Spring 5-15-2017

Population Genomics of a Baboon Hybrid Zone in Zambia

Kenneth Lyu Chiou
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Anthropology

Dissertation Examination Committee:

Jane Phillips-Conroy, Chair
Amy Bauernfeind
Clifford Jolly
Allan Larson
Amanda Melin

Population Genomics of a Baboon Hybrid Zone in Zambia

By

Kenneth L. Chiou

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2017
St. Louis, Missouri

Contents

List of Figures

List of Tables

Acknowledgments

This work would not be possible without assistance from my numerous colleagues, collaborators, friends, and family. The foundations of this research, as well as much of the data analyzed herein, are rooted in prior work in Zambia by Jane Phillips-Conroy, Cliff Jolly, Jeff Rogers, Andy Burrell, and Anna Weyher. These individuals introduced me to Zambia and inspired me to drive kilometers upon kilometers in pursuit of Zambian baboons. I wish to acknowledge in addition the early support and companionship of Christina Bergey, Zach Johnson, Linous Munsimbwe, Neeta Zambara, and Geoff Ndaradzi during our baboon trapping expeditions. I am grateful to Ngawo Namukonde for her hospitality and company during my pilot field season.

I thank my committee: Jane Phillips-Conroy, Amanda Melin, Amy Bauernfeind, Allan Larson, and Cliff Jolly. Thank you also to Bob Sussman and Jim Cheverud for their invaluable feedback early on as members of my proposal committee.

Thank you to Jane Phillips-Conroy, my advisor, for introducing me to baboons, for reading everything I write, and for encouraging me to believe in myself. I am a better academic because of her. Thank you also to Cliff Jolly, who is in all but title a second advisor to me. His presence in the field and lab always makes work more fun, and much of how I think about biological processes and evolution I attribute to him.

Three of my teachers made indelible marks on me early in my career. Anand Dacier was the first to steer me into a life of primates, Tab Rasmussen ignited my excitement about primate macroevolution, and Bob Sussman challenged me to think broadly and creatively about big systems. They are all sorely missed, and I would not be where I am without them.

I am grateful to the Zambia Wildlife Authority (now known as the Department of National

Parks & Wildlife) and the University of Zambia Department of Veterinary Medicine for granting permission for my fieldwork in Zambia. I thank in particular Griffin Shanungu, David Squarre, Kampamba Kombe, James Milanzi, Vincent Nyirenda, Kennedy Choongo, and King Nalubamba for their assistance with permits. Thank you additionally to ZAWA for sharing GIS data on my study site.

I would never have reached baboons in the field the first time, or each time thereafter, without assistance from Paul Barnes, Sophie Henley, Lackson Tembo, Rose Mwanza, Alfred Mumba, and the rest of the Pioneer Camp staff. Thank you also to Frank Willems and Jaco van Staden for help and advice with vehicle ownership and maintenance in the bush. I thank Clive Chifunte, Leevan Polomondo, Matthews Mumbi, Abel Musukwa, and Phanwell Moonga for logistical support, and Ali Zyuulu and Kelvin Ngalaba for their friendship and unfailing helpfulness during my stay in Ngoma. Thanks also to Cliff Jolly and Megan Petersdorf for their brief but memorable visit in the early days of my fieldwork and for their cheering company. I thank my primary scouts, Marthias Simayumbula and Talison Tembo, for their assistance and company and for everything that they do as wildlife police officers, as well as my field assistants, Lloyd Ngalaba, Peter Kafwabwe, and Maurice Mundia, for assistance with trapping. Thanks also to Shepherd Phiri for veterinary assistance during trapping. Finally, an enormous thank you to Andrea Porro for his friendship, hospitality, and generosity over the later stages of my fieldwork, and to Laura Sommariva, Gift Kabwe, Chiara Scatena, Mattia Colombini, Stephen Banda, and the rest of the Konkamoya Lodge staff for their support and company.

The Molecular Primatology Lab at New York University has been a second home for me for a decade and counting. I thank Todd Disotell for inviting me to perform my dissertation labwork there, for always offering his support, and for fostering the fun and collaborative atmosphere that keeps us coming back. I cannot overstate my gratitude to Andy Burrell and Christina Bergey for their help and mentorship throughout our decade of knowing each other, for being my daily sounding boards when labwork was not going so well, and for getting lunch with me every day. I thank Andy for sharing his trove of knowledge on labwork and baboons and for sharing his dataset of baboon occurrences. I thank Christina for introducing me to programming many years

ago, for all of our nerdy conversations since, and for getting me hooked on the Honest Teas that actually powered my research. I am grateful to other members of the lab through the years for their friendship and advice, especially Tony Di Fiore, Mary Blair, Luca Pozzi, Jason Hodgson, Alba Morales Jimenez, Maryjka Blaszczyk, Chris Schmitt, Lina Maria Valencia, and Megan Petersdorf.

I thank again Christina Bergey for her assistance with analysis and for contributing her ddRAD-seq bioinformatic pipeline that was indispensable for this project.

Thank you to Tony Di Fiore for his mentorship and friendship extending back to my sophomore year of college. Since then, he has been a constant mentor and collaborator in the classroom, lab, field, cloud, and beyond. I also thank Tom Igoe, my other collaborator on the Ethoinformatics project for his infectious passion for monkeys and electronics, as well as his all-around geekery. He and Tony continue to broaden the ways in which I think about data. Thank you also to Robyn Overstreet, Mike Chevett, Jonathan Cousins, Nick Sears, Nien Lam, and Sebastian Buys for their work on the project, to Jennifer Moore, Cynthia Hudson, Aaron Addison, and Chris Freeland for their support and contributions, and to all members of the Ethoinformatics Working Group for productive and stimulating discussions during our working group meetings in St. Louis and Austin.

I wish to thank the colleagues and friends that I made during my seven years at Washington University. I thank Joe Orkin, Mike Montague, and Jessica Joganic for their friendship and for our chats about genetics. A special thanks also to Marissa Milstein, Ashley van Batavia, John Willman, Chris Shaffer, Stephanie Musgrave, Crystal Koenig, Effie Robakis, Steve Goldstein, Mike Storozum, Natalie Mueller, and Natalia Guzman for their friendship.

A very special thank you to Anna Weyher for her friendship and for our numerous conversations about work, science, conservation, Africa, baboons, Game of Thrones, and everything else. Working in both Zambia and St. Louis was easier and more enjoyable because of her. I also thank Kelsey Sue for her love and companionship.

I would be nowhere without my unbelievable family. I thank Lingwei Chiou, Wei-Shing Chiou, Katie Chiou, Kevin Chiou, and Pal for their love and sacrifices. I thank Charlene Lyu and Alex Lau for their hospitality in my NYC days, and Charlene additionally for help with moving and decorating. Finally, I thank my best friend, Lisa Picascia, who has been there through it all and

tells everything like it is. I love you all.

This dissertation was generously funded by the National Science Foundation (BCS 1341018, SMA 1338524) and the Leakey Foundation. During my graduate studies, I also received a Graduate Research Fellowship from the National Science Foundation (DGE 1143954) and a dissertation writing fellowship from Washington University. Preliminary research was funded by Washington University graduate summer research funds and a National Science Foundation graduate research travel award. Portions of this dissertation work were also funded by the National Science Foundation (BCS 1260816, BCS 1029302) the Wenner-Gren Foundation, the National Geographic Society, and the NYU University Research Challenge Fund. The Genome Technology Center at NYU is supported by NIH/NCATS UL1 TR00038 and NIH/NCI P30 CA016087.

Kenneth L. Chiou

Washington University in St. Louis
May 2017

I dedicate this dissertation to my parents, Lingwei and Wei-Shing.

ABSTRACT OF THE DISSERTATION

Population Genomics of a Baboon Hybrid Zone in Zambia

by

Kenneth L. Chiou

Doctor of Philosophy in Anthropology

Anthropology

Washington University in St. Louis 2017

Professor Jane Phillips-Conroy, Chair

Hybridization is increasingly recognized as a common, important process shaping the evolution of organisms including humans. Across hybrid zones, the genomes of incipient species are mixed and recombined through hybridization and backcrossing, creating conditions ideal for evaluating the actions of natural selection on gene variants in novel genomic contexts. This dissertation aims to increase our understanding of hybridization using a Zambian baboon study system in which two species, Kinda baboons (*Papio kindae*) and grayfoot baboons (*Papio griseipes*), hybridize despite exhibiting pronounced differences in body size and behavior. Using genome-wide genotypic data prepared using double-digest RADseq, I scan for genomic regions under selection in these species and in their hybrids.

Because a large section of the hybrid zone contains groups unhabituated to human presence, I develop a new method for noninvasive genomic-scale genotyping from feces. I demonstrate that an enrichment procedure using methyl-CpG-binding-domain proteins to preferentially capture densely CpG-methylated mammalian DNA effectively partitions baboon host DNA from contaminating bacterial DNA, yielding efficient sequencing of target genomic DNA. Comparisons of same-animal double-digest RADseq libraries demonstrate high concordance between feces-derived and blood-derived genotypes.

By scanning genome-wide data for regions with significant levels of differentiation between Kinda and grayfoot baboon populations, I identify candidate genes under selection in the two species. I find evidence for selection on genes and biological pathways that underlie differences in body size between the parental species. One pathway exhibiting significantly elevated differentiation was the

JAK/STAT signaling pathway, which notably serves an important role in mediating the effects of cytokine signals on processes including epiphyseal chondrocyte proliferation essential for bone growth.

Analysis of hybrids reveals that Kinda and grayfoot baboons form a relatively wide cline in the Kafue river valley in central Zambia. Comparison of autosomal ancestry patterns to mitochondrial-DNA and Y-chromosome ancestry patterns reveals that the Kinda baboon Y chromosome has introgressed extensively across the species barrier relative to both the mitochondrial genome and the remainder of the nuclear genome. The JAK/STAT signaling pathway exhibits restricted introgression, suggesting a role in barriers to reproduction possibly due to the unusually high or low body size sexual dimorphism between male grayfoot $\times$ female Kinda and male Kinda $\times$ female grayfoot baboon mating partners. The toll-like receptor pathway exhibits enhanced introgression, suggesting adaptive introgression of pathogen defenses. Finally, the sperm tail gene *ODF2* exhibits enhanced introgression and an advantage of the grayfoot baboon variant. I suggest based on a house mouse analogy that male hybrids may be subjected to reduced sperm quality but that this effect may be mitigated or overcome by the presence of an invading Y chromosome. This effect potentially explains the extreme introgression of the Kinda baboon Y chromosome.

Chapter 1

Hybrid Zones and *Papio*: A Review

Introduction

Hybridization is increasingly recognized as a common occurrence in animals (Seehausen, 2004). In primates, the signatures of reticulate evolution can be detected in a taxonomically diverse range of genomes (Arnold & Meyer, 2006), including humans (Green et al., 2010; Reich et al., 2010; Mondal et al., 2016), and numerous cases of natural hybridization are recognized in primates (Brockelman & Gittins, 1984; Phillips-Conroy & Jolly, 1986; Bynum et al., 1997; Detwiler, 2002; Wyner et al., 2002; Kanthaswamy et al., 2008; Tung et al., 2008; Burrell et al., 2009; Cortés-Ortiz et al., 2014; Fuzessy et al., 2014), with additional cases of hybridization following anthropogenic interference (Kawamoto et al., 1999, 2001; Ohsawa et al., 2005). Hybridization's significance to primate evolution, however, remains relatively unknown (Zinner et al., 2011).

For this dissertation, I apply a genomic approach to understanding the structure of genetic variation across the Kafue National Park baboon contact zone in Zambia. As a biological anthropologist, my ultimate goal is to use baboons (genus *Papio*) as a model for understanding the impacts of historic introgression and other processes on the human lineage. Baboons are ideal models for this because of their sociality, analogous intelligence (Cheney & Seyfarth, 2008), ecological flexibility, and propensity for hybridization (Jolly, 2001). Using hybridization as a natural experiment (Harrison, 1990), I aim to disentangle the roles of different processes in shaping genetic variation. In particular, I examine the roles of local adaptation and social organization, as well as introgressive hybridization, in structuring variation within species, between species, and in interspecific hybrids.

Hybridization

Hybridization in nature allows for the study of long-term processes that are inaccessible with laboratory crosses yet central for understanding speciation (Harrison, 1990; Abbott et al., 2013). Empirical evidence supports a growing consensus that natural hybridization is a common and important evolutionary force not only in plants, but also in many animals (e.g., Phillips et al., 2004; Macholán et al., 2007; Godinho et al., 2011; Carneiro et al., 2013; Keller et al., 2013).

The term "hybrid" can be misleading given its sometimes inconsistent usage. In its most conservative definition, it can refer to the offspring of parents belonging to different species. At the opposite extreme, the term can refer to the offspring of genetically distinct parents, which is appropriate in experimental contexts in which parents are chosen based on differences in one or more traits. Individuals in natural populations, however, are by nature unique genotypes and so would universally be classified as hybrids under this definition (Harrison, 1993), making it not particularly useful outside of controlled laboratory settings. A third view defines hybrids as the offspring of individuals belonging to populations that are distinguishable by one or more heritable characters (Harrison, 1990). This definition can be consistently applied to natural populations even without agreement on species concepts or taxonomic categories.

Debate about terminology for interbreeding between species, populations, or individuals downplays the continuum made up of differing levels of genetic differentiation and distracts from the fact that gene flow can have similar consequences at all levels of this continuum. It is therefore more useful to focus on the ecological and evolutionary consequences of hybridization in order to distinguish it from concepts such as gene flow. These characteristics themselves fall on a continuum, thus rendering the distinction between hybridization and gene flow ambiguous in certain contexts. This approach is nonetheless useful for assessing the impacts of common processes while reflecting complex patterns of variation inherent in nature (Gompert & Buerkle, 2016).

The Cohesion Species Concept (Templeton, 1989) offers clarity on the relationship between hybridization and gene flow by distinguishing between genetic exchangeability and demographic exchangeability. Genetic exchangeability refers to the ability to exchange genes, which is a shared

characteristic of both species that hybridize and populations connected by gene flow. Demographic exchangeability refers to the shared intrinsic environmental tolerances of a species, also called the fundamental niche (Hutchinson, 1957). Under the Cohesion Species Concept, populations that exhibit both genetic and demographic exchangeability belong to a single species while populations that exhibit genetic but not demographic exchangeability belong to distinct species that retain the ability to hybridize. The absence of demographic exchangeability may result from ecological divergences resulting from microevolutionary processes that include genetic drift or local adaptation.

The term "hybrid" is additionally confusing because it commonly refers to the F_1 offspring of distinct parents. This leaves ambiguous the classification of F_2, F_3, etc. offspring as well as the offspring of backcrossing between "mixed" and "pure" parents. If hybridization is defined as occurring between individuals of distinct populations or species, these offspring categories would be excluded as products of intrapopulation reproduction. This could be problematic given that hybridization in nature often produces hybrid zones that after a few generations contain few F_1 individuals. One solution to this problem is to consider any individual of mixed ancestry a hybrid. An operational definition would therefore consider a hybrid any individual that is heterozygous or intermediate for at least one genotypic or phenotypic marker that is diagnosably distinct between species, but the implementation of this definition by necessity would be dependent on species concepts.

In this dissertation, hybridization is discussed in reference to two populations of baboons (Kinda and gray-footed chacma) that comprise recognizably distinct lineages (Burrell, 2009) and would therefore qualify as distinct species under the Phylogenetic Species Concept. Because the populations also exhibit substantial phenotypic differences, including extreme differences in body size, and because their distributions correspond to distinct habitat types (Ansell, 1978; Burrell, 2009), the two populations are likely not demographically exchangeable and are therefore also valid species under the cohesion species concept. Following an initial diversification of the genus in the early Pleistocene, these species most likely formed during periods of isolation and reconnection of populations resulting from savannah contractions and expansions in the Pleistocene (Zinner et al., 2009b; Sithaldeen et al., 2015). Gene flow between the species is limited in comparison to gene

flow estimates between the same species and nearby yellow baboons (Burrell, 2009). The timeline of mitochondrial divergence between the species is approximately 1.5 million years (Zinner et al., 2013), but divergence estimates for the nuclear genome are at present not well known.

Hybrid zones are relatively narrow geographic areas in which genetically distinct populations meet, mate, and produce hybrids (Barton & Hewitt, 1985, 1989). Introgression (or introgressive hybridization) describes the transfer of alleles from one population across a hybrid zone into the gene pool of a second, divergent population through hybridization and subsequent backcrossing (Barton & Hewitt, 1985, 1989; Harrison & Larson, 2014). Introgression of alleles is relative; alleles introgress with respect to other alleles in the genome. For introgression to exist, portions of the gene pools of hybridizing species must be divergent and relatively fixed. The degree of introgression across hybrid zones can range from nonexistent (e.g., hybrids are infertile) to extensive, and can vary over space and time depending on factors such as local ecology, demography, climate, geography, and social organization.

Different genomic regions can also introgress across hybrid zones at different rates, depending on their selective value in the recipient form—alleles that increase fitness introgress rapidly whereas alleles that decrease fitness introgress slowly (or not at all) relative to neutral alleles (Barton & Hewitt, 1985, 1989). Furthermore, sex-linked markers including mitochondrial or Y-linked loci can exhibit discordant patterns of introgression due to factors such as sex-biased dispersal (Petit & Excoffier, 2009), mate recognition (Grant & Grant, 1997), or the decreased fitness of heterogametic versus homogametic hybrids, also called Haldane's rule (Orr, 1997).

In the context of speciation, hybridization can yield several outcomes. Some of these involve a balance in which there is no movement towards speciation but existing differentiation is maintained. This can occur in tension zones, which are a type of hybrid zone maintained by a balance between dispersal by parental individuals into the hybrid zone and selection against hybrids (Barton & Hewitt, 1985), or in populations adapted to distinct habitats (Nosil et al., 2009). Alternatively, hybridization may result in the breakdown of barriers to gene flow, potentially resulting in the merger of species or the extinction of genetically distinct populations (Rhymer & Simberloff, 1996; Taylor et al., 2005). Finally, hybridization can result in the formation of new hybrid populations

that remain genetically distinct from both parental populations (hybrid speciation) (Mallet, 2007).

Study animals

Baboons (genus *Papio*) are large-bodied Old World primates that are distributed throughout sub-Saharan Africa and into the Arabian Peninsula. While the taxonomy of *Papio* has been historically contentious (Table 1.1), five parapatric forms are generally recognized: Guinea, hamadryas, anubis, yellow, and chacma. In recent years, the Kinda baboon has received increased recognition as a sixth major kind of baboon (Jolly, 1993, 2001; Frost et al., 2003; Leigh et al., 2003; Leigh, 2006; Zinner et al., 2009b, 2013; Jolly et al., 2011a; Dunn et al., 2013; Boissinot et al., 2014; Kopp et al., 2014; Liedigk et al., 2014; Weyher et al., 2014). Not counting hybrid zones and morphoclines between them, these six forms form stable morphotypes throughout their ranges (Jolly, 1993; Groves, 2001; Grubb et al., 2003) and are therefore considered "good" species under the Phylogenetic Species Concept. It is likely that additional subsets of some of these groups, particularly yellow baboons and chacma baboons (Jolly, 1993, 2001; Burrell, 2009), similarly constitute "good" species under the Phylogenetic Species Concept. Because of extensive evidence of hybridization between neighboring types (Jolly, 1993, 2001; Zinner et al., 2011), and because of evidence of morphoclines linking some taxa (Burrell, 2009), all of these forms would qualify as only one "good" species (*P. hamadryas s. l.*) under the Biological Species Concept.

A third common taxonomic tradition has been to recognize two baboon species based either on phenetic groupings or using the Ecological Species Concept (Van Valen, 1976), in which species are defined based on allopatric or parapatric populations that exhibit low ecological niche separation compared to other populations in their geographic range. These two resulting species are the "desert baboon", *P. hamadryas*, which is equivalent to the hamadryas baboons recognized in the five- or six-species tradition, and the "savanna baboon", *P. cynocephalus s. l.*, which encompasses all non-hamadryas baboons. These classifications emphasize the morphological and sometimes also the behavioral and ecological distinctiveness of the hamadryas baboon, but have generally lost favor as molecular phylogenies consistently show the savanna baboon group to be paraphyletic (e.g., Wildman et al., 2004; Perelman et al., 2011; Zinner et al., 2013), as well as behaviorally and

Table 1.1: Baboon (genus *Papio*) species classifications. Three classifications are listed under the Phylogenetic Species Concept (PSC), reflecting taxonomic treatments that are widely used in the literature, particularly in recent decades. These classifications differ according to the number of species elevated to full species status. Key: PSC, Phylogenetic Species Concept; ESC, Ecological Species Concept; BSC, Biological Species Concept.

<table>
<tr><th>Common name</th><th>PSC (9 spp.)</th><th>PSC (6 spp.)</th><th>PSC (5 spp.)</th><th>ESC</th><th>BSC</th></tr>
<tr><td>hamadryas baboon</td><td>hamadryas</td><td>hamadryas</td><td>hamadryas</td><td>hamadryas</td><td rowspan="10">hamadryas</td></tr>
<tr><td>anubis baboon</td><td>anubis</td><td>anubis</td><td>anubis</td><td rowspan="9">cynocephalus</td></tr>
<tr><td>Guinea baboon</td><td>papio</td><td>papio</td><td>papio</td></tr>
<tr><td>Ibean yellow baboon</td><td>ibeanus</td><td rowspan="2">cynocephalus</td><td rowspan="3">cynocephalus</td></tr>
<tr><td>typical yellow baboon</td><td>cynocephalus</td></tr>
<tr><td>Kinda baboon</td><td>kindae</td><td>kindae</td></tr>
<tr><td>gray-footed chacma baboon</td><td>griseipes</td><td rowspan="3">ursinus</td><td rowspan="3">ursinus</td></tr>
<tr><td>Ruacana chacma baboon</td><td>ruacana</td></tr>
<tr><td>Cape chacma baboon</td><td>ursinus</td></tr>
</table>

ecologically diverse (e.g., Henzi & Barrett, 2003).

Disagreements over species concepts and assignments should not distract from the fact that the biological information about baboons is more interesting, important, and real than the taxonomic decisions based on that information (Jolly, 1993). Names, however, are nevertheless necessary and useful for communication. For the sake of convenience and clarity, I follow Burrell (2009) in recognizing nine *Papio* species using the Phylogenetic Species Concept (Cracraft, 1989): the Guinea baboon (*P. papio*), the hamadryas baboon (*P. hamadryas*), the anubis baboon (*P. anubis*), the "typical" yellow baboon (*P. cynocephalus*), the Ibean yellow baboon (*P. ibeanus*), the Kinda baboon (*P. kindae*), the gray-footed chacma baboon (*P. griseipes*), the Ruacana chacma baboon (*P. ruacana*), and the Cape chacma baboon (*P. ursinus*). I recognize that the species status of *P. ibeanus*, *P. griseipes*, and *P. ruacana* are not commonly recognized and are subject to debate. Because of the prominence of *P. griseipes* in this dissertation and its distinctiveness from Cape chacma baboons (*P. ursinus*), which may be sufficient to justify elevation to species level (Jolly, 1993), I choose to use *P. griseipes* while remaining agnostic on its species status. Unless otherwise specified, I also use the common names "yellow baboon(s)" to refer to *P. cynocephalus* + *P. ibeanus* and "chacma baboon(s)" to refer to *P. griseipes* + *P. ruacana* + *P. ursinus*. Where convenient, I will use the term "grayfoot baboon(s)" synonymously with *P. griseipes*. To distinguish from other

papionins that are sometimes colloquially called "baboons", such as geladas and mandrills, I restrict the use of the term "baboon(s)" solely to genus *Papio*.

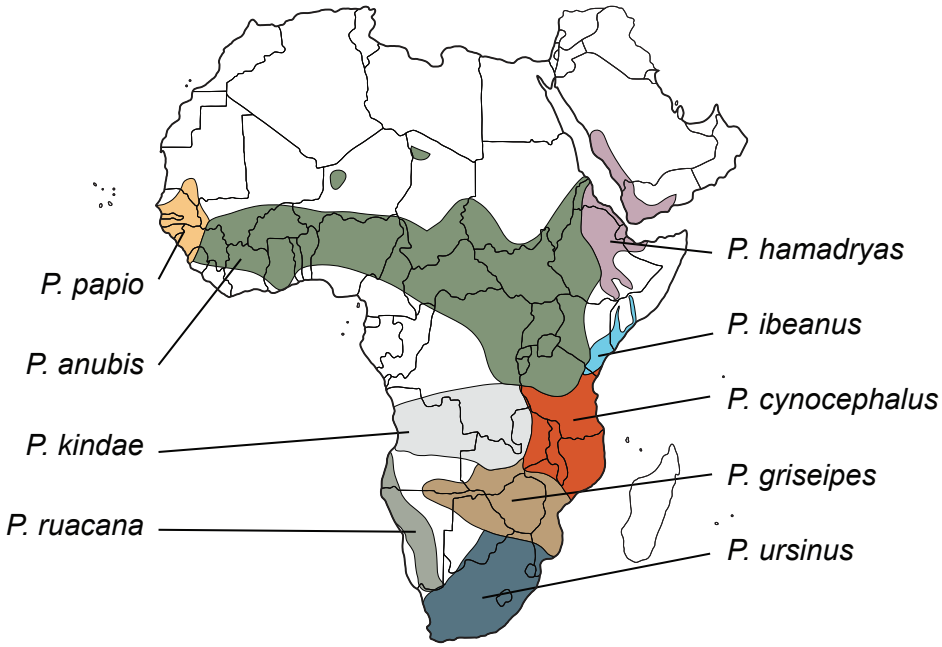

Figure 1.1: Distribution of baboons, genus *Papio*. Species distributions are adapted from Burrell (2009).

Evolution of genus *Papio*

Baboons are members of the Afro-papionin clade (subtribe Papionina), which consists of African representatives of tribe Papionini (Swedell, 2011). The extant genera in this group are *Mandrillus* (drills and mandrills), *Cercocebus* (white-eyelid mangabeys), *Lophocebus* (crested mangabeys), *Rungwecebus* (kipunjis), *Theropithecus* (geladas), and *Papio* (baboons). Interestingly, despite physical resemblances between *Mandrillus*/*Papio* and between *Cercocebus*/*Lophocebus*, the two major subclades within the Afro-papionins are *Mandrillus*/*Cercocebus* and *Lophocebus*/*Rungwecebus*/*Theropithecus*/*Papio*, a finding that is supported by numerous molecular studies (e.g., Disotell et al., 1992; Fabre et al., 2009; Perelman et al., 2011; Finstermeier et al., 2013; Guevara & Steiper, 2014; Liedigk et al., 2014; Pozzi et al., 2014).

Resolving relationships within *Lophocebus*/*Rungwecebus*/*Theropithecus*/*Papio* has been historically more difficult, suggesting a rapid diversification event circa 4 million years ago. While all

possible topologies of *Lophocebus*/*Theropithecus*/*Papio* have been proposed—*Rungwecebus* is often excluded due to its recent discovery (Jones et al., 2005; Davenport et al., 2006) and a scarcity of molecular data—recent evidence suggests that *Papio* + *Lophocebus* is the dominant relationship across the genome (Guevara & Steiper, 2014). Given that nearly half of the genome exhibits an alternative history, however, *Papio*/*Lophocebus*/*Theropithecus* probably represents a trichotomy that cannot be faithfully represented in a bifurcating pattern. For some taxonomic groups, it may ultimately be more useful to summarize species phylogenies as mosaics of conflicting gene trees reflecting complex evolutionary histories rather than to pursue a single consensus tree (Guevara & Steiper, 2014).

The new genus *Rungwecebus* lends further support to this view. Upon its discovery, the species was originally assigned to *Lophocebus* based on the presence of noncontrasting black eyelids and its arboreal behavior (Jones et al., 2005). Subsequent specimens exhibited additional morphological affinities with *Lophocebus* (Davenport et al., 2006). Molecular studies to date, however, consistently position *Rungwecebus* closest to *Papio* (Davenport et al., 2006; Burrell et al., 2009; Zinner et al., 2009a; Roberts et al., 2010), with the first voucher specimen even exhibiting mitochondrial haplotypes of nearby baboon populations (Burrell et al., 2009; Zinner et al., 2009a). Based on this latter result, Burrell et al. (2009) hypothesized that *Rungwecebus* most likely represents a distinct taxon originating from hybridization between male *Lophocebus* and female *Papio cynocephalus*. Zinner et al. (2009a) hypothesized that *Rungwecebus* represents a sister lineage to *Papio* that later underwent some introgressive hybridization with neighboring *Papio* populations. Subsequent discoveries of novel mitochondrial haplotypes in additional kipunji populations support this latter scenario (Roberts et al., 2010). In either case, single consensus bifurcating trees clearly obscure the nuances of complex but real evolutionary histories.

Phylogenetic relationships within *Papio* have been similarly contentious, again reflecting a history of extensive hybridization. A number of studies have investigated the mitochondrial phylogeny of baboons (Newman et al., 2004; Wildman et al., 2004; Burrell, 2009; Sithaldeen et al., 2009; Zinner et al., 2009b, 2013). All of these studies generally support an initial divergence between a northern clade consisting of hamadryas, Guinea, anubis, and northern yellow baboons (located in

Kenya and northern Tanzania); and a southern clade consisting of chacma, Kinda, and southern yellow baboons (located in Zambia, Malawi, and southern Tanzania). A complete mitochondrial DNA phylogeny (Zinner et al., 2013) suggests that southern chacma baboons (i.e., Cape chacma baboons excluding *P. griseipes* and possibly *P. ruacana*) may be the sister lineage to all other *Papio* (Figure 1.2). Mitochondrial phylogenies also exhibit additional paraphyly within the northern and southern clades; anubis baboons, for instance, are divided into western and eastern populations that fall on either side of an east-west divergence within the northern clade of baboons. Because of the discordance of the mitochondrial tree with morphological information, and because paraphyletic mitochondrial relationships overall reflect the clustering of geographically proximate lineages, these analyses reveal multiple incidences of introgressive hybridization leading to asymmetric cytonuclear introgression. This can occur in a scenario where hybridization followed by backcrossing between hybrid females and males of one of the parental morphotypes leads to individuals with mitochondrial DNA from one parental population and nuclear DNA from the other.

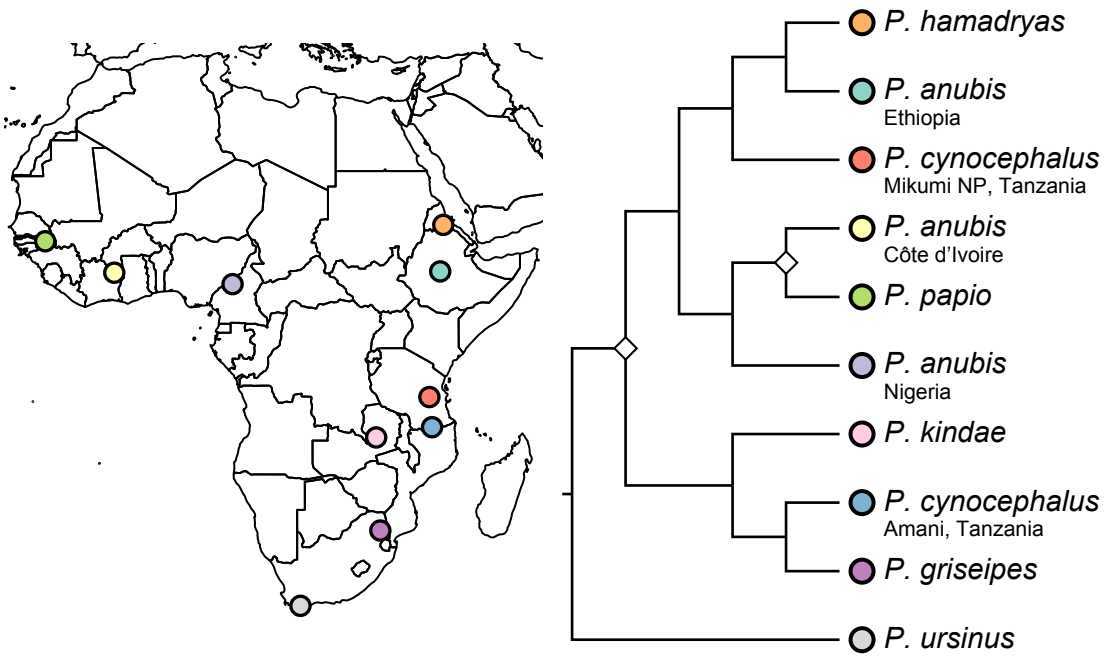

Figure 1.2: Branching patterns among baboon mitochondrial genomes, adapted from Zinner et al. (2013). Labels are modified to match the species designations used in this dissertation. For species with multiple samples, the country or locality is listed for clarity. Nodes with lower support values are indicated with diamonds.

The nuclear phylogeny of baboons is less well-known. A recent study by Boissinot et al. (2014), however, found an initial divergence between chacma baboons and all other baboons using 13 non-coding autosomal markers and 18 transposable elements (Figure 1.3). Within the second clade, they also found a subsequent divergence between yellow baboons and a remaining unresolved clade consisting of anubis, hamadryas, and Guinea baboons. Gray-footed chacma and Kinda baboons were not included in the analysis, and it remains unclear how the addition of these important populations could alter their branching patterns. The analysis, however, reveals a crown diversification of *Papio* dating to about 1.5 million years ago (0.5 – 2.5 million years 95% HPD), which is slightly more recent than mitochondrial estimates of approximately 2 million years (Burrell, 2009; Zinner et al., 2013)

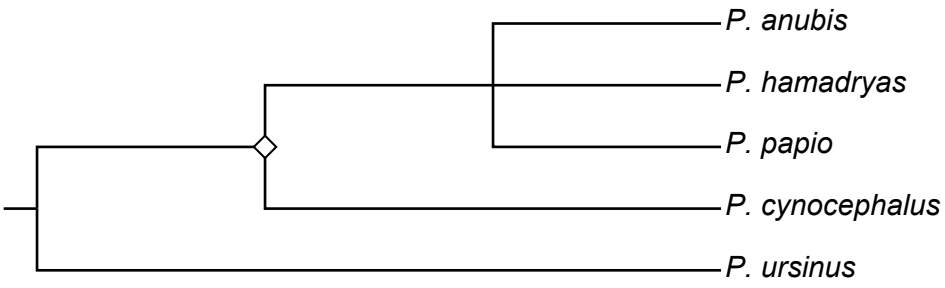

Figure 1.3: Nuclear cladogram, adapted from Boissinot et al. (2014). Species names of captive animals are assigned based on information from the Southwest National Primate Research Center and geographic provenience inferred from mitochondrial haplogroups. Labels match the species designations used in this dissertation. The *anubis-hamadryas-papio* clade is unresolved and collapsed here into a trichotomy. The remaining node with low support values is indicated with a diamond.

The earliest known fossil Afro-papionin is *Parapapio*, which first appears in East Africa around 7.4 to 5.3 million years ago (Leakey et al., 2003) and is present until approximately 2 million years ago in South Africa (Delson, 1988). Many morphs are known from this genus, with the greatest diversification occurring in southern Africa (Jablonski & Frost, 2010). *Parapapio* was a medium-to-large sized papionin with a conservative, macaque-like morphology. It is distinguished by a straight or mostly straight facial profile, a wide and relatively long muzzle, a nonprojecting supraorbital torus, weak maxillary ridges, poorly excavated or absent maxillary and mandibular fossae, and moderate sexual dimorphism (Szalay & Delson, 1979; Jablonski, 2002; Jablonski & Frost, 2010). Similar to extant *Papio*, *Parapapio* in southern Africa appears to have subsisted on

diets of herbaceous materials in savanna environments supplemented with grasses during periods in which preferred fruits were not available (Benefit & McCrossin, 1990). Molar and premolar morphology in *Parapapio* strongly resembles that of *Papio*, except that *Parapapio* exhibits lower sexual dimorphism in tooth size. Because of the dental similarities between *Papio* and *Parapapio*, and because dental evidence comprises much of the fossil evidence of *Parapapio*, taxonomic assignment of *Parapapio*/*Papio* fossils has been historically contentious with ample confusion (Szalay & Delson, 1979; Jablonski & Frost, 2010).

Another early fossil Afro-papionin is *Pliopapio alemui*, which appears in Ethiopia from 5.2 to 4.2 million years ago, and possibly as early as 5.9 million years ago (Frost et al., 2009). It is distinguished from *Papio* by a rounded muzzle, weak maxillary ridges, and poorly excavated maxillary and mandibular fossae; and from *Parapapio* by a pronounced anteorbital drop and ophryonic groove (Frost, 2001). The phylogenetic position of *Pliopapio* is unknown. Frost (2001) hypothesized that *Pliopapio* is either a stem papionin with a uniquely derived rostral profile and anteorbital drop or a stem member of the *Papio*/*Theropithecus*/*Lophocebus* clade. Because of the shared anteorbital drop, he indicated the latter relationship to be more likely. On the other hand, the similarly oblique orientation of the mandibular symphysis suggests that *Pliopapio* may be more closely related to *Parapapio* (Jablonski & Frost, 2010).

Dinopithecus ingens and *Gorgopithecus major* are two other poorly-known large-bodied thick-enameled Plio-Pleistocene Afro-papionins. *Dinopithecus* is distinguished from other papionins by its large size, broad interorbital region, robust masticatory morphology, strong nuchal crests, and large broad molars. It is distinguished from *Papio* by its lack of maxillary and mandibular fossae. *Gorgopithecus* is distinguished from other papionins by its short, high, and narrow muzzle, deep maxillary and mandibular fossae, broad interorbital region, and long calvaria. Freedman (1957) argued that similarities between the molar teeth of *Dinopithecus* and *Gorgopithecus* suggest a close affinity. These forms, which are mainly agreed to be restricted to southern Africa (but see Delson, 1984), could be descendents of southern African *Parapapio* (Jablonski, 2002).

The fossil evidence for the extant genera *Lophocebus*, *Cercocebus*, and *Mandrillus* is largely fragmentary or nonexistent (Jablonski & Frost, 2010). *Theropithecus*, on the other hand, is abundant

and geographically widespread in the Plio-Pleistocene fossil record, with several diverse morphs known (Jolly, 1972; Jablonski, 1993). The earliest known species, *T. darti*, *T. quadratirostris*, and *T. brumpti*, appear to have been largely folivorous, while the later species *T. oswaldi* was more likely graminivorous (Benefit & McCrossin, 1990). Interestingly, the decline and eventual extinction of *Theropithecus* (with the exception of extant *T. gelada*) and *Parapapio* appears to coincide with the diversification of *Papio*, with most ecological roles filled by the extinct radiations now filled by *Papio* species (Jablonski, 1993, 2002). Explanations for the differing success of these radiations center on their energetic (Lee & Foley, 1993) or social (Dunbar, 1993) adaptability in changing environments, as well as their respective abilities to live sympatrically with *Homo* (Isaac, 1977; Jablonski, 2002).

While a number of fossil specimens have been assigned to genus *Papio*, the origin of baboons in the fossil record remains contentious and poorly understood. While some specimens in the East African fossil record prior to the Middle Pleistocene have been assigned to *Papio*, the evidence is largely fragmentary and difficult to distinguish from *Dinopithecus*, *Parapapio*, and sometimes *Theropithecus* (Leakey & Delson, 1987; Jablonski & Leakey, 2008; Jablonski & Frost, 2010; Harrison, 2011). True *Papio* does not appear in East Africa until the Middle Pleistocene (Frost, 2007; Frost & Alemseged, 2007; Jablonski & Leakey, 2008; Jablonski et al., 2008).

Most of the evidence for *Papio* prior to the Pleistocene appears in the southern African fossil record (Delson, 1988; Heaton, 2006; Gilbert, 2013; Gilbert et al., 2015), which is consistent with findings from phylogenetic and phylogeographic studies suggesting a southern African origin of *Papio* (Newman et al., 2004; Wildman et al., 2004; Sithaldeen et al., 2009; Zinner et al., 2009b, 2013; Keller et al., 2010). *P. izodi*, may be the first example of true *Papio* (but see Gilbert, 2013), appearing in Taung and Sterkfontein Members 2 and 4 with an age range between 3.7 and 2.0 million years ago (Delson, 1988; Gilbert, 2013). Another species, *P. angusticeps*, appears in sites such as Kromdraai A, Cooper's A, Gladysvale, Haasgat, and Malapa and is generally dated from 2.0 to 1.5 million years ago (Delson, 1988), although it may extend back as far back as approximately 2.0 to 2.36 million years ago based on dates at Haasgat and Malapa (Adams et al., 2013; Gilbert et al., 2015).

Both *P. izodi* and *P. angusticeps* are small species with similar dentition to *Parapapio* but a steeper drop of the orbital profile. While some authors lump *P. angusticeps* into *P. izodi* (Szalay & Delson, 1979; Jablonski, 2002; Jablonski & Frost, 2010), others treat them as distinct species (Delson, 1988; McKee, 1993; Delson et al., 2000). *P. izodi* exhibits a more primitive suite of characters, including a relatively short zygomatic region, large orbits, a variable anteorbital drop, lightly excavated maxillary and mandibular fossae, lightly developed maxillary ridges, a short and broad muzzle, and relatively large teeth (Delson, 1988; McKee, 1993). *P. angusticeps*, on the other hand, exhibits a relatively tall zygomatic region, small orbits, a sharp anteorbital drop, well excavated maxillary and mandibular fossae, pronounced maxillary ridges, a long and narrow muzzle, and relatively small teeth (Gilbert, 2013; Gilbert et al., 2015). *P. angusticeps* may actually share more similarities with extant baboons, particularly *P. cynocephalus* and *P. kindae* (Delson, 1988; Gilbert, 2013), which has led some authors to place *P. angusticeps* within *P. hamadryas* (*sensu lato*) as a putative progenitor to the yellow baboon group.

P. robinsoni is a third fossil *Papio* species that may shed light on the origins of *Papio*. *P. robinsoni* is known from locales such as Swartkrans Member 1, Bolt's Farm Pit 23, Kromdraai, Drimolen, and possibly Sterkfontein with an estimated age in the range of approximately 2.6 to 1.5 million years ago (Freedman, 1957; Delson, 1984, 1988; Keyser et al., 2000; Jablonski, 2002; Herries et al., 2009). A large form, *P. robinsoni* exhibits some similarities to extant baboons including small orbits, a distinct anteorbital drop, definitive maxillary fossae, pronounced maxillary ridges, a long and narrow muzzle, and relatively small teeth. Unlike extant baboons, however, it exhibits relatively weakly excavated maxillary and mandibular fossae, maxillae that occasionally meet in the midline, and nasal bones that meet at the muzzle dorsum in males (Freedman, 1957; Gilbert, 2013).

The relationships of *P. izodi*, *P. angusticeps*, and *P. robinsoni* to extant baboons and to one another is not well understood. *P. izodi* is most commonly understood to be a sister taxon to extant baboons (Szalay & Delson, 1979; McKee, 1993; Heaton, 2006), although at least one cranio-dental cladistic analysis challenges whether it is even a crown Afro-papionin (Gilbert, 2013). The relationship of *P. angusticeps* and *P. robinsoni* to extant baboons is even less clear. Both have been argued by some authors to be part of the extant baboon radiation, with some evidence pointing to a

closer relationship between *P. angusticeps* and some extant species relative to others (Delson, 1988; Gilbert, 2013). This hypothesis, however, is challenged by recent evidence that *P. angusticeps* may have been present before 2 million years ago (Adams et al., 2013; Gilbert et al., 2015) and that the initial nuclear divergence within extant *Papio* may be closer to 1.5 million years ago (Boissinot et al., 2014) than earlier mitochondrial estimates of 1.8 to 2.0 million years ago (Newman et al., 2004; Wildman et al., 2004). As credible intervals for the divergence are still wide (0.5 – 2.5 million years 95% HPD), however, it cannot be ruled out entirely. *P. robinsoni* occupied a similar geography to extant baboons of the chacma group (including *P. ursinus* and *P. griseipes*), which diverged first from other *Papio*. Its morphology, however, resembles northern forms such as *P. anubis* more than the chacma group (Jolly, 1967), despite their later divergence.

The hypothesis that *Papio* arose in southern Africa is well-supported by phylogenetic and phylogeographic reconstructions (Newman et al., 2004; Wildman et al., 2004; Sithaldeen et al., 2009; Zinner et al., 2009b, 2010, 2013; Keller et al., 2010), by the diversity of fossil *Papio* in southern Africa prior to the middle Pleistocene, and by similarities between *Papio izodi* and *Parapapio* from South Africa, particularly *Parapapio broomi* (Williams et al., 2007). The divergence of *Papio* and dispersal north to form north-south lineages is also temporally in accordance with the expansion of savanna habitats in the Pleistocene as well as with a major radiation of antelopes (Bovidae) and the diversification of hominins (Vrba, 1999; Zinner et al., 2009b).

A partial cranium from the South African Transvaal most likely dating to the Holocene (Freedman, 1957) was assigned by Broom (1936) to *P. spelaeus*. While this specimen has larger teeth and stronger maxillary ridges, it exhibits an overall similarity to the chacma group and is now lumped by most authors with *P. ursinus*, *P. ruacana*, and *P. griseipes* (Freedman, 1957; Jablonski, 2002; Jablonski & Frost, 2010). The only *Papio* species known entirely from the fossil record that appears to fall inside the extant baboon clade is the recently discovered *P. botswanae* from !Ncumtsa, Botswana (Williams et al., 2012). Dated to at least approximately 317 thousand years ago, *P. botswanae* is currently represented by three fossils: a cranium, a frontal bone, and a juvenile mandible. While the fossils fall within the range of variation of extant species, Williams et al. (2012) argue that they are sufficiently distinct as to represent a new taxon. Distinct features of *P.*

botswanae include relatively large orbits as well as a rostrum that is broad relative to its length but narrow relative to the breadth of the cranium. Of all extant baboons, it most closely resembles the chacma group (*P. ursinus*, *P. ruacana*, and *P. griseipes*), sharing possibly derived characters such as a tall neurocranium, well-developed supraorbital torus, and strongly developed maxillary ridges. Given its geographic proximity to extant *P. griseipes* and *P. kindae*, the latter of which it does not resemble, *P. botswanae* most likely shares a sister-group relationship with the chacma group. This hypothesis is plausible given the relatively early divergence of chacma baboons (Zinner et al., 2013; Boissinot et al., 2014). Otherwise, *P. botswanae* may represent a sister taxon to extant *Papio*, despite falling within the extant range of morphometric variation.

Hybridization in *Papio*

Despite the myriad ecological, morphological, and social differences among baboon species (Jolly, 1993; Henzi & Barrett, 2003), baboons hybridize readily upon contact. Ongoing baboon hybridization has been reported between *P. hamadryas* × *P. anubis* at Awash National Park (Jolly & Brett, 1973; Nagel, 1973; Sugawara, 1979; Phillips-Conroy & Jolly, 1986), Arsi (Mori & Belay, 1990), and North Shewa (Beyene, 2007) in Ethiopia; between *P. anubis* × *P. cynocephalus* at Simba Springs (Maples & McKern, 1967; Maples, 1972), Kora (Coe, 1985), and Amboseli National Park in Kenya (Samuels & Altmann, 1986; Alberts & Altmann, 2001; Tung et al., 2008; Charpentier et al., 2012); between *P. cynocephalus* × *P. kindae* and between *P. cynocephalus* × *P. griseipes* × *P. kindae* in the Luangwa Valley of Zambia (Bergey et al., 2009; Burrell, 2009; Phillips-Conroy et al., 2009a); and between *P. kindae* × *P. griseipes* at this project's study site in the Kafue river valley of Zambia (Jolly et al., 2011a). Furthermore, phylogenetic and phylogeographic analyses indicate that introgressive hybridization has influenced the evolution of *Papio* multiple times over its evolutionary history (Wildman et al., 2004; Zinner et al., 2009b; Keller et al., 2010). In rare instances, baboons can even hybridize with other genera such as *Theropithecus* (Dunbar & Dunbar, 1974; Jolly et al., 1997) or *Lophocebus* (Burrell et al., 2009; Zinner et al., 2009a).

The Awash River contact zone has been studied for over 40 years and offers valuable insights into the dynamics of baboon hybridization (Kummer & Kurt, 1963; Kummer et al., 1970; Jolly &

Brett, 1973; Nagel, 1973; Kawai & Sugawara, 1976; Brett et al., 1977; Sugawara, 1979; Phillips-Conroy & Jolly, 1981, 1986; Phillips-Conroy et al., 1991, 1992; Nystrom, 1992; Newman, 1997; Beyene, 1998; Kaplan et al., 1999; Woolley-Barker, 1999; Jolly & Phillips-Conroy, 2003; Nystrom et al., 2004; Beehner et al., 2005; Beehner & Bergman, 2006; Bergman et al., 2008; Jolly et al., 2008, 2013; Hemmalin, 2009; Moritz et al., 2012; Bernstein et al., 2013; Bergey, 2015; Bergey et al., 2016). The Awash hybrid zone is between anubis baboons and hamadryas baboons.

Anubis baboons (Figure 1.4) are found in a variety of habitats but primarily occupy woodlands, savannas, and forests. Their social organization is characterized by multi-male/multi-female groups with female philopatry and male dispersal. Because females remain in their natal groups, they form strong bonds with other females, particularly maternal kin, as well as relatively stable female dominance hierarchies composed of ranked matrilines (Strum, 1987). Male anubis baboons compete for access to females, but only during estrus, during which females may mate with different males (Bercovitch, 1995).

Figure 1.4: Anubis baboons. Photo by Harvey Barrison. Image via Flickr (https://www.flickr.com/photos/hbarrison/7482089610) licensed under Creative Commons (CC BY-SA 2.0).

Hamadryas baboons (Figure 1.5), in contrast, are found in semi-desert environments with sparse resources. Probably as a result of their unproductive environments, hamadryas groups separate into daily subgroups that may merge at progressively higher levels of organization, forming a multilevel society (Kummer, 1968). The smallest and most cohesive level of this organization is the one-male

unit (OMU) consisting of a single male leader with one or more adult females, their offspring, and sometimes additional male followers. Males maintain the cohesion of their OMUs through aggressive herding and neck bites, thereby eliciting a following response. Male attention to females is directed through all stages of female reproductive cycles. To the extent that they are able to maintain their OMUs, males retain exclusive sexual access to females in their OMUs. Males are primarily philopatric, while females disperse from their natal OMUs. Dispersal in female hamadryas baboons, however, may not be directly comparable to other examples of sex-biased dispersal in primates as it is usually not voluntary and usually not across higher levels of hamadryas multilevel societies (Swedell et al., 2011).

Figure 1.5: Female, juvenile, and male hamadryas baboon. Photo by Dick Mudde. Public domain image via Wikimedia Commons (https://commons.wikimedia.org/wiki/File:Baviaan2.JPG).

The Awash contact zone study area is approximately 10 – 40 km long and runs along the Awash River across the Awash Falls on the southern border of Awash National Park. Approximately ten social groups with varying degrees of hybrid ancestry are located along this area, with additional hybrid groups further away from the river (Kummer, 1968; Beyene, 1993). An ecotone traverses this area, with *Acacia* woodland and well-developed gallery forest upriver and steep cliffs, scrubby vegetation, and sparse gallery forest downriver.

Contrary to early predictions, the anubis-hamadryas transition does not correspond neatly with this ecotone. Based on field observations in 1968, Nagel (1973) identified three hybrid groups located between pure anubis and pure hamadryas populations. All three groups, along with one pure anubis and one pure hamadryas group, were located downriver of the ecological boundary line at Awash Falls. He estimated that the hybrid zone was approximately 20 km long, with a sharp, steplike transition in phenotypic appearance across pure and hybrid groups. Based on these observations, Nagel concluded that there was limited gene flow across the hybrid zone (i.e., a "genetic sink").

Later expeditions in 1973 (Phillips-Conroy & Jolly, 1986), 1975 – 1976 (Kawai & Sugawara, 1976; Shotake et al., 1977), and 1978 – 1979 (Shotake, 1981; Sugawara, 1982) found that the distribution of hybrids was wider than reported by Nagel (1973), with smoother intergrading of genotypes and phenotypes between hybrid and parental populations. Because of the extensive sampling of groups and relative unambiguity of phenotypic markers used by Nagel, these results probably indicate genuine changes between 1968 and 1973 rather than observer bias or error (Phillips-Conroy & Jolly, 1986). In contradiction to Nagel's observations in 1968, Phillips-Conroy & Jolly (1986) found that:

1. By 1973, the pure anubis groups just above the Awash Falls contained young hybrids.

2. Hamadryas or hamadryas-like hybrids migrated into groups just above the falls from 1973 to 1975, with immigrants still present in 1979 (they continued to be observed following the study up to when observations ceased in 2000).

3. The pure hamadryas group located downstream of the hybrid zone in 1968 included young hybrids by 1973, as well as an anubis male.

Phillips-Conroy & Jolly (1986) explained these changes in terms of climatic effects. Based on the ecological differences between the species, anubis baboons are expected to be less resilient to periods of low habitat productivity. In times of drought, anubis males are therefore expected to experience disproportionately high mortality in their vulnerable subadult period, during which intertroop migration normally occurs. Relatively resilient hamadryas males, on the other hand, would have a higher capacity for migration, likely with a preference for richer newly-vacant habitats upstream of their previous range. During times of high habitat productivity, in contrast, anubis baboons would

be expected to migrate downstream and extend their ranges. Under this model, the hybrid zone is expected to widen during periods of climatic instability, but conversely to narrow during periods of relative stability.

The proximate mechanisms for admixture in the Awash hybrid zone have been driven by the cross-migration of anubis and hamadryas males into mixed or heterospecific groups (Sugawara, 1979; Phillips-Conroy et al., 1991; Phillips-Conroy & Jolly, 2004), as well as by group fusion (Beyene, 1993, 1998). While it was proposed that hybridization may also result from the recruitment of anubis females into OMUs by hamadryas males (Kummer et al., 1970), a lack of observational evidence as well as insights from mitochondrial DNA research indicating that phenotypically mixed groups are not necessarily mitochondrially mixed (Newman, 1997) suggests that this is not a major mechanism of hybridization. The dispersal of hamadryas males into mixed or anubis groups is probably not a behavior derived from anubis baboons through admixture, but rather an opportunistic attraction to groups with large numbers of unattached females (Phillips-Conroy & Jolly, 2004).

Longitudinal observations of baboons in the Awash contact zone have documented generations of hybrid and backcrossed individuals with no discernible physiological breakdown (Phillips-Conroy & Jolly, 1986). Nevertheless, there is likely to be some form of selection against hybrids based on the following observations (Nagel, 1973; Nystrom, 1992; Beyene, 1998; Bergman, 2000; Beehner, 2003):

1. The gradient of the phenocline has remained high over several decades. In a neutral cline, it would be expected to diminish as successive backcrossing carries foreign alleles farther into species ranges.

2. The breadth of the hybrid zone has remained narrow relative to typical baboon dispersal distances.

3. The Awash hybrid populations consist of a wide distribution of recombinant forms. The bounded hybrid superiority model, in contrast, predicts that one or a few dominant hybrid forms would be favored instead.

Any barriers to gene flow in the hybrid zone probably operate through social and behavioral

differences in mating strategies (Nagel, 1973; Sugawara, 1979, 1982, 1988). Sugawara (1988) found, for instance, that hybrid behavior generally followed morphological phenotype, with hybrid males exhibiting intermediate and therefore ineffective mating behavior. Hybrid males, for instance, attempted to form OMUs through hamadryas-like herding behaviors, but lacked the ability to elicit the requisite following responses in females needed for maintaining OMU cohesion. They were therefore readily outcompeted by other males once females came into estrus. Beehner & Bergman (2006) found in a hybrid group that female grouping behaviors similarly followed phenotypic backgrounds, with hamadryas-like females following OMU strategies, anubis-like females following non-OMU strategies, and intermediate females following loose-OMU patterns. In this group, strict-OMU females exhibited the highest reproductive success. Bergman et al. (2008), however, found in the same group that while morphological phenotypes correlated with behavioral measures, genetic ancestry did not. Moreover, based on microsatellite-derived paternity assignments, they found that some genetically and morphologically intermediate males experienced high reproductive success, indicating that the behavior of hybrid males is not an absolute barrier to admixture.

Observational research at the Awash contact zone has revealed differences in behavioral plasticity between males and females, with male behavior shown to be more inflexible than female behavior. Hamadryas males living in anubis groups, for instance, will attempt to form permanent bonds with any seemingly unattached female (Phillips-Conroy et al., 1991; Nystrom, 1992). Anubis females that were artificially transplanted into hamadryas groups, in contrast, quickly learned to follow leader males (Kummer et al., 1970). Unsurprisingly, hamadryas and anubis male mating strategies are most effective in hamadryas and anubis mating contexts, respectively (Sugawara, 1982; Nystrom, 1992). This effect is compounded by assortative mating preferences. Bergman & Beehner (2003) found that while female mating preferences in hybrid groups were assortative, male preferences were not. Sexual selection may therefore act against heterospecific male mating strategies in contact zones.

As a "natural experiment" (Harrison, 1990), analysis of the Awash hybrid zone has offered insights into the proximate causes for divergent phenotypes in the parental species. Hormonal research, for example, has revealed a positive correlation between several hormones and binding

proteins (insulin-like growth factor 1, insulin-like growth factor binding protein 3, growth hormone binding protein, and testosterone) and growth (Bernstein et al., 2013). All hormones and binding proteins except for testosterone were higher in hamadryas baboons than in anubis baboons. Neurochemical research has revealed species differences in the ratio of dopamine metabolites to serotonin metabolites, with a significantly higher ratio in hamadryas baboons (Jolly et al., 2008). Hybrid animals showed intermediate means and greater diversity in the dopamine/serotonin ratio as well as other variables, indicating a genetic basis for species differences in these neurophysiological traits (Jolly et al., 2013). Recently, based on a genome scan analysis, Bergey et al. (2016) found that the dopamine-receptor-mediated signaling pathway is highly differentiated in the two Awash species. By sequencing exomes from a subset of animals, they further discovered that the dopamine pathway contains an especially large number of variants with high functional impact or causing loss of function. Given dopamine's established relationship with social behavior and temperament as well as the known variation of dopamine metabolites between Awash species (Jolly et al., 2008), this result strongly indicates a functional role of these variants. It also suggests that the dopamine pathway and its associated effects on impulsivity may have been under differential selection in anubis and hamadryas baboons, particularly when viewed in light of their divergent social systems.

The Amboseli hybrid zone provides another valuable perspective on baboon hybridization (Maples & McKern, 1967; Samuels & Altmann, 1986, 1991; Alberts & Altmann, 2001; Charpentier et al., 2008, 2012; Tung et al., 2008). The Amboseli contact zone is between anubis baboons—described above as one of the hybridizing species in the Awash contact zone—and Ibean yellow baboons.

Yellow baboons (Figure 1.6) are found primarily in miombo (*Brachystegia*) woodlands, but also occupy grasslands, Acacia woodlands, agricultural lands, and the edges of coastal and lowland forests. Their behavior and social organization are best known through over 40 years of continuous study of the same study population at Amboseli (Alberts & Altmann, 2012). The Amboseli study population is composed mainly of yellow baboons, specifically Ibean yellow baboons that exhibit relatively coarse and wavy anubis-like mane hair compared to typical yellow baboons (Jolly, 1993). Yellow baboon social organization and behavior is broadly similar to anubis baboons, as well as other

southern species. Their social organization is characterized by multi-male/multi-female groups with female philopatry and male dispersal. Like anubis baboons, females form strong bonds and female rank is inherited matrilineally (Hausfater et al., 1982). Male-male relationships are characterized by high contest competition over estrous females, with relatively unstable ranks over the course of male lifetimes (Noë & Sluijter, 1990).

Figure 1.6: "Ibean" yellow baboons from Amboseli. Photo by Emma Gaiger. Image modified from Flickr (https://www.flickr.com/photos/121840683@N04/13525811423) licensed under Creative Commons (CC BY 2.0).

Maples & McKern (1967) documented the first evidence for hybridization between anubis and Ibean yellow baboons in the area surrounded by Simba Springs and Ithumba Hill, which are separated by approximately 10 km. Animals in the region exhibited morphological traits associated with both anubis and yellow baboons (Maples, 1972), with animals near Ithumba Hill exhibiting mostly anubis features and animals near Simba Springs exhibiting mostly yellow features.

The Amboseli basin baboon population has been studied continuously since the early 1960s (Altmann & Altmann, 1970), with nearly daily observations taken on yellow baboon groups since 1971 (Alberts & Altmann, 2012). Despite its proximity to the boundary joining anubis and yellow species distributions, anubis baboons or anubis-yellow hybrids were not detected until the early 1980s, when an unhabituated group of anubis, yellow, and phenotypically mixed juveniles moved

into the area. In December 1982, a male anubis baboon was observed immigrating into one of the Amboseli study groups, with additional anubis immigrations observed in the years following (Samuels & Altmann, 1986). The source of anubis immigrants in the Amboseli basin is believed to be the foothills of Kilimanjaro (Samuels & Altmann, 1986). From there, they likely follow the Olmolog river basin, which flows into the southwestern portion of the Amboseli basin.

In a genetic study of introgression in the greater Amboseli area, Charpentier et al. (2012) found that, concordant with phenotypic observations, the hybrid zone is relatively narrow with asymmetric gene flow from anubis into yellow populations, suggesting anubis-biased reproductive advantages. Hybrid anubis-yellow baboons appear to differ from yellow baboons in certain life history variables, including an earlier age of male dispersal/migration (Alberts & Altmann, 2001; Charpentier et al., 2008) and an earlier age of maturation (Charpentier et al., 2008). Both of these qualities may support the finding of an anubis-biased advantage in the hybrid zone. Based on longitudinal genetic data, however, Tung et al. (2008) found that the amount of anubis genetic background in hybrids decreased rather than increased over time. They therefore concluded that the reproductive and life-history advantages associated with the anubis background must interact dynamically with nonselective processes such as stochasticity in the anubis immigration rate into Amboseli.

Zambian *Papio*: diversity and distribution

Zambia is a landlocked southern African nation bordering the Democratic Republic of the Congo to the north, Tanzania, Malawi, and Mozambique to the east, Zimbabwe, Botswana, and Nambia to the south, and Angola to the west. With a total of 19 national parks (Figure 1.7), many of which have large surrounding game management areas, Zambia has one of the largest percentages of protected land area of any nation (2014 World Bank data). The national parks are a major component of a developing tourism industry, with flagship species such as lions and elephants attracting hundreds of thousands of visitors each year.

Most of Zambia lies between two major river systems: the Congo River system, which begins from its source the Chambeshi River in northeast Zambia, and the Zambezi River system including

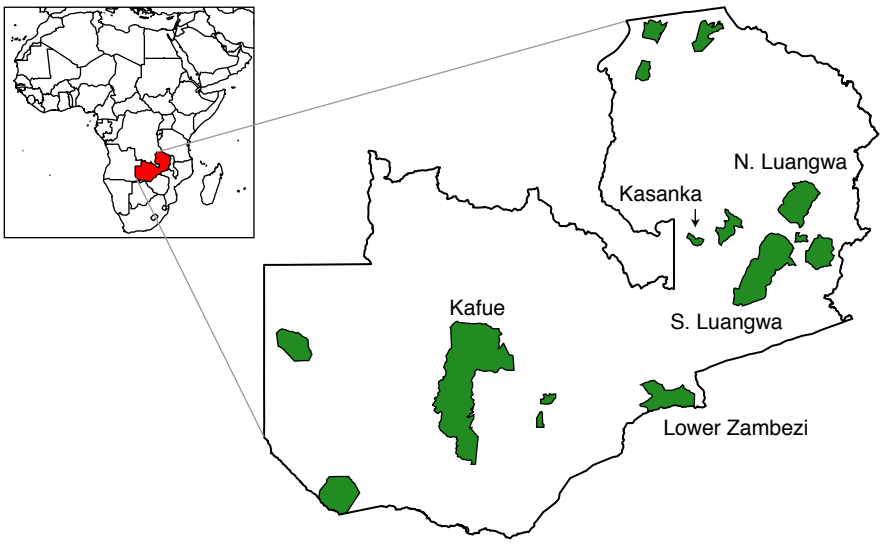

Figure 1.7: Map of national parks in Zambia. Parks are shown in green. Only parks that are mentioned in this dissertation are labeled.

its major tributaries the Kafue and Luangwa Rivers (Figure 1.8). The Zambezi River begins in northwest Zambia, enters Angola, then courses down western Zambia before turning east. Near Livingstone, it passes through a series of gorges forming the well-known Victoria Falls (Mosi-oa-Tunya) then continues through the artificial Lake Kariba into Zimbabwe and Mozambique. The Kafue River originates in the northern Solwezi district, then courses down through the Copperbelt southwest into the Itezhi Tezhi Gap—now the site of the artificial Lake Itezhi Tezhi and the Itezhi Tezhi Dam and hydropower station—where it bends east and flows through the swampy, seasonally inundated Kafue Flats floodplain to join the Zambezi. The Luangwa River begins in the northeastern Mafinga Mountains and flows southwest through a broad valley bounded on the west by the Muchinga Escarpment before joining with the Zambezi.

Zambia falls within a larger belt of Zambezian woodlands that straddles south central Africa (Happold & Lock, 2013). Much of this area is dominated by miombo woodlands, characterized by a high presence of *Brachystegia* and *Julbernardia* trees. The southern Kafue river valley marks a transition between Central African miombo and Southern African mopane woodlands, characterized by a high presence of *Colophospermum mopane* trees. The Luangwa Valley marks a transition

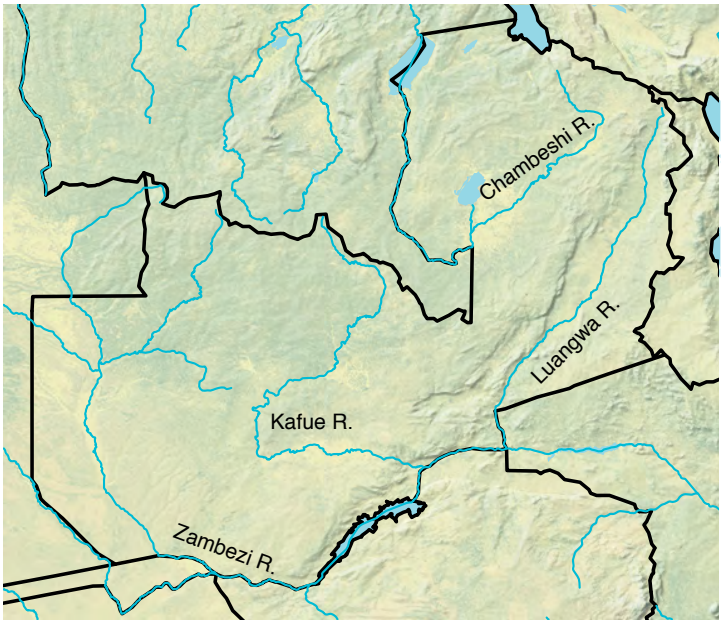

Figure 1.8: Major rivers in Zambia.

between central and eastern miombo woodlands.

