194.

Taberlet, P., Waits, L.P., and Luikart, G. (1999). Noninvasive genetic sampling: look before you leap. Trends in Ecology & Evolution *14*, 323–327.

von Thaden, A., Cocchiararo, B., Jarausch, A., Jüngling, H., Karamanlidis, A.A., Tiesmeyer, A., Nowak, C., and Muñoz-Fuentes, V. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. Scientific Reports *7*, 10768.

Thommasen, H.V., Thomson, M.J., Shutler, G.G., and Kirby, L.T. (1989). Development of DNA fingerprints for use in wildlife forensic science. Wildlife Society Bulletin (1973-2006) *17*, 321–326.

Thouless, C.R., Dublin, H.T., Blanc, J.J., Skinner, D.P., Daniel, T.E., Taylor, R.D., Maisels, F., Frederick, H., and Bouche, P. (2016). African Elephant Status Report 2016. An Update from the African Elephant Database.

Titeca, K. (2018). Illegal Ivory Trade as Transnational Organized Crime? An Empirical Study Into Ivory Traders In Uganda. The British Journal of Criminology.

Tobe, S.S., and Linacre, A. (2009). Identifying endangered species from degraded mixtures at low levels. Forensic Science International: Genetics Supplement Series *2*, 304–305.

Tobe, S.S., and Linacre, A.M. (2008). A technique for the quantification of human and non-human mammalian mitochondrial DNA copy number in forensic and other mixtures. Forensic Science International: Genetics *2*, 249–256.

Toews, D.P., and Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology *21*, 3907–3930.

TRAFFIC (2017). About 7.2 tonnes of ivory seized in Hong Kong.

Turkalo, A.K., Wrege, P.H., and Wittemyer, G. (2013). Long-term monitoring of Dzanga Bai forest elephants: forest clearing use patterns. PloS One *8*, e85154.

Turkalo, A.K., Wrege, P.H., and Wittemyer, G. (2017). Slow intrinsic growth rate in forest elephants indicates recovery from poaching will require decades. Journal of Applied Ecology *54*, 153–159.

Underwood, F.M., Burn, R.W., and Milliken, T. (2013). Dissecting the illegal ivory trade: an analysis of ivory seizures data. PloS One *8*, e76539.

UNEP, CITES, IUCN, and TRAFFIC (2013). Elephants in the dust—the African elephant crisis. A rapid response assessment (United Nations Environment Programme, GRID-Arendal Arendal).

United Nations Office on Drugs and Crime (2010). The Globalization of Crime: A Transnational Organized Crime Threat Assessment (United Nations Publications).

United Nations Office on Drugs and Crime (2013). Transnational organized crime in East Africa: a threat assessment (UNODC, Vienna).

United Nations Office on Drugs and Crime (2016). World Wildlife Crime Report: Trafficking in Protected Species (UNODC,Vienna).

Van der Merwe, N.J., Lee-Thorp, J.A., Thackeray, J.F., Hall-Martin, A., Kruger, F.J., Coetzee, H., Bell, R.H.V., and Lindeque, M. (1990). Source-area determination of elephant ivory by isotopic analysis. Nature *346*, 744.

Van Tassell, C.P., Smith, T.P., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., and Sonstegard, T.S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods *5*, 247–252.

Vande Weghe, J.P., Christy, P., Ducrocq, M., Vande Weghe, G.R., and Pauwels, O.S.G. (2016). Biodiversité des parcs nationaux et réserves du Gabon. 2. Espèces, écosystèmes et populations (Libreville, Gabon: Agence Nationale des Parcs Nationaux).

Verhegghen, A., Mayaux, P., De Wasseige, C., and Defourny, P. (2012). Mapping Congo Basin vegetation types from 300 m and 1 km multi-sensor time series for carbon stocks and forest areas estimation. Biogeosciences *9*, 5061.

Vidya, T.N.C., Kumar, V.R., Arivazhagan, C., and Sukumar, R. (2003). Application of molecular sexing to free-ranging Asian elephant (*Elephas maximus*) populations in southern India. Current Science 1074–1077.

Vigilant, L. (2002). Technical challenges in the microsatellite genotyping of a wild chimpanzee population using feces. Evolutionary Anthropology: Issues, News, and Reviews *11*, 162–165.

Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution *34*, 275–306.

Vigne, L., and Martin, E. (2017). Decline in the legal ivory trade in China in anticipation of a ban (Save the Elephants Nairobi, Kenya).

Waits, L.P., and Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. Journal of Wildlife Management *69*, 1419–1433.

Wasser, S.K., Shedlock, A.M., Comstock, K., Ostrander, E.A., Mutayoba, B., and Stephens, M. (2004). Assigning African elephant DNA to geographic region of origin: applications to the ivory trade. Proceedings of the National Academy of Sciences of the United States of America *101*, 14847–14852.

Wasser, S.K., Clark, W.J., Drori, O., Kisamo, E., Mailand, C., Mutayoba, B., and Stephens, M. (2008). Combating the illegal trade in African elephant ivory with DNA forensics. Conservation Biology *22*, 1065–1071.

Wasser, S.K., Brown, L., Mailand, C., Mondol, S., Clark, W., Laurie, C., and Weir, B.S. (2015). Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. Science *349*, 84–87.

Wilkinson, S., Wiener, P., Archibald, A.L., Law, A., Schnabel, R.D., McKay, S.D., Taylor, J.F., and Ogden, R. (2011). Evaluation of approaches for identifying population informative markers from high density SNP chips. BMC Genetics *12*, 45.

Winters, M., Torkelson, A., Booth, R., Mailand, C., Hoareau, Y., Tucker, S., and Wasser, S.K. (2018). Isolation of DNA from small amounts of elephant ivory: Sampling the cementum with total demineralization extraction. Forensic Science International *288*, 131–139.

Wittemyer, G., Northrup, J.M., Blanc, J., Douglas-Hamilton, I., Omondi, P., and Burnham, K.P. (2014). Illegal killing for ivory drives global decline in African elephants. Proceedings of the National Academy of Sciences *111*, 13117–13121.

Wozney, K.M., and Wilson, P.J. (2012). Real-time PCR detection and quantification of elephantid DNA: Species identification for highly processed samples associated with the ivory trade. Forensic Science International *219*, 106–112.

Wultsch, C., Waits, L.P., Hallerman, E.M., and Kelly, M.J. (2015). Optimizing collection methods for noninvasive genetic sampling of Neotropical felids. Wildlife Society Bulletin *39*, 403–412.

Yu, Y., Wetzler, A., Yang, X., Tang, R., and Zhang, L. (2017). Significant and timely ivory trade restrictions in both China and the United States are critical to save elephants. Conservation Letters *10*, 596–601.

Zhang, L., and Yin, F. (2014). Wildlife consumption and conservation awareness in China: a long way to go. Biodiversity and Conservation *23*, 2371–2381.

Ziegler, S., Merker, S., Streit, B., Boner, M., and Jacob, D.E. (2016). Towards understanding isotope variability in elephant ivory to establish isotopic profiling and source-area determination. Biological Conservation *197*, 154–163.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. (2009). Zero-truncated and zero-inflated models for count data. In Mixed Effects Models and Extensions in Ecology with R, (New York: Springer), pp. 261–293.

# Appendices

**Appendix 1** Description of the 107 validated SNP assays with sequence length and primer sequences. KASP is a genotyping technology based on competitive allele-specific PCR and the reaction mix contains two allele-specific forward primers and one unique reverse primer. Rescoring rates and proportion of missing genotype data per locus are given as indices of assay quality.

