Landscape genetics of Guinea baboon: assessing population structure, gene flow dynamics, and functional connectivity with molecular and spatial tools.

Helena Marisa Osório Teixeira
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
2016

Orientador
Maria Joana Ferreira da Silva, Post-Doc, CIBIO-InBIO/Cardiff University

Coorientador
José Carlos Alcobia Rogado de Brito, Senior Scientist, Associate Researcher, CIBIO-inBIO/FC-UP

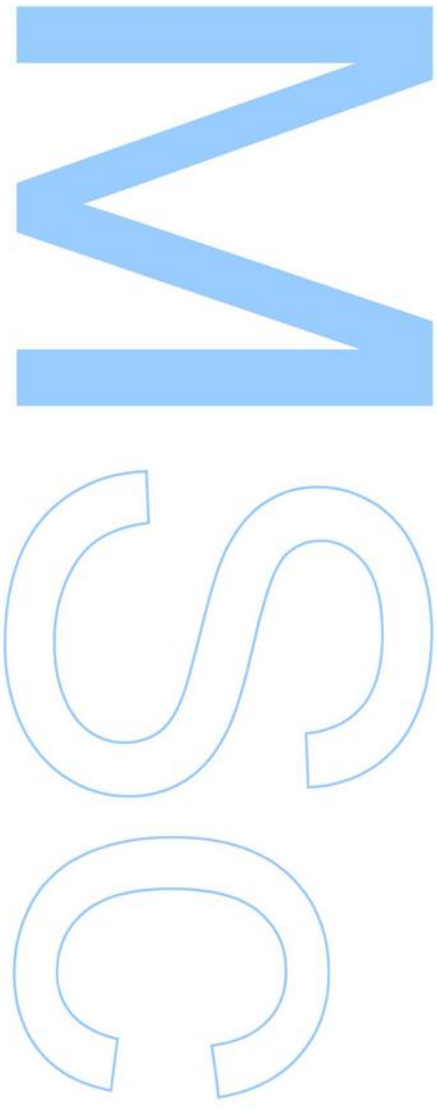
U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Em memória do meu avô, Adelino de Freitas,

com saudades infinitas.

Agradecimentos

Nos últimos anos, todos os meus passos foram desenhados para um dia chegar a este momento. Mas mais do que uma conquista pessoal, este documento é o resultado de todo o apoio científico e emocional que recebi ao longo de toda a caminhada. Foram muitos os intervenientes desta conquista, aos quais expresso o meu mais sincero voto de gratidão:

À minha orientadora, Joana Silva, por me ter aberto a porta ao maravilhoso Mundo dos babuínos e por toda a ajuda ao longo e, principalmente, na fase final da tese.

Ao meu orientador, José Carlos Brito, por me ter desafiado a desenvolver este trabalho, pela serenidade que sempre me transmitiu, pelo discernimento nos momentos de incerteza e por toda a sabedoria e ajuda.

Ao João Campos, para mim, o terceiro orientador desta tese. Por todas as vezes que se sentou ao meu lado nos barracões, pela preciosa ajuda nas análises espaciais, pela partilha de conhecimentos e, sobretudo, pela amizade.

À Cândida Vale por ter partilhado comigo o seu trabalho de modelação do Babuíno da Guiné e pelas seus conselhos.

To Gisela Kopp and Vitor Calisto for sharing their genetic data, years of work and knowledge with me.

To Clive Barlow and Dietmar Zinner who gently provided the samples for the genetic analysis. I also thank Dietmar Zinner for the opportunity to spend two months in the German Primate Centre (DPZ) and to learn primatology.

To everyone that helped during the field missions to Guinea-Bissau, Mauritania, Senegal, Guinea-Conakry and Mali.

To the Gesellschaft für Primatologie which awarded me a Robert Glaser Travel Grant to support my internship in DPZ.

To everyone that supported the genetic analyses of this project, especially the projects DESERTFLOW (PTDC/BIA-BIC/2003/2012) and PRIMATOMICS (PTDC/IVC-ANT/3058/2014), which financial support was provided by the Fundação para a Ciência

e Tecnologia (FCT), and the Born Free Foundation and Chester Zoo Conservation and Research Grant.

To everyone that helped me in the lab, mainly Rasmus Liedigk and Christian Roos (at DPZ), and Susana Lopes and Vânia Costa (at CIBIO-inBIO).

À Raquel Godinho, à Mafalda Costa, ao Guilhermo Velo-Antón e a todos os demais que gentilmente contribuiram com ideias para esta tese ou para a defesa.

Aos amigos de sempre e aos que fiz no âmbito deste mestrado e que ficaram para sempre, pelo carinho, paciência e amizade.

Ao meu Carlinhuz e à minha Anita, que possuem um estatuto tão especial que seria injusto não terem a devida distinção. Obrigada por fazerem do meu Mundo um lugar único e feliz.

À Susana Rodríguez Echeverría, por ter sido a primeira pessoa a acreditar em mim, por me ter encorajado a seguir uma carreira científica, e por ser a minha maior referência.

À minha familia, especialmente ao meu avô, que foi a pessoa que mais me mimou ao longo do meu percurso académico. Por todo o carinho com que falava de mim a toda a gente e por todo o orgulho que tinha “na neta que um dia seria doutora”. Espero que todos os meus passos o possam continuar a encher sempre de orgulho e admiração.

Resumo

A fragmentação do habitat é a maior ameaça à Biodiversidade em todo o Mundo. Apesar dos eventos naturais serem uma das causas da fragmentação, as actividades humanas são responsáveis pela transformação da paisagem numa matriz heterogénea, contendo mosaicos de habitats adequados e não adequados. Consequentemente, as populações podem tornar-se pequenas e isoladas e, em última instância, empobrecidas geneticamente. A diversidade genética é essencial para a conservação a longo-prazo das espécies, e a base para se adaptarem a potenciais alterações ambientais. A manutenção do fluxo genético entre as populações poderá evitar a erosão genética, ao prevenir o isolamento reprodutivo, a fixação e perda de alelos, e a acumulação de mutações deletérias. Assim, a persistência de uma espécie animal numa paisagem fragmentada depende da sua capacidade de dispersão na mesma, e da ocorrência de conectividade funcional entre os fragmentos de habitat. A genética da paisagem constitui um método directo para determinar a conectividade funcional. O Babuíno da Guiné (*Papio papio*) é uma espécie ecologicamente bem adaptada a uma vasta gama de habitats na África Ocidental, incluindo montanhas, bosques e florestas húmidas. As actividades humanas, como a transformação do habitat, a desflorestação e a caça, reduziram a distribuição do Babuíno da Guiné nos últimos 30 anos. No presente estudo, dados genéticos de 507 indivíduos amostrados em 10 populações ao longo da África Ocidental, foram combinados com análises espaciais para investigar as áreas mais adequadas à dispersão do Babuíno da Guiné e avaliar qual o mecanismo de isolamento que melhor explica a sua estrutura genética. Os indivíduos foram genotipados utilizando um máximo de 23 marcadores microssatélites autossómicos. Um conjunto de 15 marcadores microssatélites associados ao cromossoma Y foram testados para procurar um marcador variável para o Babuíno da Guiné. Dois modelos de nicho ecológico foram gerados separadamente, de modo a avaliar o impacto dos factores ambientais e dos factores humanos na distribuição da espécie. A teoria eléctrica de circuitos foi utilizada para prever a conectividade funcional do Babuíno da Guiné e Mantel testes simples e parciais foram utilizados para avaliar as hipóteses de isolamento por distância e de isolamento por resistência. Dos 15 marcadores microssatélites associados ao cromossoma Y testados, apenas um mostrou ser polimórfico, o que sugere baixos níveis de polimorfismo nesse cromossoma no Babuíno da Guiné. As análises genéticas revelam diferenças entre as populações localizadas nos extremos Norte, Ocidental e Oriental da distribuição da espécie. Os modelos de nicho ecológico prevêem a existência de áreas adequadas à presença do Babuíno da Guiné fora da sua actual área

de distribuição, o que implica que a distribuição IUCN deverá ser actualizada. A população localizada no Sudeste da sua distribuição, poderá representar a zona de hibridação entre o Babuíno da Guiné e o Babuíno-anúbis. As análises espaciais sugerem que a hipótese de isolamento por distância é aquela que melhor explica a estrutura genética do Babuíno da Guiné. Os factores da paisagem que poderiam actuar como barreiras ao fluxo genético não explicam totalmente as descontinuidades genéticas encontradas, enfatizando a plasticidade e habilidade deste primata em adaptar-se a paisagens heterogéneas e fragmentadas. No futuro, esforços são necessários para encontrar um marcador associado ao cromossoma Y variável para o Babuíno da Guiné. Amostras de áreas não analisas são necessárias a fim de detalhar a diversidade genética e conectividade funcional ao longo de toda a área de distribuição. Futuros estudos são, ainda, necessários para confirmar a ocorrência de hibridação entre o Babuíno da Guiné e o Babuíno-anúbis no Sudeste da distribuição do Babuíno da Guiné.

Palavras-chave: Babuíno da Guiné, Paisagem Heterogénea, Diversidade Genética, Estrutura populacional, Modelos de nicho ecológico, Conectividade Funcional, Isolamento por distância, Isolamento por resistência.

Abstract

Habitat fragmentation represents one of the greatest threats to biodiversity worldwide. Although it can be the result of natural events, human activities are largely responsible for the transformation of landscapes into heterogeneous matrixes containing mosaics of suitable and unsuitable habitats. As a consequence of fragmentation, populations may become small and isolated and, ultimately, genetically impoverished. Genetic diversity is crucial for the long-term persistence of a species and the basis for species' adaptability to potential environmental changes. Maintaining gene flow between populations may avert genetic erosion, preventing reproductive isolation and the fixation or loss of alleles and accumulation of deleterious mutations. Therefore, the persistence of animal species in fragmented landscapes is related with the animals' ability to disperse across heterogeneous landscapes and the occurrence of functional connectivity among isolated patches. Landscape genetics provides a direct way of determining functional connectivity. The Guinea baboon (*Papio papio*) is a well-adapted primate species distributed across a wide range of habitats in West Africa, including mountains, woodlands, humid high forests, and mangrove forests. Human activities, such as habitat conversion, deforestation and hunting, significantly reduced the distribution of Guinea baboons and increased fragmentation during the last 30 years. In this study, genetic data of 507 Guinea baboon individuals collected in 10 populations across West Africa were combined with spatial analyses to investigate the most suitable areas of dispersal and the isolation mechanisms related to its genetic structure. The individuals were genotyped for a maximum of 23 autosomal microsatellites *loci*. A variable Y-linked microsatellite marker was investigated by testing a set of 15 markers previously developed for humans. Two ecological niche-based models were generated to separately compare the impacts of environmental factors and human-related features on the distribution. The electrical circuit theory was used to predict the functional connectivity and simple and partial Mantel tests were used to evaluate hypotheses of isolation by distance and isolation by resistance. Of the 15 Y-linked microsatellites *loci* tested, only one was polymorphic, suggesting low levels of polymorphism associated with the Y chromosome in the Guinea baboon. Genetic analyses suggested major differences between the populations located in the northern, eastern and western extremes of the distribution. Ecological models predicted the existence of suitable areas outside the known distribution of Guinea baboon, indicating that the IUCN distribution area should be updated. The population located in the south-eastern distribution of the species likely represents the hybridization zone between Guinea and Olive baboons. The spatial analyses suggested that the

isolation by distance hypothesis was the most probable explanation for the observed genetic structure in Guinea baboon. The landscape features that could act as barriers to gene flow did not fully related with the genetic discontinuities found, which emphasizes the plasticity and ability of Guinea baboons to adapt to heterogeneous and fragmented landscapes. In the future, extensive efforts may be required to find a variable Y-chromosomal marker for Guinea baboon. Samples from currently unsampled areas are needed to detail genetic diversity and functional connectivity across the full species range. Further studies are needed to confirm the occurrence of hybridization between Guinea baboon and Olive baboon in the south-eastern distribution of Guinea baboon

Key-words: Guinea baboon, Heterogeneous landscape, Gene flow, Genetic diversity, Population structure, Ecological niche-based Models, Functional connectivity, Isolation by Distance, Isolation by Resistance.

Table of Contents

List of Figures

List of Tables

Abbreviations

IBD - isolation by distance

IBB - isolation by barrier

IBR - isolation by resistance

ENMs - Ecological niche models

MAXENT - Maximum Entropy model

LCP - Least cost pathway

km – kilometres

NT - Near Threatened

IUCN - International Union for Conservation of Nature

mtDNA – mitochondrial DNA

COI - Cytochrome C Oxidase subunit I

ADO - Allelic dropout

FA – False alleles

GPS – Global Positioning System

µl – Microliter

mg – Milligrams

PCR – Polymerase Chain Reaction

AT - Annealing Temperature

QI - Quality Index

bp – base pairs

NCBI - National Center for Biotechnology Information database

BLASTn - Basic Local Alignment Search Tool

NJ - Neighbor Joining

MCMC - Markov Chain Monte Carlo

K – Genetic number of clusters

HEW - Hardy–Weinberg equilibrium

LD - linkage disequilibrium

Na – number of different alleles per locus

Ho - observed heterozygosity

He - expected heterozygosity

F_{IS} - Inbreeding coefficient

F_{ST} – Fixation Index

%P - percentage of loci polymorphic

PA - number of private alleles

DAS – Distance Shared Allele

NNI - Nearest Neighbour Index

EGVs - ecogeographical variables

NDVI - Normalized Difference Vegetation Index

PET - Potential Evapotranspiration

WGS - World Geodetic System

UTM - Universal Transverse Mercator

PCA - Principal Component Analysis

AUC - Area Under the Curve

ROC - Receiver operating characteristic

SD - standard deviation

pID - probability of identify

pIDsibs - probability of siblings identify

DPZ – German Primate Center

1. Introduction

1.1. Connectivity

1.1.1. Landscape connectivity

Habitat fragmentation is a process by which a continuous habitat is progressively subdivided into multiple, smaller and isolated patches (Fahrig, 2003), and it represents one of the greatest threats to biodiversity worldwide (Myers et al., 2000). Fragmentation can be the result of natural events, such as geological processes and climatic fluctuations, or of human activities, such as agriculture and urbanization. Human activities may modify the landscape in a mosaic of settlements, farmlands and scattered fragments of natural ecosystems (Fahrig, 2003). Fragmentation processes drive landscape transformation from a continuous portion of natural habitats into a heterogeneous matrix containing mosaics of suitable and unsuitable habitats (Höglund, 2009) (Figure 1).

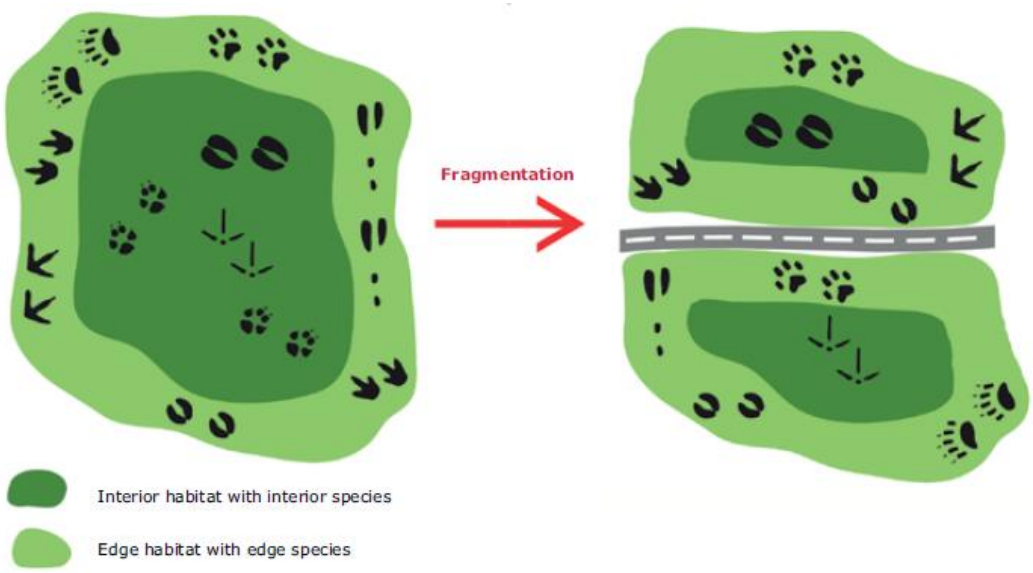


Figure 1. Hypothetical effects of habitat loss and fragmentation caused by road construction. Interior habitat area and diversity of interior species decreased as a result of fragmentation. Adapted from EEA Report (2011).

The persistence of animal species in fragmented areas is related with the animals' ability to disperse across heterogeneous landscapes and maintenance of connectivity among remaining populations (Quéméré et al., 2010a). Dispersal is the process by which animals are driven out of their birth region to reproduce in another region, allowing the incorporation of genes of one population into the gene pool of a

distinct population (Manel and Holderegger, 2013). This process is an important behaviour promoting gene flow and is one of the main determinants of genetic diversity levels and degree of population genetic structure (Manel and Holderegger, 2013). Genetic diversity is the basis for species' adaptability potential to environmental changes (Lynn et al., 2016). Decreased ability to disperse across habitat patches as consequence of habitat fragmentation can result in: (i) rapid population decline; (ii) lower genetic and phenotypic variability through the fixation or loss of alleles and accumulation of deleterious mutations; (iii) increased vulnerability to stochastic demographic fluctuations; (iv) increased risk for inbreeding depression given that mating choices may be limited to kin in isolated populations, that results in intensification of extinction risk for groups or populations (Frankham, 2002; Quéméré et al*.,* 2010a; Manel and Holderegger, 2013; Jonsson et al., 2016).

Landscape connectivity is essential for maintaining viable populations, particularly for the ones restricted to small and fragmented habitats, for which the area of suitable habitat, quality and level of connectivity between sub-populations has been reduced (Castillo et al., 2016). Also, it is important for species living in climate-extreme environments, which are naturally patchy and have temporally variable resources (Velo-Antón et al., 2013). Landscape connectivity is defined as the degree to which the landscape facilitates or impedes movements among resource patches (Taylor et al., 1993). Two components of connectivity are usually recognized: structural and functional connectivity (Manel and Holderegger, 2013). The former is related to the spatial configuration of the habitat patches in the landscape (e.g. shape, size, presence of barriers), whereas the latter refers to the response of individuals to landscape features in terms of dispersal behaviour, mortality risks and inferred costs to movement (Manel and Holderegger, 2013; Imong et al., 2014). However, both components are tightly associated. Since dispersal behaviour changes with landscape configuration, it is impossible to understand landscape connectivity without taking into account both structural and biological attributes (Baguette and Dyck, 2007).

The establishment of migration corridors have been suggested as a solution to mitigate the effects of habitat fragmentation (Stevens et al., 2006). Landscape connectivity is increasingly used to identify the most suitable areas in the matrix to restore and maintain connectivity among populations (Quéméré et al., 2010a). Dispersal corridors may enable movements of individuals among isolated populations, allowing inputs of new genes from adjacent populations, which may prevent the loss of genetic diversity and maintain the species potential to survive to environmental changes (Massot et al., 2008).

1.1.2. Landscape genetics

Understanding the patterns of gene flow requires a detailed knowledge of how landscape characteristics structure populations (Manel et al., 2003). A landscape is defined as a heterogeneous land area containing a mosaic of land cover types, which is described by its composition and configuration (Arroyo-Rodriguez et al., 2014). The former quantifies types and proportions of different forms of land cover across the landscape, and the latter quantifies the spatial arrangement of a given landscape composition.

The improvements in molecular genetic tools together with the increasingly availability of geospatial environmental data has led to the emergence of the field of landscape genetics in 2003 (Manel et al., 2003). Landscape genetics is an interdisciplinary research field that combines population genetics and landscape ecology in order to quantify the effect of landscape composition, configuration and matrix quality on spatial patterns of genetic variation (Quéméré et al., 2010a). In other words, landscape genetics allows the detection of genetic discontinuities in space and the interpretation of these discontinuities in relation with landscape features (Manel et al., 2003). In the past decade, landscape genetics tools have been used to: (i) access functional connectivity in heterogeneous and fragmented landscapes; and (ii) to identify the main barriers to gene flow (e.g. roads, rivers or mountains) and (iii) locate dispersal corridors used by the individuals (Manel et al., 2003 and Manel and Holderegger, 2013).

Landscape genetics studies often attempts to test three mechanisms leading to genetic differentiation: isolation by distance (IBD), isolation by barrier (IBB), and isolation by resistance (IBR) (Figure 2). The IBD hypothesis considers the existence of a significant positive correlation between geographic distances (often measured as Euclidean distance) and measures of genetic distance (i.e. genetic dissimilarities increase with increasing geographical distance) (Manel and Holderegger, 2013). Genetic discontinuities can also be explained by the existence of barriers in the landscape that reduces the individual's dispersal ability and, consequently, decrease gene flow between adjacent populations (Shirk et al., 2012). Under the IBB hypothesis, the populations are genetically distinct from those between barriers (e.g. rivers, mountains, roads, highways) (Shirk et al., 2012). Finally, genetic discontinuities can be explained by the existence of selective pressures across the landscape. According to the IBR hypothesis, the landscape is perceived as a complex resistant surface, in which the resistance to animals' movement is determined by landscape features (e.g. land cover, elevation, climate, slope). The IBR hypothesis is considered the most realistic one because it takes into account how organisms move through heterogeneous landscapes with varying

efficiency (McRae, 2006; Castillo et al., 2016). Therefore, contemporary landscape genetics studies have focussed on these hypotheses in order to assess how landscape structure affects gene flow across heterogeneous landscape (Zeller et al., 2012).

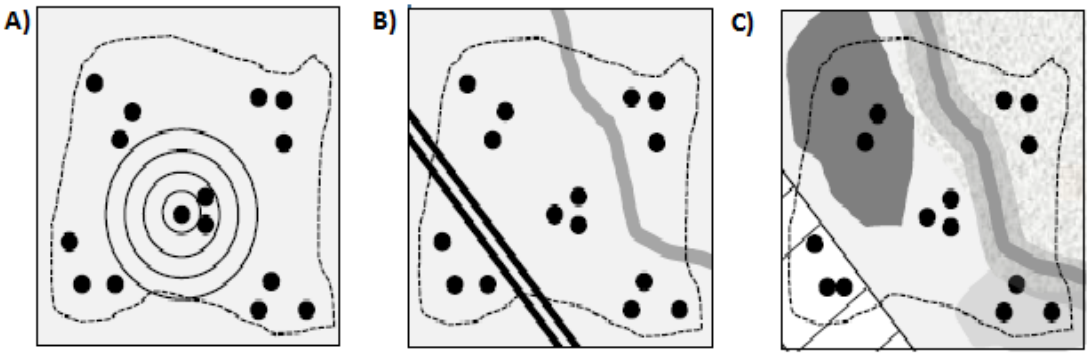


Figure 2. Representation of the three mechanisms leading to genetic differentiation. A) IBD - Isolation by distance; B) IBB - Isolation by barrier; C) IBR - Isolation by resistance. The dashed line represents the extent of the study area and dots the sampling locations. Adapted from Balkenhol and Fortin (2016).

Understanding which mechanism is driving genetic isolation and, consequently, shaping genetic structure and diversity is essential to the correct management of populations threatened by low connectivity. For instance, for populations isolated by distance, the mitigation efforts should be focussed on the establishment of stepping stone populations to reduce dispersal distances between breeding individuals; whereas, for population isolated by barrier, mitigation efforts must be focussed on implementation of structures crossing barriers (e.g. eco-passages). For populations isolated by resistance gradients, the establishing of resistance movement corridors linking habitat areas might constitute the optimal solution (Imong et al., 2004; Shirk et al., 2012; Castillo et al., 2016).

1.2. Tools to study landscape connectivity

1.2.1. Ecological Niche-based Models

The ecological niche can be defined as the bioclimatic envelop that limit and define the multidimensional space where one species is potentially able to maintain populations (Hutchinson, 1957). Ecological niche-based models (ENMs) are a class of methods that uses species occurrence data in conjunction with environmental variables to make a correlative model of the environmental preferences and tolerances of species and to predict the relative suitability of the habitat in the past, present or future environmental conditions (Warren and Seifert, 2011). Species occurrence data can be

obtained from observations during field expeditions, published studies, databases (e.g. GBIF - Global Biodiversity Information Facility) and natural history collections. Environmental data can be derived from field work, public databases (e.g. WorldClim) or from satellite imagery (Alvarado-Serrano and Knowles, 2014).

Different ENMs can be used to relate species occurrence data and environmental variables, according to data requirements. Some ENMs require the existence of presence and absence species data (e.g. Generalized Linear Model, GLM), while others can use presence-only data (e.g. Maximum Entropy model, MaxEnt). Absence data are usually difficult to obtain accurately and may be easily misidentified (Phillips et al., 2006; Alvarado-Serrano and Knowles, 2014). For this reason, MaxEnt software (Phillips et al., 2006) is one of the most popular tools for species distribution modelling. MaxEnt uses the principle of Maximum Entropy and it requires a list of species presence locations and a set of environmental predictors (e.g. precipitation, temperature) as input (Rissler, 2016). The environmental variables are associated with specific occurrence records and the suitable areas are identified by contrasting the environmental conditions from known species occurrences against the environmental variation across the study region (Phillips et al., 2006; Alvarado-Serrano and Knowles, 2014; Rissler, 2016).

Ecological niche-based models have been widely used as the first step to assess connectivity across the landscape. The resulting habitat suitability information can be translated into probable migration paths between populations using Graph based methods (Alvarado-Serrano and Knowles, 2014).

1.2.2. Graph based methods

A graph is a mathematical object comprising nodes - which represent patches - and edges - which represent paths among patches (Harary, 1969). Spatial graphs are used in landscape genetics to quantify the interactions among individuals and its dispersal behaviour (Manel and Holderegger, 2013). In this context, the landscape is considered as a spatial surface comprising a set of nodes (e.g. individuals, populations or habitat patches) connected by edges representing landscape suitability (i.e. patches used by individuals to disperse) (Manel and Holderegger, 2013). Therefore, graph based methods can measure functional connectivity, taking into account the species habitat preferences and movement behaviours (Rainey, 2009).

Animal movement is one of the most difficult behaviours to observe and quantity. Hence, resistance to movement values are used to fill this gap by providing a quantitative estimate of how the environment affect animal movement (Zeller et al., 2012). Resistance represents the willingness of an organism to cross a particular environment,

the physiological cost of moving through a particular environment, or the reduction in survival for the organism moving through a particular environment (Zeller et al., 2012). Resistance values are attributed to known or estimated relationships among the species ecological preferences and the landscape characteristics. Low resistance values denote easiness of movement while high values denote restriction to movement and the occurrence of possible dispersal barriers (Rainey, 2009; Spear et al., 2010). Consequently, the resultant resistance surface identifies different degrees of landscape permeability to species´ movement, allowing the identification of likely movement routes between habitat patches. There are two main sources of information to assign the resistance values to surface cells: *empirical data* and *quantitative data*. The former relies on the knowledge of researches about the ecology of a particular species to assign different resistance to landscape variables; while the latter use quantitative data (as presence/absence, radio telemetry or mark-recapture data) together with landscape features to obtain habitat suitability maps, where each map' pixel expresses an associated resistance value (Rainey, 2009; Spear et al., 2010; Zeller et al., 2012).

The development and use of resistance surfaces are associated with two Graph based methods: least-cost paths (LCP) and electrical circuit theory (Figure 3). The LCPs considers that the route than an individual is most likely to take when moving among patches is the one that offers the shortest cost-weighted distance. It assumes that the individual has knowledge of the surrounding landscape and will choose the minimum exposure to unsuitable habitat to disperse (Rainey, 2009; Spear et al., 2010). In the electrical circuit theory, all possible pathways across the landscape are taken into consideration. This method makes use of an analogy between the movement of charge through an electrical circuit and the movement of individuals through a landscape, in which the electrical current can be interpreted as the expected individual movement through a graph of nodes connected by resistors within a resistance surface. Resistance values are determined by the cells landscape resistance values (McRae, 2006; McRae et al., 2008). Given that current is not weighted by distance, this approach allows identifying suitable paths that were too long to be considered important routes in LCP analysis (Rainey, 2009).

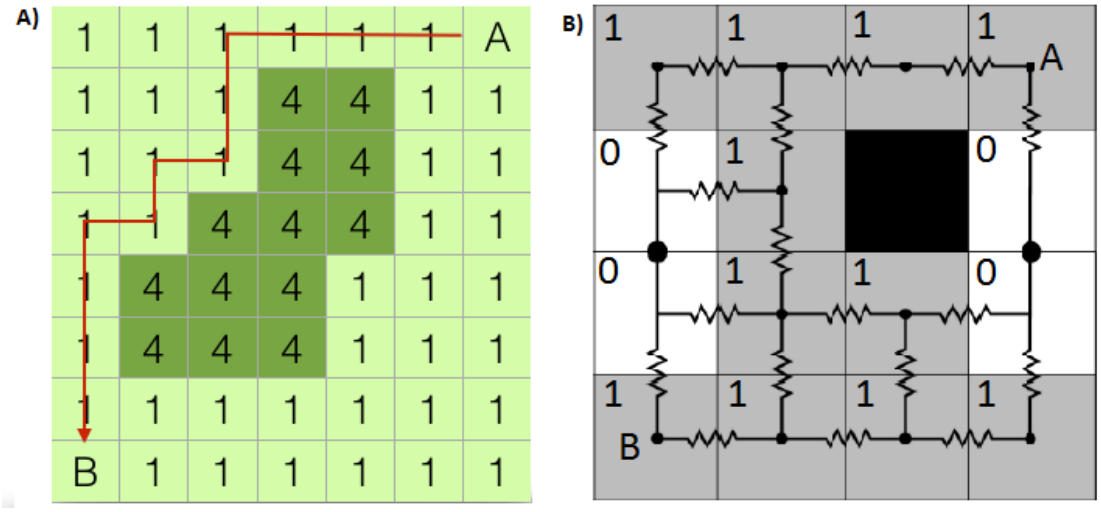


Figure 3. Representation of two resistance surfaces based in different Graph based methods. A) LCP – Least-cost pathway is the route between the points A and B with the shortest overall distance. B) In Electrical Circuit Theory, landscape is represented as an electrical circuit. The estimated distance between the points A and B is based upon the length and relative resistance of all potential paths across landscape. Adapted from Dyer (2016) and McRae (2006).

1.3. Guinea baboon as study system

Baboons (*Papio* spp.) belong to the family Cercopithecidae, also recognized as Old World Monkeys, and are a diverse and ecologically flexible group of primates. They are widely distributed in sub-Saharan Africa and are also found in South-western Arabian Peninsula (Zinner et al., 2011; 2013). This includes a large variety of habitats such as deserts, savannahs, rainforests and high-altitude mountains (Henzi and Barret, 2003). Despite extensive studies in their ecology, morphology and behaviour, the taxonomy status of the genus *Papio* is still incongruent but is generally accepted the existence of six phylogenetic species or morphotypes: Guinea baboon (*P. papio*), Olive baboon (*P. anubis*), Hamadryas baboon (*P. hamadryas*), Yellow baboon (*P. cynocephalus*), Chacma baboon (*P. ursinus*) and Kinda baboon (*P. kindae)* (Newman et al., 2004; Zinner et al., 2013) (Figure 4). Baboons shows a strong tendency for hybridization. Hybrid zones are known for Olive and northen Yellow baboon, Olive and Hamadryas baboons and Kinda and southern Yellow baboons (Zinner et al., 2011). It has been suggested that hybridization between Guinea baboon and Olive baboon also occurs at the western end of the Olive baboon range, where olive and Guinea baboons met (Zinner et al., 2009).