Zambia is home to three species of baboon: typical yellow baboons (*P. cynocephalus*), grayfooted chacma or "grayfoot" baboons (*P. griseipes*), and Kinda baboons (*P. kindae*). The distribution of Kinda baboons corresponds roughly with central miombo woodlands in the north-western half of Zambia (Figure 1.10). To the east, they form a wide morphocline in skull size with yellow baboons (Freedman, 1963; Jolly, 1965, 1993; Frost et al., 2003), with baboons intermediate in size known from northern Zambia and possibly Malawi. While there are no published accounts of Kinda or Kinda-like baboons in Tanzania, there are indications that Kinda-like intermediates may also be present at the Mahale Mountains National Park, Tanzania (Burrell, 2009). While the exact location and width of this morphocline remains unclear, genetic studies indicate that a relatively high degree of genetic exchange occurs between Kinda and yellow baboons in the Luangwa Valley (Bergey et al., 2009; Burrell, 2009).

In the southern Luangwa Valley, the distributions of yellow and grayfoot baboons approach one another, with a small degree of genetic exchange occurring between the two species (Burrell, 2009). As the boundary between the two species continues to the east all the way to the Indian Ocean, it seems likely that the the two species form additional hybrid zones elsewhere. While the

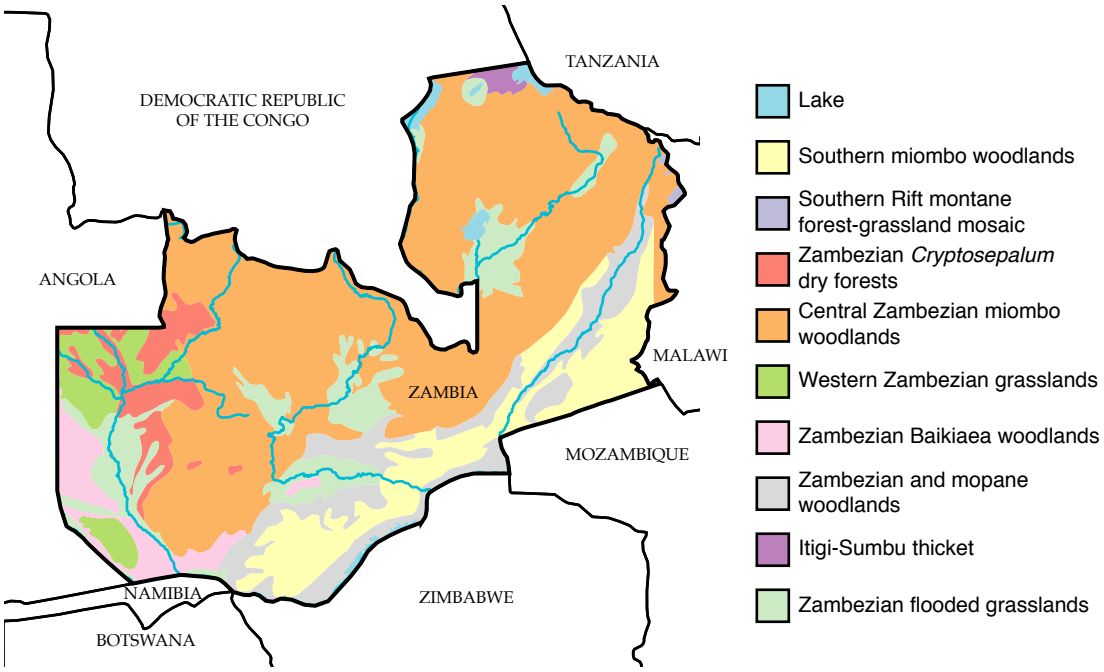

Figure 1.9: Map of terrestrial ecoregions in Zambia. Data from the World Wildlife Fund (Olson et al., 2001).

distributions of Kinda baboons and grayfoot baboons also approach one another in the Luangwa Valley, there is currently minimal evidence of genetic exchange. The two species remain in indirect genetic contact, however, via the Luangwa Valley yellow baboon population (Burrell, 2009).

The area around the capital city of Lusaka, Zambia, is urbanized today and acts as a barrier to contact between Kinda and grayfoot baboons in central Zambia, although earlier records do not indicate the presence of baboons in the area (Figure 1.10). To the west, the Kafue River forms a shallow floodplain, the Kafue Flats, that stretches for approximately 240 km. East of Kafue National Park, where the Kafue River bends east at Itezhi Tezhi, Ansell (1978) reported that the Kafue Flats and Kafue River appeared to limit the two species, with Kinda baboons on the north (and east) bank and grayfoot baboons on the south (and west) bank. West of the river in Kafue National Park, he noted that the distributions of the two species closely approached each other in the vicinity of Lubalunsuki Hill and Itumbi but found no morphological evidence for hybridization (see Study site for a description of the area).

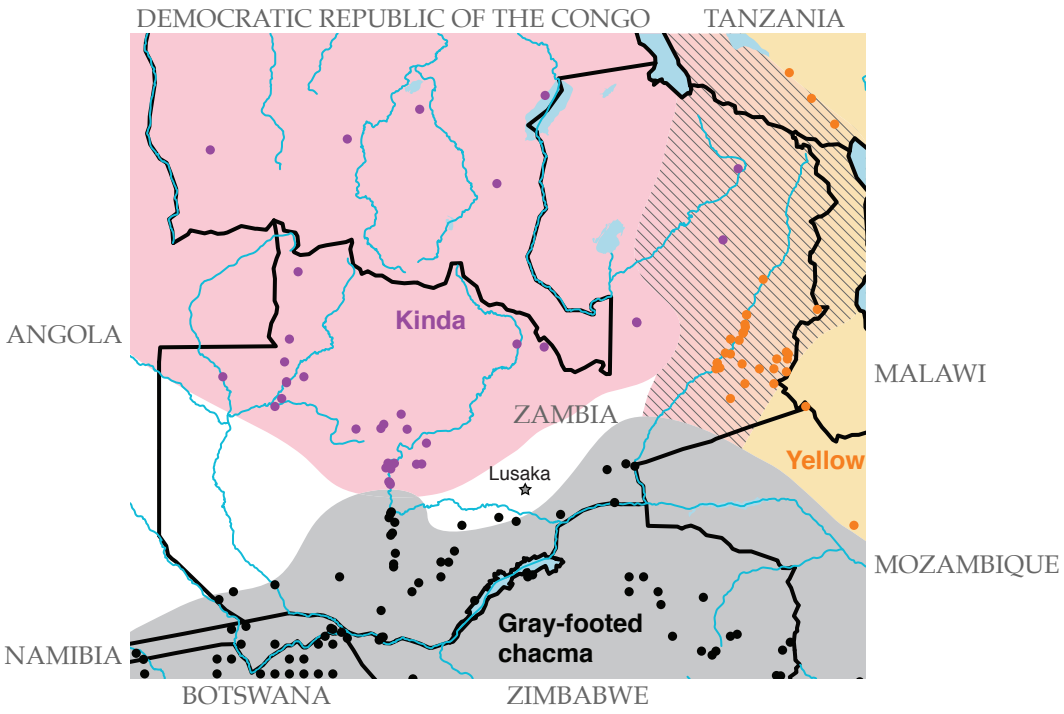

Figure 1.10: Map of baboon distributions in Zambia. Distributions are approximate, but are based on known occurrences shown here as well as on presumed geographic or ecological boundaries. The urban areas surrounding Lusaka as well as the Kafue Flats to their west, for instance, act as a barrier separating Kinda and grayfoot baboons. To the east, the area between Kinda and typical yellow baboon distributions is shown with hatching to reflect that the positions of the boundaries are uncertain. Baboon occurrences are from an unpublished dataset kindly provided by A. Burrell. Distribution areas are adapted from Burrell (2009).

The locations of Lubalunsuki Hill and Itumbi have been a minor mystery in the decades following Ansell's (1960, 1978) descriptions. Jolly et al. (2011a) could not locate these areas but, based on context, interpreted the vicinity to be near the boundary with Kinda baboons at present-day Kasabushi Camp (WP 110 in their dataset). Itumbi, however, is a former safari camp that was closed in 1959 due to trypanosomiasis cases (Mwima, 2001), as I discovered based on precise coordinates found in an internal dataset used by the Zambia Wildlife Authority. I also found Lubalunsuki Hill in a digital gazetteer as "Lubalansuk", which I infer to be the same location based on its close proximity to Itumbi and the obvious presence of a hill, confirmed by cross-referencing the coordinates with topographic information.

More recently, Jolly and I confirmed the locations of these landmarks after discovering a list of names and coordinates in an appendix included by Ansell (1978, pp. 103–112). The listed coordinates for Itumbi and Lubalunsuki Hill are less precise than mine, but unequivocally confirm the locale, which is approximately 25 km south of that inferred by Jolly et al. (2011a). The precise coordinates are 15.51020° S., 25.99219° E. for Itumbi Camp and 15.49916° S., 25.95587° E. for Lubalunsuki Hill (Figure 1.18).

Jolly et al. (2011a) stated that, based on Ansell's (1960) description, Kinda baboons might have extended as far south as the Musa River. They therefore calculated that Kinda and grayfoot baboons in Ansell's time possibly overlapped in a band of about 75 km wide without any signs of interbreeding. This distance depends on the interpretation of "Musa River". Jolly used the Musa Bridge (constructed in 1975) as a reference point, which is 75 km south of WP 110 and 50 km south of Itumbi Camp. Prior to the construction of the Itezhi Tezhi Dam, however, the Musa River extended directly from the Kafue River near Itezhi Tezhi. The old Spinal access road extended north from this point, and this is the coordinate listed as the "Musa River" by Ansell (1978). This point is 55 km south of WP 110 and 30 km south of Itumbi Camp.

Ansell's descriptions, however, call into doubt the usefulness of the Musa River as a boundary. Regarding Kinda baboons, Ansell (1960, p. 29) writes: "In S.P. [Southern Province] must occur north of the Musa and Kafue rivers, and perhaps extends farther south in Namwala and western Kalomo Districts, possibly even to the Zambezi above Livingstone, but if so then there co-existing

with *P. ursinus* [*P. griseipes*]". This description asserts only that Kinda baboons occur north of the Musa and Kafue rivers, with speculation that they may occur farther south. In his later description, Ansell (1978) omits this speculation, as well as any mention of the Musa River as a reference point. This is very likely due to his experiences as warden and wildlife biologist of Kafue National Park from 1966 – 1967 and 1967 – 1969, stationed at Ngoma (Mwima, 2001). By this later description, he clearly considers the area around Itumbi and Lubalunsuki to be the southernmost evidence of Kinda baboons, with the southernmost specimen collected near present-day Kasabushi Camp (15° 16' S., 25° 58' E.).

Based on my interpretation of these descriptions, I therefore conclude that Ansell (1978) did not observe Kinda and grayfoot baboons in an overlap zone of 75 km, but instead found that their distributions approached one another closely with possible overlap. If there was overlap, we can posit that the overlap was minimal given Ansell's station in Kafue National Park. Presumably, as part of his duties, he would have had plentiful opportunities to observe baboons via the old Spinal Road connecting Chunga and Ngoma. Given Ansell's knowledge of species differences, it is unlikely that he was significantly inaccurate in his account. Interestingly, Itumbi and Lubalunsuki Hill are 40 – 50 km north of the area in which Jolly et al. (2011a) found groups that were most phenotypically mixed. This indicates that the center of the hybrid zone has shifted considerably south since the late 1960s, implying a southward movement of Kinda baboons.

Following Ansell's work, the next observations of the Kafue National Park area were conducted by Jolly et al. (2011a) as part of a series of expeditions from 1999 – 2008 (see also Burrell, 2009). During these expeditions, baboons were observed at 48 locations in the Kafue river valley and samples were genotyped from 29 locations for the mitochondrial hypervariable region 1 (HV1) and/or the Y-linked testis-specific protein (TSPY) or DYS576 microsatellite loci. While rigorous phenotypic indices (see Nagel, 1973; Phillips-Conroy & Jolly, 1986) were not attempted, the phenotypic appearances of individuals and the phenotypic compositions of groups in general were assessed. These expeditions were the first to find phenotypic and genotypic evidence for contact and hybridization between Kinda and grayfoot baboons.

In the present day, grayfoot baboons are certainly found in the area north and east of the

Kafue River (Burrell, 2009), which Ansell (1978) previously described as a Kinda zone. It is unclear whether this reflects a change in the grayfoot distribution or if earlier surveys were incorrect. In the area west of the river, Jolly et al. (2011a) found that the animals north of approximately the -15.5° latitude were nearly universally Kinda in appearance and genotype (one animal possessed a grayfoot mitochondrion). This rough boundary is located approximately 0.5° south of the southernmost occurrence of Kinda baboons reported by Ansell (1978). The animals south of approximately the -16.0° latitude or in the area south of Lake Itezhi Tezhi between Ngoma and the Itezhi Tezhi Dam were mainly grayfoot in appearance, with some individuals showing signs of Kinda admixture. All sampled individuals in these areas possessed grayfoot mitochondria, but 5/15 males carried Kinda Y haplotypes. The area south or west of Lake Itezhi Tezhi between the -15.5° and -16.0° latitudes contained mainly animals of mixed appearance. The most northerly of these groups came from an area west of the lake, less than 1 km from the intersection of the present-day Spinal Road with the Chibila River. Further south on the Spinal Road, three phenotypically Kinda individuals were observed to the west of the lake approximately 1 km northeast of the intersection of the Spinal Road with the Mutukushi River. Further south, at the intersection of the the Musa River with the Spinal Road, animals showed a mainly Kinda appearance but exhibited some signs of admixture. All genotyped samples had Kinda mitochondria and all genotyped males had Kinda Y chromosomes.

South of the lake, two locales each contained two neighboring groups with mixed phenotypes (Jolly et al., 2011a). Two of these groups were observed on the North Nkala Road approximately 2 – 3 km south of its intersection with the Spinal Road, while the other two groups were observed on the Cordon Road approximately 0.5 km south of the Ngoma Airstrip. The North Nkala groups were predominantly grayfoot in size and proportions, but exhibited Kinda features. 17 out of 18 of these individuals carried grayfoot mitochondria but 10 out of 13 males carried Kinda Y chromosomes. The Ngoma Airstrip was the most phenotypically diverse group, with some individuals largely grayfoot in appearance and others with Kinda coloration and build but a body size larger than typical for Kinda baboons. 7 out of 11 of these individuals carried grayfoot mitochondria, with no Y-chromosome results reported.

Interestingly, Kinda Y chromosomes were more common overall, including in each of a total of

15 male baboons with Y-mitochondrial discordance in ancestry (Jolly et al., 2011a). 13 of these males were typed for the DYS576 microsatellite locus. Six alleles in total were found, with some alleles occurring in multiple groups and two groups with polymorphisms among discordant males at this locus. The prevalence of the Kinda Y chromosome in the hybrid zone, and its occurrence in every discordant male genotyped, indicates that male Kinda × female grayfoot mating and reproduction must have occurred in the patrilineal history of these baboons and may have even been favored over the reciprocal pattern of female Kinda × male grayfoot reproduction. The Y-chromosome polymorphisms found in discordant males further indicates that multiple instances of these matings occurred. While two instances of discordance with the reciprocal Kinda mitochondrial DNA/grayfoot Y chromosome pattern have since been uncovered for Ngoma Airstrip baboons (McDonald et al., 2016), the disproportionate success of male Kinda baboons remains the dominant pattern.

Previous work has shown that hybridization is rampant in baboon evolution history (see Hybridization in *Papio*). Some case studies, particularly the example of the Awash contact zone (Nagel, 1973; Phillips-Conroy & Jolly, 1986), demonstrate that successful hybridization can occur even between species that are extremely divergent in their morphology, physiology, and behavior. The Kafue contact zone adds another intriguing example of successful hybridization between species that are divergent in aspects of their biology, even occupying opposite extremes in some dimensions.

Grayfoot baboons inhabit areas of southern Zambia as well as parts of Angola, Namibia, Botswana, Zimbabwe, Mozambique, and northern South Africa (Figure 1.1). Across this distribution, they occupy primarily mopane and dry miombo woodlands as well as *Acacia* bushlands and grasslands and coastal mosaics. Grayfoot baboons are near the high extreme of both baboon body mass and sexual dimorphism, with females averaging around 14 – 15 kg and males around 26 – 30 kg (Dunbar, 1990; Delson et al., 2000). Only Cape chacma and possibly Ruacana chacma baboons surpass grayfoot baboons in mean body mass, but only slightly. Most of the knowledge about the natural history of grayfoot baboons comes from a study population at Moremi Game Reserve in the Okavango Delta of Botswana, which has been studied since 1977 by various researchers (e.g.,

Bulger & Hamilton III, 1987; Cheney & Seyfarth, 2008; Silk et al., 2009). Similar to anubis and yellow baboons, grayfoot baboons live in multi-male/multi-female social groups characterized by strong female bonds and male dispersal (Silk et al., 2009, 2010). Reproductive skew is relatively low among females (Silk et al., 2009), but high among males (Bulger, 1993).

Figure 1.11: Gray-footed chacma or "grayfoot" baboons. Photo by Michael Jansen. Image via Flickr (https://www.flickr.com/photos/brainstorm1984/11272680185) licensed under Creative Commons (CC BY-ND 2.0).

Kinda baboons inhabit areas across northern Zambia, the southern Democratic Republic of Congo, and most of Angola (Figure 1.1). They are known mainly from descriptions by Ansell (1978), from museum specimens (Jolly, 1993; Frost et al., 2003; Leigh, 2006), and from genetic studies (Bergey et al., 2009; Burrell, 2009; Jolly et al., 2011a). These studies have revealed several intriguing Kinda features. Kinda baboons are on the low extreme of baboon body mass and sexual dimorphism, which scales allometrically with body mass (Leutenegger & Cheverud, 1982). Adult Kinda females average about 10 kg and adult Kinda males average about 16 kg (J. Phillips-Conroy et al., unpubl. data).

Preliminary and ongoing information about Kinda baboon behavior has come mainly from brief observations of a study population at Chunga, Kafue National Park (Phillips-Conroy et al., 2009b; Weyher, 2010; Chiou, 2013; Weyher et al., 2014), and continuing longitudinal research of a social group at Kasanka National Park, Zambia (A. Weyher, unpubl. data). Results thus far, however, indicate that Kinda baboons are behaviorally divergent from other, more well-known ba-

boons. Unusually, adult males frequently initiate and maintain prolonged grooming associations with non-estrous females (Weyher et al., 2014), with male-to-female groomings accounting for 65% of opposite-sex grooming interactions (Weyher & Chiou, 2013). Additionally, Kinda baboons are unusual in having absent or inconspicuous female post-copulatory calls (Chiou, 2013)[1], a vocalization well-documented and mostly common in other baboons (e.g., Semple et al., 2002; Higham et al., 2009; Nitsch et al., 2011). These findings together suggest that male-female relationships in Kinda baboons are very different from the "typical" baboon pattern.

Figure 1.12: Kinda baboons. Photo by Kenneth Chiou.

Grayfoot and Kinda baboons exhibit a suite of apparently fixed phenotypic differences that are generally diagnostic (Jolly et al., 2011a). Grayfoot baboons exhibit a drab, uniform gray-brown pelage with dark ventral hair, while Kinda baboons exhibit yellow-brown pelage with light yellow ventral hair. Grayfoot baboons also exhibit a robust build with longer and more klinorhynch faces with flat muzzle profiles while Kinda baboons exhibit a gracile and lanky build with shorter and more airorhynch faces with slighly concave muzzle profiles. Grayfoot baboons have long tails with a "broken" tail carriage appearance and a flatter arch while Kinda baboons have short tails

[1]I had previously observed zero female post-copulatory calls in a one-month study of Kinda baboon vocalizations at Chunga, Kafue National Park, Zambia (Chiou, 2013). Since then, however, post-copulatory calls in females have been noted in Kinda baboon females at Kasanka National Park, Zambia, by A. Weyher. I have since observed these calls myself at Kasanka. Nevertheless, while they do occur, the calls seem to be infrequent and inconspicuous. In limited time spent at Chunga since, I have yet to observe female copulation calls.

with a high arch. Kinda baboons exhibit light pink circumorbital skin, strongly contrasting yellow cheek hair, and head hair that frequently ends in a strongly developed "mohawk" with contrasting brown-black hairs. Grayfoot baboons exhibit highly conspicuous white hair patches surrounding their muzzles and a high frequency of long, dark, and curled nape hair. Kinda baboon females exhibit small, relatively inconspicuous sexual swellings, in contrast to the large, exaggerated sexual swellings typically seen in females of other baboon species including grayfoot baboons. Curiously, Kinda baboons are also unique among baboons in having a high frequency of infants born with white or variegated coat colors.

Some of the most striking differences between Kinda and grayfoot baboons relate to body size. Grayfoot males weigh nearly twice as much as Kinda males and nearly three times as much as Kinda females. Grayfoot females weigh about one-and-a-half times more than Kinda females and nearly as much as Kinda males. Hybridization between Kinda and grayfoot baboons might be considered particularly surprising given these dramatic size differences. In a grayfoot male × Kinda female pairing, the male would weigh on average nearly three times the female's body mass. This size difference is comparable to that found in *Mandrillus* but is not seen in any other primate taxon (Smith & Jungers, 1997; Delson et al., 2000). In a grayfoot female × Kinda male pairing, in contrast, the female would weigh on average about 90% of the male's body mass, which reflects a uniformity in size not seen in any other papionin (Delson et al., 2000). Both of these pairings might therefore be susceptible to pre- and post-zygotic obstacles to successful reproduction.

The abundance of Kinda Y-chromosomes in the hybrid zone indicates that male Kinda baboons have been more successful than male grayfoot baboons in successful interspecific reproduction. This is contrary to the typical pattern in vertebrates in which the "father species" tends to be larger (Wirtz, 1999). Because most exceptions to this rule tend to occur when smaller species are disproportionately successful through a strategy of "sneaky matings" (Wirtz, 1999; Redenbach & Taylor, 2003), Jolly et al. (2011a) proposed that Kinda males may avoid contest competition with larger grayfoot males by virtue of their juvenescent appearance. Under this hypothesis, a full-grown Kinda male exhibits the morphometric characteristics of a prepubescent grayfoot male and therefore may not register as a reproductive threat. This potential phenomenon potentially interacts

with other unusual characteristics of Kinda baboons, including the propensity for males to initiate grooming bouts with females. Jolly et al. (2011a) speculated that Kinda male groomings of grayfoot females could be a pleasurable activity for females while projecting a "feminine" behavior that may seem non-threatening to grayfoot males. The close proximity stemming from the grooming context could then provide opportunities for covert matings, particularly if grayfoot males are distracted.

The reciprocal case of a grayfoot male × Kinda female mating might also disadvantage grayfoot males, again due to the juvenescent appearance of Kinda baboons (Jolly et al., 2011a). In this case, the small size of female Kinda baboons may appear prepubescent and therefore unattractive to male grayfoot baboons. This is supported by findings in other taxa in which high-ranking males generally show little interest in and may ignore the solicitations of young females (Anderson, 1986).

Other explanations might also explain the disproportionately high frequency of the Kinda Y chromosome in mixed individuals. Other pre- or postzygotic selection explanations are also possible, including genetic incompatibility between grayfoot Y chromosomes and Kinda autosomal genotypes or obstetric challenges that disproportionately disadvantage females of the smaller species, Kinda baboons.

Relevance to *Homo*

Recent years have seen major developments in our understanding of the evolutionary history of genus *Homo* prior to the disappearance of non-*sapiens* species. Largely as a result of new genetic evidence, all-or-nothing multi-species vs. single-lineage controversies have given way to more nuanced attempts to understand the importance of hybridization and gene flow between distinct, coexisting human lineages. While an increasing appreciation of hybridization has aligned paleoanthropology more closely with developments elsewhere in biology (Abbott et al., 2013), it raises a problem in that modern humans, and even the extant great apes, provide limited analogies for understanding the processes of marginal gene-flow between widespread, adaptable, parapatrically distributed populations of highly social catarrhine primates. In these areas, baboons more closely mirror past humans and therefore serve as powerful natural models (Jolly, 2001).

Introgression is a particularly salient issue for human origins research with regard to past gene

Figure 1.13: Male (above) and female (below) Kinda × grayfoot baboon hybrids. Note the strongly contrasting lightly colored cheek hair, presence of muzzle hair patches (faint in the male), presence of pink circumorbital skin (faint in both), contrasting ventral hair, intermediate builds, and absence of "mohawk" head hair. The male has a klinorhynch (downwardly flexed) face relative to Kinda males. The female's perineal swelling is near its peak but is small and weakly developed relative to typical grayfoot female swellings at equivalent stages. These individuals were photographed on the North Nkala Road (see Table 1.2). Photos by Kenneth Chiou.

flow between anatomically modern *Homo sapiens* and regional archaic humans (Disotell, 2012). This issue is at the forefront of discussions of recent human evolutionary history, most prominently in the debate over the multiregional evolution and recent African origin or "replacement" models of modern human origins. These two hypotheses differ in their predictions regarding admixture. In the multiregional model, admixture serves to maintain the unity of the modern human species as it arose by anagenesis from regional fossil hominins (i.e., *Homo erectus*) via various intermediate species (Wolpoff et al., 1984). In the replacement model, little or no admixture occurred as modern humans arose in Africa and migrated across the world, replacing local archaic species in the process (Stringer & Andrews, 1988). Earlier this decade, advances in ancient DNA technology enabled analysis of the first Neanderthal genome, which revealed evidence of 1 – 4% admixture between Neanderthals and Eurasians but none between Neanderthals and Africans (Green et al., 2010). This unexpected finding is incongruent with both models, as multiregional evolution does not predict equal admixture between Neanderthals and all Eurasians and replacement predicts little to no admixture at all. Ancient DNA analysis of another archaic human genome from Denisova Cave in Siberia revealed evidence of Denisovan contributions to modern human genomes found in island Southeast Asia and Oceania (Reich et al., 2010, 2011). Recently, genomic analysis of Asian and Pacific modern human populations revealed a small proportion of ancestry from an unknown extinct hominin in populations in South and Southeast Asia but not Europeans and East Asians (Mondal et al., 2016). These findings together indicate that gene flow with archaic species has impacted modern human genomes multiple times. They additionally highlight the importance of genomic methods for revealing episodes of admixture that are difficult or impossible to detect based on morphology alone (Templeton, 2005).

Nonhuman models are critical in order to address issues of biological variation and evolutionary history that cannot be answered through human genomics alone due to several key features of the human species. Despite their global distribution and diversity in external phenotype, human populations are highly related with little genetic variation, and their genes coalescence at extremely young dates (200 kya for mitochondrial DNA, 800 kya for most nuclear genes) (Disotell, 2012). Furthermore, while humans do exhibit geographically patterned, genetically based biological varia-

tion, this variation is clinal and continuous, and the genetic diversity existing between populations is dwarfed by the genetic diversity among individuals (Templeton, 2013). Humans also exhibit considerable social (behavioral) variation across their distribution, but the bulk of this variation is cultural and ultimately the distinction of social groups or "races" is a socially transmitted historical construction and not a reflection of genetic variation (Sussman, 2014).

Jolly (2001) has suggested that widespread papionin taxa, including baboons, are ideal models for human population structure prior to the modern human "Out of Africa" expansion. Baboons serve as useful analogs for humans because they are phylogenetically close enough to share many homologous traits, but distant enough such that they exhibit many nonhomologous (homoplastic) features of their morphology, behavior, ecology, and population structure. Baboons are distributed as morphologically diagnosable but not reproductively isolated "allotaxa" that would qualify as species under the Phylogenetic Species Concept but as subspecies under the Biological Species Concept. Ranges of baboon allotaxa come into contact at ecotones in which neighboring allotaxa inhabit similar niches but exhibit phenotypic and behavioral differences that may represent subtle adaptations to slightly different habitats. Baboon allotaxa sometimes also exhibit clear differences in social organization, and at least some of this is genetically based (Stammbach, 1987). Because allotaxa are not reproductively isolated, hybridization is possible. Its genetic outcomes, however, can range from subtle to dramatic, depending on the degree of genetic isolation between species and the consequent number and severity of premating or postmating barriers to successful reproduction. This degree of genetic isolation forms the basis for the Genetic Species Concept (Baker & Bradley, 2006), which treats phylogroups as distinct species when the integrity of their gene pools are sufficiently protected, resulting in shared evolutionary fates within but not between phylogroups.

By emphasizing genetic isolation rather than reproductive isolation, hybridization is possible between genetic species and often expected. When barriers to reproduction between genetic species are strong, the alleles that introgress across allotaxa are likely to be adaptive. In the past, Jolly (2001) has used the papionin analogy to suggest that the easiest way for early modern humans entering Europe to evolve light skin would be to acquire the necessary genes from Neanderthals rather than evolve them *de novo*. Recent genome-wide analyses of archaic ancestry (Sankararaman

et al., 2014; Vernot & Akey, 2014) provide preliminary support for this prediction, most intriguingly in *BNC2*, which is associated with skin pigmentation and freckling in Europeans. Other possible examples of adaptive introgression from archaic humans include HLA (MHC) pathogen defenses in Europeans, East Asians, and Melanesians (Abi-Rached et al., 2011), and the *EPAS1* response to hypobaric hypoxia in highland Tibetans (Huerta-Sanchez et al., 2014).

The Genetic Species Concept (Baker & Bradley, 2006) enables an explicit molecular genetic framework for interpreting baboon hybridization in the broader context of mammalian speciation and hybridization, including modern and archaic humans. The emphasis on the degree of genetic isolation is necessarily comparative and provides a basis for understanding the temporal, geospatial, and ecological dimensions of speciation in distinct lineages. In human evolution after *Homo erectus*, there has been considerable debate regarding whether genus *Homo* experienced cladogenesis with hybridization, or if modern and archaic humans are instead members of a globally distributed time-transgressive metapopulation with isolation by distance (Templeton, 2005; Holliday et al., 2014). Ancient DNA evidence from archaic humans cannot yet distinguish between these two models with confidence, although the low degree of skeletal variation across *Homo* relative to the intrapopulation variation in Pleistocene human populations argues against cladogenesis (Holliday et al., 2014; see comment by Trinkaus). If the human lineage in the past 1.5 million years has been characterized by cladogenesis with secondary contact and hybridization, then the approximately concurrent geographic expansion and diversification of baboons mirrors the history of modern and archaic humans and offers an opportunity to understand the processes of genetic isolation with the benefits of behavioral, demographic, ecological, and other data from an extant nonhuman primate model. If, on the other hand, the human lineage resembled more closely a metapopulation with isolation by distance, then the baboon and human histories in the past 1.5 million years took qualitatively different paths, and contrasting baboon and human patterns of genetic isolation in their geographic, demographic, and ecological context offers arguably an even more valuable framework for understanding the processes that distinguished human evolution from the baboon trajectory.

Study site

Kafue National Park is the largest national park in Zambia, covering an area of 22,319 km^2, which accounts for over a third of the nation's combined national park real estate. The park is situated in south-central Zambia between 14° 00' – 16° 40' S and 25° 15' – 26° 45' E and is named after the Kafue River Basin in which it is located (Figure 1.14). The area has been protected in some capacity since the early 1920s, when the Kafue Game Reserve was formed to control the attrition of wild game populations. In 1950, the Kafue National Park was formally proclaimed as a national park (IUCN category II) by the Governor of Northern Rhodesia, with accompanying relocations of villages and villagers both before and following the proclamation (see Mwima, 2001). The park attained its present status when in 1972 it was officially gazetted as a national park under the National Parks and Wildlife Act. The national park is surrounded by transitional game management areas (IUCN category VI) totaling 43,692 km^2.

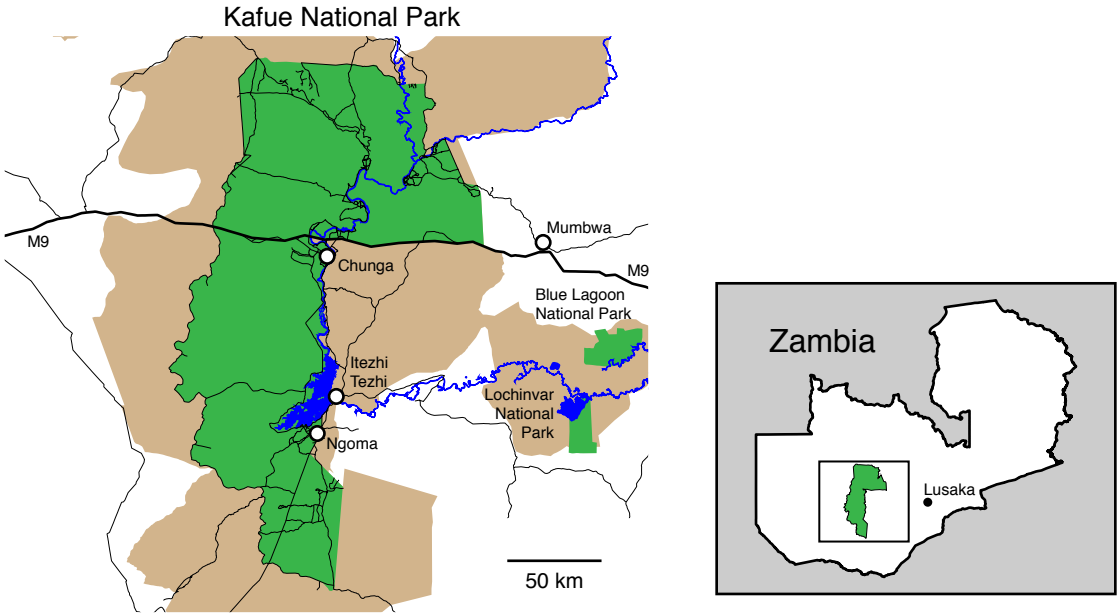

Figure 1.14: Map of Kafue National Park and surrounding areas. National park areas are shown in forest green. Game management areas are shown in tan. Populated and administrative areas are labeled.

Since its formation, the park has been managed by a series of government departments or quasi-

governmental management authorities that have undergone several restructuring episodes. Since 1999, the park has been managed by the Zambia Wildlife Authority, an autonomous agency of the Zambian government that manages most Zambian national parks and game management areas. In late 2015, the Zambia Wildlife Authority was reabsorbed into the Zambian government and renamed the Department of National Parks and Wildlife. Since 1991, the park has been divided into northern and southern administrative districts, headquartered at Chunga and Ngoma respectively. An additional office in Mumbwa approximately 30 km outside the park borders oversees Kafue National Park and other regional parks (Figure 1.14). The park is primarily accessible by the M9 (the "Great West Road"), a major highway that intersects the park, connecting Lusaka to the east with Mongu to the west.

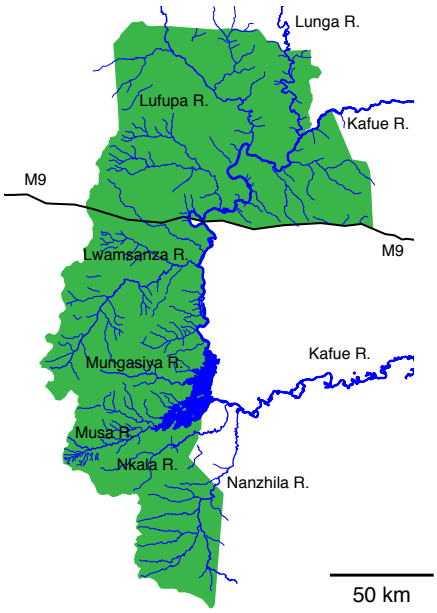

Figure 1.15: Map of Kafue National Park and its tributaries. Major tributaries are labeled.

The park encompasses 260 km of the Kafue River, which along with its tributaries stands as the dominant feature of the park (Figure 1.15). The Kafue River enters the park from the northeast, then joins with its tributary the Lunga River. The river then meanders southwest to the central portion of the park, where it is transected by the M9 road. At this intersection, the river changes direction to form the "Kafue Hook" before continuing south. At the town of Itezhi Tezhi, where

the river joins with its tributary the Musa River, the river turns east towards the Kafue Flats and Lusaka.

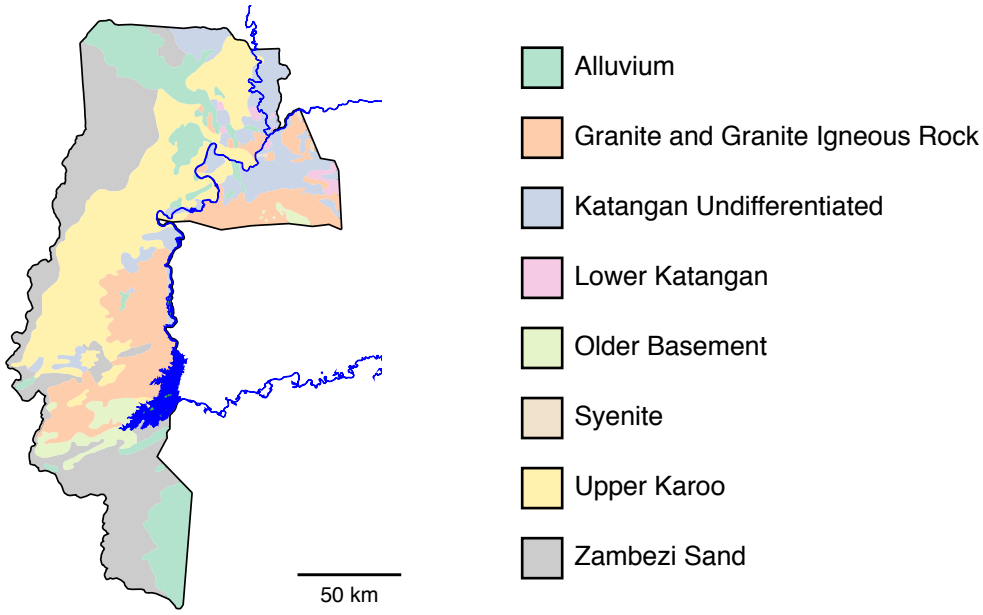

Figure 1.16: Geological map of Kafue National Park. Geological categories are adapted from Moss (1976).

The Kafue Hook is underlain by a granite massif that continues southwest into the central portion of the park (Figure 1.16). Rocks of the Katanga system are found on the margins and Karoo sediments exposed north and south of the granite massif (Fanshawe, 2010). Apart from floodplain clays, the northern end of the park consists of sandy to loamy soils, strongly leached and of low fertility. Dark gray alkaline clays occur through the Nanzhila drainage and to a lesser extent the Nkala, Musa, and Lwamsanza rivers. The remainder of the park north and south of the granite massif is covered with pallid or orange, moderately acidic Kalahari sands. These are not very fertile, having low nitrogen levels, but are well drained (Fanshawe, 2010).

Two large areas of floodplain grassland, the Busanga Plains and the Nanzhila Plains, provide extensive grazing facilities in the north and south of the park, respectively. Apart from these areas, the park is dominated by open woodlands (Figure 1.17). Kalahari woodlands line the northern portion of the park, with miombo woodland dominating much of the central portion, finally giving way to heavily degraded munga and mopane woodland in the south. See Fanshawe (2010) for

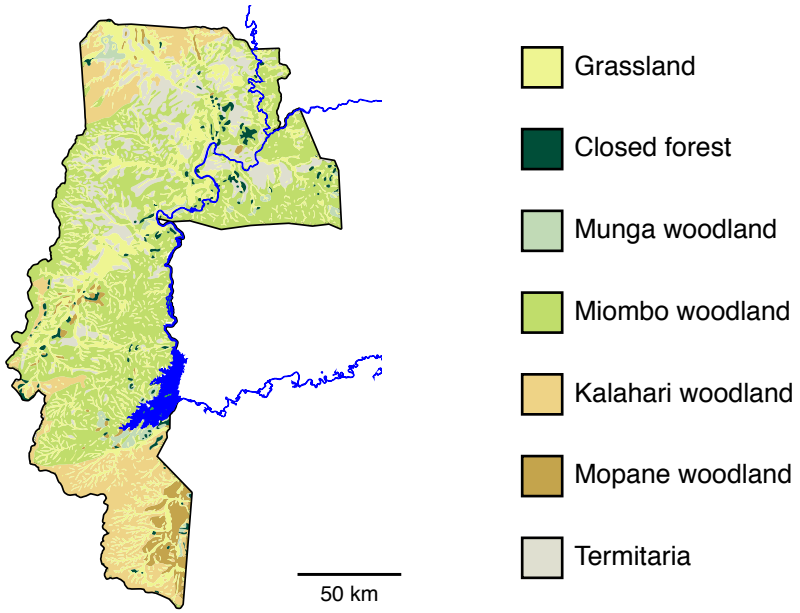

Figure 1.17: Vegetation map of Kafue National Park. Vegetation categories are adapted from from Marjokorpi & Chidumayo (1997). Because of their rarity, "closed forest" classes (*Baikiaea* forest, *Cryptosepalum* forest, Kalahari sand chipya, Lake basin chipya, and riparian forest) are subsumed under one label.

detailed descriptions, including species compositions, of these vegetation classes.

Two distinct seasons characterize the climate of Kafue National Park. A wet season extends from October to April and a dry season extends from April/May to September/October. July is the coldest month with minimum temperatures between 5°C – 7°C and maximum temperatures between 22°C – 28°C. October is the warmest month with minimum temperatures between 15°C – 18°C and maximum temperatures between 31°C – 35°C. Mean annual rainfall ranges from a mean of 510 mm in the south to 1020 mm in the north (Fanshawe, 2010).

Kafue National Park and its surrounding game management areas boast a wide diversity of wildlife species, including 158 mammals, 481 birds, 69 reptiles, 36 amphibians, and 58 fish (Mkanda & Chansa, 2011). Apart from baboons and humans, three other sympatric primates inhabit the park: vervet monkeys (or malbroucks, *Chlorocebus cynosuros*), brown greater galagos (*Otolemur crassicaudatus*), and Mohol bushbabies (*Galago moholi*). A wide variety of potential predators inhabit the area including lions, leopards, cheetahs, spotted hyenas, martial eagles, Nile crocodiles, pythons, and other snakes. While residents of this area of Zambia are not known for hunting

baboons, it is possible that baboons are occasionally hunted by nonlocal poachers (pers. obs.).

The Kafue baboon hybrid zone, and thereby the core study area for this dissertation, is centered around the artificial Lake Itezhi Tezhi, which formed following construction of the Itezhi Tezhi Dam from 1972 to 1976. The dam was built by the Zambia Electricity Supply Corporation (ZESCO) to complement the Kafue Gorge Dam 270 km upstream in response to a growing demand for electricity. The reservoir behind the dam, which has a capacity of 6 billion m^3, stores water during the wet season which is later released for the turbines at the Kafue Gorge. The two dams together are responsible for 50% of Zambia's electricity needs, with surpluses exported to Zimbabwe and South Africa. Beginning in 2011, a separate hydroelectric power station was being constructed at the Itezhi Tezhi Dam as a joint venture between ZESCO and TATA Africa. The power station was officially commissioned in February 2016.

The Kafue Gorge Dam and the Itezhi Tezhi Dam together have dramatically altered the Kafue Flats ecosystem that lies between them, with changes to landscape, vegetation, and wildlife such as the Kafue lechwe (Mumba & Thompson, 2005). Upstream of the Itezhi Tezhi Dam, the most obvious impact has been the inundation of a large area, the artificial Lake Itezhi Tezhi (Figure 1.18). The lake displaced a number of park facilities, including a "spinal" road that connected Chunga and Ngoma, the two major stations within the park. After 1976, a road with an altered course took its place, but soon fell into disrepair. A new gravel Spinal Road was finally completed in 2013, restoring direct vehicular access between the two headquarters.

Research summary

The data for this dissertation project were collected from Zambian baboons during field expeditions in 2006 – 2007, 2011 – 2012, and 2014 – 2015. Captive baboon blood and fecal samples were obtained from a baboon colony at the Southwest National Primate Research Center for the development of the methodology for this project. Wild baboon fecal samples from 2006 and 2007 were collected from Kafue National Park, Lower Zambezi National Park, and Choma during field surveys by C. Jolly, J. Phillips-Conroy, J. Rogers and colleagues. Blood samples from 2011 and 2012 were collected from northern and southern Kafue National Park, respectively, as part of a

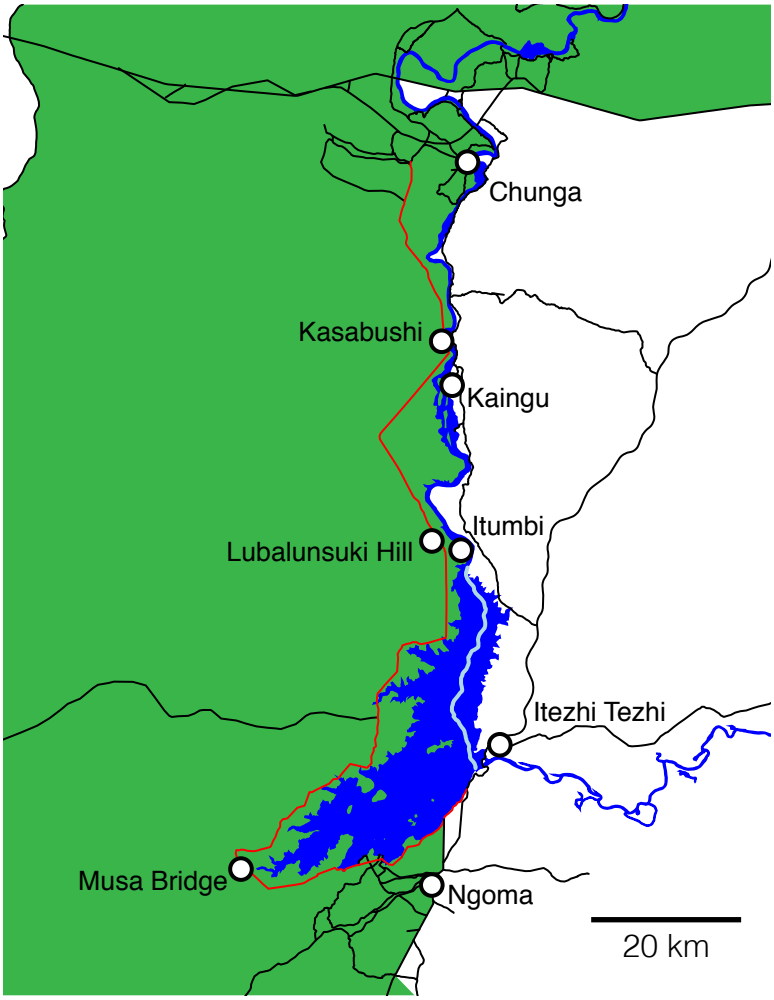

Figure 1.18: Map of Lake Itezhi Tezhi and surrounding areas. Some of the localities that are mentioned in the text (e.g., page 28) are labeled. Lake Itezhi Tezhi formed artificially following the construction of the Itezhi Tezhi Dam between 1972 and 1976. The former course of the Kafue River prior to construction of the dam is shown in light blue. The course of the Spinal Road, completed in 2013, is shown in red.

team trapping expedition led by J. Phillips-Conroy, C. Jolly, and J. Rogers, in which I participated. Fecal samples from 2014 and 2015 were collected from the Lake Itezhi Tezhi area of Kafue National Park during fieldwork conducted for this dissertation.

Because of the population genomic nature of this dissertation project, the choice of field collection methods and laboratory methods was carefully considered. As with other vertebrates, baboons have a large genome size of roughly 3 gigabases (Gb). Given current technological and financial constraints, it was not possible to sequence complete genomes of population-level samples to sufficiently high coverage without a genome complexity-reduction procedure. For this reason, samples were prepared for sequencing using double-digest RAD (restriction-associated DNA) sequencing, or ddRADseq (Peterson et al., 2012).

ddRADseq is a modification of the earlier RADseq method (Baird et al., 2008), designed for rapid and cost-effective *de novo* marker discovery and sequencing. Both RADseq and ddRADseq use restriction enzymes to digest genomic DNA into smaller fragments by cleaving DNA at enzyme-specific cut sites. In RADseq, genomic DNA is digested using a single restriction enzyme, then fragmented stochastically to a target size. Resulting "RADtag" fragments that contain restriction sites on one end are then enriched and sequenced, effectively reducing the billions of bases in a genome to a smaller subset based on proximity to restriction enzyme recognition sites. In ddRADseq, genomic DNA is digested using two separate restriction enzymes with different recognition sites, then size-selected to a target size. Resulting RADtags that are flanked by *both* restriction sites are then enriched and sequenced, effectively reducing the billions of bases in a genome to a smaller subset based on multiple enzyme recognition sites and the nucleotide distance between them. ddRADseq has two important properties that are particularly advantageous for this project. First, because RADtags are flanked by different enzyme recognition sites, it is possible to identify fragments belonging to different samples using a combinatorial indexing approach, thereby cutting costs and increasing the ability to multiplex samples. Second, the size selection procedure applies a shared bias that favors consistent genotyping of RADtags falling nearest the mean of the size distribution. ddRADseq is therefore robust to differences in sequencing coverage among samples, with the number of threshold-passing markers increasing predictably as a function of higher sequencing

coverage among samples (Peterson et al., 2012).

Collecting high-quality genomic DNA samples was a major challenge for this project. Early in my field expeditions, I attempted to collect blood DNA from baboons using a field trapping approach (Brett et al., 1977, 1982; Jolly et al., 2011b). For three months, six unhabituated baboon groups were habituated to my presence and to loose maize (corn) bait, with varying degrees of success. Three groups over this period showed no evidence of consuming any of the maize bait, which was typically eventually salvaged by birds, squirrels, or other animals. Two groups showed some evidence but not enough to begin baiting metal traps within the study period. One group, the North Nkala group, showed encouraging signs by consuming maize within a relatively small distance of observers and by frequently returning to a sleeping site located at 15.90286° S., 25.88951° E. Over ten weeks, the group was habituated to traps, to my presence, and to the sound of closing doors. At the end of the ten weeks, I judged the group to be sufficiently habituated to begin trapping. Over the course of two days, six baboons were successfully trapped using three cages. Of these six baboons, however, three escaped either by breaking the cages through force or by "rolling" the cages so that the doors slid sufficiently to enable exit. One of these escapes was by chance captured by a motion-activated camera (Figure 1.19). Following these incidents, most baboons avoided entering the traps altogether and those that did were highly aware of my presence and reacted strongly to my actions, exiting cages quickly and alarm-calling if they for instance observed me moving or, especially, approaching string triggers. Because of this, I soon called off trapping altogether.

Biopsy darting using homemade biopsy darts (Di Fiore et al., 2009) was also attempted as a method for high-quality sample collection. During the habituation to traps for the North Nkala group, the baboons were habituated to the sight and sound of an unloaded 42" JM Standard air rifle (DANINJECT), 11 mm barrel. After three months, the baboons continued to react strongly to the sound of the CO_2 release and to any movement of the barrel, including my attempts to aim the barrel furtively. I therefore decided to call off the darting without firing a single dart.

Given logistical difficulties in obtaining high-quality invasive samples, genotype data for baboon populations was in large part derived from feces. Feces were collected following established protocols

Figure 1.19: A baboon escaping a metal cage. In this case, the baboon broke the top iron mesh at its juncture with the frame, where the two were welded together. A second baboon also escaped in this manner. Image captured by a motion-activated Moultrie MCG-M990i camera.

(e.g., Burrell et al., 2009), with sample collection concentrated at dawn and dusk, when baboons were generally aggregated near sleeping sites. Fecal samples were collected opportunistically, with preference for fresh samples. 2 ml of feces were collected into 8 ml collection tubes (Sarstedt) containing approximately 4 ml of RNAlater (Ambion), based on an established ratio that has worked well for baboon research (Burrell et al., 2009). Samples were stored at ambient temperature for the duration of the field season, then frozen at -20°C for long-term storage.

Sequencing genomic-scale DNA from feces is challenging because a dominating proportion of fecal-derived DNA is bacterial in origin (Perry et al., 2010; Snyder-Mackler et al., 2016). To address this issue, I developed a new method for partitioning fecal host (baboon) DNA from bacterial DNA. For this purpose, I obtained blood and fecal samples from ten captive baboons at the Southwest National Primate Research Center. These procedures were conducted with approval from the Institutional Animal Care and Use Committee of the Texas Biomedical Research Institute (protocol no. 1403 PC 0). The new method is the subject of Chapter 2 of this dissertation.

During my dissertation fieldwork, surveys were conducted by vehicle in a 1991 Mitsubishi Pajero. During all of my surveys, in order to quantify surveying effort, I kept a GPS track log using a Garmin GPSMAP 64s, a Garmin GLO, and (in 2015) a Bad Elf 2300 GPS device. Coordinates from the Garmin GPSMAP 64s were continuously recorded at 30-second intervals and track logs were uploaded to a computer every day. Upon sightings of baboons, field observations as well as

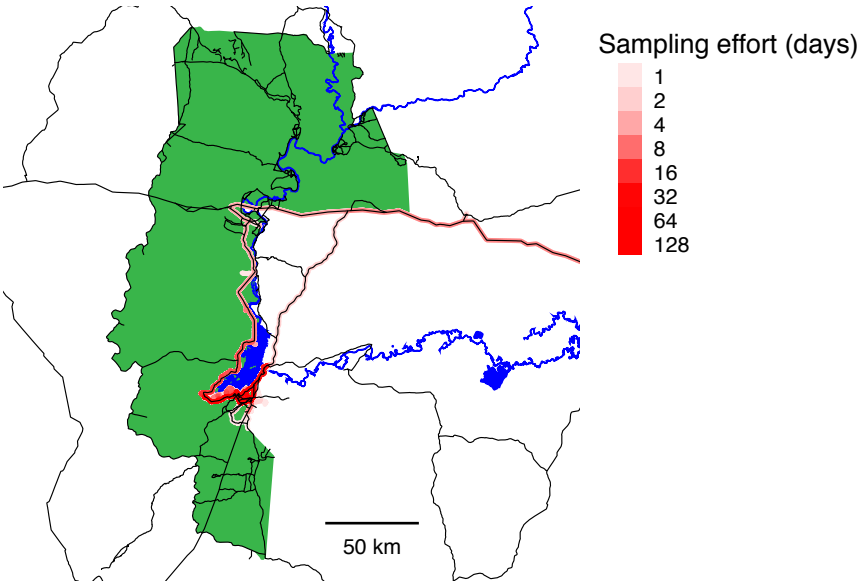

Figure 1.20: Complete track log for this project. Darker red colors indicate a greater number of days spent in an area.

start and end timestamps were recorded by hand in both a notebook and a Google Nexus 5 mobile phone using the default Google Keep note-taking application. When possible, photos were taken using a Nikon D7100 camera. After uploading photo and textual data to a computer, geographic coordinates from the temporally nearest point in the GPS track log were assigned to all photos and notes in a custom MySQL database created for this project (Figure 1.21). Because GPS coordinates were recorded at 30-second intervals, the maximum possible time difference between the recording of an observation and its assigned location is 15 seconds.

Table 1.2: Baboon study groups in Zambia that were analyzed for this dissertation. Samples that were sequenced but did not pass bioinformatic quality-control filters are excluded from this table.

Group ID	N	Locality	Latitude	Longitude	Appearance
A	71	Chunga	-15.0446°	25.9994°	Kinda
B	5	Mwengwa Rapids	-15.3115°	25.9449°	Kinda
C	3	Malala Camp	-15.7661°	25.8626°	mostly Kinda; some grayfoot
D	4	Musa Bridge	-15.9081°	25.7414°	mixed
E	5	Top Musa	-15.9013°	25.8133°	mixed
F	6	North Nkala Road	-15.9029°	25.8895°	mixed
G	12	Ngoma Airstrip	-15.9695°	25.9376°	mixed
H	4	Dendro Park	-16.1518°	26.0599°	mostly grayfoot
I	5	Nanzhila Plains	-16.2337°	25.9612°	mostly grayfoot; some mixed
J	3	Choma	-16.6383°	27.0308°	grayfoot
K	5	Lower Zambezi National Park	-15.9234°	28.9466°	grayfoot

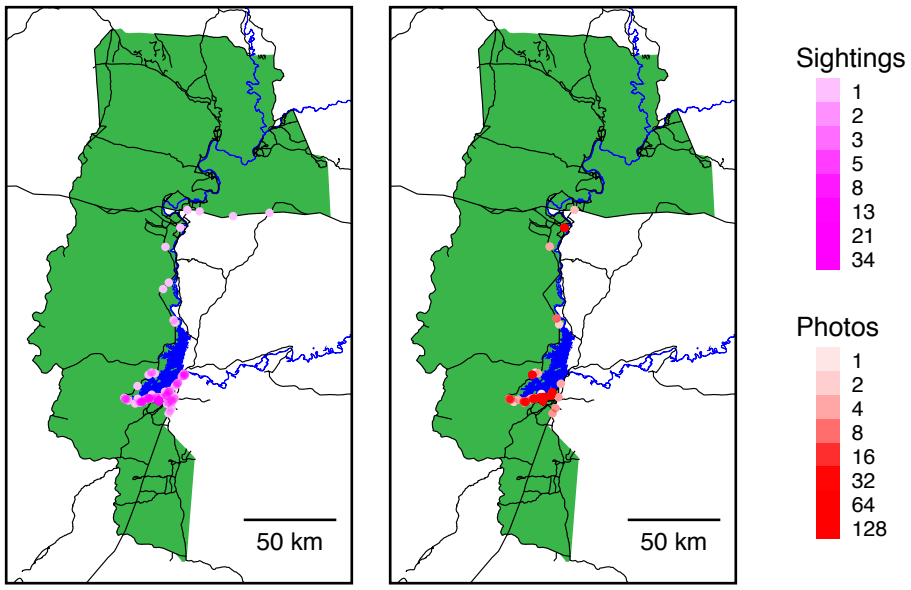

Figure 1.21: Map of sightings recorded and photos taken for this dissertation.

Samples were chosen from groups geographically distributed across the Kafue hybrid zone. Because the animals are largely unhabituated and not subjects of observational research, groups could not be positively identified and were instead assigned based on geographic proximity. For fecal samples, host DNA was enriched in order to remove contaminating bacterial sequences. All DNA was then prepared using ddRADseq and sequenced using Illumina technology. The final set of samples and study groups include only those that passed quality control filters (Table 1.2).

All procedures involving animals, including trapping, darting, and collecting noninvasive samples, were conducted with approval by the Washington University Animal Studies Committee (assurance no. A-3381-01) and following local laws in Zambia.

Overview of upcoming chapters

Apart from this introductory chapter, this dissertation is a collection of three related research articles as well as a short conclusion. Here, I provide a brief synopsis of the three articles.

In Chapter 2, I introduce a newly developed method for genotyping wild animal populations at the genomic scale. This chapter addresses a major technical issue that has until recently limited

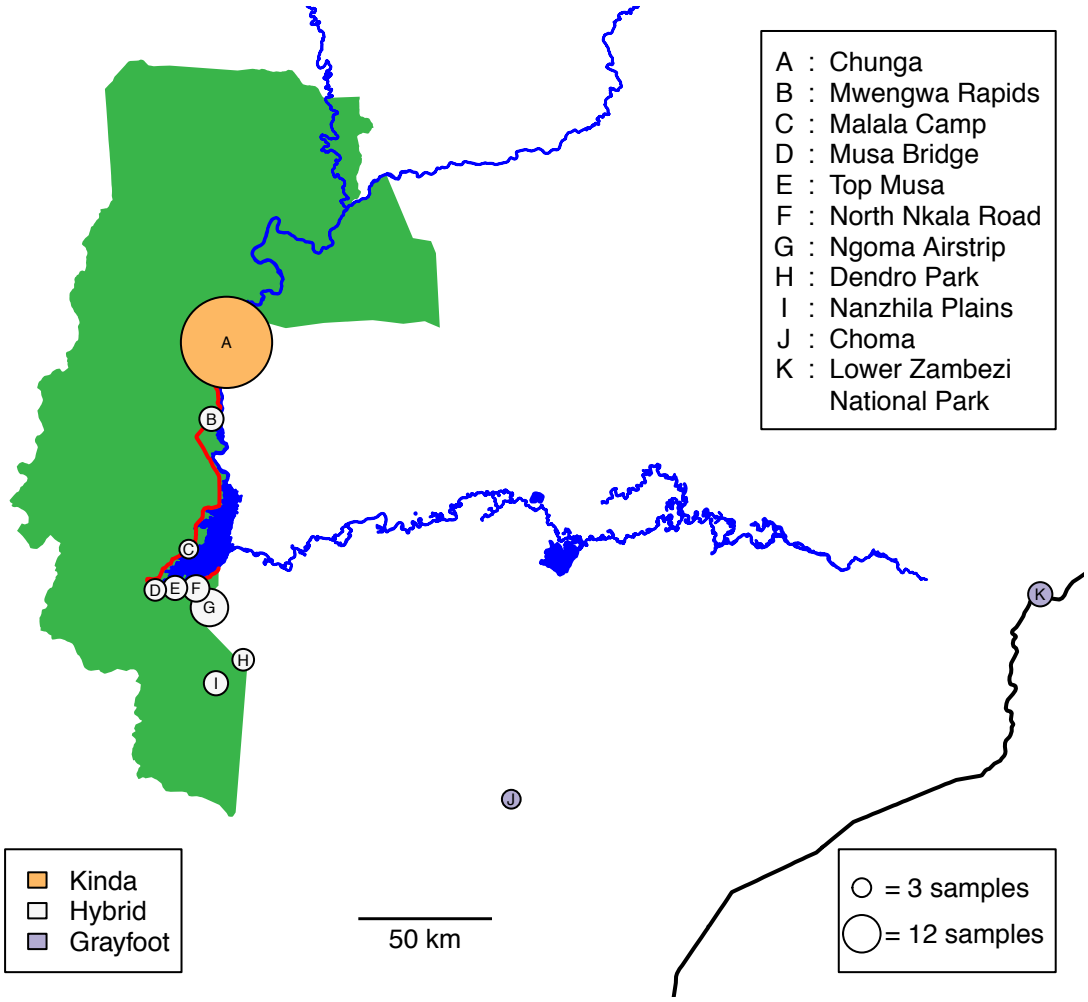

Figure 1.22: Map of samples analyzed for this dissertation.

population genomic research of nonmodel organisms for which invasive samples are not readily available. In this chapter, I describe the premise of the approach, the optimization of the laboratory protocol, and the validation of the method based on controlled comparisons between fecal- and blood-derived DNA libraries.

In Chapter 3, I use a population scan approach to identify signatures of selection between the two hybridizing species. As the species are the focus of this chapter, admixed individuals are excluded from this analysis. After genotyping > 20,000 SNPs across the genome, I calculate the fixation index (F_{ST}) as defined by Weir & Cockerham (1984) as an indicator of selection. I then assign F_{ST} estimates from SNPs to protein-coding genes and assess their significance. In order to perform hypothesis testing of gene function, I use a genome annotation that assigns semantic functional terms to genes. I then conduct statistical tests to scan for gene functions or pathways with extreme patterns in the data.

In Chapter 4, I infer patterns of ancestry across the Kafue river valley hybrid zone and assess the impacts of local adaptation and adaptive introgression on population genetic patterns in the hybrid zone. The analyses of local adaptation and adaptive introgression both use powerful Bayesian statistical techniques that identify statistical outliers while accounting for neutral patterns in the data. Using similar procedures described in the previous paragraph, I assign genomic cline statistics to protein-coding genes and test for gene functions or pathways that are associated with extreme introgression.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martínez-Rodríguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A. W., Parisod, C., Pfennig, K. S., Rice, A. M., Ritchie, M. G., Seifert, B., Smadja, C. M., Stelkens, R., Szymura, J. M., Väinölä, R., Wolf, J. B. W., & Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246.

Abi-Rached, L., Jobin, M. J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., Babrzadeh,

F., Gharizadeh, B., Luo, M., Plummer, F. A., Kimani, J., Carrington, M., Middleton, D., Rajalingam, R., Beksac, M., Marsh, S. G. E., Maiers, M., Guethlein, L. A., Tavoularis, S., Little, A.-M., Green, R. E., Norman, P. J., & Parham, P. (2011). The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science*, *334*, 89–94.

Adams, J. W., Kegley, A. D. T., & Krigbaum, J. (2013). New faunal stable carbon isotope data from the Haasgat HGD assemblage, South Africa, including the first reported values for *Papio angusticeps* and *Cercopithecoides haasgati*. *Journal of Human Evolution*, *64*, 693–698.

Alberts, S. C., & Altmann, J. (2001). Immigration and hybridization patterns of yellow and anubis baboons in and around Amboseli, Kenya. *American Journal of Primatology*, *53*, 139–154.

Alberts, S. C., & Altmann, J. (2012). The Amboseli Baboon Research Project: 40 years of continuity and change. In Kappeler, P.M., Watts, D.P. (Eds.), *Long-Term Field Studies of Primates* (pp. 261–287). Berlin: Springer-Verlag.

Altmann, S. A., & Altmann, J. (1970). *Baboon Ecology: African Field Research*. Chicago: University of Chicago Press.

Anderson, C. M. (1986). Female age: male preference and reproductive success in primates. *International Journal of Primatology*, *7*, 305–326.

Ansell, W. F. H. (1960). *The Mammals of Northern Rhodesia*. Lusaka, Zambia: Government Printer.

Ansell, W. F. H. (1978). *The Mammals of Zambia*. Chilanga, Zambia: The National Parks & Wildlife Service.

Arnold, M. L., & Meyer, A. (2006). Natural hybridization in primates: one evolutionary mechanism. *Zoology*, *109*, 261–276.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, *3*, e3376.

Baker, R. J., & Bradley, R. D. (2006). Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, *87*, 643–662.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113–148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*, 497–503.

Beehner, J. C. (2003). *Female behavior and reproductive success in a hybrid baboon group (*Papio hamadryas hamadryas *x* Papio hamadryas anubis*)* (PhD thesis). Washington University; Washington University.

Beehner, J. C., & Bergman, T. J. (2006). Female behavioral strategies of hybrid baboons in the Awash National Park, Ethiopia. In Swedell, L., Leigh, S.R. (Eds.), *Reproduction and Fitness in Baboons: Behavioral, Ecological, and Life History Perspectives* (pp. 53–79). New York: Springer.