| Locus name | Targeted SNP | Strand used for assay design | Forward primer A (5'-3') | Forward primer B (5'-3') | Common Reverse primer (5'-3') | Sequence length (bp) | Rescoring rate (%) | Missing genotype data (%) |
|---|---|---|---|---|---|---|---|---|
| CL_1 | A/C | F | AAAGCGGGGACTAACTCCCAATTA | AGCGGGGACTAACTCCCAATTC | GTTCATCATTTAGGTGGGTGTTATGGAA | 50 | 0,0 | 11,8 |
| CL_340 | A/C | F | CCAGTTCTAGAAGTACCTGTAAGGTA | CAGTTCTAGAAGTACCTGTAAGGTC | CAAGGCAGGTAAGTAGCATAGGCTT | 57 | 14,0 | 5,4 |
| CL_367 | A/G | R | AAAGGACACCCAGCGACTCCT | AGGACACCCAGCGACTCCC | GTGTGAGGAAAGTCACCCCACTTAT | 56 | 0,0 | 14,0 |
| CL_370 | C/T | F | ATCTCCCGTGAGGAGGGGC | CATCTCCCGTGAGGAGGGGT | GTGAAGCCGTGGTAGAGGGTCTT | 44 | 2,2 | 8,6 |
| CL_383 | C/T | R | TGAGCCGCCACACGCCG | CTTGAGCCGCCACACGCCA | GCGCAACATGCAGCAGCCGTT | 61 | 4,3 | 11,8 |
| CL_406 | A/C | R | TCCACGTCTAGCTTAAATGTCCCT | CCACGTCTAGCTTAAATGTCCCG | AAAGATAGGAATGACCTGAGAGTGGATTA | 63 | 2,2 | 7,5 |
| CL_453 | C/T | R | GACAGCCACAGCCCCACG | CGACAGCCACAGCCCCACA | CACCCGTGGAGTCGTAGAGCTA | 41 | 1,1 | 3,2 |
| CL_456 | C/T | R | ATTTTTGAGCCCGGTTACATATCTATC | TTTGAGCCCGGTTACATATCTATT | GGCGCCAACCCCCCAAAAGTTT | 51 | 1,1 | 10,8 |
| CL_516 | C/T | F | AATATGCTCAGGAGAACCTGAAGG | CAATATGCTCAGGAGAACCTGAAGA | CTGGGCTCCCTGCCCCCAT | 45 | 1,1 | 3,2 |
| CL_570 | A/G | R | AGGTGACATACAGTTTGAAACGGTTAT | GGTGACATACAGTTTGAAACGGTTAC | CTTCACATCTCTCCACAGCCTCTTT | 58 | 6,5 | 11,8 |
| CL_685 | C/T | F | TCCTCCACATGGTAAAGCATGATC | CTTCCTCCACATGGTAAAGCATGATT | CTCTTAGCTAACATGATGTTTTTCCACTTT | 65 | 15,1 | 10,8 |
| CL_708 | C/T | R | TTTGGCAAGATACAACAGCCAGG | GTTTTTGGCAAGATACAACAGCCAGA | GGCATCCTGACCAATCTTTCCTAACTT | 53 | 1,1 | 4,3 |
| CL_827 | G/T | F | TGCTTTGGCTTAGAGCCTCTG | CTTGCTTTGGCTTAGAGCCTCTT | CACTTAGCATCATGCCTTCTGGGAA | 50 | 1,1 | 2,2 |
| CL_1024 | C/G | F | AACAGCCTGTGAGCTCACCTAC | AACAGCCTGTGAGCTCACCTAG | CTTACCAGGGCTCCAGGGGTT | 46 | 1,1 | 12,9 |
| CL_1100 | A/C | F | CAACAGCCAGGTGGTCCCAGA | ACAGCCAGGTGGTCCCAGC | AAGGTGGGAAGCCATGTTCCTGAAA | 51 | 1,1 | 4,3 |
| CL_1104 | C/T | F | GGCTGATAGGTGCCTGCACTC | GGCTGATAGGTGCCTGCACTT | GATGGGGTGGTACACTGTGCGAA | 51 | 1,1 | 4,3 |
| CL_1134 | A/G | F | GTAACCATTGGTTTATTTCTATTTGTGCAA | AACCATTGGTTTATTTCTATTTGTGCAG | GTGGTGGAGCAAAAGGGTAGACTAA | 59 | 1,1 | 2,2 |
| CL_1165 | C/T | F | ACGTACTCCCACGACAACATTATC | GACGTACTCCCACGACAACATTATT | TCTTGGAGAAGGAGACAGAGGTGAT | 56 | 1,1 | 3,2 |
| CL_1380 | A/G | R | ACCATCAAGTAGAGGTCAACACCT | CCATCAAGTAGAGGTCAACACCC | CAGAGTTCTCATTTTTACAACTGCCATTTA | 60 | 12,9 | 23,7 |
| CL_1415 | C/T | R | ACAGGGACAGGACTATGGCAG | GACAGGGACAGGACTATGGCAA | TTCCTGGGCCTGGCTCTGCTTT | 47 | 0,0 | 3,2 |
| CL_1574 | C/T | R | ACATCGAGCTAACATTCAACAGCG | AATACATCGAGCTAACATTCAACAGCA | AGTCCAACCTCCACGACTCTAACAA | 55 | 1,1 | 4,3 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CL_1606 | A/G | R | GAATATAGAATTAAGCCTGAAACCCTCAT | ATATAGAATTAAGCCTGAAACCCTCAC | GGGAGCCTCCTTCCCTGAATGAT | 54 | 2,2 | 5,4 |
| CL_1636 | C/T | F | CAGCCCCTGCGGCCCC | CCAGCCCCTGCGGCCCT | CAAACAGCTGTTTCCAGAGGCTGAA | 57 | 1,1 | 4,3 |
| CL_1679 | A/G | R | CTGCTTGTCTTCCCAGCCTGT | CTGCTTGTCTTCCCAGCCTGC | CCCAGAGGGCCCCCGAGTT | 41 | 2,2 | 6,5 |
| CL_1743 | A/G | F | GAAGTAAAATGCTGATTAGGTCAAATGCA | AGTAAAATGCTGATTAGGTCAAATGCG | AAGGCAGGAGCTTCCCCCTGAA | 54 | 0,0 | 5,4 |
| CL_2023 | C/T | R | AGACATATGCATGCAAGTCACGTG | ATAGACATATGCATGCAAGTCACGTA | TCATTCGTGCCCCTCAGCCCTA | 72 | 17,2 | 24,7 |
| CL_2046 | C/T | R | AATGTATGTGAAAATGAAATGCAAATCTGG | CAATGTATGTGAAAATGAAATGCAAATCTGA | CAATGGGAACTTATTGAAGGGCTTCAAA | 74 | 1,1 | 12,9 |
| CL_2068 | A/G | F | GAGGAACATGTGAATGGGCGGA | AGGAACATGTGAATGGGCGGG | GCCCACAGAAGGGCACAGATGTT | 44 | 1,1 | 5,4 |
| CL_2127 | C/G | F | GAAAACAAGCTTTTCCCTCACCC | GAAAACAAGCTTTTCCCTCACCG | GCTGGAGACATTCCTTCTTCAGCAA | 48 | 0,0 | 6,5 |
| CL_2184 | A/G | R | CATTAGAAAGATTTTCATGTTATAGAGTTCTT | CATTAGAAAGATTTTCATGTTATAGAGTTCTC | CAACGGCATTATTTTAATTTTCTGGGCTTA | 71 | 0,0 | 4,3 |
| CL_2193 | A/G | R | CACCCAGCAGCAGCTGCCAT | ACCCAGCAGCAGCTGCCAC | TGCAGGACGCTGTCGTGGTCAT | 79 | 5,4 | 22,6 |
| CL_2249 | C/T | R | TTGATCATTACAACAACTGTATGAGTTAG | CTTTTGATCATTACAACAACTGTATGAGTTAA | CCCTGAGCATCAATTTCTTTGTCTGTTAA | 81 | 1,1 | 12,9 |
| CL_2425 | C/T | F | GAGGAATATTGACTAGTTACTGAC | CCTGAGGAATATTGACTAGTTACTGAT | GGACCTTTTAGAAAGGGCCACGATA | 54 | 1,1 | 11,8 |
| CL_2561 | C/T | R | ACAAGTGGTACTCACGCCACAG | GACAAGTGGTACTCACGCCACAA | CCCAGGATTTGACAGCCTGGTAAAT | 52 | 1,1 | 11,8 |