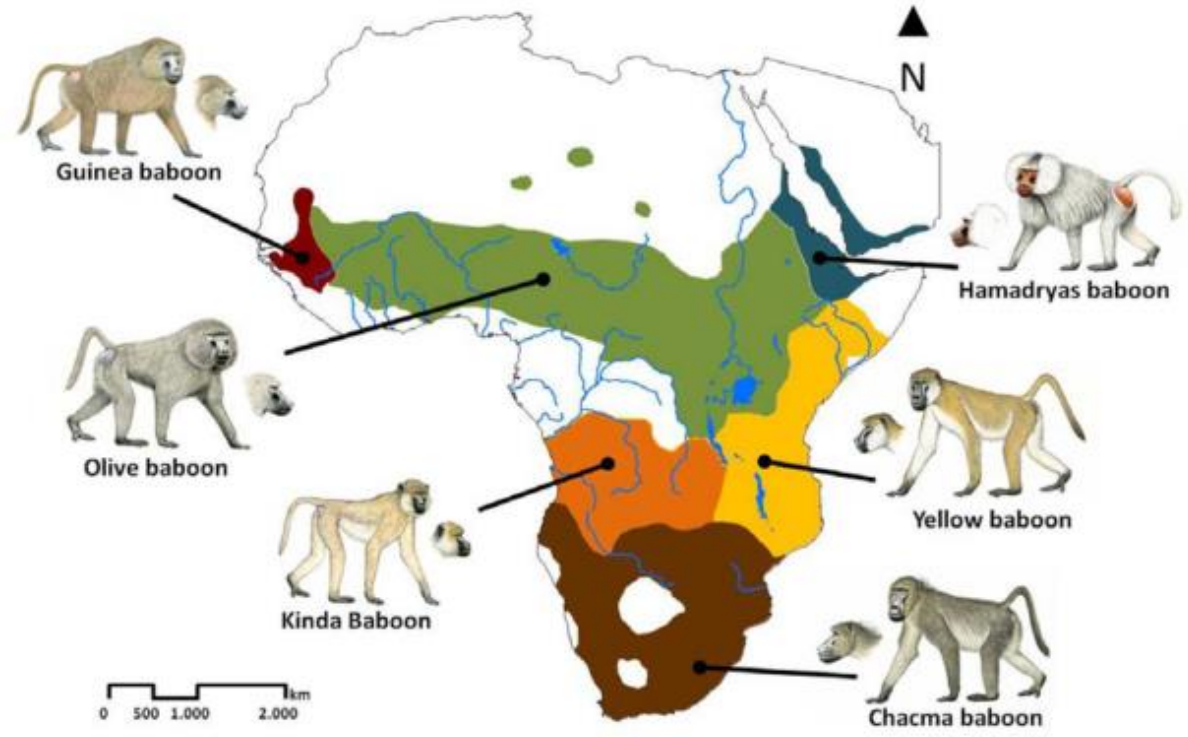


Figure 4. Geographical distribution of the six recognized baboon taxa. Adapted from Kopp (2015).

Guinea baboons (*P. papio* Desmarest, 1820) have the smallest distribution area of all baboons (ca. 250 000 Km^2). The species occur in western Africa in countries such as Mauritania, Senegal, Mali, Gambia, Guinea-Bissau, Republic of Guinea, and Sierra Leone (Figure 5). They inhabit a wide range of habitats covering very different climates, including mountains in the deserts of Mauritania, the arid Sahel woodlands, humid high forests of Guinea, and mangrove forests at sea level in Guinea-Bissau (Galat-Luong et al*.,* 2006; Ferreira da Silva et al., 2014; Kopp et al., 2015; Vale et al., 2015a). Guinea baboons occur along wide ranges of temperature (from 20 to 50 ºC) and rainfall (from less to 200 mm to 1400 mm) (Galat-Luong et al*.,* 2006). As a result of its ecological flexibility, Guinea baboons are highly opportunistic, eating a wide variety of species of plants, small mammals and birds. They can obtain food in areas when resources are scarce, the reason why they can live in areas where other primates are not able to survive (Henzi and Barret, 2003). Still, their distribution is apparently dependent on the existence of permanent water sources, especially in arid northern areas of the distribution (Vale et al., 2015a). As such, restricted dispersal ability may be expected in areas subjected to increased aridity.

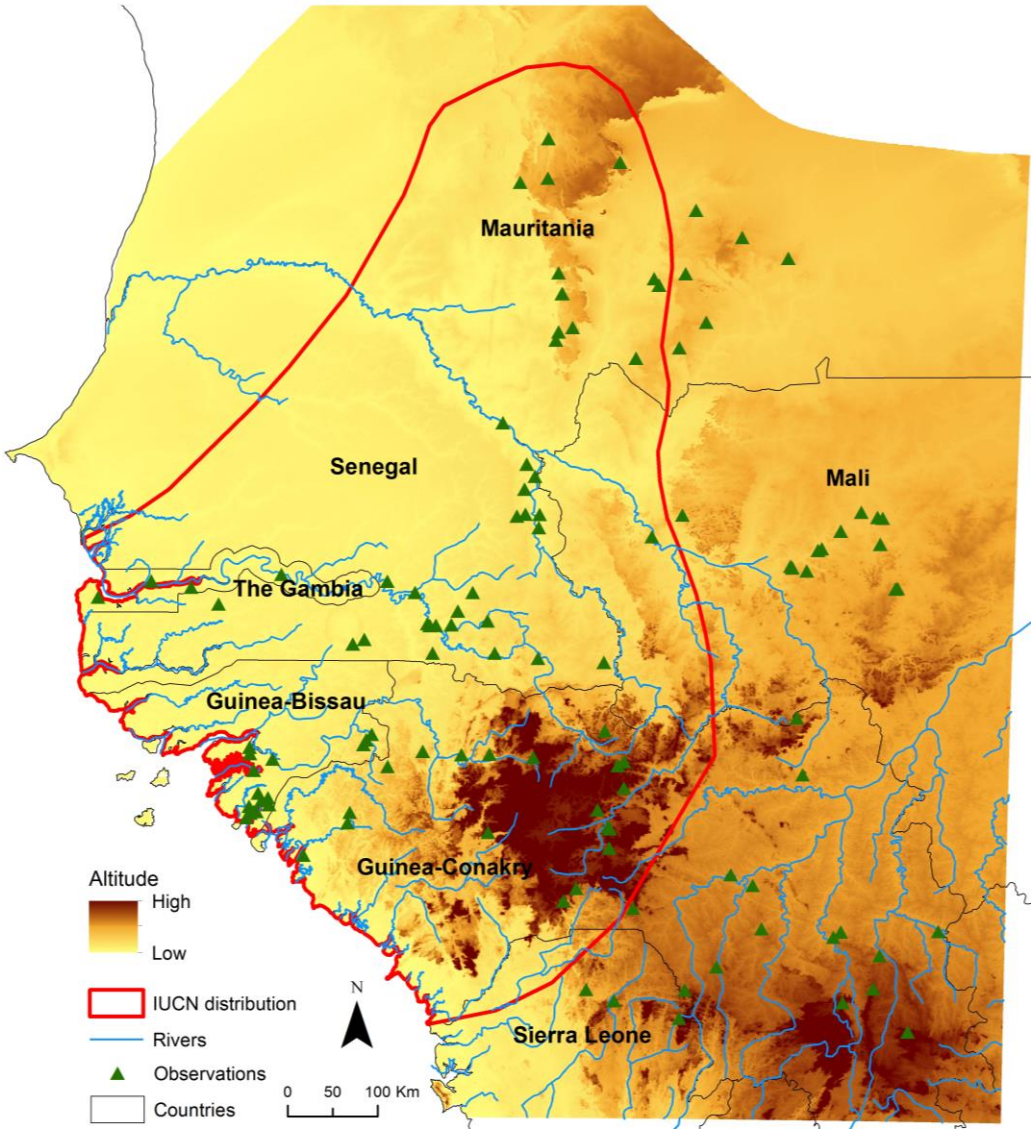


Figure 5. Distribution of Guinea baboon. Red line delimits the distribution of Guinea baboon according to IUCN (Oates et al., 2008).

Guinea baboons have a highly complex social organization, arranged in multi-layer levels (i.e. smaller social units are nested within larger ones). The basic social group is a multi-male-multi-female society but these groups aggregate into successively larger groups, which can result in communities with more than 350 individuals (Galat-Luong et al*.,* 2006; Patzelt et al., 2011; Kopp et al., 2015). Guinea baboons are able to move about 40 km per day (Galat-Luong, personal communication in Ferreira da Silva et al., 2013) and dispersal in Senegal is thought to be mostly mediated by females while the males exhibit philopatry (Kopp et al., 2015).

Guinea baboons are classified by IUCN as Near Threatened (NT) (Oates et al., 2008). Their distribution has contracted in the last 30 years and it is currently fragmented, as a result of habitat degradation, hunting, and persecution by farmers, international

trade of juveniles and meat consumption (Oates et al., 2008; Ferreira da Silva, 2012; Minhós et al., 2013a).

Given that the main threats faced by the Guinea baboon are related to human activity and human persecution, it is plausible to hypothesize that they avoid human settlements, which may affect the species dispersal pattern. Indeed, it is known that human pressure can be extremely high in specific Guinea baboon populations, which are locally considered threatened of extinction (e.g. Guinea-Bissau, Casanova and Sousa, 2007; Minhós et al., 2013a). At the same time, baboons are one of the most frequently mentioned pest species by farmers and are considered responsible for the greater portion of crop damage (Hill, 2000). This suggests that the Guinea baboon can inhabit areas surrounding croplands in order to exploit by new food resources.

1.4. Tools to access genetic connectivity in primates

1.4.1. Non-invasive DNA sampling

Many primate species are elusive, rare, occupy inaccessible areas and occur at low densities, which makes the direct study of its ecology difficult (Mittermeier et al., 2009). Also, it is estimated that 60% of the primate taxa are considered threatened with extinction (IUCN, 2015; Lynn et al., 2016). Under this context, researchers must take advantage of all the tools available to guarantee their survival. The development of non-invasive molecular methods allows assessing and monitoring wild primate populations, with minimum risks of interference for the animals (Fernando et al*.,* 2003; Lynn et al., 2016).

Non-invasive DNA sampling has been widely used in primate studies and the most common non-invasive sources of DNA includes hairs, faeces, urine, saliva, menstrual blood and male ejaculates (Lynn et al., 2016). Of these, faeces are the most widely used sources of DNA because animals defecate regularly, what means that faeces can be more easily found. Furthermore, faeces collection, storage and transport requires little technology or expense (Fernando et al*.,* 2003). DNA purification from faeces is possible because on the surface and inside of faecal pellets there are cells shed from the intestinal lining that are a useful source of host DNA (Lynn et al., 2016). However, the advantage in using faeces samples is counteracted by the challenges inherent in obtaining DNA of sufficient quality and quantity (Broquet and Petit, 2004). Faecal DNA are accompanied by a complex mixture of other compounds such as microorganisms, undigested food, digestive enzymes, mucus, bile salts and bilirubin,

which could function as chemical inhibitors and restrict the DNA extraction and amplification (Lynn et al., 2016).

Another problem associated with the use of non-invasive faecal samples is the difficulty in correctly identifying the species of origin of faeces collected in the field without directly observing the individuals. Mitochondrial DNA (mtDNA) has been widely used to identify the species of origin of unidentified faecal samples. Some of the mtDNA genes are sufficiently conserved within species but significantly differentiated across species and it study can provide a method of species DNA barcoding (Lynn et al., 2016). Particularly, the Cytochrome C Oxidase subunit I (COI) region has been successfully used to discriminate between primates' species (e.g. Minhós et al., 2013a).

1.4.2. Molecular Markers

1.4.2.1. Microsatellites *loci*

Microsatellites, also known as simple sequence length polymorphisms (SSLPs), short tandem repeats (STRs) or simple sequence repeats (SSRs), are short tandemly repeated sequences motifs consisting of repeat units of 1 to 6 base pairs (bp) in length, distributed over the entire genome (Ellegren, 2004). Changes in repeat numbers are caused by an intra-molecular mutation mechanism called DNA slippage (Schlötterer, 2000). Slippage involves DNA polymerase pausing, during which the polymerase dissociates from the DNA molecule. During the dissociation, only the terminal portion of the newly synthesized strand separates from the template, allowing the addition or subtraction of one repeat motif (Ellegren, 2004).

Microsatellite *loci* are currently the most widely used genetic markers in the study of the genetic structure of natural populations due to their high polymorphism, high variability between individuals, co-dominance and relative simplicity of assessing variation between individuals. Also, microsatellite *loci* can be used to analyse samples with low quantity and highly degraded DNA (Oosterhout et al., 2004). Microsatellites have a Mendelian heritage and are believed to be selectively neutral markers with high mutation rates. The combination of many and highly variable microsatellite *loci*, creates a unique multilocus fingerprint for each individual, allowing the study of paternity and kinship, population genetic variation and structure and gene flow between populations (Lynn et al., 2016).

Microsatellites are frequently used in studies employing non-invasive samples. However, low amounts of DNA obtained from non-invasive samples make microsatellite analysis error-prone (Taberlet et al., 1996). Allelic dropout (ADO), false alleles (FA) and

null alleles constitute the sources of microsatellite genotyping error. Allelic dropout consists in the preferential amplification of one allele in a heterozygous individual, resulting in a false homozygous individual. Replication slippage also can occur during polymerase chain reaction (PCR) of microsatellite *in vitro*, leading to the amplification of false alleles and additional stutter products that differ from the original template by multiples of the repeat unit length. When mutation occur at primers sites, some alleles cannot be amplified, resulting in null alleles (Ellegren, 2004; Broquet and Petit, 2004; Oosterhout at al., 2004). Currently, there is a large variability of methods to estimate genotyping error rates (Broquet and Petit, 2004).

1.4.2.2. Y-chromosomal markers

The use of Y-chromosomal markers has been almost restricted to evolutionary studies on humans and surveys in natural populations of other primates are still scarce (Petit et al., 2002). Recently, there was a substantial growing interest in applying Y-chromosomal markers to the study of nonhuman primates (e.g. Erler et al., 2004; Handley et al., 2006, Eriksson et al., 2006). Since the human genome has been sequenced, a large number of Y- microsatellites loci have been identified for humans. Some of these markers have been cross-amplified in non-human primates' samples (Erler et al., 2004; Handley et al., 2006). To date, few variable Y-linked markers are known for chimpanzee, orangutans, gorilla, macaque, mandrill and hamadryas baboon (e.g. Kawamoto et al., 2007, 2008; Erler et al., 2004; Gerrard and Filatov, 2005; Handley et al., 2006).

Given that a large portion of the Y chromosome does not suffer recombination and it is inherited from father to son, patterns of variation reflect the males' evolutionary history (Petit et al., 2002; Handley et al., 2006). Therefore, the study of Y-linked chromosomal markers variation can be valuable for addressing male-specific questions such as male reproductive strategies, paternity exclusion, male relatedness in multi-male groups, kin-based cooperation among males and male-specific dispersal patterns (Petit et al., 2002; Erler et al., 2004). Also, the comparison between Y-chromosome microsatellites with mitochondrial DNA sequences allows estimating the magnitude of differences in dispersal between sexes (Eiksson et al., 2006). Sex-biased dispersal (i.e. one sex shows a greater tendency to leave its natal area or to move further away) is a widespread pattern in many group-living animals, including primates (Prugnolle and Meeus, 2002; Minhós et al. 2013b). In primates, dispersal is often male-biased but some exceptions are recognized (e.g. in squirrel monkey and red colobus the dispersal is female-biased; Blair and Melnick, 2012 and Struhsaker, 2010).

The present study will be the first one to combine the recent genetic signatures of Guinea baboon (inferred by the use of microsatellites *loci*) with spatial analyses (based in ecological niche-based modelling and electrical circuit theory) to investigate its most probable areas of dispersal.

The population genetic structure and diversity of the Guinea baboon has been previously investigated by Koop (2015). However, genetic information about the South-eastern and Northern populations was lacking. In the present study, new populations of Guinea baboon will be analysed, resulting in the most complete genetic database ever used to estimate population structure and gene flow patterns across the complete range of Guinea baboon.

In many group-living primates there is a trend for dispersal to be mediated by one sex. A study led by Kopp et al. (2015), suggested that the dispersal in Guinea baboons are mediated by females but Y- chromosomal data is important to investigate male-specific population structure and unequivocally assess dispersal patterns. Given that variable Y-linked chromosomal markers are not available for Guinea baboon, a set of Y-linked microsatellite *loci* will be tested in the present study to identify variable markers.

Potential suitable areas of presence of Guinea baboon across its distribution have been mapped by Vale et al. (2015a), combining topographic, climatic, land cover, and anthropogenic variables. Given that the main threats to Guinea baboon are related to human activities, the potential impacts of landscape and human-related features in the distribution should be addressed separately. New anthropogenic variables and more observations of Guinea baboon will be used in this study to compare the effects of environmental factors and human features in its distribution.

Finally, this will be the first study using connectivity tools to investigate which isolation mechanisms (distance and resistance) are better related with the observed genetic differentiation of Guinea baboon populations.

Aims

The present study intends to investigate the effect of landscape features in connectivity between populations across a heterogeneous and fragmented landscape, using the Guinea baboon (*Papio papio)* as model system. Specifically, this study aims to:

i) Assess population structure, genetic diversity and contemporary gene flow patterns among populations;
ii) Identify the most suitable areas of presence of Guinea baboon across its distribution;
iii) Evaluate the effects of landscape features and the human-related features on gene flow between populations;
iv) Identify putative dispersal corridors.

The results of this study will bring new insights about population connectivity in West Africa landscapes and will improve our knowledge about the potential ecological corridors in the region, essential for the long-term protection of the Guinea baboon. The conclusions driven by this study will be useful for delineate conservation actions and action plans by local conservation agencies.

2. Materials and Methods

2.1. Study area

The present study area focuses on West Africa and ranges between 17º N (south of Mauritania) and 9º N (south of Guinea-Conakry). The study area comprises all countries covering the range of Guinea baboon, with the exception of Sierra Leone (Figure 6). From north to south, three major latitudinal biogeographic regions are recognized: Sahel (Sahelian Acacia Savanna), Savannah (West Sudanian Savanna), and Afrotropical (including Guinean forest-savannah mosaics, Guinean mangrove, Guinean montane forest and Western Guinean lowland forest) (Olson et al., 2001; Vale et al., 2015a). The largest rivers found in this area are the Senegal and The Gambia rivers.

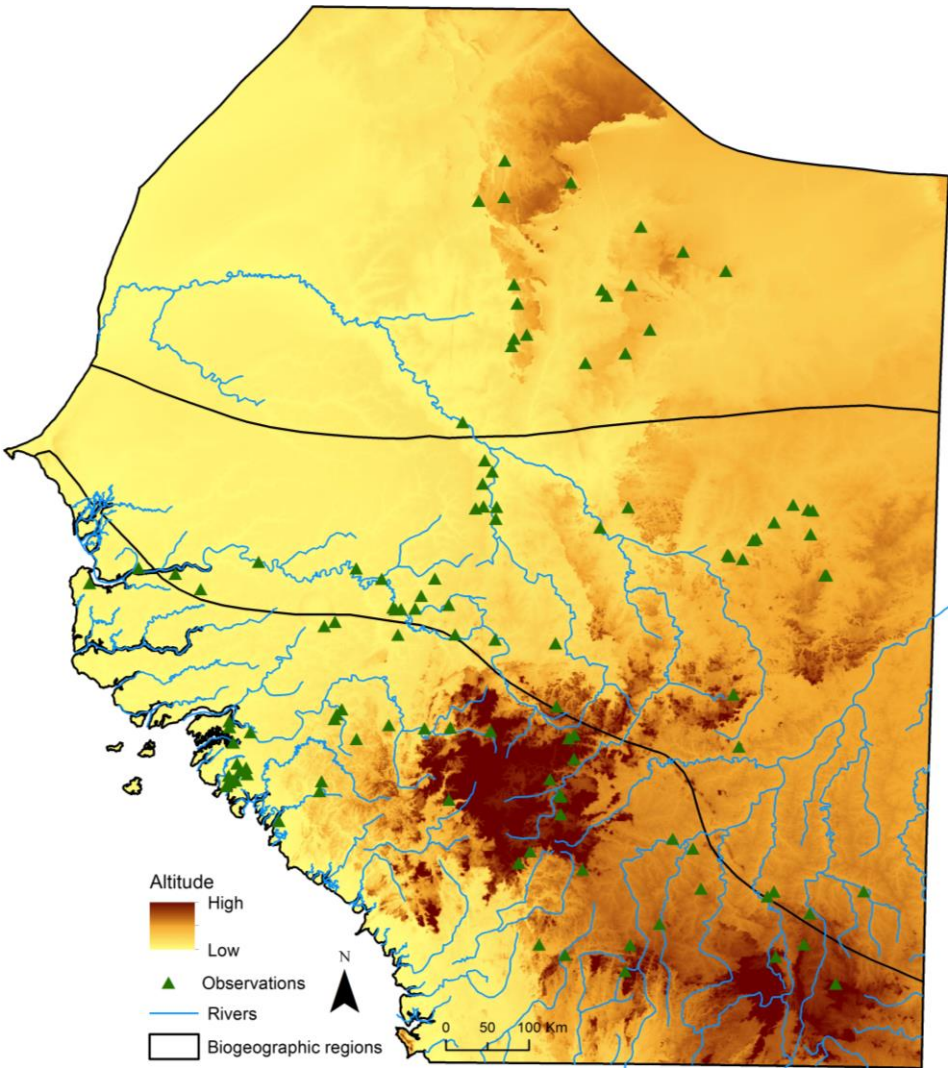


Figure 6. Study area and distribution data of Guinea baboon. Black line delimits the three biogeographic regions: Sahel, Savannah and Afrotropical; from top to bottom.

2.2. Population Genetics

2.2.1. DNA Sampling strategy and samples preservation

Faecal samples were collected non-invasively from wild Guinea baboons during field missions carried out between 2008 and 2014 by four different teams: (1) the Cognitive Ethology team (DPZ, Göttingen, Germany) that collected samples in Senegal, Guinea Conakry and Mali; (2) the BIODESERTS team (CIBIO/UP, Portugal) that collected samples in Southern Mauritania mountains; (3) MJ Ferreira da Silva (Cardiff University, United Kingdom and CIBIO/UP, Portugal) who collected samples in Guinea-Bissau; and (4) Clive Barlow, director of "Birds of The Gambia" NGO, who collected samples in The Gambia. Samples were mostly obtained from unidentified individuals, collected opportunistically in places frequently used by baboons' social groups such as foraging areas and putative sleeping sites.

Faecal samples were preserved in three different methods until DNA extraction: samples collected in The Gambia were preserved in RNAlater® (Thermo Fisher Scientific, UK), samples collected in Mauritania were preserved at room temperature and all other samples were preserved according the "Two-step" protocol (Roeder et al., 2004). The two-step protocol consists in the immediate immersion of the faeces in 99% ethanol after collection for 24 – 48 hours, followed by desiccation in silica gel (Roeder et al., 2004). All samples were stored at ambient temperature until DNA extraction.

Geographic coordinates of all samples were recorded in the field using a Global Positioning System (GPS) on the WGS84 datum.

2.2.2. Genetic data

2.2.2.1. Autosomal genotypes from previous studies

This study used unique multi-locus genotypes generated by previous studies in different laboratories: samples from Guinea-Bissau were analysed by Ferreira da Silva et al. (2014) at Cardiff University; samples from Mauritania were analysed by V. Calisto (unpublished) at CIBIO/UP, Portugal; and samples from Senegal, Guinea - Conakry and Mali were analysed by Patzelt et al. (2014), Kopp et al. (2015) and Kopp (2015) at DPZ, Göttingen, Germany.

Total genomic DNA was extracted using QIAamp® DNA Stool Mini Kit (QIAGEN®) with some modifications from the manufacturer's protocol to maximize the amount of DNA extracted. DNA extraction protocols did not vary between studies. Details of DNA extraction methods are found in Ferreira da Silva et al. (2014), Patzelt et al. (2014), Kopp et al. (2015) and Koop (2015). DNA samples were amplified for a fragment

of the hypervariable region I (HVRI) of the mitochondrial genome (D-loop) comprising 341 bp (see Kopp, 2015 for details).

Samples included in Ferreira da Silva et al. (2014) and V. Calisto (unpublished) datasets were genotyped for a maximum of 14 autosomal microsatellite *loci* using four multiplex reactions and one singlepex. Alleles range size varying between 125 and 250 bp (see Table S1, supplementary material). Samples included in Patzelt et al. (2014), Kopp et al. (2014) and Kopp (2015) datasets were genotyped for a maximum of 23 autosomal microsatellite markers (13 *loci* in common with Ferreira da Silva et al., 2014 and V. Calisto datasets) using five multiplex reactions. Allele range size varied between 94 and 254 bp (see Table 1). Most microsatellite *loci* were tetranucleotide repeats, with few dinucleotide repeat *loci*. All microsatellites were human-derived with cross-amplification in genus *Papio*. DNA amplification protocols were similar between studies, with minor differences. For details see Ferreira da Silva et al. (2014), Patzelt et al. (2014), and Kopp et al. (2014); Kopp (2015).

To assure the comparability of results between datasets generated in different laboratories, reference samples were used in all amplifications protocols to control for the effect on allele size due to different fluorescent labels and sequencing machines. Also, the allele scoring method was standardized between the laboratories and researchers to avoid errors.

The four genetic datasets were compiled in a single database containing unique Guinea baboon genotypes. This database also included one sample obtained from the Abidjan zoo (originating from Côte d'Ivoire) described morphologically as an Olive baboon but that share the same mitochondrial haplotype than Guinea baboon samples (Zinner et al., 2011; Kopp, 2015).

2.2.2.2. Autosomal genotypes generated in this study

In the present study, 97 non-invasive faecal samples were analysed in order to add new information to the database aforementioned. Of these, 25 samples were collected in The Gambia and 72 samples were collected in new sampling sites in Guinea-Conakry. Given that the exact distribution limits of the Guinea baboon are insufficiently known (Oates et al., 2008), 54 samples putatively identified as belonging to Guinea baboon and 18 samples putatively identified as belonging to Olive baboon *(P. anubis)* were analysed.

2.2.2.2.1. DNA extraction

Genomic DNA was extracted using two different protocols, according to the preservation method used to conserve the samples in the field.

The 25 faecal samples collected in Gambia preserved in RNAlater® (Thermo Fisher Scientific, UK) were extracted following the protocol of Vigilant (2011). This protocol uses 800 µl of RNAlater solution as the starting material of extraction and not the exterior part of the faecal pellet (as most faecal DNA extraction protocols) because the epithelial cells from the host are suspended in the RNAlater solution. The extraction protocol started by a centrifugation step, which intended to capture the epithelial cells in suspension in the RNAlater solution. Phosphate-buffered saline (PBS) solution was used to re-suspend the epithelial cells and wash away the remaining RNAlater solution. In the following steps, a commercial QIAGEN® lysis buffer was used to promote the cell membrane disruption, an InhibitEx tablet (QIAGEN®) was added to each sample to decrease the quantity of DNA inhibitors and Proteinase K was used to digest the animal's proteins. Finally, the resulting DNA pellet was washed in several centrifugations steps using the commercial QIAGEN® solution. The purified DNA was obtained by elution using a QIAGEN® elution buffer.

The 72 samples collected in Guinea-Conakry were preserved according the "Two-step-protocol" (Roeder et al., 2004) and were extracted using the First DNA protocol (*Gen-IAL*) used by the Cognitive Ethology team (DPZ, Göttingen, Germany). Initially, approximately 100 – 200 mg of faecal material were scraped from the pellet surface and crushed into smaller pieces. A commercial Lyse Buffer (*Gen-IAL*), Proteinase K and DTT were used to promote the cell membrane disruption and digest the animal's proteins. Thereafter, DNA was isolated from the remaining cellular constituents using chloroform and isopropanol solutions, which promote the precipitation of polysaccharides and the denaturation of proteins. In the following steps, the isolated DNA was washed several times with ethanol, in order to remove potential contaminants. The purified DNA was obtained by elution with distilled water. DNA extracts were stored at -20 ºC.

Several precautions were taken to avoid contamination during DNA extraction. All material was subjected to the UV irradiation during 30 minutes and the inside of hood was cleaned with bleach (30% dilution). Only sterile filter tips were used and a negative control was included in all procedures to test for the possible contamination with human DNA or cross-contamination between samples.

2.2.2.2.2. Microsatellites amplification

Samples were genotyped using a set of 20 autossomic microsatellites initially developed for humans and previously tested and optimized for Guinea baboon samples (Ferreira da Silva et al., 2014; Kopp et al., 2014; Kopp, 2015). Microsatellites were amplified in four multiplexes reactions (multiplex 1, 3, 4 and 5), containing between four to six different primer pairs according to the annealing temperature (AT). Table 1 gives details about the multiplexes such as the *loci* included in each multiplex and the respective annealing temperature, the repeat motif of each *loci* and the allele range size.

Table 1. Details of the Multiplexes PCRs: annealing temperature (AT, ºC), primers sequences, repeat motif, allele range size for Guinea baboon, fluorescent dye and final PCR concentration.

Multiplex	AT (ºC)	Locus	Forward (5' - 3') Reverse (5' - 3')	Repeat motif	Range	Dye	Conc (µM)
M1	57°C	D6s264	GCTGACTTTATGCTGTTCCTCCC TGAAAATTGTTCTATTCTGTGATGCC	CA	94 - 104	Fam	0.07
		D7s503	ATGACTTGGAGTAATGGG AACCTTTAATCAGGATACAGAC	CA	150 - 170	Tamra	0.15
		D12s375	TTGTTGAGGGTCTTTCTCCA TCTTCTTATTTGGAAAAGTAACCC	GATA	161 - 209	Fam	0.09
		D3s1766	ACCACATGAGCCAATTCTGT ACCCAATTATGGTGTTGTTACC	ATCT	194 - 222	Cy3	0.05
		D13s765	TGTAACTTACTTCAAATGGCTCA TTGAAACTTACAGACAGCTTG	GATA	189 - 213	Tet	0.07
M2	50°C	D14s306	TGTAACTTACTTCAAATGGCTCA TTGAAACTTACAGACAGCTTG	GATA	161 - 185	Cy3	0.08
		D1s533	CATCCCCCCCAAAAAATA TA TTGCTAATCAAATAACAATGGG	GATA	187 - 203	Fam	0.05
		D2s1326	AGACAGTCAAGAATAACTGCCC CTGTGGCTCAAAAGCTGAAT	CTAT	239-271	Tet	0.08
M3	59°C	D10s611	CATACAGGAAACTGTGTAGTGC CTGTATTTATGTGTGTGGATGG	GATA	129 - 165	Cy3	0.1
		D8s1106	TTGTTTACCCCTGCATCACT TTCTCAGAATTGCTCATAGTGC	GATA	144 - 172	Tet	0.1
		D17s791	AAAGCTACATCCAAATTAGGTAGG TGACAAAGAAACTAAAATGTCCC	GATA	160 - 170	Tamra	0.3
		D6s501	GCTGGAAACTGATAAGGGCT GCCACCCTGGCTAAGTTACT	CTAT	172 - 200	Tet	0.2
		D6s311	ATGTCCTCATTGGTGTTGTG GATTCAGAGCCCAGGAAGAT	CA	226 - 240	Tamra	0.3
M4	57°C	D5s1457	TAGGTTCTGGGCATGTCTGT TGCTTGGCACACTTCAGG	GATA	117 - 133	Fam	0.1
		D8s505	CAAAAGTGAACCCAAACCTA AGTGCTAAGTCCCAGACCAA	CA	141 - 151	Fam	0.1
		D10s143	CAGTGGACACTAAACACAATCC TAGATTATCTAAATGGTGGATTTCC	CTAT	159 - 175	Tamra	0.3
		D5s820	ATTGCATGGCAACTCTTCTC GTTCTTCAGGGAAACAGAACC	GATA	178 - 198	Fam	0.3
		D3s1768	GGTTGCTGCCAAAGATTAGA CACTGTGATTTGCTGTTGGA	GATA	193 - 221	Tet	0.05
		D7s2204	TCATGACAAAACAGAAATTAAGTG AGTAAATGGAATTGCTTGTTACC	AGAT	216 - 252	Fam	0.4
M5	58°C	D1s207	CACTTCTCCTTGAATCGCTT GCAAGTCCTGTTCCAAGTCT	CA	129 - 139	Tet	0.1

D4s243	TCAGTCTCTCTTTCTCCTTGCA TAGGAGCCTGTGGTCCTGTT	GATA	147 - 175	Fam	0.15
D1s548	GAACTCATTGGCAAAAGGAA GCCTCTTTGTTGCAGTGATT	CTAT	188 - 216	Tet	0.1
D21s114	CTCCTCCCCACTGCAGAC TCTCCAGAATCACATGAGCC	GATA	214 - 254	Fam	0.6

PCR was performed in a total volume of 10 µl, containing DNA template, 5 µl of QIAGEN Multiplex PCR Master Mix, 1 µl of Primer mix - containing the forward (F) and reverse (R) primers of all *loci* included in the multiplex PCR reaction - and RNase-free water. All PCR were run using the following program: initial denaturation step for 15 min at 95°C, followed by 40 cycles of 30 s at 94°C, 40 s at locus-specific annealing temperature (57 °C for Multiplex 1 and 4, 58 °C for Multiplex 5 and 59 °C for Multiplex 3) and 40 s at 72 °C and a final extension step for 30 min at 72 °C. PCR products were checked on 2% TAE agarose gels stained with ethidium bromide. The PCR products that successfully amplified were combined with formamide and a size standard (GeneScan 400HD ROX, Thermo Fisher Scientific) and analyzed on a 3130xl ABI automated sequencer.