Beehner, J. C., Phillips-Conroy, J. E., & Whitten, P. L. (2005). Female testosterone, dominance rank, and aggression in an Ethiopian population of hybrid baboons. *American Journal of Primatology*, *67*, 101–119.

Benefit, B. R., & McCrossin, M. L. (1990). Diet, species diversity and distribution of African fossil baboons. *Kroeber Anthropological Society Papers*, *71-72*, 77–93.

Bercovitch, F. B. (1995). Female cooperation, consortship maintenance, and male mating success in savanna baboons. *Animal Behaviour*, *50*, 137–149.

Bergey, C. M. (2015). *Population genomics of a baboon hybrid zone* (PhD thesis). New York University; New York University.

Bergey, C. M., Jolly, C. J., Phillips-Conroy, J. E., Burrell, A. S., & Disotell, T. R. (2009). Mitochondrial population structure of a baboon contact zone. *American Journal of Physical Anthropology*, *138 (suppl. 48)*, 89.

Bergey, C. M., Phillips-Conroy, J. E., Disotell R, T., & Jolly, C. J. (2016). Dopamine pathway is highly diverged in primate species that differ markedly in social behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 6178–6181.

Bergman, T. J. (2000). *Mating behavior and reproductive success of hybrid male baboons (*Papio hamadryas hamadryas *X* Papio hamadryas anubis*)* (PhD thesis). Washington University; Washington University.

Bergman, T. J., & Beehner, J. C. (2003). Hybrid zones and sexual selection: insights from the Awash baboon hybrid zone (*Papio hamadryas anubis* x *P. h. hamadryas*). In Jones, C.B. (Ed.), *Sexual Selection and Primates: New Insights and Directions* (pp. 503–537). Norman, OK: American Society of Primatologists.

Bergman, T. J., Phillips-Conroy, J. E., & Jolly, C. J. (2008). Behavioral variation and reproductive success of male baboons (*Papio anubis* × *Papio hamadryas*) in a hybrid social group. *American Journal of Primatology*, *70*, 136–147.

Bernstein, R. M., Drought, H., Phillips-Conroy, J. E., & Jolly, C. J. (2013). Hormonal correlates of divergent growth trajectories in wild male anubis (*Papio anubis*) and hamadryas (*P. hamadryas*) baboons in the Awash River Valley, Ethiopia. *International Journal of Primatology*, *34*, 732–752.

Beyene, S. (1993). Group-fusion and hybridization between anubis and hamadryas baboons at Gola, Ethiopia. *SINET: Ethiopian Journal of Science*, *16*, 61–70.

Beyene, S. (1998). *The role of female mating behavior in hybridization between anubis and hamadryas baboons in Awash, Ethiopia* (PhD thesis). Washington University; Washington Uni-

versity.

Beyene, S. (2007). Preliminary observation of a newly discovered baboon hybrid zone in North Shoa, Ethiopia. *SINET: Ethiopian Journal of Science*, *30*, 143–148.

Boissinot, S., Alvarez, L., Giraldo-Ramirez, J., & Tollis, M. (2014). Neutral nuclear variation in baboons (genus *Papio*) provides insights into their evolutionary and demographic histories. *American Journal of Physical Anthropology*, *155*, 621–634.

Brett, F. L., Jolly, C. J., Socha, W., & Wiener, A. S. (1977). Human-like ABO blood groups in wild Ethiopian baboons. *Yearbook of Physical Anthropology*, *20*, 276–289.

Brett, F. L., Turner, T. R., Jolly, C. J., & Cauble, R. G. (1982). Trapping baboons and vervet monkeys from wild, free-ranging populations. *The Journal of Wildlife Management*, *46*, 164–174.

Brockelman, W. Y., & Gittins, S. P. (1984). Natural hybridization in the *Hylobates lar* species group: implications for speciation in gibbons. In Preuschoft, H. (Ed.), *The Lesser Apes: Evolutionary and Behavioural Biology* (pp. 498–532). Edinburgh: Edinburgh University Press.

Broom, R. (1936). A new fossil baboon from the Transvaal. *Annals of the Transvaal Museum*, *18*, 393–396.

Bulger, J. B. (1993). Dominance rank and access to estrous females in male savanna baboons. *Behaviour*, *127*, 67–103.

Bulger, J., & Hamilton III, W. J. (1987). Rank and density correlates of inclusive fitness measures in a natural chacma baboon (*Papio ursinus*) troop. *International Journal of Primatology*, *8*, 635–650.

Burrell, A. S. (2009). *Phylogenetics and population genetics of central African baboons* (PhD thesis). New York University; New York University.

Burrell, A. S., Jolly, C. J., Tosi, A. J., & Disotell, T. R. (2009). Mitochondrial evidence for the hybrid origin of the kipunji, *Rungwecebus kipunji* (Primates: Papionini). *Molecular Phylogenetics and Evolution*, *51*, 340–348.

Bynum, E. L., Bynum, D. Z., & Supriatna, J. (1997). Confirmation and location of the hybrid zone between wild populations of *Macaca tonkeana* and *Macaca hecki* in Central Sulawesi, Indonesia. *American Journal of Primatology*, *43*, 181–209.

Carneiro, M., Baird, S. J. E., Afonso, S., Ramirez, E., Tarroso, P., Teotónio, H., Villafuerte, R., Nachman, M. W., & Ferrand, N. (2013). Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. *Molecular Ecology*, *22*, 2511–2525.

Charpentier, M. J. E., Fontaine, M. C., Cherel, E., Renoult, J. P., Jenkins, T., Benoit, L., Barthès, N., Alberts, S. C., & Tung, J. (2012). Genetic structure in a dynamic baboon hybrid zone corroborates behavioral observations in a hybrid population. *Molecular Ecology*, *21*, 715–731.

Charpentier, M. J. E., Tung, J., Altmann, J., & Alberts, S. C. (2008). Age at maturity in wild baboons: genetic, environmental and demographic influences. *Molecular Ecology*, *17*, 2026–2040.

Cheney, D. L., & Seyfarth, R. M. (2008). *Baboon Metaphysics: The Evolution of a Social Mind*. Chicago: University Of Chicago Press.

Chiou, K. L. (2013). A pilot description and categorization of Kinda baboon vocalizations. *American Journal of Physical Anthropology*, *150 (suppl. 56)*, 98.

Coe, M. (1985). *Islands in the Bush: A Natural History of the Kora National Reserve, Kenya*. New York: Sheridan House.

Cortés-Ortiz, L., Agostini, I., Aguilar, L. M., Kelaita, M., Silva, F. E., & Bicca-Marques, J. C. (2014). Hybridization in howler monkeys: current understanding and future directions. In Kowalewski, M.M., Garber, P.A., Cortés-Ortiz, L., Urbani, B., Youlatos, D. (Eds.), *Howler Monkeys: Adaptive Radiation, Systematics, and Morphology* (pp. 107–131). New York: Springer.

Cracraft, J. (1989). Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In Otte, D., Endler, J.A. (Eds.), *Speciation and Its Consequences* (pp. 28–59). Sunderland, MA: Sinauer Associates.

Davenport, T. R. B., Stanley, W. T., Sargis, E. J., De Luca, D. W., Mpunga, N. E., Machaga, S. J., & Olson, L. E. (2006). A new genus of African monkey, *Rungwecebus*: morphology, ecology, and molecular phylogenetics. *Science*, *312*, 1378–1381.

Delson, E. (1984). Cercopithecid biochronology of the African Plio-Pleistocene: correlation among eastern and southern hominid-bearing localities. *Courier Forschungsinstitut Senckenberg*, *69*, 199–218.

Delson, E. (1988). Chronology of South African australopith site units. In Grine, F. (Ed.), *Evolutionary History of the "Robust" Australopithecines* (pp. 317–324). New York: Aldine de Gruyter.

Delson, E., Terranova, C. J., Jungers, W. L., Sargis, E. J., Jablonski, N. G., & Dechow, P. C. (2000). Body mass in Cercopithecidae (Primates, Mammalia): estimation and scaling in extinct and extant taxa. *Anthropological Papers of the American Museum of Natural History*, *83*, 1–159.

Detwiler, K. M. (2002). Hybridization between red-tailed monkeys (*Cercopithecus ascanius*) and blue monkeys (*C. mitis*) in East African forests. In Glenn, M.E., Cords, M. (Eds.), *The Guenons: Diversity and Adaptation in African Monkeys* (pp. 79–97). New York: Kluwer Academic Publishers.

Di Fiore, A., Link, A., Schmitt, C. A., & Spehar, S. N. (2009). Dispersal patterns in sympatric woolly and spider monkeys: integrating molecular and observational data. *Behaviour*, *146*, 437–470.

Disotell, T. R. (2012). Archaic human genomics. *Yearbook of Physical Anthropology*, *55*, 24–39.

Disotell, T. R., Honeycutt, R. L., & Ruvolo, M. (1992). Mitochondrial DNA phylogeny of the Old-World monkey tribe Papionini. *Molecular Biology and Evolution*, *9*, 1–13.

Dunbar, R. I. M. (1990). Environmental determinants of intraspecific variation in body weight in baboons (*Papio* spp.). *Journal of Zoology*, *220*, 157–169.

Dunbar, R. I. M. (1993). Socioecology of the extinct theropiths: a modelling approach. In Jablonski, N.G. (Ed.), *Theropithecus: The Rise and Fall of a Primate Genus* (pp. 465–486). Cambridge, UK: Cambridge University Press.

Dunbar, R. I. M., & Dunbar, E. P. (1974). On hybridization between *Theropithecus gelada* and *Papio anubis* in the wild. *Journal of Human Evolution*, *3*, 187–192.

Dunn, J., Cardini, A., & Elton, S. (2013). Biogeographic variation in the baboon: dissecting the cline. *Journal of Anatomy*, *223*, 337–352.

Fabre, P.-H., Rodrigues, A., & Douzery, E. J. P. (2009). Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear DNA. *Molecular Phylogenetics and Evolution*, *53*, 808–825.

Fanshawe, D. B. (2010). *Vegetation descriptions of the Upper Zambezi districts of Zambia.* Occasional Publications in Biodiversity No. 20. Bulaweyo: Biodiversity Foundation for Africa.

Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., & Roos, C. (2013). A mitogenomic phylogeny of living primates. *PLoS One*, *8*, e69504.

Freedman, L. (1957). Fossil Cercopithecoidea of South Africa. *Annals of the Transvaal Museum*, *23*, 121–262.

Freedman, L. (1963). A biometric study of *Papio cynocephalus* skulls from Northern Rhodesia and Nyasaland. *Journal of Mammalogy*, *44*, 24–43.

Frost, S. R. (2001). New Early Pliocene Cercopithecidae (Mammalia: Primates) from Aramis, Middle Awash Valley, Ethiopia. *American Museum Novitates*, *3350*, 1–36.

Frost, S. R. (2007). Fossil Cercopithecidae from the Middle Pleistocene Dawaitoli Formation, Middle Awash Valley, Afar region, Ethiopia. *American Journal of Physical Anthropology*, *134*, 460–471.

Frost, S. R., & Alemseged, Z. (2007). Middle Pleistocene fossil Cercopithecidae from Asbole, Afar Region, Ethiopia. *Journal of Human Evolution*, *53*, 227–259.

Frost, S. R., Haile-Selassie, Y., & Hlusko, L. G. (2009). Cercopithecidae. In Haile-Selassie, Y., WoldeGabriel, G. (Eds.), *Ardipithecus Kadabba: Late Miocene Evidence from the Middle Awash, Ethiopia* (pp. 135–158). Berkeley: University of California Press.

Frost, S. R., Marcus, L. F., Bookstein, F. L., Reddy, D. P., & Delson, E. (2003). Cranial allometry, phylogeography, and systematics of large-bodied papionins (Primates: Cercopithecinae) inferred

from geometric morphometric analysis of landmark data. *The Anatomical Record. Part A, Discoveries in Molecular, Cellular, and Evolutionary Biology*, *275A*, 1048–1072.

Fuzessy, L. F., Silva, I. de O., Malukiewicz, J., Silva, F. F. R., Pônzio, M. do C., Boere, V., & Ackermann, R. R. (2014). Morphological variation in wild marmosets (*Callithrix penicillata* and *C. geoffroyi*) and their hybrids. *Evolutionary Biology*, *41*, 480–493.

Gilbert, C. C. (2013). Cladistic analysis of extant and fossil African papionins using craniodental data. *Journal of Human Evolution*, *64*, 399–433.

Gilbert, C. C., Steininger, C. M., Kibii, J. M., & Berger, L. R. (2015). *Papio* cranium from the hominin-bearing site of Malapa: implications for the evolution of modern baboon cranial morphology and South African Plio-Pleistocene biochronology. *PLoS One*, *10*, e0133361.

Godinho, R., Llaneza, L., Blanco, J. C., Lopes, S., Álvares, F., García, E. J., Palacios, V., Cortés, Y., Talegón, J., & Ferrand, N. (2011). Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular Ecology*, *20*, 5154–5166.

Gompert, Z., & Buerkle, C. A. (2016). What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evolutionary Applications*, *9*, 909–923.

Grant, P. R., & Grant, B. R. (1997). Hybridization, sexual imprinting, and mate choice. *The American Naturalist*, *149*, 1–28.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovič, D., Kucan, Z., Gusic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., Rasilla, M. de la, Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L. F., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., & Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*, 710–722.

Groves, C. P. (2001). *Primate Taxonomy*. Washington, DC: Smithsonian Institution Press.

Grubb, P., Butynski, T. M., Oates, J. F., Bearder, S. K., Disotell, T. R., Groves, C. P., & Struhsaker, T. T. (2003). Assessment of the diversity of African Primates. *International Journal of Primatology*, *24*, 1301–1357.

Guevara, E. E., & Steiper, M. E. (2014). Molecular phylogenetic analysis of the Papionina using concatenation and species tree methods. *Journal of Human Evolution*, *66*, 18–28.

Happold, D., & Lock, J. M. (2013). The biotic zones of Africa. In Kingdon, J., Happold, D., Hoffman, M., Butynski, T., Happold, M., Kalina, J. (Eds.), *Mammals of Africa. Vol. 1: Introductory Chapters and Afrotheria* (pp. 57–74). London: Bloomsbury Publishing.

Harrison, R. G. (1990). Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology*, *7*, 69–128.

Harrison, R. G. (1993). Hybrids and hybrid zones: historical perspective. In Harrison, R.G. (Ed.), *Hybrid Zones and the Evolutionary Process* (pp. 3–12). New York: Oxford University Press.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*, 795–809.

Harrison, T. (2011). Cercopithecoids: (Cercopithecidae, Primates). In Harrison, T. (Ed.), *Paleontology and Geology of Laetoli: Human Evolution in Context*. New York: Springer.

Hausfater, G., Altmann, J., & Altmann, S. (1982). Long-term consistency of dominance relations among female baboons (*Papio cynocephalus*). *Science*, *217*, 752–755.

Heaton, J. L. (2006). *Taxonomy of the Sterkfontein fossil Cercopithecinae: The Papionini of Members 2 and 4 (Gauteng, South Africa)* (PhD thesis). Indiana University; Indiana University.

Hemmalin, H. A. (2009). *Reproductive success of migrant hamadryas baboons in anubis baboon groups* (PhD thesis). New York University; New York University.

Henzi, S. P., & Barrett, L. (2003). Evolutionary ecology, sexual conflict, and behavioral differentiation among baboon populations. *Evolutionary Anthropology*, *12*, 217–230.

Herries, A. I. R., Curnoe, D., & Adams, J. W. (2009). A multi-disciplinary seriation of early Homo and Paranthropus bearing palaeocaves in southern Africa. *Quaternary International*, *202*, 14–28.

Higham, J. P., Semple, S., MacLarnon, A., Heistermann, M., & Ross, C. (2009). Female reproductive signaling, and male mating behavior, in the olive baboon. *Hormones and Behavior*, *55*, 60–67.

Holliday, T. W., Gautney, J. R., & Friedl, L. (2014). Right for the wrong reasons: reflections on modern human origins in the post-Neanderthal genome era. *Current Anthropology*, *55*, 696–724.

Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y., Wang, W., Zhang, X., Xu, X., Yang, H., Li, Y., Wang, J., Wang, J., & Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*, 194–197.

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

Isaac, G. L. (1977). *Olorgesailie: Archaeological Studies of a Middle Pleistocene Lake Basin in Kenya*. Chicago: Chicago University Press.

Jablonski, N. G. (2002). Fossil Old World monkeys: the late Neogene radiation. In Hartwig, W.C. (Ed.), *The Primate Fossil Record* (pp. 255–299). Cambridge, UK: Cambridge University Press.

Jablonski, N. G. (Ed.). (1993). Theropithecus*: The Rise and Fall of a Primate Genus*. Cambridge, UK: Cambridge University Press.

Jablonski, N. G., & Frost, S. R. (2010). Cercopithecoidea. In Werdelin, L., Sanders, W.J. (Eds.), *Cenozoic Mammals of Africa* (pp. 393–428). Berkeley, CA: University of California Press.

Jablonski, N. G., & Leakey, M. G. (2008). The importance of the Cercopithecoidea from the Koobi Fora Formation in the context of primate and mammalian evolution. In Jablonski, N.G., Leakey, M.G. (Eds.), *Koobi Fora Research Project: Volume 6: The Fossil Monkeys* (pp. 397–416). San Francisco: California Academy of Sciences.

Jablonski, N. G., Leakey, M. G., & Antón, M. (2008). Systematic paleontology of the cercopithecines. In Jablonski, N.G., Leakey, M.G. (Eds.), *Koobi Fora Research Project: Volume 6: The Fossil Monkeys* (pp. 103–300). San Francisco: California Academy of Sciences.

Jolly, C. J. (1965). *The origins and specializations of the long-faced Cercopithecoidea* (PhD thesis). University College London; University College London.

Jolly, C. J. (1967). The evolution of the baboons. In Vagtborg, H. (Ed.), *The Baboon in Medical Research, Vol. 2* (pp. 427–457). Austin: University of Texas Press.

Jolly, C. J. (1972). The classification and natural history of *Theropithecus* (*Simopithecus*) (Andrews, 1916) baboons of the African Plio-Pleistocene. *Bulletin of the British Museum of Natural History*, *22*, 1–123.

Jolly, C. J. (1993). Species, subspecies, and baboon systematics. In Kimbel, W.H., Martin, L.B. (Eds.), *Species, Species Concepts, and Primate Evolution* (pp. 67–107). New York: Plenum Press.

Jolly, C. J. (2001). A proper study for mankind: analogies from the papionin monkeys and their implications for human evolution. *Yearbook of Physical Anthropology*, *44*, 177–204.

Jolly, C. J., & Brett, F. L. (1973). Genetic markers and baboon biology. *Journal of Medical Primatology*, *2*, 85–99.

Jolly, C. J., & Phillips-Conroy, J. E. (2003). Testicular size, mating system, and maturation schedules in wild anubis and hamadryas baboons. *International Journal of Primatology*, *24*, 125–142.

Jolly, C. J., Burrell, A. S., Phillips-Conroy, J. E., Bergey, C., & Rogers, J. (2011a). Kinda baboons (*Papio kindae*) and grayfoot chacma baboons (*P. ursinus griseipes*) hybridize in the Kafue river valley, Zambia. *American Journal of Primatology*, *73*, 291–303.

Jolly, C. J., Phillips-Conroy, J. E., & Müller, A. E. (2011b). Trapping primates. In Setchell, J.M., Curtis, D.J. (Eds.), *Field and Laboratory Methods in Primatology* (pp. 133–145). New York: Cambridge University Press.

Jolly, C. J., Phillips-Conroy, J. E., Kaplan, J. R., & Mann, J. J. (2008). Cerebrospinal fluid

monoaminergic metabolites in wild *Papio anubis* and *P. hamadryas* are concordant with taxon-specific behavioral ontogeny. *International Journal of Primatology*, *29*, 1549–1566.

Jolly, C. J., Phillips-Conroy, J. E., Kaplan, J. R., & Mann, J. J. (2013). Monoamine neurotransmitter metabolites in the cerebrospinal fluid of a group of hybrid baboons (*Papio hamadryas* × *P. anubis*). *International Journal of Primatology*, *34*, 836–858.

Jolly, C. J., Woolley-Barker, T., Beyene, S., Disotell, T. R., & Phillips-Conroy, J. E. (1997). Intergeneric hybrid baboons. *International Journal of Primatology*, *18*, 597–627.

Jones, T., Ehardt, C. L., Butynski, T. M., Davenport, T. R. B., Mpunga, N. E., Machaga, S. J., & De Luca, D. W. (2005). The highland mangabey *Lophocebus kipunji*: a new species of African monkey. *Science*, *308*, 1161–1164.

Kanthaswamy, S., Satkoski, J., George, D., Kou, A., Erickson, B. J.-A., & Smith, D. G. (2008). Hybridization and stratification of nuclear genetic variation in *Macaca mulatta* and *M. fascicularis*. *International Journal of Primatology*, *29*, 1295–1311.

Kaplan, J. R., Phillips-Conroy, J. E., Fontenot, M. B., Jolly, C. J., Fairbanks, L. A., & Mann, J. J. (1999). Cerebrospinal fluid monoaminergic metabolites differ in wild anubis and hybrid (*Anubis hamadryas*) baboons: possible relationships to life history and behavior. *Neuropsychopharmacology*, *20*, 517–524.

Kawai, M., & Sugawara, K. (1976). Hybridization and evolution of primates. I. Social structure of hybrid baboon group between *Papio anubis* and *Papio hamadryas*. *Shizen*, *31*, 48–57.

Kawamoto, Y., Ohsawa, H., Nigi, H., Maruhashi, T., Maekawa, S., Shirai, K., & Araki, S. (2001). Genetic assessment of a hybrid population between Japanese and Taiwan macaques in Wakayama Prefecture. *Primate Research*, *17*, 13–24.

Kawamoto, Y., Shirai, K., Araki, S., & Maeno, K. (1999). A case of hybridization between the Japanese and Taiwan macaques found in Wakayama Prefecture. *Primate Research*, *15*, 53–60.

Keller, C., Roos, C., Groeneveld, L. F., Fischer, J., & Zinner, D. (2010). Introgressive hybridization in southern African baboons shapes patterns of mtDNA variation. *American Journal of Physical Anthropology*, *142*, 125–136.

Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., Wittwer, S., & Seehausen, O. (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, *22*, 2848–2863.

Keyser, A. W., Menter, C. G., Moggi-Cecchi, J., Rayne Pickering, T., & Berger, L. R. (2000). Drimolen: a new hominid bearing site in Gauteng, South Africa. *South African Journal of Science*, *96*, 193–197.

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*, 210–225.

Kummer, H. (1968). *Social Organization of Hamadryas Baboons: A Field Study*. Chicago: University of Chicago Press.

Kummer, H., & Kurt, F. (1963). Social units of a free-living population of hamadryas baboons. *Folia Primatologica*, *1*, 4–19.

Kummer, H., Gotz, W., & Angst, W. (1970). Cross-species modifications of social behavior in baboons. In Napier, J.R., Napier, P.H. (Eds.), *Old World Monkeys: Evolution, Systematics and Behavior* (pp. 351–363). New York: Academic Press.

Leakey, M. G., & Delson, E. (1987). Fossil Cercopithecidae from the Laetoli Beds. In Leakey, M.G., Harris, J.M. (Eds.), *The Pliocene Site of Laetoli, Northern Tanzania*. Oxford: Oxford University Press.

Leakey, M. G., Teaford, M. F., & Ward, C. V. (2003). Cercopithecidae from Lothagam. In Leakey, M.G., Harris, J.M. (Eds.), *Lothagam: The Dawn of Humanity in Eastern Africa* (pp. 201–248). New York: Columbia University Press.

Lee, P. C., & Foley, R. A. (1993). Ecological energetics and extinction of giant gelada baboons. In Jablonski, N.G. (Ed.), *Theropithecus: The Rise and Fall of a Primate Genus* (pp. 487–498). Cambridge, UK: Cambridge University Press.

Leigh, S. R. (2006). Cranial ontogeny of *Papio* baboons (*Papio hamadryas*). *American Journal of Physical Anthropology*, *130*, 71–84.

Leigh, S. R., Shah, N. F., & Buchanan, L. S. (2003). Ontogeny and phylogeny in papionin primates. *Journal of Human Evolution*, *45*, 285–316.

Leutenegger, W., & Cheverud, J. (1982). Correlates of sexual dimorphism in primates: ecological and size variables. *International Journal of Primatology*, *3*, 387–402.

Liedigk, R., Roos, C., Brameier, M., & Zinner, D. (2014). Mitogenomics of the Old World monkey tribe Papionini. *BMC Evolutionary Biology*, *14*, 176.

Macholán, M., Munclinger, P., Šugerková, M., Dufková, P., Bímová, B., Božíková, E., Zima, J., & Piálek, J. (2007). Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution*, *61*, 746–771.

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*, 279–283.

Maples, W. R. (1972). Systematic reconsideration and a revision of the nomenclature of Kenya baboons. *American Journal of Physical Anthropology*, *36*, 9–19.

Maples, W. R., & McKern, T. W. (1967). A preliminary report on classification of the Kenya baboon. In Vagtborg, H. (Ed.), *The Baboon in Medical Research, Vol. 2* (pp. 13–22). Austin, TX: University of Texas Press.

Marjokorpi, A., & Chidumayo, E. (1997). *Biodiversity Management in the Provincial Forestry Action Programme Area, Zambia*. Ndola: Forestry Department/FINNIDA.

McDonald, M. M., Phillips-Conroy, J., & Jolly, C. J. (2016). Mating asymmetry in the formation of the Kinda x chacma baboon hybrid zone. *American Journal of Physical Anthropology*, *159 (suppl. 62)*, 224–225.

McKee, J. K. (1993). Taxonomic and evolutionary affinities of *Papio izodi* fossils from Taung and Sterkfontein. *Palaeontologica Africana*, *30*, 43–49.

Mkanda, F. X., & Chansa, W. (2011). Changes in temporal and spatial pattern of road kills along the Lusaka-Mongu (M9) highway, Kafue National Park, Zambia. *South African Journal of Wildlife Research*, *41*, 68–78.

Mondal, M., Casals, F., Xu, T., Dall'Olio, G. M., Pybus, M., Netea, M. G., Comas, D., Laayouni, H., Li, Q., Majumder, P. P., & Bertranpetit, J. (2016). Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nature Genetics*, *48*, 1066–1070.

Mori, A., & Belay, G. (1990). The distribution of baboon species and a new population of gelada baboons along the Wabi-Shebeli River, Ethiopia. *Primates*, *31*, 495–508.

Moritz, G. L., Fourie, N., Yeakel, J. D., Phillips-Conroy, J. E., Jolly, C. J., Koch, P. L., & Dominy, N. J. (2012). Baboons, water, and the ecology of oxygen stable isotopes in an arid hybrid zone. *Physiological and Biochemical Zoology*, *85*, 421–430.

Moss, P. F. N. de V. (1976). *Kafue National Park: a Management Plan*. Lusaka: The National Parks & Wildlife Service.

Mumba, M., & Thompson, J. R. (2005). Hydrological and ecological impacts of dams on the Kafue Flats floodplain system, southern Zambia. *Physics and Chemistry of the Earth*, *30*, 442–447.

Mwima, H. K. (2001). A brief history of Kafue National Park, Zambia. *Koedoe*, *44*, 57–72.

Nagel, U. (1973). A comparison of anubis baboons, hamadryas baboons and their hybrids at a species border in Ethiopia. *Folia Primatologica*, *19*, 104–165.

Newman, T. K. (1997). *Mitochondrial DNA analysis of intraspecific hybridization in* Papio hamadryas anubis*,* P. h. hamadryas *and their hybrids in the Awash National Park, Ethiopia* (PhD thesis). New York University; New York University.

Newman, T. K., Jolly, C. J., & Rogers, J. (2004). Mitochondrial phylogeny and systematics of baboons (*Papio*). *American Journal of Physical Anthropology*, *124*, 17–27.

Nitsch, F., Stueckle, S., Stahl, D., & Zinner, D. (2011). Copulation patterns in captive hamadryas baboons: a quantitative analysis. *Primates*, *52*, 373–383.

Noë, R., & Sluijter, A. A. (1990). Reproductive tactics of male savanna baboons. *Behaviour*, *113*,

117–169.

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*, 145–156.

Nystrom, P. D. A. (1992). *Mating success of hamadryas, anubis and hybrid male baboons in a "mixed" social group in the Awash National Park, Ethiopia* (PhD thesis). Washington University; Washington University.

Nystrom, P., Phillips-Conroy, J. E., & Jolly, C. J. (2004). Dental microwear in anubis and hybrid baboons (*Papio hamadryas*, sensu lato) living in Awash National Park, Ethiopia. *American Journal of Physical Anthropology*, *125*, 279–291.

Ohsawa, H., Morimitsu, Y., Kawamoto, Y., Muroyama, Y., Maekawa, S., Nigi, H., Tori, H., Goto, S., Maruhashi, T., Nakagawa, N., Nakatani, J., Tanaka, T., Hayakawa, S., Yamada, A., Hayaishi, S., Seino, H., Saeki, M., Kawai, S., Hagiwara, H., Suzuki, K., Suzuki, K., Uetsuki, S., Okano, M., Okumura, T., Yoshida, A., & Yokoyama, N. (2005). Population explosion of Taiwanese macaques in Japan. *Natural History Journal of Chulalongkorn University*, *Suppl. 1*, 55–60.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnut, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience*, *51*, 933–938.

Orr, H. A. (1997). Haldane's rule. *Annual Review of Ecology and Systematics*, *28*, 1997.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. Â. M., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M. P. C., Silva, A., O'Brien, S. J., & Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, *7*, e1001342.

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

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

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology & Evolution*, *24*, 386–393.

Phillips, B. L., Baird, S. J. E., & Moritz, C. (2004). When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, *58*, 1536–1548.

Phillips-Conroy, J. E., & Jolly, C. J. (1981). Sexual dimorphism in two subspecies of Ethiopian baboons (*Papio hamadryas*) and their hybrids. *American Journal of Physical Anthropology*, *56*,

115–129.

Phillips-Conroy, J. E., & Jolly, C. J. (1986). Changes in the structure of the baboon hybrid zone in the Awash National Park, Ethiopia. *American Journal of Physical Anthropology*, *71*, 337–350.

Phillips-Conroy, J. E., & Jolly, C. J. (2004). Male dispersal and philopatry in the Awash baboon hybrid zone. *Primate Report*, *68*, 27–52.

Phillips-Conroy, J. E., Jolly, C. J., & Brett, F. L. (1991). Characteristics of hamadryas-like male baboons living in anubis baboon troops in the Awash hybrid zone, Ethiopia. *American Journal of Physical Anthropology*, *86*, 353–368.

Phillips-Conroy, J. E., Jolly, C. J., Burrell, A. S., Rogers, J., & Weyher, A. H. (2009a). Genetic and behavioral observations of "Kinda" baboons (*Papio cynocephalus kindae*) in Zambia. *American Journal of Physical Anthropology*, *138 (suppl. 48)*, 211.

Phillips-Conroy, J. E., Jolly, C. J., Nystrom, P., & Hemmalin, H. A. (1992). Migration of male hamadryas baboons into anubis groups in the Awash National Park, Ethiopia. *International Journal of Primatology*, *13*, 455–476.

Phillips-Conroy, J., Jolly, C. J., & Weyher, A. H. (2009b). Observations of "Kinda" baboons (*Papio cynocephalus kindae*) in Zambia: adult males as the active partner in male-female grooming dyads. *American Journal of Primatology*, *71 (suppl. 1)*, 65.

Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *75*, 165–183.

Redenbach, Z., & Taylor, E. B. (2003). Evidence for bimodal hybrid zones between two species of char (Pisces: *Salvelinus*) in northwestern North America. *Journal of Evolutionary Biology*, *16*, 1135–1148.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J.-J., Kelso, J., Slatkin, M., & Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*, 1053–1060.

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., Ko, A. M.-S., Ko, Y.-C., Jinam, T. A., Phipps, M. E., Saitou, N., Wollstein, A., Kayser, M., Pääbo, S., & Stoneking, M. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *American Journal of Human Genetics*, *89*, 516–528.

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

Roberts, T. E., Davenport, T. R. B., Hildebrandt, K. B. P., Jones, T., Stanley, W. T., Sargis, E. J.,

& Olson, L. E. (2010). The biogeography of introgression in the critically endangered African monkey *Rungwecebus kipunji*. *Biology Letters*, *6*, 233–237.

Samuels, A., & Altmann, J. (1986). Immigration of a *Papio anubis male* into a group of *Papio cynocephalus* baboons and evidence for an *anubis-cynocephalus* hybrid zone in Amboseli, Kenya. *International Journal of Primatology*, *7*, 131–138.

Samuels, A., & Altmann, J. (1991). Baboons of the Amboseli Basin: demographic stability and change. *International Journal of Primatology*, *12*, 1–19.

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., & Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*, 354–357.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*, 198–207.

Semple, S., McComb, K., Alberts, S. C., & Altmann, J. (2002). Information content of female copulation calls in yellow baboons. *American Journal of Primatology*, *56*, 43–56.

Shotake, T. (1981). Population genetical study of natural hybridization between *Papio anubis* and *P. hamadryas*. *Primates*, *22*, 285–308.

Shotake, T., Nozawa, K., & Tanabe, Y. (1977). Blood protein variations in baboons I. Gene exchange and genetic distance between *Papio anubis*, *Papio hamadryas* and their hybrid. *The Japanese Journal of Genetics*, *52*, 223–237.

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2009). The benefits of social capital: close social bonds among female baboons enhance offspring survival. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *276*, 3099–3104.

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2010). Female chacma baboons form strong, equitable, and enduring social bonds. *Behavioral Ecology and Sociobiology*, *64*, 1733–1747.

Sithaldeen, R., Ackermann, R. R., & Bishop, J. M. (2015). Pleistocene aridification cycles shaped the contemporary genetic architecture of southern African baboons. *PLoS One*, *10*, e0123207.

Sithaldeen, R., Bishop, J. M., & Ackermann, R. R. (2009). Mitochondrial DNA analysis reveals Plio-Pleistocene diversification within the chacma baboon. *Molecular Phylogenetics and Evolution*, *53*, 1042–1048.

Smith, R. J., & Jungers, W. L. (1997). Body mass in comparative primatology. *Journal of Human Evolution*, *32*, 523–559.

Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., Schlebusch, S. A., Wall, J. D., Alberts, S. C., Mukherjee, S., Zhou, X., & Tung, J. (2016). Effi-

cient genome-wide sequencing and low coverage pedigree analysis from non-invasively collected samples. *Genetics*, *203*, 699–714.

Stammbach, E. (1987). Desert, forest and montane baboons. In Smuts, B.B., Cheney, D.L., Seyfarth, R.M., Wrangham, R.W., Struhsaker, T.T. (Eds.), *Primate Societies* (pp. 112–120). Chicago: University of Chicago Press.

Stringer, C. B., & Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science*, *239*, 1263–1268.

Strum, S. C. (1987). *Almost Human: A Journey into the World of Baboons*. New York: Random House.

Sugawara, K. (1979). Sociological study of a wild group of hybrid baboons between *Papio anubis* and *P. hamadryas* in the Awash Valley, Ethiopia. *Primates*, *20*, 21–56.

Sugawara, K. (1982). Sociological comparison between two wild groups of anubis-hamadryas hybrid baboons. *African Study Monographs*, *2*, 73–131.

Sugawara, K. (1988). Ethological study of the social behavior of hybrid baboons between *Papio anubis* and *P. hamadryas* in free-ranging groups. *Primates*, *29*, 429–448.

Sussman, R. W. (2014). *The Myth of Race: The Troubling Persistence of an Unscientific Idea*. Cambridge, MA: Harvard University Press.

Swedell, L. (2011). African papionins: diversity of social organization and ecological flexibility. In Campbell, C.J., Fuentes, A., MacKinnon, K.C., Bearder, S.K., Stumpf, R.M. (Eds.), *Primates in Perspective* (pp. 241–277). New York: Oxford University Press.

Swedell, L., Saunders, J., Schreier, A., Davis, B., Tesfaye, T., & Pines, M. (2011). Female dispersal in hamadryas baboons: transfer among social units in a multilevel society. *American Journal of Physical Anthropology*, *145*, 360–370.

Szalay, F. S., & Delson, E. (Eds.). (1979). *Evolutionary History of the Primates*. New York: Academic Press.

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2005). Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, *15*, 343–355.

Templeton, A. R. (1989). The meaning of species and speciation. In Otte, D., Endler, J.A. (Eds.), *Speciation and Its Consequences* (pp. 3–27). Sunderland, MA: Sinauer Associates.

Templeton, A. R. (2005). Haplotype trees and modern human origins. *Yearbook of Physical Anthropology*, *48*, 33–59.

Templeton, A. R. (2013). Biological races in humans. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *44*, 262–271.

Tung, J., Charpentier, M. J. E., Garfield, D. A., Altmann, J., & Alberts, S. C. (2008). Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Molecular Ecology*, *17*, 1998–2011.

Van Valen, L. (1976). Ecological species, multispecies, and oaks. *Taxon*, *25*, 233.

Vernot, B., & Akey, J. M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. *Science*, *343*, 1017–1021.

Vrba, E. S. (1999). Habitat theory in relation to the evolution in African Neogene biota and hominids. In Bromage, T.G., Schrenk, F. (Eds.), *African Biogeography, Climate Change and Human Evolution* (pp. 19–39). New York: Oxford University Press.

Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Weyher, A. H. (2010). Behavioral observations of Kinda baboons (*Papio cynocephalus kindae*) in Zambia. *American Journal of Physical Anthropology*, *141 (suppl. 50)*, 241–242.

Weyher, A. H., & Chiou, K. L. (2013). Adult kinda baboon (*Papio kindae*) behavior: preliminary results from a two year study. *American Journal of Physical Anthropology*, *150 (suppl. 56)*, 289.

Weyher, A. H., Phillips-Conroy, J. E., Fourrier, M. S., & Jolly, C. J. (2014). Male-driven grooming bouts in mixed-sex dyads of Kinda baboons (*Papio kindae*). *Folia Primatologica*, *85*, 178–191.

Wildman, D. E., Bergman, T. J., al-Aghbari, A., Sterner, K. N., Newman, T. K., Phillips-Conroy, J. E., Jolly, C. J., & Disotell, T. R. (2004). Mitochondrial evidence for the origin of hamadryas baboons. *Molecular Phylogenetics and Evolution*, *32*, 287–296.

Williams, B. A., Ross, C. F., Frost, S. R., Waddle, D. M., Gabadirwe, M., & Brook, G. A. (2012). Fossil *Papio* cranium from !Ncumtsa (Koanaka) Hills, western Ngamiland, Botswana. *American Journal of Physical Anthropology*, *149*, 1–17.

Williams, F. L., Ackermann, R. R., & Leigh, S. R. (2007). Inferring Plio-Pleistocene southern African biochronology from facial affinities in *Parapapio* and other fossil papionins. *American Journal of Physical Anthropology*, *132*, 163–174.

Wirtz, P. (1999). Mother species–father species: unidirectional hybridization in animals with female choice. *Animal Behaviour*, *58*, 1–12.

Wolpoff, M. H., Wu, X., & Thorne, A. G. (1984). Modern *Homo sapiens* origins: a general theory of hominid evolution involving the fossil evidence from East Asia. In Smith, F.H., Spencer, F. (Eds.), *The Origins of Modern Humans: A World Survey of the Fossil Evidence* (pp. 411–483). New York: Alan R. Liss.

Woolley-Barker, T. (1999). *Social organization and genetic structure in a baboon hybrid zone* (PhD thesis). New York University; New York University.

Wyner, Y. M., Johnson, S. E., Stumpf, R. M., & DeSalle, R. (2002). Genetic assessment of a white-collared x red-fronted lemur hybrid zone at Andringitra, Madagascar. *American Journal of Primatology*, *57*, 51–66.

Zinner, D., Arnold, M. L., & Roos, C. (2009a). Is the new primate genus *Rungwecebus* a baboon? *PLoS One*, *4*, e4859.

Zinner, D., Arnold, M. L., & Roos, C. (2011). The strange blood: natural hybridization in primates. *Evolutionary Anthropology*, *20*, 96–103.

Zinner, D., Buba, U., Nash, S., & Roos, C. (2010). Pan-African voyagers: the phylogeography of baboons. In Sommer, V., Ross, C. (Eds.), *Primates of Gashaka* (pp. 319–358). New York: Springer.

Zinner, D., Groeneveld, L. F., Keller, C., & Roos, C. (2009b). Mitochondrial phylogeography of baboons (*Papio* spp.) - indication for introgressive hybridization? *BMC Evolutionary Biology*, *9*, 83.

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*, 133–140.

Chapter 2

Noninvasive Population Genomics from Feces

Obtaining high-quality samples from wild animals is a major obstacle for genomic studies of many taxa, particular at the population level, as collection methods for such samples are typically invasive. DNA from feces is easy to obtain noninvasively, but is dominated by a preponderance of bacterial and other non-host DNA. Because next-generation sequencing technology sequences DNA largely indiscriminately, the high proportion of exogenous DNA drastically reduces the efficiency of high-throughput sequencing for host animal genomics. In order to address this issue, we developed an inexpensive methylation-based capture method for enriching host DNA from noninvasively obtained fecal DNA samples. Our method exploits natural differences in CpG-methylation density between vertebrate and bacterial genomes to preferentially bind and isolate host DNA from majority-bacterial fecal DNA samples. We demonstrate that the enrichment is robust, efficient, and compatible with downstream library preparation methods useful for population studies (e.g., RADseq). Compared to other enrichment strategies, our method is quick and inexpensive, adding only a negligible cost to sample preparation for research that is often severely constrained by budgetary limitations. In combination with downstream methods such as RADseq, our approach allows for cost-effective and customizable genomic-scale genotyping that was previously feasible in practice only with invasive samples. Because feces are widely available and convenient to collect, our method empowers researchers to explore genomic-scale population-level questions in organisms for which invasive sampling is challenging or undesirable.

The work described in this chapter was conducted in collaboration with Christina M. Bergey and has previously been published as:

Chiou, K. L., & Bergey, C. M. (2015). FecalSeq: methylation-based enrichment for noninvasive population genomics from feces. *bioRxiv*, 032870. http://doi.org/10.1101/032870.

Introduction

The past decade has witnessed a rapid transformation of biological studies with the continuing development and implementation of massively parallel sequencing technology. This sequencing revolution, however, has thus far had a relatively muted impact on studies of wild nonmodel organisms due largely to the difficulty of obtaining high-quality samples. This problem is particularly salient for endangered animals, cryptic animals, or animals for which it is otherwise difficult, undesirable, or unethical to obtain samples invasively.

Field researchers working with nonmodel animals have explored several noninvasive sample types for DNA analysis including feces, hair, urine, saliva, feathers, skin, and nails (Kohn & Wayne, 1997). Of these, feces may be the most readily available in many taxa (Putman, 1984). Indeed, since PCR amplification of DNA from feces was first demonstrated in the 1990s (Höss et al., 1992), noninvasive genetic studies from feces have revolutionized our understanding of the evolution, population structure, phylogeography, and behavior of nonmodel organisms. PCR amplification, however, is effective only for short sequences of DNA. The ability to generate cost-effective genomic-scale data of animals from feces using massively parallel sequencing would therefore constitute an important methodological advance towards bringing nonmodel organism studies into the genomic age.

Feces presents significant challenges for genetic analysis. DNA in feces is often fragmented and low in quantity. Fecal DNA extractions are further characterized by a frequent presence of co-extracted PCR inhibitors, sometimes complicating PCR detection of genotypes (Kohn & Wayne, 1997), particularly with long amplicons. Finally, endogenous (host) DNA in feces constitutes a very low proportion, typically less than 5% (Perry et al., 2010; Qin et al., 2010; Snyder-Mackler et al., 2016), of total fecal DNA. Instead, fecal DNA contains a preponderance of DNA from exogenous (non-host) sources such as gut microbes, digesta, intestinal parasites, coprophagous animals, and other environmental organisms. Gut bacteria pose a particular challenge as they account for the highest proportion of DNA in feces (Perry et al., 2010; Qin et al., 2010).

Because of the high representation of exogenous DNA in feces, shotgun sequencing of fecal DNA

would yield only a small proportion of reads matching the host genome. For genomic studies of host organisms, particularly those targeting populations, this represents a crippling obstacle in the presence of typical financial constraints. Without an effective enrichment procedure, sequencing of fecal DNA would be less efficient than that of invasively obtained "high-quality" DNA by at least one order of magnitude regardless of improvements in sequencing throughput or cost.

Attempts to enrich host DNA from feces for genomic analysis (Perry et al., 2010; Snyder-Mackler et al., 2016) have thus far employed targeted sequence capture methodologies. Sequence capture, like PCR, enriches DNA based on sequence specificity but unlike traditional PCR can work at any scale from a single locus (Whitney et al., 2004) to a whole genome (Melnikov et al., 2011; Carpenter et al., 2013; Snyder-Mackler et al., 2016). This method involves hybridizing DNA or RNA "baits," either affixed to an array (Albert et al., 2007; Okou et al., 2007) or to magnetic beads in solution (Gnirke et al., 2009), to a mixture of target and nontarget sequences, thereby capturing targeted DNA from the mixture. Sequence capture has been used for instance to enrich human exomes (Ng et al., 2009), reduced-representation genomes (Ali et al., 2016; Hoffberg et al., 2016; Suchan et al., 2016), host DNA from ancient or museum specimens (Maricic et al., 2010; Mason et al., 2011; Bi et al., 2013; Carpenter et al., 2013), and pathogen genomes from human clinical samples (Melnikov et al., 2011). While the cost of custom oligonucleotide bait synthesis remains high, methods for transcribing custom baits from existing DNA templates (Melnikov et al., 2011; Carpenter et al., 2013) have driven costs significantly down, increasing sequence capture's appeal.

Perry et al. (2010) first successfully enriched host DNA from feces at the genomic scale. Using a modified sequence capture employing custom-synthesized baits, they were able to enrich 1.5 megabases of chromosome 21, the X chromosome, and the mitochondrial genome from fecal samples of 6 captive chimpanzees. Their protocol, however, remains prohibitively expensive for population-level analysis due to the high cost of bait synthesis. More recently, Snyder-Mackler et al. (2016) performed whole-genome capture on fecal DNA, using RNA baits transcribed *in vitro* from high-quality baboon samples to enrich host genomes from 62 wild baboons. Resulting libraries were sequenced to low coverage (mean 0.49×), but nevertheless provided sufficient information for reconstructing pedigree relationships.

Despite these methodological advances, targeted sequence capture has distinct drawbacks. To avoid the high cost of bait synthesis, RNA baits must first be transcribed from high-quality genomic DNA that is consumed by the process, limiting its appeal when working with species for which high-quality DNA is difficult to obtain or in short supply. The processes of both bait generation and hybridization with fecal DNA are labor-intensive and time-consuming, with the hybridization including an incubation step that alone takes 1 – 3 days (Snyder-Mackler et al., 2016). Because both RNA baits and the gDNA used to transcribe them are eventually depleted, the composition of RNA baits varies between bait sets, potentially impeding comparison of samples sequenced using different RNA baits and gDNA templates. *Trans* genomic captures (*i.e.* capturing DNA using baits from a different species) may complicate enrichment and introduce at least some capture biases (George et al., 2011), which will be a particular impediment for genomic studies for which high-quality DNA from related taxa is not accessible. Sequence capture may also introduce biases toward the capture of low-complexity, highly repetitive genomic regions, as well as an excess of fragments from the mitochondrial genome (Carpenter et al., 2013; Samuels et al., 2013; Snyder-Mackler et al., 2016).

The present study exploits natural, evolutionarily ancient differences in CpG-methylation densities between vertebrate and bacterial genomes to enrich the host genome from feces, making noninvasive population genomics economically and practically feasible for the first time. This method, which we call FecalSeq, uses methyl-CpG-binding domain (MBD) proteins to selectively bind and isolate DNA with high CpG-methylation density. This enrichment method is inexpensive, *de novo*, and, crucially, captures target DNA without modification, thereby enabling downstream library preparation techniques including complexity reduction-based sequencing methods such as RADseq. Because of these properties, our method is well-suited for population genomic studies requiring high sequencing coverage, including those of nonmodel organisms for which few resources (e.g., high-quality samples or reference genomes) exist.

Results

Our method is a modification of a previously described technique for enriching the microbiome from vertebrate samples containing a majority of DNA from the host organism (Feehery et al.,

2013). This technique employs a bait protein created by genetically fusing the human methyl-CpG binding domain protein 2 (MBD2) to the Fc tail of human IgG1. The resulting MBD2-Fc protein is then bound by a paramagnetic Protein A immunoprecipitation bead to create a complex that selectively binds double-stranded DNA with 5-methyl CpG dinucleotides. Because vertebrate DNA contains a high frequency of methylated CpGs (Hendrich & Tweedie, 2003; Jabbari & Bernardi, 2004) while bacterial DNA does not (Fang et al., 2012; Murray et al., 2012), this MBD bait complex selectively binds host DNA.

While Feehery et al. (2013) developed this method in order to remove contaminating host DNA for analysis of the microbiome, our strategy was to remove the contaminating fecal microbiome for analysis of host DNA. Therefore, after combining DNA with MBD baits, we retained the bound fraction with the goal of optimizing the selective recovery of host DNA (Figure 2.1). Because our aim is to genotype populations with high coverage, we used the enriched host DNA to prepare double-digest RADseq libraries (Peterson et al., 2012), although with greater sequencing investment, the method in principle should work equally well for sequencing whole genomes.

To evaluate our approach, we enriched DNA extractions from the feces of 6 captive and 46 wild baboons, which we then used to prepare and to sequence ddRADseq libraries. We also prepared ddRADseq libraries from blood-derived genomic DNA of all six captive baboons to facilitate controlled (same-individual) comparisons of blood and fecal libraries. All libraries were sequenced using Illumina sequencing.

Quantitative PCR estimates of starting host DNA proportions in fecal DNA extracts ranged widely, but were substantially lower in samples obtained from the wild (captive samples: mean 5.3%, range <0.01% – 17.4%; wild samples: mean 0.6%, range <0.01% – 4.9%; Table S2).

Based on two pilot libraries constructed from MBD-enriched fecal DNA, we found that there was large variation in the proportion of reads mapping to the baboon reference genome (mean 24.8%, range 0.7% – 81.2%; Figure S3; Table S3), with the read mapping proportion correlating with starting host DNA proportions (library A: $r^2 = 0.7338$; $p = 0.03$; library B: $r^2 = 0.9127$, $p < 0.01$). Endogenous DNA proportions on average increased 13-fold (range 4.4 – 29.6; two samples removed due to starting proportions too low to quantify).

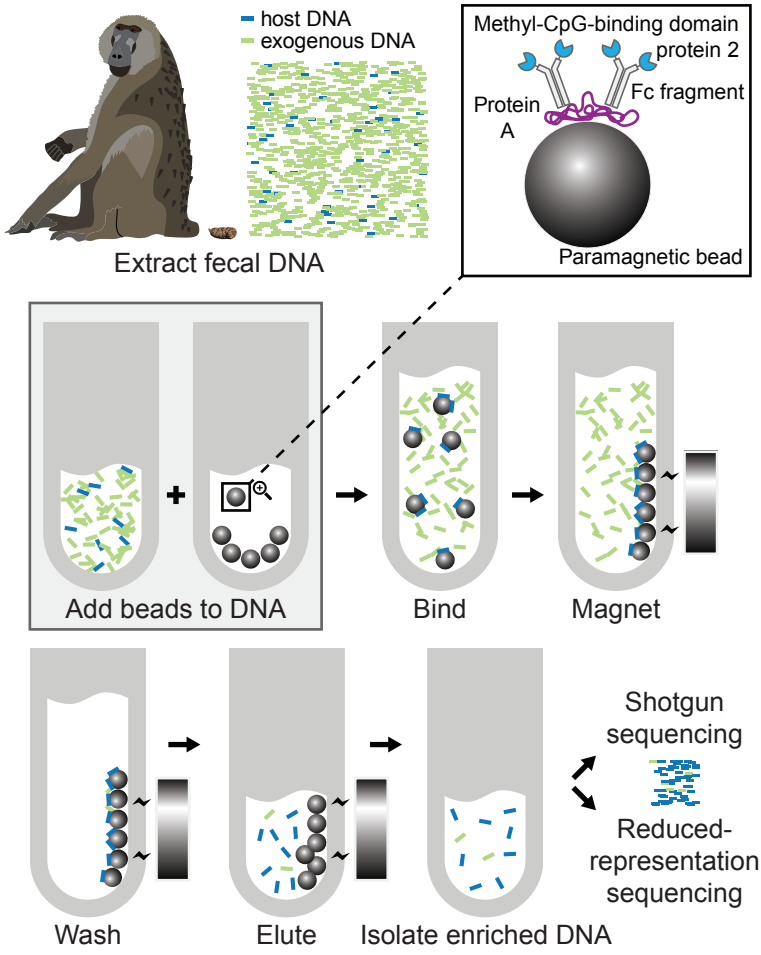

Figure 2.1: Overview of the FecalSeq method.

While some samples in our pilot libraries had high host DNA proportions following enrichment, these samples tended to already have high host DNA proportions prior to enrichment. Host DNA proportions following enrichment in the pilot libraries averaged only 4%, for instance, when samples with starting host DNA proportions greater than 1% were excluded. Because wild fecal DNA samples in our dataset on average started with less than 1% host DNA, we undertook a series of protocol optimization experiments to maximize the enrichment of these "low-quality" samples (Table S5 and Table S6).

Using a revised protocol based on our optimization experiments (Supplementary Protocol), we created and sequenced a third library from MBD-enriched fecal DNA. After noting substantial improvements in enrichment, we finally sequenced a fourth library with MBD-enriched fecal DNA from 40 wild baboons.

Despite having similar or even lower starting host DNA proportions, read mapping proportions in the third library were substantially higher than the prior two (mean 49.1%, range 8.9% – 75.3%; Figure S3; Table S3). Endogenous DNA proportions on average increased 318-fold (range 4.3 – 2632.2; one sample removed due to starting proportion too low to quantify).

The fourth library consisting entirely of fecal DNA from wild animals had the lowest starting concentrations of host DNA (mean 0.3%, range <0.01% – 3.1%). Following enrichment, however, host DNA proportions were nonetheless higher than our pilot libraries (mean 28.8%, range 1.5% – 73.6%; Figure S3; Table S3). Endogenous DNA proportions on average increased 195-fold (range 23.7 – 486.9).

Overall, the revised protocol produced substantially higher enrichment, measured as fold increases in the proportion of host DNA, particularly for samples with very low starting proportions of host DNA (Figure 2.2).While we sometimes were forced to use multiple rounds of extraction, thereby introducing variation in starting host proportions across same-individual trials, the revised protocol nonetheless exhibited robust improvement in read mapping proportions even when starting host proportions were substantially lower.

MBD binding may in principle select for genomic regions with relatively high CpG-methylation density, leading to dropout of other loci. Assessment of the concordance between blood- and feces-

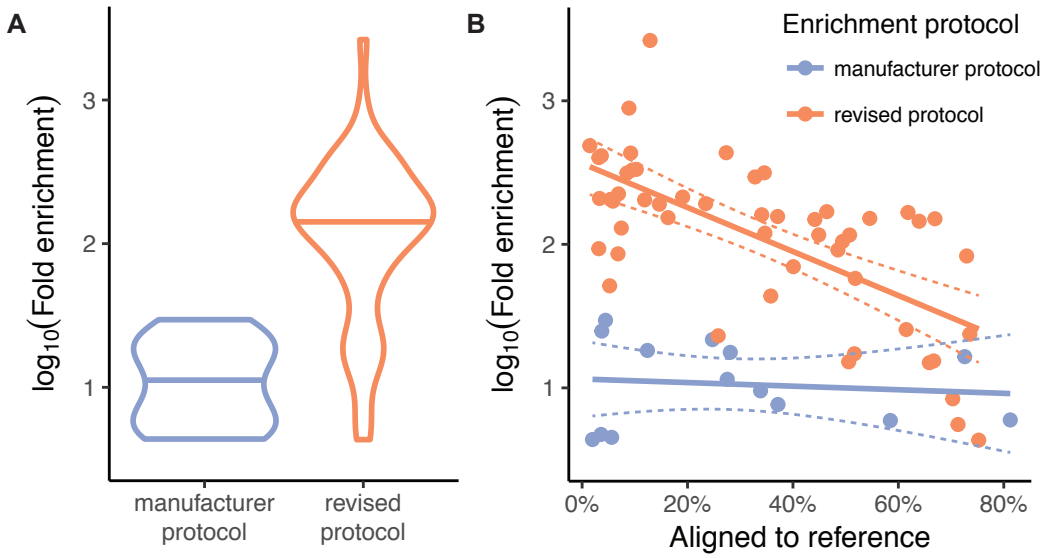

Figure 2.2: Comparison of the enrichment magnitude using the manufacturer protocol and the revised protocol. (a) Violin plots with the mean depicted show that the revised protocol results in substantially higher fold enrichment by approximately one order of magnitude. (b) A scatter plot shows that the revised protocol is particularly effective for samples with low starting quantities of host DNA. While some samples still had relatively small percentages of reads mapping to the baboon reference genome, these generally also exhibited the highest fold increases.

derived reads from the same individual was complicated by the correlation in ddRADseq between total reads and expected RADtags recovered and thereby SNPs discovered: a given RADtag is sequenced at a frequency inversely proportional to the deviation of its length from the mean of the size selection. Thus, we had to discern between dropout due to coverage-related stochasticity inherent in ddRADseq (Peterson et al., 2012) and that due to MBD enrichment. To perform this comparison, we computed the proportion of unique alleles between blood- and feces-derived RADtags from the same individual. For this test, we controlled for variation in sequencing coverage by randomly sampling reads as necessary in order to equalize total coverage among same-individual samples. Allelic dropout due to MBD enrichment would result in a higher proportion of alleles unique to blood-derived libraries relative to feces-derived libraries. We did not find a significant discrepancy (multi-sample-called SNPs: mean proportion unique alleles in blood = 2.3%, mean proportion unique alleles in feces = 2.3%; Wilcoxon signed rank test, $p = 0.97$; Figure 2.3a).

Dropout of entire RADtags is easily detectable given a reference genome or sufficient samples for comparison; dropout of a single allele at heterozygous sites is a more insidious potential bias. Allelic

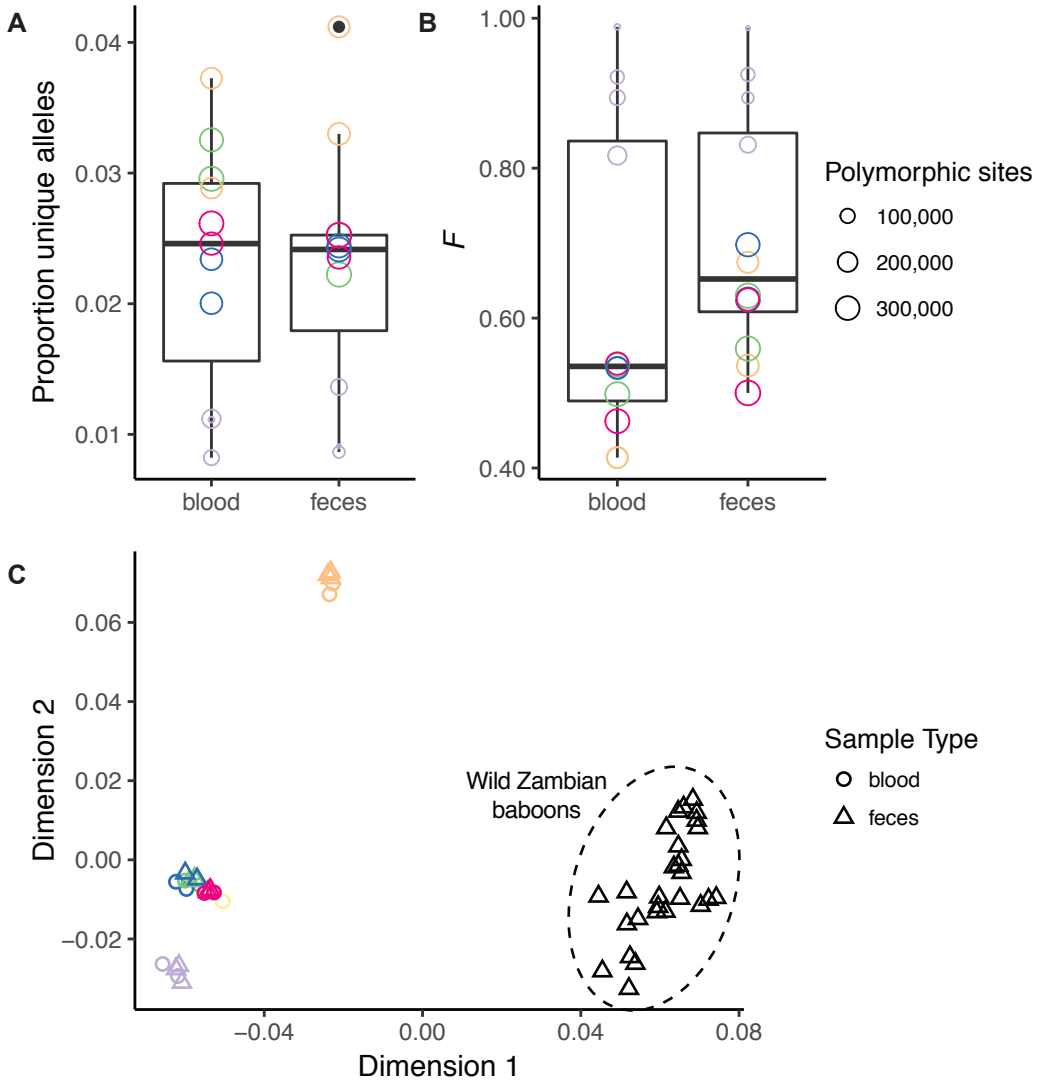

Figure 2.3: Concordance between blood- and feces-derived genotyping data from the same individuals. Colors symbolize the six captive individuals included in our study. Within these individuals, we did not find significant differences in (a) the proportion of unique alleles or (b) inbreeding coefficients from blood- and feces-derived libraries. The multidimensional scaling plot of identity-by-state shows (c) population structuring concordant with the known ancestry of animals (Table S1). Distances between feces- and blood-derived sets of genotypes from the same individual are minimal, indicating that noise added by the enrichment method is dwarfed by the population structure signal in this baboon population dataset.

dropout due to MBD enrichment would result in a decrease in heterozygosity in MBD-enriched fecal libraries. Inbreeding coefficients (F) computed from same-individual RADtags exhibited in some cases higher values for feces-derived samples (Figure 2.3b). This difference, however, was not statistically significant (mean $F_{\text{blood}} = 0.63$; mean $F_{\text{feces}} = 0.71$; Wilcoxon signed rank test, $p = 0.47$), indicating low allelic dropout attributable to the MBD enrichment. For this test, we also controlled for variation in sequencing coverage as described above.

As investigations of population structure are one potential application of our method, we visualized the wild and captive baboons' identity-by-state via multidimensional scaling (MDS) using PLINK (Purcell et al., 2007; Chang et al., 2015), and confirmed that individuals clustered by their known species or ancestry and that blood- and feces-derived reads from the same individual were close together in the MDS space (Figure 2.3c).

Stringent filtration of SNP sets, as would be implemented in a standard population genetic study, reduced the apparent biases attributable to fecal enrichment, measured both as total SNPs with a significant association with sample type (unfiltered: 25,079 out of 591,726, or 4.2%; filtered: 13 out of 7,202, or 0.2%) as well as total SNPs with significant missingness assessed via a chi-square test (unfiltered: 69,753 out of 550,224, or 12.7%; filtered: 0 out of 5,602, or 0%). Though more work is needed to quantify more exactly the extent and causal factors that lead to missingness, many population genetic analyses are robust to the low level of dropout our analyses reveal in addition to that which is inherent in the RADseq family of techniques (Gautier et al., 2013).

Discussion

Our methylation-based capture method achieves substantial enrichment of host DNA from fecal samples. Using our revised protocol developed through experimentation, we produced a mean 195-fold enrichment on our final library consisting entirely of fecal DNA obtained noninvasively under remote field conditions, with most samples nearly a decade old. A mean 28.8% of reads mapped to the baboon genome, despite starting with only a mean 0.34% of host DNA. Using fecal and blood DNA obtained from captive animals, we further demonstrate that feces-derived genotyping data following our method are concordant with corresponding data obtained from blood.

Feces are among the most readily accessible sources of information on wild animals (Kohn & Wayne, 1997), and are particularly useful for population-level studies or studies of endangered or elusive species for which obtaining high-quality samples is difficult or undesirable. By exploiting methylation differences rather than sequence differences between host and bacterial DNA, FecalSeq is a *de novo* enrichment strategy that requires neither prior genome sequence knowledge nor the use of high-quality DNA for preparation of capture baits. This enrichment is both inexpensive—we estimate a per-sample enrichment cost of $0.70 (Supplementary Note)—and replicable. The enrichment procedure is also relatively rapid and uncomplicated. Using a 96-well plate, we performed two sequential rounds of enrichment on all forty samples in our final library within a day (see Supplementary Protocol).

Importantly, FecalSeq is to our knowledge the first genomic-scale fecal DNA enrichment method that is compatible with most downstream library preparation methods for massively parallel sequencing. Through our use of ddRADseq, we demonstrate that our method facilitates low-cost high-capacity genotyping of wild populations without introducing significant bias. Further, because ddRADseq is customizable (Peterson et al., 2012), there is substantial flexibility for researchers to optimize the number of samples and the fraction of the genome sequenced for particular research questions. This is not possible for libraries prepared using targeted sequence capture, which are therefore currently limited mainly to low-coverage analyses at the population level (Snyder-Mackler et al., 2016). Transcription of sequence capture baits from reduced-representation libraries may potentially help address this problem (Ali et al., 2016; Hoffberg et al., 2016; Suchan et al., 2016), but its efficacy for fecal DNA has yet to be demonstrated.

Double-digest RADseq is possible for genotyping species with or without a reference genome. We aligned our sequencing reads to the baboon reference genome for this study, but our approach is likely also applicable to species without a reference genome. In these cases, an additional prescreening step would be necessary, in which exogenous reads are filtered out through comparison to the nearest available genome, before proceeding to clustering and variant identification as per normal reference-free ddRAD methods.