| CL_2600 | C/T | F | GCAAGGCCATTCTCAAGTTCCC | TGCAAGGCCATTCTCAAGTTCCT | CAGACCATGTGTTCTTCTATTCCATCTTT | 71 | 0,0 | 5,4 |
| CL_2831 | C/T | F | AGTCCCGATCACGTCACTCTC | CAGTCCCGATCACGTCACTCTT | AGTGGGGAGCCATGGAGGGATT | 52 | 4,3 | 7,5 |
| CL_2938 | A/T | R | GCCTGGAAGCTGAACATCCT | CTGCCTGGAAGCTGAACATCCA | CCCCTCCCCCTGGCAGTGTT | 47 | 2,2 | 2,2 |
| CL_2968 | C/T | R | ATTTGTTTTGTAATCTTTGGAGTCCCG | TATTTGTTTTGTAATCTTTGGAGTCCCA | GATTCGGAACAGTCAACAATACAGTGTTT | 74 | 1,1 | 8,6 |
| CL_3094 | A/G | F | AACGGTGCCAGCGTATCCTA | AACGGTGCCAGCGTATCCTG | GAGGTGTCAGGGCTGAAGGTGAT | 44 | 3,2 | 9,7 |
| CL_3102 | C/G | R | CCCACACTGGCCTCTGTTAAG | CCCACACTGGCCTCTGTTAAC | GGTCCGTGCGGGTCCCTGT | 41 | 0,0 | 7,5 |
| CL_3125 | A/G | F | AGATTCATGAGAATCTTAGAAAAGTTCCA | CATGAGAATCTTAGAAAAGTTCCG | CATAAGGTACAGAGCTCTGGGCAAA | 51 | 4,3 | 6,5 |
| CL_3334 | C/T | F | AGACATGGAAGTATGTTGGCAACAC | GAGACATGGAAGTATGTTGGCAACAT | CATGCCTACATATGTCCACTACAAGTTT | 104 | 9,7 | 28,0 |
| CL_3377 | A/G | F | CAGGGCTCCATAAAATGTTCAGGTA | AGGGCTCCATAAAATGTTCAGGTG | GACCTTTTCCAAGGGATTGTTAAGGTAAA | 57 | 1,1 | 4,3 |
| CL_3594 | A/G | R | CCCTGCGTCCAGGCAGTCAT | CCTGCGTCCAGGCAGTCAC | GAGTATGTGTATGTATGTTGCCTGAGTAT | 69 | 0,0 | 6,5 |
| CL_3673 | A/G | F | CATCCTCTGTAGCATCTCTGTTTTAA | CATCCTCTGTAGCATCTCTGTTTTAG | TGTTACCACTGCTCTTCTCCCCAT | 60 | 2,2 | 3,2 |
| CL_3697 | C/G | R | GTCCTTCAGTCCCCTCCG | CTGTCCTTCAGTCCCCTCCC | TCCCTGTGAAGCAAGGCCGGAT | 43 | 1,1 | 3,2 |
| CL_3702 | A/G | F | GTTAGTAACTGTGGTACACCTGCTA | AGTAACTGTGGTACACCTGCTG | TCCTATATGATGATAAGTGGGGTGGTAA | 50 | 3,2 | 4,3 |
| CL_3708 | C/T | F | GTTTCCACGTCTTTTGCTGTACC | GTGTTTCCACGTCTTTTGCTGTACT | AGAGGCCACAGATTTACTATGTCTGAAA | 54 | 0,0 | 5,4 |
| CL_3824 | A/C | R | AATAAGCACATTCCTGGGAAGAGGT | AAGCACATTCCTGGGAAGAGGG | AGCATGAGGGCAGGGCTGAGAA | 49 | 2,2 | 11,8 |
| CL_3834 | G/T | F | CACCAGGATTTCTTCCTCCTCAG | CACCAGGATTTCTTCCTCCTCAT | GCATCTTGTACGTGTATCCAGGGAT | 50 | 8,6 | 17,2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CL_4032 | A/G | F | CAAGAGATTGTGACTGTGAGGTGA | AAGAGATTGTGACTGTGAGGTGG | ACAAACAGCTGGTGGCCTAGCATA | 48 | 4,3 | 5,4 |
| CL_4133 | A/G | F | AAATTATAGCATTGAACTCCTGCCCTA | ATAGCATTGAACTCCTGCCCTG | AAGGGAGGGGACAGGCTGCTT | 47 | 0,0 | 10,8 |
| CL_4194 | A/T | R | CAGACTGTGGATTTAGTCCCTACTT | CAGACTGTGGATTTAGTCCCTACTA | TGCAAGAAAAACCTTTTCTTGAGACTTGAT | 59 | 0,0 | 14,0 |
| CL_4366 | C/T | R | GGGAGGGCATTCTCATTATCACG | GGGAGGGCATTCTCATTATCACA | TCAGAGTGTTAGGTGGGGTGGTAT | 56 | 1,1 | 14,0 |
| CL_4367 | C/G | F | AGTGGCGCCACCACAGCC | AGTGGCGCCACCACAGCG | CTTCCAACAGTGATGCTTCCCTCAA | 52 | 0,0 | 12,9 |
| CL_4629 | C/T | R | GGATTCACCAGTTGGGATTGGTG | CACCAGTTGGGATTGGTA | AGCTGCCTCTGCTGTGCCTGAA | 68 | 17,2 | 24,7 |
| CL_4719 | C/T | R | ATGTTACTTTCAGAAAGATGCTTTAAGAG | AAATATGTTACTTTCAGAAAGATGCTTTAAGAA | TCTTAGAAGCATGGAGAGTCATGTACTT | 62 | 2,2 | 22,6 |
| CL_4878 | A/G | R | AATAGCAAGACTGGAAATAACCTAAATGTT | AGCAAGACTGGAAATAACCTAAATGTC | GCTGCATAGTATTCCAAGTTTCATAACGAT | 68 | 2,2 | 15,1 |
| CL_5010 | A/G | R | GAAATTTCTAACCCCCTTGGCCT | GAAATTTCTAACCCCCTTGGCCC | AGGAGCTTCCAGAGGCCCAGAA | 48 | 0,0 | 10,8 |
| CL_5055 | A/G | F | ACCTCCACCCCAACCATCTCTA | CCTCCACCCCAACCATCTCTG | GTGTGGGCCTGCTGGGGGTT | 43 | 0,0 | 14,0 |
| CL_5374 | A/G | R | GAACTGGAGAACACCCCAT | CTGAACTGGAGAACACCCCAC | GGACTTGGGGCACAGGCTCAT | 58 | 3,2 | 15,1 |
| CL_5423 | A/C | R | TAGCCAAATGGCTGAGACTGGT | AGCCAAATGGCTGAGACTGGG | TAAGCTAGTTGCCACACACGCTCAA | 57 | 1,1 | 11,8 |
| CL_5488 | A/G | R | GACCCGCCCTTACCCAGGT | ACCCGCCCTTACCCAGGC | TCCCAGATCCTCACCCCTGCAA | 59 | 2,2 | 17,2 |
| CL_5547 | C/T | R | AAACACCCTAGCCCCGAACG | CTAAACACCCTAGCCCCGAACA | AGGCTGCACTGTTATTCTCTGCCAT | 58 | 1,1 | 12,9 |
| CL_5690 | C/T | R | GAAATGCCCTTCTATTCCTACG | CCTGAAATGCCCTTCTATTCCTACA | ATGGACAAACAACGAACAGCTGATTGAA | 53 | 1,1 | 15,1 |
| CL_5786 | A/G | F | ACACTACTCCTTGGCAACCACTA | CACTACTCCTTGGCAACCACTG | GACCTGAGCAGTATTAACCTGGCTT | 49 | 1,1 | 10,8 |
| CL_5878 | A/G | F | GTTACCACCACAGGCTACGATA | GTTACCACCACAGGCTACGATG | CTTGAGTACTCGCTTGAAACGCCAA | 60 | 0,0 | 11,8 |
| CL_6355 | G/T | F | CCATCTCCTGGAGACCAGGG | CCATCTCCTGGAGACCAGGT | CAGAGGCCCTGGTGGAGGAAA | 41 | 0,0 | 12,9 |
| CL_6363 | A/G | F | AACTACAAACTGTCTAGCATACACACA | CTACAAACTGTCTAGCATACACACG | TGTGCCATGTGACTATGTACACAGTATTT | 73 | 5,4 | 12,9 |
| CL_6583 | A/G | R | GTGGCCAGGGCCACCACT | GTGGCCAGGGCCACCACC | CTGCCCACAGCCGCGTCCAT | 50 | 8,6 | 16,1 |