In order to avoid contamination, all PCRs were performed in a dedicated room and all material used in PCRs was subject to UV decontamination. Sterile filter tips were used in all steps and a negative control was included in all reactions.

Reference samples extracted during previous studies were amplified by the multiplexes using the same conditions to standardize the allele's size between the different amplifications protocols. These included eight samples from Mauritania, six samples from Guinea-Bissau (five samples extracted from faeces and one tissue sample collected in the Guinea-Bissau bushmeat market; Minhós et al. (2013a)).

2.2.2.2.3. Genotyping procedures and quality control of genotypes

The difficulties of obtaining reliable genotypes in non-invasive samples were overcome by using a multiple-tube approach, which consists in amplifying the same DNA sample multiple times from each *locus* and comparing the different replicates to reach a consensus genotype (Taberlet et al., 1996).

The number of replicates necessary to obtain 95% confidence in the genotypes was estimated taking into account the ADO and false allele rate FA of the genetic data obtained in this study. ADO and FA were estimated using a maximum likelihood method implemented in PEDANT V1.0 (Johnson and Haydon, 2007), using two amplifications

repetitions per *loci*. GEMINI version 1.4.1 (Valière, 2002) was used to determine the minimum number of replicates that confirms an allele per *locus*. Consensus Threshold was estimated based on the ADO and FA rate previously calculated (Table S2, supplementary material). A total of 100 simulations were conducted for values between 2 and 10 replicates, resulting in a plot relating the accuracy of genotypes with the increasing number of PCRs replicates. An asymptote was reached at four PCR replicates suggesting that increasing the number of replicates would not significantly increase the reliability of genotypes. Consequently, four amplifications per *loci* were attempted for each sample. GEMINI was also used to set the genotyping rules for each *locus* (following procedures implemented in Ferreira da Silva et al., 2014).

Given the potential low quality and quantity of DNA retrieved from faeces, all DNA samples were initially amplified four times for Multiplex 1. This multiplex was chosen among the other multiplexes because the *loci* included have a higher amplification rate (Ferreira da Silva, 2012). Each sample was classified according the reliability of its genotype using a Quality Index (QI) proposed by Miquel et al. (2006). This index ranges from 0 to 1, where values below 0.50 might indicate low reliability in genotypes (Miquel et al., 2006). Following these results, all samples with a QI lower than 0.50 were excluded from the analyses and DNA samples with QI > 0.5 were amplified for the remaining multiplexes (Multiplex 3, 4 and 5). After the amplification of all *loci*, the final QI was re-calculated and the samples with values lower that 0.50 were discarded.

Samples profiles were scored following a semi-automated procedure (automated allele calling followed by visual examination) using GeneMapper v4.1 (Applied Biosystems). To reduce errors related with the inclusion of non-Guinea baboon alleles, two different panels were used: one created for the samples collected in the field as belonging to Guinea baboon and another created for samples putatively identified as Olive baboon.

Genotypes were tested for typing errors and genotyping errors using Microsatellite toolkit (Park, 2001) and Micro-checker v2.2.3 (Van Oosterhout et al., 2004), respectively. Microsatellite toolkit was also used to identify identical genotypes from the dataset. Micro-checker v2.2.3 (Van Oosterhout et al., 2004) was used to detect scoring errors attributable to stuttering and large-allele dropout (Van Oosterhout et al., 2004).

To ensure that the set of *loci* could reliably identify unique genotypes, the probability of identity (pID, Paetkau and Strobeck, 1994) and the probability of identity assuming siblings (pIDsib, Waits et al., 2001) were calculated using GenAlEx 6.5 (Peakall and Smouse, 2012). pID is the overall probability that two individuals drawn at

random from a given population share identical genotypes at all typed loci and the pIDsib is the probability that a pair of siblings share the same genotype.

2.2.2.2.4. Molecular determination of *Papio* samples

DNA samples from Guinea-Conakry included in the final dataset were identified to the species-level using a DNA barcoding procedure (following Minhós et al., 2013a). DNA extracts were amplified for approximately 700 bp fragment of the Cytochrome C Oxidase subunit I (COI), using the OWMCOI primers (Lorenz et al., 2005). PCR was performed in a total volume of 10 µl, containing DNA template, 5 µl of QIAGEN Multiplex PCR Master Mix, 0.5 µl of Forward (5' (A/G)CT (G/C)TT TTC AAC AAA (C/T)CA (C/T)AA AGA C 3') and Reverse (5' GTA (A/G)AC TTC (G/C)GG GTG (A/G)CC (A/G)AAG AA TC 3') primers and RNAse-free water. A touchdown-PCR was carried out according to the following cycling program: initial denaturation step for 15 min at 95°C, followed by 19 cycles of 30 s at 94°C, 90 s at 59 ºC and 90 s at 72°C, with decreasing annealing temperatures in decrements of 0.5 ºC per cycle, and by 21 cycles of 30 s at 94°C, 90 s at 50 ºC and 90 s at 72°C and a final extension step for 10 min at 72°C. PCR products were purified using ExoSAP-IT purification Kit (Affymetrix). Cycle sequencing was performed using the BigDye Terminator Kit (Thermo Fisher Scientific), following the manufacturers protocol. Sequencing of PCR products was performed in a 3130xl ABI automated sequencer (Applied Biosystems) and sequences were visually checked and manually edited using Geneious v4.8.5 (Biomatters).

Sequences were trimmed to 600 bp size approximately. A Basic Local Alignment Search Tool (BLASTn, Altschul et al., 1990) was performed with the trimmed sequences in National Center for Biotechnology Information database (NCBI, available in http://www.ncbi.nlm.nih.gov/) to search for the most similar sequences ("vouchers"). The sequences with higher similarities with the species vouchers (identity above? 97%) were considered as identified to the same species as the vouchers.

Given that sampling was conducted at the expected contact zone between the Guinea baboon and olive baboon (*P. anubis*), two different approaches were used to investigate the genetic distance between the Guinea and Olive baboon and the existence of potential hybrids between the two species. This is an important step because the inclusion of hybrids in the dataset could create an artificial genetic discontinuity that could be interpreted as a barrier in the connectivity analyses.

A Neighbour Joining (NJ) tree was constructed to infer the genetic distance between the Guinea baboon individuals and the individuals identified as Olive baboon

by mtDNA barcoding. The sample from the Abidjan zoo described morphologically as an Olive baboon but identified as Guinean baboon by the mtDNA barcoding (Zinner et al., 2011; Kopp, 2015) was also included in the NJ tree. The tree was constructed based on the shared alleles distance (DAS, Jin and Chakraborty, 1993) between all individuals, using the Populations v1.2.32 software (Langella, 1999).

STRUCTURE v2.3.4 (Pritchard et al., 2000) was used to identify genetic clusters assigned to different species. The Guinea baboon, Olive baboon and the Abidjan zoo genotypes were included in this analysis. Five independent simulations with 1,000,000 Markov Chain Monte Carlo (MCMC) steps following by a burn-in of 1,000,000 interactions were run to evaluate the existence of 10 genetic clusters ($1 \leq K \geq 10$). In all runs it was assumed the admixture model and the correlated frequency as allele frequency model (Falush et al., 2003). The most probable number of K was determined by estimating the highest log-likelihood [Ln P(X/K)] and by using the statistic ΔK developed by Evanno et al., (2005). STRUCTURE HARVESTER v6.8 (Earl and von Holdt, 2011; available at http://taylor0.biology.ucla.edu/struct_harvest/) was used to process STRUCTURE results.

2.2.2.3. Developing a variable Y-linked microsatellite marker

2.2.2.3.1. Bibliographic review

Initially, a bibliographic review of Y-linked microsatellite variable markers initially developed in humans and tested in non-human primates' species was conducted. A total of 72 potential polymorphic Y-linked microsatellites *loci* were found. Of these, 62 were tested in great apes species – *Gorilla sp., Pongo sp.* and *Pan sp.* (Erler et al., 2004) – and 20 were tested in Cercopithecidae species – *Mandrillus sphinx* (Erler et al., 2004), hamadryas baboon, *Papio hamadryas* (Handley et al., 2006), *Macaca fuscata* (Kawamoto et al., 2007), *Macaca fascicularis* (Kawamoto et al., 2008) and green monkey *Chlorocebus* spp. (Haus et al., 2013).

The set of Y-linked microsatellites *loci* to be tested in Guinea baboon samples were selected according to the following criteria: (1) existence of polymorphism in other primates' species, (2) phylogenetic close proximity between the species in which the genetic marker was tested and the Guinea baboon, (3) a relatively small allele size (varying between 93 and 500 bp) and (4) similar annealing temperatures.

2.2.2.3.2. Samples tested

DNA from 36 non-invasive Guinea baboon' samples previously identified as males were used. The DNA samples were gently provided by the Cognitive Ethology team (DPZ, Göttingen, Germany), MJ Ferreira da Silva (Cardiff University, United Kingdom) and V. Calisto (CIBIO/UP, Portugal). The sex of the individuals was identified using a molecular protocol designed by C. Roos (details are found in Ferreira da Silva et al. 2014 and Kopp et al. 2014). Two additional DNA tissue samples (one female and one male) collected at bushmeat markets in Bissau were also included (Minhós et al., 2013a). The inclusion of a female tissue sample in the analyses would test for male-specificity of the markers and the male tissue samples was used as positive control in all PCR optimizations.

2.2.2.3.3. Y-linked microsatellite *loci* screening

A panel of 15 Y-linked microsatellite *loci* previously used in primates' species were tested in samples of Guinea baboons. The *loci* tested were DYS391, DYS472, DYS557, DYS576, DYS645, DYS579 and DYS569, which were previously used by Handley et al. (2006) in *Papio hamadryas* and DYS470, DYS474, DYS511, DYS571, DYS574, DYS594, DYS632 and DYS643, which were previously used by Erler et al. (2004) in *Mandrillus sphinx* (Table 2).

PCRs were optimized for Guinea baboon samples. At least three different annealing temperatures were tested for each *locus* given that these markers were used in *P. papio* for the first time (see Table 2 for details). PCR was performed in a total volume of 10 µl, containing DNA template, 5 µl of QIAGEN Multiplex PCR Master Mix, 0.4 µl of each primer and RNase-free water. All PCR were run using the following program: initial denaturation step for 15 min at 94°C, followed by 40 cycles of 30 s at 94°C, 40 s at locus-specific annealing temperature and 40 s at 72°C and a final extension step for 30 min at 60°C. PCR products were checked on 2% TBE agarose gels stained with GelRed (*Biotium*). PCR products were analysed on a 3130xl ABI automated sequencer (Applied Biosystems) and scored manually using GeneMapper v4.1 (Applied Biosystems).

Initially, all *loci* were tested to check if only samples previously identified as males produced a positive result (male-specific amplification, using one female and two males samples).

The *loci* that produced a PCR product using a female sample or non-specific amplifications (i.e. multiples bands outside the size range) were discarded. In a second step, male-specific *loci* were amplified using six male samples from distant populations

(three males from Mauritania and three from Guinea-Bissau, which are approximately 700 km apart) to investigate variation in the allele's size. The *loci* that showed variation between Guinea-Bissau and Mauritania were screened for a total of 36 Guinea baboon samples representative of the entire distribution of Guinea baboon (Mauritania, Senegal, Guinea-Bissau, Guinea – Conakry, Mali and Sierra Leone).

PCR reactions for variable *loci* were repeated twice to assure reliability of the genotype.

Table 2. Details of the 15 Y-linked microsatellite *loci* tested in Guinea baboon: *locus*, primate species where it was tested, expected alleles size, primers sequences and reference of each study.

Locus	Species	Size (bp)	Forward (5' - 3') Reverse (5' - 3')	Reference
DYS391	P.hamadryas	246	CTATTCATTCAATCATACACCCA GATTCTTTGTGGTGGGTCTG	Handley et al (2006)
DYS557	P.hamadryas	500	TTTTCTGTGCCAAGCCTACA TCTAATGCACCTTGAGGGATG	Handley et al (2006)
DYS472	P.hamadryas	280	AGATTGTCCCACCTGCACTC GAGGCACTGTGTTCAGCAAA	Handley et al (2006)
DYS576	P.hamadryas	292-304	TTGGGCTGAGGAGTTC GGCAGTCTCATTTCCTGGAG	Handley et al (2006)
DYS645	P.hamadryas	320	AGCCACCTGGGTATATGAGG TGTTGCAGCTTTTCCTTCTG	Handley et al (2006)
DYS579	P.hamadryas	200	GCCAGCAGTAGACCCAGACT AGGCAGAGGTTGCAGTGAGT	Handley et al (2006)
DYS569	P.hamadryas	261	TCCATGGGATATGATGAGCA GGCAGCCTGTAGGACAGAGA	Handley et al (2006)
DYS511	Mandrillus	272	GATAGGATGGGGTGGATGTG TGTGAATTCCCCTTCTACATCTC	Erler et al (2004)
DYS574	Mandrillus	200	GGTGGGGCTTCCATATTTTT AATGTAGACGACGGGTTGATG	Erler et al (2004)
DYS594	Mandrillus	235	GATGTGCCTAATGCCACAGA CCCTGGTGTTAATCGTGTCC	Erler et al (2004)
DYS643	Mandrillus	93	AAGCCATGCCTGGTTAAACT TGTAACCAAACACCACCCATT	Erler et al (2004)
DYS470	Mandrillus	203	GGTCCTTCAGGAACCAGTTG TGGCTGTAAAACAAATATCAGCA	Erler et al (2004)
DYS474	Mandrillus	101	CCCCTGAACTTAAAAGGTGGA GGCATCTAGGTTTACTGTGAGGA	Erler et al (2004)
DYS571	Mandrillus	154	AGCCTTCAGCGACTGCTTTA AGCTGAGATCATCCCATTGC	Erler et al (2004)
DYS632	Mandrillus	140	GGCCGTTGCAAAATAAACTG TCTGGGCAACAGAAGGAGAC	Erler et al (2004)

The alleles found in polymorphic markers were sequenced to investigate the motif and pattern of variation. One allele of each size was selected and the PCR products were purified using ExoSAP-IT purification Kit (Affymetrix). Cycle sequencing was performed using the BigDye Terminator Kit (Thermo Fisher Scientific). Sequencing was performed in a 3130xl ABI automated sequencer (Applied Biosystems) and the obtained sequences were visually checked and manually aligned using Geneious v4.8.5 (Biomatters).

2.2.3. Genetic diversity of autosomal microsatellite *loci*

Genotyped individuals included in the final dataset were grouped into 10 putative geographic populations according to the type of habitat and natural features in the landscape that could constitute barriers to dispersal (e.g. rivers, mountains, arid areas) (Figure 7).

Deviations from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) across all pairs of *loci* were estimated for the overall dataset using GenAlEx 6.5 (Peakall and Smouse, 2012) and GENEPOP v4.0 (Rousset, 2008), respectively. Bonferroni correction for multiple comparisons was used to adjust significance levels (Rice 1989), considering a 95% confidence interval.

Genetic diversity levels among populations were investigated using the number of different alleles per locus (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), Inbreeding coefficient (F_{IS}), percentage of *loci* polymorphic (%P) and number of private alleles (P_A), estimated using GenAlEx 6.5 (Peakall and Smouse, 2012). Taking into account that some populations were genotyped for a lower number of *loci* (i.e. Coastal, Fouta Djalon W, Assaba and Afollé), the genetic diversity for these populations was calculated for 13 *loci*. To assure that genetic diversity does not reflect differences in number of *loci* amplified, the genetic diversity levels were also estimated considering the number of *loci* genotyped in common for all populations (13 *loci*). A Neighbour Joining (NJ) tree, based on the distance alleles shared (DAS, Jin and Chakraborty, 1993), was constructed to investigate the genetic distance between the populations using the Software Populations v1.2.32 (Langella, 1999).

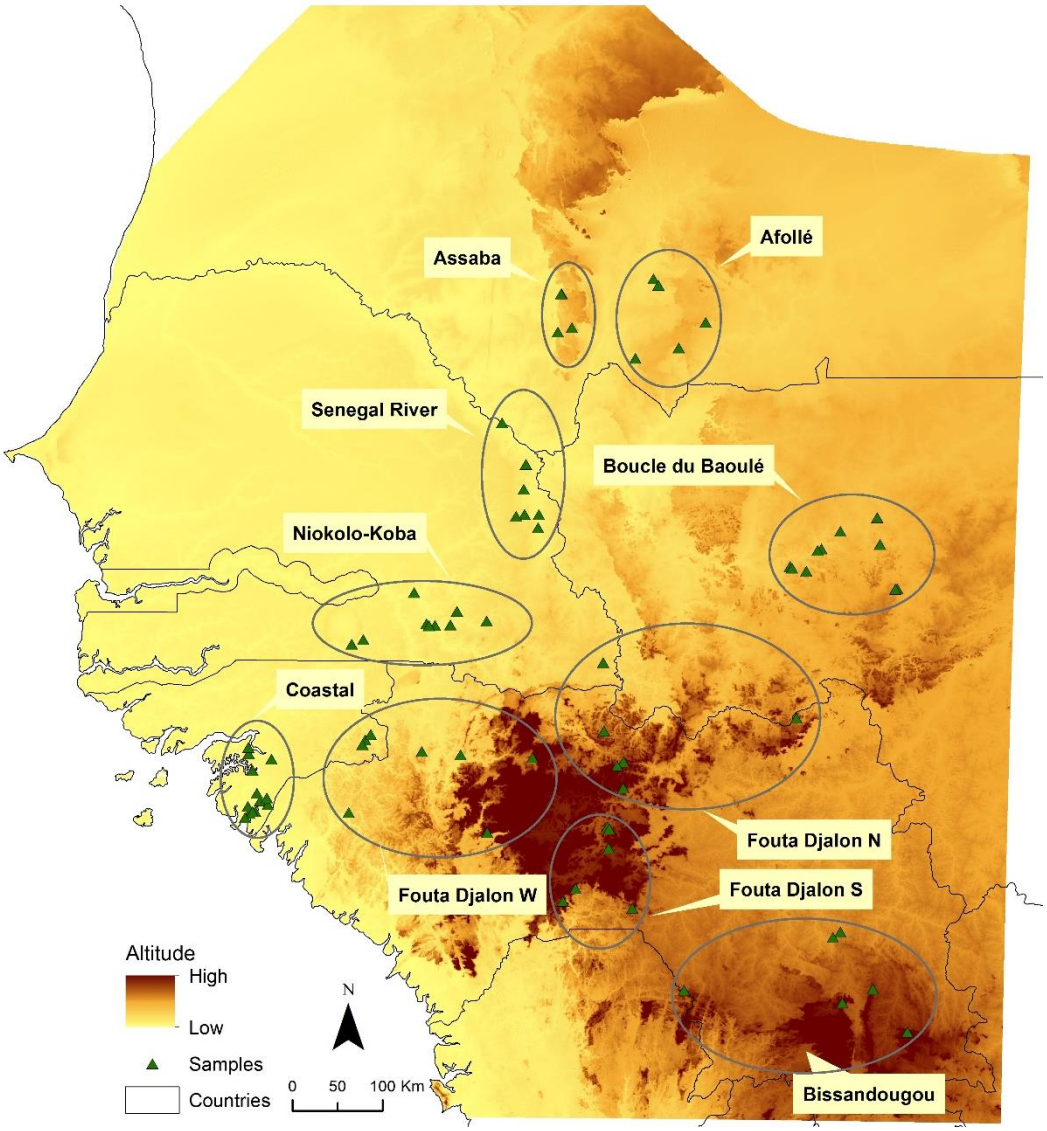


Figure 7. Spatial location of the 10 putative populations of Guinea baboon used in present study.

2.2.4. Population structure and differentiation

To investigate population genetic structure, two different individual-based Bayesian clustering methods were used: STRUCTURE v2.3.4 (Pritchard et al., 2000) and BAPS v5.2 (Corander and Marttinen, 2006).

The Bayesian clustering algorithm implemented in STRUCTURE v2.3.4 (Pritchard et al., 2000) was used to infer the most likely number of genetic clusters (K) and to assign individuals to those clusters. Preliminary runs were performed considering K values ranging from 1 to 12, but the most probable number of K was lower than K=10 in all runs. Consequently, a total of five independent simulations with 1,000,000 Markov Chain Monte Carlo (MCMC) steps following by a burn-in of 1,000,000 interactions were

run, in order to evaluate the existence of 10 genetic clusters ($1 \leq K \geq 10$). In all runs it was assumed the admixture model and the correlated frequency as allele frequency model (Falush et al., 2003). The most probable number of K was determined by estimation of the highest log-likelihood [Ln P(X/K)] and by using the statistic ΔK developed by Evanno et al. (2005), considering that the height of the modal values of ΔK indicates the strength of the population subdivision signal (Evanno et al., 2005; Bergl and Vigilant, 2007). STRUCTURE HARVESTER v6.8 (Earl and von Holdt, 2011; available at http://taylor0.biology.ucla.edu/struct_harvest/) was used to process STRUCTURE results.

The averaged individual membership coefficients (Q) across runs was ranked and plotted to identify a threshold of assignment of the individuals to each genetic cluster (Beaumont et al., 2001). Each individual was assigned to the genetic cluster for which the Q was higher than the defined threshold and the remaining individuals were considered admixed between clusters (e.g. Quéméré et al., 2010b).

To visualize genetic discontinuities and contact zones between the clusters identified by STRUCTURE, an Inverse Distance Weighted (IDW) interpolation of the individual Q values was performed using the Interpolation function implemented in ArcGIS 10.1 (ESRI 2012).

Genetic differentiation between clusters was inferred using the Wright's F-statistics (Wright, 1951), calculated in Arlequin v3.0 (Excoffier et al., 2005). Significance of pairwise F_{ST} values was calculated using 1,000 permutation tests. Deviations from Hardy–Weinberg equilibrium and estimates of genetic diversity were calculated each inferred cluster. In order to investigate the existence of genetic substructure, STRUCTURE analyses were also repeated for each of the clusters using the same parameters as before (see Quéméré et al., 2010b).

A spatial clustering of individuals was performed in BAPS v5.2 (Corander and Marttinen, 2006), following a model-based Bayesian assignment. A total of 10 independent runs were performed to assess repeatability of results, considering 15, 10, five, two and one genetic cluster, repeated five times as priors of K.

As the majority of samples were collected in places frequently used by baboons' social groups, the presence of related individuals in the dataset is likely. Given that the presence of related individuals can lead to an overestimation of the true value of K when using Bayesian clustering methods (Pritchard and Wen, 2004), two different approaches were used to investigate the effect of relatedness in pattern of genetic structure inferred. Firstly, the function "Pops Mean" in GenAlEx 6.5 (Peakall and Smouse, 2012) was used to determine whether the observed average relatedness of individuals within populations

and clusters identified by STRUCTURE was significantly higher than the average relatedness of individuals randomly assigned to populations (i.e. the average relatedness expected by chance). Secondly, CLUST_ DIST v1.0 software (Fernández et al., 2011) was used to group individuals without assumptions on Hardy-Weinberg or linkage equilibrium. The analysis in CLUST_ DIST estimates the pairwise Nei minimum distance between individuals and separate individuals in genetically homogeneous groups by maximizing the genetic distance between a predefined number of clusters (Fernández et al., 2011). CLUST_ DIST v1.0 was run for most likely number of clusters estimated by STRUCTURE and BAPS.

2.2.5. Isolation by distance (IBD)

The hypothesis of Isolation by Distance was tested to investigate if the distribution of the genetic diversity was related to geographic distance between sampling sites. Genetic distances among all pairs of individuals were measured by the shared alleles distance, calculated using Populations v1.2.32 (Langella, 1999). Linear Euclidean distances were calculated based on the geographic coordinates of all samples in R Software (R Development Core Team 2005). The correlation between Pairwise genetic distances and Euclidean geographic distances was tested using a Mantel test (Mantel, 1967) calculated in R (R Development Core Team 2005), considering the Pearson method with 10,000 permutations.

2.3. Habitat suitability mapping

2.3.1. Presence data

The region of West Africa, comprising Sahel, Savannah and Afrotropical biogeographic regions, was used for modelling the Guinea baboon' suitable areas following Vale et al. (2015a). The study area was delimitated with a buffer of 150 km around the IUCN polygon of the Guinea baboon distribution (Oates et al., 2008; Vale et al., 2015a).

A total of 362 geo-referenced observations points were used, wherein 296 records were obtained by collaborators during field work (Torres, 2007; Brito et al., 2010; Ferreira da Silva et al., 2014; Koop et al., 2014; Koop, 2015; Dietmar, unpublished; Clive Barlow, unpublished) and 66 records were obtained from the existing literature (cited in Vale et al., 2015a).

In order to avoid spatial autocorrelation among records, 107 non-clustered observations were randomly selected from clusters of species occurrence records according to the Nearest Neighbour Index (NNI) implemented in ArcGIS 10.1 (ESRI 2012). Of these, 65 observations were used for model training and 42 observations were used for model validation (see Figure 8).

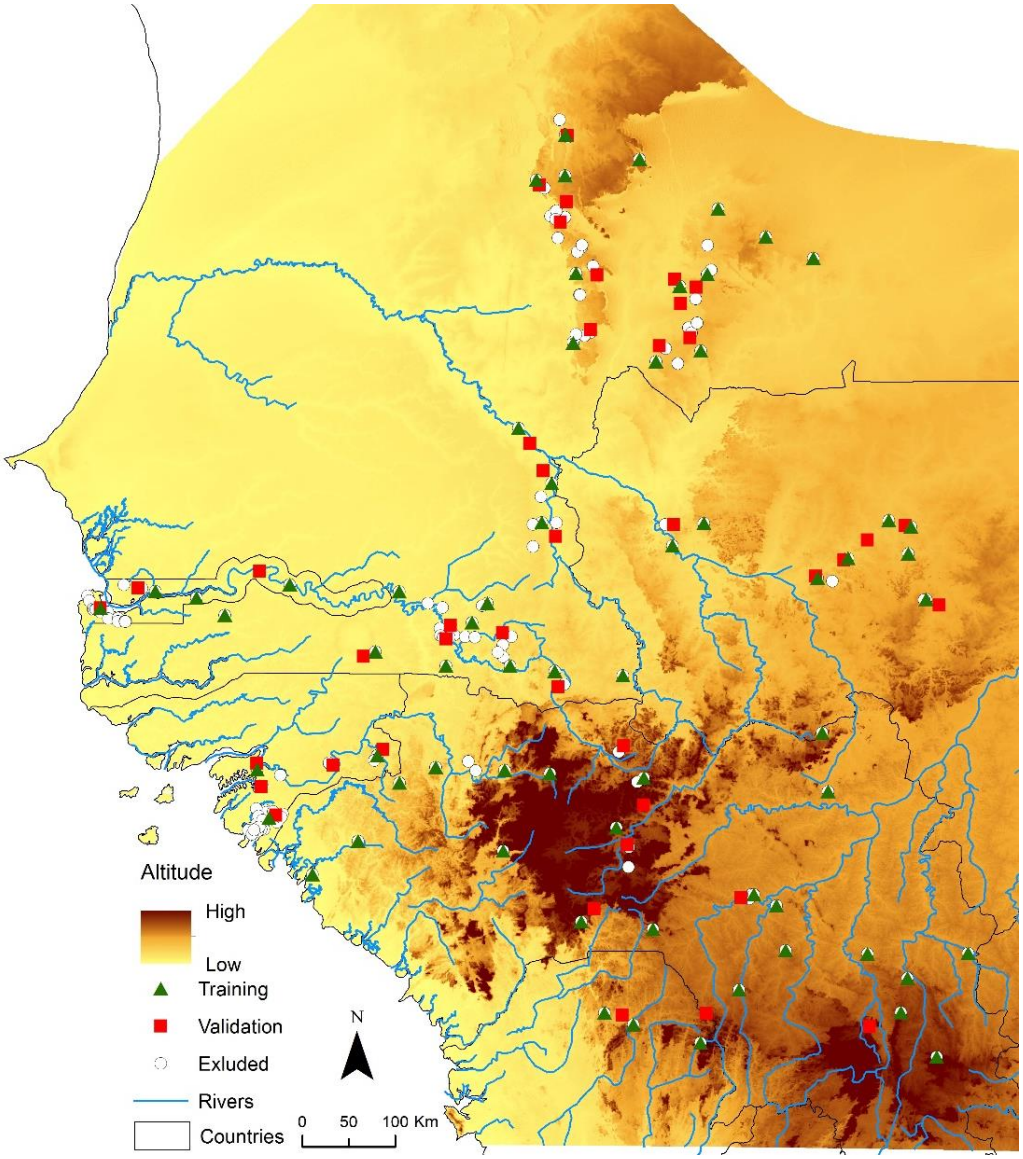


Figure 8. Study area and distribution data of *P. papio* used in the present study. Training and validation observations are represented by green and red dots, respectively. White dots represent the observations excluded from the analyses to avoid autocorrelation.

2.3.2. Ecogeographical variables (EGVs)

A set of 18 ecogeographical variables (EGVs) were selected to model the current Guinea baboon distribution at a scale of 1 km^2 resolution, including one topographic variable (slope), eight land cover variables (NDVI, Normalized Difference Vegetation Index; open broadleaved deciduous forest/woodland; closed to open shrubland; closed to open herbaceous vegetation; rocky desert; sandy desert; Gueltas; Rivers), three climatic variables (Temperature of warmest month; Precipitation of wettest month; PET, Potential Evapotranspiration) and six anthropogenic variables (toponomies; accessibility; roads; croplands; mosaic cropland/vegetation; mosaic vegetation/cropland) (Table S3, supplementary material). The EGVs were selected according to the main anthropogenic threats faced by Guinea baboon (which may affect its dispersal) and according its habitat requirements (which may facilitate its dispersal). All EGVs were projected to WGS 1984 UTM Zone 28N datum. The original categorical variables, such as land cover and anthropogenic variables were converted into continuous variables by estimating the Euclidean distance of each grid cell using ArcGIS 10.1 (ESRI 2012).

Pearson correlation index was calculated among all pairs of variables using ArcGIS 10.1 (ESRI, 2012). Given that some variables showed to be highly correlated (Pearson correlation index > 0.7; Table S4 supplementary material), an orthogonal transformation was implemented to avoid co-linearity between variables (Aguiar et al., 2016). Therefore, two Principal Component Analysis (PCA) were performed: one including the topographic, land cover and climatic variables (Figure S1, supplementary material) and a second one including only the anthropogenic related variables (Figure S2, supplementary material).

2.3.3. Modelling analysis

Two independent ecological niche-based models (ENM) were generated using Maximum entropy modelling (MAXENT, Phillips et al., 2006) to test the effect of environmental and anthropogenic features on population structure (see Figure 9). The first ENM was created using the topographic - land cover - climatic PCA layers (hereafter referred as “Environmental model”; Figure S3 and Table S5, supplementary material) while the second was created using both topographic - land cover - climatic and anthropogenic PCAs layers (hereafter referred as “Human model”; Figure S4 and Table S6, supplementary material).

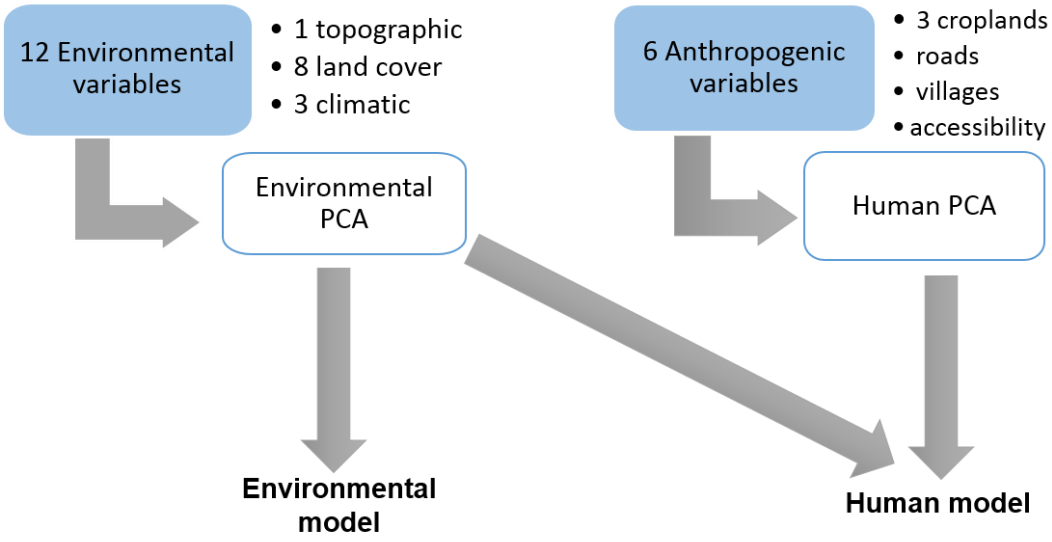


Figure 9. Framework for building the ENMs for the Guinea baboon. Details of variables can be found in Table S3 in supplementary material.