We consistently found that sequencing efficiency (percentage of reads assigned to target genome)

of MBD-enriched fecal DNA libraries correlates strongly with starting proportions of host DNA, echoing findings using other capture methods (Snyder-Mackler et al., 2016). Future attention should therefore be directed towards fecal sample collection, storage, and extraction methods that maximize the selective recovery of host nuclear DNA (e.g., Ramón-Laca et al., 2015). While we demonstrated effective genotyping of samples with often very low starting proportions of host DNA (the vast majority $< 0.5\%$), future studies may consider pre-screening extracted DNA samples using qPCR to select for samples with high starting proportions of host DNA.

Low starting proportions of host DNA present a challenge not only because they result in lower sequencing efficiency, but also because they correlate with low absolute quantities of DNA belonging to the host organism. In some cases, particularly in samples collected from wild animals under field conditions, starting proportions of host DNA were so low that only approximately 0.1 ng of target DNA was available in a 1 μg fecal DNA extract. Given the large genome sizes of vertebrates including baboons (approximately 3 Gb), substantial allelic dropout is expected in these cases. Significantly, this challenge cannot be fully addressed by this or any other enrichment method and remains an important consideration for researchers working with feces. It can be minimized, however, by optimizing the enrichment procedures to maximize the recovery of target DNA present in a fecal DNA sample, as well as by increasing the total amount of starting fecal DNA.

Because MBD enrichment partitions DNA based on CpG-methylation density, FecalSeq does not enrich hypomethylated host mitochondrial DNA (Rebelo et al., 2009). While this may be undesirable for studies requiring the matrilineally inherited marker, it also precludes the disproportionately high representation of mitochondrial DNA that is typical in libraries prepared using the targeted sequence capture approach (Perry et al., 2010; Carpenter et al., 2013; Samuels et al., 2013; Snyder-Mackler et al., 2016). FecalSeq may, however, co-enrich nuclear DNA from exogenous eukaryotes such as from plant or animal digesta. Care should therefore be taken to minimize the presence of exogenous eukaryotic tissues or cells, although the degree to which this is a problem in practice is currently unknown. As cell-wall-bound plant cells may be more likely to pass through the digestive tract intact, extraction methods that minimize lysis of cell walls should be preferred. We speculate that prey DNA from carnivorous animals may be more difficult to partition from host

DNA.

Since PCR amplification of DNA from feces was first achieved in the 1990s (Höss et al., 1992; Constable et al., 1995; Gerloff et al., 1995), noninvasive genetic studies have revolutionized our understanding of the evolution, ecology, and behavior of nonmodel organisms. By facilitating low-cost genomic-scale sequencing from feces, our method connects a community of field researchers with the benefits of massively parallel sequencing, ushering noninvasive organism studies into the genomic age.

Methods

Samples

Blood and fecal samples were collected from six captive baboons (genus *Papio*) housed at the Southwest National Primate Research Center (SNPRC) at the Texas Biomedical Research Institute. The individuals were of either *P. anubis* or hybrid ancestry (Table S1). All six baboons were fed a diet manufactured by Purina LabDiet ("Monkey Diet 15%") containing 15% minimum crude protein, 4% minimum crude fat, and 10% maximum crude fiber. In separate sedation events, blood and feces were collected from the same individual who was isolated for the duration of the sedation. Following centrifugation, the buffy coat was isolated from blood samples and stored at -80 °C. 2 ml of feces were also collected into 8 ml tubes containing 4 ml of RNAlater (Ambion). All procedures were conducted under the Texas Biomedical Research Institute IACUC protocol #1403 PC 0. Sedation and blood draws were performed under the supervision of a veterinarian and animals were returned immediately to their enclosures following recovery.

In addition, we collected or obtained fecal samples from 46 wild baboons in Zambia. Samples were collected between 2006 and 2015 from the Luangwa Valley, the Lower Zambezi National Park, Choma, or Kafue National Park and are of *P. kindae* × *P. cynocephalus*, *P. griseipes*, or *P. kindae* × *P. griseipes* ancestry (Table S1). As with the SNPRC samples, 2 ml of feces were collected into 8 ml tubes containing 4 ml of RNAlater. In contrast to the SNPRC samples, however, these samples were collected noninvasively from unhabituated animals in remote field conditions. Samples therefore

could not be attributed to particular animals, although samples were selected to avoid duplication using either field observations or geographic distance. Following collection, samples were stored without refrigeration for 1 – 6 months before being frozen at -20 °C for long-term storage.

Buffy coat extractions were performed using the QIAamp DNA Blood Mini Kit (Qiagen), following manufacturer's instructions. Fecal extractions were performed using the QIAamp DNA Stool Mini Kit (Qiagen) following manufacturer's instructions for optimizing host DNA yields. DNA concentration and yield were measured using a Qubit dsDNA BR Assay (Life Technologies). In some cases, multiple DNA extractions from the same individuals were necessary when DNA was depleted over the course of this study.

We estimated the proportion of host DNA for each fecal DNA extraction by comparing estimates of host DNA concentration obtained by quantitative PCR (qPCR) to estimates of total fecal DNA concentration obtained by Qubit. Amplification was conducted using universal mammalian *MYCBP* primers (Morin et al., 2001) and evaluated against a standard curve constructed from the liver DNA of an individual baboon obtained from SNPRC (individual #19334). Samples and standards were run in duplicate alongside positive and negative controls (see Supplementary Protocol for full details).

DNA enrichment

DNA was enriched using the NEBNext Microbiome DNA Enrichment Kit (New England BioLabs). This enrichment procedure (Feehery et al., 2013) captures eukaryotic DNA using a methylated CpG-specific binding domain protein fused to the Fc fragment of human IgG (MBD2-Fc) to selectively target sequences with high CpG methylation density.

MBD2-Fc-bound magnetic beads were prepared according to manufacturer instructions in batches ranging from 40 to 160 µl. For each n µl batch, we prebound $0.1 \times n$ µl MBD2-Fc protein to n µl protein A magnetic beads by incubating the mixture with rotation for 10 min at room temperature. The bound MBD2-Fc magnetic beads were then collected by magnet and washed twice with 1 ml ice-cold 1X bind/wash buffer before being resuspended in n µl ice-cold 1X bind/wash buffer.

As a pilot experiment, we prepared two successive libraries, library A and library B, following manufacturer's instructions for capturing methylated host DNA, with minor protocol modifications incorporated for the second pilot library (library B). Library A included MBD-enriched fecal DNA from 4 SNPRC baboons and 2 Luangwa Valley baboons, as well as blood DNA from the same SNPRC baboons. Library B included MBD-enriched fecal DNA from 4 SNPRC baboons (with two repeats from library A), 4 Kafue National Park baboons, and 2 Luangwa Valley baboons, as well as blood DNA from 2 SNPRC baboons. For each fecal DNA sample, we combined 1 – 2 μg of extracted fecal DNA with 160 μl of prepared protein-bound beads and a variable volume of ice-cold 5X bind/wash buffer for maintaining 1X concentration of bind/wash buffer. After combining beads and DNA, we incubated the mixture at room temperature with rotation for 15 min. DNA and MBD2-Fc-bound magnetic beads were then collected by magnet and the supernatant removed. For samples in library A, we washed the collected beads with 1 ml of ice-cold 1X bind/wash buffer. For samples in library B, we conducted three expanded wash steps to maximize the removal of unbound DNA. For each wash in library B, we added 1 ml of ice-cold 1X bind/wash buffer and mixed the beads on a rotating mixer for three minutes at room temperature before collecting the beads by magnet and removing the supernatant. Following the final wash, we eluted the DNA by resuspending and incubating the beads at 65 °C with 150 μL of 1X TE buffer and 15 μL of Proteinase K for 20 min with occasional mixing. The eluted DNA was then separated by magnet, purified with 1.5X homemade SPRI bead mixture (Rohland & Reich, 2012), and quantified using a Qubit dsDNA HS Assay (Life Technologies).

Our pilot sequencing results from libraries A and B revealed large variation in the percentage of reads mapping to the baboon genome, with mapping percentages ranging from 1.1% to 79.3%, with much of the variation correlating with the proportion of host DNA in the unenriched fecal DNA sample (Figure S1). To expand the utility of the enrichment protocol to all fecal DNA samples, we conducted a series of capture experiments designed to optimize the enrichment of host DNA from "low-quality" samples (i.e., samples with low proportions of host DNA). For these experiments, we artificially simulated fecal DNA by combining high-quality baboon liver or blood genomic DNA (liver: SNPRC ID #19334; blood: SNPRC ID #14068 or #25567) with *E. coli* DNA (K12 or ATCC

11303 strains) at controlled proportions. The resulting post-enrichment proportion of baboon and *E. coli* DNA was evaluated by qPCR in two analyses using (1) universal mammalian *MYCBP* (Morin et al., 2001) and (2) universal bacterial 16S rRNA (16S) (Corless et al., 2000) primers along with standards created from the same respective organisms (experiments and results are described in detail in Table S2).

We prepared a third and fourth library, libraries C and D, incorporating modifications (Supplementary Protocol) based on results from our capture optimization experiments. For these captures, we added a much smaller volume of prepared MBD2-Fc-bound magnetic beads (1 – 22 µl) based on the estimated proportion of starting host DNA, kept the capture reaction volume consistent at a relatively low 40 µl (concentrating samples as needed using a SPRI bead cleanup), added an extra wash step in which samples were resuspended in 100 µl of 1X bind/wash buffer then incubated at room temperature for 3 minutes with rotation, and eluted samples in 100 µl 2 M NaCl. For four fecal DNA samples in library C and all of library D, we serially enriched the samples by repeating the capture reaction with 30 µl of MBD-enriched DNA (post SPRI-bead cleanup). Library C included fecal DNA from 5 SNPRC baboons, 2 Kafue National Park baboons, and 1 Luangwa Valley baboon. Library D contained fecal DNA from 6 Lower Zambezi National Park baboons, 4 Choma baboons, and 30 Kafue National Park baboons.

We prepared a final library, library E, from independently extracted blood DNA from five SNPRC baboons in order to quantify the stochasticity associated with independent library preparation from independent extracts. The compositions of libraries A-E are described in detail in Table S2 and Table S3.

Library preparation and sequencing

Library preparation followed standard double-digest restriction site-associated DNA sequencing (ddRADseq) procedures (Peterson et al., 2012) with modifications to accommodate low input as described below.

For all samples, including blood DNA and MBD-enriched fecal DNA, we digested DNA with *Sph*I and *Mlu*CI (New England Biolabs), following a ratio of 1 unit of each enzyme per 20 ng

of DNA. Enzymes were diluted up to 10X using compatible diluents (New England Biolabs) to facilitate pipetting of small quantities, using an excess of enzyme if necessary to avoid pipetting less than 1 µl of the diluted enzyme mix. As the total amount of post-enrichment fecal DNA is by nature low, we adjusted adapter concentrations in the ligation reaction to ~0.1 µM for barcoded P1 and ~3 µM for P2, which correspond to excesses of adapters between $1 - 2$ orders of magnitude. Since adapter-ligated samples are multiplexed into pools in equimolar amounts, we made efforts to combine samples with similar concentrations and enrichment when known. We used the BluePippin (Sage Science) with a 1.5% agarose gel cassette for automated size selection of pooled individuals, with a target of 300 bp (including adapters) and extraction of a "tight" collection range. For PCR amplification, we ran all reactions in quadruplicate to minimize PCR biases and attempted to limit the number of PCR cycles. As the concentration of post-size-selection pools was below the limits of detection without loss of a considerable fraction of the sample, estimation of the required number of PCR cycles was difficult. We therefore iteratively quantified products post-PCR and added cycles as necessary. The total number of PCR cycles per pool is reported in Table S3, but was usually 24. Finally, libraries were sequenced using either Illumina MiSeq (libraries A-C; 2×150 paired-end) or Illumina HiSeq 2500 (library D; 2×100 paired-end) sequencing with 30% spike in of PhiX control DNA.

Analysis

We demultiplexed reads by sample and mapped them to the baboon reference genome (papAnu2; Baylor College of Medicine Human Genome Sequencing Center) using BWA with default parameters and the BWA-ALN algorithm (Li & Durbin, 2009). For every pair of blood and fecal samples from the same individual, we downsampled mapped reads to create new pairs with equal coverage in order to control for biases due to differences in sequencing depth. After realignment around indels, we identified variants using GATK UnifiedGenotyper (DePristo et al., 2011), in parallel analyses (1) calling variants in all samples at once and (2) processing each sample in isolation to avoid biasing variant calls from other samples at the expense of accuracy. Homozygous sites matching the reference genome were listed as missing when variants were inferred in single individuals. Variants

were filtered with GATK VariantFiltration (filters: `QD < 2.0`, `MQ < 40.0`, `FS > 60.0`, `HaplotypeScore > 13.0`, `MQRankSum < -12.5`, `ReadPosRankSum < -8.0`) and indels were excluded.

We digested the baboon reference genome *in silico*, tallied reads within each predicted RADtag, and gathered the following information about each region: length, GC percentage, and CpG count in region ± 5 kb. We also calculated read depth in these simulated RADtags. Distributions of blood and fecal RADtags' length, GC percentage, and local CpG density (Figure S2) were visually inspected for gross distortion due to widespread dropout.

If the fecal enrichment procedure caused widespread allelic dropout, the proportion of alleles unique to the blood samples would be higher than that to the fecal sample. We tallied these unique alleles with VCFtools (Danecek et al., 2011) and tested for an excess with a Wilcoxon signed rank test.

To quantify loss of heterozygosity due to allelic dropout, we computed the inbreeding coefficient, F for all blood-feces pairs with equalized coverage, using both the individually called and multi-sample called SNP sets. The presence of dropout is expected to inflate F. We tested for differences in paired samples' estimates of F via a Wilcoxon signed rank test. The dataset is not filtered for missingness, so sequencing errors inferred to be true variants may inflate heterozygosity estimates, thus deflating F.

To create a stringently filtered dataset with high genotyping rate, we filtered the multi-sample called SNPs in PLINK (Purcell et al., 2007; Chang et al., 2015), retaining only those genotyped in at least 90% of samples and removing samples with genotypes at fewer than 10% of sites. This filtered set was further pruned for linkage disequilibrium by sliding a window of 50 SNPs across the chromosome and removing one random SNP in each pair with $r^2 > 0.5$. Using all samples, we performed multidimensional scaling to visualize identity by state (IBS). Using just the samples that were part of the same-individual blood-feces pairs, we then performed an association test and missingness chi-square test to detect allele frequencies or missingness that correlated with sample type. We did the same with the unfiltered dataset as well. Although we had few pairs of fecal samples from the same individual, we computed distance between pairs of samples from the same individual using the stringently filtered dataset to compare distance between and within sample

types via a Wilcoxon signed rank test.

Acknowledgments

We thank Clifford Jolly and Jane Phillips-Conroy for contributing fecal samples collected in 2006 – 2007 from Zambia. We also thank the Zambia Wildlife Authority (now the Department of National Parks & Wildlife) and the University of Zambia for granting permission and providing support for fieldwork. We thank Erbay Yigit, Andrew Burrell, and Todd Disotell for helpful conversations. This study was funded by the National Science Foundation (BCS 1341018, BCS 1260816, BCS 1029302, SMA 1338524), the Leakey Foundation, the Wenner-Gren Foundation, the National Geographic Society, and the NYU University Research Challenge Fund. The Genome Technology Center at NYU is supported by NIH/NCATS UL1 TR00038 and NIH/NCI P30 CA016087. K.L.C. is supported by NSF fellowship DGE 1143954.

References

Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., Song, X., Richmond, T. A., Middle, C. M., Rodesch, M. J., Packard, C. J., Weinstock, G. M., & Gibbs, R. A. (2007). Direct selection of human genomic loci by microarray hybridization. *Nature Methods*, *4*, 903–905.

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD Capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics*, *202*, 389–400.

Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2013). Unlocking the vault: next-generation museum population genomics. *Molecular Ecology*, *22*, 6018–6032.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., Gilbert, M. T. P., Willerslev, E., Greenleaf, W. J., & Bustamante, C. D. (2013). Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American Journal of Human Genetics*, *93*, 852–864.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 559.

Constable, J. J., Packer, C., Collins, D. A., & Pusey, A. E. (1995). Nuclear DNA from primate dung. *Nature*, *373*, 393–393.

Corless, C. E., Guiver, M., Borrow, R., Edwards-Jones, V., Kaczmarski, E. B., & Fox, A. J. (2000). Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. *Journal of Clinical Microbiology*, *38*, 1747–1752.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Group, 1. G. P. A. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Angel, G. del, Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498.

Fang, G., Munera, D., Friedman, D. I., Mandlik, A., Chao, M. C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M. C., Jabado, O. J., Deikus, G., Clark, T. A., Luong, K., Murray, I. A., Davis, B. M., Keren-Paz, A., Chess, A., Roberts, R. J., Korlach, J., Turner, S. W., Kumar, V., Waldor, M. K., & Schadt, E. E. (2012). Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. *Nature Biotechnology*, *30*, 1232–1239.

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A., & Pradhan, S. (2013). A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One*, *8*, e76096.

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

George, R. D., McVicker, G., Diederich, R., Ng, S. B., MacKenzie, A. P., Swanson, W. J., Shendure, J., & Thomas, J. H. (2011). Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. *Genome Research*, *21*, 1686–1694.

Gerloff, U., Schlötterer, C., Rassmann, K., Rambold, I., Hohmann, G., Fruth, B., & Tautz, D. (1995). Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). *Molecular Ecology*, *4*, 515–518.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, *27*, 182–189.

Hendrich, B., & Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of

DNA methylation in animals. *Trends in Genetics*, *19*, 269–277.

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, *16*, 1264–1278.

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

Jabbari, K., & Bernardi, G. (2004). Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene*, *333*, 143–149.

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

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, *5*, e14004.

Mason, V. C., Li, G., Helgen, K. M., & Murphy, W. J. (2011). Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research*, *21*, 1695–1704.

Melnikov, A., Galinsky, K., Rogov, P., Fennell, T., Van Tyne, D., Russ, C., Daniels, R., Barnes, K. G., Bochicchio, J., Ndiaye, D., Sene, P. D., Wirth, D. F., Nusbaum, C., Volkman, S. K., Birren, B. W., Gnirke, A., & Neafsey, D. E. (2011). Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biology*, *12*, R73.

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

Murray, I. A., Clark, T. A., Morgan, R. D., Boitano, M., Anton, B. P., Luong, K., Fomenkov, A., Turner, S. W., Korlach, J., & Roberts, R. J. (2012). The methylomes of six bacteria. *Nucleic Acids Research*, *40*, 11450–11462.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A., & Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, *461*, 272–276.

Okou, D. T., Steinberg, K. M., Middle, C., Cutler, D. J., Albert, T. J., & Zwick, M. E. (2007). Microarray-based genomic selection for high-throughput resequencing. *Nature Methods*, *4*, 907–909.

Perry, G. H., Marioni, J. C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing

of endogenous DNA from feces. *Molecular Ecology*, *19*, 5332–5344.

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

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

Putman, R. J. (1984). Facts from faeces. *Mammal Review*, *14*, 79–97.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Pontén, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., Guchte, M. van de, Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Le Roux, K., Leclerc, M., Maguin, E., Melo Minardi, R., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., Vos, W. de, Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S. D., & Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*, 59–65.

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

Rebelo, A. P., Williams, S. L., & Moraes, C. T. (2009). *In vivo* methylation of mtDNA reveals the dynamics of protein-mtDNA interactions. *Nucleic Acids Research*, *37*, 6701–6715.

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*, 939–946.

Samuels, D. C., Han, L., Li, J., Quanghu, S., Clark, T. A., Shyr, Y., & Guo, Y. (2013). Finding the lost treasures in exome sequencing data. *Trends in Genetics*, *29*, 593–599.

Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., Schlebusch, S. A., Wall, J. D., Alberts, S. C., Mukherjee, S., Zhou, X., & Tung, J. (2016). Efficient genome-wide sequencing and low coverage pedigree analysis from non-invasively collected samples. *Genetics*, *203*, 699–714.

Suchan, T., Pitteloud, C., Gerasimova, N., Kostikova, A., Schmid, S., Arrigo, N., Pajkovic, M., Ronikier, M., & Alvarez, N. (2016). Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on museum collection specimens. *PLoS One*, *11*, e0151651.

Whitney, D., Skoletsky, J., Moore, K., Boynton, K., Kann, L., Brand, R., Syngal, S., Lawson, M., & Shuber, A. (2004). Enhanced retrieval of DNA from human fecal samples results in improved performance of colorectal cancer screening test. *Journal of Molecular Diagnostics*, *6*, 386–395.

Chapter 3

Population Signatures of Selection between Parental Species

Kinda and grayfoot baboons occupy opposite extremes of the body size distribution in extant baboons (genus *Papio*). In order to detect signatures of natural selection in these two species, we genotyped thousands of variable genome-wide autosomal markers from populations of Zambian baboons using double digest RADseq. We scanned the genome for evidence of selection by identifying regions with extreme differentiation between populations. We find evidence of selection on body size influencing multiple genes in one or both species. We also suggest that selection on body size and growth has impacted components of the JAK/STAT signaling pathway, which mediates the effect of cytokine signals on processes including epiphyseal chondrocyte proliferation essential for longitudinal bone growth.

Introduction

Body size is among the most fundamental variables influencing the diversity of biological traits in living organisms including primates (Haldane, 1926; McMahon, 1973; Martin, 1980; Fleagle, 1985). Much of the variation in primate morphology, physiology, ecology, life history, and behavior is intimately linked to variation in body size (Clutton-Brock et al., 1977; Leutenegger & Cheverud, 1982; Fleagle, 1985). While considerable attention has been dedicated to the proximate hormonal mechanisms underlying body-size evolution (Bernstein, 2010), comparatively little is known about

The work described in this chapter is the subject of a manuscript in preparation coauthored with Christina M. Bergey, Todd R. Disotell, Jeffrey Rogers, Clifford J. Jolly, and Jane E. Phillips-Conroy

the genetic basis for body-size variation in nonhuman primates, particularly at the genomic level (but see Worley et al., 2014).

Baboons (genus *Papio*) are a well-known and extensively studied genus that exhibits considerable variation in body size, ecology, morphology, and behavior (Jolly, 1993; Barrett, 2009). The impressive diversity of this genus, combined with a complex evolutionary history following a crown divergence approximately 2.0 – 1.5 million years ago (Zinner et al., 2013; Boissinot et al., 2014), have made baboons fruitful models for examining mechanisms of natural selection in wild populations (e.g., Jolly, 2001; Henzi & Barrett, 2003; Strum, 2012).

Kinda (*Papio kindae*) and gray-footed chacma baboons (*P. griseipes*), also called grayfoot baboons, are two parapatrically distributed southern African species that in some aspects occupy opposing extremes of phenotypic variation in extant baboons. Most strikingly, grayfoot baboons are one of the largest extant baboons, while Kinda baboons are the smallest (Delson et al., 2000; Jolly et al., 2011a). Adult Kinda males weigh approximately 53% of the weight of adult grayfoot males, while adult Kinda females weigh approximately 74% of the weight of adult grayfoot females. The two species differ additionally in a variety of size-correlated features, including sexual body-size dimorphism and facial length (Jolly et al., 2011a). Interestingly, these substantial body-size differences do not represent an insurmountable barrier to successful reproduction, as the two species are involved in ongoing natural hybridization in central Zambia (Jolly et al., 2011a).

In the present study, we genotype Kinda and grayfoot baboons at thousands of genome-wide polymorphic sites in order to investigate the genetic basis of divergent size-related phenotypes including body mass. We then use a genomic selection-scan approach to identify regions of the genome putatively under selection and use functional and pathway annotations to evaluate our prediction that differentiated genomic loci will show evidence of body-size adaptation.

Table 3.1: Samples included in this analysis. Only samples that passed quality and ancestry filters are listed in this table. An expanded list including sample IDs is provided in Table S8.

Locality	Latitude	Longitude	Year	Ancestry	Tissue type	N
Chunga	-15.0446°	25.9994°	2011	Kinda	blood	69
North Nkala Road	-15.9029°	25.8895°	2007	grayfoot	feces	1
Ngoma Airstrip	-15.9695°	25.9376°	2012	grayfoot	blood	4
Dendro Park	-16.1518°	26.0599°	2012	grayfoot	blood	4
Nanzhila Plains	-16.2337°	25.9612°	2007	grayfoot	feces	3
Choma	-16.6383°	27.0308°	2007	grayfoot	feces	3
Lower Zambezi National Park	-15.9234°	28.9466°	2006	grayfoot	feces	5

Methods

Samples

Genetic material from wild Zambian baboons was derived from blood samples collected between 2011 and 2012 and from fecal samples collected between 2006 and 2007 (Table 3.1). Sampling was concentrated around Kafue National Park, Zambia, where Kinda and grayfoot baboon distributions join and overlap with evidence of hybridization (Jolly et al., 2011a). Blood was drawn from animals that were trapped, tranquilized, then released following established protocols (Brett et al., 1977; Jolly et al., 2011b). We collected whole blood into evacuated tubes containing EDTA anticoagulant, then fractionated the blood by centrifugation. Plasma, erythrocyte-, and leukocyte-rich fractions were stored immediately in liquid nitrogen and subsequently at -80 °C. A small subset of whole blood was preserved on a Whatman FTA card and stored at room temperature. Feces were obtained noninvasively during field surveys and stored in RNAlater (Ambion) at room temperature, then subsequently at -20 °C. All trapping and sample collection procedures were conducted following local laws and guidelines in Zambia and with approval from the Animal Studies Committees at Washington University, New York University, and Baylor College of Medicine.

DNA isolation and library preparation

We sequenced DNA from a total of 129 animals, of which 89 remained after filtering for quality and ancestry as described in the following sections. For blood samples, we extracted DNA from stored leukocyte ($n = 45$), plasma ($n = 16$), or FTA-dried blood spot samples ($n = 26$) using the

QIAamp DNA Blood Mini Kit (Qiagen). For fecal samples, we extracted DNA using the QIAamp DNA Stool Mini Kit (Qiagen).

Because fecal DNA is dominated by a high presence of exogenous, mainly microbial DNA, we enriched host DNA from fecal samples using the FecalSeq method (Chiou & Bergey, 2015; Chapter 2), which makes use of differences in CpG methylation between vertebrate and bacterial genomes to capture relatively heavily CpG-methylated host DNA from feces. For each sample, we added 150 – 1000 ng of DNA to 1 µl of prepared MBD2-Fc/protein A beads (New England Biolabs), then washed and eluted the DNA as described by Chiou and Bergey (2015). To maximize the fraction of host DNA recovered, we performed two serial rounds of enrichment for each fecal DNA sample.

In order to sequence a representative and reproducible set of orthologous SNPs, we prepared multiplexed double-digest RADseq (ddRADseq) libraries (Peterson et al., 2012) using *Sph*I and *Mlu*CI. If the input DNA quantity was less than 20 ng, as is typically the case with enriched fecal DNA, we used 1 unit of each restriction enzyme. Otherwise, we used 1 – 10 units of each restriction enzyme following a ratio of one unit of enzyme per 20 ng of input DNA. The remainder of the ddRADseq library preparation followed standard ddRADseq procedures with modifications for low quantities when necessary (Chiou & Bergey, 2015). We sequenced finished libraries on the Illumina HiSeq 2500 or Illumina HiSeq 4000 platforms.

Sequence alignment and variant identification

Sequencing reads were demultiplexed using sabre (https://github.com/najoshi/sabre), allowing one mismatch, and adapter sequences were trimmed using fastq-mcf (arguments: `-l 15 -q 15 -w 4 -u -P 33`) (Aronesty, 2011). We then aligned reads to the baboon reference genome (papAnu2) using default settings of the Burrows-Wheeler Aligner BWA-MEM algorithm (Li & Durbin, 2009; Li, 2013a). After realignment around indels using the Genome Analysis Toolkit (GATK) (McKenna et al., 2010), we called variants on the complete dataset including all individuals using GATK UnifiedGenotyper (arguments: `-stand_call_conf 50.0 -stand_emit_conf 10.0`) and filtered variant calls using GATK VariantFiltration (filters: `QD < 2.0`, `MQ < 40.0`, `FS > 60.0`, `HaplotypeScore >`

`13.0`, `MQRankSum < -12.5`, `ReadPosRankSum < -8.0`).

Autosomal variants that passed the filters described above were further filtered to exclude variants in repetitive regions and variants with substantial missingness. Variants in repetitive regions were filtered out based on publicly available identifications obtained from UCSC Genome Browser (Meyer et al., 2013) and generated using RepeatMasker (Smit et al., 2015) and Tandem Repeats Finder (Benson, 1999) for the baboon reference genome (papAnu2). After filtering out repetitive regions, 590,666 variants remained with a genotyping rate of 23.1%. We then filtered out variants missing in $\geq 20\%$ of individuals and removed individuals with $\geq 80\%$ missing data using PLINK (`--geno 0.2 --mind 0.8`) (Purcell et al., 2007; Chang et al., 2015). After applying these filters, 24,790 variants and 123 individuals remained, with a genotyping rate of 88.8%.

For ancestry analysis, we created an additionally pruned dataset by removing variants in strong linkage disequilibrium (Bergey et al., 2016). Variants in strong linkage disequilibrium were identified based on sliding windows of 50 SNPs with a step size of 5 SNPs and an r^2 threshold of 0.5 using PLINK (`--indep-pairwise 50 5 0.5`) (Purcell et al., 2007; Chang et al., 2015). The final pruned dataset contained 14,642 variants.

Inference of ancestry

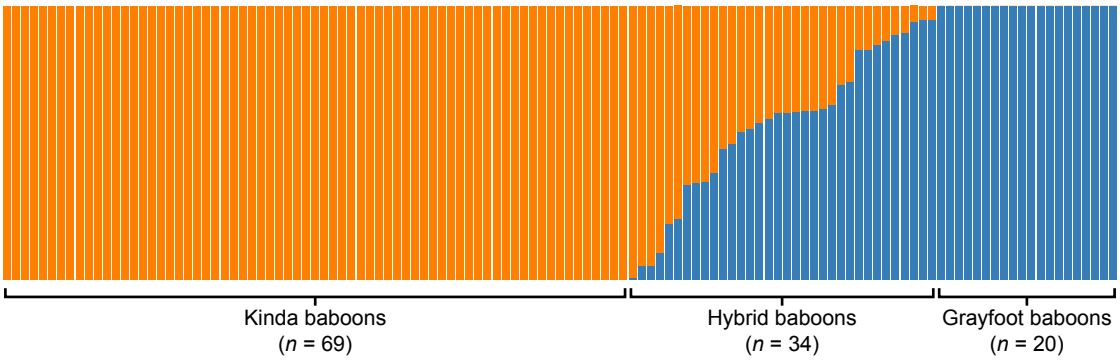

Figure 3.1: Taxonomic assignment based on ancestry estimates calculated using ADMIXTURE.

Because the focus of this analysis is on between-species differences and samples were collected from an area containing known hybrids, we took the following steps to remove individuals with

mixed ancestry from this analysis. We used an unsupervised clustering algorithm implemented in ADMIXTURE to estimate the ancestry of each individual by maximum likelihood (Alexander et al., 2009). For this analysis, we used the pruned dataset of 14,642 variants. We explored through cross validation a range of K from 1 to 10 and found that $K = 2$ corresponded best to distinguishing between Kinda and grayfoot baboon genetic backgrounds (Figure S4). We considered individuals to be of pure ancestry if their ancestry estimates inferred by ADMIXTURE were greater than 0.999. By this criterion, 34 individuals of mixed ancestry were excluded from subsequent analysis. Our final dataset contained 69 pure Kinda baboons and 20 pure grayfoot baboons (Figure 3.1).

Estimation of differentiation

Differential selection on populations leads to a significant difference in allele frequencies between populations. The well-known fixation index (F_{ST}) quantifies this pattern by comparing the variance of allele frequencies within and between populations. Comparatively large values of F_{ST} at a locus relative to neutral regions indicate a larger differentiation between populations. Regions with extremely high values of F_{ST} indicate the putative effects of directional selection and are therefore considered to be candidates for selection (Akey et al., 2002).

We first used VCFtools (Danecek et al., 2011) to calculate F_{ST}, following the equation of Weir and Cockerham (Weir & Cockerham, 1984; Cockerham & Weir, 1987), for all 24,790 SNP loci in our dataset. Next, based on the baboon reference genome (papAnu2) general transfer format (GTF) annotation file obtained from Ensembl (Cunningham et al., 2015), we assigned a single value of F_{ST} to each gene calculated as the average F_{ST} of all SNPs falling within the boundaries of a protein-coding gene $\pm$ 50 bp.

We used a permutation test to assess the significance of F_{ST} estimates for each gene. In each iteration, we randomly reassigned F_{ST} values to SNPs by sampling with replacement from all F_{ST} values in the SNP dataset. F_{ST} values were then assigned to genes following the procedure described above. Sampling was conducted first for 10,000 replicates. A $p_{F_{ST}}$ value was then calculated as the proportion of F_{ST} values in the randomized samples that were greater than the actual F_{ST} value for each gene. If significance ($p < 0.05$) could not be confidently determined after 10,000 replicates,

we allowed the analysis for remaining genes to continue to a maximum of 100 million replicates, after which all $p_{F_{ST}}$ values could be confidently distinguished from a significance threshold of $\alpha = 0.05$. This procedure increases computational efficiency by focusing computational effort on genes whose significance is more difficult to determine.

Functional enrichment analysis

Rather than test for an effect of one or a few genes with significant differentiation, enrichment analysis tests for an overall pattern of differentiation in a collection of genes classified according to known functions or roles in biological pathways. In other words, enrichment analysis tests for categories with an overall pattern of significant differentiation, even in cases in which no single gene in that category may be individually identified as statistically significant. Enrichment analyses are therefore useful for detecting cases in which sizeable phenotypic differences result from subtle directional selection on a large panel of genes, as is the case with well-known characteristics such as body height (Lango Allen et al., 2010).

In order to assign functional information to protein-coding genes, we assigned Gene Ontology (GO) annotations (Gene Ontology Consortium, 2000, 2015) to protein-coding genes based on UniProt identifications (UniProt Consortium, 2008; Magrane & UniProt Consortium, 2011). First, using the baboon reference genome coding DNA sequences (CDS) downloaded from Ensembl (Cunningham et al., 2015), we calculated open reading frames (ORFs) for all 20 autosomal chromosomes using EMBOSS `getorf` (Rice et al., 2000). Next, we assigned UniProt identifications to translated baboon ORF amino acid sequences by scanning sequences against the baboon reference proteome using HMMER3 `phmmer` (Finn et al., 2011). We excluded matches with E-values > 0.0001. For each ORF with remaining matches, we assigned a single protein with the lowest E-value. We then programmatically assigned GO functional classifications to protein matches via HTTP/REST using the UniProt API (Jain et al., 2009). Because genes may contain multiple ORFs, it was possible for genes to be assigned the same GO classifications more than once. Therefore, to avoid redundancy and thereby overrepresentation of functional classifications, we excluded duplicate assignments of GO terms for all genes.

We also assigned gene pathways to protein-coding genes using the PANTHER pathway database (Mi et al., 2013a, 2013b). Because baboon gene sets are not available in PANTHER, we first assigned baboon proteins to rhesus macaque proteins by homology using HMMER3 `phmmer` (Finn et al., 2011). As with the procedure described above for the baboon reference proteome, we excluded matches with E-values > 0.0001 and assigned a single protein with the lowest E-value. We then assigned PANTHER pathways based on UniProt identifications in the complete PANTHER sequence classifications for the rhesus macaque proteome. As with the GO analysis, we excluded duplicate assignment of PANTHER pathways to genes in order to avoid redundancy.

We used enrichment analyses to identify GO classifications or PANTHER pathways associated with more differentiated genes. For these analyses, we excluded genes lacking F_{ST} values and genes lacking either GO classifications or PANTHER pathway annotations.

Enriched classes are defined as classes that have significantly extreme F_{ST} or $p_{F_{ST}}$ values. Enrichment tests were conducted on all GO classes or PANTHER pathways using a nonparametric one-sided Wilcoxon rank sum test, evaluating the alternative hypothesis that F_{ST} or $p_{F_{ST}}$ values of genes assigned to a class or pathway were higher or lower, respectively, than F_{ST} or $p_{F_{ST}}$ values not assigned to that class or pathway. GO classes or PANTHER pathways with a p-value < 0.05 were considered enriched. Due to the exploratory nature of these analyses, correction for multiple comparisons was not conducted.

Results and Discussion

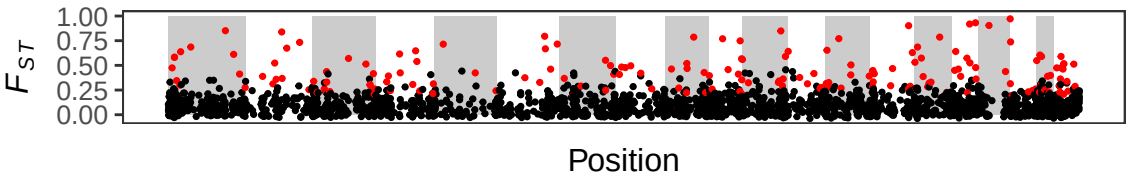

Figure 3.2: Distribution of F_{ST} for protein-coding genes. Genes identified as significant by permutation are displayed in red. Background shading indicates the position of the autosomal chromosomes in order.

Using samples from 89 wild Zambian baboons, we genotyped thousands of SNPs and used a genome-wide selection scan approach based on F_{ST} (Akey et al., 2002) to identify candidate genes

showing signatures of high differentiation between populations (Figure 3.2). Using annotation information from the baboon reference genome, we assigned a mean F_{ST} to genes based on values for all SNPs falling within or near the boundaries of a gene. Out of 27,857 genes in the baboon reference genome, we thus derived a sample of 2,213 genes, or approximately 8% of all genes. We identified through permutation 175 candidate genes with extremely high F_{ST} ($p_{F_{ST}} < 0.05$; Table S9). After controlling for multiple comparisons using the false discovery rate (FDR) (Benjamini & Hochberg, 1995), we identified 10 genes considered to be especially strong candidates for differentiation (FDR = 0.05) (Table 3.2).

Table 3.2: Genes that are significantly differentiated between Kinda and grayfoot baboon populations based on F_{ST}. Genes listed here were significant after controlling for multiple comparisons (FDR = 0.05) and are therefore considered to be strong candidates for differentiation. Asterisk (*) indicates that the gene name was not available in the baboon reference genome annotations and was instead inferred from the macaque reference genome through homology as described in the text.

Gene	Gene name	F_{ST}	$p_{F_{ST}}$	BH-adjusted $p_{F_{ST}}$
ATXN2	Ataxin 2	0.84882	0.00003	0.01660
*CTRC**	Chymotrypsin-C	0.58200	< 0.00001	< 0.00001
EFHD2	EF-hand domain family member D2	0.58200	< 0.00001	< 0.00001
FGF1	Fibroblast growth factor 1	0.66769	0.00004	0.01770
PRKCE	Protein kinase C epsilon	0.26969	0.00016	0.03934
RFX4	Regulatory factor X4	0.34189	0.00013	0.03596
UMODL1	Uromodulin-like 1	0.33847	0.00020	0.04426
Novel protein-coding	ENSPANG00000004311	0.74849	0.00012	0.03596
Novel protein-coding	ENSPANG00000027984	0.65318	0.00010	0.03596
Novel protein-coding	ENSPANG00000028262	0.61153	0.00001	0.00738

3 of the 10 genes are novel protein-coding genes lacking information about function. We assessed the remaining genes for functions that are potentially relevant in producing the morphologically and behaviorally divergent Kinda and grayfoot baboon phenotypes.

Several genes have known functions related to developmental or metabolic processes. One gene, fibroblast growth factor 1 (*FGF1*), promotes cell division in a variety of mesoderm- and neuroectoderm-derived cells, and also promotes angiogenesis (Chiu et al., 1990; Nabel et al., 1993). It is also involved, along with other fibroblast growth factors, in the induction of organs including limbs (Cohn et al., 1995). Ataxin 2 (*ATXN2*) has been associated with weight gain and insulin resistance in mice (Kiehl et al., 2006; Lastres-Becker et al., 2008) and hypertension and diabetes mellitus in humans (Levy et al., 2009; Lastres-Becker et al., 2016). Chymotrypsin-C (*CTRC*) is

a regulator of digestive enzyme activation and degradation (Batra et al., 2013). Protein kinase C epsilon (*PRKCE*) has been reported to play a role in regulating the secretion of insulin in response to glucose signals in pancreatic β cells (Warwar et al., 2008), although evidence for its role is contradictory (Schmitz-Peiffer et al., 2007). *PRKCE* also functions in neuron growth (Fagerström et al., 1996). The broader protein kinase C (PKC) family is known to play a key role in the transduction of fibroblast growth factor (FGF) signals related to osteoblast differentiation and limb development (Lu et al., 2001; Kim et al., 2003).

Two genes are relatively little-known but appear to play a role in sex-specific developmental processes. Uromodulin-like 1 (*UMODL1*) plays a role in the regulation of ovarian development in mice (Wang et al., 2012). Regulatory factor X4 (*RFX4*) is highly expressed in the central nervous system in mice and may play a role in regulating ciliogenesis by modulating Sonic hedgehog (Shh) signaling during development (Ashique et al., 2009).

Interestingly, *RFX4* in humans is expressed entirely in the testis (Morotomi-Yano et al., 2002). Testicular morphology and physiology have a well-known association with taxon-specific social behavior and levels of sperm competition in mammals (Harcourt et al., 1981; Møller, 1988, 1989; Dixson & Anderson, 2004) and baboons in particular (Jolly & Phillips-Conroy, 2003, 2006). Given the high levels of dimorphism between Kinda and grayfoot baboons in both size and behavior, differentiation of *RFX4* suggests a possible proximate mechanism. Closer inspection of SNPs in *RFX4* indicates that, in the most highly differentiated SNP located on chromosome 11 at position 105851669 on the baboon reference genome (papAnu2), the high-frequency allele in grayfoot baboons is shared with both anubis baboon and macaque outgroups (Table S10), indicating that the high differentiation of *RFX4* may be driven by directional selection on Kinda baboons. These interpretations remain speculative, however, in the absence of additional sequence and expression information from *RFX4*, as well as functional validation.

Two genes have been linked to pathways related to behavior. EF-hand domain family member D2 (*EFHD2*) plays a role in immune and brain cell function (Dütting et al., 2011). While it is best-known for its role in tau-related neurodegenerative diseases including Alzheimer's disease (Vega, 2016), it has also been implicated in pathways associated with aggression-related behavior (Malki

et al., 2014). *PRKCE* knockout mice show reduced anxiety-like behaviors and reduced levels of the stress hormones corticosterone and adrenocorticotrophic hormone (Hodge et al., 2002).

In order to analyze the gene functions associated with highly differentiated genes, we conducted enrichment analysis using Gene Ontology (GO) classifications (Gene Ontology Consortium, 2000, 2015). These analyses identified 62 enriched GO classes (Table S11), indicating that they were associated with higher genome-wide differentiation than expected by chance ($p < 0.05$ for both F_{ST} and $p_{F_{ST}}$). Because correction for multiple comparisons (Benjamini & Hochberg, 1995) was impractical due to insufficient statistical power, these results are considered preliminary but highlight avenues for future study.

The GO analyses revealed the significant enrichment of two terms broadly related to organism growth, multicellular organism growth (GO:0035264) and developmental growth (GO:0048589), which are associated in this dataset with nonoverlapping sets of genes (Table S12). Multicellular organism growth is the increase in size or mass of an entire organism rather than individual cells, while developmental growth is the increase in size or mass associated with a temporal progression across life stages. Given the extreme differences in size and developmental trajectories between Kinda and grayfoot baboons, the enrichment of both of these terms is intriguing and suggests that species differences in growth and development have a broad genetic basis involving directional selection on multiple genes.

Other interesting enriched GO terms include skeletal muscle fiber development (GO:0048741), fat cell differentiation (GO:0045444), regulation of brown fat cell differentiation (GO:0090335), multicellular organismal response to stress (GO:0033555), response to heat (GO:0009408), response to cold (GO:0009409), stem cell development (GO:0048864), somatic stem cell population maintenance (GO:0035019), organ induction (GO:0001759), corticotropin hormone secreting cell differentiation (GO:0060128), thyroid-stimulating hormone-secreting cell differentiation (GO:0060129), labyrinthine layer blood vessel development (GO:0060716), sperm motility (GO:0030317), fertilization (GO:0009566), and regulation of dopamine metabolic process (GO:0042053). In some cases, a single gene contributed to the enrichment of multiple GO terms, including *LRP11* (multicellular organismal response to stress, response to heat, response to cold) and *TTLL5* (sperm motility,

fertilization). The previously-mentioned candidate gene *FGF1* single-handedly accounted for the enrichment of four GO terms (stem cell development, organ induction, corticotropin hormone secreting cell differentiation, thyroid-stimulating hormone-secreting cell differentiation), reflecting its diverse functions. Additional caution should be therefore be exercised to avoid overinterpreting enrichment results for these terms. A list of genes associated with these terms is included in Table S12.

The enrichment of skeletal muscle fiber development (GO:0048741), fat cell differentiation (GO:0045444), and regulation of brown fat cell differentiation (GO:0090335) are potentially relevant to body-size differences between Kinda and grayfoot baboons as well as differences in body build. Kinda baboons, for instance, exhibit a lanky, gracile appearance while grayfoot baboons exhibit a more robust build (Jolly et al., 2011a). Given a paucity of data, however, the degree to which these characteristics correspond to differences in musculature or adiposity is intriguing but at present unclear.

The enrichment of regulation of dopamine metabolic process (GO:0042053) is also intriguing given previous findings that suggest that selection on the organization of dopamine-mediated reward pathways in the central nervous system underpins species-specific behavioral phenotypes in two other baboon species that differ markedly in their morphology and behavior (Jolly et al., 2008, 2013; Bergey et al., 2016). Dopamine has been broadly linked to impulsivity and reward-seeking behaviors in animals (Barron et al., 2010), and the ratio between dopamine and serotonin metabolites in cerebrospinal fluid has been found to differ (Jolly et al., 2008) between adult male hamadryas baboons, who follow and attempt to form a close exclusive bond with any unattached female, regardless of her reproductive state, and adult male anubis baboons, who exhibit interest in and compete for access to females only when the latter are periovulatory. A possible proximate relationship between the dopamine pathway and species-specific behaviors in Kinda and grayfoot baboons is less clear, but may be related to divergent grooming strategies. Unusually among baboons, male Kinda baboons appear to initiate and maintain grooming bouts across all female reproductive conditions (Weyher et al., 2014). In grayfoot baboons, male-female grooming is largely confined to the consorting context, when the female is periovulatory (Wittig et al., 2008). Future scrutiny of

Table 3.3: Pathways with enriched differentiation. Pathways that are enriched exhibit an overall shift in F_{ST} or $p_{F_{ST}}$. We considered enrichment of pathways to be significant if we found $p < 0.05$ for both F_{ST} and $p_{F_{ST}}$. Because only one pathway (JAK/STAT signaling pathway) met that criterion, pathways with $p < 0.05$ for either F_{ST} or $p_{F_{ST}}$ are also shown here.

PANTHER pathway		Components		Enrichment (p-value)	
Accession	Name	In dataset	Total known	F_{ST}	$p_{F_{ST}}$
P00038	JAK/STAT signaling pathway	3	8	0.03736	0.04720
P00026	Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	8	27	0.03955	0.08948
P06959	CCKR signaling map	16	290	0.06934	0.03410

the neurophysiology of the two species, as well as a more detailed understanding of Kinda baboon behavior, will help to clarify this relationship, which remains speculative in the absence of a larger genomic dataset and correction for multiple comparisons.

We also assessed pathway annotations using the PANTHER classification system (Mi et al., 2013a, 2013b) to identify pathways with enriched differentiation between Kinda and grayfoot baboons. These analyses identified a single enriched PANTHER pathway (Table 3.3), the JAK/STAT signaling pathway (P00038), indicating that it was associated with higher genome-wide differentiation than expected by chance ($p < 0.05$ for both F_{ST} and $p_{F_{ST}}$). As with the GO enrichment analysis, correction for multiple comparisons (Benjamini & Hochberg, 1995) was not conducted for this exploratory analysis.

The JAK/STAT signaling pathway represents a major cell-signaling mechanism, originating prior to the divergence of protostomes and deuterostomes (Liongue & Ward, 2013). Cellular responses to myriad hormones, growth factors, and cytokines are mediated by this evolutionarily conserved pathway (Kisseleva et al., 2002). JAK/STAT signaling is capable of eliciting specialized responses to extracellular signals that depend on the signal molecule and the tissue or cellular context. Predictably, mutations that affect JAK/STAT signaling, for instance by suppressing its components or by constitutively activating or failing to regulate its activity, also affect these processes and thereby influence organismal phenotypes (O'Shea et al., 2002; Rawlings et al., 2004).

The JAK/STAT pathway consists of proteins in the Janus kinase (JAK) and Signal Transducer and Activator of Transcription (STAT) families. In the canonical pathway, intracellular JAKs are

first activated by the binding of a ligand, usually a cytokine, to a cell-surface receptor. Activated JAKs then recruit STAT monomers by phosphorylating the receptor. The STAT proteins are also phosphorylated by JAK, causing them to dimerize. The STAT dimers finally detach from the receptor as fully formed transcription factors and travel to the nucleus, where they bind to DNA promoters and activate transcription of target genes. Altogether, four JAKs (JAK1, JAK2, JAK3, Tyk2) and seven STATs (STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, STAT6) are known. The JAK/STAT signaling cascade can be influenced by other components, including suppressor of cytokine signaling (SOCS) proteins, which inhibit JAKs, and protein inhibitor of activated STAT (PIAS) proteins, which inhibit STATs.

While JAK/STAT signaling underlies diverse responses including mammary gland development, hematopoiesis, and immune cell development (Watson & Burdon, 1996; Ward et al., 2000; Ghoreschi et al., 2009), the statistical enrichment of this PANTHER pathway is particularly intriguing due to its known function in organism growth. Knockout experiments in mouse models demonstrate the impacts of multiple JAKs and STATs on bone development and body mass (Li, 2013b). JAK1-deficient mice, for instance, die perinatally but are associated with significantly smaller body masses relative to their littermates (Rodig et al., 1998). The selective inactivation of STAT3 in osteoblasts is also associated with lower body mass and bone mass (Zhou et al., 2011). JAK/STAT signaling additionally has profound impacts on adipose tissue development and physiology (Richard & Stephens, 2014), although as previously mentioned, the relevance of these functions is at present unclear due to a paucity of morphometric data on adiposity.

Notably, JAK/STAT is known to serve as a signaling mechanism for the GH/IGF1 axis, which is one of the major endocrine systems regulating somatic growth and stature (Perry & Dominy, 2009). Growth hormone (GH) secreted in the anterior pituitary stimulates the production of insulin-like growth factor 1 (IGF1) through JAK/STAT signals (Zezulak & Green, 1986). Together, GH and IGF1 play essential roles in regulating bone growth through chondrocyte proliferation at epiphyseal growth plates (Nilsson et al., 1994; Vottero et al., 2013). These functions can be impacted through modifications to JAK/STAT signaling. Suppressor of cytokine signaling 2 (SOCS2), for instance, is a negative regulator of GH action through inhibition of JAK/STAT, and SOCS2 deficiency in mice

Table 3.4: Genotype frequencies for all components and genes in the JAK/STAT signaling pathway with identified variants in our dataset. Genome positions and reference alleles are given for the anubis baboon (papAnu2) genome. The reference allele for the rhesus macaque (rheMac8) genome was also obtained using a coordinate translation in liftOver (Kent et al., 2002; Rosenbloom et al., 2015). Where the reference alleles differed between genomes, both alleles are listed with the baboon allele listed first. Genotype frequencies are shown for each species in the order: homozygous for reference, heterozygous, homozygous for alternate.

Component	Gene	Position (papAnu2)		Differentiation		Allelic variants		Genotype frequencies					
				F_{ST}	$p_{F_{ST}}$	Ref.	Alt.	Kinda			grayfoot		
JAK	*JAK1*	chr1	66951070	0.07908	0.14740	A	C	65	0	0	18	0	1
		chr1	66951074	0.07908		C	A	65	0	0	18	0	1
		chr1	66951078	0.07908		C	G	65	0	0	18	0	1
		chr1	66951082	0.59910		T/C	C	0	3	62	6	7	6
PIAS	*PIAS1*	chr7	42548519	0.39690	0.06619	A	T	54	12	2	4	4	5
	PIAS4	chr19	3844945	0.21996	0.17450	C	T	67	0	0	12	3	0
STAT	*STAT3*	chr16	48997534	0.11930	0.36770	C	T	59	0	0	14	2	0

is associated with gigantism and increased longitudinal bone growth (Metcalf et al., 2000; Pass et al., 2012).

JAK/STAT function on chondrocyte proliferation has also been shown to interact with fibroblast growth hormone (FGF) signaling. FGF1, which was identified in this study as a strong candidate for differentiation as discussed above, has been shown to suppress chondrocyte proliferation in mouse and rat cells through induction of STAT1, which has anti-proliferative functions (Sahni et al., 1999).

Out of four genes in our dataset involved in the JAK/STAT pathway, none were significantly differentiated despite enrichment of the overall pathway for both F_{ST} and $p_{F_{ST}}$ (Table 3.4). The differentiation of *PIAS1* approached significance ($p_{F_{ST}} = 0.06619$). The most differentiated SNP, however, was found in *JAK1* on chromosome 1 at position 66951082 on the baboon reference genome. Interestingly, at this SNP, the alternate allele is nearly fixed in Kinda baboons and shared with a rhesus macaque outgroup, while the anubis baboon reference allele is found at high frequency in grayfoot baboons (Table 3.4). Both *JAK1*- and *PIAS1*-knockout mice are associated with significantly smaller perinatal body masses (Rodig et al., 1998; Liu et al., 2004). These results suggest a possible role for JAK/STAT signaling in divergent body size phenotypes of Kinda and grayfoot baboons. The mechanisms and the genes involved, however, remain unknown.

Conclusion

Using genome-wide polymorphism data from Zambian baboons occupying opposite extremes of the extant baboon size distribution, we were able to identify putatively selected genes that may underlie divergent phenotypes in the two species, particularly those related to body size. Our study offers the first glimpse into signatures of selection at the population genomic scale in these wild baboon species, and one of the first in baboons in general (see also Bergey et al., 2016). Due to novel advances in high-throughput genotyping from noninvasive fecal samples (Chiou & Bergey, 2015; Chapter 2), our study is also the first population genomic scan incorporating a substantial number of wild individuals sampled noninvasively. Our study therefore demonstrates the feasibility and efficacy of noninvasive techniques for addressing genomic questions in wild populations of nonmodel species (see also Snyder-Mackler et al., 2016).

Out of ten genes considered to be strongly differentiated between wild Kinda and grayfoot baboons, several are known to be involved in growth or metabolic processes and may contribute to differences in body size. Of these, *FGF1* is probably the strongest candidate given its function as a potent mitogen and its role in cell proliferation and tissue development, including its effects on chondrocyte proliferation activity that underlies organism growth. Through functional enrichment analyses using GO terms, we found robust evidence that terms related to body size differences, such as multicellular organism growth and development growth, were enriched with high F_{ST}, suggesting that body size differences have a robust genetic basis involving multiple genes. Through analysis of biological pathways using PANTHER classifications, we found that only one pathway, the JAK/STAT signaling pathway, was enriched with high F_{ST}. JAK/STAT signaling serves diverse functions, but is strongly linked to salient pathways underlying organism growth.

Additional data are needed to incorporate a higher degree of sequence information and to assess functional impacts of coding SNPs. These data will enable the generation of more detailed hypotheses that can subsequently be evaluated using functional experimentation. Assessment of additional gene sequences, for instance using whole-genome shotgun sequencing or exome capture methodologies, will incorporate genes that were not represented in our reduced-representation genome dataset.

Apart from the highly likely possibility of finding additional candidate genes, these data will also enable more detailed and higher-power enrichment analyses, which would for instance allow for corrections for multiple comparisons that were not practical for this study. Despite methodological limitations of our study, however, the consistent detection of evidence for growth-related genes using multiple complementary analyses demonstrates a robust ability to detect a genetic basis for the substantial body size differences between these two baboon species.

Acknowledgments

We thank the Zambia Wildlife Authority (now the Department of National Parks & Wildlife) and the University of Zambia for granting permission and providing support for fieldwork. We also thank Andrew Burrell for helpful conversations and advice. This study was funded by the National Science Foundation (BCS 1341018, BCS 1029302, SMA 1338524), the Leakey Foundation, and the National Geographic Society. The Genome Technology Center at NYU is supported by NIH/NCATS UL1 TR00038 and NIH/NCI P30 CA016087.

References

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, *12*, 1805–1814.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664.

Aronesty, E. (2011). *ea-utils: command-line tools for processing biological sequencing data*. https://github.com/ExpressionAnalysis/ea-utils.

Ashique, A. M., Choe, Y., Karlen, M., May, S. R., Phamluong, K., Solloway, M. J., Ericson, J., & Peterson, A. S. (2009). The Rfx4 transcription factor modulates shh signaling by regional control of ciliogenesis. *Science Signaling*, *2*, ra70.

Barrett, L. (2009). A guide to practical babooning: historical, social, and cognitive contingency. *Evolutionary Anthropology*, *18*, 91–102.

Barron, A. B., Søvik, E., & Cornish, J. L. (2010). The roles of dopamine and related compounds

in reward-seeking behavior across animal phyla. *Frontiers in Behavioral Neuroscience*, *4*, 163.

Batra, J., Szabó, A., Caulfield, T. R., Soares, A. S., Sahin-Tóth, M., & Radisky, E. S. (2013). Long-range electrostatic complementarity governs substrate recognition by human chymotrypsin C, a key regulator of digestive enzyme activation. *Journal of Biological Chemistry*, *288*, 9848–9859.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, *27*, 573–580.

Bergey, C. M., Phillips-Conroy, J. E., Disotell R, T., & Jolly, C. J. (2016). Dopamine pathway is highly diverged in primate species that differ markedly in social behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 6178–6181.

Bernstein, R. M. (2010). The big and small of it: how body size evolves. *Yearbook of Physical Anthropology*, *53*, 46–62.

Boissinot, S., Alvarez, L., Giraldo-Ramirez, J., & Tollis, M. (2014). Neutral nuclear variation in baboons (genus *Papio*) provides insights into their evolutionary and demographic histories. *American Journal of Physical Anthropology*, *155*, 621–634.

Brett, F. L., Jolly, C. J., Socha, W., & Wiener, A. S. (1977). Human-like ABO blood groups in wild Ethiopian baboons. *Yearbook of Physical Anthropology*, *20*, 276–289.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 559.

Chiou, K. L., & Bergey, C. M. (2015). FecalSeq: methylation-based enrichment for noninvasive population genomics from feces. *bioRxiv*, 032870. doi:10.1101/032870.

Chiu, I.-M., Wang, W.-P., & Lehtoma, K. (1990). Alternative splicing generates two forms of mRNA coding for human heparin-binding growth factor 1. *Oncogene*, *5*, 755–762.

Clutton-Brock, T. H., Harvey, P. H., & Rudder, B. (1977). Sexual dimorphism, socionomic sex ratio and body weight in primates. *Nature*, *269*, 797–800.

Cockerham, C. C., & Weir, B. S. (1987). Correlations, descent measures: drift with migration and mutation. *Proceedings of the National Academy of Sciences of the United States of America*, *84*, 8512–8514.

Cohn, M. J., Izpisúa-Belmonte, J. C., Abud, H., Heath, J. K., & Tickle, C. (1995). Fibroblast growth factors induce additional limb development from the flank of chick embryos. *Cell*, *80*, 739–746.

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D.,

Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Kähäri, A. K., Keenan, S., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Aken, B. L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S. M. J., Spudich, G., Trevanion, S. J., Yates, A., Zerbino, D. R., & Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research*, *43 (suppl. 1)*, D662–D669.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Group, 1. G. P. A. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

Delson, E., Terranova, C. J., Jungers, W. L., Sargis, E. J., Jablonski, N. G., & Dechow, P. C. (2000). Body mass in Cercopithecidae (Primates, Mammalia): estimation and scaling in extinct and extant taxa. *Anthropological Papers of the American Museum of Natural History*, *83*, 1–159.

Dixson, A. F., & Anderson, M. J. (2004). Sexual behavior, reproductive physiology and sperm competition in male mammals. *Physiology & Behavior*, *83*, 361–371.

Dütting, S., Brachs, S., & Mielenz, D. (2011). Fraternal twins: Swiprosin-1/EFhd2 and Swiprosin-2/EFhd1, two homologous EF-hand containing calcium binding adaptor proteins with distinct functions. *Cell Communication and Signaling*, *9*, 2.

Fagerström, S., Påhlman, S., Gestblom, C., & Nånberg, E. (1996). Protein kinase C-ϵ is implicated in neurite outgrowth in differentiating human neuroblastoma cells. *Cell Growth & Differentiation*, *7*, 775–785.

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39 (suppl. 2)*, W29–W37.

Fleagle, J. G. (1985). Size and adaptation in primates. In Jungers, W.L. (Ed.), *Size and Scaling in Primate Biology* (pp. 1–19). New York: Plenum Press.

Gene Ontology Consortium. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*, 25–29.

Gene Ontology Consortium. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Research*, *43 (suppl. 1)*, D1049–D1056.

Ghoreschi, K., Laurence, A., & O'Shea, J. J. (2009). Janus kinases in immune cell signaling. *Immunological Reviews*, *228*, 273–287.

Haldane, J. B. S. (1926). On being the right size. *Harper's Magazine*, *152*, 424–427.

Harcourt, A. H., Harvey, P. H., Larson, S. G., & Short, R. V. (1981). Testis weight, body weight and breeding system in primates. *Nature*, *293*, 55–57.

Henzi, S. P., & Barrett, L. (2003). Evolutionary ecology, sexual conflict, and behavioral differenti-

ation among baboon populations. *Evolutionary Anthropology*, *12*, 217–230.

Hodge, C. W., Raber, J., McMahon, T., Walter, H., Sanchez-Perez, A. M., Olive, M. F., Mehmert, K., Morrow, A. L., & Messing, R. O. (2002). Decreased anxiety-like behavior, reduced stress hormones, and neurosteroid supersensitivity in mice lacking protein kinase Cε. *The Journal of Clinical Investigation*, *110*, 1003–1010.

Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E., Martin, M. J., McGarvey, P., & Gasteiger, E. (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, *10*, 136.

Jolly, C. J. (1993). Species, subspecies, and baboon systematics. In Kimbel, W.H., Martin, L.B. (Eds.), *Species, Species Concepts, and Primate Evolution* (pp. 67–107). New York: Plenum Press.

Jolly, C. J. (2001). A proper study for mankind: analogies from the papionin monkeys and their implications for human evolution. *Yearbook of Physical Anthropology*, *44*, 177–204.

Jolly, C. J., & Phillips-Conroy, J. E. (2003). Testicular size, mating system, and maturation schedules in wild anubis and hamadryas baboons. *International Journal of Primatology*, *24*, 125–142.

Jolly, C. J., & Phillips-Conroy, J. E. (2006). Testicular size, developmental trajectories, and male life history strategies in four baboon taxa. In Swedell, L., Leigh, S.R. (Eds.), *Reproduction and Fitness in Baboons: Behavioral, Ecological, and Life History Perspectives* (pp. 257–275). New York: Springer.

Jolly, C. J., Burrell, A. S., Phillips-Conroy, J. E., Bergey, C., & Rogers, J. (2011a). Kinda baboons (*Papio kindae*) and grayfoot chacma baboons (*P. ursinus griseipes*) hybridize in the Kafue river valley, Zambia. *American Journal of Primatology*, *73*, 291–303.

Jolly, C. J., Phillips-Conroy, J. E., & Müller, A. E. (2011b). Trapping primates. In Setchell, J.M., Curtis, D.J. (Eds.), *Field and Laboratory Methods in Primatology* (pp. 133–145). New York: Cambridge University Press.

Jolly, C. J., Phillips-Conroy, J. E., Kaplan, J. R., & Mann, J. J. (2008). Cerebrospinal fluid monoaminergic metabolites in wild *Papio anubis* and *P. hamadryas* are concordant with taxon-specific behavioral ontogeny. *International Journal of Primatology*, *29*, 1549–1566.

Jolly, C. J., Phillips-Conroy, J. E., Kaplan, J. R., & Mann, J. J. (2013). Monoamine neurotransmitter metabolites in the cerebrospinal fluid of a group of hybrid baboons (*Papio hamadryas* $\times$ *P. anubis*). *International Journal of Primatology*, *34*, 836–858.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, *12*, 996–1006.

Kiehl, T.-R., Nechiporuk, A., Figueroa, K. P., Keating, M. T., Huynh, D. P., & Pulst, S.-M. (2006). Generation and characterization of Sca2 (ataxin-2) knockout mice. *Biochemical and Biophysical*

Research Communications, *339*, 17–24.

Kim, J.-H., Bae, S.-C., Choi, J.-Y., Kim, H.-J., & Ryoo, H.-M. (2003). The protein kinase C pathway plays a central role in the fibroblast growth factor-stimulated expression and transactivation activity of Runx2. *Journal of Biological Chemistry*, *278*, 319–326.