| CL_6769 | C/T | R | CCAGTGTTAACAGTGACATCCAG | GCCAGTGTTAACAGTGACATCCAA | GGTCAGGGAGTACCTGTTACTGATA | 49 | 2,2 | 11,8 |
| CL_6883 | A/C | F | CATTTTAGCAGCTCTCTGTTCCTCAA | TTAGCAGCTCTCTGTTCCTCAC | ACATTCAAGGGGAGCTTTTCCTAGAATT | 50 | 2,2 | 10,8 |
| CL_6951 | A/G | R | GGAGAAATGCAAAAGCCGGAT | CTGGAGAAATGCAAAAGCCGGAC | GTCCATTCGTTAGTGAGCGTGTGTA | 53 | 0,0 | 10,8 |
| CL_7275 | C/T | R | AGGGAAAGCAGGCCAGGCG | AATAGGGAAAGCAGGCCAGGCA | ACCAAGGCGCCACCTCCTCAT | 48 | 7,5 | 44,1 |
| CL_7365 | C/G | R | GTAGTTCCCACCACCCCCG | GTAGTTCCCACCACCCCCC | CCCAGGTAACAAACCTGCCCCAA | 61 | 9,7 | 10,8 |
| CL_7618 | C/T | R | GGATTGTGAGTCTAAGCCTGTCG | GTGAGTCTAAGCCTGTCA | CCTCAGGCCTCACAGAGCATCTA | 61 | 1,1 | 4,3 |
| CL_7666 | C/T | F | AGGCTTCTATTCTGAATGAGCTGC | GAGGCTTCTATTCTGAATGAGCTGT | GCTCCGAAGTCTCCAGCAGCTT | 48 | 15,1 | 8,6 |
| CL_7790 | C/T | F | GGGTGTCACTGCCTGCCC | CTGGGTGTCACTGCCTGCCT | CTAGAAATTTCATAAGTCCCTCTCCCATT | 63 | 3,2 | 4,3 |
| CL_7859 | C/T | R | GGAGTAGCGGAGGGAACCG | CAGGAGTAGCGGAGGGAACCA | ACGGCCGGCAGCACAAAACGAA | 46 | 0,0 | 10,8 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CL_8200 | A/T | R | ATTATCACAAGACCCTGTCCCATATTT | ATCACAAGACCCTGTCCCATATTA | GGTAGGAAGGGAGAACCTTTAACCTA | 59 | 0,0 | 19,4 |
| CL_8219 | A/G | R | GGCCACCGTGCTGCCGT | GGCCACCGTGCTGCCGC | ACCGCCGCCCCAGCCACTA | 46 | 0,0 | 9,7 |
| CL_8307 | C/T | R | GAATGACCTCAATCACCTGCCG | GGAATGACCTCAATCACCTGCCA | ACTGGTATTCAGCCTTGCAGGCTA | 51 | 2,2 | 5,4 |
| CL_8379 | C/T | F | CAGGAATCTGAACTGAGACCAAC | GCAGGAATCTGAACTGAGACCAAT | GAGTGGTGTTGCCATCGAGATGAAA | 59 | 2,2 | 4,3 |
| CL_8615 | A/G | R | CGTGGTCAAGAGAGATGATGACAT | GTGGTCAAGAGAGATGATGACAC | GTGTGCCCCAGCATTGCAAAGATA | 64 | 1,1 | 4,3 |
| CL_8653 | C/T | F | GGGATGGGCAGGACAGAACC | GGGATGGGCAGGACAGAACT | TTGCTTTCCTTCCTTTCGGGGGTTT | 47 | 1,1 | 5,4 |
| CL_8664 | C/T | F | CAGGGTTTTGGGTGAACCC | CCTCAGGGTTTTGGGTGAACCT | TTGGGGTGCCGTGGATGGCATT | 45 | 2,2 | 4,3 |
| CL_8743 | C/T | R | TGTAAATTATGAACTTTCTTGTTGCTTG | AACTTGTAAATTATGAACTTTCTTGTTGCTTA | TAGCTAAAGAAATGTCACCTAGGGCATTT | 62 | 1,1 | 4,3 |
| CL_8818 | A/G | F | CAGAGCCAAAAGAAACAATGTCACTA | CAGAGCCAAAAGAAACAATGTCACTG | CATTTTCAAATCCCCCCTGAGTTTTCTTT | 78 | 4,3 | 17,2 |
| CL_8848 | A/G | R | GTGTGATACAACACAATGATCATGTCAT | GTGATACAACACAATGATCATGTCAC | GGGCTGTTTAACTCTTGGAACATGATATT | 60 | 1,1 | 10,8 |
| CL_8849 | A/G | F | GGGCCTGCTCCCTTTCCCA | GGCCTGCTCCCTTTCCCG | AGATCTGTCCAAACAGAGAAAATGCCTT | 50 | 1,1 | 4,3 |
| CL_8854 | C/T | R | CCTCCATTTTGCAGAACAGCATG | ACCCTCCATTTTGCAGAACAGCATA | CCATGTTGCCAATTAACAGACATTAGGAT | 54 | 0,0 | 8,6 |
| CL_8896 | A/G | F | TGGGGATTAGTAGACCAGAAGTGA | GGGGATTAGTAGACCAGAAGTGG | ATCAGCTTGTTCCCCATCACCGTTA | 51 | 0,0 | 4,3 |
| CL_9037 | C/T | F | GAATATCAACTATCCCCATGAAGGC | ATGAATATCAACTATCCCCATGAAGGT | CCTGTTTAGGGCTTTTCAGAGAGGAA | 57 | 0,0 | 7,5 |
| CL_9115 | A/C | F | AAAGCTCAGTGGCTCAGATTAATAGA | AAGCTCAGTGGCTCAGATTAATAGC | CGTACCCTTTTTGTTTTGTAGAGTATGATT | 59 | 1,1 | 5,4 |
| CL_9162 | C/T | R | TGAGAGCTGAAGAAGCTGGCG | CTTGAGAGCTGAAGAAGCTGGCA | GGAGGCAGAGGCAGAGAAGCAA | 52 | 0,0 | 26,9 |
| CL_9201 | C/T | F | CCAAGGCCCGGCTGCTC | CTCCAAGGCCCGGCTGCTT | CTACACAATGAGGGAGTGTGTGCTT | 46 | 0,0 | 14,0 |
| CL_9263 | A/G | R | GTGCAGCTGGGGCTGGTAGT | GCAGCTGGGGCTGGTAGC | TACGTTGGCATGTTCTGCCTTGGAA | 58 | 5,4 | 12,9 |
| CL_9467 | A/C | F | GAGGCTCACTCTGCTCCAATGA | AGGCTCACTCTGCTCCAATGC | CCTGAACCCAGCCACAGGAGTT | 51 | 2,2 | 3,2 |
| CL_9552 | C/T | F | CCAGACCAAGAAGAAGTACTTTGAC | CCAGACCAAGAAGAAGTACTTTGAT | GCTTCTGCTGACGCTCCAGGTT | 59 | 3,2 | 3,2 |
| CL_9729 | A/G | R | CCCTCTTGTCTAACTACCATACT | CCCTCTTGTCTAACTACCATACC | GCTTTGAAGTACGAATGAAGTACTGACAA | 57 | 0,0 | 5,4 |
| CL_9831 | A/G | R | CAGCTAGAAGCCTAGAGCTGCT | AGCTAGAAGCCTAGAGCTGCC | AGCAGTGCACCGCTCAGAGACA | 54 | 2,2 | 3,2 |
| CL_9867 | A/G | R | CGGCATACCCGTAGTGGAAAGT | GGCATACCCGTAGTGGAAAGC | GCCTGTGGTTTGACCCCATGGTA | 48 | 2,2 | 5,4 |
| CL_9914 | A/G | F | CAGCTGAGTGACCTGCCCAGA | GCTGAGTGACCTGCCCAGG | ATCCTGGTTCTGCTCTTGGCACTAT | 49 | 2,2 | 6,5 |
| CL_10073 | C/T | R | CGGTGCATCAGCAAAGCG | AGCTCGGTGCATCAGCAAAGCA | GTTATCTTCACAATCGCTGTCGCCAT | 58 | 5,4 | 14,0 |
| CL_10251 | C/T | R | GTTTATTCTGTGTTTCCTATTTCACG | GCTGTTTATTCTGTGTTTCCTATTTCACA | CCAAACCCAAAGTATGGTCTTCAATACAA | 60 | 0,0 | 5,4 |
| CL_10364 | C/T | F | GAACTCTAGACTTAGTGGCAAGC | CTGAACTCTAGACTTAGTGGCAAGT | CCTTGAGCAGGGTTGGACATTTCTA | 52 | 0,0 | 2,2 |
| CL_11132 | C/T | F | ACTGCAAGAGTATCTGAGTGTGC | CACTGCAAGAGTATCTGAGTGTGT | GCACGCTGTCACACAGCCCATA | 47 | 1,1 | 5,4 |