The regularization multiplier, feature type, number of replicates and percentage of data for testing were set according to the number of observations available, as suggested by Vale et al. (2016). Therefore, auto-features with 0.75 of regularization multiplier with 25 replicates and 25% testing were used in the present study. Model performance was evaluated using the Area Under the Curve (AUC). This method selects the model that produces the maximum value for the area under the characteristic curve calculated using the data used in model construction (Warren and Seifert, 2011) and ranges from 0.5 (random accuracy) to a maximum value of 1 (perfect discrimination) (Jiménez-Valverde et al., 2012).

2.4. Landscape connectivity

2.4.1. Isolation by resistance (IBR)

Isolation by resistance (IBR) hypothesis was tested to investigate permeability of environmental and anthropogenic features on gene flow among Guinea baboons' individuals. Both ENMs maps obtained using Maximum entropy model ("Environmental model" and "Human model") were used to map resistance surface. Then, pairwise resistance distances between pairs of individuals from different sampled locations (N = 74, which correspond to one individual per Km^2) were calculated using CIRCUITSCAPE v4.0.5 (McRae, 2006). It was implemented the pairwise model option with focal points, connecting eight neighbours based on the average resistance (Russo et al., 2016). Genetic distances among the same pairs of individuals from different locations (N = 74)

were measured using the shared alleles distance, calculated in the Software Populations v1.2.32 (Langella, 1999).

Five different IBR hypotheses were tested for each resistance model: standard model, model^2, model^3, standard model + standard deviation (SD) and standard model - SD, making a total of 10 different resistance surfaces. The model^2 and model^3 allows enhance of the resistance values. The standard model + SD and the standard model - SD allow to incorporate the uncertainty of the MAXENT models in results, wherein the former represents the most optimistic dispersal scenario and the later the least optimistic. Mantel tests (Mantel, 1967) were used to correlate the genetic distance matrix with each landscape resistance model. Partial Mantel tests also were used to estimate the correlation of the genetic distance with the resistance distances with controlling for Euclidean distance. All Mantel tests were performed using library VEGAN version 1.6-7 (Dixon 2003, Oksanen 2005) in R (R Development Core Team 2005), following the Pearson method with 10,000 permutations.

3. Results

3.1. Population Genetics

3.1.1. DNA extraction success and genotyping

Out of the 97 faecal samples extracted in this study and amplified for Multiplex PCR 1, 33 samples collected in Guinea-Conakry showed a mean QI across *loci* above 0.50 and were amplified for the multiplexes 3, 4 and 5. The remaining samples (i.e. 39 from Guinea-Conakry and 25 faecal samples collected in The Gambia) were excluded due to non-amplification or low genotype consistency (QI < 0.5 in Multiplex PCR 1).

Out of the 20 autossomic microsatellites included in the four multiplexes, six *loci* (corresponding to *loci* included in multiplex 3, 4 and 5) were discarded due to problems during allele scoring. The markers were discarded because the peaks were too high (i.e. peak height exceeded the scale of the electropherogram) to accurately discriminate between stutters and the true alleles. The inclusion of these *loci* in the dataset could create an error in distinguishing between heterozygotic and homozygote individuals. Therefore, the samples analysed under this study were genotyped for a maximum of 14 microsatellite *loci*.

In the end of the genotyping process for 14 microsatellite *loci*, 24 genotypes with a mean QI across *loci* of between 0.5 and 0.96 (average across samples of 0.72) were obtained. Of the 24 genotypes, two genotypes were identical and were excluded from the dataset. The probability of identify (pID = 1.1×10^{-6}) and probability of identity including siblings (pIDsib = 4.6×10^{-3}) indicated that the distinction of individuals is reliable with only six microsatellite *loci*.

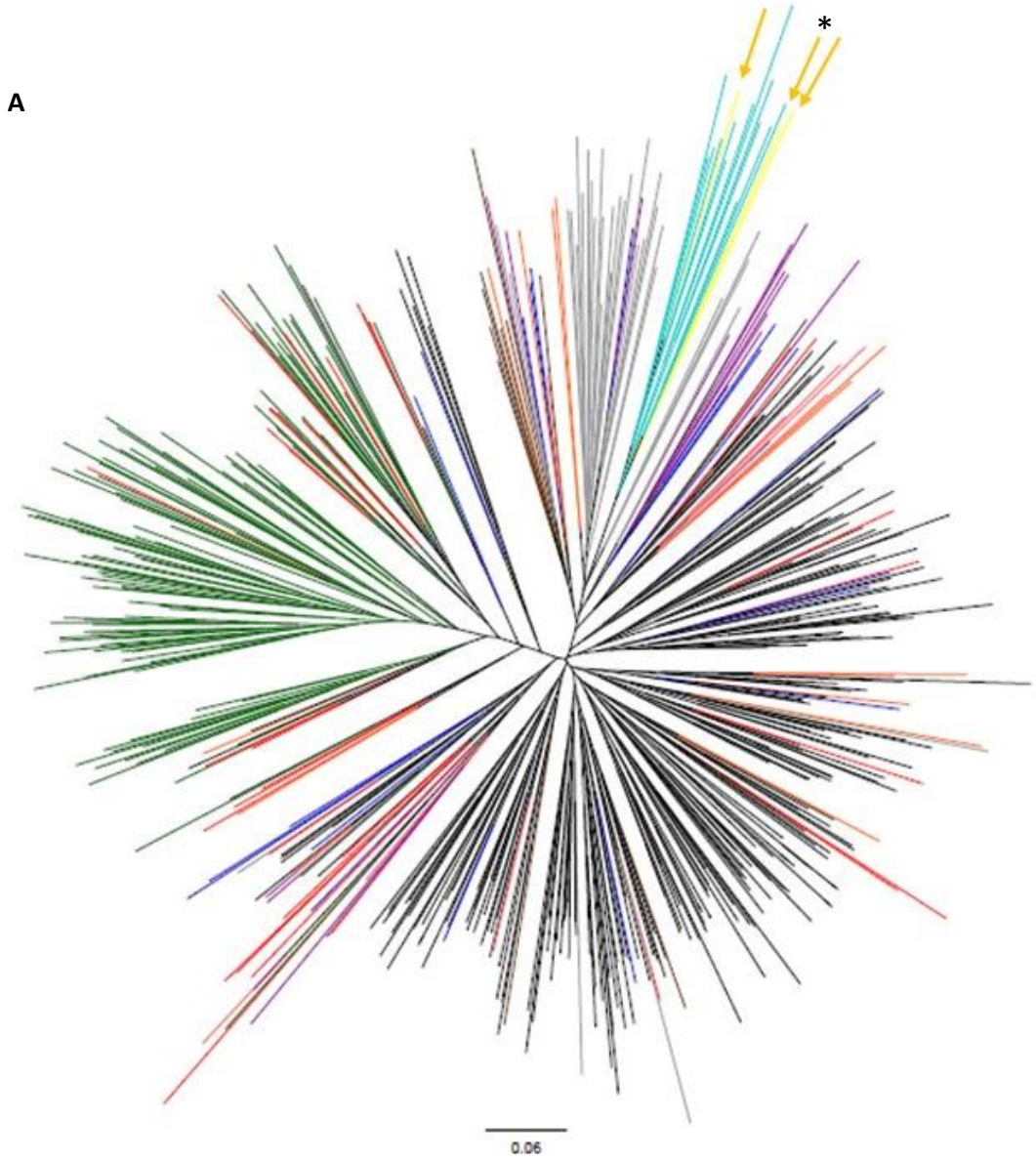
The tests carried out using MICROCHECKER did not reveal scoring errors due to stuttering or any evidence for large allele dropout.

3.1.2. Molecular determination of *Papio* samples

All 22 samples were successfully sequenced for the COI mtDNA barcoding region. The search in NCBI database for similar sequences using BLASTn revealed that: 1) six samples putatively identified as Guinea baboon showed between 99% and 100% of similarity with green monkey (*Chlorocebus sabaeus,* voucher KJ192752.1); 2) two samples putatively identified as Olive baboon showed 98% and 100% of similarity with Olive baboon (*Papio anubis*, voucher JX946197.2) and thirteen samples putatively identified as Guinea baboon and one sample putatively identified as Olive baboon

showed between 99% and 100% of similarity with Guinea baboon (*Papio papio,* vouchers AY972678.1 and JX946203.2).

The Neighbour-Joining (NJ) tree based on multilocus genotypes of 507 Guinea baboons, two individuals identified as Olive baboon by mtDNA barcoding and one individual described morphologically as an Olive baboon but identified as Guinea baboon by the mtDNA barcoding, suggests a weak genetic differentiation between the two species (Figure 10). Both Olive baboon individuals and the one from Abidjan zoo (identified in the tree using arrows) share a clade with Guinea baboon individuals from the Bissandougou population.



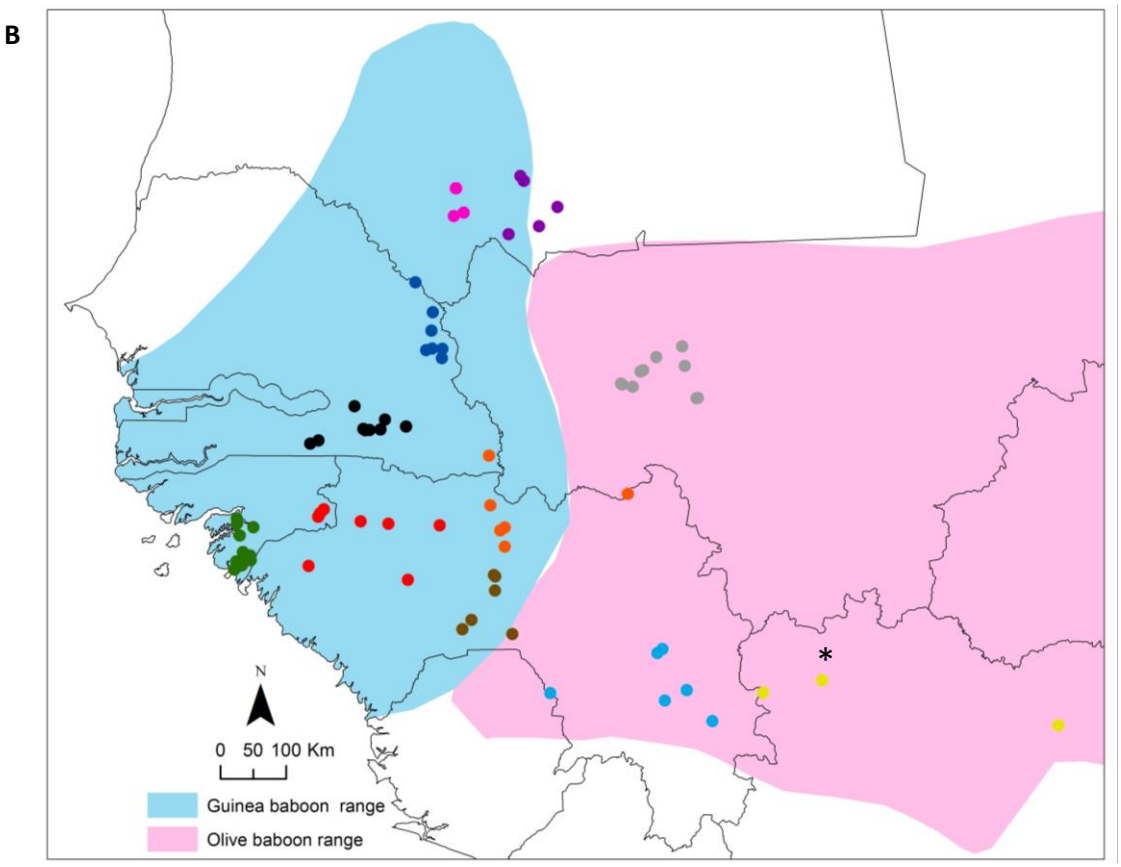


Figure 10. Results of the Neighbour-joing tree using all *Papio* genotypes. **A)** Neighbour-joining tree obtained for the 507 individuals of *P. papio*, 2 individuals of *P. anubis* and one potential hybrid calculated using distance shared alleles among multilocus genotypes. Each population of *P. papio* are represent by a different color (Coastal = green, Fouta Djalon W = red, Niokolo-Koba = black, Senegal river = dark blue, Fouta Djalon S = brown, Fouta Djalon N = orange, Assaba = pink, Afollé = purple, Boucle du Baoulé = grey, Bissandougou = light blue). Individuals identified with an arrow were identified as *P. anubis*. The potential hybrid is represented by an asterisk. **B)** Spatial location of the genotypes used in the Neighbour-joining tree. The IUCN distribution range of Guinea and Olive baboon are represented at blue and pink color, respectively.

The Bayesian clustering analysis performed in STRUCTURE to identify the genetic clusters assigned to both species did not separate the Guinea and Olive baboons in two different genetic groups (Figure 11). The highest model value in the ΔK distribution was when K = 9, which suggests the existence of nine main genetic clusters (Figure S5, supplementary material).

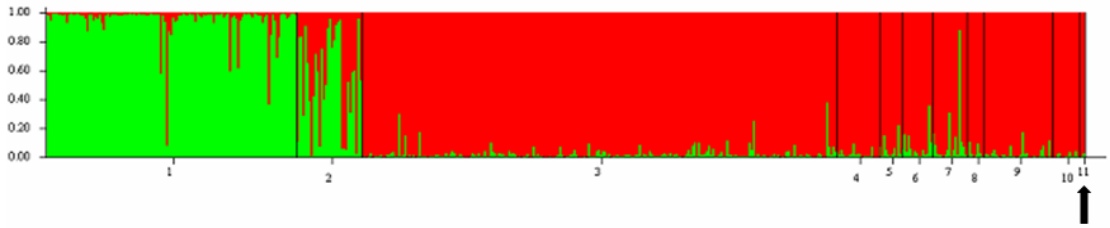


Figure 11. Bayesian admixture analysis of the 507 individuals of *P. papio*, 2 individuals of *P. anubis* and one potential hybrid, assuming clustering of K=2. Figure shows membership coefficient (Q) for an individual belong to one cluster. Each color represents one different genetic cluster and a single vertical bar represents each individual. The *P. anubis* and the potential hybrid are highlighted in the figure by an arrow.

3.1.3. Final database using autossomal microsatellite *loci*

The final dataset was formed by a total of 507 unique Guinea baboon genotypes (between six and 23 autosomal microsatellite *loci*) grouped in 10 populations (sample size ranging from eight to 233) (Figure 12). The database included 485 genotypes generated by previous studies from Guinea-Bissau, Guinea-Conakry, Senegal, Mauritania and Mali and 14 genotypes generated by the present study from Guinea-Conakry and eight genotypes obtained from the Mauritanian samples used as reference during the laboratory procedures.

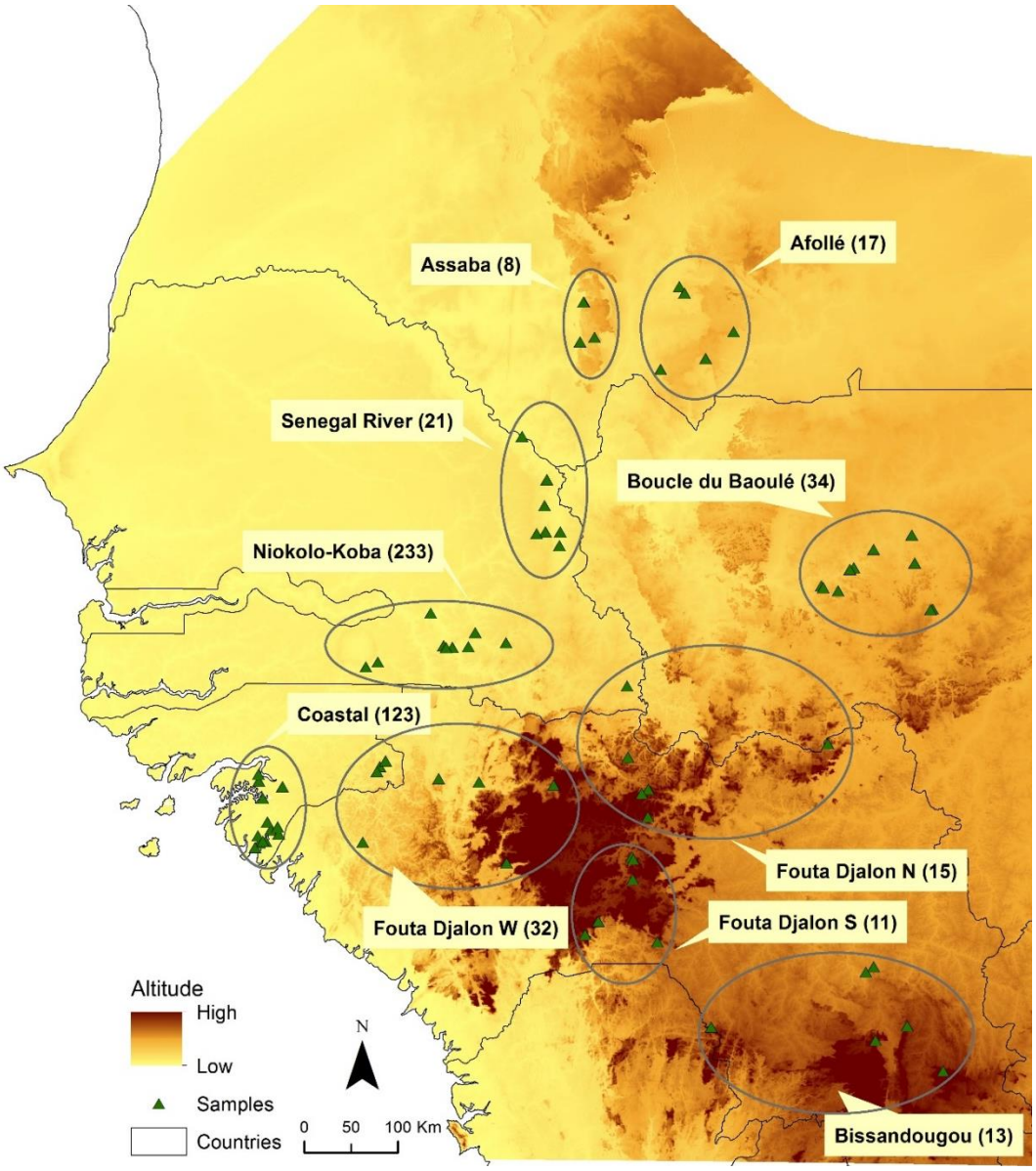


Figure 12. Spatial localization of the 10 putative populations of Guinea baboon used in present study and respective number of individuals.

3.1.4. Development of a variable Y-linked marker

Of the 15 Y-linked microsatellite markers tested in Guinea baboon samples, seven exhibited non-specific male amplification (i.e. positive amplification for the female samples) or amplification of multiples bands in males after PCR optimization and were discarded. Male-specific amplification of a single band was achieved for six *loci:* DYS470, DYS474, DYS511, DYS569, DYS57 and DYS594.

Samples from three males collected in Mauritania and three males collected in Guinea-Bissau were genotyped for the six male-specific *loci* to investigate variation in the allele's sizes. Of the six genotyped *loci*, five (DYS470, DYS474, DYS511, DYS569 and DYS594) were monomorphic and only the DYS576 showed to be potentially polymorphic for the two populations (Table 3). DYS576 screened for a total of 36 male samples collected across the entire distribution of the Guinea baboon suggested eight alleles with sizes ranging from 273 to 302 bp.

Table 3. Summary of the annealing temperatures tested by Y-linked marker and description of the obtained results. Positive results are highlighted in the table (✓).

Marker	AT (ºC)	Amplification result		
		Amplification	Male-specific	Polymorphism
DYS391	52, 54, 56, 58, 60	✓		
DYS470	58, 60, 62	✓	✓	
DYS472	56, 56, 60	Nonspecific		
DYS474	58, 60, 62	✓	✓	
DYS511	56, 58, 60	✓	✓	
DYS557	50, 52, 54	Nonspecific		
DYS569	61, 63, 65	✓	✓	
DYS571	58, 60, 62	Nonspecific		
DYS574	56, 58, 60	✓		
DYS576	50, 52, 54, 56, 58, 60	✓	✓	✓
DYS579	61, 63, 65	Nonspecific		
DYS594	56, 58, 60	✓	✓	
DYS632	58, 60, 62	Nonspecific		
DYS643	56, 58, 60	Nonspecific		
DYS645	56, 58, 60	Nonspecific		

Sequencing results from one allele of each size revealed that locus DYS576 is an imperfect microsatellite (see Figure 13). Contrary to human species, in which DYS576 varies consistently in four nucleotides (CTTT), for the Guinea baboon the repetition motif can vary between two (CT) up to eight nucleotides (CTTTTTTT). However, the repetition motif is not consistent between the eight sequenced alleles, as is exemplified in Figure 14. The comparison between sequencing results of alleles 272, 292, 300 and the allele 298 showed that: (1) the variation between the alleles 298 and 300 is inexistent (Figure14A) and (2) the variation between the alleles 298 and 292 (6 bp variation, Figure14B) and the allele 298 and 277 (21 bp variation, Figure14C) is not due to regular increases in the number of repetition motifs. These results suggest that the variation observed in fragment analyses can occur in the section of the repetition motives or in the flanking regions, which makes scoring of this marker through fragment analyses error-prone.

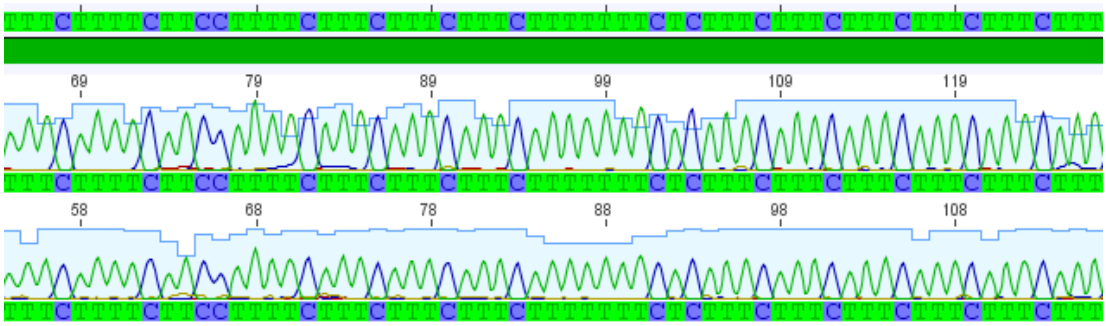


Figure 13. Partial electropherogram obtained from sequencing of two different alleles from the locus DYS576. The electropherogram indicates that this marker is an imperfect microsatellite.

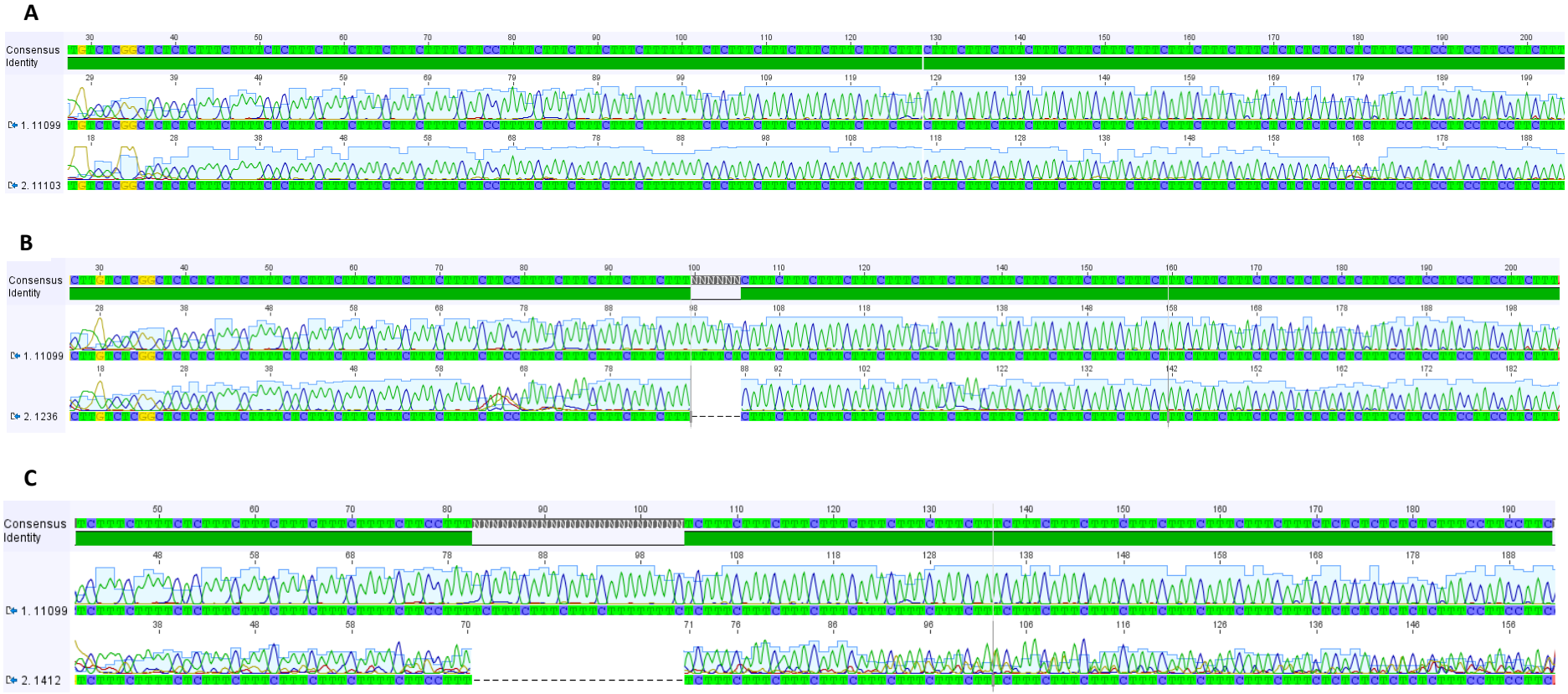


Figure 14. Partial electropherogram obtained from the sequencing of four alleles from the locus DYS576 with different size (272, 292, 298 and 300 bp). A) comparison between the allele 298 and 300 (no variation); B) comparison between the allele 298 and 292 (6 bp variation); C) comparison between the allele 298 and 277 (21 bp variation).

3.1.5. Genetic diversity

Across all samples, the number of alleles per locus ranged from five (D5s1457, D8s505, D1s207) to 11 alleles (D21s1142). The overall observed heterozygosity was of 0.56 and the total heterozygosity (heterozygosity within populations and between populations) was of 0.66 (Table S7, supplementary material).

All *loci* were polymorphic in all populations with the exceptions of the D6s311 (monomorphic in Assaba) and D3s1766 and D8s505 (monomorphic in Assaba and Afollé). The number of alleles per population ranged from 3.33 (Assaba) to 5.00 (Boucle du Baoulé) and the number of effective alleles ranged from two (Fouta Djalon S) and three (Fouta Djalon N). Observed heterozygosity (H_o) ranged between 0.48 (Assaba) to 0.60 (Bloucle du Baoulé). The H_o was lower than the expected heterozygosity (H_e) for all populations, except for Coastal, Niokolo-Koba and Assaba, which has a negative inbreeding coefficient (F_{IS}). Private alleles were found in all populations with the exception of Coastal and Fouta Djalon S. The higher number of private alleles was found in Bissandougou (PA = 14) (Table 4). No consistent departures from Hardy-Weinberg equilibrium were detected considering the ten geographically populations: a significant departure was observed for *locus* D4s243 in Niokolo-Koba and for *locus* D2s1326 in Senegal River. Linkage disequilibrium (LD) after Bonferroni correction for multiple comparisons was found for four *loci* pairs in Niokolo-Koba.

The levels of genetic diversity between the populations were similar using the same number of *loci* for all populations (13 microsatellites *loc*i; Table S8, supplementary material) and using all *loci* included in the database.

Table 4. Genetic diversity measures for each Guinea baboon population (n = 507 individuals) genotyped for 6 - 23 microsatellite loci. N = number of samples per population, Na = Number of Different Alleles, Ne = Number of effective alleles, H_o = Observed Heterozygosity, H_e = Expected Heterozygosity, F_{IS} = inbreeding coefficient, %P = Percentage of Polymorphic Loci, P_A = Private alleles, HWE = number of loci significantly departuring from HWE, p<0.05, ** p<0.01 or *** p<0.001.

Pop	N	N_a	N_e	H_o	H_e	F_{is}	%P	PA	HWE
Coastal	123	4.615	2.502	0.571 ± 0.03	0.570 ± 0.03	-0.002	100.00%	0	0
Fouta Djalon W	32	4.538	2.732	0.578 ± 0.03	0.618 ± 0.02	0.073	100.00%	1	0
Niokolo-Koba	233	5.000	2.882	0.566 ± 0.05	0.617 ± 0.03	-0.039	100.00%	5	1***
Senegal River	21	4.391	2.929	0.588 ± 0.04	0.606 ± 0.03	0.017	100.00%	3	1***
Fouta Djalon S	11	3.696	2.001	0.492 ± 0.06	0.447 ± 0.05	0.006	100.00%	0	0
Fouta Djalon N	15	4.739	3.009	0.576 ± 0.04	0.617 ± 0.03	0.081	100.00%	1	0
Assaba	8	3.333	2.295	0.478 ± 0.07	0.463 ± 0.06	-0.041	83.33%	1	0
Afollé	17	4.222	2.732	0.542 ± 0.06	0.556 ± 0.05	0.020	88.89%	4	0
Boucle du Baoulé	34	5.043	2.639	0.596 ± 0.03	0.572 ± 0.03	0.026	100.00%	3	0
Bissandougou	13	4.304	2.714	0.556 ± 0.04	0.569 ± 0.04	0.081	100.00%	14	0

Genetic distance based on the distance alleles shared (DAS) varied between 0.02 and 0.49. The lowest distance was observed between Senegal river and Fouta Djalon N and the highest distance was observed between Coastal and Bissandougou. The topology of the NJ tree based on DAS highlights the pattern of lowest genetic distance between neighbouring populations (e.g. Coastal and Fouta Djalon W) and greater genetic distance between populations from West and East (e.g. Coastal and Bissandougou) (Figure 15A).

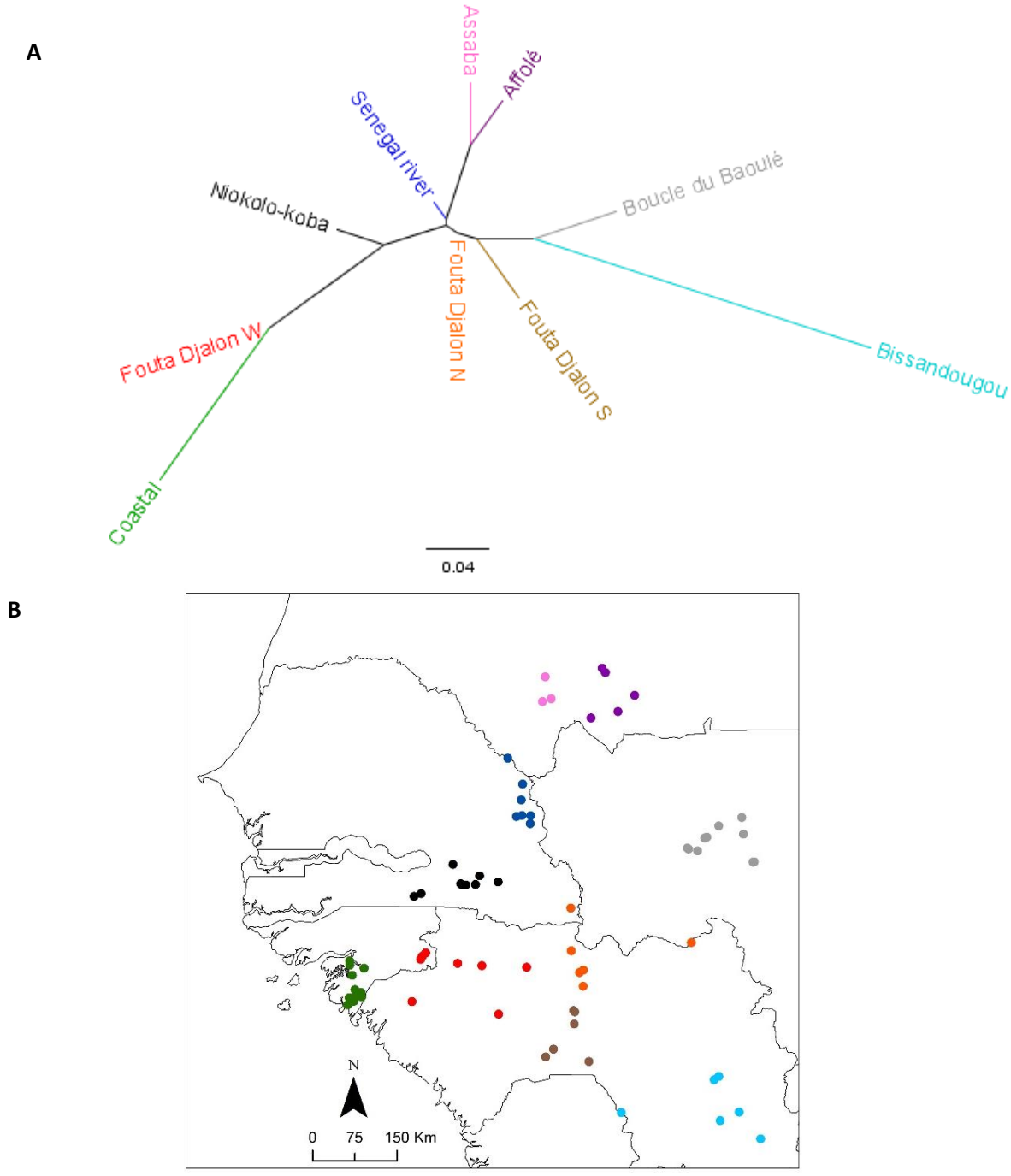


Figure 15. Neighbor-joining tree (A) and spatial location (B) of the 10 putative populations of *P. papio* based on the shared allele distance among multilocus genotypes.