Kisseleva, T., Bhattacharya, S., Braunstein, J., & Schindler, C. W. (2002). Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene*, *285*, 1–24.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A. R., Weyant, R. J., Segrè, A. V., Speliotes, E. K., Wheeler, E., Soranzo, N., Park, J.-H., Yang, J., Gudbjartsson, D., Heard-Costa, N. L., Randall, J. C., Qi, L., Vernon Smith, A., Mägi, R., Pastinen, T., Liang, L., Heid, I. M., Luan, J., Thorleifsson, G., Winkler, T. W., Goddard, M. E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y. S., Johansson, A., Zillikens, M. C., Feitosa, M. F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N. L., Hayward, C., Hottenga, J.-J., Jacobs, K. B., Knowles, J. W., Kutalik, Z., Monda, K. L., Polasek, O., Preuss, M., Rayner, N. W., Robertson, N. R., Steinthorsdottir, V., Tyrer, J. P., Voight, B. F., Wiklund, F., Xu, J., Zhao, J. H., Nyholt, D. R., Pellikka, N., Perola, M., Perry, J. R. B., Surakka, I., Tammesoo, M.-L., Altmaier, E. L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D. I., Chen, C., Coin, L., Cooper, M. N., Dixon, A. L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L. M., Kettunen, J., König, I. R., Kwan, T., Lawrence, R. W., Levinson, D. F., Lorentzon, M., McKnight, B., Morris, A. P., Müller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R. M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J. R., Turchin, M. C., Vandenput, L., Verlaan, D. J., Vitart, V., White, C. C., Ziegler, A., Almgren, P., Balmforth, A. J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J. G., Freimer, N. B., Geus, E. J. C., Glorioso, N., Haiqing, S., Hartikainen, A.-L., Havulinna, A. S., Hicks, A. A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpeläinen, T. O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.-L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Paré, G., Parker, A. N., Peden, J. F., Petersmann, A., Pichler, I., Pietiläinen, K. H., Pouta, A., Ridderstråle, M., Rotter, J. I., Sambrook, J. G., Sanders, A. R., Schmidt, C. O., Sinisalo, J., Smit, J. H., Stringham, H. M., Bragi Walters, G., Widen, E., Wild, S. H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W. L., Nelis, M., Peters, M. J., Ripatti, S., Meurs, J. B. J. van, Aben, K. K., Ardlie, K. G., Beckmann, J. S., Beilby, J. P., Bergman, R. N., Bergmann, S., Collins, F. S., Cusi, D., Heijer, M. den, Eiriksdottir, G., Gejman, P. V., Hall, A. S., Hamsten, A., Huikuri, H. V., Iribarren, C., Kähönen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L. J., Lehtimäki, T., Melander, O., Mosley, T. H., Musk, A. W., Nieminen, M. S., O'Donnell, C. J., Ohlsson, C., Oostra, B., Palmer, L. J., Raitakari, O., Ridker, P. M., Rioux, J. D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A. R., Siscovick, D. S., Stumvoll, M., Tönjes, A., Tuomilehto, J., Ommen, G.-J. van, Viikari, J., Heath, A. C., Martin, N. G., Montgomery, G. W., Province, M. A., Kayser, M., Arnold, A. M., Atwood, L. D., Boerwinkle, E., Chanock, S. J., Deloukas, P., Gieger, C., Grönberg, H., Hall, P., Hattersley, A. T., Hengstenberg, C., Hoffman, W., Lathrop, G. M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A. F., Assimes, T. L., Barroso, I., Hofman, A., Mohlke, K. L., Boomsma, D. I., Caulfield, M. J., Cupples, L. A., Erdmann, J., Fox, C. S., Gudnason,

V., Gyllensten, U., Harris, T. B., Hayes, R. B., Jarvelin, M.-R., Mooser, V., Munroe, P. B., Ouwehand, W. H., Penninx, B. W., Pramstaller, P. P., Quertermous, T., Rudan, I., Samani, N. J., Spector, T. D., Völzke, H., Watkins, H., Wilson, J. F., Groop, L. C., Haritunians, T., Hu, F. B., Kaplan, R. C., Metspalu, A., North, K. E., Schlessinger, D., Wareham, N. J., Hunter, D. J., O'Connell, J. R., Strachan, D. P., Wichmann, H.-E., Borecki, I. B., Duijn, C. M. van, Schadt, E. E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A. G., Visscher, P. M., Chatterjee, N., Loos, R. J. F., Boehnke, M., McCarthy, M. I., Ingelsson, E., Lindgren, C. M., Abecasis, G. R., Stefansson, K., Frayling, T. M., & Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*, 832–838.

Lastres-Becker, I., Brodesser, S., Lütjohann, D., Azizov, M., Buchmann, J., Hintermann, E., Sandhoff, K., Schürmann, A., Nowock, J., & Auburger, G. (2008). Insulin receptor and lipid metabolism pathology in ataxin-2 knock-out mice. *Human Molecular Genetics*, *17*, 1465–1481.

Lastres-Becker, I., Nonis, D., Eich, F., Klinkenberg, M., Gorospe, M., Kötter, P., Klein, F. A. C., Kedersha, N., & Auburger, G. (2016). Mammalian ataxin-2 modulates translation control at the pre-initiation complex via PI3K/mTOR and is induced by starvation. *Biochimica et Biophysica Acta*, *1862*, 1558–1569.

Leutenegger, W., & Cheverud, J. (1982). Correlates of sexual dimorphism in primates: ecological and size variables. *International Journal of Primatology*, *3*, 387–402.

Levy, D., Ehret, G. B., Rice, K., Verwoert, G. C., Launer, L. J., Dehghan, A., Glazer, N. L., Morrison, A. C., Johnson, A. D., Aspelund, T., Aulchenko, Y., Lumley, T., Köttgen, A., Vasan, R. S., Rivadeneira, F., Eiriksdottir, G., Guo, X., Arking, D. E., Mitchell, G. F., Mattace-Raso, F. U. S., Smith, A. V., Taylor, K., Scharpf, R. B., Hwang, S.-J., Sijbrands, E. J. G., Bis, J., Harris, T. B., Ganesh, S. K., O'Donnell, C. J., Hofman, A., Rotter, J. I., Coresh, J., Benjamin, E. J., Uitterlinden, A. G., Heiss, G., Fox, C. S., Witteman, J. C. M., Boerwinkle, E., Wang, T. J., Gudnason, V., Larson, M. G., Chakravarti, A., Psaty, B. M., & Duijn, C. M. van. (2009). Genome-wide association study of blood pressure and hypertension. *Nature Genetics*, *41*, 677–687.

Li, H. (2013a). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303.3997.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Li, J. (2013b). JAK-STAT and bone metabolism. *JAK-STAT*, *2*, e23930.

Liongue, C., & Ward, A. C. (2013). Evolution of the JAK-STAT pathway. *JAK-STAT*, *2*, e22756.

Liu, B., Mink, S., Wong, K. A., Stein, N., Getman, C., Dempsey, P. W., Wu, H., & Shuai, K. (2004). PIAS1 selectively inhibits interferon-inducible genes and is important in innate immunity. *Nature Immunology*, *5*, 891–898.

Lu, H.-C., Swindell, E. C., Sierralta, W. D., Eichele, G., & Thaller, C. (2001). Evidence for a role

of protein kinase C in FGF signal transduction in the developing chick limb bud. *Development*, *128*, 2451–2460.

Magrane, M., & UniProt Consortium. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database*, *2011*, bar009.

Malki, K., Pain, O., Du Rietz, E., Tosto, M. G., Paya-Cano, J., Sandnabba, K. N., Boer, S. de, Schalkwyk, L. C., & Sluyter, F. (2014). Genes and gene networks implicated in aggression related behaviour. *Neurogenetics*, *15*, 255–266.

Martin, R. D. (1980). Adaptation and body size in primates. *Zeitschrift für Morphologie und Anthropologie*, *71*, 115–124.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303.

McMahon, T. (1973). Size and shape in biology. *Science*, *179*, 1201–1204.

Metcalf, D., Greenhalgh, C. J., Viney, E., Willson, T. A., Starr, R., Nicola, N. A., Hilton, D. J., & Alexander, W. S. (2000). Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature*, *405*, 1069–1073.

Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., Sloan, C. A., Rosenbloom, K. R., Roe, G., Rhead, B., Raney, B. J., Pohl, A., Malladi, V. S., Li, C. H., Lee, B. T., Learned, K., Kirkup, V., Hsu, F., Heitner, S., Harte, R. A., Haeussler, M., Guruvadoo, L., Goldman, M., Giardine, B. M., Fujita, P. A., Dreszer, T. R., Diekhans, M., Cline, M. S., Clawson, H., Barber, G. P., Haussler, D., & Kent, W. J. (2013). The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Research*, *41 (suppl. 1)*, D64–D69.

Mi, H., Muruganujan, A., & Thomas, P. D. (2013a). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, *41 (suppl. 1)*, D377–D386.

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013b). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, *8*, 1551–1566.

Morotomi-Yano, K., Yano, K.-i., Saito, H., Sun, Z., Iwama, A., & Miki, Y. (2002). Human regulatory factor X 4 (RFX4) is a testis-specific dimeric DNA-binding protein that cooperates with other human RFX members. *Journal of Biological Chemistry*, *277*, 836–842.

Møller, A. P. (1988). Ejaculate quality, testes size and sperm competition in primates. *Journal of Human Evolution*, *17*, 479–488.

Møller, A. P. (1989). Ejaculate quality, testes size and sperm production in mammals. *Functional Ecology*, *3*, 91–96.

Nabel, E. G., Yang, Z.-y., Plautz, G., Forough, R., Zhan, X., Haudenschild, C. C., Maciag, T., & Nabel, G. J. (1993). Recombinant fibroblast growth factor-1 promotes intimal hyperplasia and angiogenesis in arteries *in vivo*. *Nature*, *362*, 844–846.

Nilsson, A., Ohlsson, C., Isaksson, O. G., Lindahl, A., & Isgaard, J. (1994). Hormonal regulation of longitudinal bone growth. *European Journal of Clinical Nutrition*, *48 (suppl. 1)*, S150–S158.

O'Shea, J. J., Gadina, M., & Schreiber, R. D. (2002). Cytokine signaling in 2002: new surprises in the Jak/Stat pathway. *Cell*, *109*, S121–S131.

Pass, C., MacRae, V. E., Huesa, C., Faisal Ahmed, S., & Farquharson, C. (2012). SOCS2 is the critical regulator of GH action in murine growth plate chondrogenesis. *Journal of Bone and Mineral Research*, *27*, 1055–1066.

Perry, G. H., & Dominy, N. J. (2009). Evolution of the human pygmy phenotype. *Trends in Ecology & Evolution*, *24*, 218–225.

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

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

Rawlings, J. S., Rosler, K. M., & Harrison, D. A. (2004). The JAK/STAT signaling pathway. *Journal of Cell Science*, *117*, 1281–1283.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, *16*, 276–277.

Richard, A. J., & Stephens, J. M. (2014). The role of JAK–STAT signaling in adipose tissue function. *Biochimica et Biophysica Acta*, *1842*, 431–439.

Rodig, S. J., Meraz, M. A., White, J. M., Lampe, P. A., Riley, J. K., Arthur, C. D., King, K. L., Sheehan, K. C. F., Yin, L., Pennica, D., Johnson Jr., E. M., & Schreiber, R. D. (1998). Disruption of the *Jak1* gene demonstrates obligatory and nonredundant roles of the Jaks in cytokine-induced biologic responses. *Cell*, *93*, 373–383.

Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., Harte, R. A., Heitner, S., Hickey, G., Hinrichs, A. S., Hubley, R., Karolchik, D., Learned, K., Lee, B. T., Li, C. H., Miga, K. H., Nguyen, N., Paten, B., Raney, B. J., Smit, A. F. A., Speir, M. L., Zweig, A. S., Haussler, D., Kuhn, R. M., & Kent, W. J. (2015). The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research*, *43 (suppl. 1)*, D670–D681.

Sahni, M., Ambrosetti, D.-C., Mansukhani, A., Gertner, R., Levy, D., & Basilico, C. (1999). FGF

signaling inhibits chondrocyte proliferation and regulates bone development through the STAT-1 pathway. *Genes & Development*, *13*, 1361–1366.

Schmitz-Peiffer, C., Laybutt, D. R., Burchfield, J. G., Gurisik, E., Narasimhan, S., Mitchell, C. J., Pedersen, D. J., Braun, U., Cooney, G. J., Leitges, M., & Biden, T. J. (2007). Inhibition of PKCe improves glucose-stimulated insulin secretion and reduces insulin clearance. *Cell Metabolism*, *6*, 320–328.

Smit, A. F. A., Hubley, R., & Green, P. (2015). *RepeatMasker 4.0.6*. http://repeatmasker.org.

Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., Schlebusch, S. A., Wall, J. D., Alberts, S. C., Mukherjee, S., Zhou, X., & Tung, J. (2016). Efficient genome-wide sequencing and low coverage pedigree analysis from non-invasively collected samples. *Genetics*, *203*, 699–714.

Strum, S. C. (2012). Darwin's monkey: why baboons can't become human. *Yearbook of Physical Anthropology*, *55*, 3–23.

UniProt Consortium. (2008). The universal protein resource (UniProt). *Nucleic Acids Research*, *36 (suppl. 1)*, D190–D195.

Vega, I. E. (2016). EFhd2, a protein linked to Alzheimer's disease and other neurological disorders. *Frontiers in Neuroscience*, *10*, 285.

Vottero, A., Guzzetti, C., & Loche, S. (2013). New aspects of the physiology of the GH-IGF-1 axis. *Endocrine Development*, *24*, 96–105.

Wang, W., Tang, Y., Ni, L., Kim, E., Jongwutiwes, T., Hourvitz, A., Zhang, R., Xiong, H., Liu, H.-C., & Rosenwaks, Z. (2012). Overexpression of Uromodulin-like1 accelerates follicle depletion and subsequent ovarian degeneration. *Cell Death and Disease*, *3*, e433.

Ward, A. C., Touw, I., & Yoshimura, A. (2000). The Jak-Stat pathway in normal and perturbed hematopoiesis. *Blood*, *95*, 19–29.

Warwar, N., Dov, A., Abramovitch, E., Wu, R., Jmoudiak, M., Haber, E., Cerasi, E., & Nesher, R. (2008). PKCepsilon mediates glucose-regulated insulin production in pancreatic beta-cells. *Biochimica et Biophysica Acta*, *1783*, 1929–1934.

Watson, C. J., & Burdon, T. G. (1996). Prolactin signal transduction mechanisms in the mammary gland: the role of the Jak/Stat pathway. *Reviews of Reproduction*, *1*, 1–5.

Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Weyher, A. H., Phillips-Conroy, J. E., Fourrier, M. S., & Jolly, C. J. (2014). Male-driven grooming bouts in mixed-sex dyads of Kinda baboons (*Papio kindae*). *Folia Primatologica*, *85*, 178–191.

Wittig, R. M., Crockford, C., Lehmann, J., Whitten, P. L., Seyfarth, R. M., & Cheney, D. L.

(2008). Focused grooming networks and stress alleviation in wild female baboons. *Hormones and Behavior*, *54*, 170–177.

Worley, K. C., Warren, W. C., Rogers, J., Locke, D., Muzny, D. M., Mardis, E. R., Weinstock, G. M., Tardif, S. D., Aagaard, K. M., Archidiacono, N., Rayan, N. A., Batzer, M. A., Beal, K., Brejova, B., Capozzi, O., Capuano, S. B., Casola, C., Chandrabose, M. M., Cree, A., Dao, M. D., Jong, P. J. de, Rosario, R. C.-H. del, Delehaunty, K. D., Dinh, H. H., Eichler, E. E., Fitzgerald, S., Flicek, P., Fontenot, C. C., Fowler, R. G., Fronick, C., Fulton, L. A., Fulton, R. S., Gabisi, R. A., Gerlach, D., Graves, T. A., Gunaratne, P. H., Hahn, M. W., Haig, D., Han, Y., Harris, R. A., Herrero, J., Hillier, L. W., Hubley, R., Hughes, J. F., Hume, J., Jhangiani, S. N., Jorde, L. B., Joshi, V., Karakor, E., Konkel, M. K., Kosiol, C., Kovar, C. L., Kriventseva, E. V., Lee, S. L., Lewis, L. R., Liu, Y.-S., Lopez, J., López-Otín, C., Lorente-Galdos, B., Mansfield, K. G., Marques-Bonet, T., Minx, P., Misceo, D., Moncrieff, J. S., Morgan, M. B., Nazareth, L. V., Newsham, I., Nguyen, N. B., Okwuonu, G. O., Prabhakar, S., Perales, L., Pu, L.-L., Puente, X. S., Quesada, V., Ranck, M. C., Raney, B. J., Raveendran, M., Deiros, D. R., Rocchi, M., Rodriguez, D., Ross, C., Ruffier, M., Ruiz, S. J., Sajjadian, S., Santibanez, J., Schrider, D. R., Searle, S., Skaletsky, H., Soibam, B., Smit, A. F. A., Tennakoon, J. B., Tomaska, L., Ullmer, B., Vejnar, C. E., Ventura, M., Vilella, A. J., Vinar, T., Vogel, J.-H., Walker, J. A., Wang, Q., Warner, C. M., Wildman, D. E., Witherspoon, D. J., Wright, R. A., Wu, Y., Xiao, W., Xing, J., Zdobnov, E. M., Zhu, B., Gibbs, R. A., & Wilson, R. K. (2014). The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics*, *46*, 850–857.

Zezulak, K. M., & Green, H. (1986). The generation of insulin-like growth factor-1–sensitive cells by growth hormone action. *Science*, *233*, 551–553.

Zhou, H., Newnum, A. B., Martin, J. R., Li, P., Nelson, M. T., Moh, A., Fu, X.-Y., Yokota, H., & Li, J. (2011). Osteoblast/osteocyte-specific inactivation of Stat3 decreases load-driven bone formation and accumulates reactive oxygen species. *Bone*, *49*, 404–411.

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*, 133–140.

Chapter 4

Population Genomic Analysis of the Hybrid Zone

Hybridization in nature offers unique insights into the process of natural selection in incipient species and their hybrids. In order to evaluate the patterns and actions of selection at work, we examined a recently discovered baboon hybrid zone in the Kafue river valley of Zambia, where Kinda baboons (*Papio kindae*) and grayfoot baboons (*Papio griseipes*) coexist with hybridization. We genotyped hundreds of baboons at thousands of variable genome-wide autosomal markers using double-digest RADseq. We compare ancestry patterns from this genome-wide dataset to previously reported mitochondrial-DNA and Y-chromosome ancestry patterns. In order to detect signatures of selection in the hybrid zone, we employ Bayesian methods to scan the genome for genes with extreme patterns of local adaptation or introgression. We find that the Kinda baboon Y chromosome has penetrated the species boundary to a greater extent than either mitochondrial DNA or the autosomal chromosomes. We find little evidence for a substantial effect of local adaptation on genotype frequencies in the hybrid zone. We find evidence for restricted introgression of genes in the JAK/STAT signaling pathway, indicating a role as a reproductive barrier potentially related to extreme differences in body size between parental species. We also find evidence for high introgression of toll-like receptor pathway components as well as of the sperm tail gene *ODF2*, both of which also favor grayfoot baboon variants. We suggest that, in the context of widespread hybrid dysgenesis of sperm traits, the Kinda Y chromosome may have been favored by selection due to its rescuing effect on sperm quality in hybrids.

The work described in this chapter is the subject of a manuscript in preparation coauthored with Christina M. Bergey, Andrew S. Burrell, Todd R. Disotell, Jeffrey Rogers, Clifford J. Jolly, and Jane E. Phillips-Conroy

Introduction

Among many organisms including primates, divergence with reticulate evolution is extremely common to the degree that in many lineages, hybridization during divergence may be considered a rule rather than an exception (Arnold & Meyer, 2006; Abbott et al., 2013). Natural hybridization offers a unique opportunity to evaluate the actions of selection on species-specific traits (Barton & Hewitt, 1989; Harrison, 1990) including aspects of physiology, morphology, and behavior. At the DNA level, natural selection impacts the patterns by which genetic regions are exchanged between hybridizing lineages, with the direction and rate of introgression influenced by their selective advantage or disadvantage in alternative genomic backgrounds (Gompert & Buerkle, 2011). Regions containing variants that reduce the fitness of hybrids, for example, experience limited penetration of species boundaries while regions containing variants that increase the fitness of hybrids experience more rapid penetration of species boundaries. The latter process characterizes adaptive introgression (Hedrick, 2013).

Among primates, baboons (genus *Papio*) are one of the most extensively studied taxa with respect to natural hybridization (e.g., Phillips-Conroy & Jolly, 1986; Jolly et al., 2011; Charpentier et al., 2012). Sometimes considered a single polytypic biological species due to their propensity for hybridization (e.g., Delson et al., 2000; Frost et al., 2003), a growing consensus supports the recognition of six major "forms" or phylogenetic species: hamadryas baboons (*P. hamadryas*), Guinea baboons (*P. papio*), anubis baboons (*P. anubis*), yellow baboons (*P. cynocephalus*), chacma baboons (*P. ursinus*), and Kinda baboons (*P. kindae*). Kinda baboons comprise the most recent addition to the commonly accepted five-species categorization on account of their distinct, consistently diagnosable suite of phenotypes (Jolly, 2001; Frost et al., 2003; Jolly et al., 2011; Weyher et al., 2014), as well as their genetic differentiation (Burrell, 2009; Zinner et al., 2009, 2013). We follow recent authors in recognizing them as a sixth full species, *P. kindae* (Lönnberg, 1919). Grayfooted chacma or "grayfoot" baboons (*P. griseipes*), which are usually lumped with other chacma baboons, are phenotypically distinct from southern Cape chacma baboons (Jolly, 1993), though to a lesser degree compared to differences observed among the other six species.

Our present study concentrates on a hybrid zone between Kinda and grayfoot baboons in the Kafue river valley in central Zambia. There, the distributions of the two species adjoin in and around the area of Kafue National Park, with Kinda and grayfoot baboons occupying respectively the northern and southern areas of the region. After previous speculation about potential overlap and interbreeding in the region (Ansell, 1960, 1978; Jolly, 1993; Burrell, 2009), Jolly et al. (2011) discovered that the two species did in fact hybridize, detecting individuals in multiple localities with mixed or intermediate phenotypes as well as individuals and social groups with mixed mitochondrial-DNA and Y-chromosome ancestry.

Hybridization between Kinda and grayfoot baboons is somewhat remarkable given readily apparent body-size differences between the species (Jolly et al., 2011). Grayfoot baboons are among the largest baboon species, second possibly only to the southern *P. ursinus* baboons with which they are allied. Kinda baboons, on the other hand, are the smallest baboon species, with adult males and females weighing approximately 53% and 74% of the body mass of their respective grayfoot baboon counterparts. In a male grayfoot × female Kinda baboon mating, the female partner would weigh roughly 35% of the mass of the male partner, while in a female grayfoot × male Kinda baboon mating, the female partner would weigh 88% of the mass of the male partner (Jolly et al., 2011). Both pairings occupy extremes of body size sexual dimorphism in extant papionin monkey species (Delson et al., 2000) and would therefore be susceptible to both pre- and post-zygotic obstacles to successful reproduction.

Hybridization in the Kafue river valley is also interesting in the context of behavioral differentiation between species and the interaction of competing behaviors in a hybrid zone (Bergman et al., 2008; Charpentier et al., 2012). Grayfoot baboons, like most other baboon species (but see Kummer, 1968; Fischer et al., 2017), live in multi-male/multi-female social groups characterized by strong female bonds and male dispersal (Silk et al., 2009, 2010). Mating is to a high degree determined by both male and female rank (Bulger, 1993). Kinda baboons are much less known, but appear also to live in multi-male/multi-female social groups with female bonds and male dispersal. Curiously, male Kinda baboon frequently initiate and maintain grooming interactions with non-estrous females, a behavior that is rarely seen in any other baboon species (Weyher et al., 2014),

suggesting a role for divergent strategies for reproduction in this species.

Notably, Jolly et al. (2011) found that, among hybrid males with discordant mitochondrial-DNA and Y-chromosome ancestry, the Y chromosome was exclusively inherited from Kinda baboons, a surprising finding given the diminutive size of male Kinda baboons, which would likely be physically outmatched by larger male grayfoot baboons in competition for mating opportunities. This finding instead suggests that male grayfoot baboons experience lower reproductive success in the hybrid zone. Jolly et al. (2011) hypothesized that this may be a result of either reproductive incompatibilities associated with obstetric challenges in small Kinda-like mothers or a propensity for sneak matings by male Kinda baboons, which may be facilitated by their diminutive and juvenescent appearance.

In the present study, we employ a population genomic approach to investigate genome-wide patterns of ancestry and selection in the Kafue river valley hybrid zone. We use double-digest RADseq (Peterson et al., 2012) to sequence thousands of genomic variants to high coverage, with additional FecalSeq enrichment (Chiou & Bergey, 2015) to incorporate noninvasive genomic information from unhabituated wild baboons. We use the resulting genomic dataset to characterize patterns of genomic ancestry in the hybrid zone and to identify candidate genes subject to local adaptation (Coop et al., 2010; Günther & Coop, 2013) or extreme introgression (Gompert & Buerkle, 2011).

Methods

Samples

Genetic material from wild Zambian baboons was collected between 2006 and 2015 from the Kafue river valley contact zone in Zambia (Jolly et al., 2011) (Table 4.1). Blood and fecal samples were collected and stored following procedures described in Chapter 3. All procedures were conducted following local laws and guidelines in Zambia and with approval from the Animal Studies Committees at Washington University, New York University, and Baylor College of Medicine.

Table 4.1: Samples included in this analysis. Only samples that passed filters are listed in this table. An expanded list including sample IDs is provided in Table S13.

Locality	Latitude	Longitude	Year	Appearance	Tissue type	N
Chunga	-15.0446°	25.9994°	2011	Kinda	blood	71
Mwengwa Rapids	-15.3115°	25.9449°	2014	mostly Kinda	feces	5
Malala Camp	-15.7661°	25.8626°	2014	mixed	feces	3
Musa Bridge	-15.9081°	25.7414°	2007, 2015	mixed	feces	4
Top Musa	-15.9013°	25.8133°	2014	mixed	feces	5
North Nkala Road	-15.9029°	25.8895°	2007	mixed	feces	6
Ngoma Airstrip	-15.9695°	25.9376°	2012	mixed	blood	12
Dendro Park	-16.1518°	26.0599°	2012	mostly grayfoot	blood	4
Nanzhila Plains	-16.2337°	25.9612°	2007	mostly grayfoot	feces	5
Choma	-16.6383°	27.0308°	2007	grayfoot	feces	3
Lower Zambezi National Park	-15.9234°	28.9466°	2006	grayfoot	feces	5

DNA isolation, library preparation and sequencing

We sequenced DNA from a total of 129 animals, of which 123 remained after filtering for quality as described in the following section (Table S13). For blood samples, which included leukocyte, plasma, or FTA-dried blood spots, we extracted DNA using the QIAamp DNA Blood Mini Kit (Qiagen). For fecal samples, we extracted DNA using the QIAamp DNA Stool Mini Kit (Qiagen).

Because DNA in feces contains an overwhelming majority of bacterial DNA (Perry et al., 2010; Snyder-Mackler et al., 2016), we enriched host DNA from fecal samples using FecalSeq (Chiou & Bergey, 2015; Chapter 2), which takes advantage of the high density of CpG-methylation in mammalian genomes relative to bacterial genomes to partition fecal DNA into endogenous and exogenous fractions using methyl-binding domain protein baits. We then used double-digest RADseq (ddRADseq) (Peterson et al., 2012) to sequence a representative and reproducible set of orthologous SNPs. Procedures for fecal enrichment and ddRADseq library preparation are described in Chapter 3. We sequenced finished libraries on the Illumina HiSeq 2500 or Illumina HiSeq 4000 platforms.

Bioinformatic procedures for sample demultiplexing, sequence alignment, and variant identification followed those described in Chapter 3. The final dataset contained 14,642 variants.

Detecting introgression

In an exploratory analysis, we first performed a classical multidimensional scaling (MDS) analysis to visualize identity-by-state using PLINK (Purcell et al., 2007; Chang et al., 2015) (Figure S5).

We next performed an unsupervised clustering using ADMIXTURE in order to estimate the ancestry of each individual by maximum likelihood (Alexander et al., 2009). We ran the algorithm with 10-fold cross validation to estimate the error rate using values for K ranging from 1 to 10. We found that $K = 2$ performed best for distinguishing between Kinda and grayfoot baboon genetic backgrounds (Figure S4). We also estimated standard errors for point estimates of the Q parameter using moving block bootstrapping, after calibrating the dataset such that 0 corresponded to pure Kinda baboon ancestry and 1 corresponded to pure grayfoot baboon ancestry. We evaluated concordance between the MDS and ADMIXTURE results by linear regression (Figure S6).

We compared genetic ancestry results for the present study to our previously reported ancestry results derived from matrilineal mitochondrial and patrilineal Y-chromosome markers (Jolly et al., 2011). As with the autosomal data, we assigned a hybrid index to all individuals that was either 0, indicating a Kinda baboon haplotype, or 1, indicating a grayfoot baboon haplotype. Because individuals and sampling locales differed between studies, we interpolated ancestry results from the autosomal, mitochondrial, and Y-chromosome marker sets across geographic space at 1 km^2 resolution using a Kriging surface model procedure implemented in the fields package (Nychka et al., 2015) in R (R Core Team, 2013). We then assigned for each sampling locale ($n = 40$) and marker set a single hybrid index based on the geographic interpolation results. To make up for gaps in sampling, we also randomly drew an equal number of "simulated" sampling locales sampled from the geographic background for a total of 80 locales. We then compared ancestry results from different marker sets using nonparametric Wilcoxon signed rank tests.

Local adaptation analysis

In order to identify SNPs associated with environmental variation to scan for signatures of local adaptation, we employed the Bayesian association method implemented in Bayenv (Hancock et al., 2008; Coop et al., 2010; Günther & Coop, 2013). Bayenv estimates correlations between allele fre-

quencies and environmental variables while controlling for neutral correlations due to factors such as gene flow and shared population history. To conduct this test, we first estimated a covariance matrix using the pruned dataset of 14,642 SNPs. Bayenv uses this covariance matrix as a neutral model to control for population structure; support for a linear effect of environmental variation on allele frequencies is then computed in relation to this neutral model. To assemble an environmental dataset, nineteen bioclimatic variables comprising the core BIOCLIM dataset (Table S14) (Busby, 1991) were summarized as six principal components that together account for $> 95\%$ of the variance in the study area. Bayenv was then run with 1,000,000 MCMC iterations for each SNP. We then summarized for each SNP locus i and environmental variable j a posterior Bayes factor (bf_{ij}) estimate as a measure of support for the alternative model of environment-allele frequency correlations (Coop et al., 2010).

To relate SNP patterns of local adaptation to protein-coding genes, we used the baboon reference genome (papAnu2) gene annotations obtained from Ensembl (Cunningham et al., 2015). We assigned SNPs to genes if they fell within a window defined as the boundaries of a protein-coding gene $\pm$ 50 bp. For each environmental principal component j, we then summarized Bayes factor statistics for genes calculated as the average bf_{ij} of SNPs falling within the gene's window. We refer to this statistic as BF_{ij} to distinguish it from bf_{ij} for SNP loci.

We employed a nonparametric permutation test to assess the significance of BF_{ij} estimates. In each iteration, we randomly assigned bf_{ij} values to SNPs by sampling with replacement from all bf_{ij} values in the dataset. BF_{ij} values were then calculated following the procedure described above. Sampling was conducted first for 10,000 replicates. A $p_{BF_{ij}}$ value was then calculated for each gene as the proportion of BF_{ij} values in the randomized samples that were greater than the actual BF_{ij} value for environmental variable j. If significance could not be confidently determined after 10,000 replicates, we allowed the analysis for remaining genes to continue to a maximum of 10 million replicates, after which all $p_{BF_{ij}}$ values could be confidently distinguished from the significance threshold. Because each locus was assessed six times (once for each variable), we determined significance using a Bonferroni-adjusted threshold of $\alpha = \frac{0.05}{6} = 0.008333$. We classified genes as candidates for local adaptation if $p_{BF_{ij}}$ was significant for any environmental principal component.

Genomic cline analysis

In order to identify regions with aberrant patterns of ancestry and introgression, we employed the Bayesian genomic cline method (Gompert & Buerkle, 2011) implemented in bgc (Gompert & Buerkle, 2012). Bayesian genomic clines provide a powerful statistical framework for identifying extreme patterns of introgression by first modeling neutral patterns in the dataset. The model therefore provides a useful method for identifying loci potentially associated with variation in fitness that is calibrated to the provided dataset.

As bgc requires the *a priori* specification of ancestral populations, we first classified individuals as pure or admixed based on results from the ADMIXTURE model described above. We considered individuals to be of pure ancestry if their ancestry estimates inferred by ADMIXTURE were greater than 0.999. By this criterion, we identified 69 pure Kinda baboons, 20 pure grayfoot baboons, and 34 individuals of mixed ancestry (Figure S7).

We analyzed the pruned SNP dataset in 5 runs of 400,000 MCMC iterations each, sampling every 40 iterations after discarding the first 200,000 iterations as burn in. We set Kinda baboons as species 0 and grayfoot baboons as species 1. To incorporate information about genotypic uncertainty, we ran the analysis using raw allelic read counts rather than genotype calls for all 14,642 SNPs in the dataset (arguments: `-O 0 -x 400000 -n 200000 -t 40 -N 1 -E 0.0001 -q 1 -I 1 -p 1`).

Each run was assessed for convergence visually (Figure S8 and Figure S9) and statistically using the Heidelberger & Welch (1983) convergence diagnostic as implemented in the coda package (Plummer et al., 2006) in R (R Core Team, 2013). We ran both a stationarity test and a half-width test on posterior sample estimates to ensure adequate convergence. The stationarity test uses the Cramer-von-Mises statistic to test a null hypothesis of stationarity on the posterior sample chain with successively higher fractions removed until either the null hypothesis is accepted, indicating convergence, or 50% of the chain has been discarded, indicating failure. We used an $\alpha = 0.05$ for this test. The half-width test uses the stationary portion of the posterior chain identified previously to calculate the ratio between one-half the width of the 95% confidence interval and the mean. A ratio lower than ϵ indicates that the sample length is sufficient to estimate the mean accurately while a ratio higher than ϵ indicates a sample of insufficient length. We set $\epsilon = 0.1$ for this analysis.

For each chain, we excluded loci that did not converge. We then combined MCMC samples from all chains.

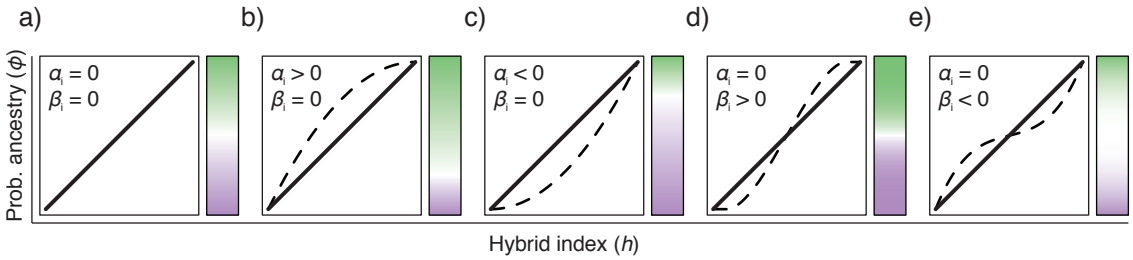

Figure 4.1: Genomic clines with hypothetical values for the α_i and β_i cline parameters. Line plots show the relationship between the genome-wide hybrid index (h) on the x-axis and the probability of ancestry at a given locus (ϕ) on the y-axis. To the right of each plot, the given pattern is illustrated using a color gradient. Under neutrality, the probability of ancestry is equivalent to an individual's hybrid index (panel A). The neutral relationship is shown in subsequent panels as a solid line while deviations are shown as dashed lines. A nonzero α_i represents a shift in the cline such that the probability of ancestry at locus i is biased towards species 1 if α_i is positive (panel B) or towards species 0 if α_i is negative (panel C). A nonzero β_i represents a change in the rate at which the probability of ancestry increases as a function of hybrid index. When β_i is positive, the transition rate increases as a function of the hybrid index corresponding to a steeper cline (panel D). When β_i is negative, the transition rate decreases as a function of the hybrid index corresponding to a wider cline (panel E). Adapted from Gompert & Buerkle (2012).

From each MCMC sample, we drew posterior estimates for the cline parameters α_i, which measures the shift of the cline (positive values: shift to species 1; negative values: shift to species 0), and β_i, which measures the gradient of the cline (positive values: steep cline; negative values: wide cline) (Figure 4.1). Because protein-coding genes are the focus of the analysis, we summarized parameter estimates by assigning a single value of α_i and β_i to protein-coding genes using a similar process to that described for the BF_{ij} values earlier. For each MCMC sample, we assigned values of α_i and β_i to each gene calculated as the mean of point parameter estimates for SNPs falling within a window defined as the boundaries of the protein-coding gene $\pm$ 50 bp. We thus derived posterior MCMC samples of cline parameter estimates for genes rather of SNPs.

In order to test for extreme patterns of introgression of protein-coding genes, we summarized the posterior chains for genes by calculating mean point estimates as well as 95% equal-tail probability (ETP) credible intervals. Following Gompert & Buerkle (2012), we characterized loci as having patterns of excess ancestry if the 95% ETP intervals for α_i or β_i did not contain 0.

Functional enrichment analysis

Instead of testing for significant or extreme signatures associated with single genes, enrichment analysis tests for extreme patterns associated overall with well-characterized biological functions or pathways. Enrichment analysis may therefore detect functions or pathways with extreme effects, even when few or none of the associated genes are individually identified as extreme.

In order to prepare gene function and pathway annotations, we annotated the baboon reference genome with Gene Ontology (GO) annotations (Gene Ontology Consortium, 2000, 2015) and the rhesus macaque reference genome with PANTHER pathway annotations (Mi et al., 2013a, 2013b), as PANTHER annotations do not exist for the baboon genome. Procedures for genome annotation followed those described in Chapter 3.

We evaluated whether genes with aberrant patterns of introgression were significantly associated with gene functions or gene pathways. For these analyses, we excluded genes lacking α_i or β_i values and genes lacking either GO or PANTHER classifications. Enrichment analyses were not conducted for the Bayenv analysis.

We used Wilcoxon rank sum tests to evaluate, for each GO class or PANTHER pathway, whether α_i or β_i values assigned to a class or pathway were more extreme than values not assigned to that class or pathway. We performed for each parameter a two-tailed test to scan for GO classes or PANTHER pathways with significantly high or low values for either parameter. Because these analyses are two-tailed, we considered GO classes or PANTHER pathways with a p-value < 0.025 to be enriched for extreme α or β. Correction for multiple comparisons was not conducted due to the exploratory nature of these analyses.

Results and Discussion

We used double-digest RADseq (Peterson et al., 2012) in combination with FecalSeq enrichment of the host genome for an important subset of DNA samples derived from feces (Chiou & Bergey, 2015; Chapter 2) to genotype Zambian baboons from the greater Kafue National Park region at thousands of polymorphic autosomal sites. After filtering out loci with insufficient quality and high

missingness, as well as loci in high linkage disequilibrium, we derived a final dataset of 123 Kinda, grayfoot, and hybrid baboons genotyped at 14,962 SNP loci. The mean sequencing depth across these loci was 41.94 reads per individual, but varied across individuals from 0.29 to 233.66, with 90% of individuals sequenced to an average of at least 5.23 reads per locus.

Hybrid zone structure

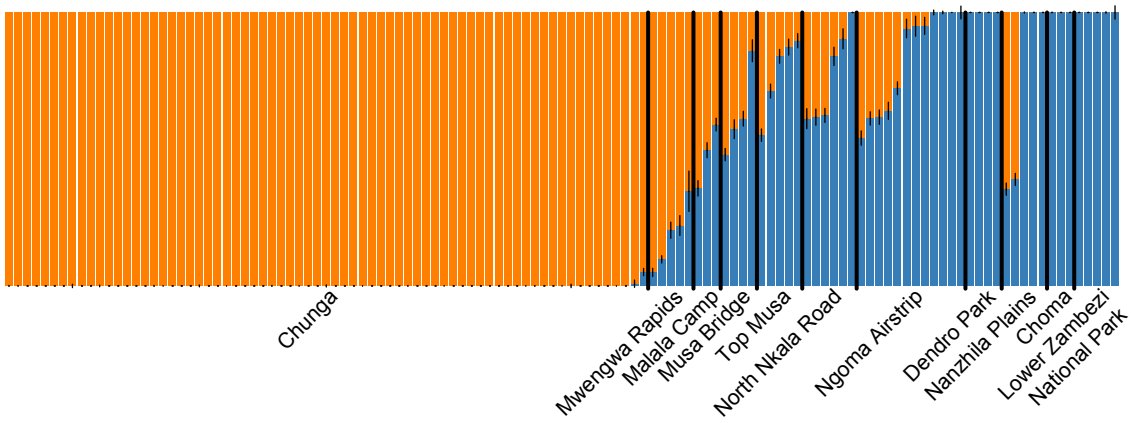

Figure 4.2: Ancestry estimation results using ADMIXTURE. Individuals are ordered according to the localities listed in Table 4.1.

We investigated population structure and estimated ancestry components using the unsupervised maximum likelihood population clustering approach implemented in ADMIXTURE (Alexander et al., 2009) (Figure 4.2 and Table S13). We used the Q parameters from the model as estimates of a hybrid index ranging from 0 to 1 for each individual, with 0 indicating pure Kinda baboon ancestry and 1 indicating pure grayfoot baboon ancestry.

The ADMIXTURE results revealed a wide cline, spanning at least 100 km on a north-south axis (Figure 4.3). As expected based on their location and physical appearance, individuals from Chunga, located north of the hybrid zone (Ansell, 1978; Jolly et al., 2011), were almost exclusively assigned pure Kinda baboon ancestry while individuals from Choma and Lower Zambezi National Park, located well south and east of the Kinda baboon distribution (Ansell, 1978), were almost exclusively assigned pure grayfoot baboon ancestry. These findings corroborate previous results based on microsatellite markers (Burrell, 2009).

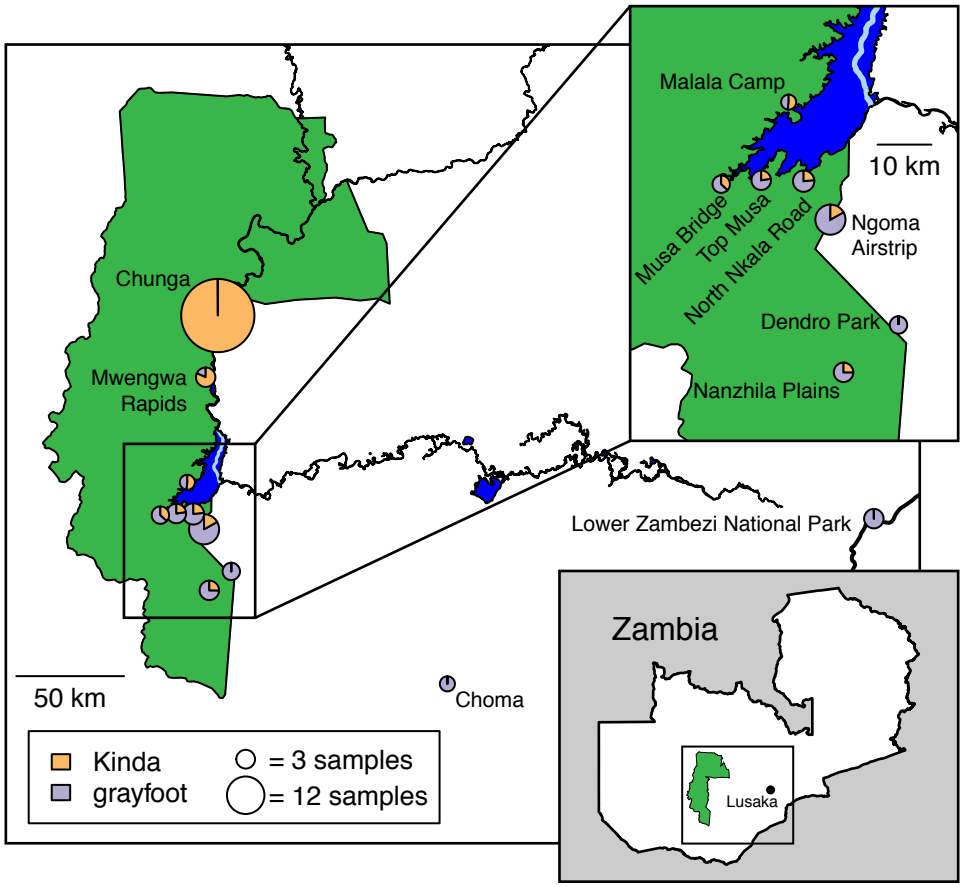

Figure 4.3: Average ADMIXTURE ancestry results (hybrid index) by locality. Circle areas are proportional to the number of individuals. The former course of the Kafue River prior to construction of the Itezhi Tezhi Dam between 1972 and 1976 is shown in light blue.

The midpoint of the cline, where genetic ancestry from Kinda and grayfoot baboons is roughly equivalent, corresponded most closely to Malala Camp, an unmanned scout camp located near the confluence of the Chibila River and Lake Itezhi Tezhi. Baboons at this location are phenotypically intermediate in size and pelage, but resemble more closely Kinda baboons in overall appearance, including presence of pink circumorbital skin, strongly contrasting cheek color, and lanky appearance (Figure S11). South of Lake Itezhi Tezhi, groups from four localities in a narrow west-east band were on average mostly grayfoot, but contained a range of individuals from those with pure or mostly pure grayfoot baboon ancestry to those with more evenly mixed ancestry. Further south, at the Dendro Park private reserve, individuals were exclusively pure grayfoot baboons. Less than 15 km southwest in the Nanzhila Plains, while a small majority of individuals were pure grayfoot baboons, the remainder were roughly evenly mixed. See Figure S10 for hybrid indices by locality and Figure 4.3 for a map of localities.

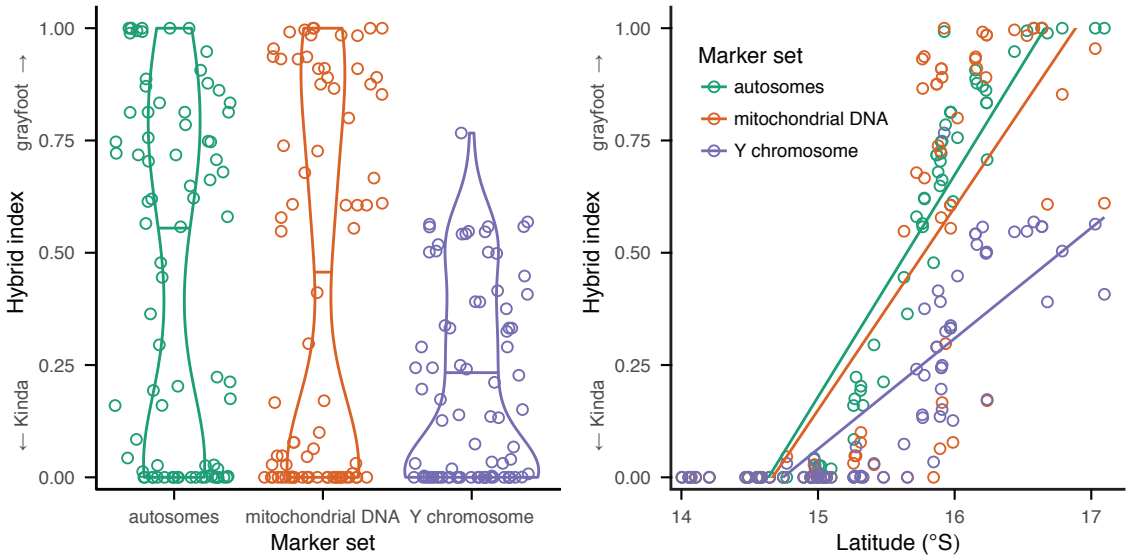

Figure 4.4: A comparison of ancestry results between autosomal, mitochondrial, and Y-chromosome marker sets. Ancestry estimates from all three marker sets were geospatially interpolated over the study area to facilitate comparison among all sample points (Figure S12 and Figure S13). Violin plots in the left pane show the distribution of hybrid indices estimated from the three marker sets, with mean estimates indicated by horizontal lines. The scatter plot in the right pane shows hybrid indices plotted against latitude. Mitochondrial and Y-chromosome data are from Jolly et al. (2011).

We compared ancestry results from our autosomal marker set to ancestry results from mitochondrial and Y-chromosome DNA using nonparametric Wilcoxon signed rank tests (Table 4.2). We

found that our ancestry results from autosomal markers were not significantly different from corresponding results from mitochondrial DNA. Ancestry results from Y-chromosome markers, however, were significantly different from ancestry results from both autosomal and mitochondrial DNA. Our results suggest that there is a sizeable average excess of Kinda baboon Y-chromosome ancestry relative to not only mitochondrial DNA, as previously reported by Jolly et al. (2011), but the remainder of the nuclear genome as well (Figure 4.4). Interestingly, the unidirectional Y introgression relative to both mitochondrial DNA and the autosomes parallels findings in another papionin group, macaques (*Macaca*), in which the *M. mulatta* Y chromosome has apparently introgressed into *M. fascicularis* populations, reaching as far south as the Isthmus of Kra in mainland southeast Asia and resulting in discordance with both the *M. fascicularis* mitochondrial genome and morphotype (Tosi et al., 2002).

Table 4.2: Comparison of ancestry results from autosomal, mitochondrial, and Y-chromosome marker sets using nonparametric Wilcoxon signed rank tests.

	Sampling locales	W	p-value
autosomes vs. mitochondrial DNA	40 real + 40 simulated	1088	0.0725
	40 real	311	0.5919
autosomes vs. Y chromosome	40 real + 40 simulated	1882	< 0.0001
	40 real	630	< 0.0001
mitochondrial DNA vs. Y chromosome	40 real + 40 simulated	1350	< 0.0001
	40 real	495	< 0.0001

As previously discussed by Jolly et al. (2011), mitochondrial/Y-chromosome discordance in Kafue National Park baboon hybrids suggests strong unidirectional Kinda baboon Y-chromosome introgression. In the absence of autosomal information, this is consistent with the common case in species in which males are the dispersing sex (Zinner et al., 2011) and indicates that the hybrid zone could be driven primarily by immigration of male Kinda baboons. A probable eventual outcome of this scenario is "nuclear swamping", by which successive backcrossing between hybrids and the influx of foreign males largely replaces the native population's nuclear genome but may retain the native population's mitochondrial gene pool. Our results, however, reveal that the Kinda baboon Y chromosome has introgressed more extensively on average than autosomal DNA, indicating either that insufficient time has passed for nuclear swamping to occur or that migration of male Kinda

baboons is not the principal factor driving the hybrid zone and other processes may be at work.

Jolly et al. (2011) proposed several hypotheses that explain strong unidirectional male Kinda baboon introgression without requiring unidirectional migration. One hypothesis is that reproductive pairings between male grayfoot baboons and female Kinda baboons are disadvantaged because of reproductive incompatibilities, perhaps associated with obstetric challenges that result from the unusually extreme size difference between potential mates (or more pertinently, between mothers and their hybrid fetuses). Another hypothesis is that the diminutive or juvenescent appearance of Kinda baboons leads to consequent issues with recognition of potential female mates and/or male competitors by male grayfoot baboons. The latter effect may be compounded by the unusual grooming behavior in Kinda baboons by which males actively initiate and maintain grooming associations with females even when females are not in estrus (Weyher et al., 2014). Both of these hypotheses are consistent with our results, as they suggest that while male grayfoot baboons are disadvantaged in the hybrid zone, subsequent backcrossing between male hybrid and female grayfoot baboons may not be similarly inhibited. Such backcrossing would counteract replacement of the grayfoot baboon nuclear genome while explaining the extreme introgression of the Kinda baboon Y chromosome.

A neutral model, however, based on sex differences in reproductive skew and the lower effective population size of the Y chromosome could also explain the unidirectional introgression of the Y chromosome. The uniparental inheritance of the Y chromosome, as well as the mitochondrial genome, results in reduced effective population sizes of both markers equaling 1/4 the effective population size of the autosomes assuming even sex ratios and the absence of sex-biased processes (Charlesworth, 2009). If male baboons have higher variance in reproductive success than females, however, the variance effective population size of the patrilineally inherited Y chromosome would be even lower than that of the matrilineally inherited mitochondrial genome. Under these conditions, genetic drift could produce large changes in Y-chromosome frequencies relatively quickly. Given that male primates tend to have higher variance in reproductive success relative to females (Kutsukake & Nunn, 2006), including grayfoot baboons (Bulger, 1993), the neutral model could plausibly explain the unidirectional introgression of the Y chromosome.

Ansell (1978), in the last published survey of Kafue National Park baboons prior to our work (Jolly et al., 2011), remarked that specimens of phenotypically grayfoot baboons were collected as far north as the vicinity of Lubalunsuki Hill and Itumbi (approximately 15.50° S., 25.97° E.). Phenotypically Kinda baboons were observed in the area as well, but no indications of interbreeding were known. It now seems likely that hybridization was already occurring given the geographic proximity of the species and the extent of the hybrid zone today, but if we assume these descriptions are otherwise accurate, then the hybrid zone has shifted considerably south in the past half-century. Baboons near Lubalunsuki Hill today resemble phenotypically Kinda baboons (Figure S14) and the center of the hybrid zone is 30 – 40 km south of Lubalunsuki. Furthermore, Kinda baboon mitochondrial haplotypes are found as far south as the Ngoma Airstrip, 55 km south of Lubalunsuki, while Y-chromosome haplotypes, as well as intermediate phenotypes, are found as farth south as the Nanzhila Plains, 80 km south of Lubalunsuki. These findings suggest either that Ansell's (1978) descriptions were inaccurate or that there has been strong directional movement of Kinda baboons, including females, into the grayfoot baboon distribution. This movement may have come from the north, but may also have come from the west, where there is an expanse of miombo forest corresponding to typical Kinda baboon habitat from which little information is available.

Between 1972 and 1976, construction of a new hydropower dam at Itezhi Tezhi flooded a large section of the Kafue River, as well as its upstream tributaries on the western bank, forming the artificial Lake Itezhi Tezhi (see Figure 4.3). The extent to which this anthropogenic event impacted baboon hybridization is unknown, but raises intriguing questions. The impacts of flooding on the local ecology might have created conditions more or less favorable to either species. The flooding would have also displaced existing groups of baboons, potentially precipitating contact and subsequent hybridization between species. Based on Ansell's (1978) descriptions, it seems likely the lake flooded primarily grayfoot baboon territory and the cline of the hybrid zone, which today begins northwest of Lake Itezhi Tezhi and extends around and to the area south of the the lake, formed in the time since. If a hybrid zone already existed, however, roughly coinciding with the present cline, then the formation of Lake Itezhi Tezhi bisected a section of the hybrid zone, creating a barrier between the area west of Lake Itezhi Tezhi and the area immediately south of the

Itezhi Tezhi Dam. A landscape genetic approach comparing baboons on either side of this barrier therefore has high potential for resolving this issue.

Signatures of local adaptation and extreme introgression

In order to identify candidate genes showing signatures of local adaptation, we used the Bayesian method implemented in Bayenv (Coop et al., 2010; Günther & Coop, 2013), which takes into account neutral covariance of allele frequencies among populations. After assessing the significance of Bayes Factor (BF_{ij}) statistics from the model by permutation, we identified 80 candidate genes exhibiting significant signatures for at least one of the 6 total environmental principal components (Table S15 and Figure S15).

In order to assess the validity of the model and rule out the possibility of a spurious fit to data with no environmental dependence (Hancock et al., 2008; Coop et al., 2010), we compared the distributions of BF_{ij} in SNPs assigned to genes and SNPs not assigned to genes. As nongenic SNPs are less likely to be functional, this comparison tests for enrichment of selection signals in the tails of the test statistic with the prediction that the signal should be stronger in genic SNPs. We did not find statistical support for this prediction (Wilcoxon rank sum test, $p = 0.1242$).

We also identified genes with extreme patterns of introgression using the Bayesian method implemented in bgc (Gompert & Buerkle, 2012). We defined genes as having excess patterns of introgression if their posterior 95% credible intervals for α_i or β_i were entirely greater than or less than 0. We thus identified 3 genes with extremely low α_i, 52 genes with extremely high α_i, 1 gene with extremely low β_i, and 5 genes with extremely high β_i (Table S16, Table S17, and Figure S16). High α_i indicates an excess probability of grayfoot baboon ancestry relative to the hybrid index while low α_i indicates an excess probability of Kinda baboon ancestry relative to the hybrid index. The high discrepancy between the number of genes with low and high α_i possibly indicates a differentially high rate of introgression of grayfoot baboon gene variants, but may also be a methodological artifact. Comparison between hybrid indices inferred by ADMIXTURE and bgc revealed a strong correlation, but the slope indicated that hybrid indices estimated by bgc were on average lower than corresponding estimates by ADMIXTURE (Figure S17). In contrast to

the unsupervised approach in ADMIXTURE, bgc requires prior specification of parental populations. Because the α_i parameter is estimated based on a comparison between the per-locus probability of ancestry and the genome-wide hybrid index estimated by bgc, an inaccurately low hybrid index may artificially inflate the number of loci with high per-locus probabilities of grayfoot baboon ancestry (see Figure 4.1). The lower sample size ($n = 20$) combined with the possibility of low levels of admixture in the grayfoot baboon parental sample may explain the systematic underestimation of grayfoot baboon ancestry by bgc.

While the above caveat applies to genes with positive α_i, it conversely gives greater credence to genes identified as having extremely negative α, indicating excess Kinda baboon ancestry. These genes were *AMPH*, *KMT2E*, and *LY96*. It also does not affect estimates of β_i, which is not affected by the direction of introgression. Extremely positive values of β_i indicate narrow clines and suggest potential roles in barriers to reproduction (Gompert & Buerkle, 2011). The genes identified were *AACS*, *LIMK1*, *LY96*, *TMEFF2*, and *TMEM178*. Extremely negative values of β_i indicate wide clines and suggest potential roles in adaptive introgression (Gompert & Buerkle, 2011). *ODF2* was the only gene identified in this category.

As with the Bayenv model, we assessed the concordance of the bgc model with biological expectations by comparing cline parameter values of genic and nongenic SNPs. In this case, we expected that genic SNPs, which are more likely to be functional, would be associated with stronger barriers to reproduction in the hybrid zone and would thereby have higher β_i values. We found statistically significant support for this prediction (Wilcoxon rank sum test, $p = 0.0443$).

Local adaptation in principle represents a barrier to introgression due to the inferiority of foreign alleles in a local environment (Kawecki & Ebert, 2004). We therefore tested the prediction that genes with signatures of local adaptation (i.e., significantly high BF_{ij}) also exhibit steeper clines (i.e., higher β_i). We did not find statistical support for this prediction (Wilcoxon rank sum test, $p = 0.3205$).

Given the overall lack of support in our Bayenv results, we conclude that the overall signal for local adaptation in our study is weak. The lack of a strong signal may reflect the weak effect of local adaptation at small geographic scales. Our core study area, for instance, is largely confined

to an area less than 150 km in length. It may also potentially reflect a weak overall effect of environment on biological variation in baboons, which has previously been suggested by Jolly (1993) and supported by analysis of behavioral (Kamilar, 2006) and anatomical data (Dunn et al., 2013). Dynamic, short-term changes in the genetic structure of our study populations, perhaps precipitated by hybridization or by other processes unrelated to environmental variation, could potentially have also temporarily broken up genotype-environment correlations. Finally, the lack of a local-adaptation signal may reflect either the low precision of our environmental dataset, which is based on global interpolations of climatological data that rely on relatively sparely distributed weather stations in developing nations such as Zambia (Hijmans et al., 2005), or the low correlation between climatic variables and non-climatic environmental variables that are more relevant to local adaptation in these populations.

Gene function and biological pathway analysis

We conducted enrichment analyses in order to identify Gene Ontology (GO) terms (Gene Ontology Consortium, 2000, 2015) and PANTHER pathways (Mi et al., 2013a, 2013b) associated with extreme overall shifts in the genomic cline parameters α_i and β_i. We identified 8 GO terms enriched for high α_i, 10 GO terms enriched for low α_i, 17 GO terms enriched for high β_i, and 9 GO terms enriched for low β_i (Table S18). We also identified 4 PANTHER pathways enriched for high α_i, 1 PANTHER pathway enriched for high β_i, and 2 PANTHER pathways enriched for low β_i (Table 4.3). We did not identify any pathways enriched for low α_i. Corrections for multiple comparisons were not applied for these exploratory analyses.

While differences between hybrid indices estimated by ADMIXTURE and bgc (Figure S17) affect the sign of α_i estimates and consequently the identification of candidate genes with extreme α_i as previously discussed, they do not affect the enrichment of α_i because enrichment analyses test for extreme parameter estimates relative to the full distribution of parameter values rather than 0.

Table 4.3: PANTHER pathways with significantly enriched α_i or β_i cline parameters. Only pathways with $p < 0.025$ for any of four one-tailed enrichment tests (high α_i, low α_i, high β_i, low β_i) are shown here. Key: +, test for high parameter value; –, test for low parameter value.

PANTHER pathway		Enrichment		
Accession	Name	Parameter	+/–	p-value
P00011	Blood coagulation	α_i	+	0.01586
P00035	Interferon-gamma signaling pathway	α_i	+	0.02408
P00054	Toll receptor signaling pathway	α_i	+	0.01134
P05911	Angiotensin II-stimulated signaling through G proteins and beta-arrestin	α_i	+	0.00762
P00038	JAK/STAT signaling pathway	β_i	+	0.01417
P00021	FGF signaling pathway	β_i	–	0.02076
P00054	Toll receptor signaling pathway	β_i	–	0.01134

JAK/STAT signaling pathway

The Janus kinase (JAK) and signal transducer and activation of transcription (STAT) signaling pathway was the only biological pathway in our analysis enriched for high β_i (Table 4.3), indicating potential involvement in barriers to reproduction. This pathway functions as a signaling mechanism for a variety of pathways and underlies diverse responses such as mammary gland development, hematopoiesis, and immune cell development (Watson & Burdon, 1996; Ward et al., 2000; Ghoreschi et al., 2009). Intriguingly, it also underlies pathways related to organism growth, for example by serving as the signaling mechanism for the growth hormone (GH) and insulin-like growth factor 1 (IGF1) axis that plays an important role in organism growth through chondrocyte proliferation and differentiation (Nilsson et al., 1994; Vottero et al., 2013). JAK/STAT signaling has also been shown to play a role in leptin-induced chondrocyte differentiation, which is mediated through STAT3 (Ben-Eliezer et al., 2007).

Notably, the JAK/STAT signaling pathway has been found to be enriched for high F_{ST} between pure Kinda and grayfoot baboon populations, indicating that its components overall exhibit high differentiation that may indicate directional selection in one or both species (Chapter 3). The enrichment of JAK/STAT signaling for high β_i in the present study suggests that it may also be involved in isolating mechanisms. The relationship of JAK/STAT components to body size differences in baboons has not yet been confirmed or characterized. If changes in the JAK/STAT

pathway are indeed related to the extreme differences in body size between Kinda and grayfoot baboons, our results suggest that these differences may be responsible for decreased fitness of hybrid offspring and that the specific genes under selection include components of the JAK/STAT signaling pathway. As suggested by Jolly et al. (2011), the selection may be strongest against reproduction between male grayfoot-like baboons and female Kinda-like baboons, possibly due to obstetric limitations differentially impacting hybrid offspring of small Kinda-like mothers or issues with mate or competitor recognition due to the juvenescent appearance of both female and male Kinda baboons. It may also, however, affect reproduction between male Kinda-like baboons and female grayfoot-like baboons, whose body size dimorphism would be on the low extreme of that found between mating partners in extant papionin monkeys (Delson et al., 2000). Barriers to reproduction are possible in both reciprocal cases but our data at present are unable to distinguish between the two.

Table 4.4: β_i parameter values for genes in the JAK/STAT signaling pathway. The point estimate (mean), percentile (relative to point parameter estimates of all genes), and posterior probability of a positive value are shown for β_i.

Component	Gene	Chromosome	Point estimate (β_i)	Percentile	Probability ($\beta_i > 0$)
JAK	*JAK1*	1	-0.04578	41.97	0.4744
PIAS	*PIAS1*	7	1.09215	98.62	0.8550
	PIAS4	19	0.86904	97.02	0.7630
STAT	*STAT2*	11	0.58100	92.09	0.6786
	STAT3	16	0.11558	59.80	0.5582

Within the JAK/STAT signaling pathway, the protein inhibitor of activated STAT (PIAS) genes *PIAS1* and *PIAS4*, as well as *STAT2*, were associated with the highest β_i values, although none were identified individually as candidates based on their posterior 95% credible intervals (Table 4.4). PIAS1 has been shown to inhibit STAT1 (Liu et al., 1998), which has anti-proliferative effects in cells including chondrocytes (Sahni et al., 1999). PIAS1 has also been shown to be a regulator of SOX9, a transcription factor that plays critical roles in developmental processes including chondrogenesis and testis determination (Oh et al., 2007).

FGF and TLR signaling pathways

We identified two pathways, the fibroblast growth factor (FGF) signaling pathway and the toll-like receptor (TLR) signaling pathway, that were enriched for negative β_i (Table 4.3), indicating that their clines were wider than average. FGF signaling activates a variety of downstream pathways and is associated with cellular outcomes including mitogenesis, differentiation, survival, apoptosis, and cell migration. Certain FGFs, including FGF1 which is highly differentiated between Kinda and grayfoot baboons (Chapter 3), play important roles in angiogenesis (Montesano et al., 1986; Eswarakumar et al., 2005). TLR signaling is involved in the recognition of pathogenic microbes and plays a role in the activation of both innate immunity and antigen-specific acquired immunity (Akira & Takeda, 2004).

Negative β_i values of these two pathways suggest that they may have a heightened ability to introgress across the hybrid zone. In the case of TLR signaling, the heightened amount of introgression may reflect adaptive introgression related to the introduction of novel pathogenic defenses. Notably, similar adaptive introgression of immune system components from archaic humans has significantly shaped the immune systems of modern human populations (Abi-Rached et al., 2011; Dannemann et al., 2016). Haplotypes in modern humans obtained through archaic human introgression include three TLRs (*TLR6-TLR1-TLR10*) demonstrably associated with microbial resistance (Dannemann et al., 2016). In our analysis, the TLR signaling pathway was also enriched for high α_i values, indicating that introgression of alleles within this pathway primarily took place from grayfoot into Kinda baboon populations.

Table 4.5: α_i and β_i parameter values for genes in the toll-like receptor signaling pathway. The point estimates (means), percentiles (relative to point parameter estimates of all genes), and posterior probabilities are shown for both cline parameters.

Component	Gene	Chromosome	Point estimate		Percentile		Probability	
			α_i	β_i	α_i	β_i	$\alpha_i > 0$	$\beta_i < 0$
MKK2	*MAP2K2*	19	0.64285	-0.59732	92.40	7.38	0.9450	0.8082
IkappaB	*NFKBIE*	4	2.02282	-1.07868	99.91	1.91	1.0000	0.7046

Two genes in the TLR signaling pathway were included in our analysis: *MAP2K2* and *NFKBIE* (Table 4.5). While neither gene encodes a TLR, they both encode kinases that play integral roles

in the downstream signaling cascade (Cohen, 2014).