**Appendix 2** Preliminary measures of polymorphism (*MAF* Minor Allele Frequency, $H_O$ observed heterozygosity, $H_E$ expected heterozygosity) and population differentiation ($F_{ST}$) within the dataset of 57 individuals split into three populations (North-East, Central and Coastal).

| Locus | Overall | | | | Populations | | |
|---|---|---|---|---|---|---|---|
| | HO | HE | Fst | MAF | MAF North-East | MAF Central | MAF Coastal |
| CL_1 | 0,180 | 0,160 | 0,029 | 0,090 | 0,075 | 0,143 | 0,063 |
| CL_340 | 0,180 | 0,190 | 0,000 | 0,109 | 0,045 | 0,133 | 0,176 |
| CL_367 | 0,190 | 0,230 | -0,037 | 0,130 | 0,143 | 0,167 | 0,088 |
| CL_370 | 0,520 | 0,470 | 0,010 | 0,370 | 0,364 | 0,367 | 0,344 |
| CL_383 | 0,060 | 0,060 | 0,024 | 0,029 | 0,000 | 0,000 | 0,094 |
| CL_406 | 0,490 | 0,480 | 0,041 | 0,404 | 0,348 | 0,333 | 0,444 |
| CL_453 | 0,370 | 0,320 | 0,017 | 0,202 | 0,152 | 0,300 | 0,167 |
| CL_456 | 0,530 | 0,480 | 0,022 | 0,396 | 0,375 | 0,467 | 0,324 |
| CL_516 | 0,200 | 0,210 | 0,056 | 0,116 | 0,130 | 0,200 | 0,000 |
| CL_570 | 0,390 | 0,430 | 0,063 | 0,314 | 0,452 | 0,333 | 0,107 |
| CL_685 | 0,080 | 0,070 | -0,025 | 0,038 | 0,024 | 0,033 | 0,063 |
| CL_708 | 0,280 | 0,470 | 0,023 | 0,386 | 0,457 | 0,200 | 0,444 |
| CL_827 | 0,140 | 0,130 | 0,030 | 0,070 | 0,065 | 0,000 | 0,139 |
| CL_1024 | 0,060 | 0,050 | -0,040 | 0,028 | 0,024 | 0,067 | 0,000 |
| CL_1100 | 0,140 | 0,130 | 0,090 | 0,071 | 0,159 | 0,000 | 0,028 |
| CL_1104 | 0,360 | 0,320 | 0,016 | 0,196 | 0,130 | 0,200 | 0,265 |
| CL_1134 | 0,320 | 0,310 | 0,101 | 0,193 | 0,152 | 0,267 | 0,139 |
| CL_1165 | 0,070 | 0,070 | 0,002 | 0,035 | 0,000 | 0,067 | 0,056 |
| CL_1380 | 0,340 | 0,390 | 0,004 | 0,264 | 0,205 | 0,250 | 0,375 |
| CL_1415 | 0,200 | 0,180 | 0,024 | 0,098 | 0,065 | 0,067 | 0,176 |
| CL_1574 | 0,110 | 0,100 | -0,009 | 0,054 | 0,043 | 0,033 | 0,088 |
| CL_1606 | 0,400 | 0,500 | 0,161 | 0,500 | 0,304 | 0,500 | 0,250 |
| CL_1636 | 0,190 | 0,170 | 0,010 | 0,096 | 0,043 | 0,167 | 0,111 |
| CL_1679 | 0,400 | 0,440 | 0,016 | 0,327 | 0,239 | 0,464 | 0,324 |
| CL_1743 | 0,050 | 0,050 | -0,041 | 0,027 | 0,045 | 0,033 | 0,000 |
| CL_2023 | 0,150 | 0,210 | 0,043 | 0,117 | 0,132 | 0,100 | 0,115 |
| CL_2046 | 0,360 | 0,340 | 0,026 | 0,218 | 0,159 | 0,367 | 0,176 |
| CL_2068 | 0,360 | 0,450 | -0,030 | 0,339 | 0,326 | 0,333 | 0,382 |
| CL_2127 | 0,090 | 0,120 | 0,009 | 0,064 | 0,109 | 0,000 | 0,063 |
| CL_2184 | 0,040 | 0,100 | 0,058 | 0,053 | 0,130 | 0,000 | 0,000 |
| CL_2193 | 0,310 | 0,290 | 0,019 | 0,177 | 0,184 | 0,167 | 0,179 |
| CL_2249 | 0,130 | 0,160 | 0,016 | 0,085 | 0,053 | 0,033 | 0,167 |
| CL_2425 | 0,160 | 0,180 | -0,005 | 0,100 | 0,159 | 0,100 | 0,029 |
| CL_2561 | 0,110 | 0,170 | -0,022 | 0,091 | 0,114 | 0,033 | 0,118 |
| CL_2600 | 0,350 | 0,390 | 0,041 | 0,264 | 0,167 | 0,433 | 0,222 |
| CL_2831 | 0,400 | 0,490 | 0,022 | 0,418 | 0,455 | 0,464 | 0,306 |
| CL_2938 | 0,450 | 0,380 | -0,005 | 0,259 | 0,261 | 0,333 | 0,176 |
| CL_2968 | 0,350 | 0,370 | -0,029 | 0,245 | 0,261 | 0,286 | 0,206 |
| CL_3094 | 0,420 | 0,460 | 0,004 | 0,365 | 0,286 | 0,433 | 0,406 |
| CL_3102 | 0,070 | 0,070 | -0,014 | 0,036 | 0,022 | 0,033 | 0,059 |
| CL_3125 | 0,410 | 0,380 | 0,055 | 0,259 | 0,152 | 0,400 | 0,294 |
| CL_3334 | 0,250 | 0,310 | 0,039 | 0,193 | 0,167 | 0,167 | 0,250 |
| CL_3377 | 0,070 | 0,070 | -0,030 | 0,035 | 0,022 | 0,033 | 0,056 |
| CL_3594 | 0,180 | 0,170 | 0,031 | 0,091 | 0,091 | 0,033 | 0,118 |
| CL_3673 | 0,530 | 0,490 | -0,034 | 0,439 | 0,435 | 0,400 | 0,472 |
| CL_3697 | 0,330 | 0,360 | 0,046 | 0,237 | 0,261 | 0,367 | 0,111 |
| CL_3702 | 0,430 | 0,460 | 0,005 | 0,357 | 0,457 | 0,267 | 0,294 |
| CL_3708 | 0,160 | 0,150 | 0,039 | 0,079 | 0,043 | 0,033 | 0,167 |
| CL_3824 | 0,440 | 0,500 | 0,008 | 0,473 | 0,364 | 0,400 | 0,500 |
| CL_3834 | 0,440 | 0,500 | -0,023 | 0,471 | 0,476 | 0,467 | 0,469 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CL_4032 | 0,090 | 0,120 | 0,009 | 0,063 | 0,022 | 0,067 | 0,118 |
| CL_4133 | 0,350 | 0,400 | -0,007 | 0,282 | 0,227 | 0,400 | 0,235 |
| CL_4194 | 0,330 | 0,400 | -0,014 | 0,278 | 0,286 | 0,367 | 0,206 |
| CL_4366 | 0,110 | 0,100 | -0,028 | 0,056 | 0,045 | 0,107 | 0,029 |
| CL_4367 | 0,460 | 0,440 | -0,006 | 0,324 | 0,286 | 0,300 | 0,412 |
| CL_4629 | 0,330 | 0,400 | 0,009 | 0,272 | 0,294 | 0,233 | 0,286 |
| CL_4719 | 0,290 | 0,250 | 0,114 | 0,146 | 0,029 | 0,100 | 0,313 |
| CL_4878 | 0,280 | 0,340 | 0,033 | 0,213 | 0,167 | 0,133 | 0,353 |
| CL_5010 | 0,220 | 0,270 | -0,045 | 0,164 | 0,159 | 0,200 | 0,147 |
| CL_5055 | 0,280 | 0,270 | 0,104 | 0,157 | 0,048 | 0,367 | 0,118 |
| CL_5374 | 0,390 | 0,480 | 0,024 | 0,398 | 0,476 | 0,333 | 0,324 |
| CL_5423 | 0,440 | 0,410 | 0,000 | 0,291 | 0,341 | 0,200 | 0,324 |
| CL_5488 | 0,290 | 0,480 | 0,055 | 0,402 | 0,250 | 0,433 | 0,438 |
| CL_5547 | 0,470 | 0,480 | 0,006 | 0,400 | 0,432 | 0,367 | 0,353 |
| CL_5690 | 0,130 | 0,320 | 0,080 | 0,198 | 0,095 | 0,179 | 0,294 |
| CL_5786 | 0,240 | 0,330 | -0,034 | 0,209 | 0,250 | 0,167 | 0,206 |
| CL_5878 | 0,090 | 0,120 | -0,012 | 0,064 | 0,114 | 0,067 | 0,000 |
| CL_6355 | 0,040 | 0,070 | -0,005 | 0,036 | 0,023 | 0,000 | 0,088 |
| CL_6363 | 0,160 | 0,150 | -0,007 | 0,082 | 0,068 | 0,167 | 0,029 |
| CL_6583 | 0,500 | 0,490 | 0,050 | 0,423 | 0,500 | 0,250 | 0,438 |
| CL_6769 | 0,250 | 0,250 | -0,023 | 0,145 | 0,182 | 0,167 | 0,088 |
| CL_6883 | 0,070 | 0,100 | -0,022 | 0,055 | 0,023 | 0,133 | 0,029 |
| CL_6951 | 0,420 | 0,450 | -0,008 | 0,336 | 0,250 | 0,433 | 0,353 |
| CL_7275 | 0,210 | 0,500 | 0,011 | 0,485 | 0,429 | 0,250 | 0,400 |
| CL_7365 | 0,250 | 0,460 | 0,025 | 0,356 | 0,474 | 0,367 | 0,235 |
| CL_7618 | 0,400 | 0,440 | 0,021 | 0,325 | 0,283 | 0,300 | 0,361 |
| CL_7666 | 0,310 | 0,490 | -0,026 | 0,445 | 0,500 | 0,429 | 0,417 |
| CL_7790 | 0,490 | 0,490 | -0,040 | 0,439 | 0,435 | 0,467 | 0,417 |
| CL_7859 | 0,530 | 0,490 | 0,036 | 0,445 | 0,432 | 0,433 | 0,294 |
| CL_8200 | 0,040 | 0,080 | 0,013 | 0,039 | 0,050 | 0,000 | 0,067 |
| CL_8219 | 0,150 | 0,170 | 0,043 | 0,091 | 0,091 | 0,167 | 0,000 |
| CL_8307 | 0,250 | 0,240 | 0,023 | 0,140 | 0,065 | 0,167 | 0,222 |
| CL_8379 | 0,280 | 0,270 | 0,044 | 0,158 | 0,087 | 0,133 | 0,278 |
| CL_8615 | 0,090 | 0,080 | 0,070 | 0,044 | 0,000 | 0,133 | 0,028 |
| CL_8653 | 0,230 | 0,280 | -0,032 | 0,170 | 0,205 | 0,167 | 0,139 |
| CL_8664 | 0,320 | 0,290 | -0,020 | 0,175 | 0,152 | 0,233 | 0,167 |
| CL_8743 | 0,330 | 0,450 | 0,023 | 0,342 | 0,370 | 0,200 | 0,444 |
| CL_8818 | 0,330 | 0,320 | 0,134 | 0,202 | 0,125 | 0,036 | 0,412 |
| CL_8848 | 0,250 | 0,300 | -0,023 | 0,182 | 0,182 | 0,133 | 0,235 |
| CL_8849 | 0,400 | 0,500 | -0,016 | 0,465 | 0,435 | 0,467 | 0,472 |
| CL_8854 | 0,140 | 0,130 | -0,036 | 0,071 | 0,087 | 0,067 | 0,059 |
| CL_8896 | 0,230 | 0,210 | -0,031 | 0,116 | 0,109 | 0,133 | 0,118 |
| CL_9037 | 0,270 | 0,240 | 0,007 | 0,136 | 0,205 | 0,133 | 0,059 |
| CL_9115 | 0,410 | 0,400 | 0,066 | 0,277 | 0,239 | 0,167 | 0,441 |
| CL_9162 | 0,040 | 0,110 | -0,016 | 0,061 | 0,029 | 0,036 | 0,118 |
| CL_9201 | 0,310 | 0,310 | -0,018 | 0,191 | 0,227 | 0,133 | 0,206 |
| CL_9263 | 0,460 | 0,460 | -0,036 | 0,361 | 0,357 | 0,400 | 0,324 |
| CL_9467 | 0,210 | 0,430 | -0,032 | 0,316 | 0,370 | 0,333 | 0,250 |
| CL_9552 | 0,180 | 0,160 | -0,013 | 0,088 | 0,065 | 0,067 | 0,139 |
| CL_9729 | 0,260 | 0,230 | 0,055 | 0,132 | 0,087 | 0,267 | 0,083 |
| CL_9831 | 0,430 | 0,460 | 0,067 | 0,357 | 0,435 | 0,167 | 0,441 |
| CL_9867 | 0,390 | 0,470 | -0,023 | 0,375 | 0,413 | 0,393 | 0,333 |
| CL_9914 | 0,200 | 0,210 | 0,080 | 0,118 | 0,190 | 0,167 | 0,000 |
| CL_10073 | 0,220 | 0,250 | -0,037 | 0,148 | 0,143 | 0,200 | 0,118 |
| CL_10251 | 0,290 | 0,450 | -0,041 | 0,339 | 0,386 | 0,300 | 0,306 |
| CL_10364 | 0,160 | 0,150 | 0,020 | 0,080 | 0,043 | 0,067 | 0,147 |
| CL_11132 | 0,240 | 0,240 | 0,153 | 0,136 | 0,159 | 0,033 | 0,147 |