3.1.6. Population structure and differentiation

The Bayesian clustering analysis performed in STRUCTURE, assuming either independent or correlated allele frequencies, suggested the presence of two genetic clusters using the ΔK method. However, while K = 2 was the highest model value in the ΔK distribution, a second elevated modal value appeared at K = 4 (Figure 16A). STRUCTURE analysis suggests a deep division at K = 2 and a subdivision at K = 4. The clustering solution with the highest log-likelihood [Ln P(X/K)] was K = 4. Figure 16B shows the results obtained from STRUCTURE assuming K = 2, 3 and 4 as the number of genetic clusters.

In the K = 2 solution, samples collected at Coastal and Fouta Djalon W were grouped together in Cluster 1, while all other samples were grouped together in a second cluster. In the K=3, samples collected in Coastal and Fouta Djalon W were grouped in one cluster, the samples collected in Niokolo-Koba were grouped in a second cluster and the other samples were grouped in a third cluster. The K = 4 solution separated the samples collected in (1) Coastal and Fouta Djalon W; (2) Assaba, Fouta Djalon W, N and S, Senegal river and Niokolo-Koba; (3) Niokolo-Koba and (4) Afollé, Boucle du Baoulé and Bissandougou.

The ranked individual Q values showed a nearly continuous distribution (except for cluster 1; see Figure S7, supplementary material). Therefore, a threshold of 0.80 was used to assign the individuals to the four genetic clusters uncover by STRUCTURE. A total of 117 individuals sampled in the Western part of the species distribution (in Coastal and Fouta Djalon W populations) were assigned to Cluster 1. Cluster 2 is formed by 59 individuals, mostly sampled in Assaba, Fouta Djalon W, N and S, Senegal river, and Niokolo-Koba populations. Cluster 3 consisted exclusively of 129 individuals from Niokolo-Koba and occupies a more central position in the species distribution. Cluster 4 is formed by 61 individuals sampled in Afollé, Boucle du Baoulé and Bissandougou populations and occupies a marginal position in eastern and northern areas of the species distribution (Figure 17). The remaining 141 individuals (Q < 0.80) were treated as a product of admixture between clusters. Higher levels of admixture were found in Fouta Djalon W (78% of individuals) and Assaba (63% of individuals) populations while lower levels of admixture were found in Bissandougou (8% of individuals) and Coastal (9% of individuals). There was no evidence of admixture in the Boucle du Baoulé population.

Inverse distance weighted interpolation implemented in ArcGIS calculated using the individual Q values estimated by STRUCTURE suggested a trend for a cline variation

pattern, where the majority of individuals from neighbouring populations were grouped together in the same cluster (Figure 17). Genetic discontinuities in genetic variation were observed in the contact zones between Clusters 1 and 2 in Guinea-Conakry and in Senegal (A), between Clusters 2 and 3 in Senegal (C), and Clusters 2 and 4 in Guinea-Conakry (B).

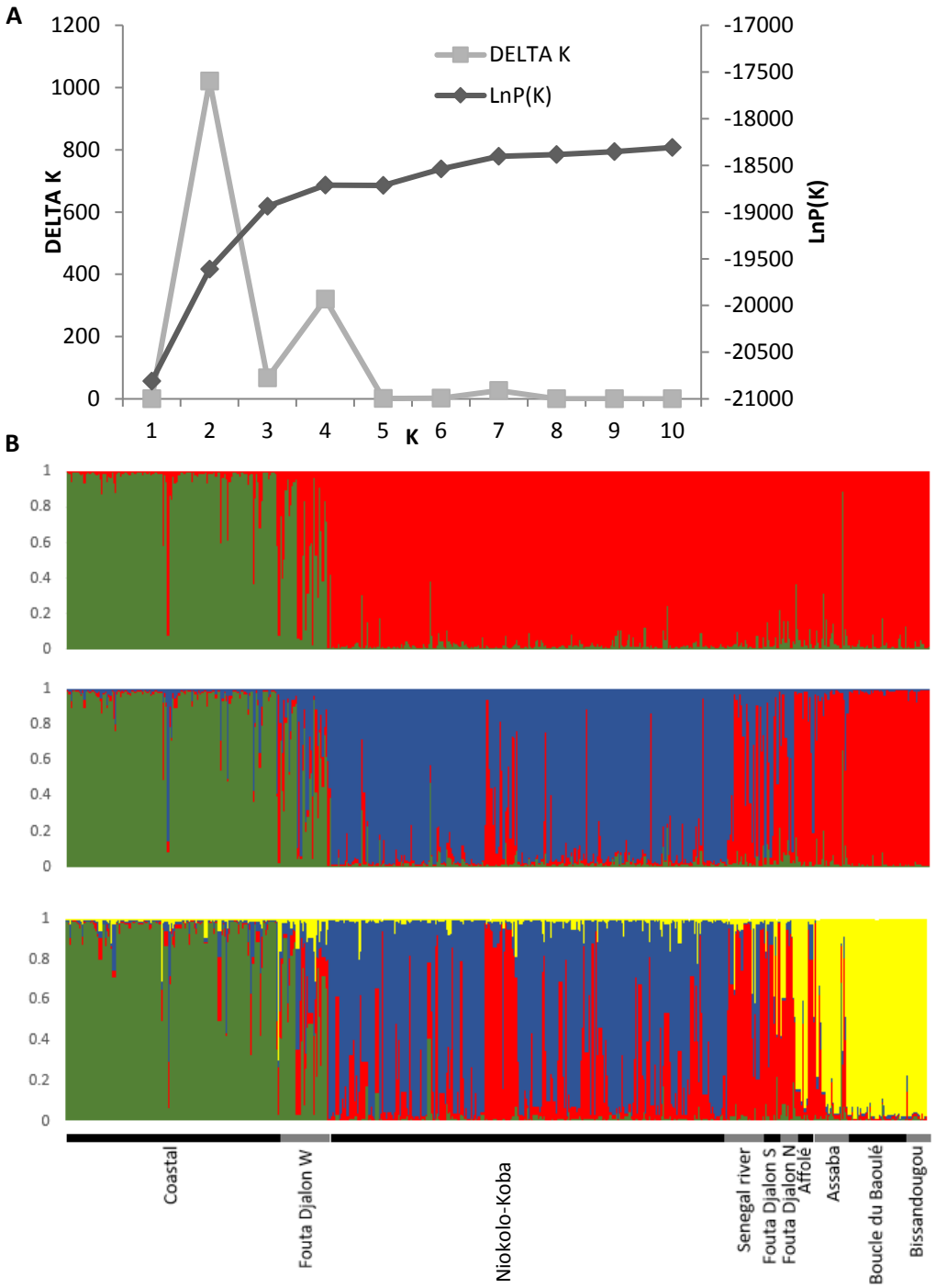


Figure 16. Results of the STRUCTURE clustering analyses using the Guinea baboon samples. A) Inference of the most probable number of clusters (K) using Delta K and mean LnP(K) values across all runs suggest the existence of 2 to 4 genetic clusters; B) Bayesian admixture analysis of *P. papio* assuming clustering of K=2, K=3 and K=4 (from top to bottom). Figure shows membership coefficient (Q) for an individual belong to one cluster. Each color represents one different genetic cluster and a single vertical bar represents each individual. The name of populations where the individuals were sampled is indicated below.

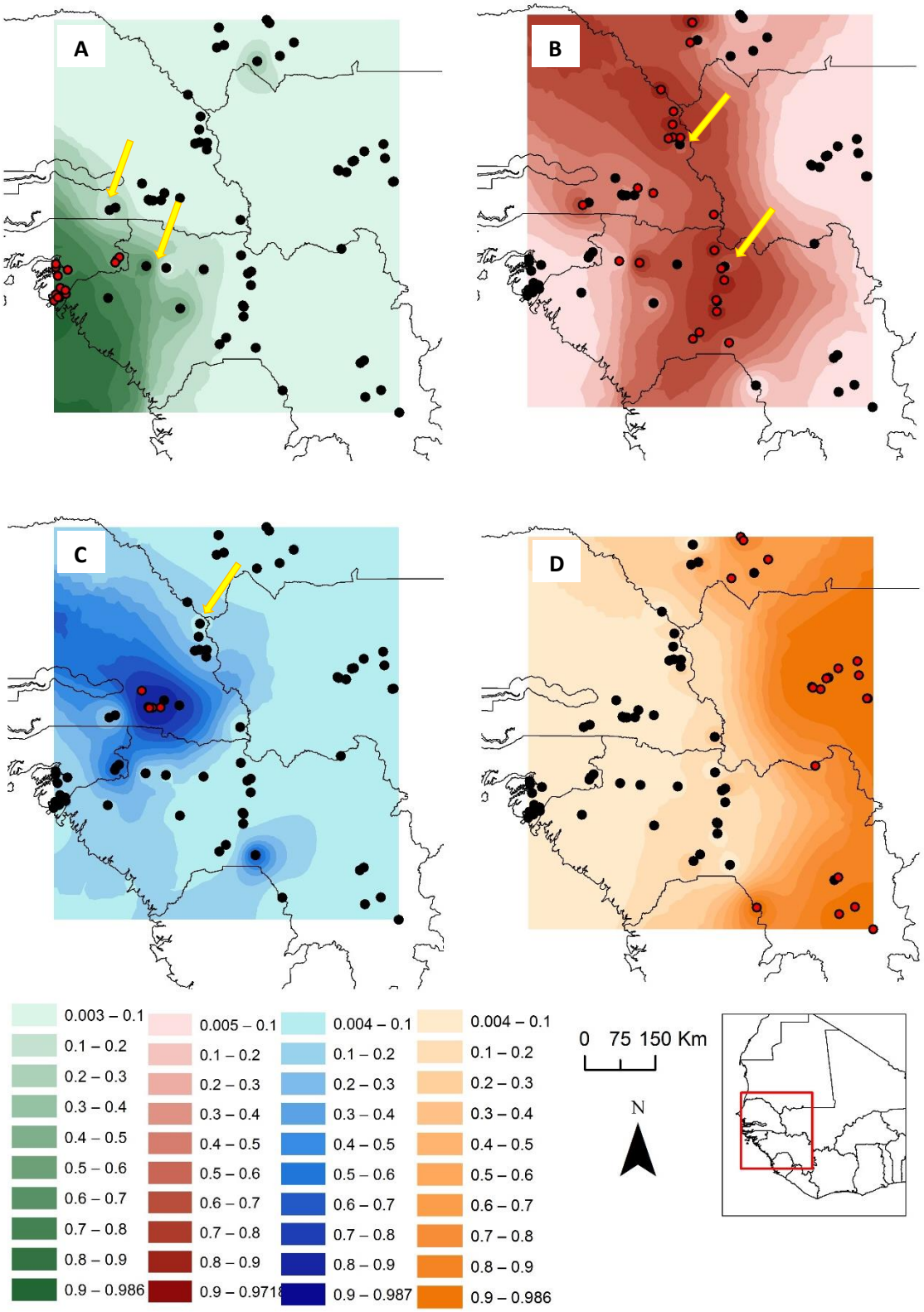


Figure 17. Interpolation (IDW) of the geographical ranges of the four genetic clusters of *P. papio* calculated using the individual Q values estimated by STRUCTURE: A) cluster 1; B) cluster 2; C) cluster 3 and D) cluster 4. Q values were divided into ten classes. Arrows indicate genetic discontinuities.

Considering individuals with Q>0.8 to each cluster, the overall F_{ST} value for Guinea baboon was 0.184 ($p<0.001$). The four genetic clusters uncovered by STRUCTURE were significantly differentiated (Table S9, supplementary material). Pairwise F_{ST} between clusters pointed to high levels of genetic differentiation between the cluster 1 and cluster 4 (F_{ST} = 0.206) whereas pairwise F_{ST} values between clusters 2, 3 and 4 were moderated (F_{ST} between 0.045 and 0.098).

The genetic diversity estimators for the four genetic clusters are given in Table 5. In general, cluster 4 (formed by eastern populations of Afollé, Boucle du Baoulé and Bissandougou) showed higher genetic diversity levels in comparison with the others clusters: the number of effective alleles (Ne = 3.3) found within this cluster was higher and displays the higher number of private alleles (PA = 34). All clusters showed similar values of observed heterozygosity but clusters 1 and 3 display a higher observed heterozygosity than expected, resulting in a negative inbreeding coefficient (F_{IS}). Departures from Hardy-Weinberg equilibrium were detected in cluster 4.

Table 5. Genetic diversity measures for each Guinea baboon cluster (n = 366 individuals; Q>0.8) genotyped for 6 - 23 microsatellite *loci*. N = number of samples per population, Na = Number of Different Alleles, Ne = Number of effective alleles, Ho = Observed Heterozygosity, He = Expected Heterozygosity, F_{IS} = inbreeding coefficient, %P = Percentage of Polymorphic Loci, PA = Private alleles, HWE = departures from HEW, p<0.05, ** p<0.01 or *** p<0.001.

Pop	N	Na	Ne	Ho	He	F	%P	PA	HWE
Cluster 1	117	4.462	2.475	0.563 ± 0.03	0.562 ± 0.03	-0.005	100.00%	2	0
Cluster 2	59	4.913	2.776	0.547 ± 0.03	0.587 ± 0.03	0.059	100.00%	10	0
Cluster 3	129	4.000	2.525	0.588 ± 0.04	0.551 ± 0.03	-0.067	100.00%	0	0
Cluster 4	61	6.174	3.322	0.589 ± 0.03	0.656 ± 0.03	0.098	100.00%	34	5***

Bayesian analyses in STRUCTURE were repeated to investigate the presence of sub-structure within each of the four clusters. For cluster 4, the highest modal value in ΔK distribution was for K=3 (Figure 18A). The first sub-cluster identified was comprised by individuals from Afollé (cluster 4a), the second sub-cluster grouped all individuals from Boucle du Baoulé (cluster 4b) and individuals from Fouta Djalon N, and the third sub-cluster grouped all individuals from Bissandougou (cluster 4c) (Figure 18B). The spatial limits of the three genetic clusters are shown in Figure 18C. Sub-structure was not found in the remaining clusters (1, 2 and 3) (Figure S8, supplementary material).

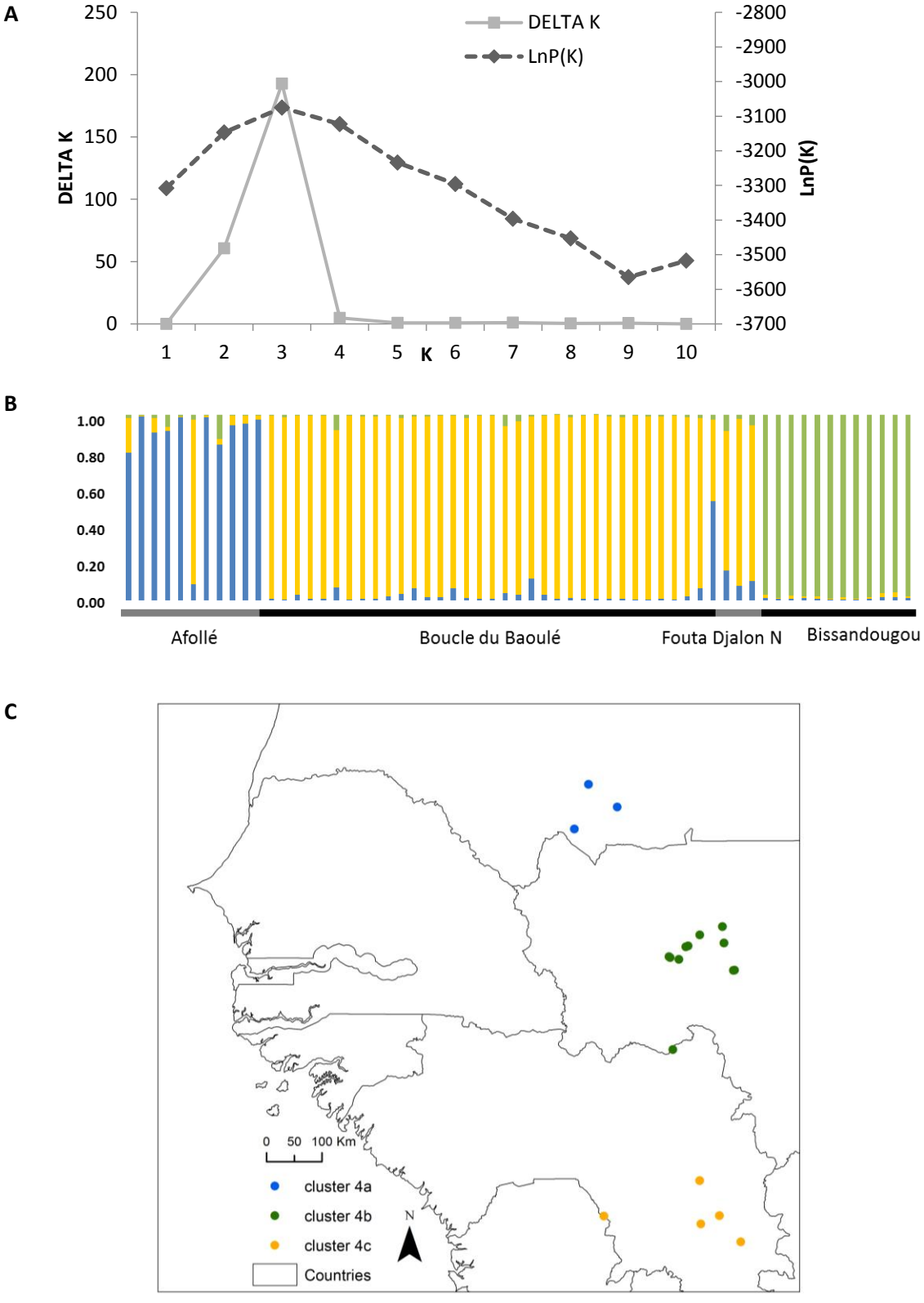


Figure 18. Results of the STRUCTURE clustering analyses using the Guinea baboon genotyped assigned to cluster 4. A) Inference of the most probable number of K using Delta K and mean LnP(K) values suggest the existence of sub-structure within the cluster 4. B) Bayesian admixture analysis performed only with individuals assigned to cluster 4 (N = 61) assuming clustering of K=3. C) Geographical ranges of the three genetic clusters of the cluster 4 calculated using the Q values estimated by STRUCTURE.

The 10 independent spatial analysis performed in BAPS suggested the presence of three genetic clusters: (1) a western cluster formed by individuals sampled in Coastal and Fouta Djalon W populations; (2) a central cluster formed by individuals sampled in the North of Assaba, Fouta Djalon W, N and S, Senegal river and Niokolo-Koba populations; (3) an eastern cluster mainly formed by individuals sampled in South of Assaba, Afollé, Boucle du Baoulé and Bissandougou populations (Figure 19). The clusters identified by BAPS were consistent with those obtained using STRUCTURE, but the latter suggested an extra cluster which was formed by samples collected exclusively at Niokolo-Koba.

The results obtained using BAPS agrees with the location of genetic discontinuities identified by the interpolation based on the individual Q values estimated by STRUCTURE: individuals with a high probability of assignment to cluster 2 and 3 were sampled in close proximity in Guinea-Conakry, Senegal North and in Mauritania.

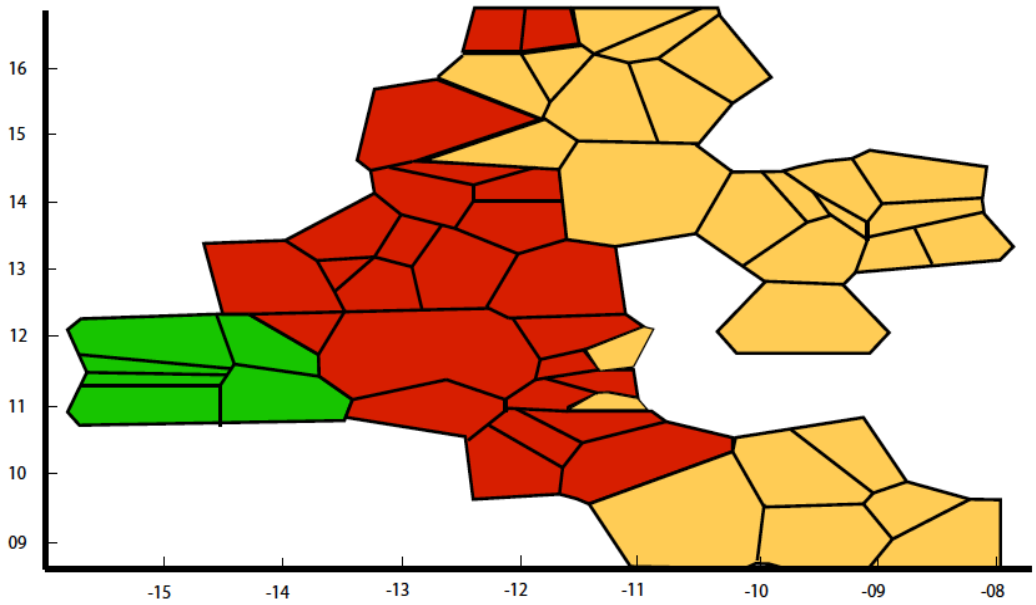


Figure 19. Spatial genetic structure using BAPS spatial analysis. BAPS suggest the presence of three different genetic clusters, indicated in the figure by different colors (Cluster 1 in blue, Cluster 2 in red, Cluster 3 in orange). The spatial location of each sample is showed above.

Levels of relatedness within the 10 geographic populations (N=507) or for cluster 2, 3 and 4 (N=366) were not significant (Figure S9, Supplementary material). However, levels of relatedness for cluster 1 (formed by Coastal and Fouta Djalon W) were significantly higher than zero (Figure S10, Supplementary material).

The CLUST_ DIST software, which is based in Nei minimum genetic distance between individuals, was run to group individuals in the most likely number of clusters estimated by STRUCTURE and BAPS. The analysis performed in CLUST_ DIST considering K = 3 and K = 4 solutions grouped the individuals collected in Coastal and Fouta Djalon W in a single cluster (Figure 20), suggesting that grouping of individuals from Coastal and Fouta Djalon W in a separate cluster by STRUCTURE and BAPs is not an artefact caused by higher relatedness levels observed in these cluster.

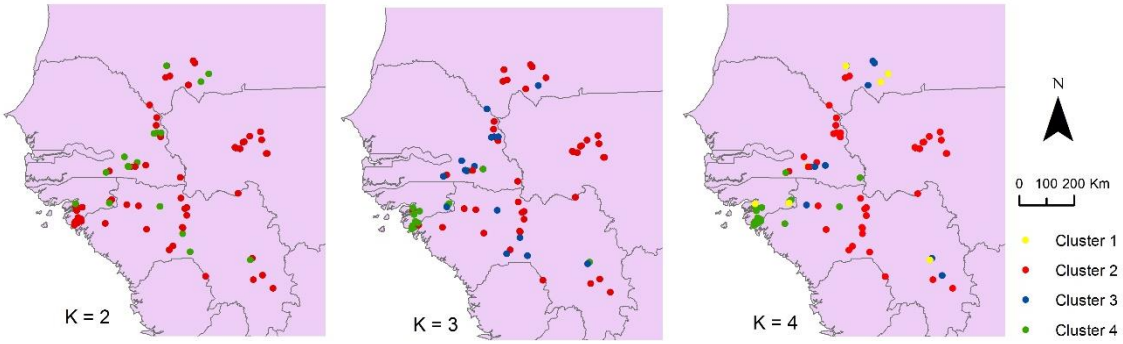


Figure 20. Clustering analysis of *P. papio* (N=507) performed in CLUST_DIST (based in Nei minimum distance) assuming clustering of K=2 to K=4.

3.1.7. Isolation by distance (IBD)

A positive and significant correlation between Euclidean geographical distances and genetic distances were observed (Mantel test, Pearson correlation coefficient R = 0.5676, p<0.001) (Figure 21), suggesting a pattern of isolation by distance. The regression line has a high intercept (DAS = 0.5), suggesting that within sampling sites pairwise genetic distances may be considerable.

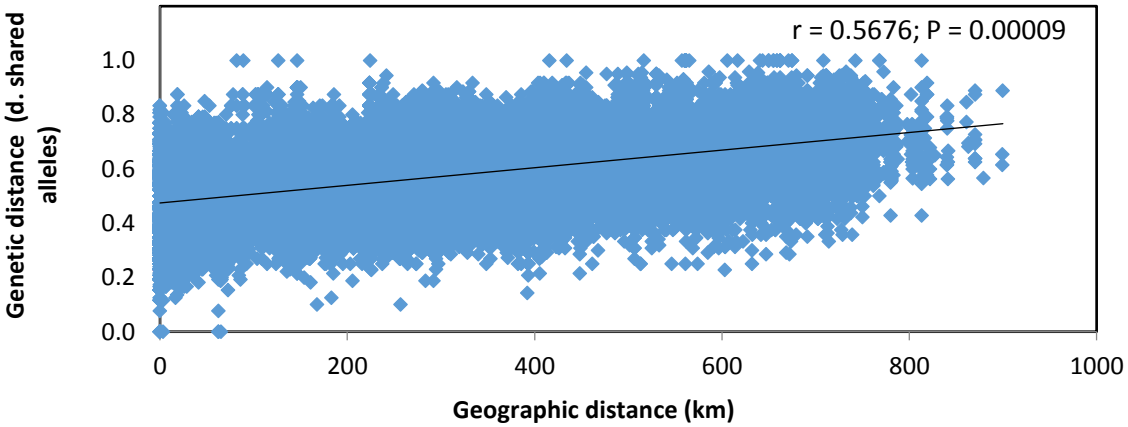


Figure 21. Analysis of isolation by distance. The present graph shows the relationship between genetic distances (distance shared alleles) as a function of geographical distance (km) for *P. papio*.

3.2. Habitat suitability modelling

The ROC plots exhibited high average AUCs with low standard deviations (SD) for environmental and human models. The average AUC for the environmental model was of 0.849 ± 0.024, whereas the AUC for the human model was of 0.903 ± 0.019, suggesting a better performance for the distribution model based on both environmental and anthropogenic variables.

The PC layers that most influenced the distribution of Guinean baboon in environmental model were PC 1 and PC 3 (Figure S3 and Table S10, Supplementary material) while PC1, PC3, PCH1 and PCH2 were more important for the human model (Figure S4 and Table S11, Supplementary material). These results suggest that the anthropogenic variables have a high percentage of contribution in ecological niche modelling of the Guinea baboon.

The environmental model predicted an almost continuous distribution of suitable habitats for species occurrence, while the human model predicted a more fragmented distribution of suitable areas, suggesting the existence of putative fragmented patches (Figure 22). Both ENMs predicted the existence of suitable areas outside the known species distribution.

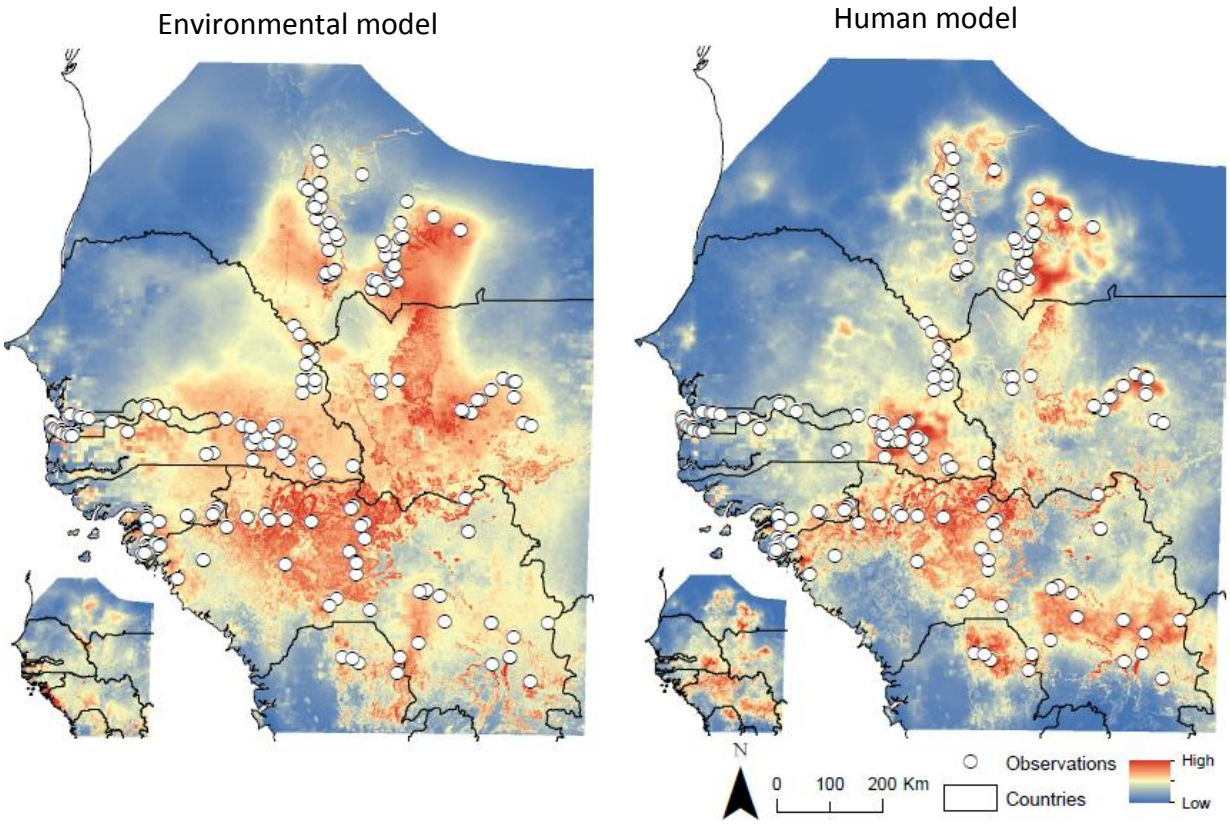


Figure 22. Guinea baboon models generated using MAXENT based on environmental features ("Environmental model", left) and based on environmental and anthropogenic features ("Human model", right). Each distribution map is accompanied by the respective standard deviation, indicated on the left of each model. High stands show the highest suitable areas for Guinean baboon.

3.3. Landscape genetics

No major observable differences were found between the environmental and human current maps resulting from the circuit theory analysis (Figure 23). Both current maps suggested higher connectivity among the western populations and a decrease in connectivity from western to eastern/northern populations. The easternmost populations of Assaba, Afollé, Boucle du Baoulé and Bissandougou seem to be unconnected with westernmost populations.