Other pathways with aberrant introgression

Apart from the TLR signaling pathway, we identified three additional pathways enriched with high α_i: the blood coagulation pathway, the interferon-gamma signaling pathway, and the angiotensin II-stimulated signaling pathway (Table 4.3). The blood coagulation pathway includes a cascade of proteins ultimately responsible for the formation of a blood clot, which is critical for preventing excess blood loss following injury, as well as the fibrinolytic system responsible for the dissolution and limitation of clot (Schenone et al., 2004). Coagulation factors interact directly with pathogens and are therefore common targets of selection, including in primates (Rallapalli et al., 2014). Interestingly, a genome-wide association study of European humans linked Neanderthal alleles to phenotypes including hypercoagulation, suggesting that adaptive introgression has affected this pathway in humans (Simonti et al., 2016). The interferon-gamma (IFNγ) signaling pathway modulates the antiproliferative effects of IFNγ, a cytokine that serves important antiviral and immunoregulatory functions (Farrar & Schreiber, 1993). The angiotensin II (ANGII)-stimulated signaling pathway is an important component of the cardiovascular system stimulated by the peptide hormone ANGII and that plays a key role in mediating immediate physiological effects of vasoconstriction and blood pressure regulation (Mehta & Griendling, 2007). As in the TLR signaling pathway, high α_i parameter values indicate that introgression of alleles in these pathways primarily occurred from grayfoot into Kinda baboon populations.

Outer dense fiber protein 2

Outer dense fiber protein 2 (ODF2) is a component of outer dense fibers (ODFs), a class of cytoskeletal structures that are specific to sperm tails (Brohmann et al., 1997; Hoyer-Fender et al., 1998; Soung et al., 2006). Notably, *ODF2* in our study was the only gene identified as having extremely low β_i, indicating a wide cline, as well as as extremely high α_i, indicating excess grayfoot baboon ancestry. These results suggest that *ODF2* is a candidate gene for adaptive introgression, with the direction of introgression primarily taking place from grayfoot baboon populations into

Kinda baboon populations.

ODF2 has been shown to be critical for the structural integrity of the sperm flagellum. Inhibition of tyrosine phosphorylation of ODF2 as well as tektin 2 (TEKT2) adversely affects sperm motility during capacitation in hamsters (Mariappa et al., 2010). In ODF2-knockout mouse lines, males with high percentages of chimerism exhibit altered sperm tail structure and function, with some sperm exhibiting additional defects including bent tails displaying abnormal motility (Tarnasky et al., 2010).

ODF2 has been subject to recent positive selection in European human populations (Voight et al., 2006). Intriguingly, it has also been associated with substantial structural differences in chimpanzee DNA as part of a suite of genes showing markedly chimpanzee-specific changes relative to humans and orangutans (Kim et al., 2011). Chimpanzee females exhibit polyandrous mating and, as would be predicted, chimpanzee males exhibit extremely large relative testis sizes compared to humans, gorillas, and orangutans (Harcourt et al., 1981). Female polyandry is predicted to promote not only the evolution of larger testes and thereby ejaculate quantities, but also higher-quality (e.g., longer tails and faster) sperm (Fitzpatrick et al., 2009; Schmera et al., 2016).

Adaptive introgression of *ODF2* is surprising at face value given that sperm traits are commonly subject to dysgenesis in the context of an alien genome (e.g., Dobzhansky, 1934). In *Lepomis* sunfish for instance, hybrid sperm are fertile in the absence of competition but are outcompeted in the presence of sperm from either parental species (Immler et al., 2011). Conspecific sperm precedence is also observed in competition between bluegill and pumpkinseed sunfish sperm for pumpkinseed sunfish eggs, but not in the reciprocal case, potentially explaining unidirectional hybridization in this system (Immler et al., 2011). In the European house mouse hybrid zone, hybrid mice exhibit significantly reduced sperm count and sperm velocity, indicating widespread hybrid dysgenesis of sperm traits (Turner et al., 2011; Albrechtová et al., 2012).

Given the costs of sperm production (Wedell et al., 2002), selection on sperm quantity and quality is expected to be particularly important in taxa with higher rates of polyandrous mating (Ginsberg & Huck, 1989). In baboons and other animals, the priority-of-access model (Altmann, 1962) posits that male access to estrous females is determined according to rank. Throughout the

estrous period, males engage in mate-guarding in order to monopolize mating access and ensure paternity. This system may be subverted, however, by mechanisms including male coalitionary behavior, sneak copulations, and female choice (Alberts et al., 2003). Grayfoot baboons have previously been shown to conform well to the priority-of-access model and exhibit one of the strongest correlations between rank and mating among baboons (Bulger, 1993). While information on Kinda baboons is scarce, emerging information suggests a different picture. Male Kinda baboons notably engage at high frequency in grooming associations with females in which they play an active role in initiating and maintaining the associations even when females are pregnant, lactating, or otherwise not in estrus (Weyher et al., 2014). Preliminary data from a Kinda baboon study population at Kasanka National Park show a relatively high frequency of matings not predicted by dominance hierarchies and demonstrate that polyandrous mating by females occurs at least some of the time (A. Weyher, pers. comm.).

These characteristics would seem to imply a heightened role of sperm competition in Kinda baboons relative to grayfoot baboons. This prediction is further supported by morphometric analyses demonstrating that male Kinda baboons have larger testes, scaled to body size, relative to all other evaluated congeneric species except anubis baboons (Phillips-Conroy et al., in prep.), corroborating parallel investigations of relative testis sizes in hamadryas baboons, which are monandrous, and anubis baboons, which are both polygynous and polyandrous (Jolly & Phillips-Conroy, 2003, see also 2006).

Our results, however, suggest that the grayfoot baboon *ODF2* variant traverses the species boundary to a greater extent than the Kinda baboon variant, contrary to our expectations. Comparison to the European house mouse hybrid zone (Albrechtová et al., 2012) offers an intriguing explanation. Like the present baboon hybrid zone (Jolly et al., 2011), the *Mus musculus musculus* Y chromosome has introgressed into *Mus musculus domesticus* populations in apparent disregard of Haldane's rule. While Albrechtová et al. (2012) found that hybrids overall exhibited decreased sperm counts, this effect was more than rescued in apparently *domesticus* males by the presence of the invading *musculus* Y chromosome. This finding is surprising given that respective *domesticus* and *musculus* sperm traits and Y chromosomes evolved in concert subject to natural selection

in the parental mouse populations, and would likely encounter Dobzhansky-Muller disadvantages when placed in a novel genetic background. The effect of the *domesticus* Y chromosome on sperm traits instead implies that there is a sperm-related advantage in the presence of the invading Y chromosome that sufficiently balances and even overcomes this effect.

The present hybrid zone, where the Kinda baboon Y chromosome experiences unidirectional introgression but the grayfoot baboon sperm-related allele experiences disproportionate success, presents a compelling parallel to the house mouse system (Albrechtová et al., 2012). Taking this comparison to its conclusion, the house mouse analogy suggests that hybrid dysgenesis of sperm-related traits may decrease reproductive fitness in the Kinda and grayfoot baboon species boundary in the Kafue river valley due to Dobzhansky-Muller incompatibilities. This effect, however, may be mitigated or even overcome by beneficial interactions with the introgressed Y chromosome of one parental species, in this case the Y chromosome of the Kinda baboon. This hypothesis could not be tested in the present study due to the lack of data on Y-chromosome ancestry and sperm traits, but offers an intriguing alternative explanation for the surprising unidirectional introgression of the Kinda baboon Y chromosome.

Conclusion

In order to study patterns of ancestry and selection in the Kafue river valley baboon contact zone, we used double-digest RADseq to genotype hundreds of Zambian baboons at thousands of autosomal SNPs with high coverage. We found that Kinda and grayfoot baboons hybridize along a wide north-south geographic cline, in agreement with previous findings (Jolly et al., 2011), and that patterns of autosomal ancestry resembled patterns of mitochondrial ancestry more closely than Y-chromosome ancestry, suggesting that the Kinda baboon Y chromosome has penetrated the species barrier to a greater degree than both the mitochondrial genome and the remainder of the nuclear genome. While we found no evidence for signatures of local adaptation, we found evidence for genes and biological pathways with extreme patterns of introgression. The JAK/STAT signaling pathway was associated with putative barriers to reproduction. Given the extreme size difference between Kinda and grayfoot baboons and the importance of the JAK/STAT signaling pathway to growth

and body size variation, the JAK/STAT signaling pathway may underlie selection against either extremely large or extremely small size differences between male and female reproductive partners. The toll-like receptor pathway was associated with putative adaptive introgression, mirroring similar findings in humans and suggesting positive selection for novel pathogenic defenses. Finally, *ODF2* was identified as an intriguing candidate for adaptive introgression. Given its connection to sperm tail morphology and motility, our results suggest an effect in the hybrid zone favoring the grayfoot baboon sperm variant, which is potentially conditional on the presence of the invading Kinda baboon Y chromosome.

The low variance effective population size of the Y chromosome, which is reduced relative to autosomes due to the uniparental inheritance of the Y chromosome and further reduced by high reproductive skew (Charlesworth, 2009), may predict Y-chromosome introgression even in the absence of selection. It does not, however, predict the direction of introgression. Our findings in this study and previous studies (Jolly et al., 2011) suggest that the Kinda baboon Y chromosome is more prevalent in the hybrid zone, contradicting expectations by indicating that male Kinda baboons are more successful than grayfoot baboons approximately twice their body size at passing on their Y chromosomes. Given the importance of male competition in baboons (Barton et al., 1996), it seems unlikely that the Y chromosome would evolve under completely neutral conditions in the hybrid zone. For genetic drift to explain these findings, the strength of drift likely had to counteract and overcome the directional effect of selection. This is most plausible in cases where both the Y-chromosome effective population size is small and the strength of selection is weak, possibly due to mitigating factors such as directional reproductive incompatibilities or directionally impaired mate recognition systems (Jolly et al., 2011). Future research must weigh the relative strength of these two forces in order to better understand the unidirectional Y introgression in the baboon hybrid zone. Our findings, however, suggest one possible mechanism by which selection may act specifically to increase the frequency of the Kinda baboon Y chromosome relative to the mitochondrial genome and remainder of the nuclear genome due to epistatic selection on sperm traits in the Kafue river valley hybrid zone.

Acknowledgments

We thank the Zambia Wildlife Authority (now the Department of National Parks & Wildlife) and the University of Zambia for granting permission and providing support for fieldwork. We also thank Anna Weyher for sharing unpublished information about Kinda baboon behavior. This study was funded by the National Science Foundation (BCS 1341018, BCS 1029302, SMA 1338524), the Leakey Foundation, and the National Geographic Society. The Genome Technology Center at NYU is supported by NIH/NCATS UL1 TR00038 and NIH/NCI P30 CA016087.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martínez-Rodríguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A. W., Parisod, C., Pfennig, K. S., Rice, A. M., Ritchie, M. G., Seifert, B., Smadja, C. M., Stelkens, R., Szymura, J. M., Väinölä, R., Wolf, J. B. W., & Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246.

Abi-Rached, L., Jobin, M. J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., Babrzadeh, F., Gharizadeh, B., Luo, M., Plummer, F. A., Kimani, J., Carrington, M., Middleton, D., Rajalingam, R., Beksac, M., Marsh, S. G. E., Maiers, M., Guethlein, L. A., Tavoularis, S., Little, A.-M., Green, R. E., Norman, P. J., & Parham, P. (2011). The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science*, *334*, 89–94.

Akira, S., & Takeda, K. (2004). Toll-like receptor signalling. *Nature Reviews Immunology*, *4*, 499–511.

Alberts, S. C., Watts, H. E., & Altmann, J. (2003). Queuing and queue-jumping: long-term patterns of reproductive skew in male savannah baboons, *Papio cynocephalus*. *Animal Behaviour*, *65*, 821–840.

Albrechtová, J., Albrecht, T., Baird, S. J. E., Macholán, M., Rudolfsen, G., Munclinger, P., Tucker, P. K., & Piálek, J. (2012). Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *279*, 4803–4810.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664.

Altmann, S. A. (1962). A field study of the sociobiology of rhesus monkeys, *Macaca mulatta*. *Annals of the New York Academy of Sciences*, *102*, 338–435.

Ansell, W. F. H. (1960). *The Mammals of Northern Rhodesia*. Lusaka, Zambia: Government Printer.

Ansell, W. F. H. (1978). *The Mammals of Zambia*. Chilanga, Zambia: The National Parks & Wildlife Service.

Arnold, M. L., & Meyer, A. (2006). Natural hybridization in primates: one evolutionary mechanism. *Zoology*, *109*, 261–276.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*, 497–503.

Barton, R. A., Byrne, R. W., & Whiten, A. (1996). Ecology, feeding competition and social structure in baboons. *Behavioral Ecology and Sociobiology*, *38*, 321–329.

Ben-Eliezer, M., Phillip, M., & Gat-Yablonski, G. (2007). Leptin regulates chondrogenic differentiation in ATDC5 cell-line through JAK/STAT and MAPK pathways. *Endocrine*, *32*, 235–244.

Bergman, T. J., Phillips-Conroy, J. E., & Jolly, C. J. (2008). Behavioral variation and reproductive success of male baboons (*Papio anubis* × *Papio hamadryas*) in a hybrid social group. *American Journal of Primatology*, *70*, 136–147.

Brohmann, H., Pinnecke, S., & Hoyer-Fender, S. (1997). Identification and characterization of new cDNAs encoding outer dense fiber proteins of rat sperm. *Journal of Biological Chemistry*, *272*, 10327–10332.

Bulger, J. B. (1993). Dominance rank and access to estrous females in male savanna baboons. *Behaviour*, *127*, 67–103.

Burrell, A. S. (2009). *Phylogenetics and population genetics of central African baboons* (PhD thesis). New York University; New York University.

Busby, J. R. (1991). BIOCLIM - a bioclimate analysis and prediction system. *Plant Protection Quarterly*, *6*, 8–9.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 559.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*, 195–205.

Charpentier, M. J. E., Fontaine, M. C., Cherel, E., Renoult, J. P., Jenkins, T., Benoit, L., Barthès, N., Alberts, S. C., & Tung, J. (2012). Genetic structure in a dynamic baboon hybrid zone corroborates behavioral observations in a hybrid population. *Molecular Ecology*, *21*, 715–731.

Chiou, K. L., & Bergey, C. M. (2015). FecalSeq: methylation-based enrichment for noninvasive population genomics from feces. *bioRxiv*, 032870. doi:10.1101/032870.

Cohen, P. (2014). The TLR and IL-1 signalling network at a glance. *Journal of Cell Science*, *127*, 2383–2390.

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*, 1411–1423.

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Kähäri, A. K., Keenan, S., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Aken, B. L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S. M. J., Spudich, G., Trevanion, S. J., Yates, A., Zerbino, D. R., & Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research*, *43 (suppl. 1)*, D662–D669.

Dannemann, M., Andrés, A. M., & Kelso, J. (2016). Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *American Journal of Human Genetics*, *98*, 22–33.

Delson, E., Terranova, C. J., Jungers, W. L., Sargis, E. J., Jablonski, N. G., & Dechow, P. C. (2000). Body mass in Cercopithecidae (Primates, Mammalia): estimation and scaling in extinct and extant taxa. *Anthropological Papers of the American Museum of Natural History*, *83*, 1–159.

Dobzhansky, T. (1934). Studies on hybrid sterility. I. Spermatogenesis in pure and hybrid *Drosophila pseudoobscura*. *Zeitschrift für Zellforschung und Mikroskopische Anatomie*, *21*, 169–223.

Dunn, J., Cardini, A., & Elton, S. (2013). Biogeographic variation in the baboon: dissecting the cline. *Journal of Anatomy*, *223*, 337–352.

Eswarakumar, V. P., Lax, I., & Schlessinger, J. (2005). Cellular signaling by fibroblast growth factor receptors. *Cytokine & Growth Factor Reviews*, *16*, 139–149.

Farrar, M. A., & Schreiber, R. D. (1993). The molecular cell biology of interferon-γ and its receptor. *Annual Review of Immunology*, *11*, 571–611.

Fischer, J., Kopp, G. H., Dal Pesco, F., Goffe, A., Hammerschmidt, K., Kalbitzer, U., Klapproth, M., Maciej, P., Ndao, I., Patzelt, A., & Zinner, D. (2017). Charting the neglected West: the social system of Guinea baboons. *Yearbook of Physical Anthropology*, *63*, 15–31.

Fitzpatrick, J. L., Montgomerie, R., Desjardins, J. K., Stiver, K. A., Kolm, N., & Balshine, S. (2009). Female promiscuity promotes the evolution of faster sperm in cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 1128–1132.

Frost, S. R., Marcus, L. F., Bookstein, F. L., Reddy, D. P., & Delson, E. (2003). Cranial allometry,

phylogeography, and systematics of large-bodied papionins (Primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. *The Anatomical Record. Part A, Discoveries in Molecular, Cellular, and Evolutionary Biology*, *275A*, 1048–1072.

Gene Ontology Consortium. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*, 25–29.

Gene Ontology Consortium. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Research*, *43 (suppl. 1)*, D1049–D1056.

Ghoreschi, K., Laurence, A., & O'Shea, J. J. (2009). Janus kinases in immune cell signaling. *Immunological Reviews*, *228*, 273–287.

Ginsberg, J. R., & Huck, U. W. (1989). Sperm competition in mammals. *Trends in Ecology & Evolution*, *4*, 74–79.

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*, 2111–2127.

Gompert, Z., & Buerkle, C. A. (2012). bgc: software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*, 1168–1176.

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*, 205–220.

Hancock, A. M., Witonsky, D. B., Gordon, A. S., Eshel, G., Pritchard, J. K., Coop, G., & Di Rienzo, A. (2008). Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genetics*, *4*, e32.

Harcourt, A. H., Harvey, P. H., Larson, S. G., & Short, R. V. (1981). Testis weight, body weight and breeding system in primates. *Nature*, *293*, 55–57.

Harrison, R. G. (1990). Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology*, *7*, 69–128.

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*, 4606–4618.

Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, *31*, 1109–1144.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

Hoyer-Fender, S., Petersen, C., Brohmann, H., Rhee, K., & Wolgemuth, D. J. (1998). Mouse Odf2 cDNAs consist of evolutionary conserved as well as highly variable sequences and encode outer dense fiber proteins of the sperm tail. *Molecular Reproduction and Development*, *51*, 167–175.

Immler, S., Hamilton, M. B., Poslusny, N. J., Birkhead, T. R., & Epifanio, J. M. (2011). Post-mating reproductive barriers in two unidirectionally hybridizing sunfish (Centrarchidae: *Lepomis*). *Journal of Evolutionary Biology*, *24*, 111–120.

Jolly, C. J. (1993). Species, subspecies, and baboon systematics. In Kimbel, W.H., Martin, L.B. (Eds.), *Species, Species Concepts, and Primate Evolution* (pp. 67–107). New York: Plenum Press.

Jolly, C. J. (2001). A proper study for mankind: analogies from the papionin monkeys and their implications for human evolution. *Yearbook of Physical Anthropology*, *44*, 177–204.

Jolly, C. J., & Phillips-Conroy, J. E. (2003). Testicular size, mating system, and maturation schedules in wild anubis and hamadryas baboons. *International Journal of Primatology*, *24*, 125–142.

Jolly, C. J., & Phillips-Conroy, J. E. (2006). Testicular size, developmental trajectories, and male life history strategies in four baboon taxa. In Swedell, L., Leigh, S.R. (Eds.), *Reproduction and Fitness in Baboons: Behavioral, Ecological, and Life History Perspectives* (pp. 257–275). New York: Springer.

Jolly, C. J., Burrell, A. S., Phillips-Conroy, J. E., Bergey, C., & Rogers, J. (2011). Kinda baboons (*Papio kindae*) and grayfoot chacma baboons (*P. ursinus griseipes*) hybridize in the Kafue river valley, Zambia. *American Journal of Primatology*, *73*, 291–303.

Kamilar, J. M. (2006). Geographic variation in savanna baboon (*Papio*) ecology and its taxonomic and evolutionary implications. In Lehman, S.M., Fleagle, J.G. (Eds.), *Primate Biogeography: Progress and Prospects* (pp. 169–200). New York: Springer.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241.

Kim, R. N., Kim, D.-W., Choi, S.-H., Chae, S.-H., Nam, S.-H., Kim, D.-W., Kim, A., Kang, A., Park, K.-H., Lee, Y. S., Hirai, M., Suzuki, Y., Sugano, S., Hashimoto, K., Kim, D.-S., & Park, H.-S. (2011). Major chimpanzee-specific structural changes in sperm development-associated genes. *Functional & Integrative Genomics*, *11*, 507–517.

Kummer, H. (1968). *Social Organization of Hamadryas Baboons: A Field Study*. Chicago: University of Chicago Press.

Kutsukake, N., & Nunn, C. L. (2006). Comparative tests of reproductive skew in male primates: the roles of demographic factors and incomplete control. *Behavioral Ecology and Sociobiology*, *60*, 695–706.

Liu, B., Liao, J., Rao, X., Kushner, S. A., Chung, C. D., Chang, D. D., & Shuai, K. (1998). Inhibition of Stat1-mediated gene activation by PIAS1. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 10626–10631.

Lönnberg, E. (1919). Contributions to the knowledge about the monkeys of Belgian Congo. *Revue*

Zoologique Africaine, *7*, 107–154.

Mariappa, D., Aladakatti, R. H., Dasari, S. K., Sreekumar, A., Wolkowicz, M., Hoorn, F. van der, & Seshagiri, P. B. (2010). Inhibition of tyrosine phosphorylation of sperm flagellar proteins, outer dense fiber protein-2 and tektin-2, is associated with impaired motility during capacitation of hamster spermatozoa. *Molecular Reproduction and Development*, *77*, 182–193.

Mehta, P. K., & Griendling, K. K. (2007). Angiotensin II cell signaling: physiological and pathological effects in the cardiovascular system. *American Journal of Physiology. Cell Physiology*, *292*, C82–C97.

Mi, H., Muruganujan, A., & Thomas, P. D. (2013a). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, *41 (suppl. 1)*, D377–D386.

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013b). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, *8*, 1551–1566.

Montesano, R., Vassalli, J.-D., Baird, A., Guillemin, R., & Orci, L. (1986). Basic fibroblast growth factor induces angiogenesis *in vitro*. *Proceedings of the National Academy of Sciences of the United States of America*, *83*, 7297–7301.

Nilsson, A., Ohlsson, C., Isaksson, O. G., Lindahl, A., & Isgaard, J. (1994). Hormonal regulation of longitudinal bone growth. *European Journal of Clinical Nutrition*, *48 (suppl. 1)*, S150–S158.

Nychka, D., Furrer, R., Paige, J., & Sain, S. (2015). *fields: tools for spatial data*. doi:10.5065/D6W957CT.

Oh, H. J., Kido, T., & Lau, Y. F. C. (2007). PIAS1 interacts with and represses SOX9 transactivation activity. *Molecular Reproduction and Development*, *74*, 1446–1455.

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

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

Phillips-Conroy, J. E., & Jolly, C. J. (1986). Changes in the structure of the baboon hybrid zone in the Awash National Park, Ethiopia. *American Journal of Physical Anthropology*, *71*, 337–350.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11.

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

R Core Team. (2013). R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org.

Rallapalli, P. M., Orengo, C. A., Studer, R. A., & Perkins, S. J. (2014). Positive selection during the evolution of the blood coagulation factors in the context of their disease-causing mutations. *Molecular Biology and Evolution*, *31*, 3040–3056.

Sahni, M., Ambrosetti, D.-C., Mansukhani, A., Gertner, R., Levy, D., & Basilico, C. (1999). FGF signaling inhibits chondrocyte proliferation and regulates bone development through the STAT-1 pathway. *Genes & Development*, *13*, 1361–1366.

Schenone, M., Furie, B. C., & Furie, B. (2004). The blood coagulation cascade. *Current Opinion in Hematology*, *11*, 272–277.

Schmera, D., Pizá, J., Reinartz, E., Ursenbacher, S., & Baur, B. (2016). Breeding system, shell size and age at sexual maturity affect sperm length in stylommatophoran gastropods. *BMC Evolutionary Biology*, *16*, 89.

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2009). The benefits of social capital: close social bonds among female baboons enhance offspring survival. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *276*, 3099–3104.

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2010). Female chacma baboons form strong, equitable, and enduring social bonds. *Behavioral Ecology and Sociobiology*, *64*, 1733–1747.

Simonti, C. N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D. S., Chisholm, R. L., Crosslin, D. R., Hebbring, S. J., Jarvik, G. P., Kullo, I. J., Li, R., Pathak, J., Ritchie, M. D., Roden, D. M., Verma, S. S., Tromp, G., Prato, J. D., Bush, W. S., Akey, J. M., Denny, J. C., & Capra, J. A. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. *Science*, *351*, 737–741.

Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., Schlebusch, S. A., Wall, J. D., Alberts, S. C., Mukherjee, S., Zhou, X., & Tung, J. (2016). Efficient genome-wide sequencing and low coverage pedigree analysis from non-invasively collected samples. *Genetics*, *203*, 699–714.

Soung, N.-K., Kang, Y. H., Kim, K., Kamijo, K., Yoon, H., Seong, Y.-S., Kuo, Y.-L., Miki, T., Kim, S. R., Kuriyama, R., Giam, C.-Z., Ahn, C. H., & Lee, K. S. (2006). Requirement of hCenexin for proper mitotic functions of polo-like kinase 1 at the centrosomes. *Molecular and Cellular Biology*, *26*, 8316–8335.

Tarnasky, H., Cheng, M., Ou, Y., Thundathil, J. C., Oko, R., & Hoorn, F. A. van der. (2010). Gene trap mutation of murine Outer dense fiber protein-2 gene can result in sperm tail abnormalities in mice with high percentage chimaerism. *BMC Developmental Biology*, *10*, 67.

Tosi, A. J., Morales, J. C., & Melnick, D. J. (2002). Y-chromosome and mitochondrial markers

in *Macaca fascicularis* indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in the Isthmus of Kra. *International Journal of Primatology*, *23*, 161–178.

Turner, L. M., Schwahn, D. J., & Harr, B. (2011). Reduced male fertility is common but highly variable in form and severity in a natural house mouse hybrid zone. *Evolution*, *66*, 443–458.

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*, e72.

Vottero, A., Guzzetti, C., & Loche, S. (2013). New aspects of the physiology of the GH-IGF-1 axis. *Endocrine Development*, *24*, 96–105.

Ward, A. C., Touw, I., & Yoshimura, A. (2000). The Jak-Stat pathway in normal and perturbed hematopoiesis. *Blood*, *95*, 19–29.

Watson, C. J., & Burdon, T. G. (1996). Prolactin signal transduction mechanisms in the mammary gland: the role of the Jak/Stat pathway. *Reviews of Reproduction*, *1*, 1–5.

Wedell, N., Gage, M. J. G., & Parker, G. A. (2002). Sperm competition, male prudence and sperm-limited females. *Trends in Ecology & Evolution*, *17*, 313–320.

Weyher, A. H., Phillips-Conroy, J. E., Fourrier, M. S., & Jolly, C. J. (2014). Male-driven grooming bouts in mixed-sex dyads of Kinda baboons (*Papio kindae*). *Folia Primatologica*, *85*, 178–191.

Zinner, D., Arnold, M. L., & Roos, C. (2011). The strange blood: natural hybridization in primates. *Evolutionary Anthropology*, *20*, 96–103.

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.

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*, 133–140.

Chapter 5

Summary and Conclusions

The frequency and impacts of hybridization are increasingly appreciated in the evolution of numerous lineages including humans. In this dissertation, I aimed to examine the patterns and processes of introgressive hybridization in a Zambian baboon study system in which the species exhibit numerous phenotypic dissimilarities including extreme differences in body size. In order to accomplish this, I employed a population genetic approach using thousands of genome-wide variants to characterize the population structure of the hybrid zone and to scan for regions exhibiting signatures of selection.

As is true for many studies of wild populations, sampling constraints from unhabituated animals represented a substantial obstacle, particularly for addressing the genomic-scale questions posed by this dissertation (Kohn & Wayne, 1997; Perry et al., 2010). When this study was initiated, high-throughput sequencing of baboon DNA was feasible in practice only with invasively obtained samples such as blood while population-level sampling of unhabituated baboons was feasible in practice only with noninvasively obtained samples such as feces. In order to address the aims of this dissertation, I developed in collaboration with Christina Bergey a novel technique for high-throughput sequencing of baboon DNA from feces (Chiou & Bergey, 2015). The development of this method is the subject of Chapter 2 of this dissertation.

The primary obstacle barring efficient sequencing of mammalian DNA from feces was the dominating presence of exogenous DNA of mainly bacterial origin. Our technique, which we call "FecalSeq", addresses this challenge by enriching mammalian DNA through a methylation-based capture procedure that takes advantage of substantial differences in CpG methylation density be-

tween mammalian and bacterial genomes (Feehery et al., 2013). We demonstrate using dozens of fecal samples that FecalSeq effectively partitions host DNA from contaminating bacterial DNA, enabling efficient and inexpensive high-throughput sequencing of the target host genome, particularly in combination with additional downstream genome complexity-reduction techniques such as ddRADseq (Peterson et al., 2012). By comparing genotyping data in individuals with both blood- and feces-derived genotypes, we demonstrate that FecalSeq produces genotyping data that are highly concordant with data derived from blood.

I next used FecalSeq and ddRADseq to derive genome-wide genotyping data from animals within and surrounding the Kafue National Park baboon hybrid zone in Zambia, where Kinda baboons (*Papio kindae*) and grayfoot baboons (*Papio griseipes*) meet (Jolly et al., 2011). Using an unsupervised population clustering approach (Alexander et al., 2009), I identified samples from individuals with pure Kinda and grayfoot ancestry. I then used a population scan approach to identify regions that were extremely differentiated between populations (Akey et al., 2002). This analysis is the subject of Chapter 3 of this dissertation.

Given the extreme differences in body size between Kinda and grayfoot baboons, I expected to find differentiation of genes and pathways with known effects on body size. I identified based on F_{ST} several candidate genes related to growth or metabolic processes. These included *FGF1*, which is a potent mitogen that is also involved in organ induction. Gene Ontology analysis identified an overall signal for differentiation in genes related to multicellular organism growth and developmental growth. PANTHER pathway analysis identified a signal for differentiation in genes belonging to the JAK/STAT signaling pathway. JAK/STAT signaling mediates signals for a variety of pathways, but plays an important role in longitudinal growth through chondrocyte proliferation.

I finally used a combined dataset of pure Kinda baboons, pure grayfoot baboons, and baboons of mixed ancestry to characterize the structure of the Kafue National Park baboon hybrid zone and to identify genomic regions putatively under selection. These analyses are the subject of Chapter 4 of this dissertation.

I found that in accordance with previous reports (Jolly et al., 2011), Kinda and grayfoot baboons form a wide geographic cline spanning at least 100 km. By comparing ancestry patterns from

my genome-wide autosomal dataset to previously reported ancestry patterns from mitochondrial DNA and the Y chromosome, I found that that autosomal ancestry largely resembled mitochondrial ancestry, while the Y chromosome exhibited a significantly different pattern with extreme introgression of the Kinda baboon Y chromosome. Bayesian methods for detecting local adaptation (Coop et al., 2010) failed to detect candidates for local adaptation, while Bayesian methods for detecting extreme introgression (Gompert & Buerkle, 2011) identified candidates for both isolating mechanisms and adaptive introgression. Notably, JAK/STAT signaling was associated with overall steep clines, suggesting that it may underlie selection against hybrid offspring of parents with extremely large or extremely small body size differences. The toll-like receptor pathway was associated with overall wide clines and excess grayfoot baboon ancestry, suggesting adaptive introgression of pathogen-related defenses mirroring similar findings in humans (Dannemann et al., 2016). The sperm-tail gene *ODF2* was also identified as a candidate for adaptive introgression with excess grayfoot baboon ancestry.

At the start of this dissertation project, it was evident that Kinda and grayfoot baboons exhibited marked phenotypic differences particularly with regard to body size. I therefore expected to find signatures of selection acting on body size in both parental species and in hybrids. Despite an approach that included a relatively small fraction of known genes, I nevertheless found robust signals of selection on body size in both parental species and hybrids. This includes selection on components of the JAK/STAT signaling pathway. JAK/STAT signaling plays a critical role in organism growth by mediating cytokine signals in multiple growth-related pathways. In my analyses, JAK/STAT signaling was associated with both extreme differentiation among pure species and with reduced introgression in hybrids. This suggests that selection on body size differences between Kinda and grayfoot baboons has acted on areas of the genome involved in JAK/STAT signaling. It also suggests that, as predicted, there is selection against body size differences in the hybrid zone (Jolly et al., 2011).

While selection on JAK/STAT signaling is compelling given its critical role in growth-related pathways, it serves diverse functions and is involved in numerous other pathways. My analyses revealed significant enrichment of differentiation (F_{ST}) between the genomes of Kinda and grayfoot

baboons in JAK/STAT signaling. My analyses, however, did not necessarily include the functional variants directly under selection. In fact, it is more likely that my analyses revealed candidate regions in linkage disequilibrium with functional variants under selection rather than the targets of selection themselves. Future study should aim to sequence more completely components of the JAK/STAT signaling pathway in Kinda and grayfoot baboon genomes, and to use haplotype trees and tree scanning methods to find variants associated with the hypothesized contrasting phenotypes. Aside from revealing potential selection on additional components not included in my dataset, such an approach will highlight variants in coding or regulatory (e.g., promoter or enhancer) regions that putatively underlie differences in body size. Once these variants are identified, functional cellular assays may yield valuable information to evaluate the hypothesis that these genes underlie body size differences in these two species.

One of the most striking findings in the Kafue National Park baboon hybrid zone is the extreme introgression of the Kinda baboon Y chromosome. This finding was previously reported by Jolly et al. (2011) based on numerous cases of mitochondrial DNA and Y-chromosome discordance. Jolly et al. (2011) interpreted these findings mainly in terms of a distinct male Kinda baboon advantage in the hybrid in traits related to body size. For instance, the large grayfoot baboon body size phenotype may introduce obstetric challenges disproportionately affecting male grayfoot $\times$ female Kinda baboon reproduction. The small appearance of male Kinda-like baboons may also allow them to "sneak" copulations without detection by much larger male grayfoot-like baboons, who may not recognize them as mature adults. Conversely, the small appearance of adult female Kinda-like baboons and the undeveloped appearance of their peak sexual swellings may also appear to be immature to male grayfoot-like baboons, compounding this effect.

The disproportionate success of male Kinda baboons would explain the high degree of introgression of the Kinda Y chromosome relative to the mitochondrial genome. If male Kinda baboons are disproportionately successful, then this success should also be reflected in a higher frequency of Kinda baboon autosomal variants relative to mitochondrial DNA. I therefore predicted that autosomal ancestry estimates would be intermediate between mitochondrial and Y-chromosome ancestry estimates. Curiously, I did not find evidence to support this prediction, but instead found that

autosomal ancestry was similar to mitochondrial DNA ancestry while Y-chromosome ancestry was dissimilar to both. This finding implies that the Kinda baboon Y chromosome has penetrated the species boundary to a high degree relative to both mitochondrial DNA and the remainder of the nuclear genome.

Three hypotheses might explain this finding. First, the extreme introgression of the Kinda baboon Y chromosome could result from strong genetic drift due to the small variance effective population size of the Y chromosome, which could be reduced even further in cases with high male reproductive skew. Alternatively, male Kinda baboons might be disproportionately successful as suggested by Jolly et al. (2011), but without sustained immigration into the hybrid zone, repeated backcrossing between male hybrid and female grayfoot baboons might have sufficiently reduced the proportion of Kinda baboon autosomal alleles while retaining a relatively high representation of the Kinda baboon Y chromosome. Alternatively, the extreme introgression of the Kinda baboon Y chromosome may indicate selection acting directly on the Y chromosome. Interestingly, this scenario is supported in the European house mouse hybrid zone, where the high penetration of the *Mus m. musculus* Y chromosome relative to *Mus m. domesticus* is explained to a high degree by its effect in "rescuing" the otherwise high hybrid dysgenesis of sperm traits (Albrechtová et al., 2012). Curiously, my analysis of the hybrid zone also revealed selection acting to favor grayfoot baboon sperm-related variants in *ODF2*, posing an intriguing possibility that a similar epistatic interaction between grayfoot baboon *ODF2* and Kinda baboon Y chromosomes may be under positive selection in the context of widespread sperm-related hybrid dysgenesis. Notably, this hypothesis would explain the mitochondrial/Y-chromosome discordance reported by Jolly et al. (2011) without implying an advantage of male Kinda baboon phenotypes related to body size.

The inclusion of mitochondrial and Y-chromosome genotypes for my samples was not feasible due to logistical constraints and a direct comparison of individual-level ancestry for all three marker sets was therefore impossible. Instead, ancestry comparisons were achieved indirectly through interpolation of ancestry estimates from all three marker sets, with subsequent comparison of interpolated estimates assigned to geographic localities. In order to test the latter hypothesis proposed above, future study should test for a relationship between sperm-related genes and the Y chro-

mosome, with the prediction that grayfoot baboon sperm-related variants are disproportionately associated with Kinda baboon Y chromosomes. Importantly, for this hypothesis to be parsimonious, the strength of directional selection must exceed the strength of genetic drift.

This dissertation makes a substantial methodological contribution by facilitating population-level genomic analyses of wild animals without the logistical challenges inherent to invasive sampling. The genomic approach employed in this dissertation identifies genetic regions of interest that may be more directly targeted in future studies for further analysis. It is therefore only an initial step in deconstructing the genomic architecture of fitness-related traits. Future functional studies are necessary to corroborate and elucidate the functions of candidate genes and pathway families identified in this dissertation.

This dissertation partially addresses basic questions about the Kafue National Park baboon hybrid zone while raising a larger number of new ones. What is the age of the Kafue National Park hybrid zone? How many waves of hybridization have occurred? What is the relationship of the hybrid zone to the ecology of the area? To what extent has the hybrid zone been influenced by anthropogenic forces such as agriculture or the formation of the Itezhi Tezhi Dam? What is the structure of the hybrid zone in the wider region, including the area east of the Kafue River and the areas west of the more developed eastern area of the national park?

Most of these questions are important not only for understanding the process of divergence and speciation in baboons, but also the evolution of modern and archaic humans in the Pleistocene. The African expansion and diversification of baboons in the past 1.5 - 2.0 million years mirrors the expansion of *Homo* following *Homo erectus*, as does the occurrence of secondary contact and genetic exchange among populations. One critical question is whether the cladogenesis that characterized baboon diversification also characterized the human expansion. This question is the subject of considerable debate in the literature (Holliday et al., 2014) but has profound implications for the application of the baboon analogy. If humans expanded with cladogenesis, then baboon evolution resembles human evolution and the study of extant baboon populations may be an effective proxy for modeling the dynamics of speciation and hybridization in modern and archaic populations in the Pleistocene. If humans expanded without cladogenesis, however, then baboon evolution offers

a model of human "alternative history" and future attention should focus on the processes and conditions that led to cladogenesis in baboons but not in humans. Future work on the Kafue National Park hybrid zone will require not only more comprehensive genomic information, but also information about grayfoot baboon and particularly Kinda baboon behavior, ecology, and physiology in order to lend future valuable insights necessary for addressing these lingering questions.

References

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, *12*, 1805–1814.

Albrechtová, J., Albrecht, T., Baird, S. J. E., Macholán, M., Rudolfsen, G., Munclinger, P., Tucker, P. K., & Piálek, J. (2012). Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *279*, 4803–4810.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664.

Chiou, K. L., & Bergey, C. M. (2015). FecalSeq: methylation-based enrichment for noninvasive population genomics from feces. *bioRxiv*, 032870. doi:10.1101/032870.

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*, 1411–1423.

Dannemann, M., Andrés, A. M., & Kelso, J. (2016). Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *American Journal of Human Genetics*, *98*, 22–33.

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A., & Pradhan, S. (2013). A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One*, *8*, e76096.

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*, 2111–2127.

Holliday, T. W., Gautney, J. R., & Friedl, L. (2014). Right for the wrong reasons: reflections on modern human origins in the post-Neanderthal genome era. *Current Anthropology*, *55*, 696–724.

Jolly, C. J., Burrell, A. S., Phillips-Conroy, J. E., Bergey, C., & Rogers, J. (2011). Kinda baboons (*Papio kindae*) and grayfoot chacma baboons (*P. ursinus griseipes*) hybridize in the Kafue river

valley, Zambia. *American Journal of Primatology*, *73*, 291–303.

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

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

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

Appendices

Appendix A

Supplementary Figures

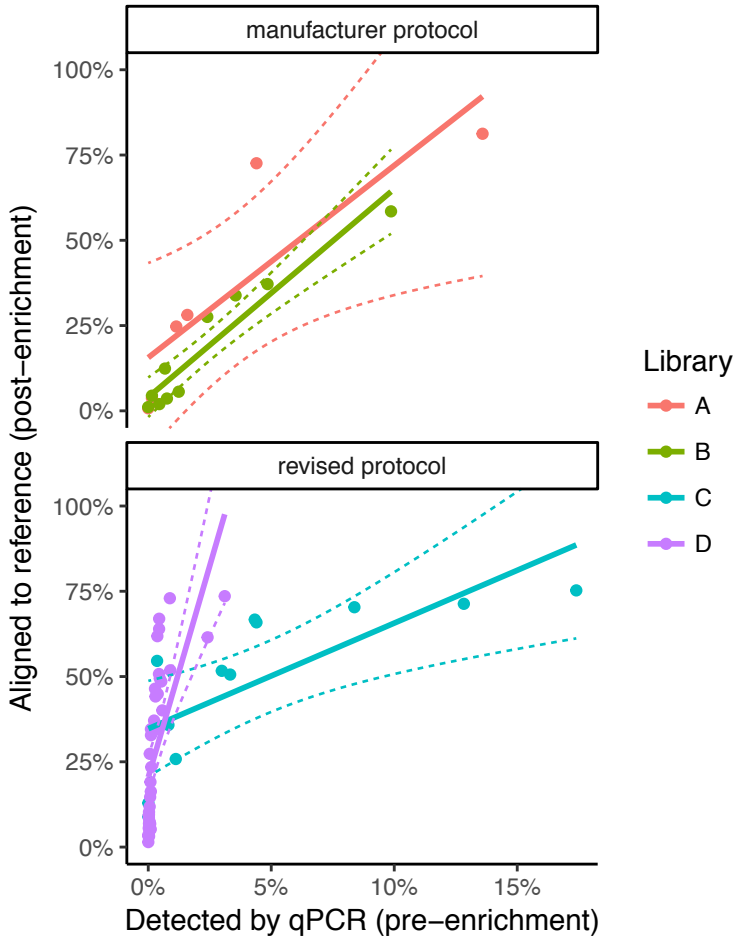

Figure S1: Relationship between pre-enrichment host percentage, as estimated by quantitative PCR, and post-enrichment host percentage, as estimated by alignment of sequencing reads to the baboon reference genome.

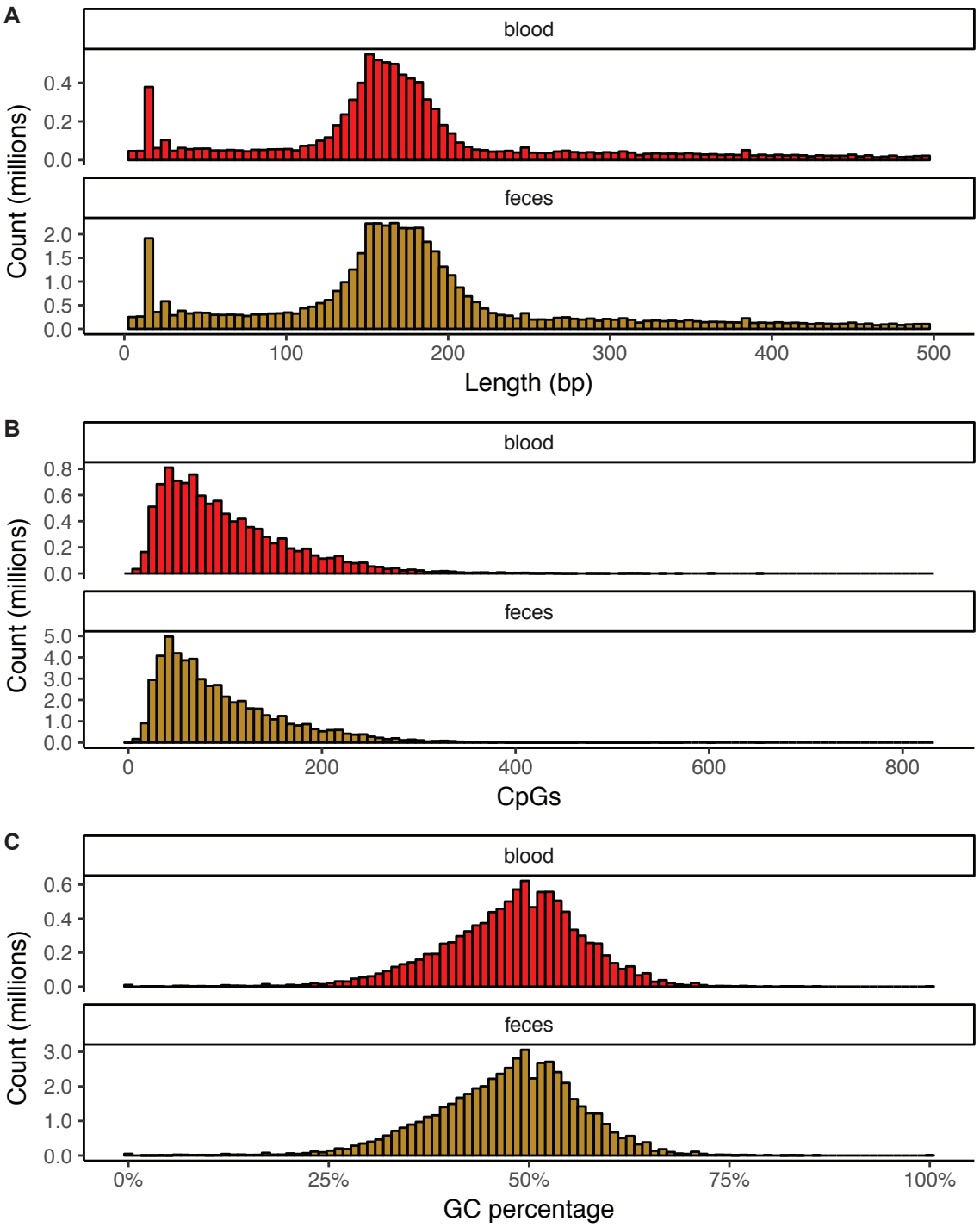

Figure S2: Combined distributions of (a) RADtag lengths, (b) CpG counts (within the boundaries of the sequenced RADtag ± 5,000), and (c) GC percentages in sequenced libraries.

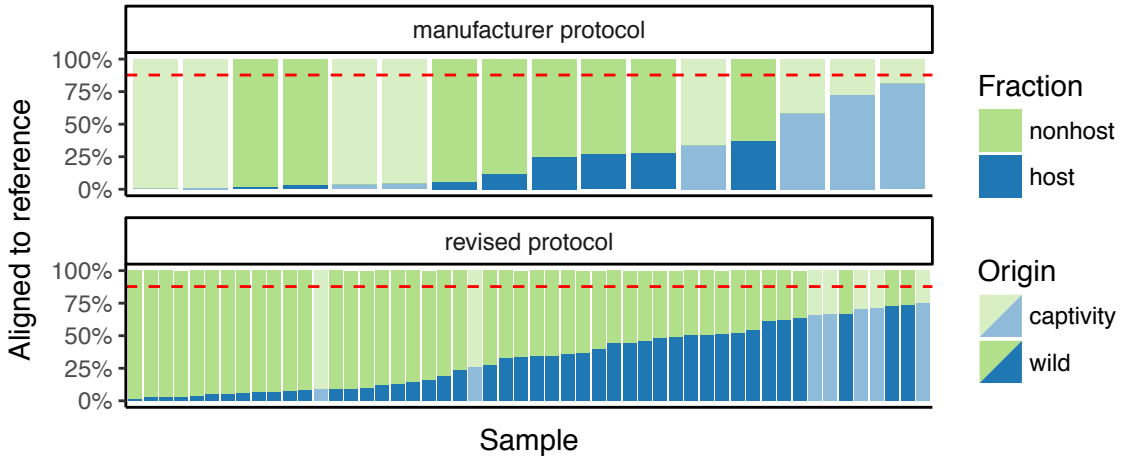

Figure S3: Percentage of reads mapping to the baboon reference genome (papAnu2) for all samples included in this study. Sixteen samples were enriched using the manufacturer protocol and 52 using the revised protocol.

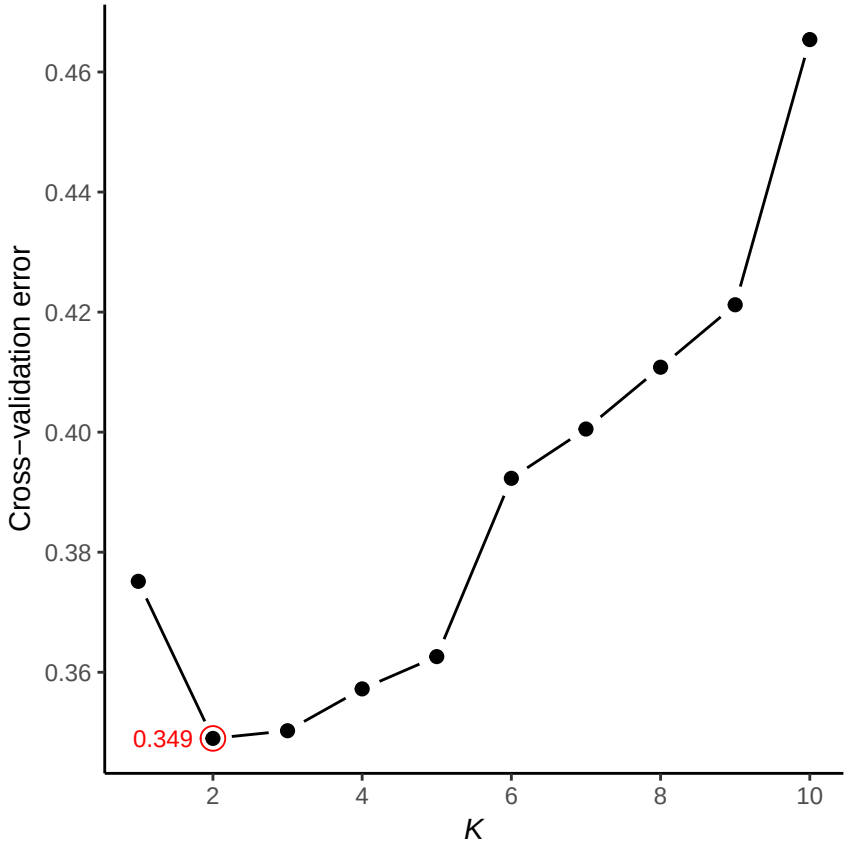

Figure S4: Cross-validated error results from ADMIXTURE runs in which K varied from 1 to 10.

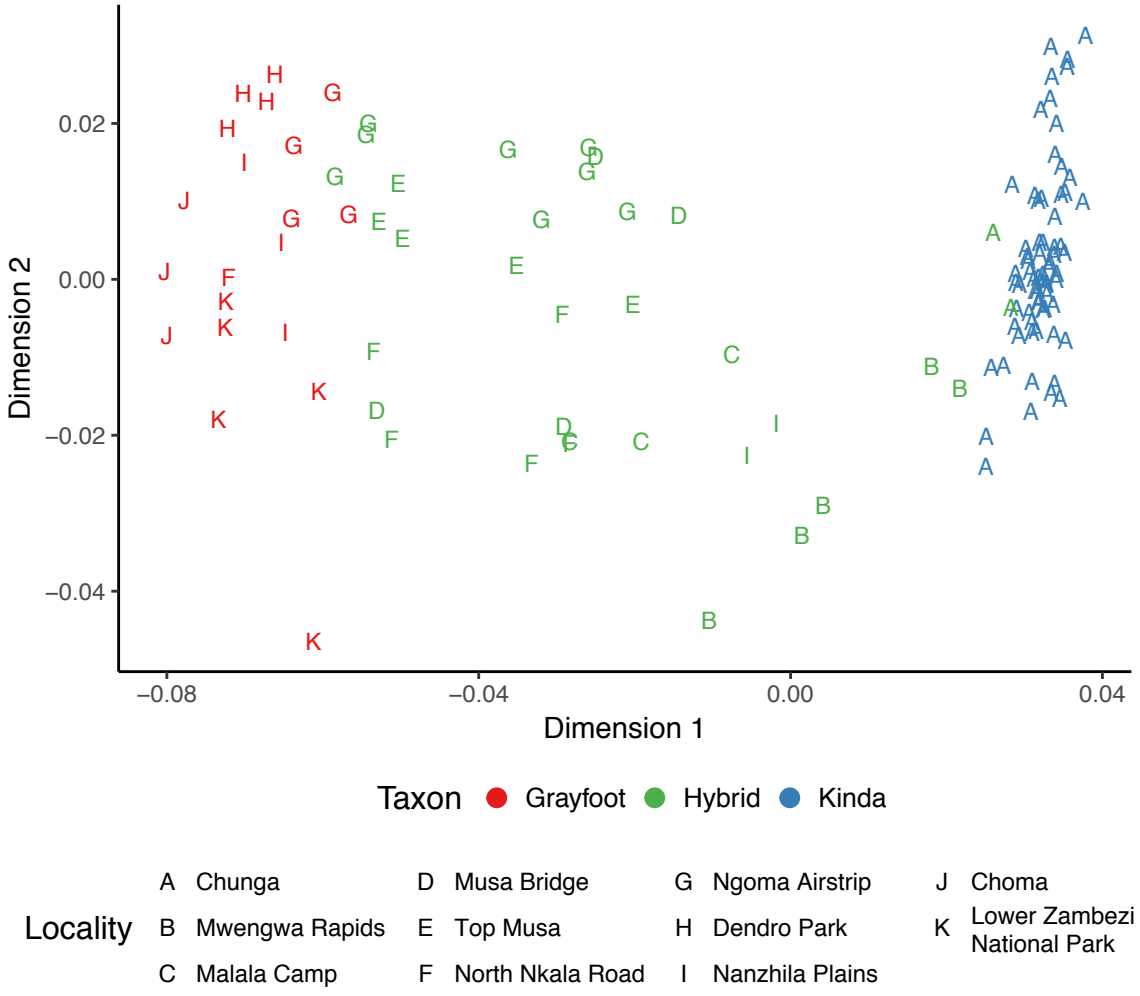

Figure S5: Multidimensional scaling of identity-by-state: first two dimensions. Points are colored based on their taxonomy inferred using ADMIXTURE (see Figure S7).

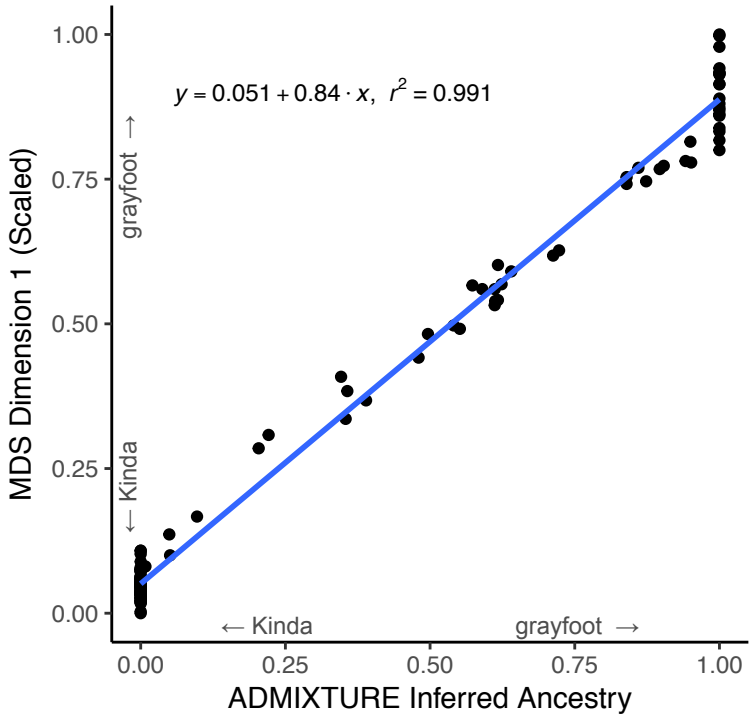

Figure S6: Concordance between multidimensional scaling and ADMIXTURE analysis. The first dimension of the MDS was first scaled from 0 to 1.

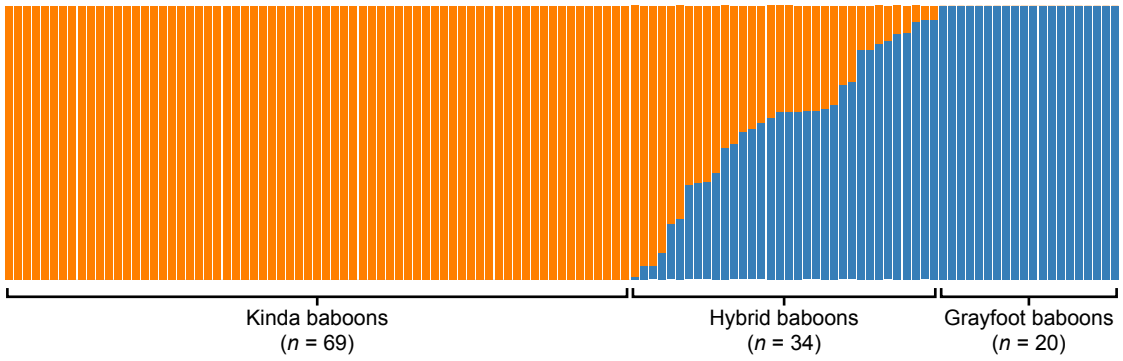

Figure S7: Taxonomic assignment based on ancestry estimates calculated using ADMIXTURE.

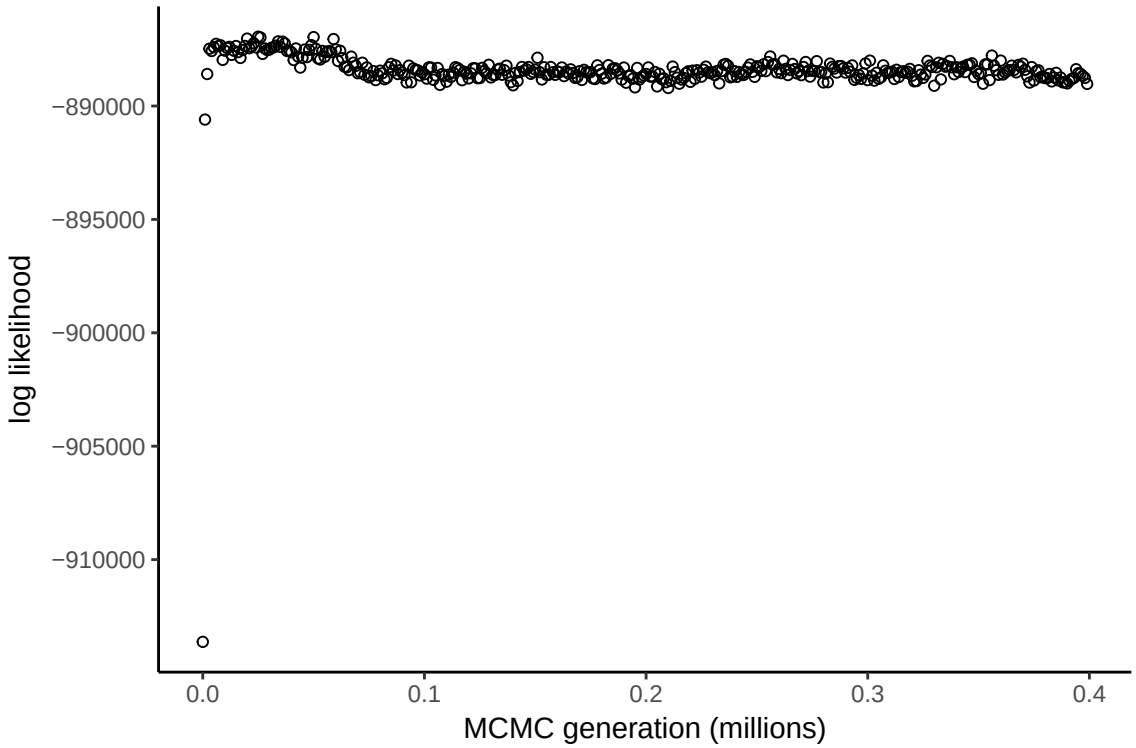

Figure S8: A representative example of a converged complete bgc chain of 400,000 generations. For this plot, chains are shown every 1,000 generations.

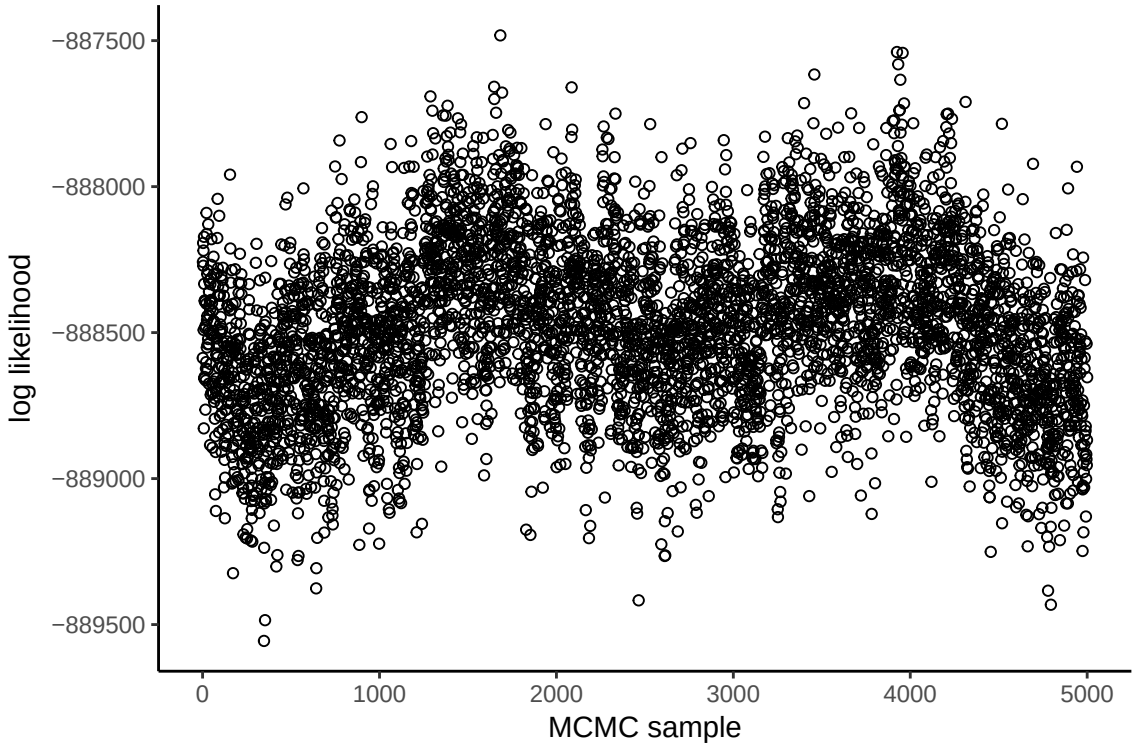

Figure S9: A representative example of a converged posterior BGC chain of 5,000 samples. This chain represents the last 200,000 generations of the chain shown in Figure S8, sampled every 40 generations.

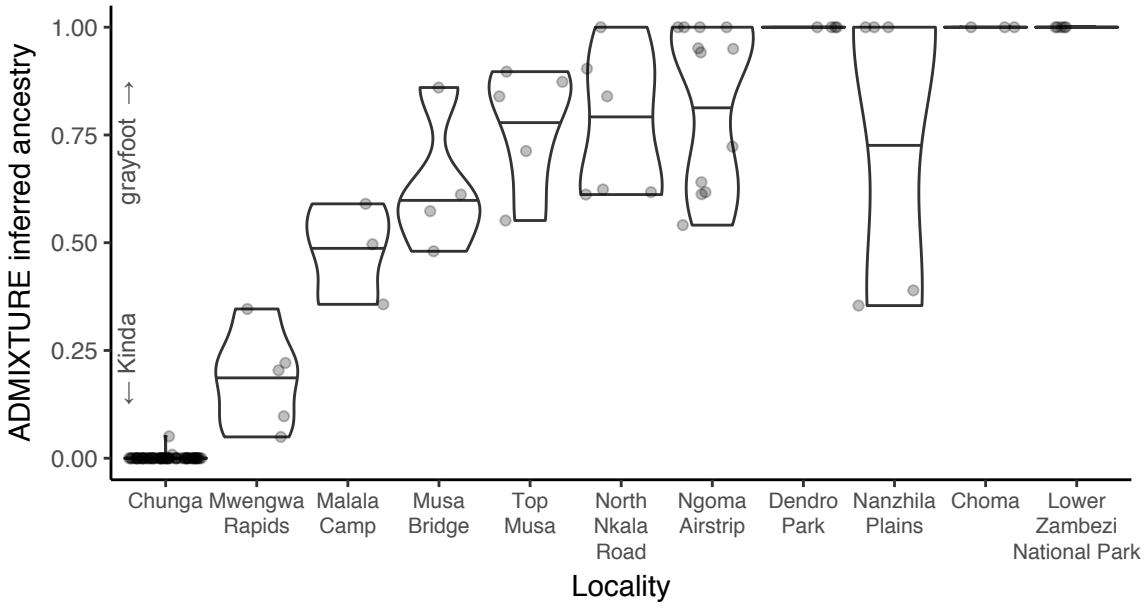

Figure S10: ADMIXTURE ancestry estimation results by locality.

Figure S11: Baboons from Malala Camp resemble Kinda baboons superficially, but are intermediate in size and darker in coloration. In the right pane, the male in the foreground superficially resembles a grayfoot baboon.