**Appendix 3** Location of the 107 validated SNPs within *Loxodonta africana* assembly (LoxAfr 3.0, June 2007). The position within the scaffolds, and when applicable, gene name and ID number and corresponding product were obtained using NCBI BLAST. Seven SNPs were located within the coding DNA sequence (CDS) of the gene.

| Locus name | Loxafr 3.0 Scaffold | Start location | End location | Gene name | NCBI Gene ID | Product | SNP in CDS |
|---|---|---|---|---|---|---|---|
| CL_1 | 4 | 76852835 | 76852982 | | | | |
| CL_340 | 12 | 18371452 | 18371599 | ARIH2 | 100657240 | 266 bp E3 ubiquitin-protein ligase ARIH2 isoform X1/X2 | No |
| CL_367 | 45 | 4170197 | 4170344 | SNX8 | 100655101 | sorting nexin 8 | No |
| CL_370 | 26 | 23043829 | 23043976 | SLC25A42 | 100665615 | mitochondrial coenzyme A transporter SLC25A42 | Yes |
| CL_383 | 163 | 678604 | 678751 | | | | |
| CL_406 | 54 | 3977959 | 3978404 | | | | |
| CL_453 | 10 | 60121443 | 60121590 | | | | |
| CL_456 | 118 | 1049195 | 1049342 | | | | |
| CL_516 | 69 | 11227713 | 11227860 | FRMD1 | 100669987 | FERM domain-containing protein 1 | No |
| CL_570 | 11 | 6881254 | 6881401 | FARP1 | 100671176 | FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived) | No |
| CL_685 | 11 | 50373352 | 50373499 | | | | |
| CL_708 | 11 | 50692979 | 50693126 | | | | |
| CL_827 | 63 | 7679420 | 7679566 | | | | |
| CL_1024 | 105 | 513838 | 513985 | | | | |
| CL_1100 | 7 | 74872253 | 74872400 | | | | |
| CL_1104 | 7 | 74005149 | 74005296 | | | | |
| CL_1134 | 33 | 17823777 | 17823924 | | | | |
| CL_1165 | 26 | 21731614 | 21731761 | GNB1 | 100673391 | guanine nucleotide binding protein (G protein), beta polypeptide 1 | Yes |
| CL_1380 | 54 | 11896241 | 11896388 | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CL_1415 | 72 | 7289573 | 7289720 | EBF3 | 100665830 | transcription factor COE3 | No |
| CL_1574 | 26 | 52294774 | 52294921 | PPM1H | 100664904 | protein phosphatase 1H | No |
| CL_1606 | 10 | 68562491 | 68562638 | | | | |
| CL_1636 | 0 | 125838096 | 125838243 | KCNK5 | 100669713 | potassium channel subfamily K member 5 | No |
| CL_1679 | 76 | 7624208 | 7624355 | AMOTL2 | 100670180 | angiomotin-like protein 2 | No |
| CL_1743 | 69 | 6879230 | 6879377 | | | | |
| CL_2023 | 58 | 4929965 | 4930112 | DSCAML1 | 100676259 | Down syndrome cell adhesion molecule like 1 | No |
| CL_2046 | 1 | 64345260 | 64345407 | | | | |
| CL_2068 | 63 | 7283377 | 7283524 | LRIT1 | 100677675 | leucine-rich repeat, immunoglobulin-like and transmembrane domains 1 | No |
| CL_2127 | 34 | 23647405 | 23647552 | ST3GAL3 | 100668469 | CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase | No |
| CL_2184 | 24 | 14933120 | 14933267 | | | | |
| CL_2193 | 96 | 644047 | 644194 | DHRS7B | 100675045 | dehydrogenase/reductase SDR family member 7B | No |
| CL_2249 | 24 | 32999938 | 33000085 | EDA | 100666834 | ectodysplasin-A | No |
| CL_2425 | 19 | 36243153 | 36243300 | MROH8 | 104846046 | protein MROH8 | No |
| CL_2561 | 49 | 4778472 | 4778619 | TNRC6C | 100671590 | trinucleotide repeat-containing gene 6C protein | No |
| CL_2600 | 5 | 71457409 | 71457556 | MAGI2 | 100666628 | membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2 | No |
| CL_2831 | 47 | 12807894 | 12808041 | | | | |
| CL_2938 | 32 | 24769369 | 24769516 | LOC100658789 | 100658789 | uncharacterized protein C1orf228 homolog | No |
| CL_2968 | 11 | 9507799 | 9507946 | UGGT2 | 100672519 | UDP-glucose glycoprotein glucosyltransferase 2 | No |
| CL_3094 | 13 | 10931115 | 10931262 | SLC26A9 | 100662751 | solute carrier family 26 member 9 | Yes |
| CL_3102 | 18 | 420129 | 420276 | | | | |
| CL_3125 | 18 | 24915209 | 24915356 | DHX15 | 100655268 | putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 | No |
| CL_3334 | 40 | 13540880 | 13541027 | | | | |
| CL_3377 | 27 | 32053321 | 32053468 | | | | |
| CL_3594 | 68 | 555985 | 556132 | | | | |
| CL_3673 | 4 | 16047337 | 16047484 | LOC100668429 | 100668429 | NACHT, LRR and PYD domains-containing protein 7-like | No |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CL_3697 | 105 | 849381 | 849528 | | | | |
| CL_3702 | 54 | 13546559 | 13546706 | | | | |
| CL_3708 | 3 | 53879928 | 53880075 | | | | |
| CL_3824 | 34 | 16335133 | 16335280 | | | | |
| CL_3834 | 11 | 2956312 | 2956459 | | | | |
| CL_4032 | 18 | 2339429 | 2339576 | ADD1 | 100668539 | alpha-adducin | No |
| CL_4133 | 29 | 2271058 | 2271205 | | | | |
| CL_4194 | 40 | 5861189 | 5861336 | | | | |
| CL_4366 | 12 | 15593298 | 15593442 | | | | |
| CL_4367 | 9 | 53302178 | 53302325 | | | | |
| CL_4629 | 18 | 4894279 | 4894409 | DGKQ | 100676890 | diacylglycerol kinase theta | No |
| CL_4719 | 14 | 8569169 | 8569316 | | | | |
| CL_4878 | 38 | 2929153 | 2929300 | CWH43 | 100677091 | PGAP2-interacting protein | No |
| CL_5010 | 45 | 2903030 | 2903177 | LOC100676817 | 100676817 | uncharacterized protein C7orf50 homolog | No |
| CL_5055 | 79 | 208492 | 208639 | | | | |
| CL_5374 | 83 | 76682 | 76829 | LOC100669046 | 100669046 | T-cell receptor beta-1 chain C region-like | No |
| CL_5423 | 19 | 48225089 | 48225236 | | | | |
| CL_5488 | 2 | 21485322 | 21485469 | | | | |
| CL_5547 | 7 | 4094797 | 4094944 | | | | |
| CL_5690 | 18 | 20538405 | 20538552 | KCNIP4 | 100663977 | Kv channel-interacting protein 4 | No |
| CL_5786 | 61 | 5892680 | 5892827 | | | | |
| CL_5878 | 31 | 18255733 | 18255880 | SGCA | 100669677 | alpha-sarcoglycan | No |
| CL_6355 | 18 | 38485848 | 38485995 | | | | |
| CL_6363 | 0 | 61504808 | 61504954 | | | | |
| CL_6583 | 9 | 81721134 | 81721281 | KIF26A | 100657429 | kinesin-like protein KIF26A | Yes |
| CL_6769 | 73 | 9250020 | 9250167 | CACNG2 | 100671412 | voltage-dependent calcium channel gamma-2 subunit | No |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CL_6883 | 106 | 358541 | 358688 | | | | |
| CL_6951 | 113 | 2952879 | 2953024 | | | | |
| CL_7275 | 28 | 10768415 | 10768561 | | | | |
| CL_7365 | 121 | 1573529 | 1573677 | | | | |
| CL_7618 | 63 | 13308326 | 13308473 | TBRG4 | 100671408 | protein TBRG4 | No |
| CL_7666 | 10 | 60468324 | 60468471 | | | | |
| CL_7790 | 226 | 185127 | 185274 | | | | |
| CL_7859 | 72 | 303422 | 303569 | HTRA1 | 100675235 | serine protease HTRA1 | No |
| CL_8200 | 28 | 30549239 | 30549386 | | | | |
| CL_8219 | 70 | 1471021 | 1471168 | FAM222A | 100656053 | protein FAM222A | Yes |
| CL_8307 | 12 | 42774714 | 42774861 | ITPR1 | 100665028 | inositol 1,4,5-trisphosphate receptor type 1 | No |
| CL_8379 | 75 | 7789300 | 7789447 | BMP7 | 100666495 | bone morphogenetic protein 7 | No |
| CL_8615 | 8 | 78449021 | 78449168 | ST3GAL1 | 100667581 | CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1 | No |
| CL_8653 | 65 | 12734846 | 12734993 | ARMC5 | 100663645 | armadillo repeat-containing protein 5 | No |
| CL_8664 | 126 | 1577595 | 1577742 | | | | |
| CL_8743 | 5 | 22968201 | 22968348 | DPY19L2 | 100668800 | probable C-mannosyltransferase DPY19L2 | No |
| CL_8818 | 1 | 57437643 | 57437790 | NRG2 | 100670957 | pro-neuregulin-2, membrane-bound isoform | No |
| CL_8848 | 28 | 2078310 | 2078457 | | | | |
| CL_8849 | 10 | 67065346 | 67065493 | WDFY4 | 100663596 | WD repeat- and FYVE domain-containing protein 4 | No |
| CL_8854 | 117 | 2679136 | 2679283 | | | | No |
| CL_8896 | 3 | 79580074 | 79580219 | KIF5C | 100658646 | kinesin heavy chain isoform 5C isoform X1 | No |
| CL_9037 | 9 | 42077648 | 42077795 | | | | |
| CL_9115 | 6 | 44126042 | 44126189 | | | | |
| CL_9162 | 21 | 35383824 | 35383971 | | | | |
| CL_9201 | 43 | 2706151 | 2706298 | MMP2 | 100659450 | 72 kDa type IV collagenase | No |
| CL_9263 | 31 | 12365891 | 12366038 | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CL_9467 | 13 | 16077912 | 16078059 | | | | |
| CL_9552 | 1 | 95311297 | 95311444 | STK10 | 100664614 | serine/threonine-protein kinase 10 | Yes |
| CL_9729 | 107 | 1197980 | 1198127 | | | | |
| CL_9831 | 8 | 2426825 | 2426972 | | | | |
| CL_9867 | 8 | 32108053 | 32108200 | RALYL | 100661139 | RALY RNA binding protein-like | No |
| CL_9914 | 34 | 8918891 | 8919038 | | | | |
| CL_10073 | 2 | 46495691 | 46495838 | LRP1 | 100660360 | prolow density lipoprotein receptor-related protein 1 | Yes |
| CL_10251 | 70 | 1740330 | 1740477 | UBE3B | 100656625 | ubiquitin-protein ligase E3B | No |
| CL_10364 | 25 | 35086962 | 35087109 | PHLDB2 | 100656595 | pleckstrin homology-like domain family B member 2 | No |
| CL_11132 | 79 | 1692805 | 1692952 | | | | |

**Appendix 4** A modified protocol for DNA extraction from faecal samples collected using a swab. Faeces were scrubbed using a buccal swab (Isohelix, Cell projects) and preserved into storage buffer (500 µl of LS buffer and 25 µl of proteinase K, Stabilizing kit, Isohelix, Cell Projects). The protocol is derived from the QIAamp Fast Stool Mini kit (51604) protocol (QIAGEN) and the Isohelix DNA Isolation kit (DDK-50) protocol (Cell Project).

| Step | Description |
|---|---|
| 1 | Vortex the 2-ml tube containing the sample (swab tip in buffer solution). Centrifuge during 2 min (14,100 g). Discard the swab. |
| 2 | Pipette supernatant into a clean 1.5ml tube (~ 450ul). Add 250 µl InhibitEx buffer and vortex immediately for at least one minute. Leave for a minute at ambient temperature. |
| 3 | Centrifuge for 2 min (14,100 g). Pipette supernatant into a new 1.5 ml tube containing 25 µl of proteinase K. Vortex. |
| 4 | Incubate at 56°C for 1 hour. |
| 5 | Centrifuge for 2 min (14,100 g). If stool particles deposited at bottom of tube, pipette supernatant into a clean 1.5 ml tube. |
| 6 | Add 500 µl CT solution and vortex immediately. |
| 7 | Pipette 600 µl into the QIAamp spin column. Centrifuge 1 min (14,100 g). Place the pin column into a new collection tube and repeat until all the lysate has been loaded on the spin column. Place the spin column into a new collection tube. |
| 8 | Add 500 µl buffer AW1. Centrifuge 1 min (14,100 g). Place the spin column into a new collection tube. |
| 9 | Add 500 µl buffer AW2. Centrifuge 1 min (14,100 g). Place the spin column into a new collection tube. |
| 10 | Centrifuge 3 min (14,100 g). Place the spin column into a clean 1.5 ml tube. |
| 11 | Add 75 µl buffer ATE. Incubate 2 min at 56°C. Centrifuge 1 min (14,100 g). |