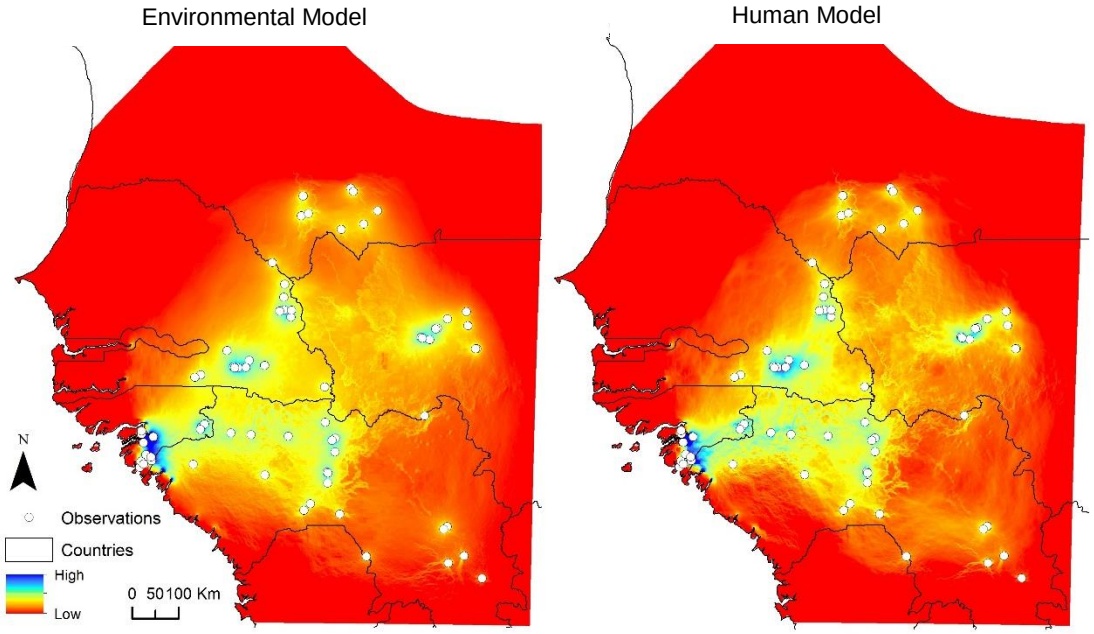


Figure 23. Current maps output from CIRCUITSCAPE based on environmental model (left) and based on human model (right). Results showing cumulative resistance between populations of Guinea baboon based on the combined effect of environmental variables. High stands for the highest current flow.

Mantel tests showed that geographical distance exhibited the highest correlation coefficient with genetic distance ($r = 0.568$, $p<0.001$; Figure 24). All Mantel tests conducted with environmental matrixes indicated a significant correlation between genetic distance and presence of environmental barriers, while tests conducted with human matrixes showed a weak positive correlation between the genetic differentiation and presence of human settlements. However, after controlling for geographical distance, partial Mantel test revealed that both environmental and human correlations were not significant.

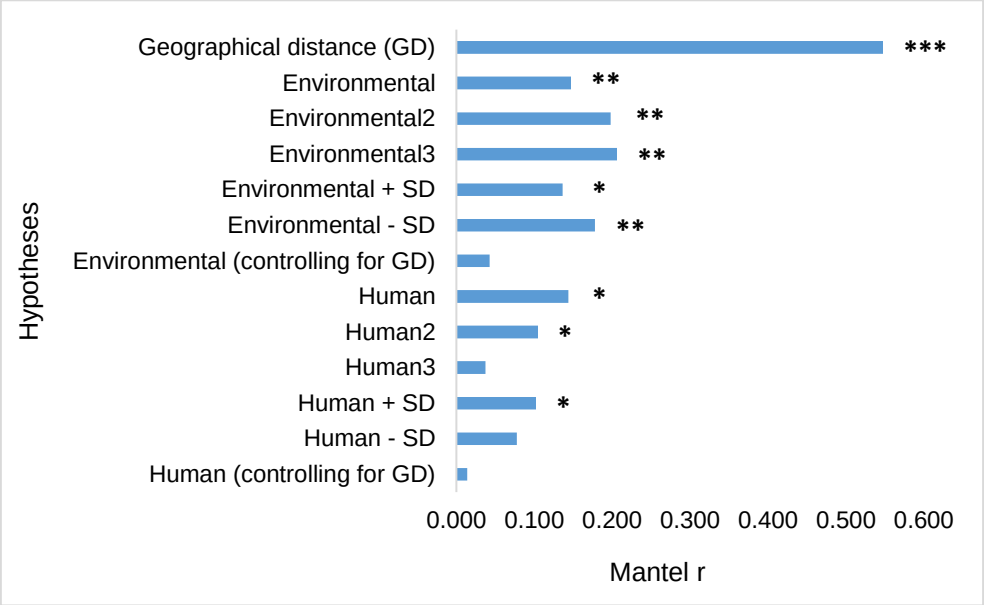


Figure 24. Results of Mantel tests testing the correlations (r) between the pairwise genetic distance (distance shared alleles) and models of spatial resistance (geographical distance, Environmental resistance model and Human resistance model). Significant values indicated with * p<0.05, ** p<0.01 or *** p<0.001.

4. Discussion

This is the first landscape genetics study of the Guinea baboon (*Papio papio),* an ecologically plastic primate that inhabits heterogeneous landscapes in West-Africa*.* To investigate the potential effects of environmental and anthropogenic features in connectivity between populations at the species range scale, the most updated and complete genetic and spatial databases were assembled. A total of 507 individuals sampled in 10 geographically distinct populations distributed across the range of the species, with the exception of Sierra Leone and The Gambia, were genotyped for a maximum of 23 microsatellites *loci.* Species-specific habitat suitability models were built using 362 observations. For the first time, distinct models based on environmental factors and on human-related features were created separately, allowing comparing the species response to both types of features in terms of distribution and dispersal. In addition, variable Y-linked microsatellite markers for Guinea baboon were investigated.

4.1. Sources of uncertainty

Several factors may have introduced bias in the results and, consequently, in the interpretations that will be discussed below. A possible source of uncertainty in the present study is related with the likely hybridization between *P. papio* and *P. anubis* and potential pitfalls of analyses.

4.1.1. Hybridization between *P. papio* and *P. anubis*

Hybridization between *P. papio* and *P. anubis* may potentially occur in the study area. The analyses based on multilocus genotypes revealed weak genetic distances between the two species. For instance, STRUCTURE clustering analyses did not separate the two species assuming the K = 2 clustering and *P. anubis* samples were grouped with *P. papio* individuals in the NJ tree. However, the weak genetic distances between the two species can be misleading by several methodological factors, namely the dissymmetry in the number of samples of Guinea and Olive baboon used in STRUCTURE analysis (N = 507 and N = 2, respectively) or the lower number of genetic markers used to genotyped the Olive baboons (i.e. 14 microsatellites *loci*). For instance, Evanno et al. (2005) suggested that the performed of STRUCTURE decreases when used with different sample sizes and with low number of *loci*. Alternatively, the weak genetic distances observed between samples assigned to both species could be related

with a phylogenetic close relation between the two species and a great similarity in the genome between them (Zinner et al., 2013). Still, if some of the samples identified as Guinea baboon by mtDNA corresponded to hybrid individuals between the two species, it was considered that the mother of that individual (who transmitted the mtDNA genome) used the specific area to live or disperse. As such, even if the weak genetic distances observed between the two species suggest the presence of potential hybrids, their presence should not have biased the habitat suitability and connectivity analyses. Therefore, the obtained results of the spatial analysis are valid to infer the Guinea baboon habitat suitability and the connectivity between the habitat patches.

4.1.2. Potential pitfalls of analyses

The successful dispersal of animals across the landscape is determined by many biological processes, including its feeding ecology, foraging, social organisation, predation risk and reproduction (Taylor et al., 2011; Zeller et al., 2012). Consequently, the correct prediction of the impact of landscape fragmentation and heterogeneity on functional connectivity of Guinea baboon requires specific ecological and behavioural data that are not yet available for large scales. In fact, one of the major limitations of landscape genetics is related with lack of information on species-specific movement and dispersal behaviours. Movement dynamics are central to landscape genetics studies but obtaining reliable movement data of wild animals remains challenging (Zeller et al., 2012; Dexter et al., 2016). Technological advances have been allowing remote monitoring of animal movement, for instance by the installation of motion-activated cameras or by tracking the animals using radio-telemetry. However, the use of remote technology is unable to address movements and dispersal integrated in a large time scale (e.g. decades or centuries) and is extremely costly and time-consuming (Dexter et al., 2016).

Availability of data on the impacts of human activities, such as hunting and human persecution, in the survival of Guinea baboons (Imong et al., 2014) constitutes another constraint to the connectivity analyses. Although there are data on the conservation status and threats faced by the Guinea baboon from Guinea-Bissau, Guinea-Conakry and Senegal (Casanova and Sousa, 2007; Galat-Luong et al., 2006; Ferreira da Silva et al., 2013, 2014), there is little information in the remaining areas of the distribution. Consequently, it is difficult to hypothesize *a priori* how human activities could affect functional connectivity in Guinea baboon and to design specific tests.

Ecological models suggested that the presence of Guinea baboon is highly related with water availability and less with vegetation type, which corroborated previous

studies (Vale et al., 2015a) based on distinct datasets of habitat variables (habitat classification by Bicheron et al., 2008). Taken into account that the most relevant landscape features for habitat selection might not be the same as for dispersal (Zeller et al., 2012), the areas selected by Guinea baboons for living and for dispersing can exhibit discrepancies. Such differences may also be emphasised by the high ecological adaptability of the species, where individuals may display a higher plasticity in the choice of areas to disperse. As such, Guinea baboon can use more areas to disperse than those predicted by the ENMs.

The software CIRCUITSCAPE has been widely used to predict connectivity in heterogeneous landscapes using electronic circuit theory (Shah and McRae, 2008). However, the computational algorithm used by CIRCUITSCAPE considers pairwise resistance distances between individuals from locations separated by at least 1 km^2. For that reason, it was randomly selected one genotyped individual whenever several individuals were available in each 1x1km grid cell. As such, it is possible that the selected individual exhibited the lowest or highest genetic distance within a 1x1km grid cell, thus deflating or inflating, respectively, genetic distances in relation to other sampled localities. Also, of the 507 individuals genotyped only 74 individuals were used in the connectivity analyses, which mean that the genetic variation within each location (i.e. samples collected at distances < 1km) was ignored. This simplification required by the computational algorithm could have caused a fluctuation on the real genetic variation in each sampled locality.

Another potential source of uncertainty in the connectivity analysis is related with the exclusive use of the shared alleles (DAS) genetic distance. The DAS has been used in other landscape genetics studies (e.g. Russo et al., 2016) because it is based on the stepwise mutation model (Goldstein et al., 1995). Comparisons with other type of genetic distance measures (such as Nei' minimum genetic distance) were not performed to investigate the presence of incongruence between results. Therefore, it is unclear if the connectivity results vary according to the genetic distance used in analysis.

The suitability of Mantel and partial Mantel tests in landscape genetics has been questioned. Some authors have defended that the use of this method of testing correlations between two distance matrices can result in type I error, i.e. in the incorrect rejection of the null hypothesis resulting in "false positives" (e.g. Cushman et al., 2013). Conversely, other authors (e.g. Castillo et al., 2014; 2016) defend that Mantel tests can be successfully used to identify the underlying landscape variables influencing gene flow, except when the aim of the study is to determine the magnitude of resistance for individual landscape variables (i.e. the most significant landscape features shaping gene

flow). In the present study, the Mantel tested revealed, for instance, a significant correlation between genetic distance and presence of natural barriers. But after controlling for geographical distance, Mantel test revealed that correlations were not significant. Given that false positive results may occur, the true significance of these results cannot be assured.

4.2. Distribution patterns

The ecological niche-based models (ENM) predicted the existence of suitable areas outside the known distribution of Guinea baboon. According to IUCN (Oates et al., 2008), this primate is confined to an area between southern Mauritania and the north-western Sierra Leone (Oates et al., 2008). However, the results of this study points to a greater area than previously considered, which includes new suitable areas outside the range defined by IUCN to the east of its currently known distribution. These areas comprise three eastern populations: Afollé (Mauritania), Boucle du Baoulé (Mali) and Bissandougou (Guinea-Conakry). All samples collected in these populations were identified as Guinea baboon by mtDNA barcoding. This pattern was also noted by Kopp (2015), who suggested that individuals sampled in Boucle du Baoulé were phenotypically identifiable as Guinea baboons. Consequently, the results of the current study indicate that the IUCN distribution of Guinea baboon should be updated.

The ENMs based in human-related features predicted a relatively fragmented distribution of suitable areas for Guinea Baboon, which is in agreement with previous modelling exercises (Vale et al., 2015a). Strong connections between species presence and water availability (particularly in the Sahel region, where the aridity index is higher than in other regions) have been observed, as well as weak links between species presence and vegetation type or environmental variability, such as precipitation or temperature (Vale et al., 2015a). The present study reinforces that the Guinea baboon is an ecologically plastic species well adapted to distinct environmental conditions within the species range (Vale et al., 2015a) and suggests that ENMs based in human-related features constitute the most realistic approach to model the distribution of suitable areas for this primate species.

4.3. Genetic diversity, population structure and habitat suitability

The genetic and spatial analyses here developed (i.e. STRUCTURE, BAPS, ENMs) suggest that there are major differences between the populations located in the northern, eastern and western extremes of the Guinea baboon distribution. This geographic pattern was also found by Kopp (2015) using a sub-set of the genetic data of this study and using different Bayesian clustering methods (i.e. TESS, DAPC and sPCA). Regardless of the algorithm employed, the results of both studies points to the existence of clear geographic pattern in the genetic structure of Guinea baboon.

In northern areas, the populations in Mauritania displayed the lowest levels of genetic diversity. The Assaba and Afollé populations exhibited observed heterozygosities of 0.478 and 0.542 and percentages of polymorphic *loci* of 83.3% and 88.9%, respectively. In Mauritania, Vale et al. (2015a) showed that the distribution of this primate is strongly dependent on permanent water sources and the species is known to inhabit areas in close proximity to rock pools, locally known as *gueltas* (Brito et al., 2010). Consequently, the probability of its occurrence in Mauritania decreases abruptly whit increasing distance to *gueltas* (Vale et al. 2015a). Considering the general aridity of the lowlands surrounding Assaba and Afollé mountains, is expected that dispersal for a water-dependent species is rare or limited to specific areas in the Sahel that contain permanent water sources (Brito et al., 2010; Vale et al. 2015b). Given that the maintenance of gene flow is related to the ability of Guinea baboons to disperse across fragmented suitable patches, isolation may result in a decline of genetic variation within habitat fragments though the fixation or loss of alleles and accumulation of deleterious mutations (Frankham, 2002; Quéméré et al*.,* 2010; Manel et al*.,* 2013). This interpretation is corroborated by the connectivity models, which predicted that populations in Assaba and Afollé should be relatively isolated, and by the STRUCTURE Bayesian analysis, which separated the individuals' samples in Assaba and Afollé in distinct genetic clusters. High isolation levels between Assaba and Afollé mountains have also been observed in other mountain-restricted species inhabiting Mauritania, such as in Tilapia fishes (Dilyte, 2014), agamid lizards (Gonçalves et al., 2012), crocodiles (Velo-Antón et al., 2014), and gundis (Sousa, 2016). In all these examined taxa, genetic structure was found between Assaba and Afollé mountains, which is most likely explained by the presence of the arid Karakoro river valley, a largely unsuitable area for most taxa that probably restricts gene flow between those mountains. Taken together, the genetic and spatial results suggest limited dispersal of Guinea baboon across arid areas.

The eastern population of Bissandougou in Guinea-Conakry displayed the highest number of private alleles (PA = 14 while other populations varies between 0 and 5). This population falls outside the IUCN Guinea baboon range (Oates et al., 2008) and could potentially be located in the contact zone between the distributions of Guinea and Olive baboons (Zinner et al., 2009; 2011; Kopp, 2015). Although the individuals sampled at Bissandougou were identified as Guinea baboons by mtDNA barcoding, hybrid individuals between the two species could have been sampled. Therefore, the higher number of private alleles could be interpreted as stemming from admixture between Guinea and Olive baboons (Kopp, 2015). Dispersal in the Guinea baboon is thought to be mostly mediated by females, whereas in the olive baboon, dispersal is male-biased (Kopp et al., 2015). Therefore, if female Guinea baboons disperse eastwards into the areas occupied by Olive baboon groups it would be expected the capture of the Guinea baboon mitochondrial genome by Olive baboon populations, while if male Olive baboons disperse westwards into the areas occupied by Guinea baboon groups it would be expected the swamping of nuclear Olive baboon genome (Kopp, 2015. The former hypothesis is supported by the observed clustering in the NJ tree of the individuals sampled in Bissandougou with those identified as Olive baboon using mtDNA barcoding. These results indicate that hybridization between the two species should be explored in the southern and eastern areas of the distribution of Guinea baboon, particularly in the Bissandougou population.

The westernmost costal population in Guinea-Bissau is the most differentiated one and exhibited the lowest levels of population admixture. These results could be explained by the central–marginal hypothesis (CMH), which suggests that marginal populations should exhibit lower genetic diversity, lower gene flow rates, and highest genetic differentiation in comparison to central populations (Eckert et al., 2008; Munwes et al., 2010; Micheletti and Storfer, 2015). Connectivity studies across the entire distribution in another primate species are still scarce. However, the CMH hypothesis has been observed in other animal species inhabiting wide geographical ranges. For instance, Micheletti and Storfer (2015) used a combination of population genetics and landscape tools to test the CMH in the Streamside salamander (*Ambystoma barbouri*), and confirmed a decrease in genetic diversity, effective population size and genetic connectivity from central to peripheral populations across the distribution of the species. Although the genetic differentiation of the western population may be explained by the CMH, the connectivity models obtained for Guinea baboon predicted genetic connectivity between the western and central populations. The disparate genetic and spatial results, suggesting lack of dispersal and landscape connectivity, respectively, may be related

with external relevant factors not included in the spatial analysis. Other environmental factors and/or human-related factors relevant to the dispersal of Guinea baboon may still need to be included in spatial analyses, in order to fully derive accurate measures of landscape connectivity. The obtained results support the differentiation of the western population and suggest that further studies are necessary to realize if this population must be considered as a unit of conservation.

The global population structure pattern found reflects a spatially restricted gene flow for the Guinea baboon. This pattern could be the result of the specie's dispersal behaviour: this primate could move about 40 km per day (Galat-Luong, personal communication in Ferreira da Silva et al., 2013) and occupies large home-ranges (25 km^2, Fickenscher, 2010). For species with a wide geographic range, such as the Guinea baboon, gene flow is expected to occur mostly between neighbouring populations, following a stepping stone model (e.g. Phillipsen et al., 2015). Higher levels of admixture were found in adjacent populations, whereas lower levels of admixture were found in the marginal populations (such as Coastal, Bissandougou and Boucle du Baoulé), suggesting asymmetrical gene flow according to geographical distance and population distribution. In fact, the isolation by distance hypothesis was corroborated by a significant Mantel test. Taken together, the results of the present study suggest that the genetic structure found in Guinea baboon is mostly related with the isolation by distance hypothesis.

4.4. Landscape connectivity

Landscape features that could act as barriers to gene flow did not fully explained the genetic discontinuities found. Rivers have been recognised as exerting the greatest barrier effects to gene flow in primate species, such as *Propithecus tattersalli* (Quéméré et al., 2010) and *Gorilla gorilla gorilla* (Fünfstück et al., 2014). Nevertheless, rivers do not seem to act as a barrier to gene flow in Guinea baboon. This can be explained by the strong seasonality in water level during the dry season in West Africa (Campos et al., 2012) and by the capability of baboons to swim across rivers (Zinner et al., 2011; Ferreira da Silva et al., 2014; Kopp, 2015. In fact, the connectivity models suggested that the Guinea baboons could cross rivers while dispersing, thus assuring the gene flow between populations.

Differences between the environmental and human connectivity models were expected given the potential negative impacts of human activities on functional connectivity of primate species (Imong et al., 2014; Ruiz-Lopez et al., 2015). For

instance, dispersal among some populations of the Cross River Gorillas (*Gorilla gorilla diehli*) in Nigeria-Cameroon region is affected by increasing human disturbance (Imong et al., 2014). Likewise, the landscape genetic study led by Ruiz-Lopez et al. (2015) using the Udzungwa red colobus monkey (*Procolobus gordonorum*) in Tanzania as case-study clearly demonstrated the negative influence of anthropogenic activities in species dispersal abilities. However, no major differences were found between the environmental and human connectivity models in the present study. These results suggested that some human-related features incorporated in the landscape connectivity model may exert positive impacts in the estimates of population connectivity. This is probably the case of croplands. Guinea baboons are highly opportunistic omnivorous species that are able to exploit food resources in a wide variety of habitats and to incorporate a variety of cultivated plant species in their diet such as rice, maize, groundnuts (Oates et al., 2008; Henzi and Barret, 2003). Given the seasonality of natural food resources in West Africa, the agricultural crops surrounding areas of natural habitat can be an opportunistic source of food for this primate (Hill, 2000). At the same time, the crop-raiding behaviour of the Guinea baboon most likely increases the number of negative interactions events with villagers (Hill, 2000; Casanova et al., 2012). Therefore, the existence of croplands in the landscape can have a positive impact in connectivity between populations but only if dispersing animals are not hunted by humans.

4.5. Future research

4.5.1. Developing of a variable Y-linked microsatellite marker

The present study indicated that extensive efforts may be required to find a variable Y-chromosomal marker for Guinea baboon. Of the15 Y-linked microsatellites *loci* tested in Guinea baboon, only the DYS576 was polymorphic for a set of 36 male samples. This result suggests low levels of polymorphism associated with the Y chromosome for the Guinea baboon, which is in concordance with results obtained in other Cercopithecidea species, namely in Mandrillus (*Mandrillus sphinx*, Erler et al., 2004) and the Hamadryas baboon (*Papio hamadryas*, Handley et al., 2006; Städele et al., 2015). In a study led by Erler et al. (2004), 136 human Y-chromosomal microsatellites were analysed in five species of nonhuman primates, including mandrills. However, the authors only identified three polymorphic *loci* for mandrills (DYS472, DYS557 and DYS576) whereas for the others four species, between 17 and 56 alleles were found (in orangutans and central chimpanzees, respectively). Handley et al. (2006) tested seven human-derived Y-linked microsatellites *loci* in a set of 97 male Hamadryas baboons from

Saudi-Arabia and found that DYS576 was the only polymorphic marker. Likewise, Städele et al. (2015) tested a panel of 26 human Y-linked markers in 126 males of Hamadryas baboons from Ethiopia and found the same polymorphic marker (DYS576).

In mammals, the lower level of polymorphism in the Y chromosome in comparison with the other regions of the nuclear genome may be explained by the Y lower effective population size (i.e. Y chromosomes are four times less frequent than an autosomal gene within a population). Consequently, the Y chromosome is more sensitive to genetic drift and more affected by demographic events, such as recent bottlenecks or past demographic expansions (Petit et al., 2002; Handley et al., 2006).

Despite the results of polymorphism in *locus* DYS576 in Guinea baboons, the sequencing of each allele revealed that the variation occurs in flanking areas of the fragment and not as a result of adding or subtracting the repeat motif of the microsatellite. As such, this marker is not suitable to be genotyped through fragment size analyses. However, the obtained result is promising because primers closer to the repeat motif could be designed for *Papio papio* to amplify this locus in the future using Next Generation Sequencing (NGS) tools.

4.5.2. Landscape genetics analyses

Future studies will be required to investigate the genetic diversity and genetic connectivity in The Gambia. The ENM based on human-related features suggested a complete isolation of populations in The Gambia. Given that the ENMs predict how habitat quality can affect genetic diversity and functional connectivity (Micheletti and Storfer, 2015), a lack of connectivity between The Gambia and the surrounding populations is expected. However, lack of genetic data prevented testing if this population is functionally connected to others or not. Samples collected in The Gambia were initially included in the analyses but they were excluded due to non-amplification. A likely explanation of lack of amplification could be the storage of the faecal samples in RNAlater® for a long period (i.e. 12 months, C. Barlow pers. communication) before DNA extraction. Soto-Calderón et al. (2009) tested the effect of three storage media (silica, RNAlater® and ethanol) on mitochondrial and nuclear marker amplification success using faecal DNA samples from different African ungulates. The authors found that samples preserved in RNAlater® over longer storage periods yielded lower amplification success for nuclear DNA when compared with other preservation methods. Consequently, efforts to generate genetic data from The Gambia are required to obtain information about the connectivity of Guinea baboon in its entire distribution.

In the present study, landscape and human-related features does not completely explain functional connectivity of the Guinea baboon. A possible explanation could be related to the wide environmental variation across the specie's distribution area, covering at least three biogeographical regions: Sahel, Savannah and Afrotropical (Olson et al., 1998). The important factors associated with dispersal may vary across the distinct regions because individuals may differ in their responses to particular environmental conditions due local adaptation (Zeller et al., 2012). At the same time, these regional differences may be undetected when analyses are performed across the whole distribution area (Quemere et al., 2010) As such, Phillipsen et al. (2015) suggested that the spatial scale at which landscape genetics analysis are performed should be adjusted according the dispersal ability of the study model. In the future, the modelling of suitable areas for Guinea baboons and the subsequent connectivity analyses should be performed separately for each of the three biogeographic regions where species is present.

5. Concluding remarks

The present study is the first one using Guinea baboon as study model to investigate the functional connectivity in a heterogeneous landscape. In general, this study provided new information about: (1) the distribution area of Guinea baboons, (2) the potential hybridization zone between Guinea and Olive baboons, (3) the genetic diversity and functional connectivity between Guinea baboon populations; (4) the isolation mechanism (distance or resistance) that best explained the observed genetic differentiation in this primate; and (5) a variable Y-linked microsatellite marker for the Guinea baboon. The main conclusions of this study are:

(1) Ecological models predicted the existence of suitable areas outside the known distribution of Guinea baboon, indicating that the IUCN distribution area should be updated;
(2) The south-eastern distribution of the Guinea baboon (particularly the Bissandougou population) likely represents the hybridization zone between Guinea and Olive baboons, but further studies are needed to confirm such hypothesis;
(3) Isolation by distance hypothesis is the most probable explanation for the genetic differentiation between the Guinea baboon populations;
(4) Gene flow among Guinea baboon populations does not appear to be strongly limited by landscape resistance. These results emphasise the ability of Guinea baboons to adapt to human-fragmented landscape and to live alongside of human populations;
(5) Variation on the Y chromosome seems to be low for the Guinea baboon but a variable Y-linked microsatellite *loci* was identified. After designing more specific primers for the Guinea baboon, this finding could represent an opportunity to investigate the patterns of male-specific dispersal in the future.

In conclusion, the present study achieves important results about the functional connectivity of Guinea baboon. These results will allow determining which habitat patches should be the priority focus of conservation in future wildlife management works.

6. References

Aguiar, L. M. S., Bernard, E., Ribeiro, V., Machado, R. B., & Jones, G. (2016). Should I stay or should I go? Climate change effects on the future of Neotropical savannah bats. *Global Ecology and Conservation*, *5*, 22–33. http://doi.org/10.1016/j.gecco.2015.11.011.

Alvarado-Serrano, D. F., & Knowles, L. L. (2014). Ecological niche models in phylogeographic studies: Applications, advances and precautions. *Molecular Ecology Resources*, *14*(2), 233–248. http://doi.org/10.1111/1755-0998.12184.

Arroyo-Rodríguez, V., & Fahrig, L. (2014). Why is a landscape perspective important in studies of primates? *American Journal of Primatology*, *76*(10), 901–909. http://doi.org/10.1002/ajp.22282.

Baguette, M., & Van Dyck, H. (2007). Landscape connectivity and animal behavior: Functional grain as a key determinant for dispersal. *Landscape Ecology*, *22*(8), 1117–1129. http://doi.org/10.1007/s10980-007-9108-4.

Balkenhol, N., & Fortin, M. J. Basics of Study Design: Sampling Landscape Heterogeneity and Genetic Variation for Landscape Genetic Studies. *Landscape Genetics: Concepts, Methods, Applications*, 58-76.

Bergl, R. A., & Vigilant, L. (2007). Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (*Gorilla gorilla diehli*). *Molecular Ecology*, *16*(3), 501–516. http://doi.org/10.1111/j.1365-294X.2006.03159.x.

Bicheron, P., Defourny, P., Brockmann, C., Schouten, L., Vancutsem, C., Huc, M., Bontemps, S., Leroy, M., Achard, F., Herold, M., Ranera, F., & Arino, O. (2008). GLOBCOVER, products description and validation report. Medias-France and Postel, Toulouse, France. http://postel.obs-mip.fr/?-land-cover- 68- [accessed 9 May 2012].

Blair, M.E. and Melnick, D.J. (2012). Genetic evidence for dispersal by both sexes in the Central American squirrel monkey, *Saimiri oerstedii citrinellus*. *American Journal*

*of Primatol*ogy 74:37–47.

Broquet, T., & Petit, E. (2004). Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology*, *13*(11), 3601–3608. http://doi.org/10.1111/j.1365-294X.2004.02352.x.

Campos, J.C., Sillero, N., Brito, J.C. (2012). Normalized Difference Water Indexes have dissimilar performances in detecting seasonal and permanent water in the Sahara-Sahel transition zone. *Journal of Hydrology*, 464-465: 438-446.

Casanova, C., Sousa, C., & Costa, S. (2014). Are animals and forests forever? Perceptions of wildlife at Cantanhez Forest National Park, Guinea-Bissau Republic. Memória - Special issue in anthropology and environment. *Sociedade de Geografia de Lisboa, Lisboa*, 40.

Castillo, J.A., Epps C.W., Davis, A.R., Cushman, S.A. (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology*, 23, 843–856.

CGIAR (2016). Global Aridity and PET Database (CGIAR-CSI). http://www.cgiar-csi.org/data/global-aridity-and-pet-database.

CIESIN-ITOS (2013). Global Roads Open Access Data Set, Version 1 (gROADSv1). Center for International Earth Science Information Network - CIESIN - Columbia University, and Information Technology Outreach Services - ITOS - University of Georgia. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4VD6WCT.

Cushman, S., Wasserman T., Landguth E., Shirk A. (2013) Re-evaluating causal modelling with mantel tests in landscape genetics. Diversity, 5,51–72.

Dexter, C. E. A., Appleby, R. G. A., Edgar, J. P. B., Scott, J. A., & Jones, D. N. A. (2016). Using complementary remote detection methods for retrofitted eco-passages : a case study for monitoring individual koalas in south-east Queensland. *Wildlife Research*, 43(5), 369-379.

Dieringer, D., & Schlötterer, C. (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, *3*(1), 167-169.

Dilyte, J. (2014). Population structure and gene flow in desert environment: an application of molecular tools to isolated fish populations in West Africa. MSc thesis. University of Porto.

Dyer, R. (2016). Applied Population Genetics. http://popgen.bio.vcu.edu/ecological-distance.html

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188. http://doi.org/10.1111/j.1365-294X.2007.03659.x.

EEA (2011). Annual report 2010 and Environmental statement 2011. ISSN 1561-2120

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, *5*(6), 435–445. http://doi.org/10.1038/nrg1348.

Eriksson, J., Siedel, H., Lukas, D., Kayser, M., Erler, A., Hashimoto, C., Vigilant, L. (2006). Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (*Pan paniscus*). *Molecular Ecology*, *15*(4), 939–949. http://doi.org/10.1111/j.1365-294X.2006.02845.x.

Erler, A., Stoneking, M., & Kayser, M. (2004). Development of Y-chromosomal microsatellite markers for nonhuman primates. *Molecular Ecology*, *13*(10), 2921–2930. http://doi.org/10.1111/j.1365-294X.2004.02304.x.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. http://doi.org/10.1111/j.1365-294X.2005.02553.x.

Fahrig L. (2003). Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology and Evolution* 34: 487–515.

Fernando, P., Vidya, T. N. C., Rajapakse, C., Dangolla, A., & Melnick, D. J. (2003). Reliable noninvasive genotyping: Fantasy or reality? *Journal of Heredity*, *94*(2), 115–123. http://doi.org/10.1093/jhered/esg022.

Ferreira da Silva, M.J. (2012). Hunting pressure and the population genetic patterns and sex-mediated dispersal in the Guinea baboon in Guinea-Bissau. PhD thesis, University of Cardiff.

Ferreira da Silva, M. J., Casanova, C., & Godinho, R. (2013). On the western fringe of baboon distribution: mitochondrial D-loop diversity of Guinea Baboons (*Papio papio* Desmarest, 1820) (Primates: Cercopithecidae) in Coastal Guinea-Bissau, western Africa. *Journal of Threatened Taxa*, *5*(10), 4441–4450. http://doi.org/10.11609.

Ferreira da Silva, M. J., Godinho, R., Casanova, C., Minhós, T., Sá, R., & Bruford, M. W. (2014). Assessing the impact of hunting pressure on population structure of Guinea baboons (*Papio papio*) in Guinea-Bissau. *Conservation Genetics*. http://doi.org/10.1007/s10592-014-0621-0.