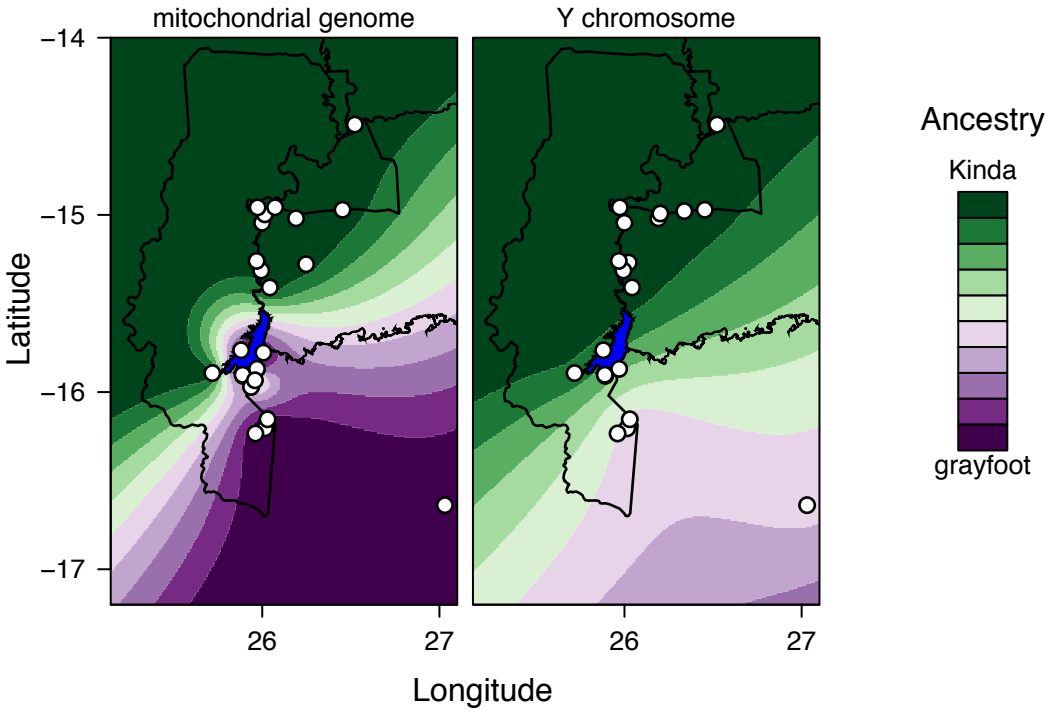

Figure S12: Geospatial interpolation of ancestry estimates for mitochondrial DNA and the Y chromosome. Mitochondrial and Y genotypes from Jolly et al. (2011) were interpolated over the study area at 1 km^2 resolution using a Kriging surface model procedure.

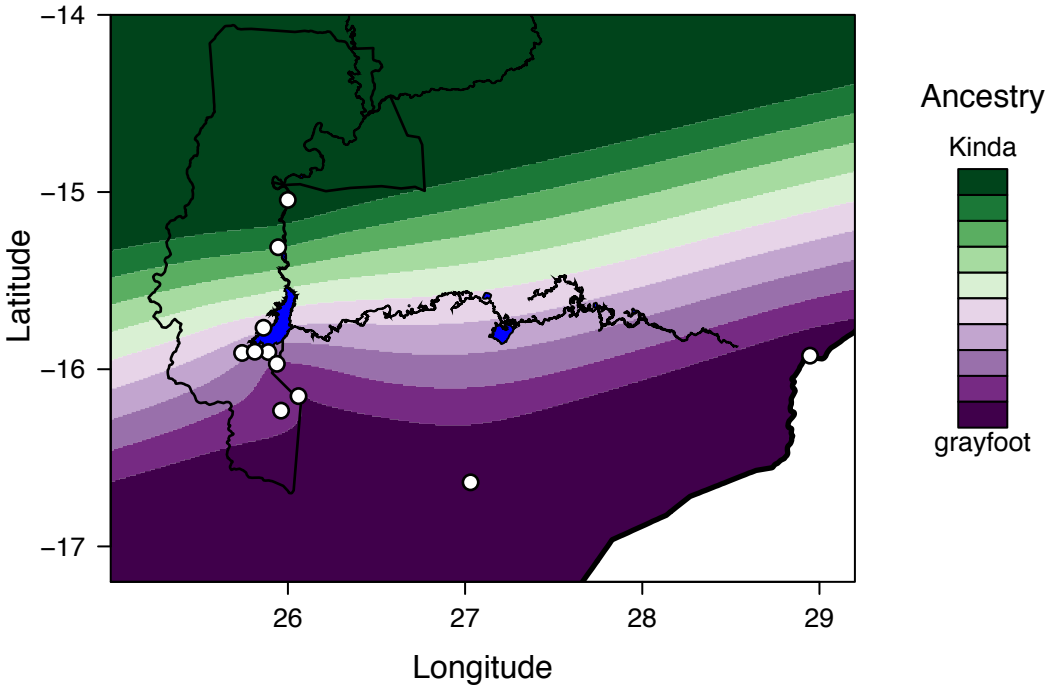

Figure S13: Geospatial interpolation of ancestry estimates for autosomal SNPs genotyped by ddRADseq. Ancestry estimates obtained from ADMIXTURE analysis (Figure 4.2) were interpolated over the study area at 1 km^2 resolution using a Kriging surface model procedure. The geographic limit of this figure has been extended to the east in order to include the grayfoot baboon sites of Choma and Lower Zambezi National Park. As a consequence of this, ancestry estimates over much of the eastern half of this map are extrapolated and should be interpreted with caution, as they are almost certainly inaccurate.

Figure S14: Baboons from Lubalunsuki Hill in the present day resemble Kinda baboons.

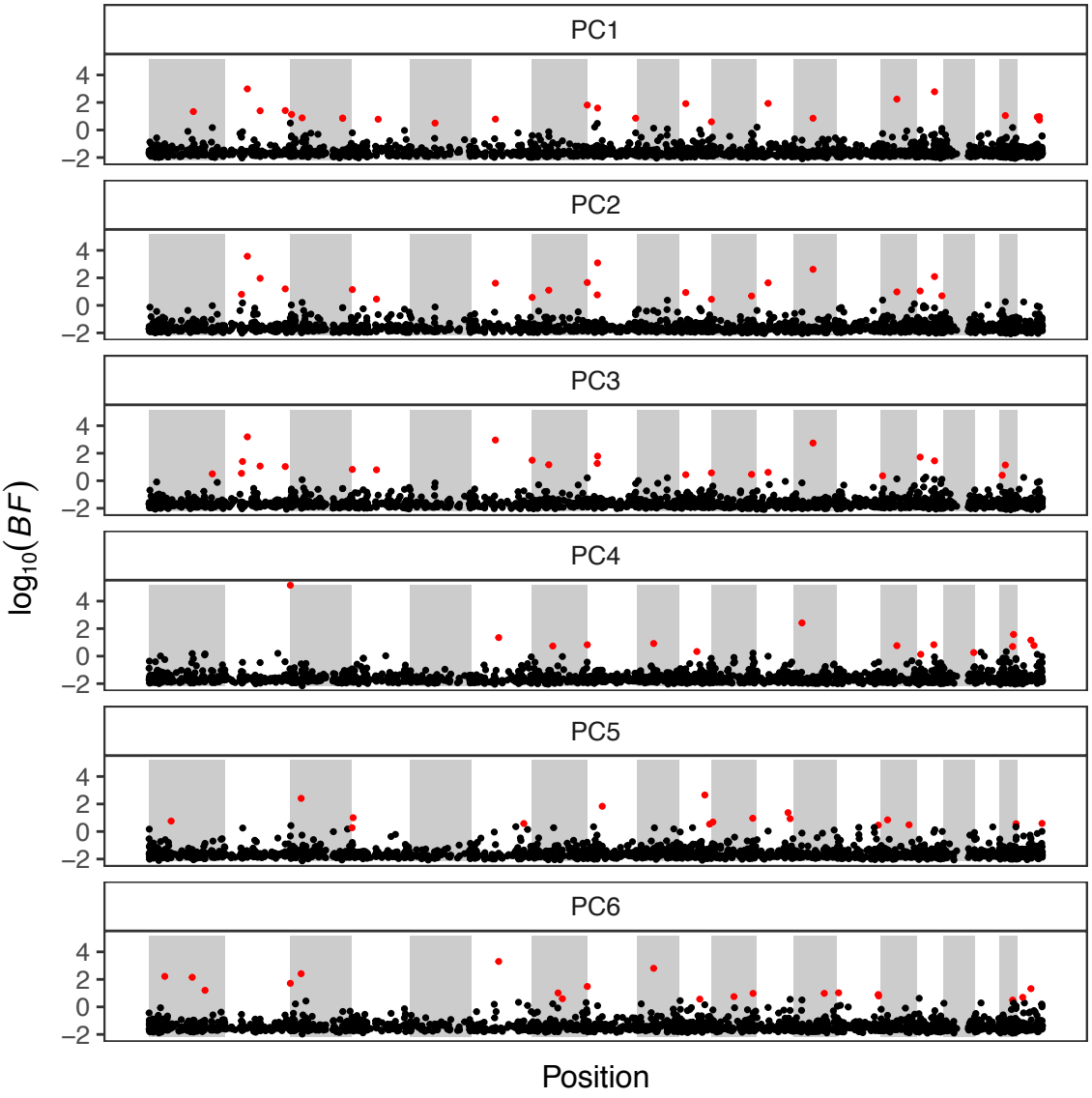

Figure S15: Distribution of Bayes factor (BF_{ij}) estimates for protein-coding genes over six environmental principal components. High BF_{ij} indicates high correlation with a variable after controlling for neutral covariance in the data. Genes identified as significant for a variable by permutation are displayed in red. Background shading indicates the position of the autosomal chromosomes in order.

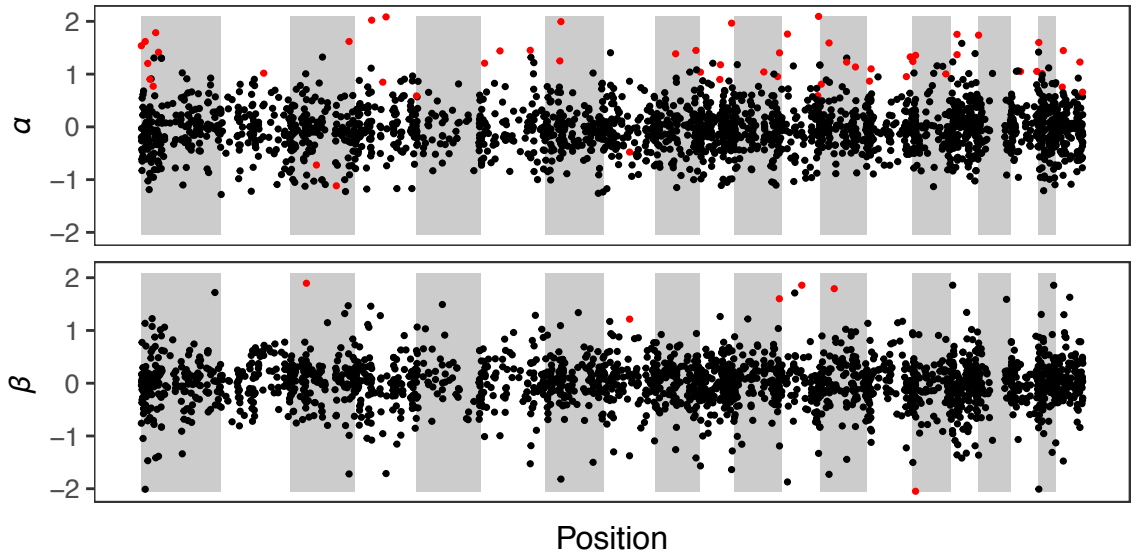

Figure S16: Distribution of α_i and β_i estimates for protein-coding genes. Genes identified as extreme for a parameter by 95% Bayesian credible intervals are displayed in red. Background shading indicates the position of the autosomal chromosomes in order.

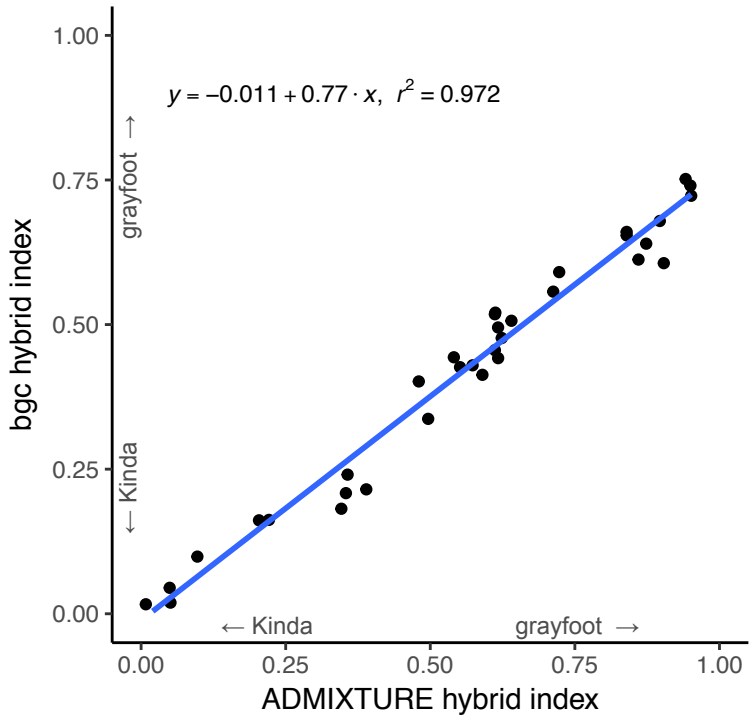

Figure S17: Concordance between hybrid indices estimated using ADMIXTURE and bgc.

Appendix B

Supplementary Tables

Table S1: Animals sequenced for this study.

Individual	Locale	Taxon	Sex	Origin	Year
SNPRC #13245	SNPRC	*Papio anubis*	male	captivity	2014
SNPRC #14068	SNPRC	*Papio anubis*	male	captivity	2014
SNPRC #25567	SNPRC	*Papio anubis* × *Papio ursinus*	female	captivity	2014
SNPRC #27278	SNPRC	*Papio anubis* × *Papio cynocephalus*	male	captivity	2014
SNPRC #27958	SNPRC	*Papio anubis*	female	captivity	2014
SNPRC #28064	SNPRC	*Papio anubis*	female	captivity	2014
BZ06-051	South Luangwa NP	*Papio kindae* × *Papio cynocephalus*	unknown	wild	2006
BZ06-053	South Luangwa NP	*Papio kindae* × *Papio cynocephalus*	unknown	wild	2006
BZ06-066	South Luangwa NP	*Papio kindae* × *Papio cynocephalus*	unknown	wild	2006
BZ06-148	North Luangwa NP	*Papio kindae* × *Papio cynocephalus*	unknown	wild	2006
BZ06-218	Lower Zambezi NP	*Papio griseipes*	unknown	wild	2006
BZ06-220	Lower Zambezi NP	*Papio griseipes*	unknown	wild	2006
BZ06-221	Lower Zambezi NP	*Papio griseipes*	unknown	wild	2006
BZ06-224	Lower Zambezi NP	*Papio griseipes*	unknown	wild	2006
BZ06-225	Lower Zambezi NP	*Papio griseipes*	unknown	wild	2006
BZ06-227	Lower Zambezi NP	*Papio griseipes*	unknown	wild	2006
BZ07-001	Choma	*Papio griseipes*	unknown	wild	2007
BZ07-004	Choma	*Papio griseipes*	unknown	wild	2007
BZ07-005	Choma	*Papio griseipes*	unknown	wild	2007
BZ07-007	Choma	*Papio griseipes*	unknown	wild	2007
BZ07-029	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-030	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-032	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-034	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-035	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-039	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-041	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-042	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-045	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-047	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
BZ07-100	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2007
Chiou-14-001	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014

continued on next page

Table S1 – continued from previous page

Individual	Locale	Taxon	Sex	Origin	Year
Chiou-14-003	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-004	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-005	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-030	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-036	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-039	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-041	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-042	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-044	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-050	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-054	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-056	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-057	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-058	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-059	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-065	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-14-069	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2014
Chiou-15-003	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2015
Chiou-15-004	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2015
Chiou-15-005	Kafue NP	*Papio kindae* × *Papio griseipes*	unknown	wild	2015

Table S2: Fecal DNA enrichment results. Key: ID, capture experiment ID; Lib, library ID; Ind, individual (see Table S1); PHB, percent host DNA before; TD, total fecal DNA used (ng); BV, bead volume used (µl); TV, total reaction volume (µl); TY, total DNA yield (ng); NE, number of enrichment steps; PHA, percent host DNA after; NDB, *n*-fold decrease in bacterial DNA.

ID	Lib	Ind	PHB	TD	BV	TV	TY	NE	PHA	NDB
A01F.T002	A	SNPRC #14068	0.15%	1,000.00	160.00	163.00	82.83	single	6.79%	-
A02F.T005	A	SNPRC #25567	4.40%	2,000.00	160.00	174.80	455.40	single	50.90%	-
A03F.T007	A	SNPRC #27278	0.00%	1,000.00	160.00	165.10	191.40	single	6.38%	-
A04F.T009	A	SNPRC #27958	13.59%	2,000.00	160.00	172.40	584.10	single	47.50%	-
A05F.B051	A	BZ06-051	1.59%	1,800.00	160.00	285.00	122.43	single	-	-
A06F.B053	A	BZ06-053	1.14%	1,365.00	160.00	535.00	119.79	single	-	-
B01F.T001	B	SNPRC #13245	3.55%	1,000.00	160.00	170.90	72.80	single	19.45%	-
B02F.T002	B	SNPRC #14068	0.15%	1,000.00	160.00	164.40	16.08	single	8.34%	-
B03F.T007	B	SNPRC #27278	0.00%	1,000.00	160.00	165.40	26.80	single	12.68%	-
B04F.T010	B	SNPRC #28064	9.87%	1,000.00	160.00	177.50	78.80	single	125.38%	-
B05F.C030	B	Chiou-14-030	1.24%	1,000.00	160.00	285.00	38.16	single	14.31%	-
B06F.C050	B	Chiou-14-050	2.40%	1,000.00	160.00	198.10	13.28	single	10.81%	-
B07F.C065	B	Chiou-14-065	0.76%	1,000.00	160.00	285.00	22.96	single	25.44%	-
B08F.C069	B	Chiou-14-069	0.45%	1,000.00	160.00	211.20	15.92	single	17.88%	-
B09F.B066	B	BZ06-066	4.85%	1,000.00	160.00	285.00	39.20	single	38.72%	-
B10F.B148	B	BZ06-148	0.68%	1,000.00	160.00	292.10	9.92	single	9.07%	-
C01F.T001	C	SNPRC #13245	4.33%	1,000.00	6.92	40.00	20.80	single	65.00%	1.89
C02F.T002	C	SNPRC #14068	1.12%	1,000.00	1.80	40.00	6.80	single	6.41%	10.52
C03F.T005	C	SNPRC #25567	8.38%	1,000.00	13.40	40.00	29.48	single	170.96%	6.38
C04D.T002	C	SNPRC #14068	0.01%	8,000.00	1.00	40.00	5.24	serial	0.00%	24.46
C05F.T009	C	SNPRC #27958	17.40%	800.00	22.28	40.00	36.00	single	152.22%	13.29
C06F.T010	C	SNPRC #28064	4.40%	1,000.00	7.04	40.00	15.12	single	138.62%	19.02
C07F.C050	C	Chiou-14-050	2.99%	1,000.00	4.78	40.00	5.24	single	14.05%	7.92
C08D.C050	C	Chiou-14-050	0.82%	600.00	1.00	40.00	1.02	serial	0.00%	11.08
C09F.B051	C	BZ06-051	3.33%	500.00	2.66	40.00	2.92	single	49.45%	2.64
C10F.T005	C	SNPRC #25567	12.83%	1,000.00	1.50	40.00	14.08	single	221.02%	10.31
C11D.C069	C	Chiou-14-069	0.00%	1,000.00	1.00	40.00	0.76	serial	24.47%	6.81
C12D.B051	C	BZ06-051	0.36%	600.00	1.00	40.00	1.18	serial	14.76%	2.10
D01D.B220	D	BZ06-220	0.03%	1,000.00	1.00	40.00	0.70	serial	-	-
D02D.J001	D	BZ07-001	0.02%	1,000.00	1.00	40.00	2.39	serial	-	-
D03D.J007	D	BZ07-007	0.12%	947.20	1.00	40.00	0.95	serial	-	-
D04D.J029	D	BZ07-029	0.00%	1,000.00	1.00	40.00	2.25	serial	-	-
D05D.J032	D	BZ07-032	0.88%	1,000.00	1.00	40.00	0.97	serial	-	-
D06D.J034	D	BZ07-034	0.47%	1,000.00	1.00	40.00	2.23	serial	-	-
D07D.J039	D	BZ07-039	0.06%	1,000.00	1.00	40.00	1.05	serial	-	-
D08D.C057	D	Chiou-14-057	0.24%	1,000.00	1.00	40.00	0.90	serial	-	-
D09D.C003	D	Chiou-14-003	0.03%	1,000.00	1.00	40.00	0.88	serial	-	-
D10D.C044	D	Chiou-14-044	0.03%	913.96	1.00	40.00	0.93	serial	-	-
D11D.C041	D	Chiou-14-041	0.03%	1,000.00	1.00	40.00	1.31	serial	-	-
D12D.C042	D	Chiou-14-042	0.53%	1,000.00	1.00	40.00	0.78	serial	-	-
D13D.B221	D	BZ06-221	0.09%	548.96	1.00	40.00	0.66	serial	-	-
D14D.B227	D	BZ06-227	0.03%	1,000.00	1.00	40.00	0.52	serial	-	-
D15D.J004	D	BZ07-004	0.39%	1,000.00	1.00	40.00	0.69	serial	-	-
D16D.J005	D	BZ07-005	0.21%	1,000.00	1.00	40.00	0.46	serial	-	-
D17D.J030	D	BZ07-030	0.27%	988.80	1.00	40.00	0.68	serial	-	-
D18D.J035	D	BZ07-035	0.44%	1,000.00	1.00	40.00	0.70	serial	-	-
D19D.J041	D	BZ07-041	0.44%	1,000.00	1.00	40.00	0.53	serial	-	-
D20D.C001	D	Chiou-14-001	0.01%	1,000.00	1.00	40.00	0.64	serial	-	-

continued on next page

Table S2 – continued from previous page

ID	Lib	Ind	PHB	TD	BV	TV	TY	NE	PHA	NDB
D21D.C004	D	Chiou-14-004	0.01%	1,000.00	1.00	40.00	0.53	serial	-	-
D22D.C065	D	Chiou-14-065	0.08%	1,000.00	1.00	40.00	0.48	serial	-	-
D23D.C030	D	Chiou-14-030	0.30%	629.26	1.00	40.00	0.63	serial	-	-
D24D.C039	D	Chiou-14-039	0.29%	1,000.00	1.00	40.00	0.57	serial	-	-
D25D.B218	D	BZ06-218	0.06%	893.52	1.00	40.00	< 0.40	serial	-	-
D26D.B224	D	BZ06-224	0.03%	795.70	1.00	40.00	< 0.40	serial	-	-
D27D.B225	D	BZ06-225	0.06%	1,000.00	1.00	40.00	< 0.40	serial	-	-
D28D.J042	D	BZ07-042	0.11%	1,000.00	1.00	40.00	< 0.40	serial	-	-
D29D.J045	D	BZ07-045	0.03%	985.60	1.00	40.00	< 0.40	serial	-	-
D30D.J047	D	BZ07-047	0.02%	1,000.00	1.00	40.00	< 0.40	serial	-	-
D31D.J100	D	BZ07-100	0.08%	1,000.00	1.00	40.00	0.42	serial	-	-
D32D.C036	D	Chiou-14-036	0.37%	154.76	1.00	40.00	0.42	serial	-	-
D33D.C054	D	Chiou-14-054	0.44%	1,000.00	1.00	40.00	< 0.40	serial	-	-
D34D.C056	D	Chiou-14-056	0.11%	1,000.00	1.00	40.00	< 0.40	serial	-	-
D35D.C058	D	Chiou-14-058	0.57%	627.80	1.00	40.00	< 0.40	serial	-	-
D36D.C059	D	Chiou-14-059	0.90%	383.98	1.00	40.00	< 0.40	serial	-	-
D37D.H003	D	Chiou-15-003	0.10%	220.46	1.00	40.00	< 0.40	serial	-	-
D38D.H004	D	Chiou-15-004	3.11%	186.88	1.00	40.00	< 0.40	serial	-	-
D39D.H005	D	Chiou-15-005	0.11%	511.00	1.00	40.00	< 0.40	serial	-	-
D40D.C005	D	Chiou-14-005	2.41%	182.50	1.00	40.00	< 0.40	serial	-	-

Table S3: Library preparation and sequence mapping results. Key: ID, experiment ID (see Table S2); T, tissue type; PHB, percent host DNA before; TD, total DNA used (ng); PID, pool ID; PC, total number of PCR amplification cycles; TR, total number of sequencing reads; RM, number of reads mapping to the baboon reference genome (papAnu2); PRM, percentage of reads mapping to the baboon reference genome (papAnu2).

ID	T	PHB	TD	PID	PC	TR	RM	PRM
A01F.T002	feces	0.1500%	83.00	A1	24	264,158	9,856	3.73%
A02F.T005	feces	4.4000%	200.00	A1	24	2,607,006	1,892,463	72.59%
A03F.T007	feces	0.0000%	191.00	A1	24	2,040,002	15,224	0.75%
A04F.T009	feces	13.5900%	200.00	A2	24	4,104,470	3,334,314	81.24%
A05F.B051	feces	1.5900%	122.00	A2	24	916,680	257,714	28.11%
A06F.B053	feces	1.1400%	120.00	A2	24	591,626	146,285	24.73%
A07B.T002	blood	-	200.00	A3	24	2,683,338	2,384,501	88.86%
A08B.T005	blood	-	200.00	A3	24	745,504	681,730	91.45%
A09B.T007	blood	-	200.00	A4	24	2,128,974	1,886,608	88.62%
A10B.T009	blood	-	200.00	A4	24	1,189,390	1,087,777	91.46%
B01F.T001	feces	3.5500%	60.06	B1	24	4,882,156	1,652,610	33.85%
B02F.T002	feces	0.1500%	13.27	B1	24	2,507,424	111,206	4.44%
B03F.T007	feces	0.0000%	22.11	B1	24	2,461,390	27,185	1.10%
B04F.T010	feces	9.8700%	65.01	B1	24	5,444,582	3,184,008	58.48%
B05F.C030	feces	1.2400%	33.39	B1	24	619,862	34,761	5.61%
B06F.C050	feces	2.4000%	11.62	B1	24	2,123,294	584,719	27.54%
B07F.C065	feces	0.7600%	20.09	B1	24	1,241,912	44,732	3.60%
B08F.C069	feces	0.4500%	13.93	B1	24	1,203,334	23,642	1.96%
B09F.B066	feces	4.8500%	34.30	B1	24	1,563,428	581,005	37.16%
B10F.B148	feces	0.6800%	8.68	B1	24	501,188	62,126	12.40%
B11B.T001	blood	-	196.50	B2	24	2,216,126	1,819,922	82.12%
B12B.T010	blood	-	201.50	B2	24	2,092,674	1,780,702	85.09%
C01F.T001	feces	4.3300%	17.68	C1	20	1,782,002	1,188,499	66.69%
C02F.T002	feces	1.1200%	5.78	C2	24	1,290,578	333,245	25.82%
C03F.T005	feces	8.3800%	25.06	C1	20	1,871,868	1,316,267	70.32%
C04D.T002	feces	0.0100%	3.41	C3	26	1,841,762	163,819	8.89%
C05F.T009	feces	17.4000%	30.60	C1	20	3,116,288	2,345,065	75.25%
C06F.T010	feces	4.4000%	12.85	C1	20	1,469,246	967,950	65.88%
C07F.C050	feces	2.9900%	4.45	C2	24	3,570,776	1,845,128	51.67%
C08D.C050	feces	0.8200%	0.67	C4	26	2,059,728	737,230	35.79%
C09F.B051	feces	3.3300%	2.48	C2	24	1,816,378	918,754	50.58%
C10F.T005	feces	12.8300%	11.97	C1	20	2,068,478	1,475,254	71.32%
C11D.C069	feces	0.0049%	0.49	C5	26	1,514,706	195,365	12.90%
C12D.B051	feces	0.3600%	0.77	C6	26	1,896,580	1,035,140	54.58%
D01D.B220	feces	0.0289%	0.53	D1	22	5,875,038	558,679	9.51%
D02D.J001	feces	0.0158%	1.79	D1	22	1,449,446	47,799	3.30%
D03D.J007	feces	0.1220%	0.71	D1	22	3,243,182	760,274	23.44%
D04D.J029	feces	0.0030%	1.69	D1	22	1,542,546	22,534	1.46%
D05D.J032	feces	0.8793%	0.73	D1	22	9,398,314	6,856,889	72.96%
D06D.J034	feces	0.4713%	1.67	D1	22	2,656,920	1,313,120	49.42%
D07D.J039	feces	0.0629%	0.79	D1	22	2,230,514	609,824	27.34%
D08D.C057	feces	0.2374%	0.67	D1	22	24,351,758	9,032,717	37.09%
D09D.C003	feces	0.0339%	0.66	D1	22	4,453,940	140,853	3.16%
D10D.C044	feces	0.0292%	0.70	D1	22	2,137,704	124,844	5.84%
D11D.C041	feces	0.0271%	0.98	D1	22	1,140,122	96,922	8.50%
D12D.C042	feces	0.5310%	0.59	D1	22	10,202,338	4,952,523	48.54%
D13D.B221	feces	0.0894%	0.49	D2	22	5,147,652	982,161	19.08%

continued on next page

Table S3 – continued from previous page

ID	T	PHB	TD	PID	PC	TR	RM	PRM
D14D.B227	feces	0.0309%	0.39	D2	22	1,828,496	188,056	10.28%
D15D.J004	feces	0.3858%	0.52	D2	22	3,606,424	1,619,089	44.89%
D16D.J005	feces	0.2120%	0.35	D2	22	4,681,458	1,595,305	34.08%
D17D.J030	feces	0.2748%	0.51	D2	22	9,260,376	4,298,954	46.42%
D18D.J035	feces	0.4370%	0.52	D2	22	14,641,030	7,429,804	50.75%
D19D.J041	feces	0.4441%	0.40	D2	22	2,734,646	1,830,636	66.94%
D20D.C001	feces	0.0079%	0.48	D2	22	3,011,748	95,667	3.18%
D21D.C004	feces	0.0088%	0.40	D2	22	2,685,536	97,831	3.64%
D22D.C065	feces	0.0770%	0.36	D2	22	6,516,948	954,911	14.65%
D23D.C030	feces	0.2962%	0.47	D2	22	4,200,672	1,854,549	44.15%
D24D.C039	feces	0.2900%	0.43	D2	22	15,175,272	5,263,567	34.69%
D25D.B218	feces	0.0581%	0.23	D3	26	2,624,536	311,121	11.85%
D26D.B224	feces	0.0307%	0.23	D3	26	3,581,376	246,941	6.90%
D27D.B225	feces	0.0573%	0.23	D3	26	16,512,734	1,228,293	7.44%
D28D.J042	feces	0.1062%	0.23	D3	26	6,967,098	1,135,899	16.30%
D29D.J045	feces	0.0262%	0.23	D3	26	2,307,634	124,309	5.39%
D30D.J047	feces	0.0213%	0.23	D3	26	5,640,310	519,703	9.21%
D31D.J100	feces	0.0793%	0.31	D3	26	1,219,418	83,022	6.81%
D32D.C036	feces	0.3708%	0.32	D3	26	7,119,672	4,402,290	61.83%
D33D.C054	feces	0.4415%	0.23	D4	22	7,013,280	4,484,858	63.95%
D34D.C056	feces	0.1098%	0.23	D4	22	7,949,918	2,748,447	34.57%
D35D.C058	feces	0.5733%	0.23	D4	22	10,539,604	4,223,730	40.07%
D36D.C059	feces	0.8954%	0.23	D4	22	5,962,728	3,090,413	51.83%
D37D.H003	feces	0.1017%	0.23	D4	22	1,418,168	74,131	5.23%
D38D.H004	feces	3.1094%	0.23	D4	22	9,161,532	6,739,523	73.56%
D39D.H005	feces	0.1113%	0.23	D4	22	1,784,298	585,250	32.80%
D40D.C005	feces	2.4120%	0.23	D4	22	2,621,020	1,611,725	61.49%
E01B.T001	blood	-	200.00	E1	12	2,326,792	2,027,296	87.13%
E02B.T002	blood	-	200.00	E1	12	1,249,950	1,095,499	87.64%
E03B.T005	blood	-	200.00	E2	12	4,812,938	4,192,408	87.11%
E04B.T009	blood	-	200.00	E1	12	1,986,292	1,746,868	87.95%
E05B.T010	blood	-	200.00	E2	12	4,091,500	3,597,699	87.93%

Table S4: DNA samples used for controlled experiments. Artificial "fecal" DNA was prepared by manually mixing DNA samples in controlled proportions. Artificial methylated DNA was also prepared using amplicons of lambda phage DNA (with known sequence) and methyltransferase enzymes with specific recognition sites. 5,012 bp amplicons were prepared using the primers `/5Biosg/GTTCTGCACTGACAGATTAAAACTCG` and `CTGCTCATTAATATACTTCTGGGTTCC`, 15,089 bp amplicons were prepared using the primers `/5Biosg/GAGTGAATATATCGAACAGTCAGG` and `GTGTCATATTTCACTTCCGTACC`, and 10,144 bp amplicons were prepared using the primers `/5BiosG/ATAAAGATGAGACGCTGGAGTACA` and `GCGATAACCAGGTAAAATTTTCCG`. Key: ID, prepared DNA sample ID; DNA1, input DNA sample 1; DNA2, input DNA sample 2; PH, percentage of "host" (baboon) DNA; PB, percentage of bacterial DNA; L, length of DNA amplicon; Enz, methyltransferase enzyme(s) used; MD, CpG methylation density.

ID	DNA1	DNA2	PH	PB	L	Enz	MD
PB01	K12 *E. coli*	none	0.0%	100.00%	-	-	-
PB02	ATCC 11303 *E. coli*	none	0.0%	100.00%	-	-	-
PH01	Baboon blood	none	100.0%	0.0%	-	-	-
PH02	Baboon liver	none	100.0%	0.0%	-	-	-
AF01	Baboon blood	K12 *E. coli*	2.0%	98.0%	-	-	-
AF02	Baboon blood	K12 *E. coli*	0.2%	99.8%	-	-	-
AF03	Baboon blood	K12 *E. coli*	50.0%	50.0%	-	-	-
AF04	Baboon blood	K12 *E. coli*	2.0%	98.0%	-	-	-
AF05	Baboon blood	K12 *E. coli*	50.0%	50.0%	-	-	-
AF06	Baboon blood	K12 *E. coli*	5.0%	95.0%	-	-	-
AF07	Baboon blood	K12 *E. coli*	10.0%	90.0%	-	-	-
AF08	Baboon blood	ATCC 11303 *E. coli*	2.0%	98.0%	-	-	-
AF09	Baboon liver	ATCC 11303 *E. coli*	2.0%	98.0%	-	-	-
AF10	Baboon liver	ATCC 11303 *E. coli*	50.0%	50.0%	-	-	-
AF11	Baboon liver	ATCC 11303 *E. coli*	0.5%	99.5%	-	-	-
AF12	Baboon liver	ATCC 11303 *E. coli*	2.0%	98.0%	-	-	-
AF13	Baboon liver	ATCC 11303 *E. coli*	5.0%	95.0%	-	-	-
AF14	Baboon liver	ATCC 11303 *E. coli*	2.0%	98.0%	-	-	-
AF15	Baboon liver	ATCC 11303 *E. coli*	0.5%	99.5%	-	-	-
AF16	Baboon liver	ATCC 11303 *E. coli*	2.0%	98.0%	-	-	-
AF17	Baboon liver	ATCC 11303 *E. coli*	5.0%	95.0%	-	-	-
AF18	Baboon liver	ATCC 11303 *E. coli*	0.5%	99.5%	-	-	-
CD01	Lambda cl857 phage	none	0.0%	0.0%	5,012	*Hha*I	3.6
CD02	Lambda cl857 phage	none	0.0%	0.0%	5,012	-	0.0
CD03	Lambda cl857 phage	none	0.0%	0.0%	5,012	*Hha*I	3.6
CD04	Lambda cl857 phage	none	0.0%	0.0%	5,012	*Hha*I + *Hpa*II	7.2
CD05	Lambda cl857 phage	none	0.0%	0.0%	5,012	*Hha*I	3.6
CD06	Lambda cl857 phage	none	0.0%	0.0%	15,089	*Hha*I	6.9
CD07	Lambda cl857 phage	none	0.0%	0.0%	10,144	*Hha*I	6.3
CD08	Lambda cl857 phage	none	0.0%	0.0%	10,144	*Hha*I + *Hpa*II	17.7

Table S5: Controlled DNA enrichment experiments. DNA enrichment was simulated from artificial "fecal" samples. In some cases, additional DNA was included. A number of variables described in Supplementary Protocol were tuned to evaluate their impact on enrichment results (see Table S6). Key: ID, experiment ID; SID, experiment set ID; DNA1, input DNA sample 1 (see Table S4 or this table); TD1, total amount of sample 1 (ng); PH, percentage of "host" (baboon) DNA in sample 1; DNA2, input DNA sample 2 (see Table S4); TD2, total amount of sample 2 (ng); BV, volume of protein A beads used (µl); PV, volume of MBD-Fc protein used (µl); NC, NaCl concentration of reaction (µM); TV, total volume of reaction (µl); NW, number of washes; NCW, NaCl concentration of each wash (µM); WV, volume of each wash (µl); EM, Elution method. For elutions in TE, proteinase K was added at a ratio of 1 µl proteinase K to 10 µl 1X TE.

ID	SID	DNA1	TD1	PH	DNA2	TD2	BV	PV	NC	TV	NW	NCW	WV	EM
X001	S01	AF01	1,000.00	2.0%	-	0.00	160.0	16.00	150	166.20	0	-	-	150 µl TE
X002	S01	AF01	2,000.00	2.0%	-	0.00	160.0	16.00	150	172.40	0	-	-	150 µl TE
X003	S01	AF02	1,000.00	0.2%	-	0.00	160.0	16.00	150	164.00	0	-	-	150 µl TE
X004	S01	AF01	1,000.00	2.0%	-	0.00	320.0	32.00	150	326.20	0	-	-	150 µl TE
X005	S01	AF01	1,000.00	2.0%	-	0.00	80.0	8.00	150	86.20	0	-	-	150 µl TE
X006	S01	AF01	1,000.00	2.0%	-	0.00	40.0	4.00	150	46.20	0	-	-	150 µl TE
X007	S02	PB01	1,000.00	0.0%	-	0.00	160.0	16.00	150	163.60	0	-	-	150 µl TE
X008	S02	AF01	1,000.00	2.0%	-	0.00	16.0	1.60	150	22.20	0	-	-	150 µl TE
X009	S02	AF03	40.00	50.0%	-	0.00	80.0	8.00	150	130.00	0	-	-	150 µl TE
X010	S02	AF03	40.00	50.0%	-	0.00	40.0	4.00	150	90.00	0	-	-	150 µl TE
X011	S02	AF03	40.00	50.0%	-	0.00	16.0	1.60	150	66.00	0	-	-	150 µl TE
X012	S02	AF03	40.00	50.0%	-	0.00	8.0	0.80	150	58.00	0	-	-	150 µl TE
X013	S03	PB01	1,000.00	0.0%	-	0.00	160.0	0.00	150	163.60	0	-	-	150 µl TE
X014	S03	PB01	1,000.00	0.0%	-	0.00	40.0	0.00	150	47.20	0	-	-	150 µl TE
X015	S03	PB01	200.00	0.0%	-	0.00	40.0	0.00	150	47.20	0	-	-	150 µl TE
X016	S03	PB01	1,000.00	0.0%	-	0.00	40.0	32.00	150	47.20	0	-	-	150 µl TE
X017	S04	AF04	1,000.00	2.0%	-	0.00	1.0	0.10	150	7.20	0	-	-	150 µl TE
X018	S04	PB01	1,000.00	0.0%	-	0.00	40.0	4.00	150	43.60	0	-	-	150 µl TE
X019	S04	PB01	1,000.00	0.0%	-	0.00	40.0	4.00	300	58.20	0	-	-	150 µl TE
X020	S04	AF05	1,000.00	50.0%	-	0.00	40.0	4.00	150	105.80	0	-	-	150 µl TE
X021	S05	AF04	1,000.00	2.0%	-	0.00	40.0	4.00	150	46.20	0	-	-	150 µl TE
X022	S05	AF04	1,000.00	2.0%	-	0.00	40.0	4.00	150	46.20	0	-	-	150 µl TE
X023	S05	X001	29.16	51.0%	-	0.00	40.0	4.00	150	78.60	0	-	-	150 µl TE
X024	S05	X006	21.48	44.1%	-	0.00	40.0	4.00	150	78.40	0	-	-	150 µl TE
X025	S06	AF04	1,000.00	2.0%	CD01	500.00	40.0	4.00	150	48.90	0	-	-	150 µl TE
X026	S06	AF06	1,000.00	5.0%	-	0.00	40.0	4.00	150	49.90	0	-	-	150 µl TE
X027	S06	AF07	1,000.00	10.0%	-	0.00	40.0	4.00	150	56.10	0	-	-	150 µl TE
X028	S06	AF04	1,000.00	2.0%	CD02	500.00	40.0	4.00	150	48.70	0	-	-	150 µl TE
X029	S07	AF04	1,000.00	2.0%	CD01	2,500.00	40.0	4.00	150	58.70	0	-	-	150 µl TE
X030	S07	AF04	1,000.00	2.0%	-	0.00	1.0	0.10	150	7.20	0	-	-	150 µl TE
X031	S07	AF04	1,000.00	2.0%	-	0.00	1.0	0.10	150	100.00	0	-	-	150 µl TE
X032	S07	AF04	1,000.00	2.0%	CD01	500.00	1.0	0.10	150	9.70	0	-	-	150 µl TE
X033	S08	AF08	1,000.00	2.0%	-	0.00	8.0	0.80	150	13.00	0	-	-	150 µl TE
X034	S08	AF08	1,000.00	2.0%	-	0.00	4.0	0.40	150	9.00	0	-	-	150 µl TE
X035	S08	AF08	1,000.00	2.0%	-	0.00	2.0	0.20	150	7.00	0	-	-	150 µl TE
X036	S08	AF08	1,000.00	2.0%	-	0.00	0.5	0.05	150	6.00	0	-	-	150 µl TE
X037	S08	PB02	980.00	0.0%	-	0.00	1.0	0.10	150	7.20	0	-	-	150 µl TE
X038	S08	PH01	20.00	100.0%	-	0.00	1.0	0.10	150	7.20	0	-	-	150 µl TE
X039	S09	AF08	1,000.00	2.0%	CD03	1,000.00	40.0	4.00	150	56.30	0	-	-	150 µl TE
X040	S09	AF08	1,000.00	2.0%	CD03	500.00	40.0	4.00	150	50.60	0	-	-	150 µl TE
X041	S10	AF09	1,000.00	2.0%	CD04	500.00	40.0	4.00	150	80.00	0	-	-	150 µl TE
X042	S10	AF09	1,000.00	2.0%	CD04	500.00	16.0	1.60	150	32.00	0	-	-	150 µl TE
X043	S11	PH02	1,000.00	100.0%	-	0.00	1.0	0.10	150	12.13	0	-	-	150 µl TE

continued on next page

Table S5 – continued from previous page

ID	EID	DNA1	TD1	PH	DNA2	TD2	BV	PV	NC	TV	NW	NCW	WV	EM
X044	S11	PH02	2,000.00	100.0%	-	0.00	1.0	0.10	150	23.36	0	-	-	150 µl TE
X045	S11	PH02	1,000.00	100.0%	-	0.00	2.0	0.20	150	13.13	0	-	-	150 µl TE
X046	S11	PH02	1,000.00	100.0%	-	0.00	4.0	0.40	150	15.13	0	-	-	150 µl TE
X047	S11	PH02	1,000.00	100.0%	-	0.00	8.0	0.80	150	19.13	0	-	-	150 µl TE
X048	S11	PH02	1,000.00	100.0%	-	0.00	16.0	1.60	150	27.13	0	-	-	150 µl TE
X049	S12	PH02	112.00	100.0%	-	0.00	1.0	0.10	150	2.25	0	-	-	150 µl TE
X050	S12	PH02	112.00	100.0%	-	0.00	1.0	0.10	150	2.25	0	-	-	40 µl TE
X051	S12	PH02	112.00	100.0%	-	0.00	1.0	0.10	150	2.25	0	-	-	150 µl TE
X052	S12	PH02	112.00	100.0%	-	0.00	1.0	0.10	150	2.25	0	-	-	60 µl TE
X053	S13	PB02	1,000.00	0.0%	CD05	1,000.00	40.0	4.00	150	80.00	0	-	-	150 µl TE
X054	S13	AF10	2,000.00	50.0%	-	0.00	40.0	4.00	150	80.00	0	-	-	150 µl TE
X055	S14	PB02	1,000.00	0.0%	CD06	1,000.00	40.0	4.00	150	80.00	0	-	-	150 µl TE
X056	S14	AF11	1,000.00	0.5%	CD06	1,000.00	40.0	4.00	150	80.00	0	-	-	150 µl TE
X057	S14	AF12	1,000.00	2.0%	CD06	1,000.00	40.0	4.00	150	80.00	0	-	-	150 µl TE
X058	S14	AF13	1,000.00	5.0%	CD06	1,000.00	40.0	4.00	150	80.00	0	-	-	150 µl TE
X059	S15	PB02	1,000.00	0.0%	-	0.00	40.0	4.00	150	80.00	1	150	80	see Table S7
X060	S15	PH02	250.00	100.0%	-	0.00	40.0	4.00	150	80.00	1	150	80	see Table S7
X061	S16	AF14	1,000.00	2.0%	-	0.00	40.0	4.00	150	80.00	0	-	-	2 M NaCl
X062	S16	AF14	1,000.00	2.0%	-	0.00	40.0	1.00	150	80.00	0	-	-	2 M NaCl
X063	S16	AF14	1,000.00	2.0%	CD07	1,000.00	40.0	4.00	150	80.00	0	-	-	2 M NaCl
X064	S16	AF14	1,000.00	2.0%	CD07	1,000.00	40.0	4.00	150	80.00	0	-	-	2 M NaCl
X065	S16	AF14	1,000.00	2.0%	-	0.00	40.0	4.00	200	80.00	0	-	-	2 M NaCl
X066	S16	AF14	1,000.00	2.0%	-	0.00	40.0	4.00	150	80.00	1	200	200	2 M NaCl
X067	S17	AF14	1,000.00	2.0%	-	0.00	40.0	4.00	150	80.00	0	-	-	150 µl TE
X068	S17	AF14	1,000.00	2.0%	-	0.00	40.0	4.00	150	80.00	0	-	-	2 M NaCl
X069	S17	AF14	1,000.00	2.0%	-	0.00	40.0	0.25	150	80.00	0	-	-	2 M NaCl
X070	S17	AF14	1,000.00	2.0%	-	0.00	40.0	0.25	150	80.00	1	150	100	2 M NaCl
X071	S17	AF14	1,000.00	2.0%	-	0.00	40.0	0.25	150	80.00	1	200	100	2 M NaCl
X072	S17	AF14	1,000.00	2.0%	CD07	1,000.00	40.0	0.25	150	80.00	1	200	100	2 M NaCl
X073	S18	AF14	1,000.00	2.0%	-	0.00	40.0	1.00	150	80.00	1	150	100	2 M NaCl
X074	S18	AF14	1,000.00	2.0%	-	0.00	40.0	1.00	150	80.00	1	200	100	2 M NaCl
X075	S18	AF14	2,000.00	2.0%	-	0.00	40.0	1.00	150	80.00	1	150	100	2 M NaCl
X076	S19	AF15	1,000.00	0.5%	-	0.00	40.0	1.00	150	80.00	1	150	100	2 M NaCl
X077	S19	AF15	1,000.00	0.5%	-	0.00	40.0	1.00	150	80.00	2	150	100	2 M NaCl
X078	S19	AF15	1,000.00	0.5%	-	0.00	40.0	1.00	150	80.00	3	150	100	2 M NaCl
X079	S20	AF15	1,000.00	0.5%	-	0.00	40.0	1.00	150	80.00	1	150	80	2 M NaCl
X080	S20	AF15	1,000.00	0.5%	-	0.00	40.0	1.00	150	80.00	1	150	40	2 M NaCl
X081	S20	AF15	1,000.00	0.5%	CD07	40.00	40.0	1.00	150	80.00	1	150	80	2 M NaCl
X082	S20	PB02	1,000.00	0.0%	CD07	40.00	40.0	1.00	150	80.00	1	150	80	2 M NaCl
X083	S21	AF14	1,000.00	2.0%	-	0.00	40.0	1.00	150	80.00	1	150	80	2 M NaCl
X084	S21	AF14	1,000.00	2.0%	-	0.00	40.0	1.00	150	80.00	1	150	200	2 M NaCl
X085	S21	AF14	1,000.00	2.0%	-	0.00	40.0	4.00	150	80.00	1	150	80	2 M NaCl
X086	S21	AF15	1,000.00	0.5%	-	0.00	40.0	4.00	150	80.00	1	150	80	2 M NaCl
X087	S21	AF15	1,000.00	0.5%	CD08	20.00	40.0	4.00	150	80.00	1	150	80	2 M NaCl
X088	S21	PB02	1,000.00	0.0%	CD08	20.00	40.0	4.00	150	80.00	1	150	80	2 M NaCl
X089	S22	AF15	1,000.00	0.5%	-	0.00	40.0	1.00	150	80.00	1	150	80	2 M NaCl
X090	S22	AF15	1,000.00	0.5%	CD08	20.00	40.0	1.00	150	80.00	1	150	80	2 M NaCl
X091	S22	AF14	1,000.00	2.0%	CD08	20.00	40.0	1.00	150	80.00	1	150	80	2 M NaCl
X092	S22	PB02	1,000.00	0.0%	CD08	20.00	40.0	1.00	150	80.00	1	150	80	2 M NaCl
X093	S23	AF15	1,000.00	0.5%	-	0.00	40.0	0.25	150	80.00	1	150	100	2 M NaCl
X094	S23	AF15	1,000.00	0.5%	-	0.00	40.0	0.25	150	80.00	2	150	100	2 M NaCl
X095	S23	AF15	1,000.00	0.5%	-	0.00	40.0	0.25	150	80.00	3	150	100	2 M NaCl

continued on next page

Table S5 – continued from previous page

ID	EID	DNA1	TD1	PH	DNA2	TD2	BV	PV	NC	TV	NW	NCW	WV	EM
X096	S23	AF15	1,000.00	0.5%	-	0.00	40.0	0.25	150	80.00	4	150	100	2 M NaCl
X097	S24	AF15	1,000.00	0.5%	-	0.00	40.0	0.25	150	80.00	1	150	100	2 M NaCl
X098	S24	AF15	1,000.00	0.5%	-	0.00	2.5	0.25	150	80.00	1	150	100	2 M NaCl
X099	S24	AF16	1,000.00	2.0%	-	0.00	40.0	0.25	150	80.00	1	150	100	2 M NaCl
X100	S24	AF17	1,000.00	5.0%	-	0.00	40.0	0.25	150	80.00	1	150	100	2 M NaCl
X101	S25	AF15	1,000.00	0.5%	-	0.00	2.5	0.25	150	40.00	1	150	100	2 M NaCl
X102	S25	AF16	1,000.00	2.0%	-	0.00	10.0	1.00	150	40.00	1	150	100	2 M NaCl
X103	S25	AF17	1,000.00	5.0%	-	0.00	25.0	2.50	150	40.00	1	150	100	2 M NaCl
X104	S25	AF15	1,000.00	0.5%	-	0.00	2.5	0.25	150	40.00	1	150	40	2 M NaCl
X105	S25	AF16	1,000.00	2.0%	-	0.00	3.2	0.32	150	40.00	1	150	100	2 M NaCl
X106	S25	AF15	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	150	100	2 M NaCl
X107	S26	AF15	1,000.00	0.5%	-	0.00	2.5	0.25	150	40.00	1	150	100	2 M NaCl
X108	S26	AF15	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	150	100	2 M NaCl
X109	S26	AF15	1,000.00	0.5%	-	0.00	2.5	0.25	150	40.00	1	150	100	2 M NaCl
X110	S27	AF17	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	150	100	2 M NaCl
X111	S27	AF17	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	200	100	2 M NaCl
X112	S27	AF17	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	350	100	2 M NaCl
X113	S27	AF17	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	450	100	2 M NaCl
X114	S27	AF17	4,000.00	0.5%	-	0.00	3.2	0.32	150	40.00	1	150	100	2 M NaCl
X115	S27	AF16	1,000.00	2.0%	-	0.00	3.2	0.32	150	40.00	1	150	100	2 M NaCl
X116	S28	AF18	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	150	100	2 M NaCl
X117	S28	AF18	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	200	100	2 M NaCl
X118	S28	AF18	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	350	100	2 M NaCl
X119	S28	AF18	1,000.00	0.5%	-	0.00	0.8	0.08	150	40.00	1	450	100	2 M NaCl
X120	S28	X107	2.40	27.0%	-	0.00	1.3	0.13	150	40.00	1	150	100	2 M NaCl

Table S6: Controlled DNA enrichment experiment results. Percentages of host and bacterial DNA before and after enrichment experiments listed in Table S5 were estimated by qPCR using host- and bacteria-specific primers (see Supplementary Protocol). Key: ID, experiment ID (see Table S5); PHB, percentage of host DNA before; PHA, percentage of host DNA after; PBA, percentage of bacterial DNA after; TY, total DNA yield (ng); HY, estimated host DNA yield (ng); BY, estimated bacterial DNA yield (ng).

ID	PHB	PHA	PBA	TY	HY	BY
X001	2.0%	50.98%	56.74%	38.88	19.82	22.06
X002	2.0%	42.23%	49.32%	73.60	31.08	36.30
X003	0.2%	5.10%	101.77%	20.32	1.04	20.68
X004	2.0%	45.10%	62.70%	39.20	17.68	24.58
X005	2.0%	60.54%	46.75%	32.64	19.76	15.26
X006	2.0%	44.13%	45.74%	28.64	12.64	13.10
X007	0.0%	2.72%	110.93%	38.24	1.04	42.42
X008	2.0%	51.49%	51.28%	28.20	14.52	14.46
X009	50.0%	120.19%	6.43%	17.24	20.72	1.11
X010	50.0%	132.00%	3.18%	19.00	25.08	0.60
X011	50.0%	154.86%	1.86%	11.52	17.84	0.21
X012	50.0%	140.93%	1.76%	14.12	19.90	0.25
X013	0.0%	-	-	4.28	-	-
X014	0.0%	-	-	5.28	-	-
X015	0.0%	-	-	2.84	-	-
X016	0.0%	-	-	44.80	-	-
X017	2.0%	60.02%	3.18%	16.36	9.82	0.52
X018	0.0%	-	-	18.56	-	-
X019	0.0%	-	-	9.68	-	-
X020	50.0%	86.25%	0.48%	464.00	400.20	2.23
X021	2.0%	58.69%	29.84%	24.40	14.32	7.28
X022	2.0%	41.37%	56.03%	23.88	9.88	13.38
X023	51.0%	29.76%	15.36%	0.50	0.15	0.08
X024	44.1%	2.19%	8.89%	0.56	0.01	0.05
X025	2.0%	11.68%	10.47%	114.00	13.32	11.94
X026	5.0%	70.00%	18.42%	54.40	38.08	10.02
X027	10.0%	98.29%	9.19%	93.60	92.00	8.60
X028	2.0%	46.22%	37.09%	35.48	16.40	13.16
X029	2.0%	3.90%	3.26%	346.80	13.53	11.31
X030	2.0%	47.61%	4.07%	24.28	11.56	0.99
X031	2.0%	40.93%	11.19%	23.60	9.66	2.64
X032	2.0%	47.85%	1.96%	25.12	12.02	0.49
X033	2.0%	45.76%	41.10%	24.04	11.00	9.88
X034	2.0%	39.32%	28.48%	18.12	7.12	5.16
X035	2.0%	74.85%	13.87%	15.04	11.26	2.09
X036	2.0%	60.00%	16.84%	12.80	7.68	2.16
X037	0.0%	0.00%	73.67%	2.56	0.00	1.89
X038	100.0%	115.83%	0.01%	12.76	14.78	0.00
X039	2.0%	7.90%	6.71%	158.80	12.54	10.66
X040	2.0%	10.89%	10.28%	98.80	10.76	10.16
X041	2.0%	7.67%	9.32%	232.80	17.86	21.70
X042	2.0%	35.24%	38.99%	41.60	14.66	16.22
X043	100.0%	-	-	38.40	-	-
X044	100.0%	-	-	42.40	-	-
X045	100.0%	-	-	78.80	-	-
X046	100.0%	-	-	141.60	-	-
X047	100.0%	-	-	282.40	-	-

continued on next page

Table S6 – continued from previous page

ID	PHB	PHA	PBA	TY	HY	BY
X048	100.0%	-	-	456.00	-	-
X049	100.0%	-	-	17.52	-	-
X050	100.0%	-	-	14.64	-	-
X051	100.0%	-	-	1.89	-	-
X052	100.0%	-	-	3.19	-	-
X053	0.0%	-	2.59%	347.20	-	9.00
X054	50.0%	-	2.19%	496.00	-	10.86
X055	0.0%	-	10.08%	206.40	-	20.80
X056	0.5%	0.88%	6.01%	220.40	1.95	13.24
X057	2.0%	5.32%	3.21%	204.40	10.88	6.56
X058	5.0%	12.17%	2.88%	222.40	27.06	6.40
X059	0.0%	-	-	-	-	-
X060	100.0%	-	-	-	-	-
X061	2.0%	60.57%	19.32%	18.36	11.12	3.55
X062	2.0%	83.19%	12.39%	14.40	11.98	1.78
X063	2.0%	2.10%	0.25%	680.00	14.28	1.70
X064	2.0%	1.97%	0.30%	656.00	12.92	1.97
X065	2.0%	58.82%	18.97%	16.32	9.60	3.10
X066	2.0%	101.43%	5.27%	9.76	9.90	0.51
X067	2.0%	55.65%	16.44%	16.28	9.06	2.68
X068	2.0%	65.89%	21.69%	17.24	11.36	3.74
X069	2.0%	78.68%	17.98%	12.76	10.04	2.29
X070	2.0%	128.57%	3.58%	6.44	8.28	0.23
X071	2.0%	114.15%	2.56%	6.36	7.26	0.16
X072	2.0%	28.33%	0.82%	26.12	7.40	0.21
X073	2.0%	144.68%	4.86%	8.64	12.50	0.42
X074	2.0%	133.61%	4.28%	7.20	9.62	0.31
X075	2.0%	137.91%	6.36%	15.88	21.90	1.01
X076	0.5%	237.32%	30.14%	2.84	6.74	0.86
X077	0.5%	357.50%	24.50%	2.40	8.58	0.59
X078	0.5%	270.16%	17.74%	2.48	6.70	0.44
X079	0.5%	234.81%	33.86%	3.16	7.42	1.07
X080	0.5%	257.07%	44.62%	3.68	9.46	1.64
X081	0.5%	20.03%	2.15%	29.16	5.84	0.63
X082	0.0%	0.00%	3.59%	26.76	0.00	0.96
X083	2.0%	240.00%	8.42%	9.00	21.60	0.76
X084	2.0%	243.69%	10.49%	8.24	20.08	0.86
X085	2.0%	233.55%	20.94%	9.36	21.86	1.96
X086	0.5%	227.66%	66.81%	3.76	8.56	2.51
X087	0.5%	43.31%	12.76%	16.44	7.12	2.10
X088	0.0%	0.00%	14.38%	14.28	0.00	2.05
X089	0.5%	224.26%	45.29%	2.72	6.10	1.23
X090	0.5%	46.20%	8.26%	16.32	7.54	1.35
X091	2.0%	68.14%	3.79%	19.40	13.22	0.74
X092	0.0%	0.00%	7.89%	15.56	0.00	1.23
X093	0.5%	86.44%	7.81%	2.64	2.28	0.21
X094	0.5%	73.31%	4.84%	2.36	1.73	0.11
X095	0.5%	106.31%	3.65%	1.74	1.85	0.06
X096	0.5%	114.75%	3.60%	1.42	1.63	0.05
X097	0.5%	127.97%	40.42%	2.36	3.02	0.95
X098	0.5%	130.90%	28.13%	2.88	3.77	0.81
X099	2.0%	106.34%	8.35%	13.56	14.42	1.13

continued on next page

Table S6 – continued from previous page

ID	PHB	PHA	PBA	TY	HY	BY
X100	5.0%	135.67%	5.32%	17.44	23.66	0.93
X101	0.5%	49.11%	40.65%	2.48	1.22	1.01
X102	2.0%	83.82%	8.76%	13.60	11.40	1.19
X103	5.0%	96.18%	13.18%	20.40	19.62	2.69
X104	0.5%	91.02%	50.19%	2.16	1.97	1.08
X105	2.0%	74.20%	6.46%	12.56	9.32	0.81
X106	0.5%	-	-	-	-	-
X107	0.5%	139.38%	72.94%	3.20	4.46	2.33
X108	0.5%	101.79%	46.19%	2.68	2.73	1.24
X109	0.5%	124.12%	45.74%	2.72	3.38	1.24
X110	0.5%	98.30%	32.05%	2.24	2.20	0.72
X111	0.5%	138.57%	21.16%	1.96	2.72	0.41
X112	0.5%	128.51%	40.00%	1.48	1.90	0.59
X113	0.5%	137.04%	83.33%	1.08	1.48	0.90
X114	0.5%	113.78%	21.12%	7.84	8.92	1.66
X115	2.0%	141.15%	2.73%	10.40	14.68	0.28
X116	0.5%	60.37%	53.60%	2.72	1.64	1.46
X117	0.5%	92.22%	45.12%	2.42	2.23	1.09
X118	0.5%	68.13%	30.86%	1.78	1.21	0.55
X119	0.5%	77.34%	66.41%	2.05	1.59	1.36
X120	27.0%	170.44%	3.55%	0.55	0.94	0.02

Table S7: Controlled DNA enrichment elution series. After hybridizing DNA to MBD-bound beads, bound DNA was eluted in a series with progressively higher NaCl concentrations. The quantity of DNA in each elution was then quantified by Qubit. Key: ID, experiment ID (see Table S5); EN, elution number (elution 0 represents a wash); NC, NaCl concentration of reaction (μM); EV, Elution volume (μl); EY, elution DNA yield (ng); CY, cumulative DNA yield including previous elutions in the series (ng).

ID	E	NC	EV	EY	CY
X059	0	150	80	864.00	864.00
X059	1	200	80	27.04	891.04
X059	2	350	80	2.93	893.97
X059	3	450	80	1.17	895.14
X059	4	600	80	0.00	895.14
X059	5	1000	80	0.00	895.14
X059	6	2000	80	0.00	895.14
X060	0	150	80	62.56	62.56
X060	1	200	80	15.76	78.32
X060	2	350	80	49.44	127.76
X060	3	450	80	67.84	195.60
X060	4	600	80	100.80	296.40
X060	5	1000	80	34.24	330.64
X060	6	2000	80	2.26	332.90

Table S8: Full list of samples included in this analysis. Samples that were sequenced but failed quality-control filters or exhibited hybrid ancestry are not included in this table.

Sample ID	Tissue type	Locality	Ancestry
BZ11-001	leukocyte	Chunga	Kinda
BZ11-002	leukocyte	Chunga	Kinda
BZ11-003	leukocyte	Chunga	Kinda
BZ11-004	leukocyte	Chunga	Kinda
BZ11-005	leukocyte	Chunga	Kinda
BZ11-006	leukocyte	Chunga	Kinda
BZ11-007	leukocyte	Chunga	Kinda
BZ11-008	leukocyte	Chunga	Kinda
BZ11-009	leukocyte	Chunga	Kinda
BZ11-010	leukocyte	Chunga	Kinda
BZ11-011	leukocyte	Chunga	Kinda
BZ11-012	leukocyte	Chunga	Kinda
BZ11-013	FTA blood spot	Chunga	Kinda
BZ11-014	leukocyte	Chunga	Kinda
BZ11-015	leukocyte	Chunga	Kinda
BZ11-016	leukocyte	Chunga	Kinda
BZ11-017	leukocyte	Chunga	Kinda
BZ11-018	leukocyte	Chunga	Kinda
BZ11-019	leukocyte	Chunga	Kinda
BZ11-020	leukocyte	Chunga	Kinda
BZ11-021	FTA blood spot	Chunga	Kinda
BZ11-022	FTA blood spot	Chunga	Kinda
BZ11-023	FTA blood spot	Chunga	Kinda
BZ11-024	leukocyte	Chunga	Kinda
BZ11-025	leukocyte	Chunga	Kinda
BZ11-026	FTA blood spot	Chunga	Kinda
BZ11-028	leukocyte	Chunga	Kinda
BZ11-029	leukocyte	Chunga	Kinda
BZ11-030	leukocyte	Chunga	Kinda
BZ11-031	leukocyte	Chunga	Kinda
BZ11-032	leukocyte	Chunga	Kinda
BZ11-033	leukocyte	Chunga	Kinda
BZ11-034	leukocyte	Chunga	Kinda
BZ11-035	leukocyte	Chunga	Kinda
BZ11-036	leukocyte	Chunga	Kinda
BZ11-037	leukocyte	Chunga	Kinda
BZ11-038	leukocyte	Chunga	Kinda
BZ11-039	leukocyte	Chunga	Kinda
BZ11-040	leukocyte	Chunga	Kinda
BZ11-041	leukocyte	Chunga	Kinda
BZ11-042	leukocyte	Chunga	Kinda
BZ11-043	FTA blood spot	Chunga	Kinda
BZ11-045	leukocyte	Chunga	Kinda
BZ11-046	FTA blood spot	Chunga	Kinda
BZ11-047	leukocyte	Chunga	Kinda
BZ11-048	FTA blood spot	Chunga	Kinda
BZ11-050	leukocyte	Chunga	Kinda
BZ11-051	FTA blood spot	Chunga	Kinda
BZ11-052	FTA blood spot	Chunga	Kinda
BZ11-053	leukocyte	Chunga	Kinda

continued on next page

Table S8 – continued from previous page

Sample ID	Tissue type	Locality	Ancestry
BZ11-054	leukocyte	Chunga	Kinda
BZ11-056	leukocyte	Chunga	Kinda
BZ11-058	leukocyte	Chunga	Kinda
BZ11-059	leukocyte	Chunga	Kinda
BZ11-061	FTA blood spot	Chunga	Kinda
BZ11-062	FTA blood spot	Chunga	Kinda
BZ11-063	FTA blood spot	Chunga	Kinda
BZ11-064	FTA blood spot	Chunga	Kinda
BZ11-065	FTA blood spot	Chunga	Kinda
BZ11-066	FTA blood spot	Chunga	Kinda
BZ11-067	FTA blood spot	Chunga	Kinda
BZ11-068	FTA blood spot	Chunga	Kinda
BZ11-070	FTA blood spot	Chunga	Kinda
BZ11-071	FTA blood spot	Chunga	Kinda
BZ11-072	FTA blood spot	Chunga	Kinda
BZ11-073	FTA blood spot	Chunga	Kinda
BZ11-074	FTA blood spot	Chunga	Kinda
BZ11-075	FTA blood spot	Chunga	Kinda
BZ11-076	FTA blood spot	Chunga	Kinda
BZ07-042	feces	North Nkala Road	grayfoot
BZ12-003	plasma	Ngoma Airstrip	grayfoot
BZ12-006	plasma	Ngoma Airstrip	grayfoot
BZ12-008	plasma	Ngoma Airstrip	grayfoot
BZ12-009	plasma	Ngoma Airstrip	grayfoot
BZ12-030	plasma	Dendro Park	grayfoot
BZ12-031	plasma	Dendro Park	grayfoot
BZ12-032	plasma	Dendro Park	grayfoot
BZ12-033	plasma	Dendro Park	grayfoot
BZ07-029	feces	Nanzhila Plains	grayfoot
BZ07-032	feces	Nanzhila Plains	grayfoot
BZ07-034	feces	Nanzhila Plains	grayfoot
BZ07-004	feces	Choma	grayfoot
BZ07-005	feces	Choma	grayfoot
BZ07-007	feces	Choma	grayfoot
BZ06-218	feces	Lower Zambezi National Park	grayfoot
BZ06-220	feces	Lower Zambezi National Park	grayfoot
BZ06-221	feces	Lower Zambezi National Park	grayfoot
BZ06-225	feces	Lower Zambezi National Park	grayfoot
BZ06-227	feces	Lower Zambezi National Park	grayfoot

Table S9: Full list of genes with significant F_{ST} ($p_{F_{ST}} < 0.05$) prior to correction for multiple comparisons. Asterisks (*) indicate that the gene name was not available in the baboon reference genome annotations and was instead inferred from the macaque reference genome through homology as described in the text.