**Appendix 5** Effects of storage time and faecal quality on probability to extract elephant DNA from faeces.

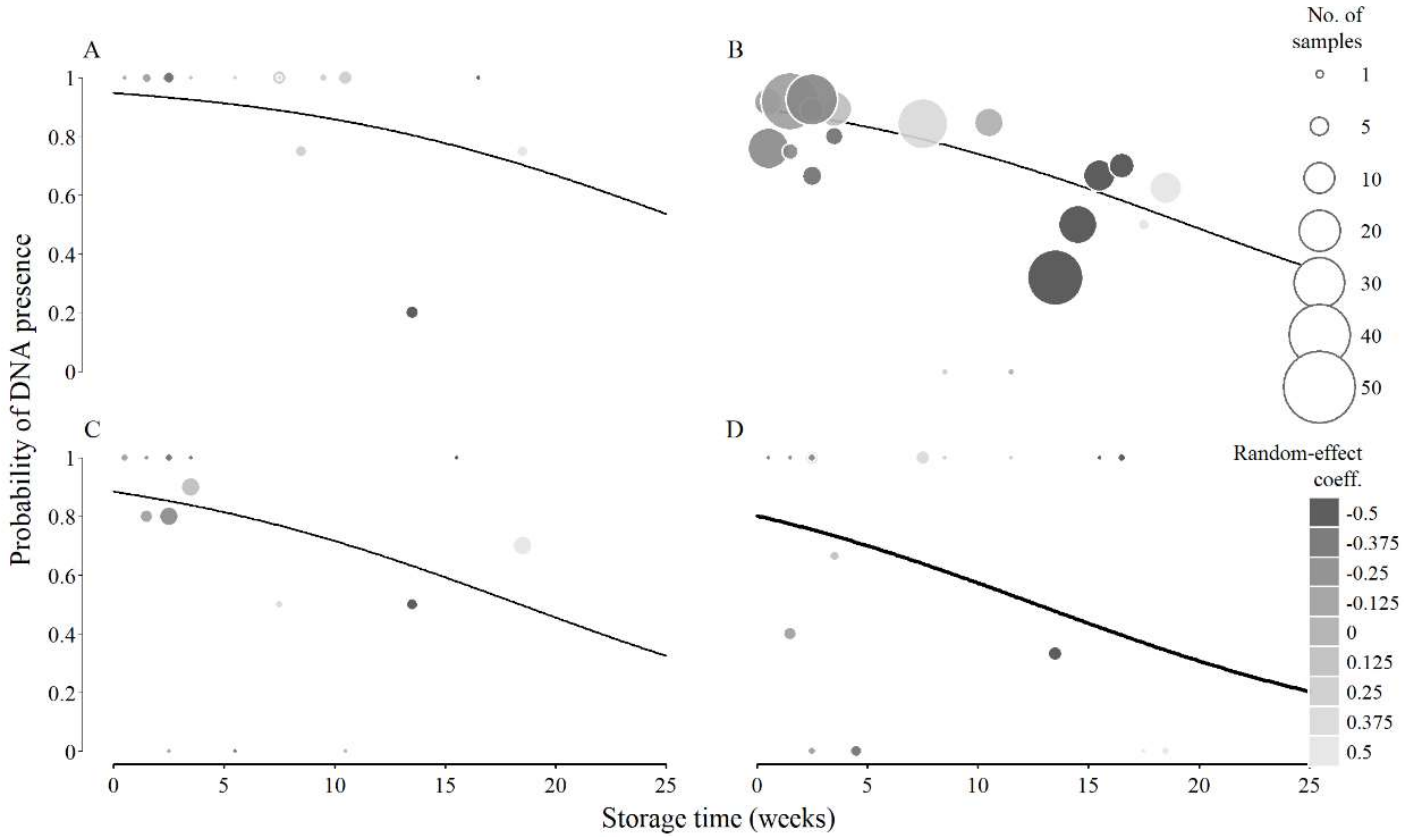


Figure A5.1 Observed and predicted probability to extract elephant DNA from faeces per week of storage for different faecal qualities. The four categories of faecal quality were: (A) very fresh (collected within 1 hour after defecation), (B) fresh (collected within 24 hours after defecation), (C) reduced surface (less than 24 hours old but partly destroyed by insects or exposed to direct sunlight), and (D) degraded (collected between 24 and 48 hours after defecation or found after rain or partly immersed in water). Observed data are represented by circles proportional to the number of samples collected and coloured according to random-effect coefficients for sampling locations. Details of the binomial generalized linear mixed model are given in table 3.

**Appendix 6** Effects of storage time and faecal quality on elephant DNA concentration.

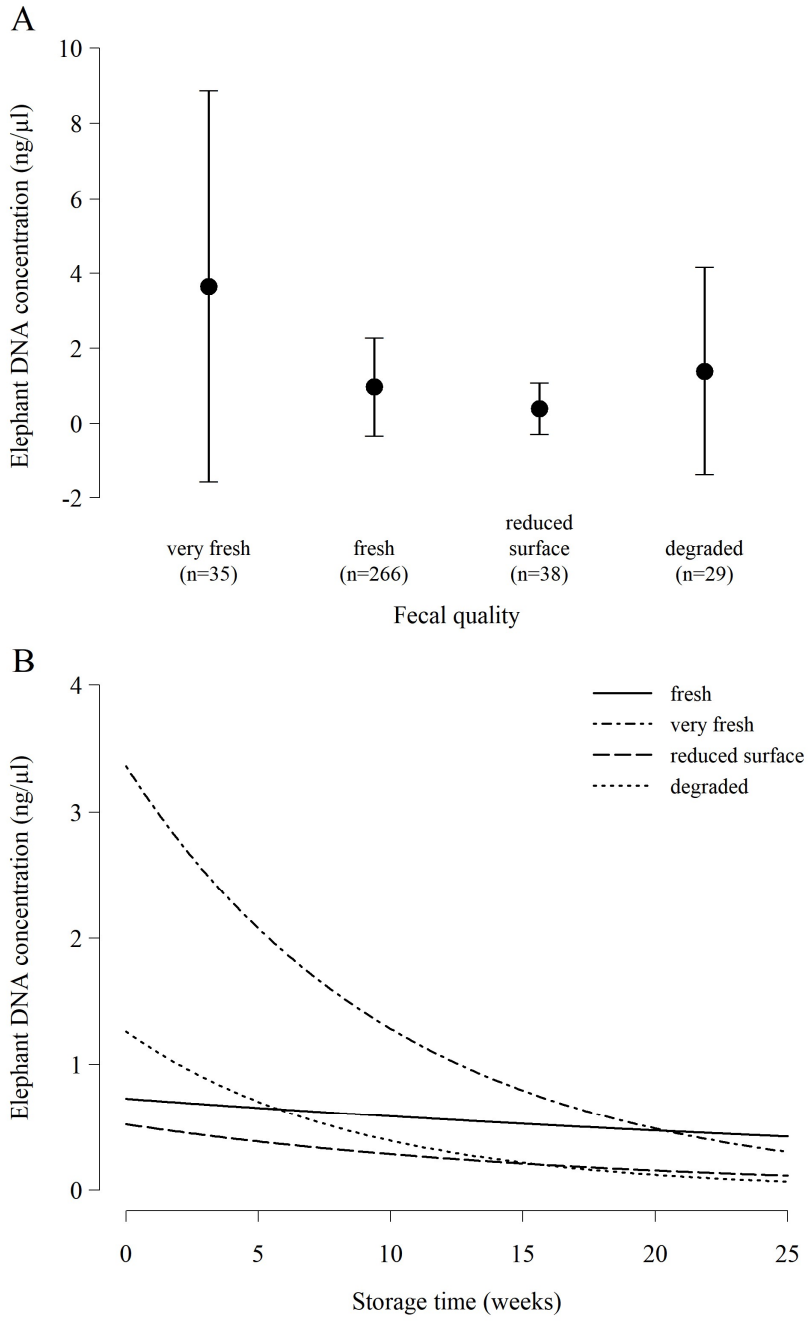


Figure A6.1 (A) Mean elephant DNA concentration (± SD) and (B) predicted elephant DNA concentration per week of storage for four categories of faecal quality.

Table A6.1 Summary of the best truncated negative binomial generalized linear mixed model using 396 faecal DNA extracts. Faecal quality was categorized into four groups: very fresh, fresh (reference category), reduced surface and degraded. Sampling location was included as random effect.

| Variable | Coeff. (±SE) | Z | p-value |
|---|---|---|---|
| Fixed effects | | | |
| Intercept | 6.428 ±0.131 | 49.23 | <0.001 |
| Storage time | -0.126 ±0.079 | -1.59 | 0.112 |
| Very fresh | 0.959 ±0.123 | 7.80 | <0.001 |
| Reduced surface | -0.627 ±0.153 | -4.11 | <0.001 |
| Degraded | -0.176 ±0.160 | -1.10 | 0.272 |
| Storage time:very fresh | -0.442 ±0.118 | -3.74 | <0.001 |
| Storage time:reduced surface | -0.230 ±0.147 | -1.57 | 0.117 |
| Storage time:degraded | -0.562 ±0.200 | -2.81 | 0.005 |
| Random effects | | | |
| No. groups | 9 | | |
| Variance | 0.126 | | |
| SD | 0.355 | | |

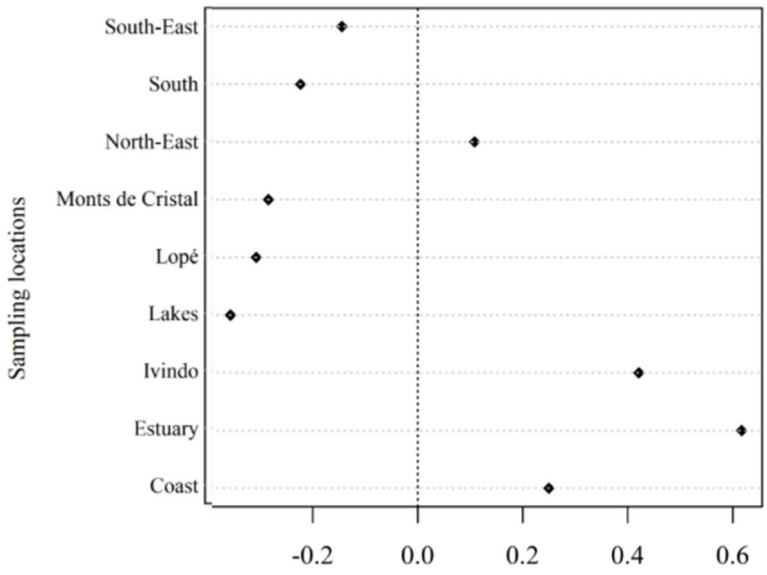


Figure A6.2 Coefficients of random effects for the 9 sampling locations in the best truncated negative binomial generalized linear mixed model for the effects of storage time and faeces quality on elephant DNA concentration.