Fickenscher, G. H. (2010) Genetic population structure of free-ranging Guinea baboons (*Papio papio*) in the Niokolo-Koba National Park, Senegal. PhD thesis. Gottingen University.

Fünfstück, T., Arandjelovic, M., Morgan, D. B., Sanz, C., Breuer, T., Stokes, E. J., Vigilant, L. (2014). The genetic population structure of wild western lowland gorillas (*Gorilla gorilla gorilla*) living in continuous rain forest. *American Journal of Primatology*, *76*(9), 868–878. http://doi.org/10.1002/ajp.22274.

Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, *139*(1), 463–471.

Gonçalves, D. V., Brito, J. C., Crochet, P. A., Geniez, P., Padial, J. M., & Harris, D. J. (2012). Phylogeny of North African Agama lizards (Reptilia: *Agamidae*) and the role of the Sahara Desert in vertebrate speciation. *Molecular Phylogenetics and*

Evolution, 64(3), 582-591.

Hill, C. M. (2000). Conflict of interest between people and baboons: Crop raiding in Uganda. *International Journal of Primatology*, *21*(2), 299–315. http://doi.org/10.1023/A:1005481605637.

Höglund, J. (2009). Evolutionary conservation genetics. *Oxford Biology CN - QH456 .H65 2009*, 189. http://doi.org/10.1093/acprof:oso/9780199214211.001.0001.

Hutchinson, G. E. (1957). Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology, 22, 415–427.

Imong, I., Robbins, M. M., Mundry, R., Bergl, R., & Kühl, H. S. (2014). Informing conservation management about structural versus functional connectivity: A case-study of Cross River gorillas. *American Journal of Primatology*, *76*(10), 978–988. http://doi.org/10.1002/ajp.22287.

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, *21*(4), 498–507. http://doi.org/10.1111/j.1466-8238.2011.00683.x.

Jin, L., & Chakraborty, R. (1994). Estimation of genetic distance and coefficient of gene diversity from single-probe multilocus DNA fingerprinting data. *Molecular Biology and Evolution*, 11(1), 120-127.

Kopp, G. H., Ferreira da Silva, M. J., Fischer, J., Brito, J. C., Regnaut, S., Roos, C., & Zinner, D. (2014). The Influence of Social Systems on Patterns of Mitochondrial DNA Variation in Baboons. *International Journal of Primatology*, *35*(1), 210–225. http://doi.org/10.1007/s10764-013-9725-5.

Kopp, G. (2015). Gene Flow Dynamics in Baboons – The Influence of Social Systems. PhD thesis, University of Göttingen.

Kopp, G. H., Fischer, J., Patzelt, A., Roos, C., & Zinner, D. (2015). Population genetic insights into the social organization of Guinea baboons (*Papio papio*): Evidence

for female-biased dispersal. *American Journal of Primatology*, *77*(8), 878–889. http://doi.org/10.1002/ajp.22415.

Lawson Handley, L. J., Hammond, R. L., Emaresi, G., Reber, a, & Perrin, N. (2006). Low Y chromosome variation in Saudi-Arabian hamadryas baboons (*Papio hamadryas hamadryas*). *Heredity*, *96*(4), 298–303. http://doi.org/10.1038/sj.hdy.6800803.

Manel, S., & Holderegger, R. (2013). 10 Years of Landscape Genetics. *Trends in Ecology and Evolution*, *28*(10), 614–621.

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, *18*(4), 189–197. http://doi.org/10.1016/S0169-5347(03)00008-9.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer research* 27:209–220.

McRae, B. H. (2006). Isolation By Resistance. *Evolution*, *60*(8), 1551–1561. http://doi.org/10.1111/j.0014-3820.2006.tb00500.x.

Micheletti, S. J., & Storfer, A. (2015). A test of the central-marginal hypothesis using population genetics and ecological niche modelling in an endemic salamander (*Ambystoma barbouri*). *Molecular Ecology*, *24*(5), 967–979. http://doi.org/10.1111/mec.13083.

Minhós, T., Nixon, E., Sousa, C., Vicente, L. M., Ferreira da Silva, M. F., Sá, R., & Bruford, M. W. (2013b). Genetic evidence for spatio-temporal changes in the dispersal patterns of two sympatric African colobine monkeys. *American Journal of Physical Anthropology*, 150(3), 464–474. http://doi.org/10.1002/ajpa.22223.

Minhós, T., Wallace, E., da Silva, M. J. F., Sá, R. M., Carmo, M., Barata, A., & Bruford, M. W. (2013a). DNA identification of primate bushmeat from urban markets in Guinea-Bissau and its implications for conservation. *Biological conservation*, 167, 43-49.

Miquel, C., Bellemain, E., Poillot, C., Bessière, J., Durand, A., & Taberlet, P. (2006). Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. *Molecular Ecology Notes*, *6*(4), 985–988. http://doi.org/10.1111/j.1471-8286.2006.01413.x.

Munwes, I., Geffen, E., Roll, U., Friedmann, A., Daya, A., Tikochinski, Y., & Gafny, S. (2010). The change in genetic diversity down the core-edge gradient in the eastern spadefoot toad (*Pelobates syriacus*). *Molecular Ecology*, *19*(13), 2675–2689. http://doi.org/10.1111/j.1365-294X.2010.04712.x.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. http://doi.org/10.1038/35002501.

NGA (2016). National Geospatial-Intelligence Agency GEOnet Names Server (GNS). http://geonames.nga.mil/gns/html.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, *51*(11), 933. http://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.

Patzelt, A., Kopp, G. H., Ndao, I., Kalbitzer, U., Zinner, D., & Fischer, J. (2014). Male tolerance and male-male bonds in a multilevel primate society. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(41), 14740–14745. http://doi.org/10.1073/pnas.1405811111.

Petit, E., Balloux, F. F., & Excoffier, L. (2002). Mammalian population genetics: why not Y? *Trends in Ecology & Evolution*, *17*(1), 28–33.

Phillipsen, I. C., Kirk, E. H., Bogan, M. T., Mims, M. C., Olden, J. D., & Lytle, D. A. (2015). Dispersal ability and habitat requirements determine landscape-level genetic patterns in desert aquatic insects. *Molecular Ecology*, *24*(1), 54–69. http://doi.org/10.1111/mec.13003.

Prugnolle, F., & de Meeus, T. (2002). Inferring sex-biased dispersal from population

genetic tools: a review. *Heredity*, *88*, 161–165. http://doi.org/10.1038/sj/hdy/6800060.

Quéméré, E., Crouau-Roy, B., Rabarivola, C., Louis, E. E., & Chikhi, L. (2010a). Landscape genetics of an endangered lemur (*Propithecus tattersalli*) within its entire fragmented range. *Molecular Ecology*, *19*(8), 1606–1621. http://doi.org/10.1111/j.1365-294X.2010.04581.x.

Quéméré, E., Louis, E. E., Ribéron, A., Chikhi, L., & Crouau-Roy, B. (2010b). Non-invasive conservation genetics of the critically endangered golden-crowned sifaka (*Propithecus tattersalli*): High diversity and significant genetic differentiation over a small range. *Conservation Genetics*, *11*(3), 675–687. http://doi.org/10.1007/s10592-009-9837-9.

Rissler, L. J. (2016). Union of phylogeography and landscape genetics. *Proceedings of the National Academy of Sciences*, *113*(29), 201601073. http://doi.org/10.1073/pnas.1601073113.

Ruiz-Lopez, M. J., Barelli, C., Rovero, F., Hodges, K., Roos, C., Peterman, W. E., & Ting, N. (2016). A novel landscape genetic approach demonstrates the effects of human disturbance on the Udzungwa red colobus monkey (*Procolobus gordonorum*). *Heredity*, *116*(2), 167-176.

Russo, I.-R. M., Sole, C. L., Barbato, M., von Bramann, U., Bruford, M. W., Jones, K. E., Roger, J. H. (2016). Landscape determinants of fine-scale genetic structure of a small rodent in a heterogeneous landscape (Hluhluwe-iMfolozi Park, South Africa). *Scientific Reports*, 6, 29168. http://doi.org/10.1038/srep29168.

Schlotterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, *109*(6), 365–371. http://doi.org/10.1007/s004120000089.

Shah, V. B., & Supercomputing, I. (2008). Circuitscape : A Tool for Landscape Ecology, *Proceedings of the 7th Python in Science Conference*, 62–65.

Shirk, A. J., Cushman, S. A., & Landguth, E. L. (2012). Simulating pattern-process relationships to validate landscape genetic models. *International Journal of*

Ecology, *2012*. http://doi.org/10.1155/2012/539109.

Soto-Calderón, I. D., Ntie, S., Mickala, P., Maisels, F., Wickings, E. J., & Anthony, N. M. (2009). Effects of storage type and time on DNA amplification success in tropical ungulate faeces: technical advances. *Molecular Ecology Resources*, *9*(2), 471–479. http://doi.org/10.1111/j.1755-0998.2008.02462.x.

Sousa F.A.V. (2015). Biogeography and phylogenetic position of a Sahara-Sahel mountain endemic, Felovia vae (*Ctenodactylidae*). MSc thesis. University of Porto.

Städele, V., Van Doren, V., Pines, M., Swedell, L., & Vigilant, L. (2015). Fine-scale genetic assessment of sex-specific dispersal patterns in a multilevel primate society. *Journal of Human Evolution*, *78*, 103–113. http://doi.org/10.1016/j.jhevol.2014.10.019.

Stevens, V. M., Verkenne, C., Vandewoestijne, S., Wesselingh, R. A., & Baguette, M. (2006). Gene flow and functional connectivity in the natterjack toad. *Molecular Ecology*, *15*(9), 2333–2344. http://doi.org/10.1111/j.1365-294X.2006.02936.x.

Struhsaker, T.T. (2010). The red colobus monkeys: variation in demography, behavior and ecology of endangered species. Oxford: Oxford University Press.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic acids research*, 24(16), 3189-3194.

Taylor, A. C., Walker, F. M., Goldingay, R. L., Ball, T., & van der Ree, R. (2011). Degree of landscape fragmentation influences genetic isolation among populations of a gliding mammal. *PLoS ONE*, *6*(10), 1–9. http://doi.org/10.1371/journal.pone.0026651.

Vale, C. G., Campos, J. C., Silva, T. L., Gonçalves, D. V., Sow, A. S., Martínez-Freiría, F., Brito, J. C. (2016). Biogeography and conservation of mammals from the West Sahara-Sahel: an application of ecological niche-based models and GIS. *Hystrix, the Italian Journal of Mammalogy*, *27*(1). http://doi.org/10.4404/HYSTRIX-27.1-

11659.

Vale, C. G., Ferreira da Silva, M. J., Campos, J. C., Torres, J., & Brito, J. C. (2015a). Applying species distribution modelling to the conservation of an ecologically plastic species (*Papio papio*) across biogeographic regions in West Africa. *Journal for Nature Conservation*, *27*, 26–36. http://doi.org/10.1016/j.jnc.2015.06.004.

Vale, C. G., Pimm, S. L., & Brito, J. C. (2015b). Overlooked mountain rock pools in deserts are critical local hotspots of biodiversity. *PLoS ONE*, *10*(2). http://doi.org/10.1371/journal.pone.0118367.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535–538. http://doi.org/10.1111/j.1471-8286.2004.00684.x.

Velo-Antón, G., Parra, J. L., Parra-Olea, G., & Zamudio, K. R. (2013). Tracking climate change in a dispersal-limited species: Reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology*, *22*(12), 3261–3278. http://doi.org/10.1111/mec.12310.

Velo-Antón, G., Godinho R., Campos J.C., Brito J.C. (2014). Should I stay or should I go? Dispersal and population structure in small, isolated desert populations of West African crocodiles. *PLoS One*, 9(4), e94626.

Vigilant, L. (2011) Protocol for isolation of DNA from silica & RNA later preserved faeces. Vigilant Lab

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*(2), 335–342. http://doi.org/10.1890/10-1171.1.

Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology*, *27*(6), 777–797. http://doi.org/10.1007/s10980-012-9737-0.

Zinner, D., Groeneveld, L. F., Keller, C., & Roos, C. (2009). Mitochondrial phylogeography of baboons (*Papio* spp.): indication for introgressive hybridization? *BMC Evolutionary Biology*, *9*, 83. http://doi.org/10.1186/1471-2148-9-83.

Zinner, D., Buba, U., Nash, S., & Roos, C. (2011). Pan-African voyagers: the phylogeography of baboons. In Primates of Gashaka (pp. 319-358). *Springer New York*.

Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L. F., & Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. *American Journal of Physical Anthropology*, *150*(1), 133–140. http://doi.org/10.1002/ajpa.22185.

7. Supplementary material

Table S1. Details of the Multiplexes PCRs used by Ferreira da Silva et al. (2014): annealing temperature (AT), primers sequences, repeat motif, allele range size for Guinea baboon, fluorescent dye, and final PCR concentration.

Multiplex	AT (ºC)	Locus	Forward (5' - 3') / Reverse (5' - 3')	Repeat motif	Range	Dye	Conc. (µM)
M1	57ºC	D7s503	ATGACTTGGAGTAATGGG AACCTTTAATCAGGATACAGAC	CA	142-156	NED	0.6
		D12s375	TTGTTGAGGGTCTTTCTCCA	GATA	164-184	PET	0.1
		D3s1766	TCTTCTTATTTGGAAAAGTAACCC ACCACATGAGCCAATTCTGT	ATCT	192-208	FAM	0.1
		D13s765	ACCCAATTATGGTGTTGTTACC TGTAACTTACTTCAAATGGCTCA TTGAAACTTACAGACAGCTTG	GATA	200-212	NED and TAMRA	0.2
M2	50°C	D14s306	TGTAACTTACTTCAAATGGCTCA TTGAAACTTACAGACAGCTTG	GATA	161-181	FAM	0.2
		D1s533	CATCCCCCCCCAAAAAATA TA	GATA	187-203	HEX	0.4
		D2s1326	TTGCTAATCAAATAACAATGGG AGACAGTCAAGAATAACTGCCC CTGTGGCTCAAAAGCTGAAT	CTAT	192-208	FAM	0.3
M3	59°C	D10s611	CATACAGGAAACTGTGTAGTGC CTGTATTTATGTGTGTGGATGG	GATA	129-137	FAM	0.1
		D8s1106	TTGTTTACCCCTGCATCACT TTCTCAGAATTGCTCATAGTGC	GATA	149-161	VIC	0.1
		D6s501	GCTGGAAACTGATAAGGGCT GCCACCCTGGCTAAGTTACT	CTAT	171-187	FAM	0.5
M4	57°C	D5s1457	TAGGTTCTGGGCATGTCTGT TGCTTGGCACACTTCAGG	GATA	125-137	PET	0.1
		D3s1768	GGTTGCTGCCAAAGATTAGA CACTGTGATTTGCTGTTGGA	GATA	193-212	VIC	0.1
		D7s2204	TCATGACAAAACAGAAATTAAGTG AGTAAATGGAATTGCTTGTTACC	AGAT	230-250	FAM	0.4
SINGLEPLEX	58°C	D4s243	TCAGTCTCTCTTTCTCCTTGCA TAGGAGCCTGTGGTCCTGTT	GATA	152-172	FAM	0.2

Table S2. Allelic dropout and false allele's rates per locus estimated by Pedant Software for the 14 microsatellites *loci* amplified in this study.

Multiplex	**Locus**	**Allelic dropout**	**False alleles**
M1	D6s264	0.028	0.000
	D7s503	0.062	0.000
	D12s375	0.026	0.000
	D13s765	0.144	0.000
	D3s1766	0.069	0.000
M3	D10s611	0.214	0.000
	D17s791	0.071	0.013
	D6s311	0.179	0.000
M4	D5s1457	0.099	0.000
	D8s505	0.072	0.026
	D10s1432	0.152	0.035
	D3s1768	0.130	0.015
	D7s2204	0.107	0.012
M5	D21s1442	0.282	0.000

Table S3. Description of the ecogeographical variables (EGVs) used in the present study to model the Guinea baboon suitable areas and respective reference source.

EGVs	Description	Reference
Slope	Slope	USGS 2006
D01_CROP	Croplands	Bicheron et al. 2008
D02_CRVE	Mosaic cropland (50-70%) /vegetation (20-50%)	Bicheron et al. 2008
D03_VECR	Mosaic vegetation (50-70%) / cropland (20-50%)	Bicheron et al. 2008
D07_OPBD	Open (15-40%) broadleaved deciduous forest/woodland (>5m)	Bicheron et al. 2008
D09_COSH	Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)	Bicheron et al. 2008
D10_COHE	Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)	Bicheron et al. 2008
D15_ROCK	Consolidated bare areas (hardpans, gravels, bare rock, stones, boulders)	Bicheron et al. 2008
D16_SAND	Non-consolidated bare areas (sandy desert)	Bicheron et al. 2008
NDVI	Normalized Difference Vegetation Index	GIMMS 2008
PET	Potencial Evapotranspiration	CGIAR 2016
Bio 5	Temperature of warmest month	Wordclim
Bio 13	Precipitation of wettest month	Wordclim
d. rivers	Distance to rivers	Digititezed from Michelin Africa North & West Map 741
d. gueltas	Distance to gueltas	Vale et al. 2015b
d. roads	Distance to roads	CIESIN-ITOS 2013
d. accessibility	Distance to accessibility	NGA 2016
d. toponomies	Distance to villages	NGA 2016

Table S4. Correlation matrix between the 18 EGVs used in the present study. Variables highly correlated (Pearson correlation index > 0.7) are showed in bold. d07 = D. open broadleaved deciduous forest/woodland, d09 = D. closed to open shrubland, d10 = D. closed to open herbaceous cover, d15 = D. rocky desert, d16 = D. sandy desert, NDVI = Normalized Difference Vegetation Index, TWM = Temperature of warmest month, PWM = Precipitation of wettest month, PET = Potential Evapotranspiration, d1 = D. croplands, d2 = D. mosaic cropland/vegetation, d3 = D. mosaic vegetation/cropland.

	Slope	D07_OPBD	D09_COSH	D10_COHE	D14_BARE	D15_ROCK	D. rivers	D. gueltas	NDVI	TWM	PWM	PET	D. Roads	D. Toponomies	D. accessibility	D01_CROP	D02_CRVE	D03_VECR
Slope	1																	
D07_OPBD	-0.197	1																
D09_COSH	-0.150	**0.916**	1															
D10_COHE	0.393	-0.587	-0.476	1														
D15_ROCK	0.327	-0.689	-0.574	0.694	1													
D16_SAND	0.281	**-0.722**	-0.614	0.636	**0.974**	1												
D. rivers	-0.183	**0.842**	**0.758**	-0.508	-0.612	-0.639	1											
D. gueltas	0.240	-0.678	-0.583	0.564	**0.909**	**0.947**	-0.615	1										
NDVI	0.273	**-0.893**	**-0.813**	0.695	**0.824**	**0.835**	**-0.728**	**0.751**	1									
TWM	-0.329	0.670	0.554	-0.521	**-0.752**	**-0.777**	0.618	**-0.847**	-0.668	1								
PWM	0.301	-0.670	-0.574	**0.755**	**0.834**	**0.811**	-0.588	**0.764**	**0.746**	-0.655	1							
PET	-0.044	-0.613	-0.636	0.201	0.281	0.361	-0.494	0.276	0.547	-0.076	0.166	1						
D. Roads	-0.091	0.639	0.687	-0.230	-0.344	-0.377	0.599	-0.367	-0.491	0.409	-0.354	-0.418	1					
D. Toponomies	-0.127	**0.725**	**0.765**	-0.318	-0.441	-0.480	0.651	-0.425	-0.634	0.457	-0.453	-0.512	0.684	1				
D. accessibility	-0.175	0.656	0.539	-0.300	-0.439	-0.451	**0.806**	-0.469	-0.483	0.559	-0.456	-0.278	0.507	0.494	1			
D01_CROP	0.271	-0.008	0.122	0.382	0.558	0.522	-0.027	0.535	0.207	-0.407	0.443	-0.219	0.160	0.202	0.005	1		
D02_CRVE	-0.089	**0.795**	**0.891**	-0.320	-0.312	-0.370	0.693	-0.316	-0.619	0.356	-0.374	-0.636	0.685	**0.775**	0.507	0.390	1	
D03_VECR	-0.120	0.599	**0.701**	-0.162	-0.283	-0.302	0.536	-0.250	-0.464	0.342	-0.359	-0.385	0.699	**0.813**	0.506	0.294	**0.789**	1

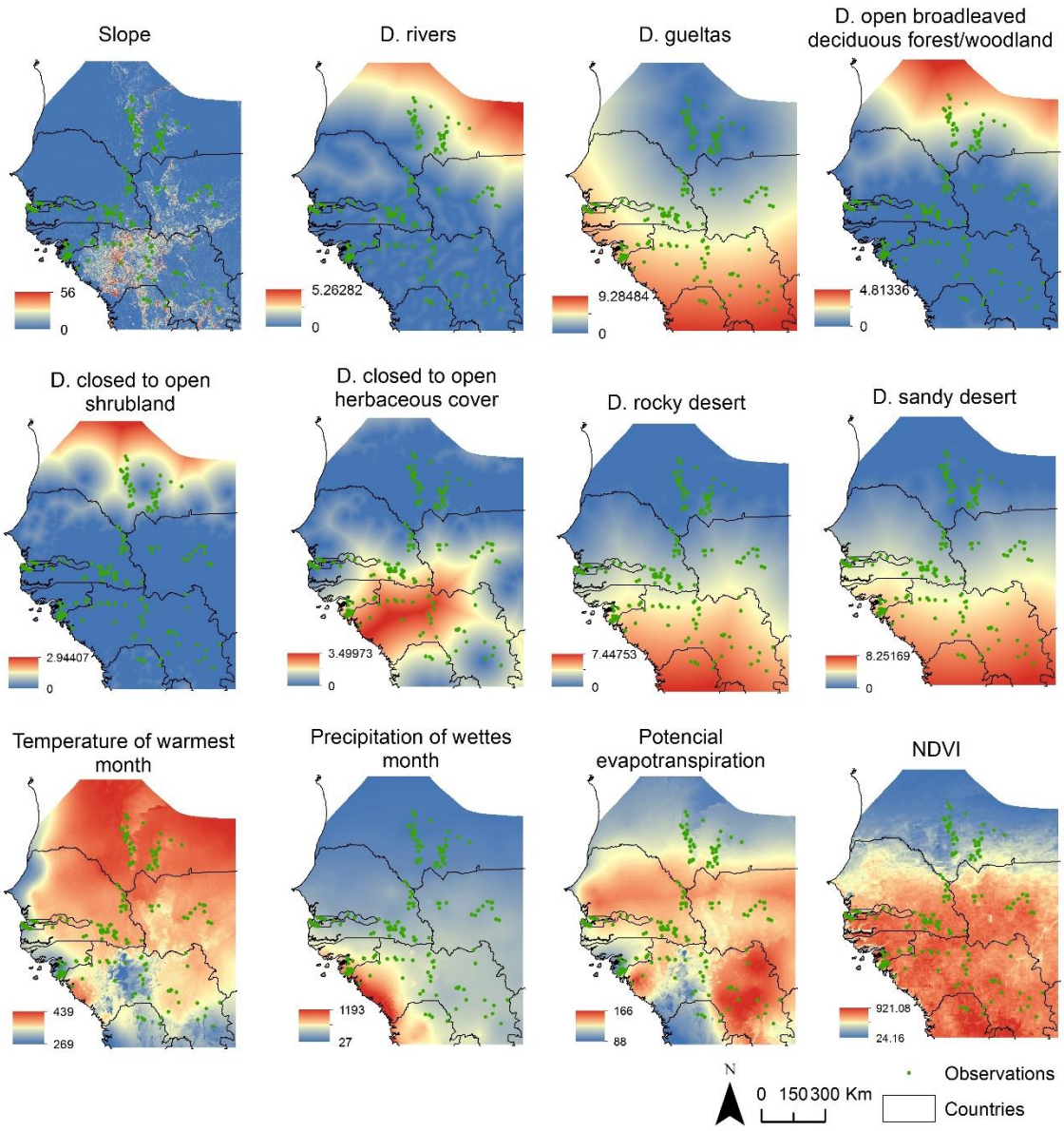


Figure S1. Spatial representation of the Environmental ecogeographical variables used in this study to model the location of the Guinea baboon suitable areas.

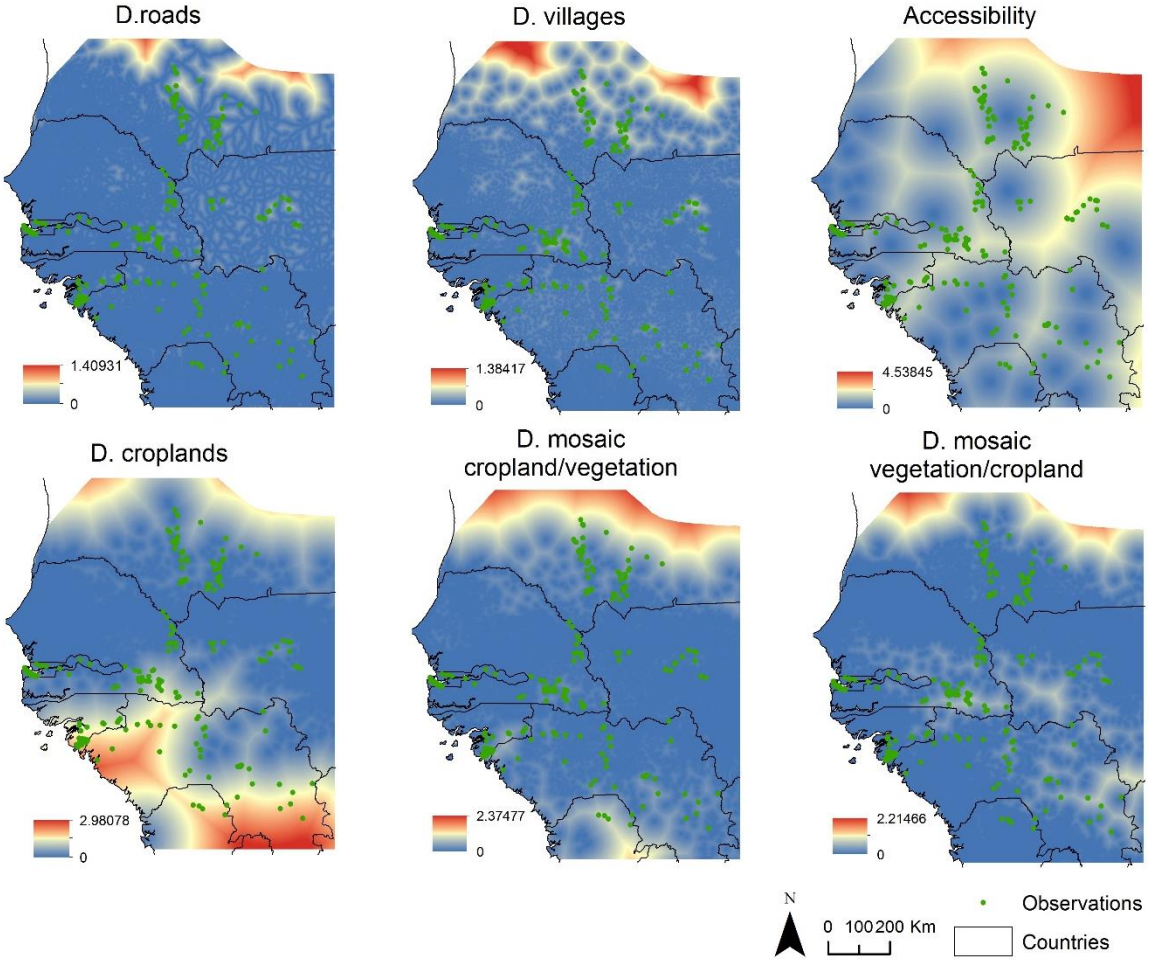


Figure S2. Spatial representation of the Human ecogeographical variables used in this study to modelling the Guinea baboon suitable areas.

Table S5. Description of eigenvalues and eigenvectors obtained for the Environmental PC used in the present study to modelling the Guinea baboon suitable areas.

PC Layer	1	2	3	4	5	6	7	8	9	10	11	12
Eigenvalues	5.640	1.278	0.659	0.469	0.430	0.227	0.159	0.091	0.063	0.039	0.028	0.010
% of eigenvalues	62.028	14.058	7.251	5.153	4.732	2.494	1.746	1.002	0.696	0.424	0.304	0.111
Eigenvectors												
PET	0.174	0.551	0.094	0.579	0.028	-0.224	0.276	0.276	0.298	-0.015	0.152	0.085
bio 5	-0.294	0.227	0.166	0.288	-0.449	-0.316	-0.395	-0.134	-0.372	0.048	-0.367	-0.018
bio 13	0.313	-0.187	-0.032	-0.193	-0.380	-0.246	-0.496	0.346	0.504	0.032	0.045	0.036
D07_OPBD	-0.331	-0.258	-0.075	0.106	-0.082	-0.262	0.064	0.138	-0.165	0.564	0.602	-0.008
D09_COSH	-0.299	-0.336	-0.098	0.060	-0.137	-0.416	0.420	-0.357	0.458	-0.131	-0.246	0.037
D10_COHE	0.272	-0.185	0.273	-0.005	-0.661	0.269	0.495	0.145	-0.195	0.013	-0.020	-0.058
D15_ROCK	0.333	-0.218	-0.169	0.207	-0.049	-0.190	-0.088	-0.255	-0.301	-0.413	0.322	0.543
D16_SAND	0.338	-0.158	-0.226	0.232	0.062	-0.226	-0.028	-0.104	-0.154	-0.150	0.074	-0.799
d. Gueltas	0.323	-0.183	-0.313	0.077	0.219	-0.193	0.149	0.272	-0.223	0.426	-0.542	0.228
d. Rivers	-0.226	-0.398	-0.144	0.624	-0.001	0.515	-0.207	0.174	0.163	-0.075	-0.101	0.017
NDVI	0.343	0.092	0.045	0.153	-0.075	0.230	-0.126	-0.662	0.230	0.530	0.028	0.038
Slope	0.122	-0.351	0.820	0.123	0.363	-0.180	-0.078	0.020	-0.009	0.019	-0.049	-0.003

Table S6. Description of eigenvalues and eigenvectors obtained for the Human PC used in the present study to modelling the Guinea baboon suitable areas.