Gene	F_{ST}	$p_{F_{ST}}$
ABCC6	0.29510	0.03379
ADAM19	0.46159	0.04911
*AFDN**	0.31298	0.01436
AHCTF1	0.41161	0.02608
ALDH7A1	0.32752	0.03378
ALLC	0.31691	0.02241
ALMS1	0.40316	0.00149
ALPK2	0.23213	0.03372
*AQP7**	0.31983	0.02067
ARCN1	0.26019	0.00734
ATP9B	0.26107	0.02414
ATXN2	0.84882	0.00003
BCAS3	0.57289	0.02749
BCL9L	0.90226	0.00153
BMP7	0.25936	0.00059
C16orf62	0.36272	0.00939
*CACNA1D**	0.52348	0.03682
CACNA2D4	0.55962	0.02934
CAPN9	0.27515	0.01603
CD226	0.24492	0.00979
CFAP46	0.39592	0.03196
CHST11	0.32745	0.00345
CIB3	0.30536	0.02744
COL27A1	0.57369	0.00249
COPG2	0.51315	0.03892
*CTRC**	0.58200	<0.00001
DENND6B	0.56703	0.00256
DIS3L2	0.27283	0.02388
DNA2	0.46812	0.04726
DNER	0.27585	0.04829
DPP6	0.20804	0.04012
ECE2	0.83854	0.00330
EDIL3	0.37468	0.04026
EFHD2	0.58200	<0.00001
EHD2	0.39810	0.03011
EP300	0.41123	0.02664
EPB41L4B	0.38656	0.03501
ESRRB	0.24742	0.04475
ETV7	0.39259	0.03271
FAF2	0.71554	0.01160
*FAM149A**	0.24276	0.00791
*FAM210A**	0.73778	0.01023
FAM3D	0.31282	0.02341
*FBRSL1**	0.64186	0.01737
FBXO10	0.28294	0.02814
FCHSD2	0.46819	0.04708
FDXR	0.93147	0.00050
FGF1	0.66769	0.00004
FLII	0.46240	0.04869

continued on next page

Table S9 – continued from previous page

Gene	F_{ST}	$p_{F_{ST}}$
FNTA	0.49563	0.04092
FTO	0.33726	0.02929
GAS6	0.31437	0.00045
GGNBP2	0.22363	0.03609
GLCCI1	0.57015	0.02784
GRAMD1B	0.28781	0.01144
GRM1	0.53937	0.03379
GTF3C1	0.46246	0.04879
H2AFY2	0.22263	0.04506
HSPG2	0.34571	0.00094
HTRA1	0.22455	0.02425
IK	0.79478	0.00539
IP6K1	0.35887	0.04662
IPO9	0.85115	0.00291
IQSEC1	0.38822	0.03422
IQSEC3	0.20883	0.01999
KCNK13	0.49790	0.04120
KDM2A	0.44769	0.01671
KMT2C	0.41598	0.02534
KMT5B	0.44284	0.01807
LAMP3	0.36783	0.00367
LCMT1	0.47434	0.00262
LDLR	0.22747	0.02134
LLGL2	0.47763	0.01119
LPP	0.25330	0.03910
LRP11	0.64726	0.01656
LSM12	0.52516	0.03500
LY96	0.42064	0.00678
MAD1L1	0.26257	0.00400
MED13L	0.29158	0.01028
MED20	0.24371	0.02413
MEGF11	0.19554	0.01331
MOB3B	0.32995	0.00299
MTCL1	0.19511	0.02164
MTO1	0.42405	0.00642
MYL1	0.41734	0.00747
MYO7B	0.39156	0.01150
*MYO9B**	0.58992	0.02486
NAA35	0.22633	0.02278
NCOR2	0.59053	0.00209
NDUFS8	0.28450	0.01752
NMUR1	0.26947	0.03795
NOL10	0.32559	0.00224
NTNG2	0.26701	0.00112
NTRK3	0.29037	0.00293
NUDT7	0.28728	0.03905
NUGGC	0.47754	0.01010
NUP93	0.23881	0.04300
*ODC1**	0.29926	0.00502
ODF2	0.68581	0.01385
PACS1	0.33250	0.00289
PANK2	0.76970	0.00787

continued on next page

Table S9 – continued from previous page

Gene	F_{ST}	$p_{F_{ST}}$
PATJ	0.68590	0.01376
PCDH7	0.71470	0.01159
PCID2	0.97146	0.00033
PDGFRL	0.48296	0.04418
PERP	0.24191	0.03173
PHC3	0.67413	0.01424
PLAUR	0.22857	0.03708
PLEKHA5	0.32369	0.01197
PRDM10	0.62931	0.01903
PRKCE	0.26969	0.00016
PRRC2A	0.33359	0.03057
PTEN	0.78622	0.00603
PUM3	0.78612	0.00612
PYDC1	0.51555	0.03711
RAB11FIP4	0.38858	0.03391
RAD51	0.36802	0.04253
RBFOX1	0.47494	0.01168
RBM33	0.26127	0.04292
RFC5	0.36285	0.04428
RFX4	0.34189	0.00013
RHBDF2	0.36213	0.00897
RPS6KA2	0.21401	0.04906
RUNDC1	0.91873	0.00070
RYR1	0.38968	0.03413
SCO1	0.64062	0.01769
*SDR42E2**	0.47514	0.04549
SETD3	0.40980	0.02626
SLC24A3	0.45728	0.00364
SLC45A1	0.47461	0.04562
SLC6A13	0.35348	0.04959
SPHKAP	0.26428	0.02997
*SRCAP**	0.20734	0.04156
SRGN	0.51940	0.03680
*ST8SIA1**	0.46287	0.04803
SYT9	0.29371	0.00615
TET3	0.50453	0.03953
TF	0.73293	0.01061
TMEFF2	0.31345	0.04249
TMEM178A	0.77110	0.00755
*TMPRSS9**	0.54854	0.03136
TNKS	0.40762	0.00892
*TRBV6-1**	0.30260	0.04972
TRIM62	0.63802	0.01795
TRIM72	0.51555	0.03821
TTLL5	0.55135	0.03075
UBAC1	0.53118	0.03537
UBAC2	0.43741	0.01960
UBR3	0.37741	0.00705
UMODL1	0.33847	0.00020
UQCRC2	0.59024	0.02441
URB1	0.22569	0.01335
VAC14	0.51243	0.00138

continued on next page

Table S9 – continued from previous page

Gene	F_{ST}	$p_{F_{ST}}$
WBSCR17	0.23422	0.03062
WDFY2	0.90447	0.00129
WNT7B	0.23625	0.03621
WRAP53	0.44955	0.01752
XYLT1	0.22870	0.00975
ZBTB3	0.41557	0.02462
ZDHHC14	0.30653	0.00096
ZFPM2	0.26225	0.04439
ZNF3	0.33029	0.00026
ZNF485	0.27073	0.00497
ZNF536	0.25102	0.03087
ZNF564	0.60466	0.02290
ZSCAN25	0.43668	0.01962
ENSPANG00000004311	0.74849	0.00012
ENSPANG00000010457	0.61534	0.00159
ENSPANG00000017858	0.22827	0.04616
ENSPANG00000023891	0.30260	0.04967
ENSPANG00000027984	0.65318	0.00010
ENSPANG00000028076	0.27890	0.04559
ENSPANG00000028262	0.61153	0.00001
ENSPANG00000028279	0.42387	0.02326
ENSPANG00000028627	0.42785	0.00637

Table S10: Genotype frequencies for genes considered to be strong candidates for differentiation (see Table 3.2). Genome positions and reference alleles are given for the anubis baboon (papAnu2) genome. The reference allele for the rhesus macaque (rheMac8) genome was also obtained using a coordinate translation in liftOver (Kent et al., 2002; Rosenbloom et al., 2015). Where the reference alleles differed between genomes, both alleles are listed with the baboon allele listed first. Asterisks (*) indicate that the coordinate could not be converted to rheMac8, and the reference allele from rheMac2 (an older assembly) is shown instead. Genotype frequencies are shown for each species in the order: homozygous for reference, heterozygous, homozygous for alternate.

Gene	Position (papAnu2)		Differentiation		Allelic variants		Genotype frequencies					
			F_{ST}	$p_{F_{ST}}$	Ref.	Alt.	Kinda			grayfoot		
ATXN2	chr11	110620892	0.84882	0.00003	G	A	54	13	0	0	0	15
	chr11	110620934	0.84882		A	G	0	13	54	15	0	0
EFHD2	chr1	17875364	0.61707	< 0.00001	C	G	64	2	0	4	3	2
	chr1	17875397	0.71940		T	C	66	0	0	4	3	2
	chr1	17875411	0.71940		T	G	66	0	0	4	3	2
	chr1	17875416	-0.00265		G/A	A	9	57	0	3	3	3
	chr1	17875436	0.71940		C/T	T	66	0	0	4	3	2
	chr1	17875441	0.71940		C	T	66	0	0	4	3	2
FGF1	chr6	136137340	0.23022	0.00004	T/T*	C	12	25	26	0	1	15
	chr6	136137374	0.87739		T/T*	C	62	1	0	2	3	11
	chr6	136137439	0.89546		T/T*	G	62	1	0	1	4	11
PRKCE	chr13	44855469	0.51137	0.00016	A/G	G	6	25	36	10	0	1
	chr13	44855470	0.51137		C	G	6	25	36	10	0	1
	chr13	44855471	0.51137		A	G	6	25	36	10	0	1
	chr13	44855507	0.51137		G/C	C	6	25	36	10	0	1
	chr13	44898572	0.71643		A	C	66	0	0	5	4	4
	chr13	44898586	-0.01417		C	T	65	2	0	13	0	0
	chr13	44898587	0.52273		G	A	61	6	0	5	4	4
	chr13	44898613	0.51940		G	A	67	0	0	8	3	2
	chr13	44898617	0.51940		G	A	67	0	0	8	3	2
	chr13	44898696	-0.00921		T	A	63	3	1	13	0	0
	chr13	45024148	-0.01925		A	C	57	7	1	12	1	0
	chr13	45024166	-0.01605		G	A	62	2	1	13	0	0
	chr13	45024206	0.71375		A	C	65	0	0	5	4	4
	chr13	45024265	-0.01925		C/T	T	57	7	1	12	1	0
	chr13	45053068	0.17177		C	T	68	0	0	12	0	1
	chr13	45053077	0.08873		C	T	67	1	0	12	0	1
	chr13	45053155	0.17177		G	A	68	0	0	12	0	1
	chr13	45053159	0.17177		G	C	68	0	0	12	0	1
	chr13	45053165	0.11253		T	G	44	20	4	13	0	0
	chr13	45150212	0.00257		C	T	55	4	0	16	0	0
	chr13	45150236	-0.01482		C	T	58	1	0	16	0	0

continued on next page

Table S10 – continued from previous page

Gene	Position (papAnu2)		Differentiation		Allelic variants		Genotype frequencies					
			F_{ST}	$p_{F_{ST}}$	Ref.	Alt.	Kinda			grayfoot		
RFX4	chr11	105851669	0.72300		C	T	2	13	51	11	3	0
	chr11	105851730	-0.01749		A	G	65	1	0	14	0	0
	chr11	105870425	0.68939		C	T	65	0	0	5	4	3
	chr11	105870454	-0.01573		C	T	63	2	0	12	0	0
	chr11	105870471	0.44224		C	A	66	0	0	9	1	2
	chr11	105870476	0.09165	0.00013	T	C	44	21	1	12	0	0
	chr11	105870506	0.44224		T	G	66	0	0	9	1	2
	chr11	105870520	0.44224		C	T	66	0	0	9	1	2
	chr11	105870526	0.44224		A	G	66	0	0	9	1	2
	chr11	105870586	0.44224		C	T	66	0	0	9	1	2
	chr11	105870588	0.07881		C	G	48	14	4	12	0	0
UMODL1	chr3	4512016	0.40211		A/T	G	68	1	0	10	0	3
	chr3	4512024	0.46734		T/A	G	69	0	0	10	0	3
	chr3	4512027	0.46734		A/T	G	69	0	0	10	0	3
	chr3	4512050	0.46734		G/C	C	69	0	0	10	0	3
	chr3	4512065	0.06414		C/A	T	52	15	2	13	0	0
	chr3	4512066	0.01904	0.00020	G/C	A	68	1	0	13	0	0
	chr3	4512075	0.00457		C/A	T	63	6	0	13	0	0
	chr3	4512076	0.46734		G/C	A	69	0	0	10	0	3
	chr3	4512080	0.46734		C/G	T	69	0	0	10	0	3
	chr3	4512092	0.46734		A/T	G	69	0	0	10	0	3
	chr3	4512098	0.46734		T/A	C	69	0	0	10	0	3
ENSPANG-00000004311	chr10	86441991	0.74849	0.00012	G	A	50	16	2	0	1	13
	chr10	86441993	0.74849		T	A	50	16	2	0	1	13
ENSPANG-00000027984	chr13	5031394	0.75948		C	G	58	7	2	2	1	12
	chr13	5031410	0.44058	0.00010	G	A	67	0	0	10	3	2
	chr13	5031434	0.75948		T	A	58	7	2	2	1	12
ENSPANG-00000028262	chr1	185450553	0.19649		A	G	65	1	0	13	0	2
	chr1	185450577	0.86013	0.00001	G	A	66	0	0	3	3	9
	chr1	185450607	0.57834		C	T	52	12	2	3	2	9
	chr1	185450667	0.81117		T	C	66	0	0	4	3	7

Table S11: Gene Ontology (GO) terms with significantly enriched differentiation. Terms that are enriched exhibit an overall shift in either F_{ST} or $p_{F_{ST}}$. Only terms with $p < 0.05$ for both F_{ST} or $p_{F_{ST}}$ are shown here. Key: BP, biological process; CC, cellular component; MF, molecular function.

GO term			Enrichment (p-value)	
Accession	Name	Ontology	F_{ST}	$p_{F_{ST}}$
GO:0000186	activation of MAPKK activity	BP	0.00929	0.01147
GO:0000973	posttranscriptional tethering of RNA polymerase II gene DNA at nuclear periphery	BP	0.04187	0.04509
GO:0001759	organ induction	BP	0.04662	0.04208
GO:0006302	double-strand break repair	BP	0.03340	0.02854
GO:0006939	smooth muscle contraction	BP	0.02450	0.02647
GO:0007608	sensory perception of smell	BP	0.02267	0.02821
GO:0008333	endosome to lysosome transport	BP	0.01818	0.02853
GO:0009408	response to heat	BP	0.02872	0.03863
GO:0009409	response to cold	BP	0.00907	0.01411
GO:0009566	fertilization	BP	0.01657	0.01398
GO:0010001	glial cell differentiation	BP	0.01427	0.01267
GO:0010804	negative regulation of tumor necrosis factor-mediated signaling pathway	BP	0.01357	0.01302
GO:0016486	peptide hormone processing	BP	0.04288	0.04984
GO:0016973	poly(A)+ mRNA export from nucleus	BP	0.04187	0.04509
GO:0021940	positive regulation of cerebellar granule cell precursor proliferation	BP	0.04662	0.04208
GO:0030317	sperm motility	BP	0.02972	0.03697
GO:0030336	negative regulation of cell migration	BP	0.00174	0.00358
GO:0030512	negative regulation of transforming growth factor beta receptor signaling pathway	BP	0.00602	0.00198
GO:0032088	negative regulation of NF-kappaB transcription factor activity	BP	0.03088	0.04750
GO:0032497	detection of lipopolysaccharide	BP	0.02865	0.02418
GO:0033555	multicellular organismal response to stress	BP	0.00848	0.01221
GO:0035019	somatic stem cell population maintenance	BP	0.02792	0.01891
GO:0035264	multicellular organism growth	BP	0.03040	0.02989
GO:0035457	cellular response to interferon-alpha	BP	0.03086	0.02853
GO:0042053	regulation of dopamine metabolic process	BP	0.03934	0.04095
GO:0042991	transcription factor import into nucleus	BP	0.04272	0.04962
GO:0043488	regulation of mRNA stability	BP	0.04187	0.04509
GO:0045444	fat cell differentiation	BP	0.01849	0.02701
GO:0046668	regulation of retinal cell programmed cell death	BP	0.04662	0.04208
GO:0046827	positive regulation of protein export from nucleus	BP	0.04754	0.03285
GO:0046855	inositol phosphate dephosphorylation	BP	0.03938	0.02396
GO:0046856	phosphatidylinositol dephosphorylation	BP	0.04214	0.02782
GO:0048589	developmental growth	BP	0.02438	0.02470
GO:0048678	response to axon injury	BP	0.04662	0.04208
GO:0048741	skeletal muscle fiber development	BP	0.01735	0.01964
GO:0048864	stem cell development	BP	0.04662	0.04208
GO:0051881	regulation of mitochondrial membrane potential	BP	0.03131	0.02769
GO:0051895	negative regulation of focal adhesion assembly	BP	0.02456	0.03393
GO:0051897	positive regulation of protein kinase B signaling	BP	0.04885	0.02395
GO:0060070	canonical Wnt signaling pathway	BP	0.00650	0.00775
GO:0060128	corticotropin hormone secreting cell differentiation	BP	0.04662	0.04208
GO:0060129	thyroid-stimulating hormone-secreting cell differentiation	BP	0.04662	0.04208
GO:0060716	labyrinthine layer blood vessel development	BP	0.04191	0.03962
GO:0071033	nuclear retention of pre-mRNA at the site of transcription	BP	0.04187	0.04509

continued on next page

Table S11 – continued from previous page

GO term			Enrichment (p-value)	
Accession	Name	Ontology	F_{ST}	$p_{F_{ST}}$
GO:0090090	negative regulation of canonical Wnt signaling pathway	BP	0.02020	0.01760
GO:0090267	positive regulation of mitotic cell cycle spindle assembly checkpoint	BP	0.02062	0.02787
GO:0090335	regulation of brown fat cell differentiation	BP	0.04241	0.04866
GO:0090394	negative regulation of excitatory postsynaptic potential	BP	0.01716	0.02062
GO:1902414	protein localization to cell junction	BP	0.02441	0.01283
GO:2000117	negative regulation of cysteine-type endopeptidase activity	BP	0.04187	0.04509
GO:0000932	cytoplasmic mRNA processing body	CC	0.03332	0.04029
GO:0005635	nuclear envelope	CC	0.04635	0.02608
GO:0005643	nuclear pore	CC	0.03716	0.03073
GO:0016235	aggresome	CC	0.04026	0.03311
GO:0016282	eukaryotic 43S preinitiation complex	CC	0.01964	0.03353
GO:0016592	mediator complex	CC	0.03753	0.02428
GO:0030016	myofibril	CC	0.01045	0.01295
GO:0070390	transcription export complex 2	CC	0.04187	0.04509
GO:0001104	RNA polymerase II transcription cofactor activity	MF	0.02570	0.01375
GO:0001786	phosphatidylserine binding	MF	0.01551	0.03024
GO:0001875	lipopolysaccharide receptor activity	MF	0.02865	0.02418
GO:0043027	cysteine-type endopeptidase inhibitor activity involved in apoptotic process	MF	0.04154	0.03363

Table S12: Genes associated with Gene Ontology (GO) terms enriched for differentiation as discussed in the text. Asterisks (*) indicate that the gene name was not available in the baboon reference genome annotations and was instead inferred from the macaque reference genome through homology as described in the text.

GO term		Gene	Enrichment (p-value)	
Accession	Name		F_{ST}	$p_{F_{ST}}$
GO:0035264	multicellular organism growth	*ANKRD11*	0.12089	0.45990
		GNAL	0.16787	0.21680
		HEG1	0.09551	0.55590
		MTOR	0.14410	0.32770
		UNC79	0.29201	0.05873
		WWTR1	0.17157	0.16200
GO:0048589	developmental growth	*CHST11*	0.32745	0.00345
		DMBX1	0.34161	0.05513
		GLI3	0.12795	0.40280
		ZMIZ1	0.14726	0.31800
		*PPP1R13L**	0.21184	0.17750
GO:0048741	skeletal muscle fiber development	*DNER*	0.27585	0.04829
		RYR1	0.38968	0.03413
GO:0045444	fat cell differentiation	*ALMS1*	0.40316	0.00149
		BBS2	0.14209	0.32310
		GRK5	0.10961	0.60620
		PIAS1	0.39690	0.06619
		*CF7L2**	0.17320	0.19580
GO:0090335	regulation of brown fat cell differentiation	*FTO*	0.33726	0.02929
		LEP	0.17948	0.22590
GO:0033555	multicellular organismal response to stress	*LRP11*	0.64726	0.01656
		PTEN	0.78622	0.00603
GO:0009408	response to heat	*LRP11*	0.64726	0.01656
		TRPV2	0.18884	0.19220
GO:0009409	response to cold	*LRP11*	0.64726	0.01656
		SLC27A1	0.29156	0.11970
		ZNF516	0.18343	0.18960
GO:0048864	stem cell development	*FGF1*	0.66769	0.00004
GO:0035019	somatic stem cell population maintenance	*BCL9L*	0.90226	0.00153
		PRDM16	0.16000	0.15060
		ZGLP1	0.17252	0.24480
GO:0001759	organ induction	*FGF1*	0.66769	0.00004
GO:0060128	corticotropin hormone secreting cell differentiation	*FGF1*	0.66769	0.00004
GO:0060129	thyroid-stimulating hormone-secreting cell differentiation	*FGF1*	0.66769	0.00004
GO:0060716	labyrinthine layer blood vessel development	*FBXW8*	0.18538	0.16220
		HS6ST1	0.29516	0.05515
GO:0030317	sperm motility	*TTLL5*	0.55135	0.03075
		*DNAH1**	0.18832	0.17600
GO:0009566	fertilization	*CD226*	0.24492	0.00979
		SPTBN4	0.15902	0.27260
		TTLL5	0.55135	0.03075
GO:0042053	regulation of dopamine metabolic process	*CHRNB2*	0.34932	0.08275
		*PARK2**	0.18016	0.15290

Table S13: Full list of samples included in this analysis. Samples that were sequenced but failed quality-control filters are not included in this table. Hybrid indices and ancestry determinations are based on ADMIXTURE analysis.

Sample ID	Sample type	Locality	Hybrid index	Ancestry
BZ11-001	leukocyte	Chunga	0.0000	Kinda
BZ11-002	leukocyte	Chunga	0.0000	Kinda
BZ11-003	leukocyte	Chunga	0.0000	Kinda
BZ11-004	leukocyte	Chunga	0.0000	Kinda
BZ11-005	leukocyte	Chunga	0.0000	Kinda
BZ11-006	leukocyte	Chunga	0.0000	Kinda
BZ11-007	leukocyte	Chunga	0.0000	Kinda
BZ11-008	leukocyte	Chunga	0.0000	Kinda
BZ11-009	leukocyte	Chunga	0.0000	Kinda
BZ11-010	leukocyte	Chunga	0.0000	Kinda
BZ11-011	leukocyte	Chunga	0.0000	Kinda
BZ11-012	leukocyte	Chunga	0.0000	Kinda
BZ11-013	FTA blood spot	Chunga	0.0000	Kinda
BZ11-014	leukocyte	Chunga	0.0000	Kinda
BZ11-015	leukocyte	Chunga	0.0000	Kinda
BZ11-016	leukocyte	Chunga	0.0000	Kinda
BZ11-017	leukocyte	Chunga	0.0000	Kinda
BZ11-018	leukocyte	Chunga	0.0000	Kinda
BZ11-019	leukocyte	Chunga	0.0000	Kinda
BZ11-020	leukocyte	Chunga	0.0000	Kinda
BZ11-021	FTA blood spot	Chunga	0.0000	Kinda
BZ11-022	FTA blood spot	Chunga	0.0000	Kinda
BZ11-023	FTA blood spot	Chunga	0.0000	Kinda
BZ11-024	leukocyte	Chunga	0.0000	Kinda
BZ11-025	leukocyte	Chunga	0.0000	Kinda
BZ11-026	FTA blood spot	Chunga	0.0000	Kinda
BZ11-028	leukocyte	Chunga	0.0000	Kinda
BZ11-029	leukocyte	Chunga	0.0000	Kinda
BZ11-030	leukocyte	Chunga	0.0000	Kinda
BZ11-031	leukocyte	Chunga	0.0000	Kinda
BZ11-032	leukocyte	Chunga	0.0000	Kinda
BZ11-033	leukocyte	Chunga	0.0000	Kinda
BZ11-034	leukocyte	Chunga	0.0000	Kinda
BZ11-035	leukocyte	Chunga	0.0000	Kinda
BZ11-036	leukocyte	Chunga	0.0000	Kinda
BZ11-037	leukocyte	Chunga	0.0000	Kinda
BZ11-038	leukocyte	Chunga	0.0000	Kinda
BZ11-039	leukocyte	Chunga	0.0000	Kinda
BZ11-040	leukocyte	Chunga	0.0000	Kinda
BZ11-041	leukocyte	Chunga	0.0000	Kinda
BZ11-042	leukocyte	Chunga	0.0000	Kinda
BZ11-043	FTA blood spot	Chunga	0.0000	Kinda
BZ11-045	leukocyte	Chunga	0.0000	Kinda
BZ11-046	FTA blood spot	Chunga	0.0000	Kinda
BZ11-047	leukocyte	Chunga	0.0000	Kinda
BZ11-048	FTA blood spot	Chunga	0.0000	Kinda
BZ11-050	leukocyte	Chunga	0.0000	Kinda
BZ11-051	FTA blood spot	Chunga	0.0000	Kinda
BZ11-052	FTA blood spot	Chunga	0.0000	Kinda

continued on next page

Table S13 – continued from previous page

Sample ID	Sample type	Locality	Hybrid index	Ancestry
BZ11-053	leukocyte	Chunga	0.0000	Kinda
BZ11-054	leukocyte	Chunga	0.0000	Kinda
BZ11-056	leukocyte	Chunga	0.0000	Kinda
BZ11-057	leukocyte	Chunga	0.0508	hybrid
BZ11-058	leukocyte	Chunga	0.0000	Kinda
BZ11-059	leukocyte	Chunga	0.0000	Kinda
BZ11-061	FTA blood spot	Chunga	0.0000	Kinda
BZ11-062	FTA blood spot	Chunga	0.0000	Kinda
BZ11-063	FTA blood spot	Chunga	0.0000	Kinda
BZ11-064	FTA blood spot	Chunga	0.0000	Kinda
BZ11-065	FTA blood spot	Chunga	0.0000	Kinda
BZ11-066	FTA blood spot	Chunga	0.0000	Kinda
BZ11-067	FTA blood spot	Chunga	0.0000	Kinda
BZ11-068	FTA blood spot	Chunga	0.0000	Kinda
BZ11-069	FTA blood spot	Chunga	0.0082	hybrid
BZ11-070	FTA blood spot	Chunga	0.0000	Kinda
BZ11-071	FTA blood spot	Chunga	0.0000	Kinda
BZ11-072	FTA blood spot	Chunga	0.0000	Kinda
BZ11-073	FTA blood spot	Chunga	0.0000	Kinda
BZ11-074	FTA blood spot	Chunga	0.0000	Kinda
BZ11-075	FTA blood spot	Chunga	0.0000	Kinda
BZ11-076	FTA blood spot	Chunga	0.0000	Kinda
Chiou-14-030	feces	Mwengwa Rapids	0.2039	hybrid
Chiou-14-036	feces	Mwengwa Rapids	0.2210	hybrid
Chiou-14-039	feces	Mwengwa Rapids	0.0494	hybrid
Chiou-14-041	feces	Mwengwa Rapids	0.3463	hybrid
Chiou-14-042	feces	Mwengwa Rapids	0.0973	hybrid
Chiou-14-001	feces	Malala Camp	0.5902	hybrid
Chiou-14-005	feces	Malala Camp	0.4964	hybrid
Chiou-14-065	feces	Malala Camp	0.3570	hybrid
Chiou-14-050	feces	Musa Bridge	0.6117	hybrid
Chiou-15-003	feces	Musa Bridge	0.8600	hybrid
Chiou-15-004	feces	Musa Bridge	0.4801	hybrid
Chiou-15-005	feces	Musa Bridge	0.5732	hybrid
Chiou-14-054	feces	Top Musa	0.8969	hybrid
Chiou-14-056	feces	Top Musa	0.5515	hybrid
Chiou-14-057	feces	Top Musa	0.8734	hybrid
Chiou-14-058	feces	Top Musa	0.7126	hybrid
Chiou-14-059	feces	Top Musa	0.8396	hybrid
BZ07-039	feces	North Nkala Road	0.9038	hybrid
BZ07-041	feces	North Nkala Road	0.6236	hybrid
BZ07-042	feces	North Nkala Road	1.0000	grayfoot
BZ07-045	feces	North Nkala Road	0.8396	hybrid
BZ07-047	feces	North Nkala Road	0.6173	hybrid
Chiou-14-069	feces	North Nkala Road	0.6117	hybrid
BZ12-001	plasma	Ngoma Airstrip	0.5408	hybrid
BZ12-002	plasma	Ngoma Airstrip	0.7229	hybrid
BZ12-003	plasma	Ngoma Airstrip	1.0000	grayfoot
BZ12-004	plasma	Ngoma Airstrip	0.9415	hybrid
BZ12-005	plasma	Ngoma Airstrip	0.9498	hybrid
BZ12-006	plasma	Ngoma Airstrip	1.0000	grayfoot
BZ12-007	plasma	Ngoma Airstrip	0.6403	hybrid

continued on next page

Table S13 – continued from previous page

Sample ID	Sample type	Locality	Hybrid index	Ancestry
BZ12-008	plasma	Ngoma Airstrip	1.0000	grayfoot
BZ12-009	plasma	Ngoma Airstrip	1.0000	grayfoot
BZ12-010	plasma	Ngoma Airstrip	0.6125	hybrid
BZ12-011	plasma	Ngoma Airstrip	0.6172	hybrid
BZ12-012	plasma	Ngoma Airstrip	0.9511	hybrid
BZ12-030	plasma	Dendro Park	1.0000	grayfoot
BZ12-031	plasma	Dendro Park	1.0000	grayfoot
BZ12-032	plasma	Dendro Park	1.0000	grayfoot
BZ12-033	plasma	Dendro Park	1.0000	grayfoot
BZ07-029	feces	Nanzhila Plains	1.0000	grayfoot
BZ07-030	feces	Nanzhila Plains	0.3894	hybrid
BZ07-032	feces	Nanzhila Plains	1.0000	grayfoot
BZ07-034	feces	Nanzhila Plains	1.0000	grayfoot
BZ07-035	feces	Nanzhila Plains	0.3541	hybrid
BZ07-004	feces	Choma	1.0000	grayfoot
BZ07-005	feces	Choma	1.0000	grayfoot
BZ07-007	feces	Choma	1.0000	grayfoot
BZ06-218	feces	Lower Zambezi National Park	1.0000	grayfoot
BZ06-220	feces	Lower Zambezi National Park	1.0000	grayfoot
BZ06-221	feces	Lower Zambezi National Park	1.0000	grayfoot
BZ06-225	feces	Lower Zambezi National Park	1.0000	grayfoot
BZ06-227	feces	Lower Zambezi National Park	1.0000	grayfoot

Table S14: Bioclimatic variables used for this analysis.

Variable	Description
Bio01	Annual mean temperature (°C)
Bio02	Mean diurnal temperature range (mean(period max – min)) (°C)
Bio03	Isothermality (Bio02 ÷ Bio07)
Bio04	Temperature seasonality (coefficient of variation)
Bio05	Max temperature of warmest week (°C)
Bio06	Min temperature of coldest week (°C)
Bio07	Temperature annual range (Bio05 – Bio06) (°C)
Bio08	Mean temperature of wettest quarter (°C)
Bio09	Mean temperature of driest quarter (°C)
Bio10	Mean temperature of warmest quarter (°C)
Bio11	Mean temperature of coldest quarter (°C)
Bio12	Annual precipitation (mm)
Bio13	Precipitation of wettest week (mm)
Bio14	Precipitation of driest week (mm)
Bio15	Precipitation seasonality (coefficient of variation)
Bio16	Precipitation of wettest quarter (mm)
Bio17	Precipitation of driest quarter (mm)
Bio18	Precipitation of warmest quarter (mm)
Bio19	Precipitation of coldest quarter (mm)

Table S15: Genes with significantly high BF_{ij}. Asterisks (*) indicate that the gene name was not available in the baboon reference genome annotations and was instead inferred from the macaque reference genome through homology as described in the text.

Gene (i)	Chromosome	Boundaries		Variable (j)	BF_{ij}	$p_{BF_{ij}}$
KDM4A	1	45942958	46000149	PC6	9.18736	0.00100
uncharacterized	1	64367458	64427173	PC5	2.13704	0.00620
CGN	1	125306625	125334403	PC6	8.59800	0.00121
KCNN3	1	128523961	128685063	PC1	3.81540	0.00350
NAV1	1	162148067	162315226	PC6	3.33637	0.00317
KIAA1614	1	183371046	183412218	PC3	1.64620	0.00719
IQSEC1	2	47221072	47393899	PC2	2.23248	0.00539
				PC3	1.71613	0.00615
EMC3	2	50188261	50211494	PC3	4.05550	0.00253
FOXP1	2	63989206	64390504	PC1	19.82744	0.00024
				PC2	35.46995	0.00013
				PC3	24.29732	0.00023
EPHB3	2	100963472	100975588	PC1	4.03667	0.00319
				PC2	7.13465	0.00135
				PC3	2.90270	0.00379
FBXL2	2	173530079	173639117	PC1	4.11000	0.00300
				PC2	3.32080	0.00380
				PC3	2.80810	0.00363
PCBP3	3	811964	904160	PC4	173.17040	0.00002
				PC6	5.53678	0.00167
UMODL1	3	4479719	4541151	PC1	3.10494	0.00434
uncharacterized	3	32076147	32128246	PC5	11.13348	0.00075
				PC6	11.14555	0.00064
*NGRN**	3	34942623	34954844	PC1	2.41438	0.00572
KLHDC10	3	151859502	151988616	PC1	2.36130	0.00587
TMEM209	3	151958927	151999858	PC1	2.36130	0.00594
PTPRN2	3	178675260	179670322	PC5	1.31523	0.00797
WDR60	3	179945922	180029924	PC2	3.15625	0.00350
				PC3	2.27813	0.00492
GMDS	4	1480008	2104188	PC5	2.72516	0.00211
SLC17A5	4	69294582	69357884	PC2	1.57789	0.00793
				PC3	2.21083	0.00529
uncharacterized	4	74131666	74132205	PC1	2.18443	0.00602
SRD5A3	5	73186214	73209888	PC1	1.65070	0.00789
ENC1	6	68719875	68721644	PC1	2.20258	0.00698
				PC2	5.01890	0.00218
				PC3	19.24186	0.00040

continued on next page

Table S15 – continued from previous page

Gene (i)	Chromosome	Boundaries		Variable (j)	BF_{ij}	$p_{BF_{ij}}$
EDIL3	6	77971114	78415215	PC4	3.85125	0.00279
				PC6	27.18721	0.00028
CYFIP2	6	150720429	150851890	PC5	1.78930	0.00650
NDN	7	411412	412377	PC2	1.78767	0.00719
				PC3	4.44374	0.00230
PML	7	48444304	48497153	PC2	3.01375	0.00358
				PC3	3.20430	0.00315
AKAP13	7	59967547	60182646	PC4	2.08170	0.00552
LRRK1	7	75226584	75365243	PC6	2.75220	0.00510
HECTD1	7	87770557	87878098	PC6	1.81482	0.00786
MARK3	7	159589739	159703932	PC1	6.12425	0.00187
				PC2	5.25054	0.00236
				PC4	2.28140	0.00443
				PC6	4.40749	0.00237
PNOC	8	26377294	26387735	PC2	2.14049	0.00576
				PC3	3.51269	0.00308
KIF13B	8	27138761	27328391	PC1	4.92911	0.00283
				PC2	21.97450	0.00053
				PC3	5.99603	0.00163
DKK4	8	40659278	40662492	PC5	6.26166	0.00181
TSNARE1	8	137130058	137254716	PC1	2.36267	0.00553
uncharacterized	9	49531383	49566001	PC4	2.50635	0.00468
				PC6	16.51772	0.00057
ZMYND8	10	16915681	17060143	PC1	6.78099	0.00178
				PC2	2.54971	0.00480
				PC3	1.55352	0.00755
KIF16B	10	49075231	49383323	PC4	1.39805	0.00776
APMAP	10	57545597	57575139	PC6	1.75964	0.00806
EIF4ENIF1	10	72034857	72092546	PC5	14.21900	0.00069
ATXN10	10	85856097	86032210	PC5	1.71000	0.00688
SLC6A13	11	239050	279778	PC1	1.81284	0.00783
				PC2	1.55850	0.00784
				PC3	1.77440	0.00593
FGF6	11	4501152	4512737	PC5	1.98884	0.00679
MYRFL	11	65241954	65334710	PC6	2.10940	0.00725
FBXO21	11	116366356	116417988	PC2	1.97592	0.00638
				PC3	1.59119	0.00712
COQ5	11	119768501	119791403	PC5	2.62056	0.00485
CAMKK2	11	120537234	120593057	PC6	2.66789	0.00414

continued on next page

Table S15 – continued from previous page

Gene (i)	Chromosome	Boundaries		Variable (j)	BF_{ij}	$p_{BF_{ij}}$
SSB	12	31670197	31748320	PC1	6.90752	0.00174
				PC2	5.14832	0.00204
				PC3	1.84901	0.00622
TM4SF20	12	89223352	89240189	PC5	3.94070	0.00320
DGKD	12	95183530	95298392	PC5	2.54481	0.00483
*DRC1**	13	25421094	25476564	PC4	11.19500	0.00054
uncharacterized	13	57229239	57275823	PC1	2.34644	0.00593
				PC2	13.72475	0.00064
				PC3	15.49804	0.00082
INPP4A	13	89866737	89934586	PC6	2.68045	0.00442
*LOC721299**	14	3126889	3154824	PC6	2.78804	0.00445
KCNJ1	14	117794364	117860715	PC5	1.59018	0.00822
				PC6	2.45524	0.00622
PRDM10	14	118801938	118861158	PC6	2.20620	0.00699
*PRRT1**	15	6414964	6418756	PC3	1.43464	0.00783
KIF12	15	20683079	20695132	PC5	2.33596	0.00576
MOB3B	15	47681585	47885267	PC1	9.42180	0.00125
				PC2	2.66670	0.00450
				PC4	2.13520	0.00534
PCSK5	15	82652020	83116524	PC5	1.62777	0.00795
WRAP53	16	7375692	7395868	PC2	2.84078	0.00404
				PC3	5.57029	0.00198
STX8	16	8926934	9256383	PC4	1.14159	0.00616
WIPF2	16	46910048	46975649	PC4	2.29530	0.00448
STAT3	16	48978970	49012645	PC1	16.13338	0.00021
				PC2	8.12002	0.00118
				PC3	4.25684	0.00275
TIMP2	16	70204701	70224681	PC2	2.00434	0.00590
COL4A2	17	87342023	87539696	PC4	1.29616	0.00420
VAV1	19	6525684	6607298	PC3	1.49240	0.00807
*MYO9B**	19	15621123	15738599	PC1	2.85140	0.00478
				PC3	3.14360	0.00328
*CEACAM5**	19	36999858	37057164	PC4	2.01103	0.00312
				PC6	1.64437	0.00715
CCDC61	19	39480480	39499745	PC4	4.84788	0.00182
uncharacterized	19	46689989	46690089	PC5	1.75954	0.00720
uncharacterized	19	46692046	46692130	PC5	1.60410	0.00768
NDE1	20	14587216	14642823	PC6	2.00800	0.00773

continued on next page

Table S15 – continued from previous page

Gene (i)	Chromosome	Boundaries		Variable (j)	BF_{ij}	$p_{BF_{ij}}$
CES5A	20	38402501	38436739	PC4	3.18773	0.00136
				PC6	3.76712	0.00223
CDH11	20	47152411	47324537	PC4	2.16438	0.00494
WDR59	20	56574087	56689100	PC1	2.56140	0.00537
CENPN	20	62768678	62789675	PC1	2.65119	0.00540
*PKD1L2**	20	62862425	62970721	PC1	2.05316	0.00756
ZC3H18	20	70286072	70347821	PC5	1.80139	0.00651

Table S16: Genes with excess α_i. Asterisks (*) indicate that the gene name was not available in the baboon reference genome annotations and was instead inferred from the macaque reference genome through homology as described in the text.

Gene (i)	Chromosome	Boundaries		Mean (α_i)	ETP interval (α_i)	
		High α_i				
RGS12	1	555231	707923	1.53853	0.12780	3.13951
SLC45A1	1	11154262	11183152	1.61807	0.41679	2.45595
SLC25A34	1	18194591	18200737	1.20260	0.10914	2.28004
LDLRAD2	1	24044439	24052967	0.90149	0.05009	1.78411
LAPTM5	1	32963659	32988931	0.77023	0.03108	1.45115
C1orf109	1	39951245	39966524	1.78544	0.60010	3.21693
TESK2	1	47583596	47696639	1.41475	0.03193	2.83091
PHC3	2	115535587	115621735	1.01787	0.16558	2.19253
SLC37A3	3	162125892	162193524	1.61862	0.42885	2.89215
NFKBIE	4	43632831	43641209	2.02282	0.85927	3.49348
uncharacterized	4	74131666	74132205	0.84749	0.04108	1.61821
AKIRIN2	4	83037303	83064962	2.08446	0.89284	3.57656
uncharacterized	5	1127274	1400939	0.58709	0.08970	1.05584
SEMA5A	6	8812682	9007710	1.20747	0.12999	2.19952
DHX29	6	51293228	51348456	1.43984	0.24456	2.91424
IK	6	134169964	134181145	1.45201	0.24013	2.71444
LCTL	7	40880562	40900235	1.25034	0.26853	2.47289
KIF23	7	43868409	43905208	1.99501	0.81166	3.50669
FAM149B1	9	56429241	56507955	1.38599	0.14251	2.65677
PLPP4	9	112062913	112196960	1.44887	0.25589	2.70259
CFAP46	9	124314486	124435692	1.03731	0.09627	1.88325
SLC24A3	10	52238911	52441012	0.89863	0.19719	1.68283
XRN2	10	54011481	54100124	1.17762	0.21637	2.07172
KIAA1644	10	84493086	84562612	1.96598	0.52424	3.30072
XPC	11	82092158	82129435	1.04051	0.05027	2.19713
COQ5	11	119768501	119791403	0.95241	0.07774	1.95525
TMEM132B	11	124695755	124989612	1.40371	0.20318	2.89844
STAM2	12	14362981	14415522	1.75917	0.23931	2.84203
*AGAP1**	12	97496718	97711301	0.59133	0.06243	1.14762
*UBE2F**	12	99766986	99843407	2.09450	0.80998	3.60063
ALLC	13	3065128	3116554	0.80901	0.21460	1.59779
uncharacterized	13	24314237	24346739	1.59115	0.36145	3.07258
C2orf78	13	72653999	72665448	1.22736	0.31994	2.18866
TGFBRAP1	13	96592596	96634535	1.13548	0.28426	2.06213
KDM2A	14	6836879	6978648	0.86676	0.03077	1.82306
INCENP	14	11575470	11598846	1.09960	0.03423	2.18011
BCL9L	14	108226812	108247635	0.95260	0.09708	2.21301
PRDM10	14	118801938	118861158	1.32657	0.31555	2.39229
NACC2	15	2212006	2260850	1.24021	0.18861	2.16712
ODF2	15	9550178	9597858	1.35574	0.20521	2.15592
GOLM1	15	92537623	92608307	1.00344	0.01040	2.23289
TRPV2	16	15731378	15751538	1.75441	0.80284	2.94720
PIGL	16	15853278	15962381	1.36898	0.40052	2.56130
LATS2	17	140678	231436	1.73844	0.53363	2.95804
RNF138	18	24060644	24097045	1.04932	0.03257	2.33077
ZNF236	18	68967234	69085587	1.05171	0.01722	2.17271
*TMPRSS9**	19	2175904	2187829	1.59976	0.31975	2.50440
ZNF564	19	11317666	11340206	1.01027	0.02432	2.13787

continued on next page

Table S16 – continued from previous page

Gene (i)	Chromosome	Boundaries		Mean (α_i)	ETP interval (α_i)	
ABCC6	20	15088676	15163648	0.75447	0.08758	1.41665
IQCK	20	17906074	18044544	1.44633	0.31753	2.48762
PLCG2	20	63514137	63700068	1.22891	0.39000	1.98859
CBFA2T3	20	70602695	70700524	0.65815	0.09181	1.26826
		Low α_i				
AMPH	3	72517333	72771775	-0.72320	-1.41438	-0.05225
KMT2E	3	127167177	127242094	-1.11702	-2.08411	-0.16206
LY96	8	69788797	69883096	-0.47783	-1.02987	-0.02982

Table S17: Genes with excess β_i.

Gene (i)	Chromosome	Boundaries		Mean (β_i)	ETP interval (β_i)	
		High β_i				
LIMK1	3	45157337	45194630	1.89539	0.08140	4.03588
LY96	8	69788797	69883096	1.21470	0.01039	2.65116
AACS	11	124402598	124484755	1.60177	0.02148	3.38445
TMEFF2	12	53718921	53960589	1.85727	0.49071	3.47751
TMEM178A	13	38564332	38615864	1.79232	0.11404	3.94489
		Low β_i				
ODF2	15	9550178	9597858	-2.04875	-4.57410	-0.01293

Table S18: Gene Ontology (GO) terms with significantly enriched α_i or β_i cline parameters. Only terms with $p < 0.025$ for any of four one-tailed enrichment tests (high α_i, low α_i, high β_i, low β_i) are shown here. Key: BP, biological process; CC, cellular component; MF, molecular function; +, test for high parameter value; −, test for low parameter value.

GO term			Enrichment		
Accession	Name	Ontology	Parameter	+/−	p-value
GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	BP	α_i	+	0.02335
GO:0006672	ceramide metabolic process	BP	α_i	+	0.02312
GO:0009395	phospholipid catabolic process	BP	α_i	+	0.01799
GO:0045095	keratin filament	CC	α_i	+	0.02375
GO:0005242	inward rectifier potassium channel activity	MF	α_i	+	0.01883
GO:0008536	Ran GTPase binding	MF	α_i	+	0.02211
GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	MF	α_i	+	0.01260
GO:0016829	lyase activity	MF	α_i	+	0.01438
GO:0006351	transcription, DNA-templated	BP	α_i	−	0.00198
GO:0006355	regulation of transcription, DNA-templated	BP	α_i	−	0.00067
GO:0006470	protein dephosphorylation	BP	α_i	−	0.01700
GO:0006935	chemotaxis	BP	α_i	−	0.00812
GO:0016051	carbohydrate biosynthetic process	BP	α_i	−	0.01855
GO:0019752	carboxylic acid metabolic process	BP	α_i	−	0.02383
GO:0045010	actin nucleation	BP	α_i	−	0.01290
GO:0005267	potassium channel activity	MF	α_i	−	0.01219
GO:0016817	hydrolase activity, acting on acid anhydrides	MF	α_i	−	0.01139
GO:0016831	carboxy-lyase activity	MF	α_i	−	0.02383
GO:0000724	double-strand break repair via homologous recombination	BP	β_i	+	0.01885
GO:0003333	amino acid transmembrane transport	BP	β_i	+	0.01639
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	BP	β_i	+	0.00705
GO:0006865	amino acid transport	BP	β_i	+	0.01890
GO:0008284	positive regulation of cell proliferation	BP	β_i	+	0.02478
GO:0010508	positive regulation of autophagy	BP	β_i	+	0.01824
GO:0045995	regulation of embryonic development	BP	β_i	+	0.01893
GO:0046034	ATP metabolic process	BP	β_i	+	0.01448
GO:0048500	signal recognition particle	CC	β_i	+	0.00761
GO:0000166	nucleotide binding	MF	β_i	+	0.02182
GO:0003697	single-stranded DNA binding	MF	β_i	+	0.02009
GO:0003700	transcription factor activity, sequence-specific DNA binding	MF	β_i	+	0.01733
GO:0004181	metallocarboxypeptidase activity	MF	β_i	+	0.02323
GO:0008312	7S RNA binding	MF	β_i	+	0.00705
GO:0015171	amino acid transmembrane transporter activity	MF	β_i	+	0.01639
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	MF	β_i	+	0.01431
GO:0019789	SUMO transferase activity	MF	β_i	+	0.00906
GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	BP	β_i	−	0.02079
GO:0006836	neurotransmitter transport	BP	β_i	−	0.00595
GO:0006915	apoptotic process	BP	β_i	−	0.00861
GO:0051014	actin filament severing	BP	β_i	−	0.02296
GO:0030126	COPI vesicle coat	CC	β_i	−	0.00989
GO:0004003	ATP-dependent DNA helicase activity	MF	β_i	−	0.01647

continued on next page

Table S18 – continued from previous page

GO term			Enrichment		
Accession	Name	Ontology	Parameter	+/−	p-value
GO:0004252	serine-type endopeptidase activity	MF	β_i	–	0.00820
GO:0005242	inward rectifier potassium channel activity	MF	β_i	–	0.01331
GO:0005328	neurotransmitter:sodium symporter activity	MF	β_i	–	0.00595

Appendix C

Supplementary Protocol

FecalSeq enrichment protocol

Portions of this protocol are modified from the NEBNext Microbiome DNA Enrichment Kit manual (New England Biolabs cat. #E2612S)

Materials and reagents

- Extracted fecal-derived DNA of known quantity
- NEBNext Microbiome DNA Enrichment Kit (New England Biolabs; cat. #E2612S or #E2612L)
- Rotating mixer
- Magnetic rack for 1.5/2.0 ml microcentrifuge tubes
- 5 M NaCl

Before beginning

1. Extract and prepare DNA samples

   While any fecal DNA (fDNA) extraction method should in principle be compatible with the MBD enrichment, methods that maximize the recovery of host DNA are preferable. Bead-beating methods that increase total DNA yield from feces, for example, should be avoided because the mechanical disruption increases the yield of cell-wall-bound DNA (i.e., from bacteria or plants) while fragmenting host DNA.

   We suggest aiming for a total yield of 1 μg of DNA for all samples in a maximum volume of 30 μl each, although we have had success with as little as 500 ng (the yield of host DNA is likely more important than the yield of total fDNA). If the volume is greater than 30 μl, the DNA can be concentrated via a bead cleanup (Auxiliary protocol A).

   Prior to enrichment, DNA should be quantified for the total yield (e.g., by fluorometer or spectrophotometer). Ideally, the host DNA should also be quantified by qPCR (Auxiliary protocol B).

2. Calculate the required volume of MBD2-Fc-bound magnetic beads (hereafter referred to as "MBD beads") for each enrichment reaction, as well as the total volume for a set of reactions as follows.

   As an approximate rule, prepare 1 µl of MBD beads for every 6.25 ng of target host DNA in each enrichment reaction. If samples contain less than 6.25 ng of host DNA or if the amount of host DNA is not quantified, prepare 1 µl of MBD beads.

   We recommend preparing batches of MBD beads (see step 5) with a minimum volume of 40 µl, as lower volumes preclude adequate mixing. If a smaller volume is needed, leftover unused MBD beads can be stored at 4 °C for up to a week.

3. Resuspend protein A magnetic beads by gently pipetting the mixture up and down until the suspension is homogenous, or by slowly rotating the mixture at 4 °C for 15 minutes. **Do not vortex**.

4. Prepare 1X bind/wash buffer by diluting 1 part 5X bind/wash buffer with 4 parts DNase-free water. As a general rule, the volume of 1X bind/wash buffer needed can be calculated as:

   2.5 ml + 1.2 ml × [number of enrichment reactions]

   The amount of 1X bind/wash buffer depends on the total volume of MBD beads and the total number of enrichment reactions. MBD beads can be prepared with a maximum volume of 160 µl in a single reaction. As very small volumes (1 – 8 µl) of beads are needed for our enrichment method, a single bead preparation reaction is nearly always sufficient. If more beads are needed, increase the number of bead preparation reactions and adjust the volume of 1X bind/wash buffer accordingly. Alternatively, for volumes up to 320 µl, prepare an additional 1 ml of 1X bind/wash buffer per bead preparation reaction and add an extra wash step (see step 14).

   2.5 ml of 1X bind/wash buffer are required for a single bead preparation reaction up to 160 µl. Prepare an additional 1.2 ml of 1X bind/wash buffer per enrichment reaction. This number takes into account the volume needed to prepare 2 M NaCl elution buffer in the following step.

   Keep 1X bind/wash buffer on ice throughout the MBD bead preparation. For the wash steps following the capture reaction, 1X bind/wash buffer can be at room temperature.

5. Prepare 2 M NaCl elution buffer by diluting 5 M NaCl with 1X bind/wash buffer. 100 µl of 2 M NaCl elution buffer are needed per enrichment reaction.

   1X bind/wash buffer has a NaCl concentration of 150 mM. 1 ml of 2 M NaCl elution buffer can be prepared by adding 370 µl of 5 M NaCl with 630 µl of 1X bind/wash buffer.

Preparing MBD beads

6. If preparing 40 µl of MBD beads, add 4 µl of MBD2-Fc protein to 40 µl of protein A magnetic beads in a 1.5 ml microcentrifuge tube. For preparing other volumes (n µl) of MBD beads, add $n/10$ µl MBD2-Fc protein to n µl of protein A magnetic beads.

As a rule, we do not prepare less than 40 µl of MBD beads due to diminished efficiency of both rotational mixing and magnetic separation at low volumes.

7. Mix the bead-protein mixture by rotating the tube in a rotating mixer for 10 minutes at room temperature.

8. Briefly spin the tube and place on the magnetic rack for 2 – 5 minutes until the beads have collected to the wall of the tube and the solution is clear.

9. Carefully remove and discard the supernatant with a pipette without disturbing the beads.

10. Add 1 ml of 1X bind/wash buffer (kept on ice) to the tube to wash the beads. Pipette up and down a few times to mix.

11. Mix the beads by rotating the tube in a rotating mixer for 3 minutes at room temperature.

12. Briefly spin the tube and place on the magnetic rack for 2 – 5 minutes until the beads have collected to the wall of the tube and the solution is clear.

13. Carefully remove and discard the supernatant with a pipette without disturbing the beads.

14. Repeat steps 10 – 13.

    If preparing between 160 µl and 320 µl of beads, repeat steps 10 – 13 twice for a total of three washes to ensure the removal of unbound MBD2-Fc protein.

15. Remove the tube from the rack and add n µl (determined in step 6) of 1X bind/wash buffer to resuspend the beads. Mix by pipetting the mixture up and down until the suspension is homogenous.

Capture methylated host DNA

Since reaction volumes are well under 100 µl, multiple enrichment reactions can be processed together in a microplate, with pipetting steps conducted using a multichannel pipettor. Compatible rotating mixers and magnetic separators would also be required. Here, we proceed to describe the capture procedure using a 1.5 ml tube.

The total volume of the capture reaction is an important consideration. We have observed decreased DNA binding efficiency when the concentration of MBD beads or DNA in the capture reaction is low. We therefore recommend maintaining a total reaction volume of approximately 40 µl, as we have experienced consistent success with this volume even when adding as little as 1 µl of MBD beads. Decreasing the reaction volume may result in decreased efficacy of rotational mixing. It is a good idea to keep the volume of all reactions consistent as this facilitates processing of many samples and, if DNA amounts and bead volumes are kept consistent, serves as a control for the effects of bead or DNA concentration on enrichment efficiency. Our subsequent procedures assume a reaction volume of 40 µl (not including MBD beads). If using other reaction volumes, pay particular attention to notes following each step in this section.

16. Aliquot 8 µl of 5X bind/wash buffer to a 1.5 ml microcentrifuge tube

    For reaction volumes other than 40 µl, tune the volume of 5X bind/wash buffer to maintain 1X concentration and adjust accordingly the volume of DNase-free water added in step 17. The volume of MBD beads should be excluded from this calculation as prepared MBD beads are already at 1X concentration.

    We recommend equilibrating 5X bind/wash buffer to room temperature prior to aliquoting for more accurate pipetting.

17. Add up to 30 µl of DNA (prepared in step 1) to the tube. Bring the total volume to 40 µl with DNase-free water.

    For reaction volumes other than 40 µl, adjust the volume of DNase-free water added to reach the target volume. Be sure to maintain 1X bind/wash concentration.

18. Add MBD beads to the tube using the volume determined in step 2. Pipette the mixture up and down or swirl a few times to mix.

    As an approximate rule and as stated above, add 1 µl of MBD beads for every 6.25 ng of target host DNA in each enrichment reaction. If samples contain less than 6.25 ng of host DNA or if the amount of host DNA is not quantified, add 1 µl of MBD beads.

19. Incubate the reaction for 15 minutes at room temperature with rotation.

20. Following incubation at room temperature, briefly spin the tube and place on the magnetic rack for 5 minutes until the beads have collected to the wall and the solution is clear.

21. Carefully remove the supernatant with a pipette without disturbing the beads. The supernatant is enriched for microbial DNA and may be saved and purified by bead cleanup (Auxiliary protocol A). Otherwise, discard the supernatant.

22. Add 1 ml of 1 bind/wash buffer (kept at room temperature) to wash the beads.

    If processing in a microplate, decrease the volume of wash buffer to 100 µl.

23. Carefully remove and discard the wash buffer with a pipette without disturbing the beads.

24. *Optional.* Add 100 µl of 1X bind/wash buffer (kept at room temperature) to the beads. Pipette the mixture up and down a few times to mix.

    We have found that an additional wash with 100 µl of 1X bind/wash buffer followed by rotation (steps 24 – 27) substantially improved enrichment. To skip this wash, proceed to step 28.

25. Mix the beads by rotating the tube in a rotating mixer for 3 minutes at room temperature.

26. Briefly spin the tube and place on the magnetic rack for 2 – 5 minutes until the beads have collected to the wall of the tube and the solution is clear.

27. Carefully remove and discard the supernatant with a pipette without disturbing the beads.

Eluting captured host DNA

The NEBNext Microbiome Enrichment Kit includes an elution protocol for captured DNA that includes digestion of DNA-bound MBD beads with proteinase K and elution with TE buffer. We have found that elution with 2 M NaCl is just as effective, is less time consuming, and conserves proteinase K. Most importantly, we have found that DNA samples eluted with 2 M NaCl and purified by bead cleanup can be further enriched in a repeat enrichment reaction. DNA samples eluted with proteinase K and TE buffer and purified by bead cleanup in contrast produced miniscule yields following a repeat enrichment reaction.

28. Add 100 µl of 2 M NaCl (prepared in step 5 and kept at room temperature) to the beads. Pipette the mixture up and down a few times to mix.

    If large numbers of samples are being processed, considering lowering the elution volume such that the combined volume of DNA and SPRI beads (see Auxiliary protocol A; step 1) does not exceed the capacity of microplate wells and thereby preclude the ability to parallelize bead cleanups.

29. Mix the beads by rotating the tube in a rotating mixer for 3 minutes at room temperature.

30. Briefly spin the tube and place on the magnetic rack for 2 – 5 minutes until the beads have collected to the wall of the tube and the solution is clear.

31. Carefully remove the supernatant to a fresh microcentrifuge tube and discard beads.

32. Proceed to bead cleanup to purify sample (Auxiliary protocol A).

Auxiliary protocols

Auxiliary protocol A: Bead cleanup

Portions of this protocol are modified from Pacific Biosciences protocol # 001-252-177-03.

Materials and reagents

- Pre-washed magnetic SPRI beads, prepared following Rohland & Reich (2012)
- 70% ethanol, freshly prepared
- 1X TE buffer
- Magnetic stand
- Centrifuge

Procedures

1. Add 1.5X – 1.8X volume of pre-washed magnetic beads to DNA in a 1.5 ml tube.

   If the combined volume of beads and DNA does not exceed the capacity of the tube or well, large numbers of bead cleanups can be conducted in parallel on a microplate.

2. Mix the bead/DNA solution thoroughly by pipetting up and down several times.

3. Vortex the beads for 5 minutes.

4. Briefly spin the tube and place on the magnetic rack for 5 minutes or until the solution is clear.

5. Carefully remove and discard the supernatant without disturbing the beads.

6. Wash beads with freshly prepared 70% ethanol. Wait 1 minute, then pipette and discard the ethanol.

   Use a sufficient volume of 70% ethanol to completely cover the bead pellet (e.g., 100 µl for microplates and 400 µl for 1.5 ml tubes). Slowly dispense the 70% ethanol against the side of the tube opposite the beads. Do not disturb the bead pellet.

7. Repeat step 6 above.

8. Remove residual 70% ethanol and air-dry the bead pellet for 1 minute.

   Spin at full speed for 2 minutes in order to collect residual 70% ethanol. Then place on the magnetic rack for 30 seconds before pipetting the residual 70% ethanol and air-drying for 1 minute.

9. Resuspend the beads in 30 – 40 µl of 1X TE buffer or another suitable DNA stabilization buffer.

10. Vortex for 1 minute, then incubate for 2 minutes. Spin the sample at full speed to pellet beads. Return to the magnet and collect the supernatant in a new 1.5 ml microcentrifuge tube.

11. Following bead cleanup, quantify with a fluorometer or spectrophotometer. Validate enrichment by qPCR (Auxiliary protocol B). Enriched DNA can be sequentially enriched by repeating the enrichment protocol adding 30 µl of the enriched product to the FecalSeq enrichment protocol: step 17.

Auxiliary protocol B: qPCR estimation of enrichment

Materials and reagents

- Extracted fDNA of known quantity
- 2X SYBR Green master mix (e.g., Qiagen cat. #204143 or ThermoFisher Scientific cat. #A25780)
- Taxon-specific primers
- DNA standards
  For host quantification, standards can be created by performing a dilution series (i.e., 10 ng/µl, 1 ng/µl, 0.1 ng/µl, 0.01 ng/µl) of high-quality gDNA (such as blood or liver DNA) from a suitable taxon.
- qPCR instrument

Procedures

1. Run samples and standards at least in duplicate. We also recommend running a positive and negative control with each set of quantifications.

2. Use primers specific to the analysis

    a. The proportion of host DNA can be quantified by comparing qPCR results using host-specific primers to the absolute quantification estimated by some independent means (e.g., fluorometer or spectrophotometer). For our baboon DNA quantifications, we use universal mammal primers for the *MYCBP* gene (Morin et al., 2001):

    b. Enrichment of DNA captured with MBD beads can be quantified as above using host-specific primers with enriched methylated host DNA. Alternatively, enrichment can be estimated by observing the *n*-fold decrease in quantified levels from unenriched to enriched samples using the universal 16s rRNA primer (Corless et al., 2000). 1 µl of unenriched DNA can be diluted to the concentration of the enriched sample prior to qPCR to standardize concentrations. Because MBD enrichment can in principle be biased towards densely methylated areas of the host genome, we prefer the latter method for estimating enrichment success.

Primer ID	Type	Locus	Sequence	Reference
cmycF cmycR	mammalian	*MYCBP*	`GCCAGAGGAGGAACGAGCT` `GGGCCTTTTCATTGTTTTCCA`	Morin et al., 2001
16s_F 16s_R	bacterial	16S rRNA	`CCATGAAGTCGGAATCGCTAG` `GCTTGACGGGCGGTGT`	Corless et al., 2000

3. Set up qPCR reactions in a 20 µl total volume containing 1X of SYBR Green master mix, 0.5 mM of each primer, and 1 µl of DNA.

4. Run samples in the qPCR instrument at 95 °C for 15 minutes, followed by 50 cycles of 94 °C for 15 seconds, 59 °C (for all primers specified above; adjust for other primers) for 25 seconds, and 72 °C for 20 seconds.

References

Corless, C. E., Guiver, M., Borrow, R., Edwards-Jones, V., Kaczmarski, E. B., & Fox, A. J. (2000). Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. *Journal of Clinical Microbiology*, *38*, 1747–1752.

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

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*, 939–946.

Appendix D

Supplementary Note

In its advertised use, the Microbiome DNA Enrichment Kit (New England Biolabs cat. #E2612S) contains enough reagents to enrich six samples, assuming 160 µl of protein A beads and 16 µl of MBD2-Fc protein are used per sample.

For FecalSeq, each reaction can be scaled down significantly. Assuming that fecal DNA samples on average contain 2.5% host DNA, we estimate that each reaction will require on average 4 µl of protein A beads and 0.4 µl of MBD2-Fc protein. This represents a scaling-down by a factor of 40. Therefore, a Microbiome DNA Enrichment Kit contains enough protein A beads and MBD2-Fc protein to support a total of 240 (6×40) enrichments.

240 enrichments at \$168 / kit (university rate) = \$0.70 per enrichment.