Appendix 7 Map of the spatial principal component lagged scores for forest elephant samples across Gabon.

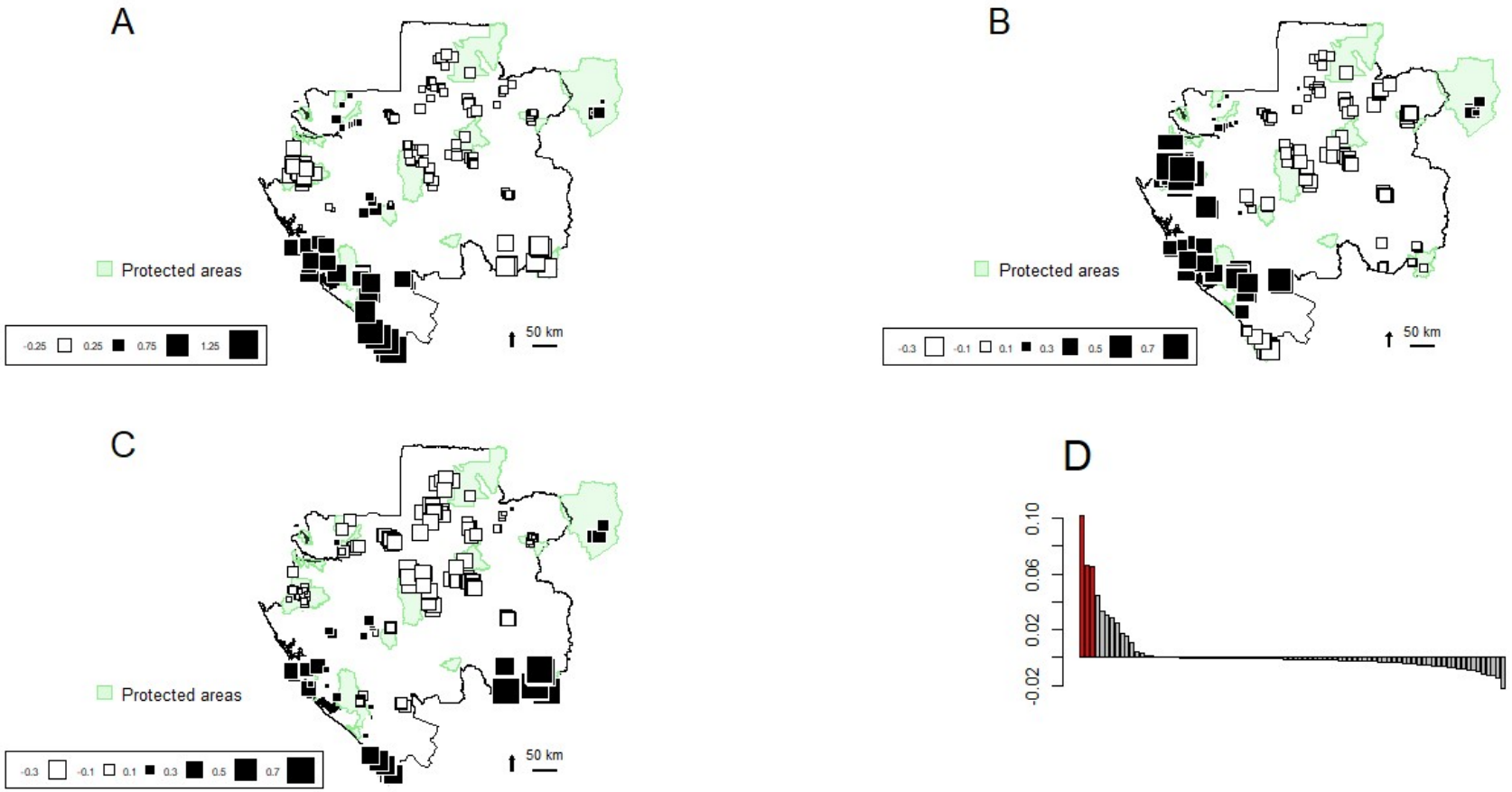


Figure A7.1 Map of the first (A), second (B) and third (C) principal components of sPCA across forest elephants in Gabon. Negative and positive values of spatially lagged scores are represented by white and black squares, respectively (size is proportional to absolute values). D) Barplot of eigenvalues corresponding to global (positive) and local (negative) patterns. The three first components (in red) were retained).

Appendix 8 Scatterplots of principal component analysis for the genetic structure of forest elephants in Gabon using 86 SNP markers.

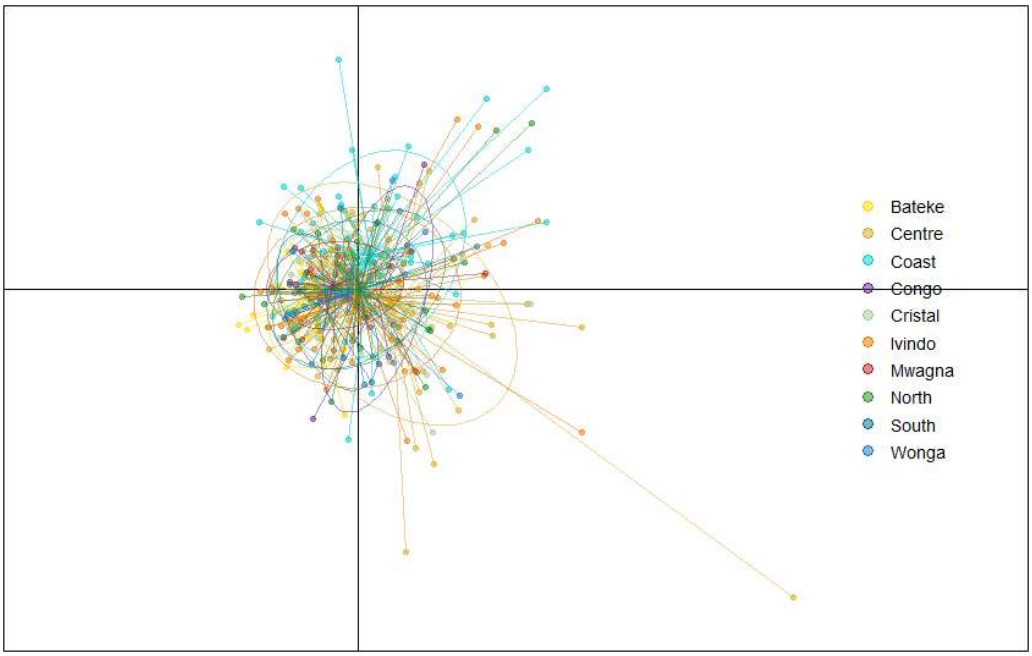


Figure A8.1 First two principal components (PC1 and PC2). Each sample is represented by a dot and coloured according to its sampling location. Inertia ellipses include 67% of the samples for each location.

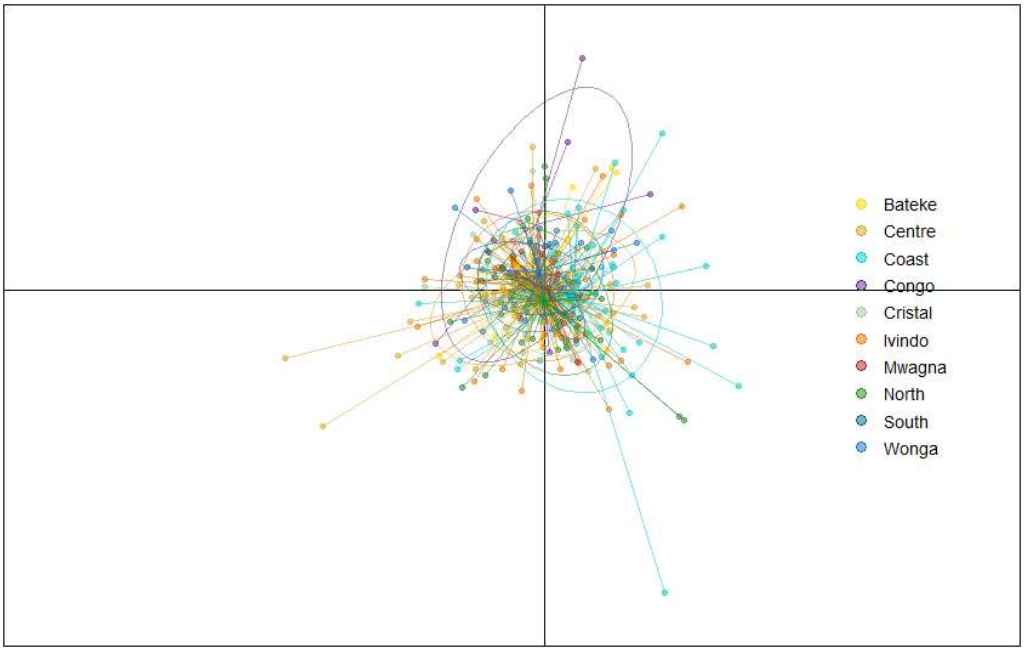


Figure A8.2 Second and third principal components (PC2 and PC3). Each sample is represented by a dot and coloured according to its sampling location. Inertia ellipses include 67% of the samples for each location.