PC Layer	1	2	3	4	5	6
Eigenvalues	2.806	0.780	0.409	0.262	0.155	0.135
% of eigenvalues	61.711	17.149	8.999	5.769	3.399	2.974
Eigenvectors						
D01_CROP	-0.180	-0.879	-0.363	-0.119	0.169	0.144
D02_CRVE	-0.472	-0.137	0.043	0.153	-0.828	-0.218
D03_VECR	-0.473	-0.030	0.198	0.258	0.502	-0.645
d.accessability	-0.343	0.430	-0.828	0.050	0.075	0.056
d.roads	-0.434	0.141	0.212	-0.859	0.056	0.077
d.toponomies	-0.463	0.060	0.309	0.395	0.156	0.711

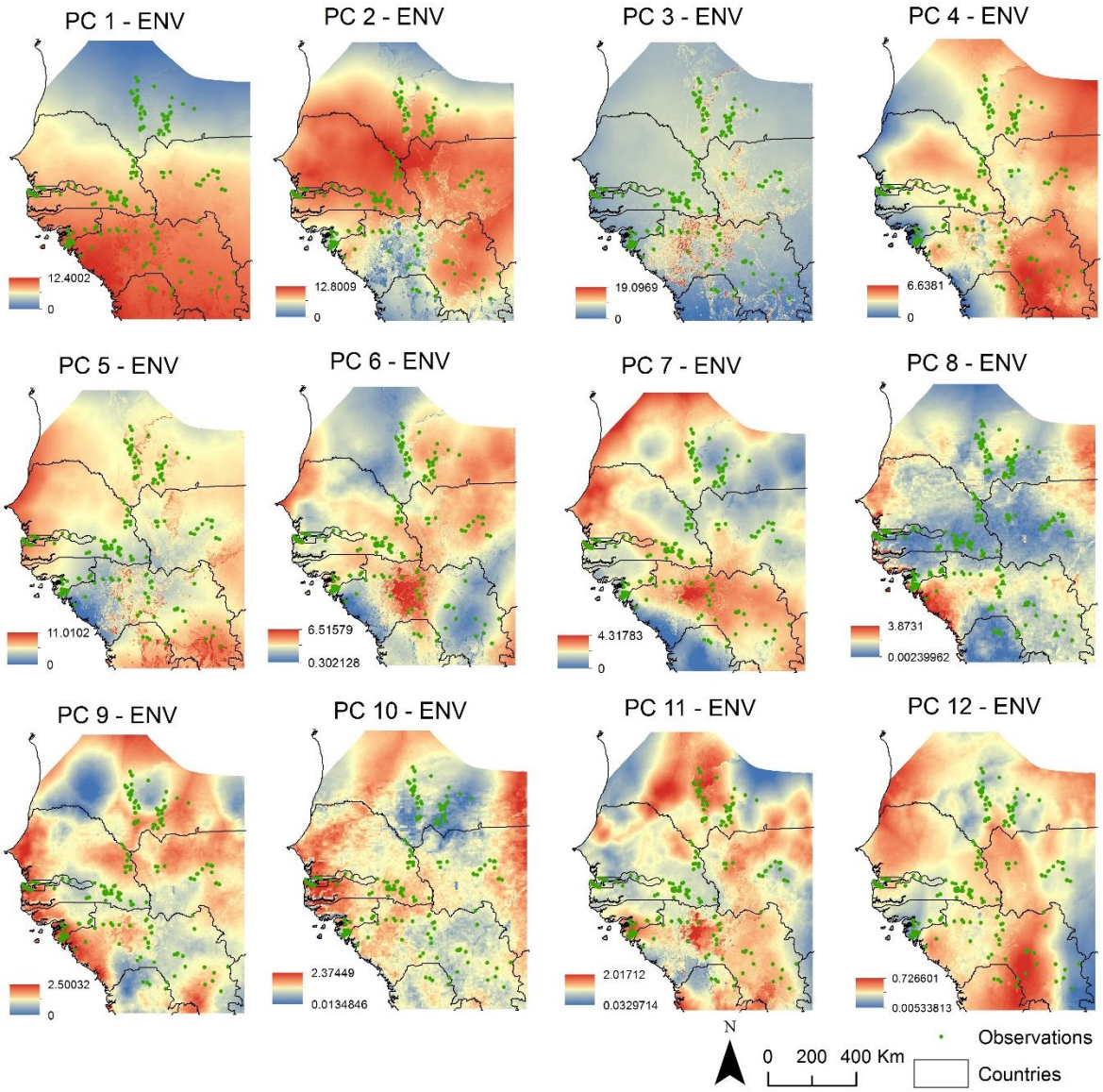


Figure S3. Spatial representation of the 12 environmental PC layers used to model Guinea baboon distribution.

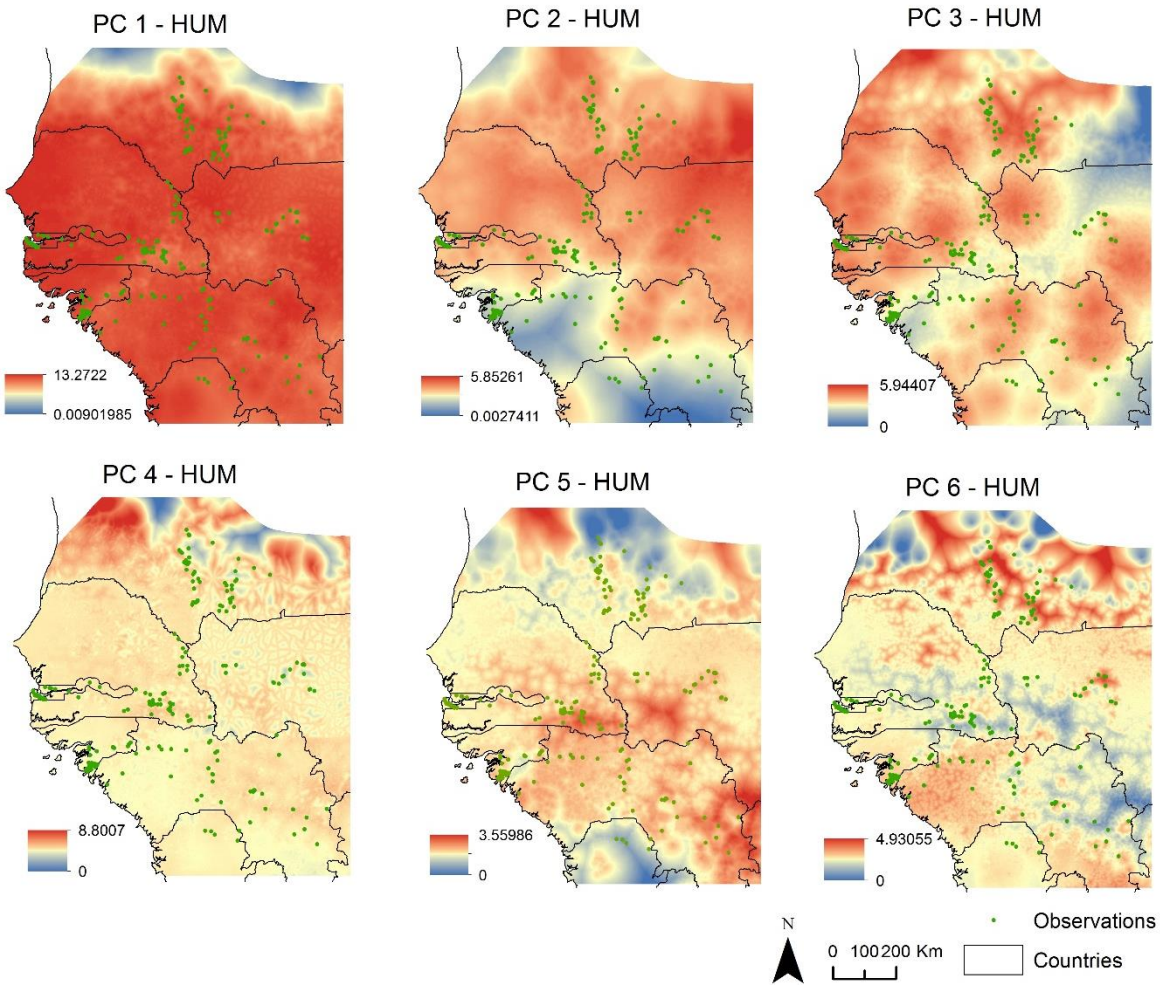
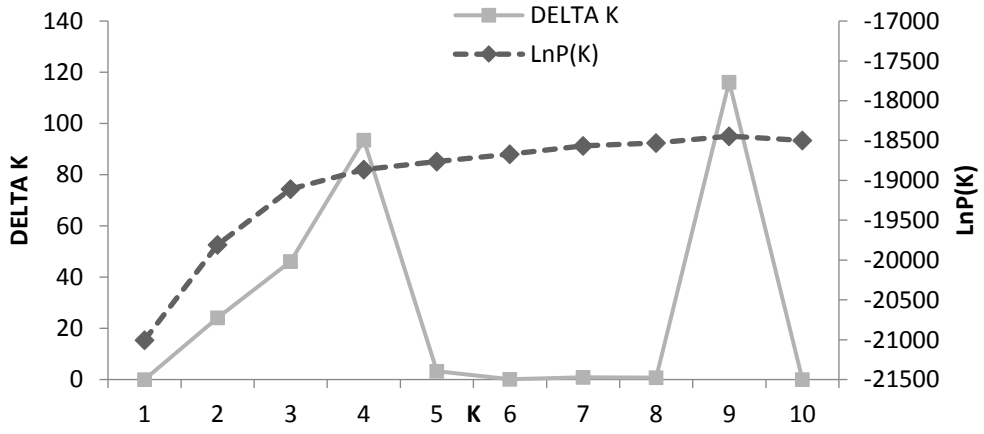


Figure S4. Spatial representation of the six Human PC layers used to model Guinea baboon distribution

Table S7. Nei's estimation of heterozygosity for the 23 microsatellites used in this study. Ho = observed heterozygosity, Hs = heterozygosity within populations, Dst = heterozygosity between populations, Ht = total heterozygosity (Hs + Dst).

Locus	Ho	Hs	Dst	Ht
D6s264	0.525	0.582	0.095	0.677
D7s503	0.702	0.721	0.122	0.843
D12s37	0.681	0.713	0.086	0.799
D3s176	0.266	0.303	0.166	0.469
D13s76	0.502	0.497	0.022	0.519
D14s30	0.558	0.63	0.049	0.679
D1s533	0.521	0.527	0.019	0.546
D2s132	0.57	0.617	0.017	0.635
D10s61	0.429	0.558	0.075	0.634
D8s110	0.589	0.598	0.066	0.663
D17s79	0.64	0.628	0.061	0.69
D6s501	0.652	0.693	0.028	0.721
D6s311	0.493	0.533	0.153	0.685
D5s145	0.556	0.551	0.061	0.611
D8s505	0.217	0.211	0	0.211
D10s14	0.556	0.671	0.009	0.68
D5s820	0.803	0.753	0.052	0.805
D3s176	0.608	0.696	0.059	0.755
D7s220	0.653	0.705	0.066	0.77
D1s207	0.394	0.386	0.039	0.425
D4s243	0.664	0.727	0.029	0.756
D1s548	0.744	0.71	0.084	0.795
D21s11	0.655	0.796	0.016	0.812
Overall	0.564	0.6	0.06	0.66

A



B

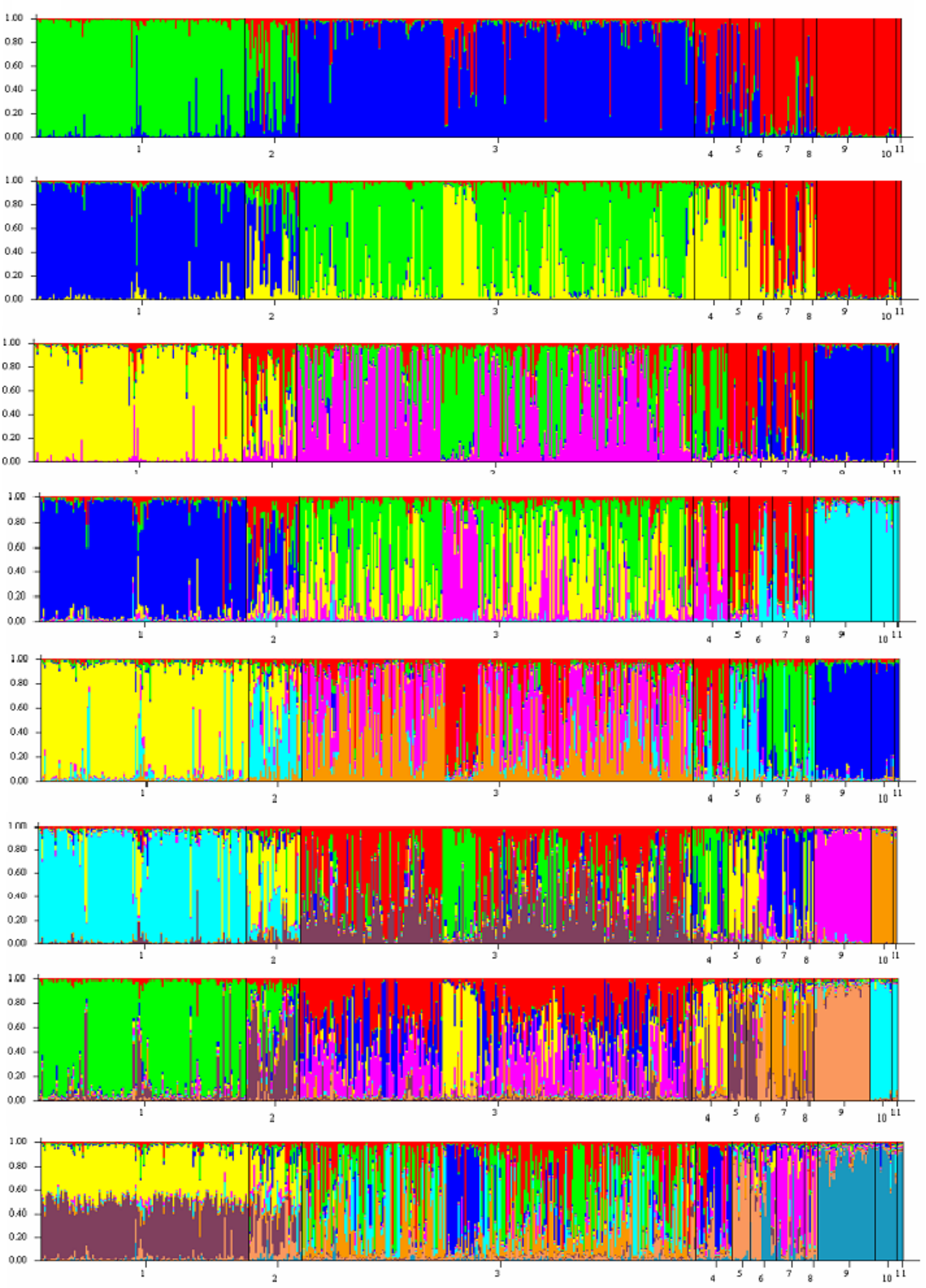


Figure S5. Results of the Bayesian STRUCTURE analysis using the 507 individuals of *P. papio*, two individuals of *P. anubis* and one potential hybrid. A) Inference of the most probable number of clusters (K) using Delta K and mean LnP(K) values across all runs suggest the existence of 9 genetic clusters. B) Bayesian admixture analysis assuming clustering of K= 3 to K = 10 (from top to bottom). Figure shows membership coefficient (Q) for an individual belong to one cluster. Each color represents one different genetic cluster and a single vertical bar represents each individual. The number ranging between 1-10 represents de ten populations of *P.papio* and the population 11 the *P.anubis*.

Table S8. Genetic diversity measures for each Guinea baboon population (n = 507 individuals) genotyped for 6 - 23 microsatellite loci. N = number of samples per population, Na = Number of Different Alleles, Ne = Number of effective alleles, Ho = Observed Heterozygosity, He = Expected Heterozygosity, F_{IS} = inbreeding coefficient, %P = Percentage of

Pop	N	Na	Ne	Ho	He	F	%P	PA	HWE
Coastal	123	4.615	2.502	0.571	0.570	-0.002	100.00%	0	0
Fouta Djalon W	32	4.538	2.732	0.578	0.618	0.073	100.00%	1	0
Niokolo-Koba	233	5.615	2.666	0.602	0.581	-0.037	100.00%	5	1***
Senegal River	21	4.462	2.867	0.583	0.588	0.011	100.00%	3	1***
Fouta Djalon S	11	3.923	2.772	0.572	0.572	-0.015	100.00%	0	0
Fouta Djalon N	15	4.769	3.090	0.552	0.624	0.144	100.00%	1	0
Assaba	8	3.692	2.527	0.506	0.530	0.038	92.31%	1	0
Afollé	17	4.308	2.744	0.541	0.554	0.021	92.31%	4	0
Boucle du Baoulé	34	5.462	3.032	0.612	0.637	0.043	100.00%	3	0
Bissandougou	13	4.231	2.853	0.546	0.621	0.112	100.00%	14	0

Polymorphic Loci, PA = Private alleles, HWE = departures from HEW, p<0.05, ** p<0.01 or *** p<0.001.

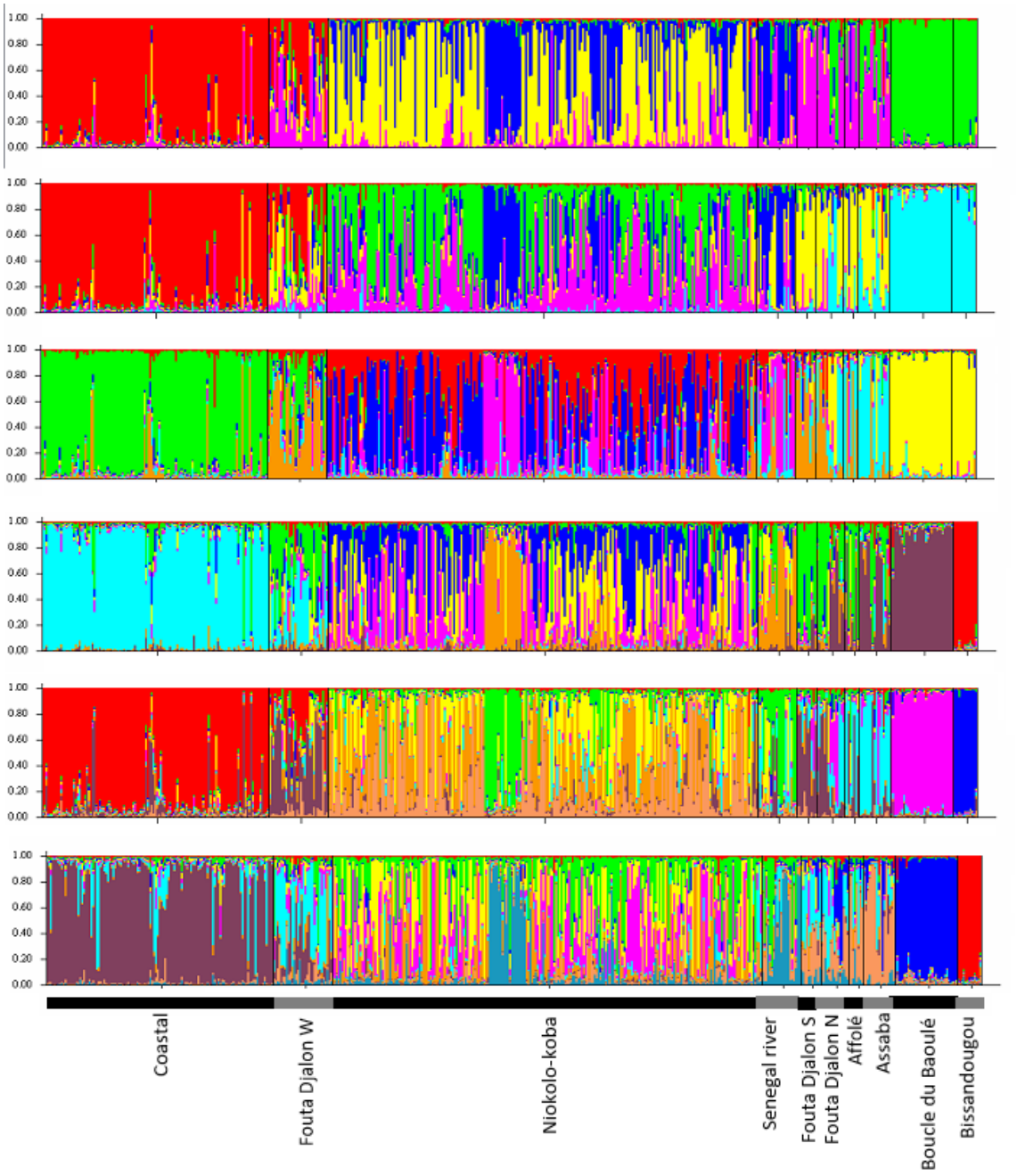


Figure S6. Bayesian admixture analysis of *P. papio* assuming clustering of K=5 to K=10 (from top to bottom). Figure shows membership coefficient (Q) for an individual belong to one cluster. Each color represents one different genetic cluster and a single vertical bar represents each individual. The name of populations where the individuals were sampled is indicated below**.**

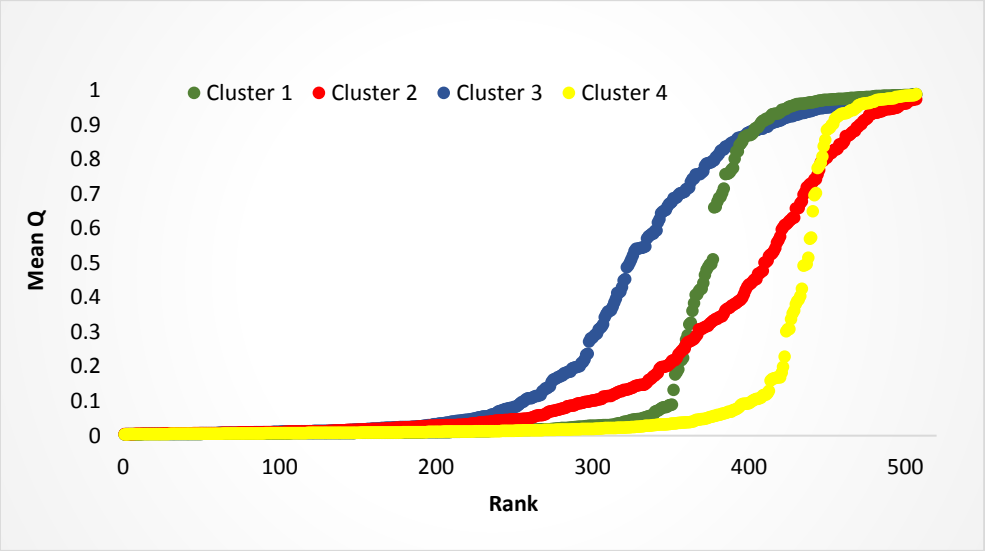


Figure S7. Ranked mean membership coefficient (Q) across runs for each cluster considering K=4. Each color represents the four different clusters obtained from the software STRUCTURE.

Table S9. Pairwise F_{ST} values between genetic cluster uncovered by STRUCTURE (N = 366 individuals; Q>0.80). Significant values indicated with * p<0.05, ** p<0.01 or *** p<0.001.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	0			
Cluster 2	0.189***	0		
Cluster 3	0.179***	0.045***	0	
Cluster 4	0.206***	0.068***	0.098***	0

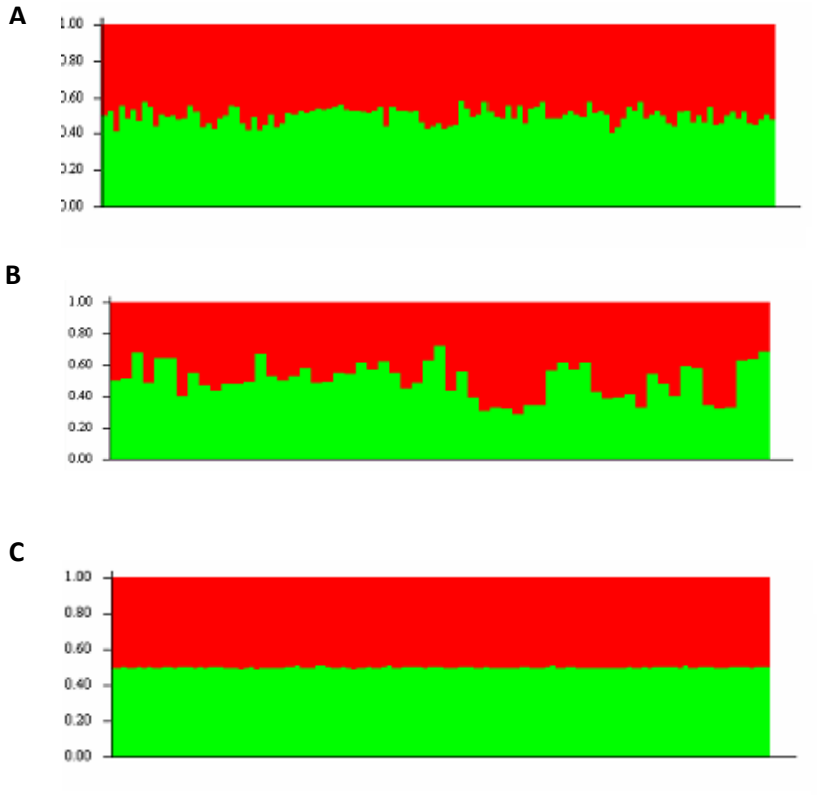


Figure S8. There is no evidence of existence of sub-structure within the clusters 1, 2 and 3. Bayesian admixture analysis performed only with individuals assigned to A) cluster 1 (N=117); B) cluster 2 (N=59); C) cluster 3 (N=129), considering K = 2 solution.

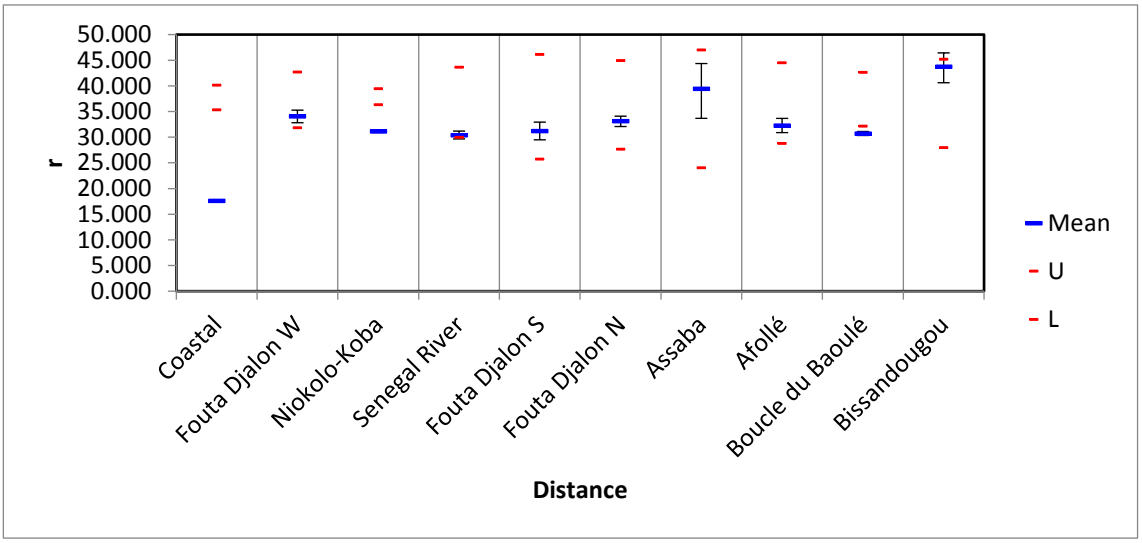


Figure S9. Observed average relatedness of Guinea baboon individuals within populations are lower than the average relatedness expected by chance. U = upper error and L = lower error.

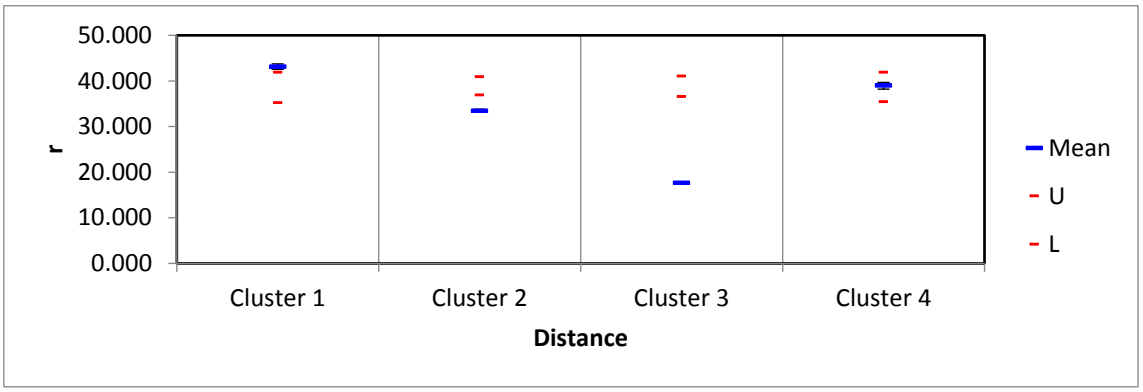


Figure S10. Observed average relatedness of Guinea baboon individuals within clusters are lower than the average relatedness expected by chance, except for the cluster 1. U = upper error and L = lower error.

Table S10. Relative contributions of the 12 environmental PC layers (pc1 – pc12) used to build the "Environmental model" using MAXENT. Values shown are averages over replicate runs.

Variable	% contribution	Permutation importance
pc3	28.8	21
pc1	17.4	24
pc9	10.5	5.5
pc12	8.5	13.5
pc7	7.1	9.1
pc10	6.6	8.8
pc2	6.2	5.8
pc4	5.4	2.2
pc11	3.3	3.7
pc5	3	3.7
pc6	2.7	2.2
pc8	0.5	0.5

Table S11. Relative contributions of the 12 environmental PC layers (pc1 – pc12) plus the 6 Human PC layers (pch1 – pch6) used to build the "Human model" using MAXENT. Values shown are averages over replicate runs.

Variable	% contribution	Permutation importance
pc1	13.7	13.8
pch1	13.4	19.5
pc3	12.9	9.2
pch2	10.8	10
pc9	8.7	6.4
pc12	6.1	5.6
pc2	5.6	6.9
pc7	4.6	6
pc4	4.5	3.3
pc10	3.6	2.6
pc5	3.1	1.1
pch5	2.6	4.8
pch3	2.6	2.6
pc6	2.5	2
pch6	2.4	2.7
pc11	2.4	1.6
pch4	0.3	0.9
pc8	0.2	0.7

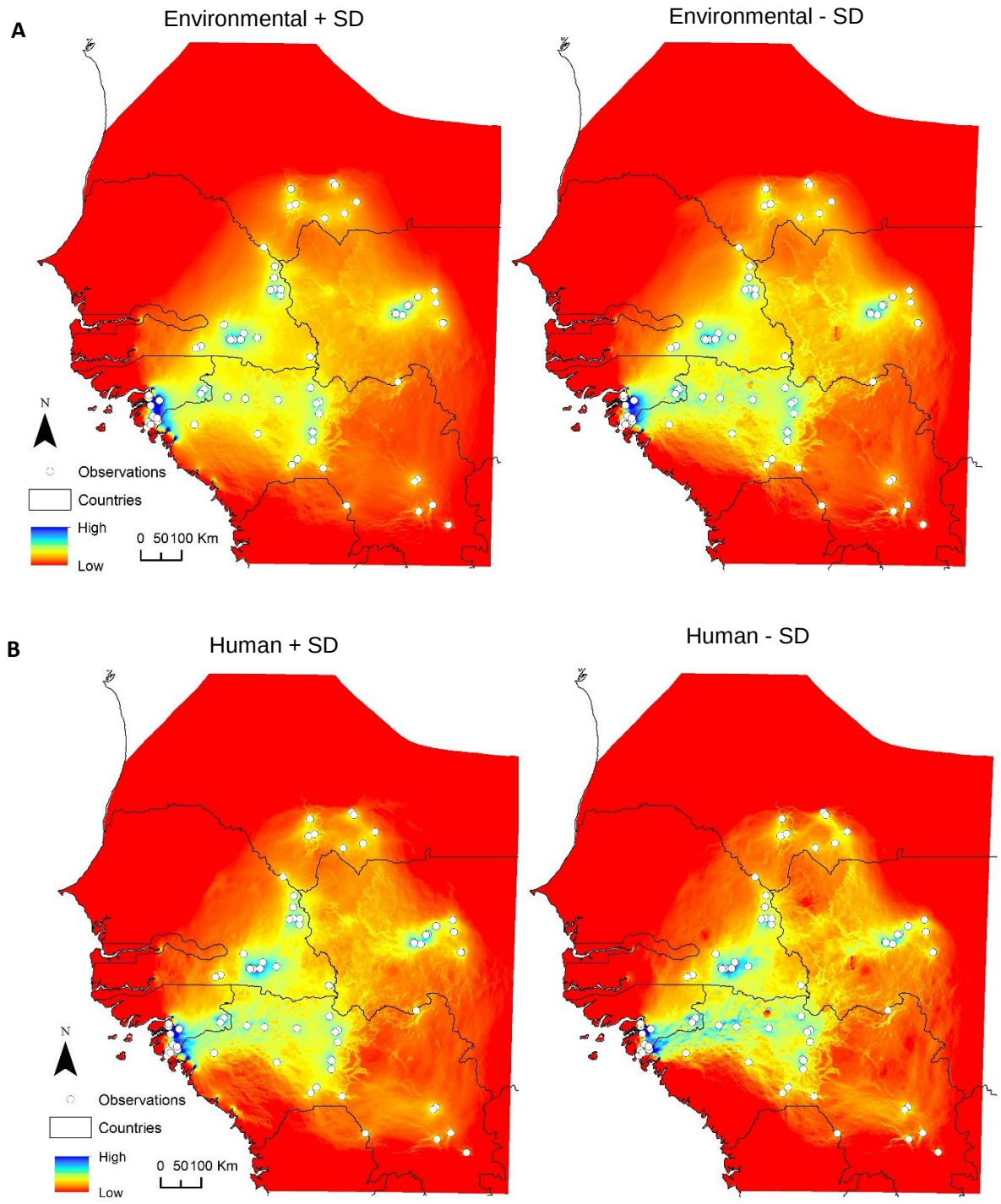


Figure S11. Current maps output from CIRCUITSCAPE based on A) environmental + SD (left) and environmental - SD (left); B) human + SD (left) and human - SD (left). Results showing cumulative resistance between populations of Guinea baboon based on the combined effect of environmental variables. High stands for the highest current flow